The Low-pH Unfolded State of the C-Terminal Domain of the Ribosomal Protein L9 Contains Significant Secondary Structure in the Absence of Denaturant but Is No More Compact Than the Low-pH Urea Unfolded State†

Bing Shan,‡ Shibani Bhattacharya,§ David Eliezer,*,|| and Daniel P. Raleigh*,‡,⊥

Department of Chemistry, State University of New York, Stony Brook, New York 11794-3400, New York Structural Biology Center, New York, New York 10027-7556, Department of Biochemistry and Program in Structural Biology, Weill Cornell Medical College, New York, New York 10032, and Graduate Program in Biochemistry and Structural Biology and Graduate Program in Biophysics, State University of New York, Stony Brook, New York 11794

Received April 18, 2008; Revised Manuscript Received June 26, 2008

ABSTRACT: There is considerable interest in the properties of the unfolded states of proteins, particularly unfolded states which can be populated in the absence of high concentrations of denaturants. Interest in the unfolded state ensemble reflects the fact that it is the starting point for protein folding as well as the reference state for protein stability studies and can be the starting state for pathological aggregation. The unfolded state of the C-terminal domain (residues 58−149) of the ribosomal protein L9 (CTL9) can be populated in the absence of denaturant at low pH. CTL9 is a 92-residue globular α, β protein. The low-pH unfolded state contains more secondary structure than the low-pH urea unfolded state, but it is not a molten globule. Backbone (^{1}H, ^{13}C, and ^{15}N) NMR assignments as well as side chain $^{13}C_\beta$ and $^1H_\beta$ assignments and ^{15}N R_2 values were obtained for the pH 2.0 unfolded form of CTL9 and for the urea unfolded state at pH 2.5. Analysis of the deviations of the chemical shifts from random coil values indicates that residues that comprise the two helices in the native state show a clear preference for adopting helical φ and ψ angles in the pH 2.0 unfolded state. There is a less pronounced but nevertheless clear tendency for residues 107−124 to preferentially populate helical φ and ψ values in the unfolded state. The urea unfolded state has no detectable tendency to populate any type of secondary structure even though it is as compact as the pH 2.0 unfolded state. Comparison of the two unfolded forms of CTL9 provides direct experimental evidence that states which differ significantly in their secondary structure can have identical hydrodynamic properties. This in turn demonstrates that global parameters such as R_h or R_g are very poor indicators of "random coil" behavior.

Characterization of the unfolded state ensemble of proteins is an essential step in the development of a complete description of the protein folding process (*1, 2*). A detailed understanding of the unfolded state, particularly the nature and energetics of any residual interactions, is important because the unfolded state is the thermodynamic reference state for protein engineering studies, as well as the starting state for protein folding, and can be the starting state for pathological protein aggregation. Along these lines, recent work has highlighted the fact that mutations can exert significant effects upon the energetics of the unfolded state ensemble and can also alter a protein's propensity to aggregate (*3–7*). Interest in the unfolded state has increased dramatically with the realization that actual unfolded state ensembles can, and often do, deviate significantly from the classic random coil model (*3, 5–13*). It is now recognized that the properties and structural propensities of the unfolded state can vary widely depending upon experimental conditions, ranging from relatively expanded ensembles with little propensity to form persistent structure to more compact ensembles containing elements of native and non-native structure (*14–17*). The latter are often found under more native conditions, while the former are often populated under denaturing conditions. The unfolded ensemble populated under native conditions is clearly the most physiologically relevant state, but it is difficult to characterize since it is normally sparsely populated because of the cooperativity of folding and the fact that the free energy balance favors the native state. These considerations have led to increased interest in proteins whose unfolded states can be accessed in the absence of denaturant.

† This work was supported by NIH Grants GM70941 (D.P.R.) and AG019391 and AG025440 (D.E.). D.P.R., S.B., and D.E. are members of the New York Structural Biology Center, which is a STAR center supported by the New York State Office of Science, Technology and Academic Research, is supported by NIH Grant P41 GM66354, and received funds from the National Institutes of Health, the Keck Foundation, New York State, and the NYC Economic Development Corporation for the purchase of 900 MHz spectrometers.

* To whom correspondence should be addressed. D.P.R.: telephone, (631) 632-9547; fax, (631) 632-7960; e-mail, draleigh@notes.cc.sunysb.edu. D.E.: telephone, (212) 746-6557; fax, (212) 746-4843; e-mail, dae2005@med.cornell.edu.

‡ Department of Chemistry, State University of New York.

§ New York Structural Biology Center.

|| Weill Cornell Medical College.

⊥ Graduate Program in Biochemistry and Structural Biology and Graduate Program in Biophysics, State University of New York.

10.1021/bi8006862 CCC: $40.75
Published on Web 08/16/2008

There is also considerable interest in characterizing the conformational propensities of the unfolded state ensemble at the level of individual residues. One particularly important question is whether unfolded state structures are limited to native interactions or whether they include contributions from non-native interactions. A second important issue is whether global parameters which report on compactness such as the radius of hydration (R_h)[1] or the radius of gyration (R_g) can be used to test if the unfolded state conforms to the random coil model. This is important because it is well-known that R_h and R_g should obey a characteristic power law dependence on the number of residues in a random coil chain. However, computational studies have suggested that proteins can obey the "random coil" power law yet contain significant secondary structure (*8, 14, 16, 18, 19*). Thus, R_h and R_g may not be good probes of deviations from random coil behavior.

NMR is ideally suited for investigations of the unfolded state since it is currently the only method which can provide residue specific information for every residue in the unfolded protein (*15, 17, 20–22*). The assignment of the NMR spectra of unfolded states can still be challenging because of limited spectral resolution. Fortunately, modern triple-resonance methods can often overcome these problems and have opened the door to a detailed characterization of the unfolded state (*23*). Here we report the use of NMR in characterizing the unfolded state of CTL9 in the absence of denaturant, and in the presence of 7.6 M urea. CTL9 is interesting in this regard since the low-pH unfolded state in the absence of urea is no more compact than the urea unfolded state as judged by their respective radii of hydration.

CTL9 is a 92-residue globular protein derived from the ribosomal protein L9 (Figure 1). The protein adopts an interesting fold made up of two α-helices and an unusual three-stranded mixed parallel, antiparallel β-sheet. The folding kinetics and equilibrium unfolding thermodynamics have been analyzed, and equilibrium unfolding is well-described as being two-state (*24, 25*). The domain can be unfolded by heat, by denaturant, or by low pH. The acid-induced unfolded state contains residual secondary structure as judged by CD, but it is not a molten globule (*26*). The goal of this work is to compare the secondary structure propensities of an unfolded state of CTL9 that is populated in the absence of denaturant and the one which is populated under strongly denaturing conditions. The terms unfolded state, denatured state ensemble, and denatured state are all found in the literature, and there is no convention with respect to when a particular terminology should be used. We use the term pH 2.0 unfolded state here to refer to the unfolded state populated in the absence of denaturant at pH 2.0.

