THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Conformational Dynamics of the Trp-Cage Miniprotein at Its Folding Temperature

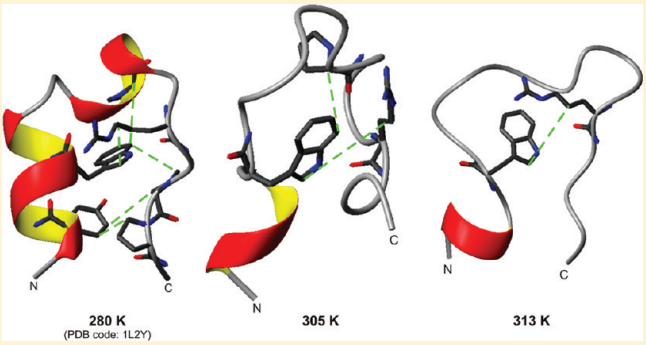
Anna Hałabis,[†] Wioletta Żmudzińska,[†] Adam Liwo,[‡] and Stanisław Ołdziej*[,†]

[†]Laboratory of Biopolymer Structure, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, Kładki 24, 80-922 Gdańsk, Poland

[‡]Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland

Supporting Information

ABSTRACT: The folding temperature of the trp-cage miniprotein was determined to be in the range 311−317 K depending on the method used. Our study is focused on determining the structure and dynamics of the polypeptide chain close to its unfolding or melting temperature. At T = 305 K, Trp6−Arg16 and Trp6−Pro12 long-range interactions are observed, and at T = 313 K, only the Trp6−Arg16 interactions remain, while all of mentioned interactions are observed in the native state of the protein. Partial (at T = 305 K) and complete (at T = 313 K) melting of the N-terminal α-helix is observed, manifested by the appearance of minor sets of signals in NMR spectra. Our key findings are: (*i*) conformational phase transition (melting point) could be described as a cooperative breaking of the Trp6−Pro12 long-range hydrophobic interaction and the melting of the N-terminal α-helix; (ii) many ROE signals corresponding to local or short-range interactions vanish rapidly with temperature increase; however, long-range interaction such as Trp6−Arg16 remains until 313 K. The presence of the native long-range interaction at 313 K makes that conformational ensemble resemble a very diffuse native state structure, but it is not a simple mixture of the folded and unfolded states, as could be expected on the basis of the common two-state folding mechanism.



INTRODUCTION

Since Anfinsen[1] in his experiment demonstrated that a protein can fold spontaneously into a unique functional three-dimensional structure (the native state), researchers have tried to understand details of the protein-folding process. One of the methods for understanding how the protein folds is to design amino acid sequences not observed in nature, which could self-organize into a unique three-dimensional structure. In 1997, Dahiyat and Mayo redesigned the natural zinc-finger motif (which is stabilized by a zinc ion in natural proteins) to engineer a small 28-residue mini-protein termed FSD (first sequence designed).[2] Using the same computational technology, several variants of the FSD sequence were designed; all of these designed sequences share the same three-dimensional structure but differ in the thermodynamics stability. Their experimentally measured stabilities correlated well with those predicted during the computational sequence design.[3] The successful design of the FSD mini-protein pushed forward further research in the direction of redesigning the existing sequences, or structural motifs to create artificial proteins. In 2002, Neidigh and co-workers, by truncation and mutation of a naturally occurring 39-residue peptide, engineered a 20-residue mini-protein named tryptophan-cage (trp-cage).[4]

Because of its small size (20 residues), compact structure, sequence consisting of only naturally occurring amino acids, no disulfide-bridges, and no tendency to oligomerization, the trp-cage became an attractive object for the study of the mechanism and kinetics of protein folding. The three-dimensional structure of the trp-cage was determined by using NMR spectroscopy at a temperature of T = 280 K, and it is shown in Figure 1.[4] The structure of this miniprotein is composed of a short N-terminal α-helix (residues 2−9), a turn structure (residues 10−16), and a poliproline II structure at the C-terminus (residues 17−20). The structure is stabilized by long-range hydrophobic interactions between the aromatic Tyr3 and Trp6 side chains with the side chains of Pro12, Pro18, and Pro19. Moreover, some theoretical studies emphasize the importance of the Asp9−Arg16[5,6] salt-bridge in the overall structure stabilization; however, experimental results do not clearly confirm these theoretical predictions.[7,8]

The spectroscopic data from NMR and CD experiments[4,9,10] suggested that trp-cage folds/unfolds by a simple two-state mechanism.[11] On the other hand, the folding time determined by using tryptophan fluorescence is about 4 μs, which make trp-

Special Issue: Harold A. Scheraga Festschrift

Received: December 31, 2011
Revised: April 12, 2012
Published: April 12, 2012

   dx.doi.org/10.1021/jp212630y | J. Phys. Chem. B 2012, 116, 6898−6907

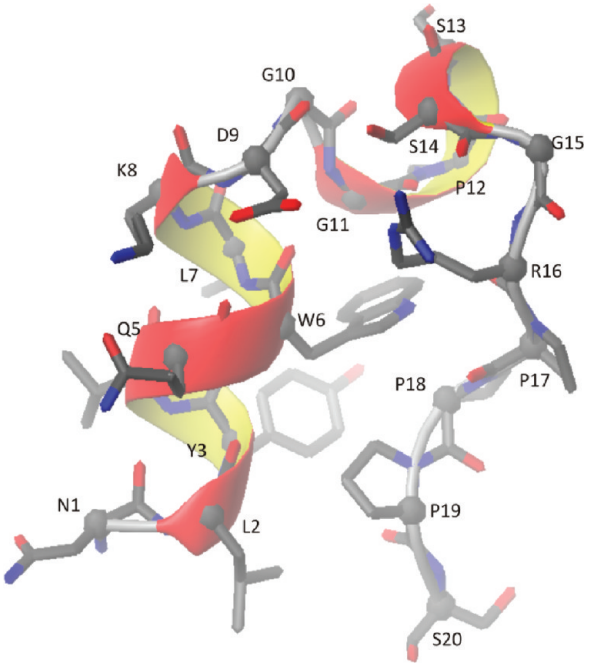


Figure 1. Three-dimensional structure of the trp-cage miniprotein.[4]

cage the fastest folding protein[12] and strongly suggests a downhill folding mechanism.[11] However, many theoretical studies rather predict a slower folding time of 14 μs[13] or even 100 μs[14] and this supports a two-state folding mechanism. Some experimental results suggest that the folding mechanism of the trp-cage protein is even more complex and involves intermediates in the folding pathway.[15] It is extremely intriguing that even for such a small model system like trp-cage protein we are not able to find a consensus folding/unfolding mechanism.

