Structural and Dynamic Characterization of the Acid-Unfolded State of hUBF HMG Box 1 Provides Clues for the Early Events in Protein Folding[†,‡]

Xuecheng Zhang, Yingqi Xu, Jiahai Zhang, Jihui Wu,* and Yunyu Shi*

Hefei National Laboratory for Physical Sciences at Microscale, School of Life Science, University of Science and Technology of China, Hefei, Anhui 230027, People's Republic of China

Received February 1, 2005; Revised Manuscript Received March 16, 2005

ABSTRACT: To understand the events that occur in the early stages of the folding of hUBF HMG box 1, we characterized its pH 2.1 unfolded state in detail with NMR. Through a triple resonance strategy, the assignments of complete backbone and some side chains were achieved. Then, significant conformational information was extracted from secondary chemical shifts, interresidual ^{1}H-^{1}H NOEs, $^3J_{HNHA}$ coupling constants, amide proton temperature coefficients, and ^{15}N relaxation data. The secondary chemical shifts for ^{13}CA, ^{13}CB, ^{13}CO, ^{1}HA, and ^{1}HN indicate that the residues between 64 and 78 exhibit a substantial preference for helical structure in the acid-unfolded state, which is also evidenced by the relatively more negative deviations of $^3J_{HNHA}$ and amide proton temperature coefficients from their corresponding random-coil values and particularly confirmed by the strongest sequential $d_{NN}(i, i + 1)$ proton NOEs along the region. Following this region until residue 82 is a segment that tends to form a turn-like structure, which is unstable and exchanges between alternative states. In addition, some evidences imply that the regions 18−28 and 38−43 also possess propensities for helical structure but to a different less degree than the region 64−78. The polypeptide backbone dynamics investigated using reduced spectral density function shows apparent motional restrictions in residual structural regions and to less extent at some hydrophobic residues. On the basis of the results presented herein, we propose a potential protein-folding pathway on which these residual structures play a role of initiation site in the early folding stages.

Elucidating the mechanism of protein folding requires complete characterization of all of the states of protein present along protein-folding pathways, including the native, the partially folded intermediate, and the fully unfolded states (*1*−*5*). Many small proteins adopt partially folded states at equilibrium under certain denaturing conditions such as high concentrations of guanidine hydrochloride (GdnHCl),[1] elevated temperature, extreme pH, etc. There is strong evidence for the structural correspondence between these equilibrium states and kinetic intermediates during the protein-folding process (*3*−*6*). Although equilibrium states do not necessarily resemble kinetic intermediates, a close relationship between them is indisputable because of the same principles governing the behavior of proteins. In these states, residual structures represent local preferences and may be initiation sites for folding of proteins. Therefore, detailed structural and dynamic characterization of denatured states formed at equilibrium can give significant contribution to the elucidation of initial events during a folding process (*3*−*8*).

The structures of folded proteins can be determined by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, but it is difficult to characterize unfolded proteins because of their conformational heterogeneity and flexibility. From the X-ray crystallography point of view, the difficulty is that it is nearly impossible to crystallize unstructured proteins. As for NMR, the primary problem arises from the poor chemical-shift dispersions especially for ^{1}H resonance in unfolded states, which hampers resonance assignments and the following in-depth studies. Nonetheless, the problem can be circumvented by making use of relatively better spectral dispersion of ^{15}N and ^{13}CO nuclei. Once resonance assignments are obtained, NMR can give a great deal of (less specific but accurate) information of a conformational ensemble. Thus, overwhelming crystallography spectroscopy, NMR is extensively adopted as the leading technique in studies of non-native structures (*1*−*6*). In resent years, many denatured states of proteins have been directly characterized on the individual residue level by labeling proteins with ^{15}N and ^{13}C coupled with using multidimensional heteronuclear NMR methods (*9*−*17*).

Human upstream binding factor (hUBF) is a nuclear transcription factor. It plays a critical role in the activation of ribosomal RNA gene transcription. It contains six tan-

[†] This work was supported by the Chinese National Fundamental Research Project, Grant G1999075605 and 2002CB713806; the Chinese National Natural Science Foundation, Grant 30270293 and 30121001; the Key Project of the National High Technology Research and Development Program of China, Grant 2002BA711A13; and the Pilot Project of the Knowledge Innovation Program of the Chinese Academy of Science, Grant KSCX1-SW-17.

[‡] The resonance assignments for the ^{1}H, ^{15}N, and ^{13}C chemical shifts of the main chain and some side chains for the acid-unfolded hUBF HMG box 1 at pH 2.1 have been deposited in the BioMagResBank (http://www.brmb.wisc.edu) under BMRB accession number 6473.

* To whom correspondence should be addressed. Telephone: +86-551-3607464. Fax: +86-551-3601443. E-mail: wujihui@ustc.edu.cn (J.W.); yyshi@ustc.edu.cn (Y.S.).

[1] Abbreviations: NOE, nuclear Overhauser enhancement; NMR, nuclear magnetic resonance; MG, molten globule; GdnHCl, guanidine hydrochloride; CD, circular dichroism; ANS, 1-anilino-8-aphthalene sulfonate; CSI, chemical-shift index.

10.1021/bi0501939 CCC: $30.25
Published on Web 05/12/2005

demly arranged HMG (high mobility group) box domains, among which hUBF HMG box 1 is the first one. hUBF HMG box 1 is small (about 90 residues) and can bind to the ribosomal promoter specifically (*18*−*20*). The three-dimensional structure determined by NMR shows that its native form adopts a twisted L shape consisting of three α helices, with a hydrophobic core embraced by these helices and stabilized by extensive hydrophobic interactions between nonpolar residues around the core (*21*). HMG boxes have been found in many other proteins from different species, and they all adopt a similar fold like that of hUBF HMG box 1. Thus, hUBF HMG box 1 can be regarded as a very good model for the study of protein folding. In a previous work (*22*), we have reported its acid-induced unfolding in the presence of a low concentration of salt at equilibrium. At about pH 2, the protein is unfolded largely, and as pH decreases further to about 0.2 or lower, it partially refolds to a molten globule (MG)-like state (A state). The acid-unfolded state (U state) at about pH 2 and the A state have been characterized by circular dichroism (CD), 1-anilino-8-aphthalene sulfonate (ANS) fluorescence, and 2D NMR spectroscopy. A fact that attracts our attention is that unlike the fully unfolded state produced by GdnHCl, the acid-unfolded state of hUBF HMG box 1 still retains some residual local structures (*22*).