RESULTS AND DISCUSSION

Sequence Specific Assignments of CTL9 in the pH 2.0 Acid Unfolded State and in the pH 2.5, 7.6 M Urea Unfolded State. The unfolded state of CTL9 populated at pH 2.0 in the absence of urea and the pH 2.5, 7.6 M urea unfolded state were characterized using ^{15}N, ^{1}H, ^{13}C triple-resonance methods. We choose pH 2.5 for the urea unfolded state

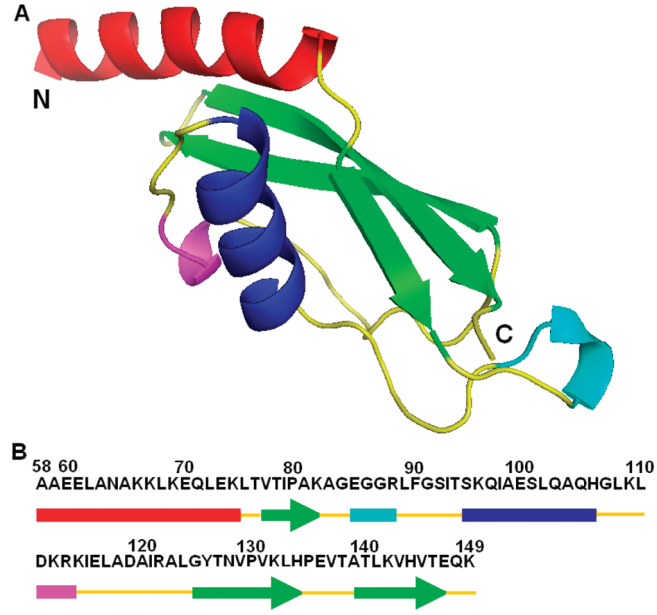

FIGURE 1: Structure of CTL9. (A) Ribbon diagram of residues 58−149 of L9 [Protein Data Bank (PDB) entry 1DIV]. The N- and C-termini are labeled. (B) The primary sequence of CTL9 is shown together with a schematic representation of the different elements of secondary structure (arrows represent β-strands and colored cylinders α-helices and 3_{10}-helices). The ribbon diagram was constructed using PyMOL.

studies since low-pH solutions with high urea concentrations require the addition of significant amounts of acid because of the buffering capability of urea near pH 2.0. Thus, adjusting the pH to 2.0 would lead to a noticeable increase in ionic strength. The pK_a values of all of the acidic residues in the urea unfolded state are expected to be very close to model compound values; thus, the choice of pH 2.0 or 2.5 should not affect the ionization state of any titratable side chains for the high-concentration urea studies. The HSQC spectrum of the pH 2.0 unfolded state of CTL9 is typical of that expected for an unfolded protein (Figure 2). A number of clearly resolved resonances are observed, but a large number of peaks are clustered together in a limited frequency range, spanning less than 1 ppm in the ^{1}H domain but covering approximately 20 ppm in the ^{15}N dimension. The limited dispersion is typical of unfolded proteins but is worse in the case of CTL9 because of the small number of aromatic residues in the sequence. The peaks are sharp as expected for a monomeric protein. Previously reported PFG-NMR diffusion experiments have shown that CTL9 is monomeric under the conditions of these studies (*26*). The HSQC spectrum of the urea unfolded state is also poorly resolved (Figure 2); nevertheless, nearly complete ^{13}C, ^{1}H, ^{15}N backbone $^{13}C_\beta$ and $^{1}H_\beta$ assignments could be obtained. We also assigned the native state spectrum at pH 3.8 to provide a basis for comparison with the unfolded state data. The folded state assignments have been reported for CTL9 at pH 5.5. However, complete $^{13}C_\beta$ assignments are not available in the literature (*27*). In addition, pH 3.8 is closer to the pH used for the unfolded state studies.

HNCACB and CBCACONH triple-resonance experiments were used to establish backbone connectivities for both states. The HNCO and HNCACO experiments were used to confirm these assignments and obtain ^{13}C carbonyl assignments. $^{1}H_\alpha$ and $^{1}H_\beta$ chemical shifts were obtained

[1] Abbreviations: CTL9, C-terminal domain of the ribosomal protein L9; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt; PFG-NMR, pulsed field gradient nuclear magnetic resonance; R_g, radius of gyration; R_h, radius of hydration.

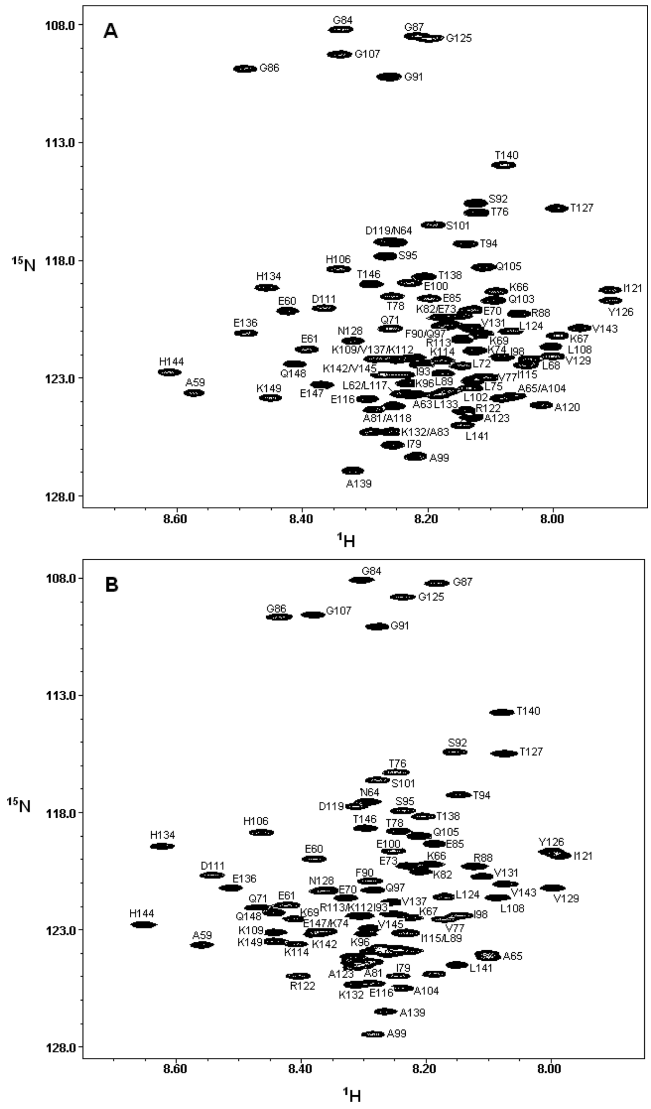

FIGURE 2: (A) ^{15}N−^{1}H HSQC spectrum of the pH 2.0 unfolded state of CTL9. (B) ^{15}N−^{1}H HSQC spectrum of the pH 2.5, 7.6 M urea unfolded state of CTL9. Peaks are labeled in both spectra. Spectra were recorded at 25 °C.

from analysis of the HBHACONH spectrum, while $^{13}C_\beta$ assignments were determined using the HNCACB experiment. HSQC-TOCSY experiments were used to confirm the assignments and aided in the assignments of resonances in residues immediately preceding prolines. Complete backbone assignments were determined for 89 of 92 residues, in the pH 2.0 unfolded state. The sole exceptions are the three prolines (P80, P130, and P135), for which we obtained ^{13}C carbonyl, $^{13}C_\alpha$, $^{13}C_\beta$, $^{1}H_\alpha$, and $^{1}H_\beta$ assignments but not ^{15}N assignments. ^{15}N and amide ^{1}H assignments were obtained for more than 85% of the non-proline residues in the urea unfolded state, and $^{13}C_\alpha$, $^{1}H_\alpha$, $^{13}C_\beta$, and $^{1}H_\beta$ assignments were obtained for all but three of the non-glycine residues. ^{13}C carbonyl assignments were obtained for 89 of the 92 residues. No peaks could be detected for L62, L102, and L117. Overall, at least partial assignments were obtained for 89 of 92 residues. Native state assignments at pH 3.8 were obtained using the same strategy and are used here of comparison with our unfolded state assignments. All of the assignments have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu/).