Recently, in our laboratory, we performed a series of NMR experiments in a wide range of temperatures (T = 283−323 K) and we proposed a very detailed picture of the mechanism of the folding of β-hairpin-forming peptides of a length from 8 up to 34 residues.[16−20] At selected temperatures, we recorded the NMR spectra, and on the basis of the restrains derived from the ROESY (NOESY) spectra, we determined the structure, as well as the conformational dynamics, of these peptides at a given temperature. We are aware that, especially for flexible peptides with dynamic structure, the measurements cannot provide individual distinct conformations but only an average picture of the conformational ensemble at a given temperature. The reason for this is the long mixing time used in the NMR experiment. Nevertheless, we are able to characterize the dynamics of long-range interactions responsible for structure stabilization very precisely. We can also determine how the occurrence of some long-range interactions between hydrophobic side chains is related to the folding transition.[16−20] Our experiments show that clean, interpretable NMR spectra without signal broadening induced by increased system dynamic and proton exchange could be registered up to 313 K (at higher temperatures, fast amide proton exchange makes NMR spectra not interpretable).[18,19] The trp-cage system is perfectly suited for performing a detailed analysis of the thermal folding/unfolding process by applying the methodology described above. The folding temperatures reported for trp-cage in the literature are T = 311 K,[7] T = 313 K,[15] and T = 315 K[4] with a mean value of T = 313 K, which is the highest temperature at which NMR experiment could be reasonably performed and the interproton distances for trp-cage miniprotein might be obtained (see above). The structure of the trp-cage was determined at T = 280 K;[4] some NMR data related to the structure at T = 300 K are also known from the literature.[7] Therefore, in the present study, we recorded the NMR spectra at T = 305 K and T = 313 K, where the second temperature value is close to the folding temperature. The structural information obtained at temperatures close/equal to the folding temperature is extremely valuable from the point of view of understanding the trp-cage protein folding mechanism.

MATERIALS AND METHODS

Peptide Synthesis. The linear peptide was prepared on a Milipore synthesizer employing standard solid-phase Fmoc-amino acid chemistry. The resin used for the synthesis (Tenta Gel R RAM resin, 1 g, capacity 0.18 mmol/g) was treated with piperidine (20%) in DMF. All amino acids were coupled by using the DIPCI/HOBt methodology. The coupling reaction time was 1 h. Piperidine (20%) in DMF was used to remove the Fmoc group at all steps. After deprotection of the last Fmoc N-terminal group, the resin was washed with methanol and dried *in vacuo*. The peptide was cleaved from the resin using 88/5/5/2 TFA/water/phenol/triisopropylsilane mixture in 10 mL/g of the resin at room temperature for 2 h. Then, the resin was separated from the mother liquid, excess solvent was evaporated, and, finally, the peptide was precipitated with diethyl ether. The crude trp-cage was dissolved in pure water and purified by reverse phase HPLC C18 prep-scale column using gradients of water/acetonitrile (having 0.1% and 0.085% TFA, respectively). The collected fractions were lyophilized, and their identities and molecular weight were established from an analytical RP-HPLC and MALDI-TOF/ESI LC MS IT TOF analysis.

^{1}H NMR Spectroscopy. The NMR spectra of trp-cage at 305 and 313 K were measured on a VARIAN 500 MHz spectrometer. The following 2D H NMR spectra were recorded: DQF-COSY,[21] TOCSY[22] (80 ms), and ROESY[23] (200 ms) at each analyzed temperature. The samples were dissolved in H_2O/D_2O (9:1 by vol) (pH ∼5.5), and the concentration of each of the samples was 2.3 mM. The spectra were processed by using the VARIAN 4.3 software and analyzed with the XEASY[24] program. The spectra were calibrated against DSS (sodium 4,4-dimethyl-4-silapentane-1-sulfonate) signal.[25] Proton signals were assigned on the basis of the TOCSY spectra. We tried to use the NOESY spectra for structure determination. The NOESY spectra were used in the previous works related to the trp-cage protein;[4,7,8] however, at the elevated temperatures used in this work, we observed a rapid decrease of NOE signal intensities, regardless of the mixing time used in NOESY experiments (50, 100, 150, 200, 300, and 350 ms). The 20-residue trp-cage mini-protein is on the borderline of the applicability of the NOESY spectroscopy, which does not work for medium-size compounds[26] and, additionally, the increase of temperature probably leads to a change of rotational correlation time in such manner that it leads to reducing the NOE effects observed in the NOESY spectra. Therefore, in our study, we used the ROESY spectra. The sequential analysis of the trp-cage was confirmed by the ROESY spectra.[23] The coupling constants between HN and H_α protons ($^3J_{HN/H\alpha}$) of trp-cage were obtained from two-dimensional DQF-COSY and one-dimensional ^{1}H NMR spectra. The intensities of ROE signals were estimated from

dx.doi.org/10.1021/jp212630y | J. Phys. Chem. B 2012, 116, 6898−6907

the ROESY spectra. The ROE interproton cross-peaks of trp-cage at 305 and 313 K were derived from 2D ^{1}H NMR ROESY. In the first step, the ROESY peak volumes were converted to upper distance bounds by using CALIBA[27] of the DYANA package.[28] In the next step, the torsion angles were generated by using the HABAS[29] algorithm of the DYANA package,[28] based on the Bystrov–Karplus[30] equation. The upper distance and torsional-angle limits were used as restraints in molecular dynamics calculations.

Three-Dimensional Structure Calculations. Molecular dynamics simulations were carried out with the AMBER force field by using the AMBER 9.0 package[31] with amber-99 force-field.[32] All simulations were performed in a periodic box of TIP3P water[33] at constant volume with the particle-mesh Ewald procedure[34,35] for long-range electrostatic interactions. The interproton distances and the dihedral angles were restrained with the force constants k = 20 kcal/(mol × Å^2) and k = 2 kcal/(mol × rad^2), respectively. The dihedral angles ω were restrained with a center at 180° and k = 10 kcal/(mol × rad^2). The improper dihedral angles centered at the C_α atoms (defining the chirality of amino acid residues) were restrained with k = 50 kcal/(mol × rad^2). The time-averaged restraint method (TAV)[31,36,37] was used to include experimental values for the calculations, with interproton-distance restraints (calculated from the intensities of the ROE signals) and dihedral-angle restraints (calculated from the coupling constants). The numbers of distance restraints were 171 and 105 for the trp-cage at T = 305 K and T = 313 K, respectively. Two sets of separate simulations, using the restraints from the NMR data collected at T = 305 K and T = 313 K, were run for trp-cage. MD simulations with time-averaged restraints, at these two different temperatures, were carried out with a time step of 2 fs, and the total duration of the run was 4 ns. For each temperature (T = 305 K and T = 313 K), we carried out two separate TAV MD simulations with NMR restrains: in the first MD simulation, the starting structure contained the Asp9–Arg16 salt-bridge, while in the second one the structure did not contain the Asp9–Arg16 salt-bridge. The coordinates were saved every 2000 steps of MD simulations. For every NMR restraint set, four independent TAV MD simulations were run at the following temperatures: T = N, 400, 500, and 600 K (where N is the temperature at which the NMR experiment was performed, 305 or 313 K) for two independent sets of calculations. Owing to running simulations at many temperatures, we provide an extensive sampling of the conformational space. For every trajectory, 300 final snapshots were collected for further analysis. The structures from four trajectories, obtained from simulations performed using the same NMR restraint set, were combined together. After TAV MD simulations, we obtained two sets of 1200 conformations (four runs, with 300 conformations from every run) corresponding to two NMR restraint sets recorded at 305 and 313 K. Two sets of conformations were clustered separately, with the use of the MOLMOL[38] program (Figures 5 and 6). A C_α–RMS deviation cutoff of 5.0 Å was used in the clustering procedure. The same procedure was used for the analysis of much more flexible peptides studied before.[16−19] The clustering procedure provided 12 (where the starting structure involved the Asp9–Arg16 salt-bridge), 7 (where the starting structure did not involve the Asp9–Arg16 salt bridge), and 10 (where the starting structure involved the Asp9–Arg16 salt-bridge), 10 (where the starting structure did not involve the Asp9–Arg16 salt-bridge) families of conformations for simulations which used NMR data recorded at 305 and 313 K, respectively.