In the current work, we used the multidimensional heteronuclear NMR method to characterize structure and dynamics of the acid-unfolded hUBF HMG box 1 at pH 2.1 on the residue level. The chemical-shift data involving five nuclei show that in several regions there are propensities for certain secondary structures more or less. The interresidual ^{1}H-^{1}H nuclear Overhauser enhancements (NOEs), $^3J_{HNHA}$ coupling constants, and amide proton temperature coefficients data provide complementary evidence for the conclusion drawn from the chemical-shift analysis. Moreover, the dynamic data analyzed with the reduced spectral density function have given insight into the backbone motions and identified the motional restrictions in the segments that correspond well to the residual local structures indicated by other NMR parameters.

MATERIALS AND METHODS

Protein Production. hUBF HMG box 1 uniformly labeled with ^{15}N and doubly labeled with ^{15}N and ^{13}C was expressed and purified with a method as described previously (*23*). After purification, the protein was dialyzed against distilled water and then lyophilized.

NMR Sample Preparation. NMR samples were prepared by dissolving 6 mg of lyophilized protein directly into a 400 μL solution of 10 mM Na_2HPO_4/KH_2PO_4 in 95% ^1H_2O/5% ^2H_2O, the pH of which was adjusted to 1.7 using HCl in advance. The final pH of solution was 2.1. Then, the samples were centrifugated to remove precipitates.

NMR Spectroscopy. All NMR experiments employed pulsed field gradients and gradient-selected sensitivity enhancement. The ^{1}H resonance was referenced to water at 30 °C (4.726 ppm). All heteronuclear dimensions were referenced indirectly based on the water signal. The carriers for ^{1}H, ^{15}N, ^{13}CA ^{13}CB, and ^{13}CO were respectively set at the water resonance, 119.036, 45.693, and 176.662 ppm. Two-dimensional ^{1}H-^{15}N HSQC spectra were acquired with 1024 (t_2) and 256 (t_1) complex points and with spectra widths 6009.615 (^{1}H) and 973.094 Hz (^{15}N). In all 3D experiments, the spectral widths along the ^{1}H and ^{15}N dimension are the same as in 2D ^{1}H-^{15}N HSQC spectra, and the number of complex points along t_3 for most experiments is the same, 1024, except when specially mentioned. ^{15}N TOCSY−HSQC was acquired with 180 and 50 complex points in the t_1 and t_2 dimension. ^{15}N NOESY−HSQC was acquired with 200 (t_1) and 60 (t_2) complex points. Spectral width along ^{1}H of both spectra was 5401.026 Hz. CBCA(CO)NH and HNCACB spectra were acquired with spectral width of 9659.503 Hz (^{13}C) and with 44 (t_2), 80 (t_1) and 44 (t_2), 100 (t_1) complex points, respectively. HNCO and HN(CA)-CO spectra were recorded with the spectral width of 2263.980 Hz (^{13}C) and with 44 (t_2), 60 (t_1) and 44 (t_2), 70 (t_1) complex points, respectively. HNN and HN(C)N (*24*) spectra were acquired with 1536, 42, and 42 complex points in the three dimensions. ^{15}N HMQC−NOESY−HSQC spectrum was acquired with 60 (t_2) and 60 (t_1) complex points. The HNHA experiment was carried out with a spectral width of 4200.798 Hz (^{1}H) and with 1536 (t_3), 90 (t_2), and 50 (t_1) complex points. A 61-ms mixing time was used in the ^{15}N TOCSY−HSQC experiment, and a 250 ms mixing time was used in both ^{15}N NOESY−HSQC and ^{15}N/^{15}N HMQC−NOESY−HSQC experiments. All of the above experiments were performed on a Bruker DMX600 spectrometer at 30 °C.

Temperature coefficients for amide protons were obtained from a series of ^{1}H-^{15}N HSQC spectra recorded at 3 °C intervals between 21 and 45 °C. All spectra were acquired on a Bruker DMX500 spectrometer with 1024 (t_2) and 300 (t_1) complex points, and the spectral widths were 5000 (^{1}HN) and 912.305 Hz (^{15}N).

^{15}N T_1 and T_2 relaxation times and steady-state ^{1}H-^{15}N NOEs were measured on a Bruker DMX500 spectrometer at 30 °C. For the T_1 measurements, 6 time points were collected with delays of 11.2, 61.6, 142, 243, 364, and 525 ms; duplicate data were acquired at the third delay time point to allow estimation of errors. For T_2, 6 time points with delays of 17.6, 35.2, 52.8, 70.4, 105.6, and 140.8 ms were acquired. The steady-state ^{1}H-^{15}N NOE was measured from duplicate pairs of ^{1}H-^{15}N spectra recorded with and without amide proton saturation. All ^{1}H-^{15}N HSQC spectra for measurements of T_1 and T_2 were recorded by using 1536 and 220 complex points in the t_2 and t_1 dimension, with spectral widths of 7002.801 (HN) and 912.305 Hz (^{15}N). A recycle delay of 1 s was used for T_1 and T_2 measurements. The ^{1}H-^{15}N HSQC spectra for measurements of ^{1}H-^{15}N NOE were acquired with 1024 (t_1) and 400 (t_2) complex points and with the same spectral widths as T_1 and T_2. Saturation was achieved by using a train of 120° pulses with 5 s of recycle delay.