Analysis of ^{13}C, ^{1}H, and ^{15}N Chemical Shifts Provides Evidence of Native Secondary Structure in the Unfolded State in the Absence of Denaturant. It is well-known that secondary chemical shifts, i.e., the deviations between observed and random coil chemical shifts, are very sensitive to secondary structure. We analyzed the deviations between our experimentally determined chemical shifts and the random coil values reported for pH 2.3 in 8 M urea (*28*), as reported in NMRView. Since these chemical shifts were measured at pH 2.3, no pH correction is needed. Sequence-dependent corrections of ^{13}C carbonyl chemical shifts were made using reported protocols (*29*).

Figure 3 shows plots of the deviations of the C_α ^{1}H chemical shifts. C_α ^{1}H shifts are sensitive to secondary structure and are shifted upfield in α-helices and downfield in β-sheets. The first panel of the figure displays the deviations for the native state at pH 3.8; the second shows the plot for the pH 2.0 unfolded state, and the third panel displays the deviations for the urea unfolded state at pH 2.5. The deviations are large for the folded state, spanning the range from −0.73 to 1.05 ppm. The secondary shifts of the folded state are consistent with the known secondary structure of the protein. As expected, the deviations are much smaller for the pH 2.0 unfolded state but are clearly different from zero. The sequences corresponding to the first and second α-helices display contiguous negative deviations. For α-helix 1, the average deviation is −0.05 ppm, while it is −0.09 ppm for α-helix 2. For comparison, both of these regions in the native state show average deviations of −0.23 ppm. The deviations from random coil values obtained for the sequence corresponding to the β-sheet exhibit no particular trend. A consecutive set of negative deviations (−0.07 ppm in average) is also observed from residue 112 to 114, a segment forming a 3_{10}-helix in the native state. The plot for the urea unfolded state (Figure 3C) displays even smaller negative deviations of C_α ^{1}H chemical shifts, as compared with the pH 2.0 unfolded state. The averages are −0.04 and −0.06 ppm for α-helix 1 and α-helix 2 regions, respectively, while the segment corresponding to the 3_{10}-helix does not display any obvious trend. Taken in isolation, the C_α ^{1}H deviations suggest a tendency to populate helical regions of the φ and ψ plot in the pH 2.0 unfolded state for the residues which are helical in the native state. In the urea unfolded state, the tendency to form α-helical structure is much weaker. However, C_α ^{1}H secondary shifts are usually small for unfolded proteins and therefore may not be sensitive to small populations of residual secondary structure in a disordered protein. A more accurate and complete picture can be obtained by analyzing the $^{13}C_\alpha$, $^{13}C_\beta$, and ^{13}C carbonyl shifts.

Figure 4 displays plots of the deviations of the $^{13}C_\alpha$ (Figure 4A), carbonyl ^{13}C (Figure 4B), and $^{13}C_\beta$ chemical shifts (Figure 4C) from random coil for the pH 2.0 unfolded state. $^{13}C_\alpha$ and ^{13}CO shifts are expected to exhibit positive secondary shifts in α-helices and negative secondary shifts in β-strands. The pattern of the $^{13}C_\alpha$ secondary shifts is consistent with the C_α ^{1}H secondary shifts. The largest deviations are found in the regions which correspond to the two native state α-helices. The average deviation for the segment corresponding to α-helix 1 is 0.32 ppm, while it is 0.62 ppm for α-helix 2. Positive deviations are also observed

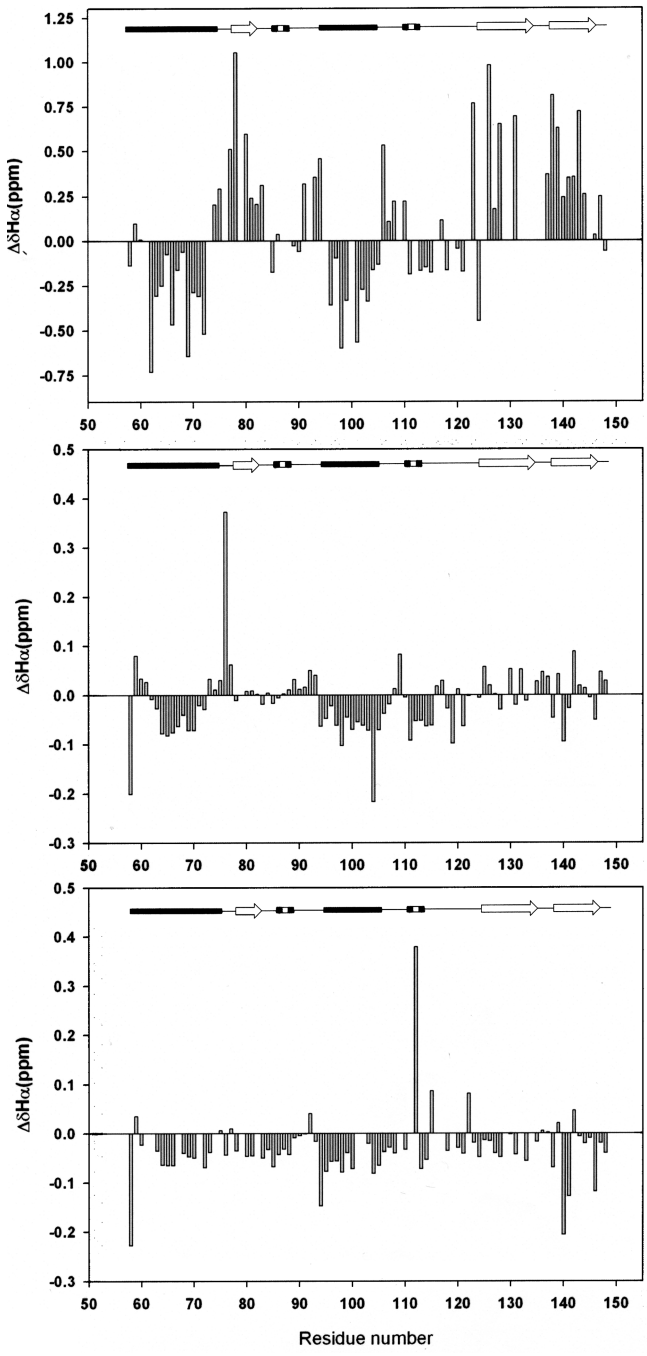

FIGURE 3: Plots of the deviations of the measured C_α ^1H chemical shifts and random coil chemical shifts. Data plotted as observed minus random coil. Random coil values in acidic (pH 2.3) 8 M urea were used (*28*) together with sequence specific corrections (*29*): (A) deviations for the native state (pH 3.8), (B) deviations for the pH 2.0 unfolded state, and (C) deviations for the pH 2.5, 7.6 M urea unfolded state. A schematic representation of the elements of secondary structure of the native state of CTL9 is shown at the top of each panel (arrows represent β-strands, filled cylinders α-helices, dashed cylinders 3_{10}-helices, and single lines loop regions).