RESULTS AND DISCUSSION

NMR Measurements. *Key Interactions at Folding Transition Temperature.* The chemical shifts of the proton resonances for the trp-cage at 305 and 313 K are listed in Tables I and II in the Supporting Information. The rotating frame Overhauser enhancement (ROE) effects corresponding to the interproton distances and the $^3J_{HN/H\alpha}$ coupling constants of trp-cage at 305 and 313 K are shown in Figure 2, and long-range interactions are summarized in Table 1.

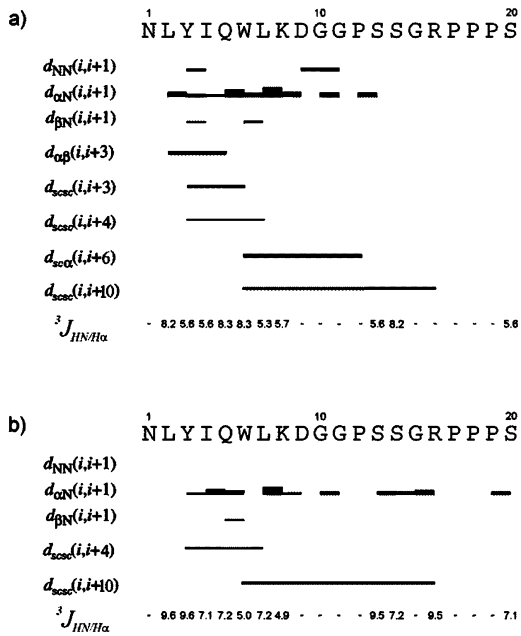


Figure 2. ROE effects corresponding to the interproton distances and the $^3J_{HN/H\alpha}$ coupling constants of trp-cage at (a) T = 305 K and (b) T = 313 K. The thickness of the bars reflects the strength of the ROE correlation as strong, medium, or weak.

Table 1. Atoms of Residues at Separation $|i - j| > 1$ between Which the ROE Peaks Were Found at 305 and 313 K for the trp-Cage Miniprotein

ROE peaks between residues $\|i - j\| > 1$	
305 K	313 K
αL2 – βQ5	
δY3 – ε_3W6	εY3 – δ_1L7
εY3 – δ_1L7	δ_1W6 – γR16
δ_1W6 – γR16	
ζ_2W6 – αP12	

We identified 171 and 105 ROE signals for spectra registered at 305 and 313 K, respectively. For comparison, 169 and 152 NOE signals were identified at 280[4] and 300 K,[7] respectively. As seen, the number of signals used to derive geometrical restrains decreases with temperature, but this is not a straightforward relationship. However, it should be kept in mind that we are considering all the observed Overhauser effects which are measured by two different techniques ROESY (this work) and NOESY[4,7] acquired under slightly different conditions (pH, concentration) in the described experiments. However, when we look at the number of $|i - j| > 1$ (where i

dx.doi.org/10.1021/jp212630y | J. Phys. Chem. B 2012, 116, 6898–6907

and j are the residue numbers) interactions which define structure, we have 64 such signals at 280 K,[4] 32 at 300 K,[7] and 5 and 2, respectively, at 305 and 313 K (see Table 1). A dramatic decrease in the number of medium- and long-range interactions clearly shows that the mobility of the trp-cage protein increases with temperature, but interestingly, some long-range native interactions are presented at elevated temperatures of 305 and 313 K (see Table 1).

In our previous studies of the β-hairpin forming peptides,[16−20] owing to recording NMR at exactly or near melting point temperature, we are able to identify key interactions, the disruption of which defines the transition between the folded and unfolded state. In all the β-hairpin-forming peptides studied, those key interactions were hydrophobic interactions between aromatic side chains. As mentioned in the Introduction, the structure of the trp-cage protein at T = 280 K[4] is stabilized by the following two clusters of long-range interactions: (a) Tyr3 with Pro19 and Pro18; (b) Trp6 with Pro12, Arg16, and Pro18. Except for the Trp6−Arg16 interaction, all of the other interactions mentioned above are purely hydrophobic. With temperature increase to T = 300 K[7] and further (see Table 1), one cluster of hydrophobic interactions associated with the Tyr3 side chain completely vanishes. Temperature increase to T = 300 K and above leads to the decrease of the number of observed interactions involving the Trp6 aromatic side chain (see Table 1). If occurrence of the melting temperature is associated with breaking the long-range hydrophobic interactions, as our previous studies suggested,[16−20] then the only one long-range ($|i - j| > 4$) hydrophobic interaction observed at 305 K is Trp6−Pro12 and it is not observed at 313 K (see Table 1). One possibility is that the breaking/formation of the Trp6−Pro12 interaction could be responsible for the phase transition in the thermal folding/unfolding of the trp-cage protein. The Trp6−Arg16 long-range interaction is present at all temperatures from T = 280 K up to T = 313 K[4,7] (see Table 1); however, this interaction is not a hydrophobic one but it is the interaction between the positively charged guanidine group and the π-electrons of the aromatic indole ring.

Because there is some disagreement regarding the melting point temperature determination and the reported melting temperatures vary from T = 311 K to T = 317 K,[4,7,15,39] our results could be interpreted in the following way: if the folding temperature is closer to 311 K, then it could be concluded that the Trp6−Pro12 hydrophobic interaction is the key interaction, whose stability determines the transition temperature and the presence of the Trp6−Arg16 interaction rather reflects some additional effects. The presence of the Trp6−Arg16 interactions could reflect the observation that the Trp6 indole ring at T = 313 K is still in some sort of a "cage" formed by a peptide chain. We did not observe any long-range ROE contacts (except Trp6−Arg16) at 313 K; however, it is still clearly visible that the protons of some residues are in close contact with the indole ring. The chemical shift of the H_α proton of the Pro18 residue is 2.66 ppm at 280 K[4] and 3.79 and 3.84 ppm at T = 305 and T = 313 K, respectively (see the Supporting Information). For comparison, the chemical shifts of the H_α proton of residue Pro19 are 4.37, 4.27, and 4.33 ppm at T = 280,[4] 305, and 313 K, respectively. The downshifting of the chemical shift of the H_α proton of residue Pro18 is a clear indication of the influence of the aromatic ring. Another clear influence of aromatic electrons is stereoselective downshifting of one of the H_α protons of the Gly11 residue. The differences between the chemical shifts of the H_α protons of the Gly11 residue are 2.12, 0.58, and 0.73 ppm at T = 280, 305, and 313 K, respectively, while the expected values are in the range 0.1−0.0 ppm. For both Pro18 and Gly11 residues, the influence of the indole ring on chemical shifts decreases with temperature increase; however, it is still substantial at T = 313 K. Moreover, Ahmed and co-workers,[15] using UV resonance Raman spectroscopy, demonstrated that the tryptophan side chain is well shielded from the solvent and even at T = 343 K the shielding is higher than at T = 280 K. The formation of the "cage" around the indole moiety by polypeptide chain could bring the Trp6 and Arg16 residues close in space without too much implication on the structure, and only moderate implication on the structure stability. Mutation of the Arg16 residue to lysine slightly increases the melting temperature (by about 4−6°), and Arg16Orn (ornithine) mutation does not change the melting temperature.[8] Recently, Rovo and co-workers reported a single point mutation of the Arg16 to homo-Arg (homoarginie−arginine side-chain is elongated by one $-CH_2-$ unit) and alanine. The melting temperatures for those constructs are not determined, but it was shown that these mutations do not affect the structure of the constructs and that the Trp6 residue is always in contact with residue 16, regardless of residue type.[40]