Data Analysis. NMR data were processed and analyzed using the programs NMRPipe, NMRDraw, Sparky, and NMRView. Data were typically apodized using a 54° shifted $\sin^2$ function before zero-filling and Fourier transformation in all dimensions. Time domain data in the ^{15}N and ^{13}C dimensions were linear-predicted to increase the number of data points. In general, baseline corrections were not necessary. Secondary chemical-shift data were obtained using the chemical-shift index (CSI) in NMRview as described by Yao et al. (*25*−*28*). Maximum peak heights were used in

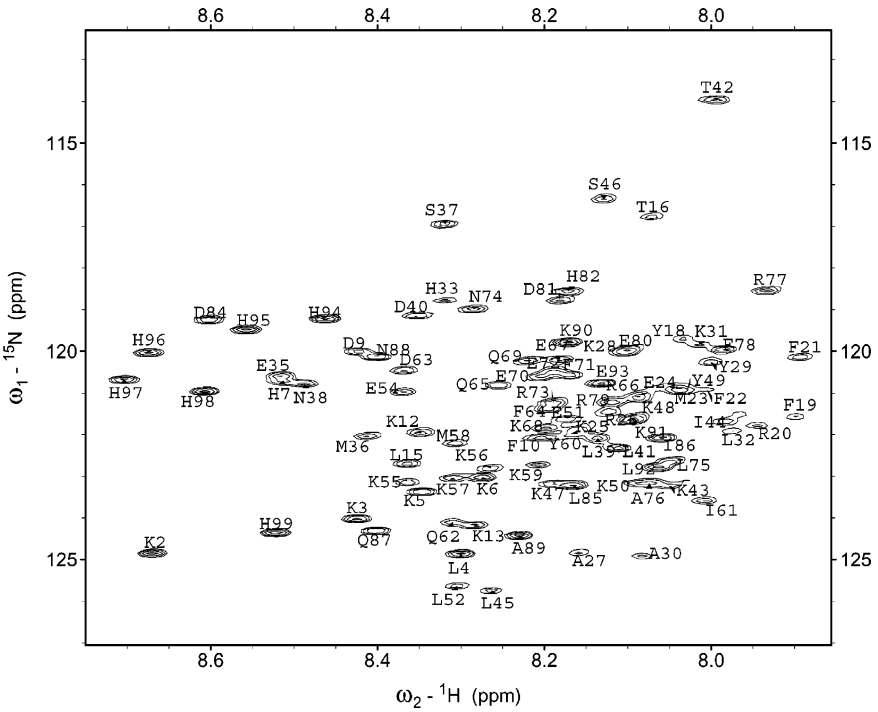

FIGURE 1: ^{1}H-^{15}N HSQC spectrum (600 MHz) for the acid-unfolded hUBF HMG box 1 in 10 mM Na_2HPO_4/KH_2PO_4 at pH 2.1 and 30 °C.

processing the R_1 and R_2 and steady-state ^{1}H-^{15}N NOE data. Decays were fit to monoexponential functions. Spectral densities were calculated using method 1 as described by Farrow et al. (*29*). Temperature coefficients were calculated from a linear regression to all of the temperature points.

RESULTS

Resonance Assignments Obtained Using Heteronuclear Resonance Method. The 600 MHz ^{1}H-^{15}N HSQC spectrum of the acid-unfolded hUBF HMG box 1 at pH 2.1 and 30 °C is shown in Figure 1. The narrow resonance dispersion of the acid-unfolded hUBF HMG box 1 (the chemical shifts of ^{1}HN are within a range of only 1 ppm) is characteristic of a denatured protein. However, the chemical-shift dispersion of ^{15}N is still large, and the cross-peaks are sufficiently well-resolved for assignment. Because of the nature of an unfolded state, the chemical shifts of the side chains of the acid-unfolded hUBF HMG box 1 are very close to random-coil values, which makes determining spin system types with ^{15}N-TOCSY, HNCACB, and CBCA(CO)NH easy. Because the chemical shifts of ^{15}N and ^{13}CO are more sensitive to the local amino acid sequence rather than structural context and therefore remain well-resolved even in fully unfolded states (*1*−*5*), most sequential connectivities can be obtained from HNCACB and CBCA(CO)NH or HNCO and HN(CA)CO, and the ambiguity derives from the resonance overlap of CA/CB or CO can be resolved in another ^{15}N dimension by HNN and HN(C)N (*24*). Using the above assignment strategy, 91 of all 92 non-Pro residues including His tag have been assigned, except that the first Met cannot be observed in the ^{1}H-^{15}N HSQC spectrum.

Secondary Chemical Shifts Indicate Three Regions with Different Degrees of Propensities for Helical Structure. Deviations of chemical shifts from random-coil values are an excellent indicator of the conformational propensities of the polypeptide chain (*1*−*5*, *30*−*32*). However, there are many sets of random-coil values obtained under different experimental conditions, and using an appropriate reference dataset is imperative to interpret chemical shift data correctly. The dataset determined for Ac-GGXGG-NH_2 peptides in 8 M urea at pH 2.3 (*25*) was adopted as a random-coil reference in the present work, and the corresponding correction factors (*26*) were used to correct local sequence-dependent contributions. In addition, the influence of Pro to the preceding residue has also been taken into account (*27*), and the errors derive from the differences in the acquisition temperature and pH were processed by the inherent correction function of the CSI in NMRView. ^{13}CA is the most sensitive and useful probe for the measure of residual secondary structure in unfolded states. ^{13}CO is also a very good probe for conformational preference, provided that its dependence on local amino acid sequence is corrected. ^{13}CB shifts show systematic differences between helices and sheets but with much more overlap in the observed distributions. Protons including ^{1}HA and ^{1}HN are somewhat less useful for detection of secondary structural propensities because their chemical shifts are influenced to a degree by local dipolar effects, solvent, pH, temperature, etc. (*3*, *32*−*34*). Generally speaking, if the residue is located in the helix, its chemical shifts of ^{1}HA, ^{1}HN, and ^{13}CB will go upfield from random-coil values, while ^{13}CA and ^{13}CO go downfield; in a β sheet or other extended configurations, the situation reverses (*3*, *30*−*32*).