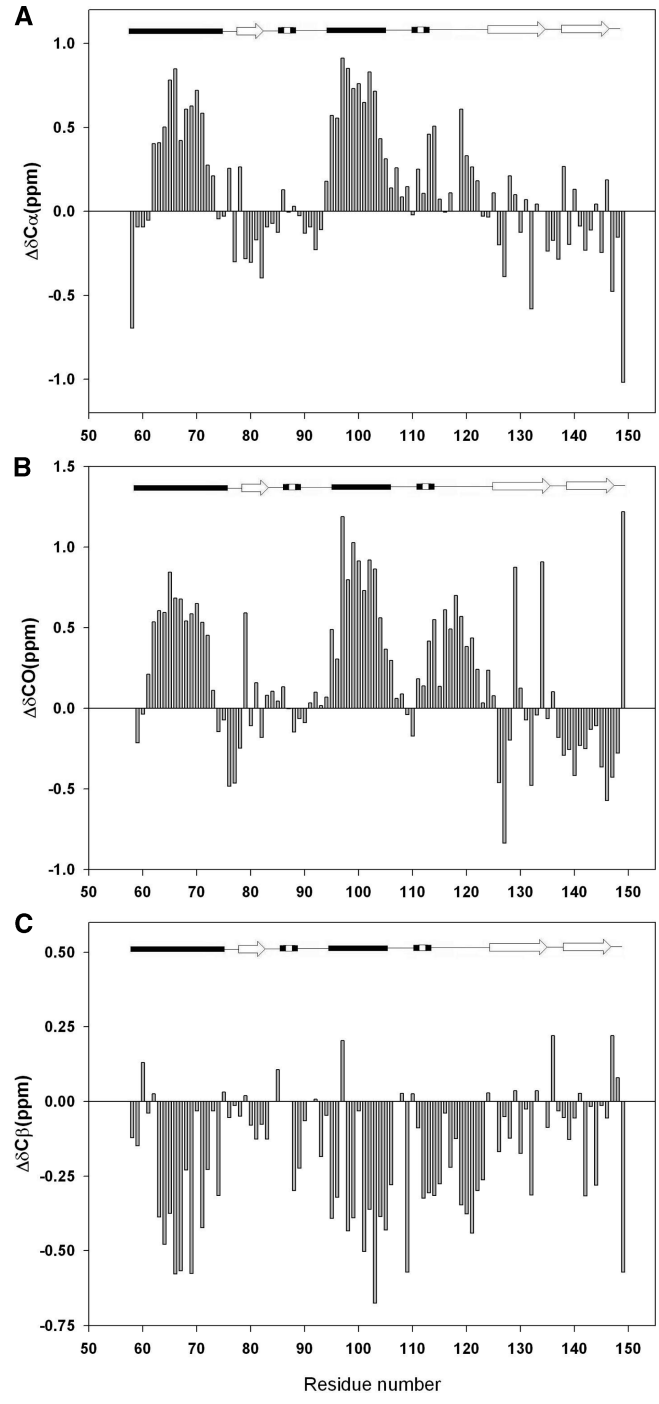

FIGURE 4: Plots of the deviations of the measured ^{13}C chemical shifts of the pH 2.0 unfolded state of CTL9 from random coil chemical values. Data plotted as observed minus random coil: (A) $^{13}C_\alpha$ chemical shift deviations, (B) carbonyl ^{13}C chemical shift deviations, and (C) $^{13}C_\beta$ chemical shift deviations. Sequence-dependent corrections were made (*29*). A schematic diagram of the elements of secondary structure of the native state of CTL9 is shown at the top of each panel.

for the loop region which connects α-helix 2 to β-strand 2 (residues 107−124); however, they are somewhat smaller (0.19 ppm). No strong trend is observed for the β-sheet regions. A similar pattern was observed for the ^{13}CO secondary shifts. The average is 0.41 ppm for α-helix 1 and 0.70 ppm for α-helix 2. Smaller positive deviations (0.28 ppm) are observed in the loop region between α-helix 2 and β-strand 2, in agreement with the smaller $^{13}C_\alpha$ and $^1H_\alpha$ secondary shifts for the region.

$^{13}C_\beta$ secondary shifts are also sensitive to secondary structure, but the sign of the deviation is reversed relative to that observed for $^{13}C_\alpha$ shifts (*28, 30*). The $^{13}C_\beta$ secondary shifts are consistent with the C_α ^1H, $^{13}C_\alpha$, and ^{13}CO secondary shifts. Negative secondary shifts are detected for the segments corresponding to two α-helices in the native state, and for the connection between α-helix 2 and β-strand 2. The

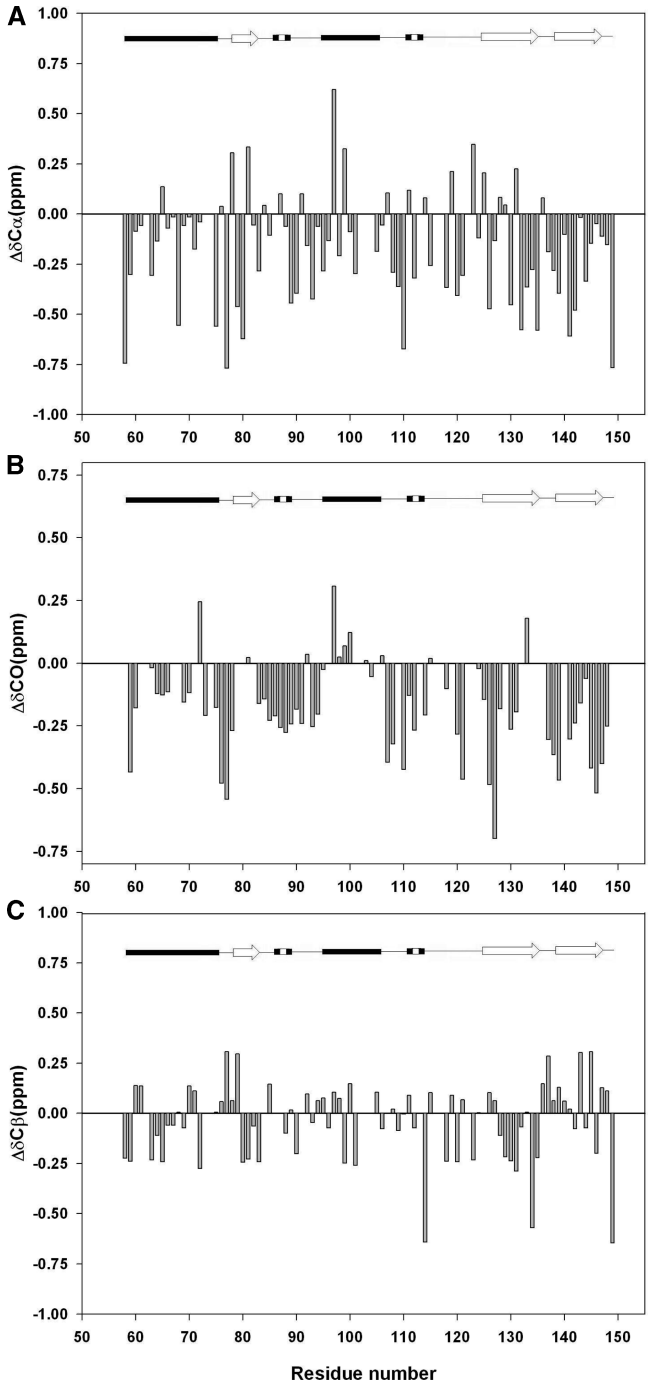

FIGURE 5: Plots of the deviations of the measured ^{13}C chemical shifts of the pH 2.5, 7.6 M urea unfolded state of CTL9 from random coil chemical values. Data plotted as observed minus random coil: (A) $^{13}C_\alpha$ chemical shift deviations, (B) carbonyl ^{13}C chemical shift deviations, and (C) $^{13}C_\beta$ chemical shift deviations. Random coil values in acidic (pH 2.3) 8 M urea were used (*28*), and sequence-dependent corrections were made (*29*). A schematic diagram of the elements of secondary structure of the native state of CTL9 is shown at the top of each panel.

average values for α-helical regions are −0.26 and −0.33 ppm, respectively, while the average for the connection region is −0.23 ppm. No consecutive regions of positive deviations are observed for segments corresponding to the three β-strands.