Taking into account the fact that a polypeptide chain could wrap around the side chain of the Trp6 residue at T = 313 K and that the Pro12 residue is a part of this wrapping (see the paragraph above), we cannot expect that breaking the Trp6−Pro12 hydrophobic interaction could change the thermodynamic properties of the investigated system dramatically. If the polypeptide chain is wrapped around the indole moiety, then breaking the Trp6−Pro12 interaction only slightly changes the indole environment. For the β-hairpin forming peptides which are similar in size,[16−20] breaking the hydrophobic interactions leads to a substantial change of the heat capacity of the system at the folding temperature [the heat capacity change is measured by differential scanning calorimetry (DSC)]. However, the heat-capacity peak related to the folding transition of the trp-cage is rather weak and rises only about 30% above the thermal fluctuation at that temperature.[39] Small changes in the heat capacity fit well to small rearrangements of the indole surrounding (breaking the Trp6−Pro12 interaction with polypeptide chain still wrapped around tryptophan side-chain) without substantial changes of the protein solvation shell. The UV resonance Raman spectra also revealed that at temperatures above T = 293 K the distance between the Trp6 side chain and the proline imide bond decreases to some extent with temperature increase above T = 215 K.[15] Specifically, the Pro12 residue was indicated as the one which interacts with Trp6 more tightly as temperature increases (see Figure 8 in ref 15). Additionally, some theoretical studies[5] supported by experimental data[41] emphasize the importance of the Trp6−Pro12 interaction in the folding/unfolding transition. We gathered the experimental evidence that the Trp6−Pro12 interaction is probably involved in the phase-transition process; however, we were unable to find such evidence for the importance of the Trp6−Arg16 interaction. We could only assume that if the Trp6−Arg16 interaction is related to the folding transition then breaking of this interaction cannot substantially change shielding of the indole ring against solvent.[15]

Stability of the N-Terminal α-Helix. For trp-cage at T = 305 K and T = 313 K, the $H_\alpha(i) - H_N(i + 1)$ sequential ROEs are

dx.doi.org/10.1021/jp212630y | J. Phys. Chem. B 2012, 116, 6898−6907

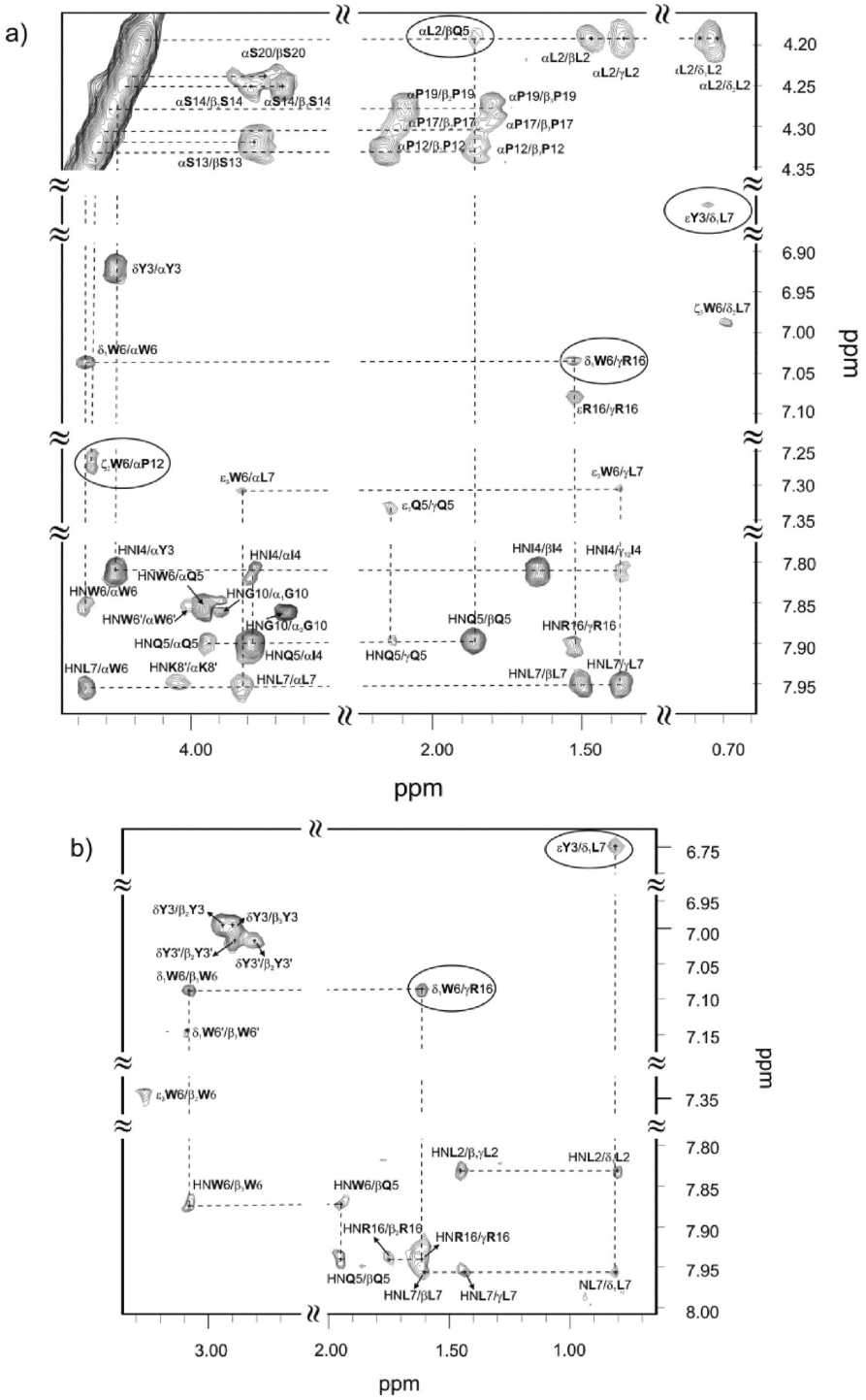


Figure 3. Fragments of the ROESY spectra of trp-cage recorded at (a) T = 305 K and (b) T = 313 K. ROE signals corresponding to long-range interactions listed in Table 1 are elipsed.

almost always observed but their strength is variable (see Figure 2). We also observed quite strong $H_N(i) - H_N(i + 1)$ sequential ROEs between Tyr3/Ile4, Asp9/Gly10, and Gly10/Gly11 at T = 305 K, i.e., mainly in the N-terminal part of the molecule. No such ROE peaks are observed in the spectra recorded at T = 313 K. The $H_\beta(i) - H_N(i + 1)$ connectivities are much weaker but are also found between Tyr3/Ile4 and Trp6/Leu7 at T = 305 K and between Gln5/Trp6 at T = 313 K. Long-range-distance ROE connectivities are also observed; these occur for the $H_\alpha(i) - H_\beta(i + 3)$, $H_{sc}(i) - H_{sc}(i + 3)$, $H_{sc}(i) - H_{sc}(i + 4)$, $H_{sc}(i) - H_\alpha(i + 6)$, and $H_{sc}(i) - H_{sc}(i + 10)$ proton pairs at T = 305 K and for the $H_{sc}(i) - H_{sc}(i + 4)$ and $H_{sc}(i) - H_{sc}(i + 10)$ proton pairs at T = 313 K [where the subscript "sc" denotes side-chain proton(s)], respectively (Figures 2 and 3, Table 1). As is clearly seen from Figure 2a, a regular pattern of ROE connectivities characteristic of α-helix is observed for residues 2−7 at T = 305 K; such a pattern is not observed at T = 313 K (see Figure 2b). The data presented in Figure 2 support the previous finding from NMR studies[7] that the N-terminal α-helix melts with temperature increase. Our