The secondary chemical-shift results for the acid-unfolded hUBF HMG box 1 are shown in Figure 2. Consecutive pronounced downfield secondary ^{13}CA and ^{13}CO shifts are observed for 63−79 residues, indicating that the polypeptide backbone in this region significantly populates the α region of (Φ, Ψ) space. Consistent upfield shifts are observed for ^{13}CB, ^{1}HA, and ^{1}HN in the same region, providing further evidence for the presence of the helical structure. In regions 18−28 and 38−43, the ^{13}CA chemical shifts also go

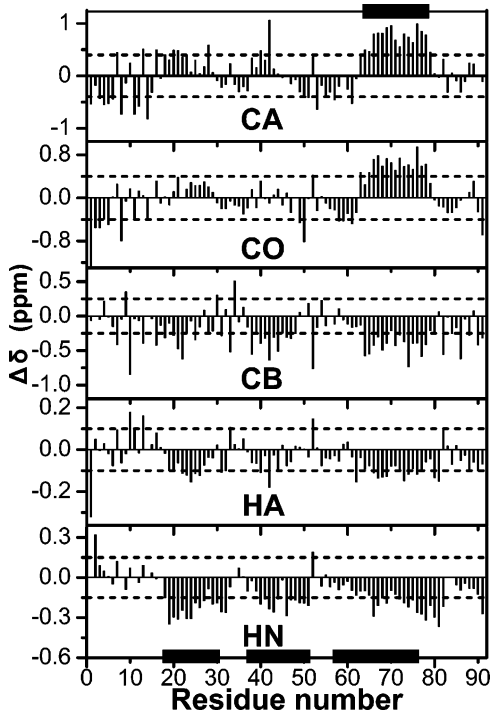

FIGURE 2: Chemical-shift deviations from the sequence-corrected random-coil values (*25*−*27*) for the acid-unfolded hUBF HMG box 1. The black bars at the bottom panel indicate the positions of the helices in the native structure, and the bar at the top panel denotes the helical region determined by the consensus result of CSI for the acid-unfolded state. The broken lines indicate the threshold values for CSI.

downfield with respect to random-coil values. However, relative to 63−79, the ^{13}CA in these two regions go downfield to a less extent, implicating corresponding degree weaker preferences for helical structure in these regions. These structural tendencies of 18−28 and 38−43 are less apparent in regards to ^{13}CO chemical shifts but reinforced by the pronounced upfield secondary ^{1}HA, ^{1}HN, and ^{13}CB shifts. Notably, the region 63−79 corresponds to most parts of the third helix in the native structure but extends beyond the C end of it by 3 more residues, while the regions 18−28 and 38−43 respectively match the N ends of the first and second helix in the native protein. In other regions of the protein, upfield secondary ^{13}CA and ^{13}CO shifts indicate the preference for extended conformations that populate the β region of (Φ, Ψ) space.

Relative Intensities of Sequential NOEs Give Hints for Regular Structures. Different secondary structures are associated with specific patterns of NOEs (*35*). Because of the collapsed chemical-shift dispersion for ^{1}H in denatured protein, some proton resonances in acid-unfolded hUBF HMG box 1 overlap each other severely, which makes it difficult to assign ^{1}H-^{1}H NOEs unambiguously from ^{15}N NOESY−HSQC. The problem can be at least partly resolved by using the HMQC−NOESY−HSQC experiment, which makes use of the intrinsically well-dispersed chemical shifts of ^{15}N even in the unfolded state to resolve ^{1}HN in ^{15}N dimension. Figure 3 shows most observable NOEs in the acid-unfolded hUBF HMG box 1. Both sequential $d_{NN}(i, i + 1)$and $d_{\alpha N}(i, i + 1)$ NOEs are observed almost throughout the sequence, but no long-range NOEs could be unambiguously assigned, reflecting extensive conformational averaging along the polypeptide chain. These results are expected for an unfolded state (*1*−*3*, *5*). Although there are no characteristically long- and medium-range NOEs available to extract conformational information of the regular structure, the relative intensity of sequential NOEs could play a similar role (*2*−*4*, *36*−*38*). In general, $d_{NN}(i, i + 1)$ NOEs are particularly strong in folded helix, while $d_{\alpha N}(i, i + 1)$ NOEs are relatively more intense in a β sheet and other extended conformations. In the region of 63−79, most available sequential $d_{NN}(i, i + 1)$ NOEs are among the strongest, in accordance with the significant propensity for a helical structure of this region indicated by the secondary chemical-shift results. In two other regions that possess a weaker preference for helical structure indicated by chemical shifts, the intensities of the $d_{NN}(i, i + 1)$ NOEs are mostly among the strongest or substrongest accordingly. What deserves special attention is the region between residue 79 and 82. It immediately follows the helical region 63−79, and sequential 79−80 and 80−81 $d_{NN}(i, i + 1)$ NOEs are likewise among the strongest. Whereas the ^{13}CA and ^{13}CO chemical shifts show that the preceding helical structure fades away in this region toward the following extend structure. In particular, there are two $i-i+2$ NOE connections between E80 and H82. A reasonable interpretation for these results is that the region 79−82 forms a turn-like structure as the boundary of the preceding helical structure (*36*). This assumption is supported by the later determination of temperature coefficients. In addition, strangely, at the C terminal of the sequence, the intensities of the sequential $d_{NN}(i, i + 1)$ NOEs are abnormally of medium intensity.