Figure 5 displays plots of the secondary shifts for the $^{13}C_\alpha$ (Figure 5A), carbonyl ^{13}C (Figure 5B), and $^{13}C_\beta$ (Figure 5C) chemical shifts for the pH 2.5, 7.6 M urea unfolded state. No consecutive set of residues can be found to display large secondary shifts. This indicates that the presence of a high concentration of denaturant abolishes most of the residual structure in the pH 2.0 unfolded state, making the urea unfolded state very different from the pH 2.0 unfolded state in terms of the amount of residual secondary structure. Consideration of the individual plots of secondary shifts gives rise to a self-consistent picture. Residues corresponding to α-helix 1 and α-helix 2 have a tendency to preferentially populate the helical region of the Ramachandran plot in the pH 2.0 unfolded state in the absence of urea. Also, there are suggestive hints about α-helical propensity in the sequence between α-helix 2 and β-strand 2. The results are quite different for the pH 2.5, urea unfolded state. The deviations from random coil values are much smaller, and no clear trends are detected.

Individual secondary shifts are useful, but recent work has shown that combining these data can provide more reliable information. A common approach is to calculate the difference between $^{13}C_\alpha$ secondary shifts and $^{13}C_\beta$ secondary shifts. This method has the advantage of cancelling out any uncertainty in the chemical shift reference. Positive $\Delta\delta C_\alpha - \Delta\delta C_\beta$ values indicate α-helical structure, and negative values indicate a propensity to adopt β-strand structure (*31*). Figure 6 displays the plot of the difference in $^{13}C_\alpha$ and $^{13}C_\beta$ secondary shifts. Several obvious trends are readily apparent for the pH 2.0 unfolded state. Positive values are observed for α-helix 1 and α-helix 2. The average for α-helix 1 is 0.58 ppm, and that for α-helix 2 is 0.95 ppm. A set of consecutive positive deviations are observed for the loop between α-helix 2 and β-strand 2 as well. The average for this region, residues 107−124, is 0.43 ppm. The difference is much smaller for the urea unfolded state, and no clear pattern is detected. The average deviation for the helix 1 region is −0.10 ppm, while it is −0.02 ppm for helix 2 and only −0.09 ppm for residues 107−124.

Recently, Forman-Kay and co-workers have used the weighted average of the secondary shifts to probe secondary structure where the weighting factors reflect the relative sensitivity of the secondary shifts to the type of structure. The method, known as SSP analysis, has been applied to the synucleins (*31*). Figure 7 shows the results of the SSP analysis applied to the pH 2.0 unfolded state and to the urea unfolded state. In this analysis, the higher the value of the SSP score, the greater the tendency to adopt helical φ and ψ angles. A value of 1.0 is expected for a residue which adopts helical φ and ψ angles 100% of the time and a value of −1.0 for a residue which exclusively populates the β-sheet region. The analysis is based upon the assumption that an observed secondary shift, relative to the average secondary shift for fully formed secondary structure, represents the fractional population of the structure at a particular site. As such, SSP scores should exhibit a correspondence with the propensity to adopt secondary structure.

The regions corresponding to α-helix 1 and α-helix 2 clearly have the highest SSP scores for the pH 2.0 unfolded state, with average scores of 0.28 ± 0.07 and 0.32 ± 0.07, respectively. The value for the connection linking α-helix 2 and β-strand 2 is 0.19. On the surface, the positive SSP score for this segment might imply that non-native structure is populated in this region. However, it is also possible that the native fold could generate positive SSP scores. Calcula-

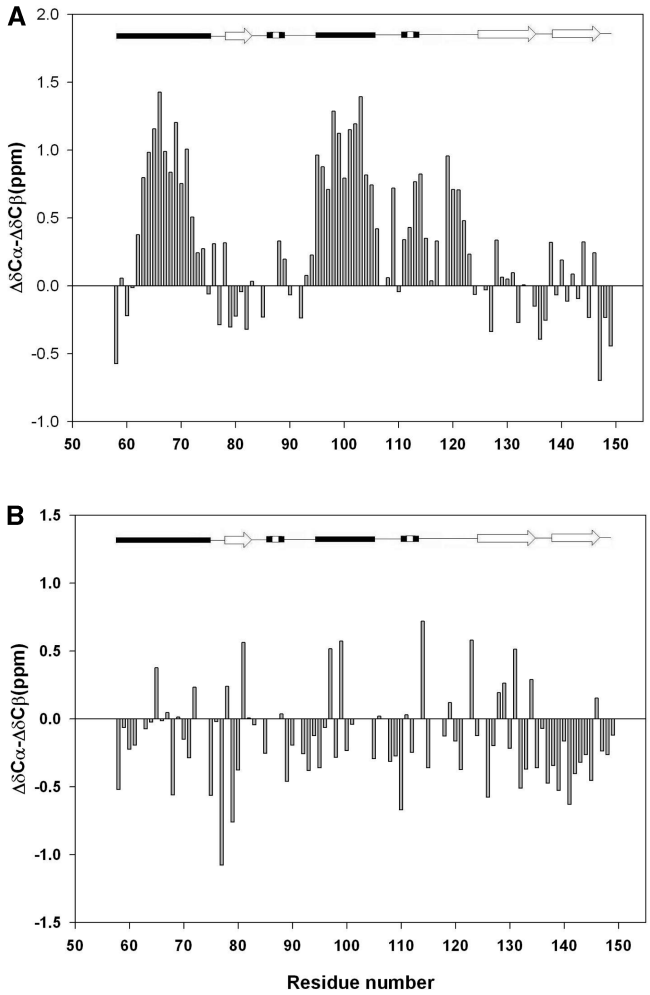

FIGURE 6: Plot of the difference in the secondary shifts of the $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts for (A) the pH 2.0 unfolded state of CTL9 and (B) the pH 2.5, 7.6 M urea unfolded state. A schematic diagram of the elements of secondary structure of the native state of CTL9 is shown at the top of each panel.

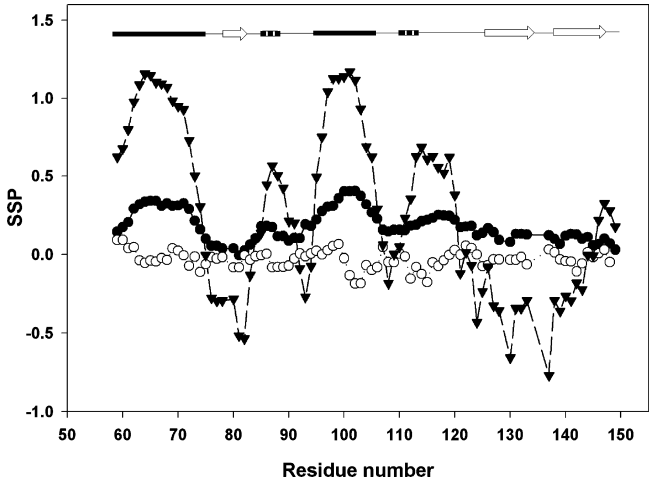

FIGURE 7: SSP analysis of the pH 2.0 (●) and pH 2.5, 7.6 M urea (○) unfolded state of CTL9 conducted using the method of Forman-Kay and co-workers (*26*). The calculation for the pH 3.8 native state is included for comparison (▼). Positive values indicate a propensity to populate the helical region of the φ and ψ map, while negative values indicate a preference for the β-sheet region. A schematic diagram of the elements of secondary structure of the native state of CTL9 is shown at the top. $^{13}C_\alpha$, $^{13}C_\beta$, and C_α ^1H chemical shifts were used in the analysis.