dx.doi.org/10.1021/jp212630y | J. Phys. Chem. B 2012, 116, 6898−6907

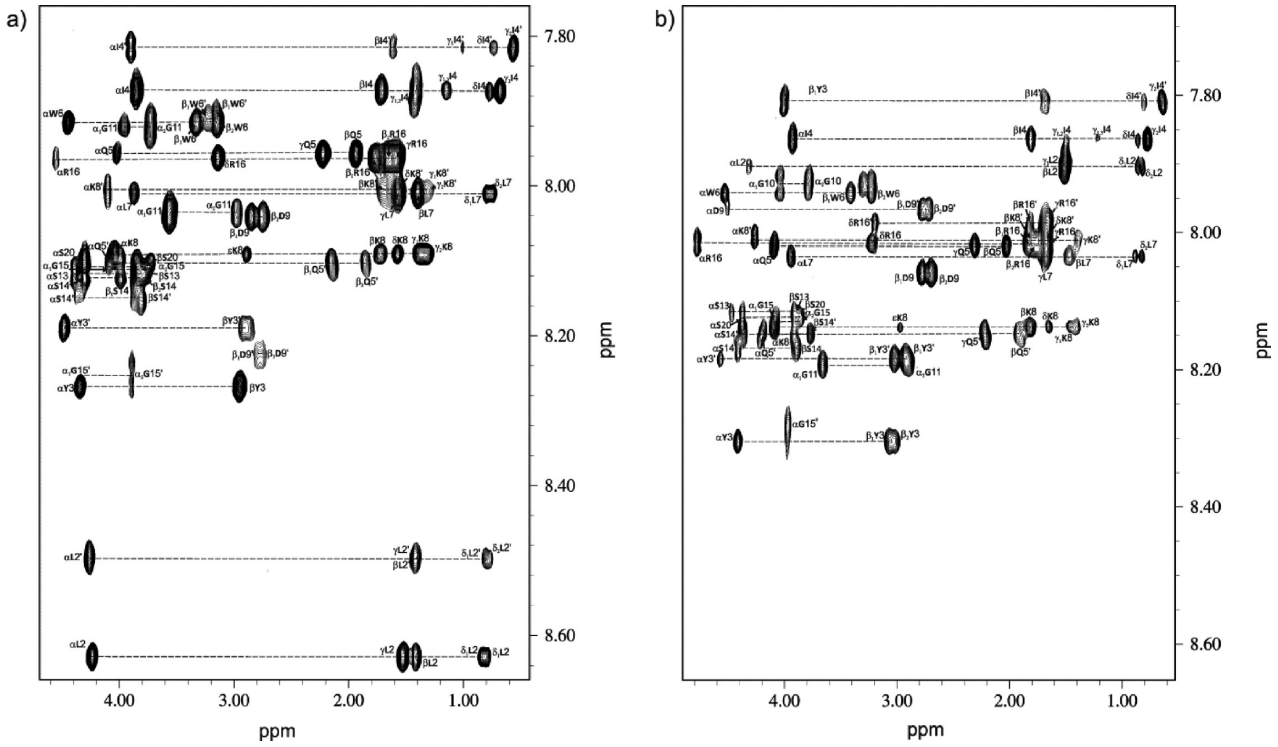


Figure 4. Amino acid spin systems in the TOCSY spectra of trp-cage in H_2O (t_m = 80 ms; the diagnostic region) at (a) T = 305 K and (b) T = 313 K. Black and gray signals correspond to the major and minor set of signals, respectively (′ denotes minor conformation in the proton assignment).

data clearly indicate that, at T = 313 K, the N-terminal α-helix is completely melted. With temperature increase, the overall helicity of the trp-cage decreases, which is clearly seen at CD spectra[4,7,41] and in the Raman UV resonance spectra.[15] It should be noted that the melting temperature of the trp-cage was determined by using the CD and Raman-UV resonance spectra by observation of the overall helicity of the investigated system.[4,7,15] The importance of the N-terminal α-helix stability in the stability of the whole trp-cage protein was investigated by Barua and co-workers[41] who concluded that the stabilization or destabilization of the helical part is strongly positively correlated with the thermal stability of the whole protein. An interesting finding is that the complete melting of the N-terminal α-helix (as it is seen in Figure 2) is correlated with the disappearance of the Trp6−Pro12 interaction, as we described in the previous section. Therefore, unlike peptide systems studied previously,[16−20] the conformational-folding transition is not determined only by the breaking/formation of the one key interaction, but as more complicated event involves the breaking/formation of key long-range interactions coupled with the melting/formation of some secondary structure elements.

Conformational Ensemble at the Folding-Transition Temperature Implication on Folding Mechanism. In the TOCSY spectra of trp-cage at T = 305 K and T = 313 K (Figure 4), we observed a minor set of signals for residues Asn1 (only at T = 305 K), Leu2, Tyr3, Ile4, Gln5, Trp6, Lys8, Asp9, Ser14 (only at T = 305 K), Gly15, and Arg16. The minor amino-acid spin systems for each residue were usually much weaker than the major ones, yet they are clearly visible (see Figure 4). For the minor sets of signals, we observed only (i) − (i + 1) ROE connectivities and none of minor signals give long-range ROE effects. We did not observe signals which could be interpreted as a trace of cis−trans isomerization of the X-Pro peptide bonds. The presence of such isomerization was detected at low temperatures for some variants of trp-cage protein.[8]

Recently, Rovo and co-workers[40] showed the appearance of two sets of signals at T = 280 K for some variants of the trp-cage sequence. A major and a minor set of signals were observed for residues Leu2, Tyr3, Ile4, Gln5, Leu7, Gly10, and Ser14; the minor set of signals did not indicate any long-range interproton signals. The appearance of the minor set of signals was interpreted as a manifestation of a possible alternative fold of trp-cage which appeared in the low temperature. However, for the original trp-cage sequence,[4] the appearance of any minor conformation was not reported before.

The appearance of two sets of signals at elevated temperatures could be interpreted in different ways. One way of interpretation is that, according to the two-state folding theory, a mixture of folded and unfolded states of the protein chain should be observed.[42] Major sets of signals could correspond to the folded state because we observed signals related to long-range interaction, while the minor set of signals can be ascribed to the unfolded state (lack of long-range interactions). According to the two-state folding theory, the overall concentration of the unfolded state should increase with increasing temperature. To assess the fraction of the major and minor conformation, we analyzed the volume of selected signals in the TOCSY spectra. For analysis, we selected only such signals which do not overlap one with another at both temperatures, to reduce the possibility of errors in the volume measurement. The data related to the peak volume are summarized in Table 2. We found that, according to some NMR signals, the fraction of the major and minor conformation is almost unaffected by temperature change (HN/Hα 3, HN/Hα 5, Hε/δ 16); for other signals, we observed an increase of the fraction of the minor conformation upon temperature

dx.doi.org/10.1021/jp212630y | J. Phys. Chem. B 2012, 116, 6898−6907

Table 2. Relative Peak Volumes for Selected Signals in TOCSY Spectra for the Minor and Major Conformation of trp-Cage Protein Recorded at T = 305 and 313 K

	relative peak volume (in %) (′ denotes minor conformation)	
peak name	305 K	313 K
HN/Hα 3	80	81
HN/Hα 3′	20	19
HN/Hα 4	74	58
HN/Hα 4′	26	42
HN/Hα 5	67	66
HN/Hα 5′	33	34
HN/Hβ_2 6; Hδ1/Hε_1 6	70	52
HN/Hβ_2 6′; Hδ1/Hε_1 6′	30	48
Hε/δ 16	67	66
Hε/δ 16′	33	34

increase (HN/Hα 4, HN/Hβ_2 6, Hδ1/Hε_1 6). The data shown in Table 2 do not support the two-state folding mechanism for the trp-cage protein. Although we observed two sets of signals in the NMR spectra and those sets could be associated with the folded and unfolded state, respectively, the temperature dependence of the fractions of these two alternative conformational states is different from that expected on the basis of the two-state theory.