Conformational Averaged Coupling Constants Are Insensitive to Local Structural Preference. Conformational averaging in unfolded states frequently results in intermediate values of $^{3}J_{HNHA}$ that are not particularly valuable as diagnostics of secondary structural propensities. Thus, coupling constants are rarely as useful as chemical shifts in the detection of residual structure (*1*, *3*−*4*). This is exactly true in the case of the acid-unfolded hUBF HMG box 1, which is illustrated by Figure 4. Because of resonance overlap, coupling constants for several residues could not be correctly measured. Of those available $^{3}J_{HNHA}$, most are located in the range between 6.5 and 8.5 Hz, indicating a conformational averaged extended structure that populates the β region of (Φ, Ψ) space (*39*−*40*). However, in the regions 64−78 and 18−28, although limited coupling constants are available because of resonance overlap, the negative or slightly positive deviations from random-coil values (*41*) of them still somewhat imply that these regions have a stronger tendency to populate the α region of (Φ, Ψ) space, in agreement with the secondary chemical-shift results for these regions.

Reduced Temperature Coefficients and Curved Temperature Dependence Signifying Ordered Structure and Conformational Heterogeneity. Amide proton temperature coefficients can provide valuable information about hydrogen-bonding interactions and solvent sequestration in unfolded proteins. The chemical shift of an amide proton that is protected from exchange with solvent protons by hydrogen bonds is less sensitive to changes in temperature than that of a proton that can exchange with the bulk solvent. Abnormally low-temperature coefficients relative to random-coil values are an indication of local structure and interaction (*3*, *42*−*44*). The upper panel of Figure 5 shows the deviations

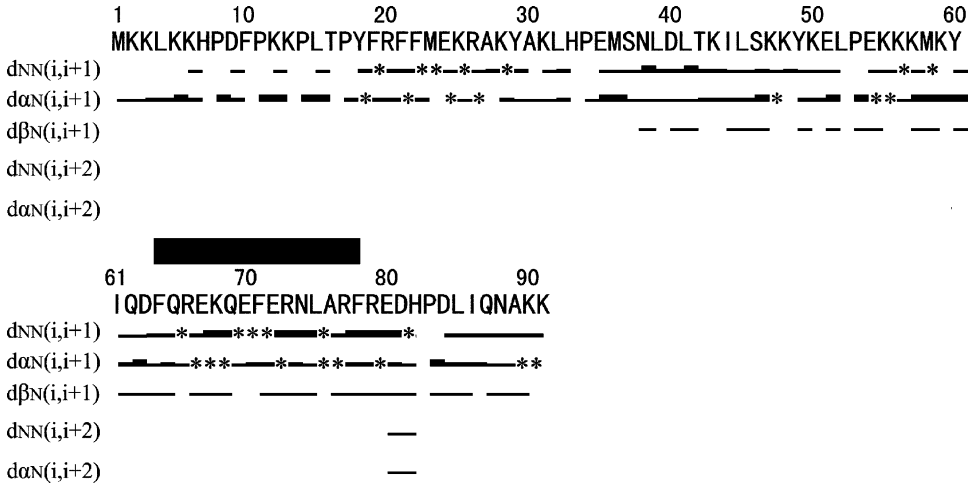

FIGURE 3: Summary of interresidual ^{1}H-^{1}H NOEs observed in the acid-unfolded hUBF HMG box 1. The thickness of the bars indicates the relative intensities of the NOEs. Asterisks denote NOEs that are not observable or ambiguous because of resonance overlap. The bar above the sequence number indicates the position of the helical structure in the acid-unfolded protein determined by CSI.

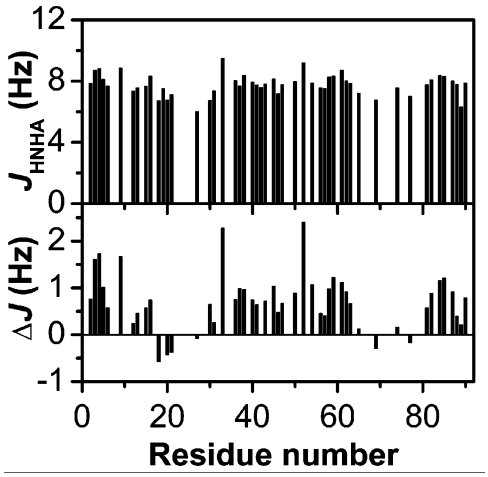

FIGURE 4: $^3J_{HNHA}$ coupling constants obtained from the HNHA experiment and the deviations from predicted random-coil values (*41*) for the acid-unfolded hUBF HMG box 1.

from random-coil values of ^{1}HN temperature coefficients (*45*) measured in the temperature range from 21 to 45 °C at 3 °C intervals. Over this range, the far-UV CD spectrum of acid-unfolded hUBF HMG box 1 changes little (data not shown), suggesting that structural alterations are minimal. The regions 67−82, 19−33, and to a less extent 39−42 all display consecutive reduced temperature coefficients, supporting the previous conclusion that there are some nonrandom structures around these regions. Abnormal temperature dependences of amide protons were observed for residues D81 and H82, which are shown in the middle and bottom panels of Figure 5. Curved temperature dependence is a symbol of conformational heterogeneity, i.e., exchange between alternative states (*44*). As previously presumed, the region 79−82 probably forms a turn-like structure following the end of helical structure 64−78 in the acid-unfolded hUBF HMG box 1. Thus, the curved temperature dependences could arise from the inherent instability of a boundary segment.

Dynamic Data Show Motional Restrictions for Residual Structures. NMR relaxation experiments have the potential to provide valuable insights into the internal molecular dynamics of unfolded and partly folded states. Local variations in backbone dynamics are correlated with propensities

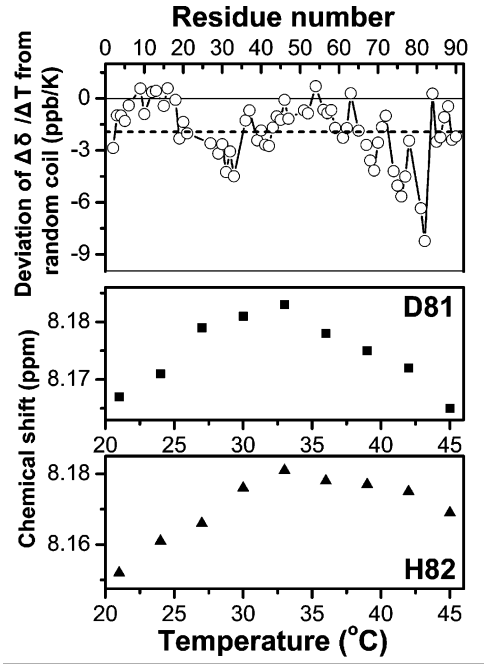

FIGURE 5: Temperature coefficient deviations from random-coil values (*45*) (upper panel) for the acid-unfolded hUBF HMG box 1 and the chemical shifts of D81 (middle panel) and H82 (bottom panel) as functions of the temperature. The broken line indicates the average value of all deviation values.