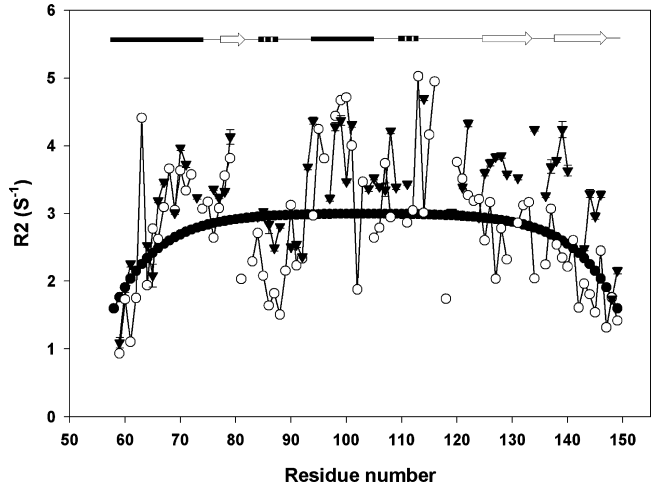

FIGURE 8: Plot of ^{15}N R_2 rates for the pH 2.0 unfolded state (○) and the pH 2.5, 7.6 M urea unfolded state (▼). The solid line is the best fit to the phenomenological model of Schwalbe and co-workers (*5*).

tion of the SSP score for the pH 3.8 native state gives an average of 0.25 for residues 107−124, indicating that positive SSP values in this segment can result from native state structure; thus, it is formally not clear if the chemical shift deviations between residues 107 and 124 represent a tendency to populate native or non-native structure (Figure 7). However, the key result is that the SSP analysis is completely consistent with all other measured parameters and indicates a tendency to form structure. The plot clearly reveals that helix propensity is significantly higher in the pH 2.0 unfolded state for regions of the protein that are helical in the native state. In contrast, the urea unfolded state shows no statistically significant tendency to populate nonrandom structure. The SSP values are -0.01 ± 0.06 for α-helix 1, -0.05 ± 0.09 for α-helix 2, and -0.04 ± 0.07 for the connection between α-helix 2 and β-strand 2. Forman-Kay and colleagues have pointed out that ^1H C_α secondary shifts can sometimes be unreliable and discussed how their inclusion or exclusion can alter the SSP analysis (*31*). Consequently, we repeated the analysis without including the ^1H C_α secondary shifts. The overall shape of the plot is essentially unaltered, and the conclusions are not affected (Supporting Information).

^{15}N R_2 relaxation rates have been used to detect apparent deviations from random coil behavior in unfolded proteins. Schwalbe and co-workers have pioneered the approach (*5, 32*). In their analysis, relaxation data are fit to a purely phenomenological model which assumes that the R_2 value at a particular residue is correlated with the values of R_2 at residues adjacent in primary sequence. The length of the correlated segment can, at some level, be viewed as being related to the persistence length. For a chain with only local interactions, a characteristic inverted U shape is predicted. Deviations from the predicted curve are typically taken to represent persistent clustering of hydrophobic side chains, although an atomic-level description of the origin of the effect is not available. Figure 8 compares plots of the measured R_2 values and the R_2 values calculated using the random coil model. The calculated R_2 values follow an inverted U-shaped curve, showing uniform relaxation rates in the middle of the polypeptide chain, with an average of 2.74 s^{-1}. Smaller rates are observed at both termini of the chain. There are no significant differences between the experimental data for the pH 2.0 unfolded state and the urea unfolded state.

There are deviations from the phenomenological "random coil" model for both data sets, but these are smaller than those that have been observed for other proteins (*32, 33*). The simplest interpretation of the relaxation data is that there are no long-lived clusters in the unfolded state(s); of course, one should bear in mind that the analysis involves comparison of experimental data to a phenomenological model.

CONCLUSIONS

We have obtained essentially complete ^{1}H, ^{13}C, and ^{15}N backbone assignments, as well as $^{13}C_{\beta}$ and ^{1}H C_{β} assignments for the pH 2.0 unfolded state of CTL9 and virtually complete assignments for the pH 2.5 urea unfolded state. The secondary shifts, the differences between the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ secondary shifts, and the SSP analysis all yield a self-consistent picture, namely, that residues which are helical in the native state have a propensity to preferentially populate helical φ and ψ angles in the pH 2.0 unfolded state. Residues 107−124, which form the connection between the second α-helix and the second β-strand, also exhibit a clear, albeit weakened, tendency to preferentially sample the helical region of the Ramachandran plot. Thus, the unfolded state of CTL9 contains significant secondary structure in the absence of denaturant. There is a very good correlation between the regions predicted by the chemical shift analysis to sample helical φ and ψ angles and the regions predicted to have the highest helical propensity by the AGADIR algorithm of Serrano and co-workers (*34*). AGADIR predicts that the first and second α-helices and the loop connecting the second α-helix to the second β-sheet have the highest helical propensity in CTL9 (Supporting Information). The urea unfolded state is very different and displays, as judged by the NMR analysis, no tendency to preferentially populate secondary structure.

It is difficult to precisely deduce the fractional population of a particular type of secondary structure from analysis of chemical shift data alone, but the SSP method appears to be the most promising approach. It is thus interesting to compare the SSP analysis to estimates of the helical content based upon CD measurements. The mean residue ellipticity of the pH 2.0 unfolded state is -4900 deg cm^{-1} mol^{-1} at 222 nm. The apparent fractional population of helix can be estimated from CD spectra if values for a fully helical peptide and for an unstructured peptide are known. The largest ambiguity arises in the choice of the ellipticity value for the random coil state. If we use the actual ellipticity measured in 8 M urea, we obtain an estimate of 8.2% α-helix. If, instead, we use empirical relations derived by Baldwin and co-workers, we obtain a value of 12.6% α-helix (*35*). The SSP analysis leads to a somewhat higher predicted fraction of α-helix. The average SSP score for the entire sequence is 0.21, which in the simplest interpretation corresponds to 21% α-helix (a SSP score of 1.0 is taken to represent a 100% tendency to populate helical φ and ψ angles). The discrepancy may reflect difficulties associated with using the SSP methodology to precisely quantify the relative population of secondary structure. Of course, difficulties in interpreting the CD spectra of partially structured polypeptides could also play a factor. However, it is worth bearing in mind that CD and chemical shifts have a fundamentally different sensitivity to helical structure, and we believe this is the most likely explanation. Secondary shifts are largely local in origin, and a short segment of chain that exclusively populates the helical region of $\varphi-\psi$ space would give rise to SSP scores near 1.0. In contrast, the CD signal of α-helices is well-known to depend on the length of the helix, and short helices can give rise to much weaker intensity at 222 nm than longer helices (*36*). Thus, the difference between the NMR and CD analysis could reflect a tendency for individual residues or short sections of the chain to adopt helical φ and ψ angles even though there is a much more modest propensity for a consecutive set of segment of residues to cooperatively sample helical conformations. Details aside, the NMR studies and CD measurements both demonstrate that the low-pH unfolded state of CTL9 clearly deviates from the classic random coil model.