Lack of long-range interproton interactions in the minor set of signals is the only argument for the interpretation of those signals as corresponding to the unfolded state (see paragraph above). The Rovo and co-workers[40] interpretation could be used that minor sets of signals correspond to some alternative fold of trp-cage which appeared at elevated temperatures. A small concentration of those species (alternative fold) causes the long-range interproton interactions to not show up. If we assume that the major and minor sets of signals correspond to the two alternative conformations, then the simple two-state model does not work anymore for trp-cage folding/unfolding, that leading to a very complicated folding mechanism in which the folded state unfolds following two distinct pathways. On the other hand, the chemical shifts of the minor and major signal sets are not substantially different, so we cannot expect that a possible alternate fold will be very different, especially keeping in mind that the indole ring should be shielded from the solvent.

The third possible explanation for appearance of two sets of signals is probably the most plausible one. The major and minor sets of signals appear only for some residues (those of the N-terminal α-helix and some residues from the turn region), so the minor and major conformations share some common residues which display only one set of signals. Interestingly, at T = 305 K, the Trp6 residue displays two sets of signals; however, part of the aromatic ring shows only one set of chemical shifts (ξ and η protons; see Table 1 in the Supporting Information) and we observed long-range ROE effects between ξ_2Trp6 and αPro12 (see Table 1). Because Pro12 displays only one set of signals, this could provide some evidence that the minor and major sets of signals correspond to two substantially distinct conformational ensembles, which share a common structural element which is long-range Trp6–Pro12 interaction. We could suppose that, at T = 305 K, the conformational ensemble includes two distinct families of conformations which share a common structural element (long-

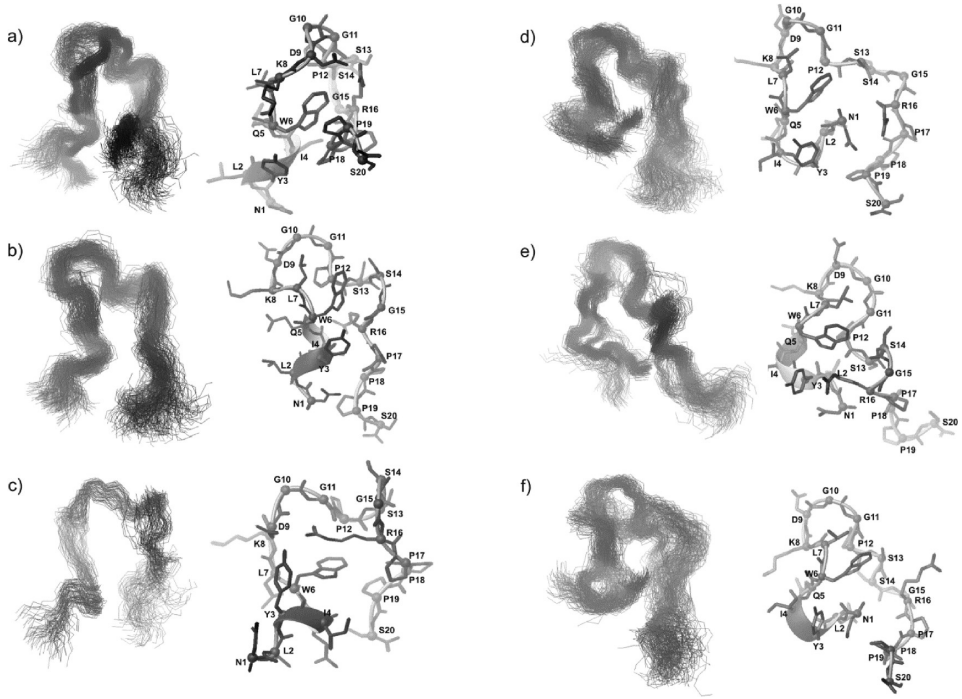


Figure 5. Three major families of conformations of trp-cage obtained by time-averaged MD methodology and restrains from NMR measurements at T = 305 K. The left column shows all conformations from a family (only backbones are shown for clarity), and the right column shows the lowest energy conformation from the corresponding family (all heavy atoms are shown). 1200 conformations were subjected to a cluster analysis, leading to the following relative population numbers and percentages of each clustered family: (a) 456 (38,0%), (b) 292 (24,3%), (c) 86 (7,2%), (d) 331 (27,6%), (e) 299 (24,9%), (f) 254 (21,2%) [(a, b, c) the MD starting structure had an Asp9–Arg16 salt-bridge; (d, e, f) the MD starting structure did not have an Asp9–Arg16 salt bridge].

dx.doi.org/10.1021/jp212630y | J. Phys. Chem. B 2012, 116, 6898–6907

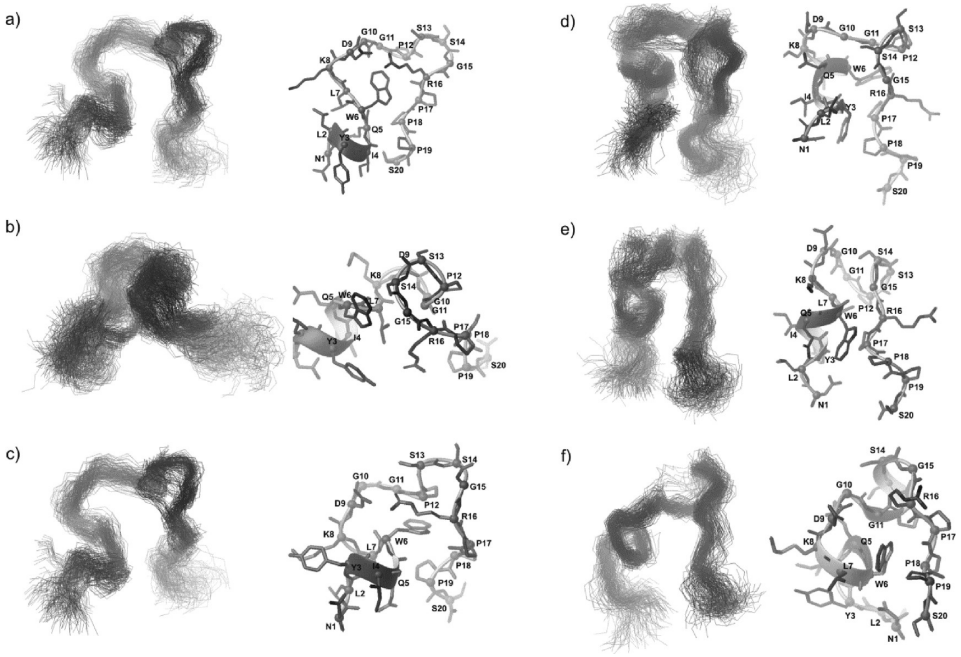


Figure 6. Three major families of conformations of trp-cage obtained by using time-averaged MD methodology and restrains from NMR measurements at T = 313 K. The left column shows all conformations from a family (only backbones are shown for clarity), and the right column shows the lowest energy conformation from the corresponding family (all heavy atoms are shown). 1200 conformations were subjected to a cluster analysis, leading to the following relative population numbers and percentages of each clustered family: (a) 390 (32,5%), (b) 319 (26,6%), (c) 187 (15,6%), (d) 413 (34,4%), (e) 229 (19,1%), (f) 188 (15,7%) [(a, b, c) the MD starting structure had the Asp9−Arg16 salt-bridge; (d, e, f) the MD starting structure did not have the Asp9−Arg16 salt-bridge].