for local compaction of the polypeptide chain that results in restriction of backbone motions (*1*−*5*). Backbone dynamics are most commonly investigated by measurements of R_1 and R_2 relaxation rates and ^{1}H-^{15}N NOE in uniformly ^{15}N-labeled protein. These data are sensitive to motions on a range of time scales. The transverse relaxation rate R_2 is more sensitive to lower frequency (nanosecond) motions and also reflects contributions from slower milli- to microsecond exchange processes. Whereas, the ^{1}H-^{15}N NOE is typically most sensitive to higher-frequency (picosecond) motions of the backbone, with lower values indicating increased local flexibility of the polypeptide. The R_1 relaxation rate by itself does not discriminate effectively between faster and slower motions within the pico- to nanosecond time scale (*3*−*4*, *46*).

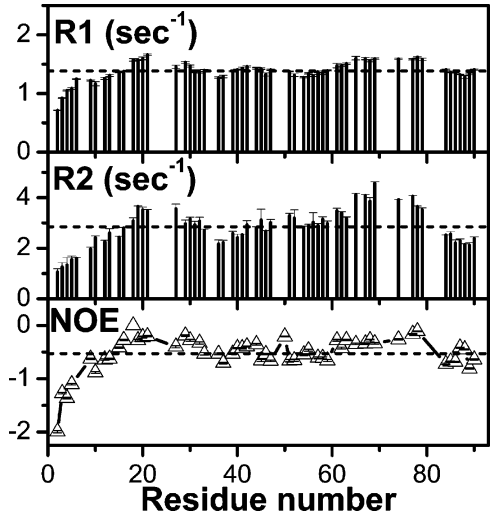

FIGURE 6: ^{15}N R_1 and R_2 relaxation rates and ^1H-^{15}N NOEs for the acid-unfolded hUBF HMG box 1 at 500 MHz. The broken lines indicate the average values of R_1 and R_2 rates and ^1H-^{15}N NOEs.

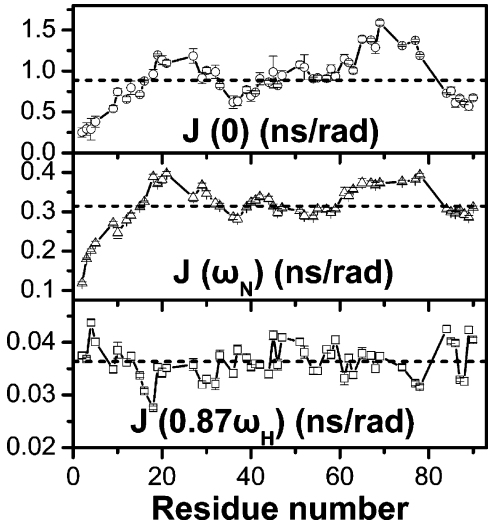

FIGURE 7: Spectral density function $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ values for the acid-unfolded hUBF HMG box 1 at 500 MHz. The broken lines indicate the average values of $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$.

The relaxation rates R_1 and R_2 and the heteronuclear ^1H-^{15}N NOE have been measured for all nonoverlapped cross-peaks in the HSQC spectrum at 500 MHz, and the results are shown in Figure 6. Three relaxation parameters all fluctuate little along the sequence, except for at the N terminal, where relaxation is decreased as a result of increased conformational flexibility (the C terminal His tag has not been plotted in the graph). In addition, all of the heteronuclear ^1H-^{15}N NOEs are negative. These results indicate large amplitude fluctuations of most of the polypeptide chain on a subnanosecond time scale, which is expected for a highly expanded state (*46*−*48*). However, small degrees of motional restrictions are presented in the regions 61−78 and 18−32, illustrated by relatively larger R_2 and R_1 and to a less extent ^1H-^{15}N NOEs.

Analysis of the dynamics of folded proteins is commonly interpreted within the framework of the model-free formalism (*49*). However, because of the wide variation in local correlation times between residues in the more structured regions and those in the less structured and unstructured regions, it is unlikely that the model-free approach can provide accurate analysis of the data for unfolded states. A reduced spectral density method avoids the assumptions inherent in model-free analysis that the molecule tumbles isotropically with a single correlation time τ_m by direct mapping of the spectral density (*3*−*5*, *29*). The values of $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ determined to sample the spectral density at 0, 50, and 435 MHz are shown in Figure 7. $J(0.87\omega_H)$ reflects the highest frequency motions and is largely determined by the heteronuclear NOEs. $J(\omega_N)$ is most sensitive to R_1 and, like R_1, does not discriminate well between faster and slower motions. $J(0)$ is most sensitive to R_2, reflecting slower (nanosecond) motions, with possible contributions from exchange processes on the milli- to microsecond time scale (*3*, *29*, *46*). In accordance with the heteronuclear ^1H-^1H NOEs data, the spectral density function of $J(0.87\omega_H)$ is flat and extended, indicating that high-frequency fluctuations occur throughout the whole peptide backbone. While, corresponding to the relaxation rates R_1 and R_2 data, consecutive increased values of $J(\omega_N)$ and $J(0)$ in some regions implicate less mobility on slower motions in these segments. The motional restrictioned regions agree well with the residual local structures determined by other NMR parameters, and the region between residues 64 and 78 behaves slightly more like a rigid segment than other regions.