At first glance, the observation of helical structure in the low-pH unfolded state may seem to be at variance with hydrodynamic measurements. Those studies showed the pD 2.1 unfolded state has the same radius of hydration ($R_h = 33.5$ Å) as the urea unfolded state at pD 3.8 ($R_h = 33.6$ Å) even though the urea unfolded state has much less, if any, residual helical structure (*26*). In addition, the R_2 data collected here for the pH 2.0 unfolded state and the pH 2.5, urea unfolded state are broadly similar. The apparent discrepancy simply reflects the fact that global parameters such as R_h and R_g can be insensitive to the presence of even significant amounts of secondary structure. This effect has been predicted on the basis of computational studies by Pande and co-workers (*19*). Those workers compared the calculated scattering profiles of random-flight chains to identical sequences that contained segments of α-helical structure. They found that the calculated values of R_g were very similar even though the sequences have quite different amounts of secondary structure. CTL9 provides a striking experimental example of this principle.

It is interesting to consider the results in light of a recently published analysis of the transition state of the folding of CTL9 (*25*). That study made use of φ value approaches and found that the largest φ values were localized in the β-hairpin encompassing β-strands 2 and 3 and the connecting loop. Small φ values were observed in mutations in both helix 1 and helix 2. Small φ values, in the absence of the unfolded state effects, indicate that the region being probed is no more structured in the transition state than in the unfolded state. However, small φ values do not necessarily mean that the region of the protein being studied is devoid of structure in the transition state. Instead, small φ values could arise because the region is similarly structured in both the unfolded state and the transition state (*37*); thus, there is a structural ambiguity in the interpretation of small φ values. The analysis presented here indicates that the two helices are partially formed in the unfolded state and thus also partially structured in the transition state for folding.

MATERIALS AND METHODS

Protein Expression and Purification. Uniformly ^{15}N-labeled CTL9 and ^{13}C- and ^{15}N-labeled CTL9 were expressed in *Escherichia coli* BL21 cells in M9 minimal medium. The medium for expressing the ^{15}N-labeled protein contains 0.8 g/L $^{15}NH_4Cl$ as the sole nitrogen source and 10 g/L glucose as the sole carbon source, while 0.8 g/L $^{15}NH_4Cl$ and 10 g/L [^{13}C]glucose were used in the expression of ^{13}C- and ^{15}N-labeled CTL9. Ampicillin was added to the main medium at a concentration of 100 mg/L. The cells were grown at 37 °C until the optical density (OD) at 600 nm reached 0.7, at

which point 100 mg/L IPTG was added to induce expression. After 4 h, cells were harvested and then lysed by sonication. The protein was purified as described previously (*24*). The molecular mass of the protein was confirmed by mass spectrometry.

NMR Sample Preparation. Protein samples for NMR experiments were prepared in 10% D_2O at pH 2.0 without urea or at pH 2.5 in 7.6 M urea, both at a concentration of approximately 1 mM. The urea concentration was determined by refractometry. The sample used for the native state assignments was prepared in 20 mM sodium acetate and 100 mM NaCl (pH 3.8).

NMR Spectroscopy. All heteronuclear NMR experiments were performed on uniformly ^{15}N-labeled or ^{13}C- and ^{15}N-labeled protein samples. All NMR spectra were recorded on a 700 or 800 MHz Bruker spectrometer with a cryoprobe at the New York Structural Biology Center at 25 °C. In all NMR experiments, the ^1H dimension was centered at the water resonance and the ^{15}N offset frequency was set to 118.0 ppm.

$^{15}N-^1H$ correlated heteronuclear single-coherence (HSQC) spectra were collected using 1024 × 256 complex points with eight scans per increment. The spectral widths of the pH 2.0 sample were 7183.9 and 2027.2 Hz for the ^1H and ^{15}N dimensions, respectively. The pH 2.5, 7.6 M urea sample had spectral widths of 6009.6 Hz (^1H) and 1561.0 Hz (^{15}N).

The following set of triple-resonance experiments performed on a ^{13}C- and ^{15}N-labeled sample of CTL9 were used to generate backbone assignments: HNCO, HN(CA)CO, HNCACB, and CBCACONH. In the HNCACB and CBCACONH experiments, the ^{13}C offsets were 39.0 ppm, whereas the ^{13}C offsets were set at 176.0 ppm in the HNCO and HN(CA)CO experiments. For the pH 2.0 unfolded CTL9, the HNCACB and CBCACONH spectra were recorded with 1024 × 80 × 160 complex points, with spectral widths of 7183.9, 2027.2, and 13076.2 Hz in the ^1H, ^{15}N, and ^{13}C dimensions, respectively. The HNCO and HN(CA)CO spectra were acquired with 1024 (^1H) × 72 (^{15}N) × 144 (^{13}C) complex points, with spectral widths of 7183.9, 2027.2, and 3018.4 Hz for ^1H, ^{15}N, and ^{13}C, respectively. For the pH 2.5, 7.6 M urea sample, the data sets for the HNCACB and CBCACONH spectra comprised 1024 (^1H) × 72 (^{15}N) × 128 (^{13}C) complex points. The spectral widths were 6009.6, 1561.0, and 13210.04 Hz in the ^1H, ^{15}N, and ^{13}C dimensions, respectively. The HNCO and HN(CA)CO spectra were acquired with 1024 (^1H) × 72 (^{15}N) × 128 (^{13}C) complex points, with spectral widths of 6009.6, 1561.0, and 2816.9 Hz for ^1H, ^{15}N, and ^{13}C, respectively.

Three-dimensional HBHACONH experiments were carried out to determine the $^{13}C_\alpha$ ^1H and $^{13}C_\beta$ ^1H chemical shifts. For the pH 2.0 unfolded sample, the HBHACONH spectrum was collected with 1024 × 80 × 160 complex points, with spectral widths of 7183.9 Hz (direct ^1H dimension), 2027.2 Hz (^{15}N), and 7200.72 Hz (indirect ^1H dimension). For the pH 2.5, 7.6 M urea unfolded sample, the spectral widths were 6009.6 Hz (direct ^1H dimension), 1561.0 Hz (^{15}N), and 7001.1 Hz (indirect ^1H dimension) with 1024 (direct ^1H dimension) × 72 (^{15}N) × 128 (indirect ^1H dimension) complex points.

A three-dimensional TOCSY-HSQC experiment was performed on a ^{15}N-labeled sample for the pH 2.0 unfolded state. A mixing time of 75 ms was used. The spectrum was recorded using a data matrix of 1024 (direct ^1H dimension) × 128 (^{15}N) × 256 (indirect ^1H dimension). Spectral widths were 8012.8 Hz in the direct ^1H dimension, 2432.9 Hz in the ^{15}N dimension, and 8001.6 Hz in the indirect ^1H dimension.

Measurement of Transverse Relaxation Rates. The relaxation experiments were carried out as described previously (*38*). The spectra were collected at 10 delay times: 16.32, 32.64, 48.96, 65.28, 81.60, 97.92, 114.24, 130.56, 146.88, and 163.20 ms. Repeat measurements were taken at 32.64, 48.96, 97.92, and 163.2 ms to allow estimation of uncertainty. Each T_2 experiment was conducted with four scans using 1024 × 256 complex points. The spectral widths were 9615.4 Hz (^1H) and 2027.7 Hz (^{15}N) for the pH 2.0 unfolded state, and they were 7183.9 Hz (^1H) and 2027.6 Hz (^{15}N) for the pH 2.5, 7.6 M urea unfolded state. A recycle delay of 3 s was used.