range Trp6−Pro12 interaction) related to the native state, so both families of conformations could be called the "native folded" at this temperature. The presented data emphasizes the importance of the Trp6−Pro12 interaction as a descriptor of the folding/unfolding transition (see section "Key Interactions at Folding Transition Temperature"). At T = 313 K, the Trp6−Pro12 interaction is broken, which leads to the appearance of minor and major sets of signals for the ξ and η protons at the indole ring (see Table 2 in the Supporting Information), and it is hard to draw a reasonable conclusion about the nature of the minor conformation.

All of the presented possible explanations of the appearance of the major and minor signal sets in the NMR spectra at T = 305 K and T = 313 K demonstrate that the common two-state model used to explain the folding/unfolding transition of trp-cage[4] does not apply in this case. Many previous studies questioned the simple two-state model and proposed more complicated models which include intermediates.[15,40] Our results do not support the two-state model, and we do not find any evidence on the presence of the intermediates on the pathway of the thermal unfolding of the trp-cage protein. Our finding rather supports the downhill folding mechanism[43−47] for the trp-cage protein, which implies that, during unfolding/folding up to the folding transition temperature, the conformational ensemble at a given temperature is not a mixture of fully folded and fully unfolded conformation (the two-state model) but rather is a uniform set conformation which shares some structural feature related to the native state. In the present study, we showed that the long-range Trp6−Pro12 and Trp6−Arg16 interactions are those structural elements which are present in the conformational ensembles at elevated temperatures (close to the transition temperature), but they are also present in the native state. Locally (residually), the protein melts and fast conformational equilibrium leads to disappearance of ROE/NOE signals in the NMR spectra, but vanishing of local interaction is not associated with vanishing some long-range interactions and this is what we observed in our experiments. The two-state model assumes that, during unfolding, the rate of the change of conformation-dependent quantities (such as the ROE/NOE signals) is at least comparable, but we observed that it is not true in the case of local and nonlocal interproton distances (see Table 1).

MD Simulations and Conformational Analysis. The data summarized in Figure 2 (we used only the data related to major sets of signals; see section "Conformational Ensemble at the Folding-Transition Temperature Implication on Folding Mechanism") were used to carry out MD simulations with time-averaged restrains, to determine the structure and conformational dynamics of trp-cage at 305 and 313 K. In Figures 5 and 6, the most populated conformational families clustered by using the MOLMOL[38] program collected at T = 305 K and T = 313 K, respectively, are presented. After DYANA calculations, we obtained 50 structures of trp-cage that could be divided into two main groups—one group of structures involved a Asp9−Arg16 salt-bridge, and the second one did not. Therefore, we decided to run two separate sets of MD TAV simulations for trp-cage with NMR restrains of each temperature (305 and 313 K). In the first MD simulation, the starting structures possessed the Asp9−Arg16 salt-bridge (the final MD structures are shown in panels a, b, and c of Figures 5 and 6), while the starting structures of the second set did not possess this salt bridge (the final MD structures are shown in panels d, e, and f of Figures 5 and 6). Three main families of conformations were selected as representative of the ensembles for each set of MD simulations, which represent about 70% of all obtained structures (see Figures 5 and 6). It can be seen

dx.doi.org/10.1021/jp212630y | J. Phys. Chem. B 2012, 116, 6898−6907

that, for the trp-cage at T = 305 K, the conformations of each family are similar to those obtained at T = 280 K[4] (see Figure 1). Admittedly, the N-terminal α-helix is not completely disrupted. It can clearly be seen that the conformations of almost each family possess an α-helix in this part of the molecule. An analogous situation occurs at T = 313 K, but appearance of the α-helix at the N-terminus is rather an artifact of the force-field used because NMR data does not support such a structure. At both temperatures, the protein chain bends (rapidly changes its direction) exactly at the same position as observed at T = 280 K[4] (Asp9, Gly10, Gly11 region). The structure of the trp-cage at T = 305 K is stabilized by several long-range interactions: Leu2−Gln5, Tyr3−Trp6, Tyr3−Leu7, Trp6−Pro12, and Trp6−Arg16 (see Table 1), which are also observed at T = 280 K[4] and T = 300 K.[7] The structure is much more dynamic at T = 305 K because of a smaller number of ROE contacts, as compared to those observed at T = 280 K. At T = 313 K, the structure becomes even more flexible than at T = 305 K; the number of ROE connectivities still decreases (only two long-range signals: Tyr3−Leu7, Trp6−Arg16; see Table 1).

According to some studies,[5,7] the Asp9−Arg16 salt-bridge is one of the main factors stabilizing the trp-cage three-dimensional structure. However, our results do not support this finding. Even when the starting MD structure contains the Asp9−Arg16 salt-bridge, the final structures do not always preserve it (see structures a, b, and c in Figures 5 and 6). Moreover, when the starting structure did not have the preformed Asp9−Arg16 salt-bridge, this salt-bridge did not form spontaneously in simulations. The results of our simulations clearly indicate that the Asp9−Arg16 salt bridge is less likely to occur when melting of the N-terminal α-helix is progressing with temperature increase. Our results support recent reports in which the Asp9−Arg16 interaction is pointed out as not very important for the trp-cage structure stabilization.[41]

As mentioned above, the overall shape of the structures presented at Figures 5 and 6 resembles those observed at 280 K[4] (see Figure 1) with the turn region (at residues 10−16) as the most conservative structural elements, which does not change substantially with temperature. The U-shape of the structure makes the Trp6 residue well shielded from the solvent. The shielding of Trp6 from the solvent at the folding-transition temperature is in agreement with results obtained by Ahmed and co-workers.[15] UV resonance Raman spectra clearly indicate that the Trp6 side chain is well shielded from the solvent at T = 313 K and a substantial shielding effect is observed up to T = 343 K.[15] A plausible explanation of this shielding effect is that the overall structure of trp-cage at very high temperatures still resembles those observed at T = 313 K (Figure 6), and the key factor responsible for shielding is a well organized structure in the turn region (at residues 10−16). The presence of the well organized structural element well above the folding temperature (in the unfolded state) was also observed in our earlier studies on the β-hairpin forming peptides.[16−19]