DISCUSSION

Residual Nativelike Structure in the Acid-Unfolded hUBF HMG Box 1. As appropriate, random-coil shift reference values (*25*) were used and sequence-dependent corrections (*26*−*27*) were made. The secondary chemical-shift data for the acid-unfolded hUBF HMG box 1 show convincing evidence for the substantial propensity for helical structure in the region 64−78 (Figure 2). However, significant deviations of chemical shifts from random-coil values just indicate preferential population of the α regions of (Φ, Ψ) space, rather than the presence of rigid folded structures (*1*−*5*). Thus, the fact that there is no characteristic medium- and long-range NOEs of α helix observed in this helical region (Figure 3) is logical. Nonetheless, the increased preference for helical conformation is still strongly implicated by the appearance of the strongest sequential $d_{NN}(i, i + 1)$ NOEs in the region 64−78. Furthermore, from the available data for this region, negative or weaker positive deviations of $^3J_{HNHA}$ as well as greater negative deviations of the temperature coefficient from corresponding random-coil values give complementary evidence for the propensity for the helical structure. The population of dihedral angles in the α and β regions of (Φ, Ψ) space can be calculated from the magnitude of the secondary chemical shifts for defined regions (*28*, *32*, *47*). The mean populations of helical conformation for the region 64−78 estimated from ^{13}CA and ^{13}CO are 27 and 30%, respectively (using helical secondary shifts of 2.8 and 2.1 ppm for ^{13}CA and ^{13}CO). While, a larger figure 35% is obtained when the populations of the helix are estimated from ^1HA (helical secondary shifts of 0.3), probably because of its sensitivity to other influences rather than secondary structure.

A less substantial preference for the helical structure is exhibited in other regions in the acid-unfolded hUBF HMG

box 1. In the region between residues 18 and 28, the interresidual NOEs data (Figure 3) show only slight propensity for the helical structure. However, secondary chemical shifts, $^3J_{HNHA}$, and temperature coefficients in this region are more or less similar with those in the region 64−78 (Figure 4), suggesting a little weaker propensity for helical conformation of the region 18−28. The preference for helical structure in the region of residues between 38 and 43 is even more ambiguous than the region 18−28. Its conformational propensity is only mildly indicated by secondary chemical shifts, ^{1}H-^{1}H NOEs, and amide proton temperature coefficients. All of the three helical regions in the acid-unfolded hUBF HMG box 1 correspond to a portion of the three helixes in the folded protein, respectively, implicating that the former could be initiation sites for the formation of the latter in the protein-folding process, and the different degree propensities for the helical structure of them probably reflect the different rates of their formation during the folding reaction.

Residual Non-native Structures. The interresidual NOEs data for the acid-unfolded hUBF HMG box 1 (Figure 3) show that the strongest sequential $d_{NN}(i, i + 1)$ NOEs are not only presented in the region 64−76 (a portion of the third helix in native folded structure) but also presented in the region 77−81. This implies that in the acid-unfolded state the 4−5 residues immediately following the third helix have considerable propensities to form a helix, which is also exhibited in the native state (*21*). However, the secondary chemical shifts for ^{13}CA and ^{13}CO change from downfield for residues 64−79 to upfield for residues 80−82. Particularly, there are two $i-i+2$ NOE connections existed between E80 and H82. Probably, the residues 79−82 form a turn-like structure as the boundary between the preceding helical structure and the following extended structure (*35*−*36*, *50*), and the curved temperature dependence of D81 and H82 (Figure 5) is an embodiment of their conformational exchange between turn-like and extended structures (*44*). In addition, strangely, besides the three helical regions noted above, the C terminal of the sequence exhibits a very mild propensity for helical structure, which is slightly presented by its downfield ^{1}HA and ^{1}HN chemical shifts, intermediate intense $d_{NN}(i, i + 1)$, and negative temperature coefficient deviations from random-coil values.

Turns as well as nascent helices are excellent candidates for the initiation of protein folding and are frequently populated in unfolded states of proteins (*2*, *36*, *50*−*51*). The appearance of the non-native turn-like structure 79−82 and the very faint helical C end in the acid-unfolded hUBF HMG box 1 may provide one more example for the above viewpoint and supports the argument that short segments of local structure are not necessarily presented in the final folded form of the protein (*36*). The intrinsically plastic nature of both native and non-native structures in unfolded proteins would allow local interactions determined by the amino acid sequence bias the conformational search toward specific structured forms, even in the absence of stabilization by long-range interactions with the remainder of the protein (*2*, *4*, *51*). As the protein folding proceeds, the non-native initiation structures would dissolve to form native folded structures. However, on the other hand, the non-native structures in the acid-unfolded hUBF HMG box 1 may be the results of the non-native pH and the influence from the His tag and will not appear in the kinetic intermediate states on the actual protein-folding pathway.

The formation of both native and non-native residual structures is likely to be the result of inherent conformational tendency and near local interactions of residues (*2*, *50*), because there is no long-range interaction unambiguously identified between these regions from any NMR parameters including dynamic data.

Backbone Dynamics. Like many other unfolded proteins that possess little ordered structures (*12*, *47*), the acid-unfolded hUBF HMG box 1 is characterized by relatively uniform backbone dynamics (Figure 6). The negative ^{1}H-^{15}N NOEs immediately signify that the polypeptide backbone undergoes large amplitude fluctuations on a subnanosecond time scale throughout its sequence, in accordance with the flat $J(0.87\omega_H)$ spectral density function that is most sensitive to picosecond time-scale motions (*29*, *46*). Nevertheless, larger than average values of relaxation parameters are observed in several regions (Figure 6), implicating restrictions of slower motions. These motional restrictions are more evident from the $J(\omega_N)$ and $J(0)$ data (Figure 7).