Date Processing and Analysis. All spectra were processed using NMRPipe (*39*), and chemical shift assignments were made using NMRView (*40*). All chemical shifts were referenced to the absolute frequency of DSS at 0 ppm. Sequence-dependent corrections of the chemical shifts were made using methods developed by Schwarzinger, Wright, Dyson, and co-workers (*29*). Random coil values in acidic (pH 2.3) 8 M urea (*28*) were used to calculate the secondary chemical shifts, except that the values used for Glu and Val $^{13}C_\beta$ random coil shifts are those reported by the NMRView algorithm. SSP analysis was performed using the method of Forman-Kay as described previously (*31*). The software provided at their website (http://pound.med.utoronto.ca/software.html) was used. R_2 relaxation rates were determined using the automated program in NMRView by fitting the peak intensities to eq 1 which describes a two-parameter exponential decay:

$$I(t) = I_0 \exp(-t/T_2) \quad (1)$$

where $I(t)$ is the peak intensity after a delay of time t and I_0 is the intensity at the time zero. The R_2 rates were analyzed using a simple model by fitting the experimental R_2 rates to eq 2:

$$R_2(i) = R_2(\text{int}) \sum_{j=1}^{N} \exp\left(-\frac{|i-j|}{\lambda}\right) \quad (2)$$

where $R_2(i)$ is the experimental R_2 value for residue i, R_2(int) is the intrinsic relaxation rate which depends on the temperature and viscosity of the solution, λ is the persistence length of the chain, and N is the total number of residues in the protein (*5*).

ACKNOWLEDGMENT

We thank Prof. David Hoffman for providing the pH 5.5 native state assignments of CTL9, and we thank Dr. Ying Li for her contributions to the initial stages of this project.

SUPPORTING INFORMATION AVAILABLE

A figure showing the results of the SSP analysis conducted with the exclusion of the C_α ^1H chemical shifts, a figure showing the results of the AGADIR analysis, and a table of ^{15}N R_2 values measured for the pH 2.0 and 2.5, 7.6 M urea unfolded states and the assignments for the pH 2.0 unfolded

state and the pH 2.5, 7.6 M urea unfolded state. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Baldwin, R. L. (2002) A new perspective on unfolded proteins. *Adv. Protein Chem. 62*, 361–367.
2. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol 6*, 197–208.
3. Cho, J. H., Sato, S., and Raleigh, D. P. (2004) Thermodynamics and kinetics of non-native interactions in protein folding: A single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *J. Mol. Biol. 338*, 827–837.
4. Mok, Y. K., Elisseeva, E. L., Davidson, A. R., and Forman-Kay, J. D. (2001) Dramatic stabilization of an SH3 domain by a single substitution: Roles of the folded and unfolded states. *J. Mol. Biol. 307*, 913–928.
5. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a non-native protein. *Science 295*, 1719–1722.
6. Cho, J. H., and Raleigh, D. P. (2005) Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. *J. Mol. Biol. 353*, 174–185.
7. Pace., C. N., Alston, R. W., and Shaw, K. L. (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci. 9*, 1395–1398.
8. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins. *Annu. Rev. Biochem. 60*, 795–825.
9. Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, D. J. (2001) NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry 40*, 3561–3571.
10. Shortle, D. R. (1996) Structural analysis of non-native states of proteins by NMR methods. *Curr. Opin. Struct. Biol. 6*, 24–30.
11. Choy, W.-Y., and Forman-Kay, J. D. (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol. 308*, 1011–1032.
12. Lindorff-Larson, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C. M., Poulsen, F. M., and Vendruscolo, M. (2004) Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein. *J. Am. Chem. Soc. 126*, 3291–3299.
13. Zhang, O., and Forman-Kay, J. D. (1997) NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. *Biochemistry 36*, 3959–3970.
14. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasen, M. Z., Pande, V. S., Ruczinski, I., Doniach, S., and Plaxco, K. W. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12491–12496.
15. Dyson, J. H., and Wright, P. E. (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv. Protein Chem. 62*, 311–340.
16. Fitzkee, N. C., and Rose, G. D. (2004) Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12497–12502.
17. Mittag, T., and Forman-Kay, J. D. (2007) Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol. 17*, 3–14.
18. Goldenberg, D. P. (2003) Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol. 326*, 1615–1633.
19. Zagrovic, B., and Pande, V. S. (2003) Structural correspondence between the α-helix and the random-flight chain resolves how the unfolded proteins can have native-like properties. *Nat. Struct. Biol. 10*, 955–961.
20. Neri, D., Billeter, M., Wider, G., and Wuthrich, K. (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science 257*, 1559–1563.
21. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V., and Fersht, A. R. (2005) Solution structure of a protein denatured state and folding intermediate. *Nature 437*, 1053–1056.
22. Eliezer, D. (2007) Characterization residual structure in disordered protein states using nuclear magnetic resonance. *Methods Mol. Biol. 350*, 49–67.
23. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1996) *Protein NMR spectroscopy: Principles and practice*, Academic Press, Inc., San Diego.
24. Sato, S., and Raleigh, D. P. (2002) pH-dependent stability and folding kinetics of a protein with an unusual α-β topology: The C-terminal domain of the ribosomal protein L9. *J. Mol. Biol. 345*, 163–173.
25. Li, Y., Gupta, R., Cho, J. H., and Raleigh, D. P. (2007) Mutational analysis of the folding transition state of the C-terminal domain of ribosomal protein L9: A protein with an unusual β-sheet topology. *Biochemistry 46*, 1013–1021.
26. Li, Y., Picart, F., and Raleigh, D. P. (2005) Direct characterizatoin of the folded, unfolded and urea-denatured states of the C-terminal domain of the ribosomal protein L9. *J. Mol. Biol. 349*, 839–846.
27. Hoffman, D. W., Cameron, S. C., Davies, S., White, S. W., and Ramakrishnan, V. (1996) Ribosomal Protein L9: A structure determination by the combined use of X-ray crystallography and NMR spectroscopy. *J. Mol. Biol. 264*, 1058–1071.
28. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Wright, P. E., and Dyson, H. J. (2001) Random coil chemical shifts in acidic 8M urea: Implementation of random coil shift data in NMRView. *J. Biol. NMR 18*, 43–48.
29. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc. 123*, 2970–2978.
30. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR 5*, 67–81.
31. Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci. 15*, 2795–2804.
32. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of Lysozyme in 8 M urea. *Biochemistry 36*, 8977–8991.
33. Platt, G. W., McParland, V. J., Kalverda, A. P., Homans, S. W., and Radford, S. E. (2005) Dynamics in the unfolded state of β2-microglobulin studied by NMR. *J. Mol. Biol. 346*, 279–294.
34. Lacroix, E., Viguera, A. R., and Serrano, L. (1998) Elucidating the folding problem of α-helices: Local motifs, long-range electrostatics, ionic strength dependence and prediction of NMR parameters. *J. Mol. Biol. 284*, 173–191.
35. Luo, P., and Baldwin, R. L. (1997) Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry 36*, 8413–8421.
36. Chin, D. H., Woody, R. W., Rohl, C. A., and Baldwin, R. L. (2002) Circular dichroism spectra of short, fixed-nucleus alanine helices. *Proc. Natl. Acad. Sci. U.S.A. 99*, 15416–15421.
37. Luisi, D. L., Kuhlman, B., Sideras, K., Evans, P. A., and Raleigh, D. P. (1999) Effects of varying the local propensity to form secondary structure on the stability and folding kinetics of a rapid folding mixed α/β protein: Characterization of a truncation mutant of the N-terminal domain of the ribosomal protein L9. *J. Mol. Biol. 289*, 167–174.
38. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry 33*, 5984–6003.
39. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multi-dimensional spectral processing system based on Unix Pipes. *J. Biomol. NMR 6*, 277–293.
40. Johnson, B. A. (2004) Using NMRView to visualize and analyze NMR spectra of macromolecules. *Methods Mol. Biol. 278*, 313–352.

BI8006862