CONCLUSIONS

We interpreted 2D-^{1}H NMR spectra of the trp-cage protein at T = 305 K and T = 313 K. The phase-transition temperature (melting point) for the investigated miniprotein is reported to be in the range T = 313−317 K;[4,7,15] therefore, the results of our measurements give a detailed picture of the structure and dynamics of the polypeptide in the conformational phase-transition state. On the basis of data acquired in this work as well as on the data from the literature,[4,5,7,41] we identified the Trp6−Pro12 hydrophobic contact as a key interaction whose breaking/formation is responsible for the conformational phase transition (the melting point); however, the breaking/formation of the Trp6−Pro12 interaction is coupled with the melting/formation of the N-terminal α-helix (see Figure 2). We conclude that the unfolding/folding for the trp-cage protein occurs by synchronizing the breaking/formation of the Trp6−Pro12 interaction and the melting/formation of the N-terminal α-helix. It is interesting that the reported findings are in agreement with the results of our earlier studies of β-hairpin forming peptides.[16−20] In the case of β-hairpin forming peptides, we found that the breaking/formation of one long-range hydrophobic interaction is responsible for unfolding/folding transition. However, unlike peptides, proteins do possess secondary structure elements, which makes some parts of the sequence less flexible; thus, breaking/formation of some key interactions should be associated with conformational rearrangement of parts of the sequence, where the residues responsible for key interaction are located. A general conclusion from this study on the trp-cage protein as well as from the studies of β-hairpin-forming peptides[16−20] is that breaking/formation of specific long-range hydrophobic interaction is a key and the most important feature of unfolding/folding transition. However, presented results also showed that in the case of trp-cage unfolding the phase transition is not so sharp as in the case of β-hairpin-forming of similar size which may suggested that breaking of some of the hydrophobic interactions does not lead to substantial change in the solvation shell of the protein.[15] In other words, the folding transition is not associated with large conformational changes.

When studying the trp-cage protein at the folding transition temperature, we found that the polypeptide chain of this protein is very dynamic. The dynamics of the protein is manifested by a low number of long-range interactions observed in the ROESY spectra (see Table 1) as compared to lower temperatures[5,7] and by occurrence of the second sets of signals for some residues (see Figure 4). The commonly accepted two-state model[42] of protein folding implies that, at the folding-transition temperature, the conformational ensemble is composed of 50% folded and 50% unfolded conformation. However, the data from our experiments do not support such composition of conformational ensemble at the folding-transition temperature. For the trp-cage miniprotein at folding-transition temperature, one family of conformations dominates and the shape of this dominant conformation resembles the native conformation of the protein, because of the presence of long-range interactions characteristic of the native state (see Figures 1, 5, and 6). This finding is in agreement with the experimental[43,44] and theoretical studies[45−47] on other proteins, whose conformational ensembles at the folding-transition temperature are best described as collections of residually folded conformations (downhill folding mechanism) but not as mixtures of two or more distinct conformations (two-state folding mechanism). Our results also showed that data related to the conformational state of the protein backbone (CD spectra, chemical shifts) as a function of temperature are not sufficient to judge about the folding mechanism and some information about long-range interactions is necessary.[48] However, most of the studies of the folding mechanism rely solely on the data from CD

spectroscopy or analysis of the chemical-shift changes induced by temperature.[42]

■ ASSOCIATED CONTENT

Supporting Information

Tables showing chemical shifts. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: stan@biotech.ug.gda.pl.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Dr. Paweł Sowiński for help in the NMR measurements. This research was conducted by using the computational resources of the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, Poland.

■ REFERENCES

(1) Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 1309−1314.
(2) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87.
(3) Sarisky, C. A.; Mayo, S. L. *J. Mol. Biol.* **2001**, *307*, 1411−1418.
(4) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425−430.
(5) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci.U.S.A.* **2006**, *43*, 15859−15864.
(6) Bungan, M. R.; Yang, X.; Savern, J. G.; Gai, F. *J. Phys. Chem. B* **2006**, *110*, 3759−3763.
(7) Hudaky, P.; Starner, P.; Farkas, V.; Varadi, G.; Toth, G.; Perczel, A. *Biochemistry* **2008**, *47*, 1007−1016.
(8) Williams, D. W.; Bryne, A.; Stewart, J.; Andersen, N. H. *Biochemistry* **2011**, *50*, 1143−1152.
(9) Gellman, S. H.; Woolfson, D. N. *Nat. Struct. Biol.* **2002**, *9*, 408−410.
(10) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952−12953.
(11) Drer, R. B. *Curr. Opin. Struct. Biol.* **2007**, *17*, 38−47.
(12) Kubelka, J.; Eaton, W. A.; Hofrichter, J. *J. Mol. Biol.* **2003**, *329*, 625−630.
(13) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.
(14) Snow, C. D.; Zagorovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548−14549.
(15) Ahmed, Z.; Beta, I. A.; Mikhonin, A. V.; Asher, S. A. *J. Am. Chem. Soc.* **2005**, *127*, 10943−10950.
(16) Skwierawska, A.; Makowska, J.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 931−953.
(17) Skwierawska, A.; Żmudzińska, W.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 637−654.
(18) Lewandowska, A.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. *Biopolymers* **2010**, *93*, 469−480.
(19) Lewandowska, A.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 723−737.
(20) Lewandowska, A.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. *Biophys. Chem.* **2010**, *151*, 1−9.
(21) Piantini, U.; Sørensen, O. W.; Ernst, R. R. *J. Am. Chem. Soc.* **1982**, *104*, 6800−6801.
(22) Bax, A.; Freeman, R. *J. Magn. Reson.* **1985**, *65*, 355−360.
(23) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *63*, 207−213.
(24) Bartels, C.; Xia, T.; Billeter, M.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **1995**, *6*, 1−10.
(25) Tiers, G. V. D.; Coon, R. I. *J. Org. Chem.* **1961**, *26*, 2097−2098.
(26) Rule, G. S.; Hitchens, T. K. In *Fundamentals of Protein NMR Spectroscopy*; Focus on structural biology, Vol. *5*; Kaptein, R., Ed.; Springer: Dordrecht, The Netherlands, 2006; Chapter 16, pp 364−365.
(27) Güntert, P.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1991**, *217*, 517−530.
(28) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283−298.
(29) Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **1991**, *1*, 447−456.
(30) Bystrov, V. F. *Prog. NMR Spectrosc.* **1976**, *10*, 41−81.
(31) Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California, San Francisco: San Francisco, CA, 2006.
(32) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(33) Jorgensen, W. L.; Chandresekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(34) Ewald, P. P. *Ann. Phys.* **1921**, *64*, 253−287.
(35) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(36) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289−294.
(37) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 457−479.
(38) Koradi, R.; Billeter, M; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51−55.
(39) Streicher, W. W.; Makhatadze, G. I. *Biochemistry* **2007**, *46*, 2876−2880.
(40) Rovo, P.; Farkas, V.; Hegyi, O.; Szolomajer-Csikos, O.; Toth, G. K.; Perczel, A. *J. Pept. Sci.* **2011**, *17*, 610−619.
(41) Barua, B.; Lin., J. W.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H. *PEDS* **2008**, *21*, 171−185.
(42) Sosnick, T. R.; Barrick, D. *Curr. Opin. Struct. Biol.* **2011**, *21*, 12−24.
(43) Munoz, V.; Sanchez-Ruiz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17646−17651.
(44) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2006**, *442*, 317−321.
(45) Zou, G.; Wang, J.; Wang, W. *Proteins* **2006**, *63*, 165−173.
(46) Pitera, J. W.; Swope, W. C.; Abraham, F. F. *Biophys. J.* **2008**, *94*, 4837−4846.
(47) Maisuradze, G. G.; Liwo, A.; Ołdziej, S.; Scheraga, H. A. *J. Am. Chem. Soc.* **2010**, *132*, 9444−9452.
(48) Schuler, B.; Eaton, A. W. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16−26.