Larger than average values of $J(\omega_N)$ and $J(0)$ are observed in the regions 60−78 and 18−31 and to a less extent in the region 38−43, indicating restrictions of subnanosecond time-scale motions. The same conclusion can be drawn from the results of R_1 and R_2. Relative to other residual structural regions in the acid-unfolded hUBF HMG box 1, in 64−78, the $J(0)$ values are larger than expected on the basis of $J(\omega_N)$ and $J(0.87\omega_H)$, reflecting contribution from slow motion on the milli- to microsecond time-scale exchange processes. All of the above regions that exhibit apparent motional restrictions correspond well to the residual helical structures determined by secondary chemical shifts and other NMR parameters, which indicate the association between motional restriction and local structure. The slow motion on the milli- to microsecond time scale of the region 64−78 may be attributed to its relatively stronger propensity for a regular structure. Besides the regions where residual secondary structures are present, there are mild motional restrictions on the exchange process time scale at positions around residues Y49 and Y61, which are exhibited by the slightly larger than average $J(0)$ and R_2 values. Y49, Y61, as well as F21 are the main contributors of the hydrophobic core in the native structure (*20*); therefore, the motional restrictions of the former two residues may arise from transient long- or short-range hydrophobic interactions (*17*). The hydrophobic interactions may also explain the fact that although the tendency for the helical structure of the region around F21 is much less substantial than the region 64−78, it possesses motional restriction very close to the latter. All motional restrictions can be observed from ^{1}H-^{15}N NOEs data too, just to a much less extent.

Implications for the Protein Folding. In a previous work, we have presumed a folding pathway for hUBF HMG box 1 on the basis of the acid-unfolding behavior of it (*22*). In this potential folding pathway, the acid-unfolded state appears before the MG-like state formed at pH 0.2 in the early stage. CD data show that there are still some local structures retained in the acid-unfolded protein, which is confirmed by NMR data in the present work. These residual secondary structures respectively correspond to a portion of the three

helixes in the native structure. Besides nativelike secondary structures, there is also a non-native turn-like structure in the acid-unfolded protein. Probably, these residual structures act as initiation sites in the protein-folding process (*2*, *4*, *36*, *50*−*51*). When the folding proceeds, these nascent helices and turn-like structure may develop into nativelike secondary structures such as those found in the MG-like state formed at pH 0.2. The positions of residual helical structure in the acid-unfolded protein are far away from each other in the native structure, and no obvious evidence shows that there is tertiary interaction between them in the acid-unfolded state, suggesting that the tendency toward the secondary structure of residues is governed by their intrinsic nature and local rather than long distance interactions (*2*, *50*). This is also supported by the fact that the near-UV CD spectrum of the acid-unfolded hUBF HMG box 1, like that of the GdnHCl fully unfolded state, shows no rigid tertiary structure (*22*). However, deduced from the dynamic data, weak hydrophobic interactions may exist transiently between side chains of some aromatic residues during the conformational search process. To detect such short-lived tertiary interactions as well as characteristic medium- and long-range NOEs in the helical structure, using more specific and sensitive NMR experiments and searching for more favorable conditions that can stabilize the low populated structures are needed.

On the basis of the result presented herein plus the previous data (*22*), the events that happened along the protein-folding process may be proposed as follows. Driven by the intrinsic tendency of local structure formation, residues around 64−78, 18−28, and 38−43 sample helical conformation to a more or less extent. Although unstable, these nascent secondary structures as well as short-lived hydrophobic interactions significantly reduce the space of conformational search (*2*, *50*, *51*), which make the growing up of new secondary structures based on the formed structure easier. In other words, these residual structures may act as initiation sites, on to which the formation of further helical structures and thereafter/or at the same time the packing of the tertiary structure take place. This structural development could be achieved by a diffusion-collision mechanism. Whereas, the non-native structures will melt during the conformational search process as the protein folding proceeds. At a later folding stage, the secondary and tertiary structures condense and the protein could form a MG-like state, like that formed at pH 0.2, which will rearrange the orientations of its secondary structures and finally collapse into the native folded structure.

In conclusion, in the current work, we have identified several regions that have propensities to form a nativelike helical structure or non-native turn-like structure in acid-unfolded hUBF HMG box 1. These residual helical regions in addition to some hydrophobic residues present different degrees of motional restrictions, which are indicated by dynamic data. Presumably, these nascent helices may play an important role in the initiation of protein folding to form correct secondary structures and then a correct fold of the whole molecule. Nevertheless, because our presumption is based on the results from the samples under non-native conditions at equilibrium, to obtain more detailed and exact information of the folding pathway of hUBF HMG box 1 awaits further especially kinetic studies.

ACKNOWLEDGMENT

We thank Dr. F. Delaglio and Prof. A. Bax for providing the software NMRPipe, Prof. T. D. Goddard and Prof. D. G. Kneller for providing Sparky, and Dr. B. A. Johnson for providing NMRView.

REFERENCES

1. Shortle, D. R. (1996) Structural analysis of non-native states of proteins by NMR methods, *Curr. Opin. Struct. Biol. 6*, 24−30.
2. Dyson, H. J., and Wright, P. E. (1998) Equilibrium NMR studies of unfolded and partially folded proteins, *Nat. Struct. Biol. 5*, 499−503.
3. Dyson, H. J., and Wright, P. E. (2001) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states, *Methods Enzymol. 339*, 258−270.
4. Juneja, J., and Udgaonkar, J. B. (2003) NMR studies of protein folding, *Curr. Sci. 84*, 157−172.
5. Dyson, H. J., and Wright, P. E. (2004) Unfolded proteins and protein folding studied by NMR, *Chem. Rev. 104*, 3607−3622.
6. Wuthrich, K. (1994) NMR assignment as a basis for structural characterization of denatured states of globular proteins, *Curr. Opin. Struct. Biol. 4*, 93−99.
7. Ptitsyn, O. B. (1995) Structures of folding intermediates, *Curr. Opin. Struct. Biol. 5*, 74−78.
8. Brockwell, D., Smith, D. A., and Radford, S. E. (2000) Protein folding mechanisms: New methods and emerging ideas, *Curr. Opin. Struct. Biol. 10*, 16−25.
9. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) Structural and dynamic characterization of partially folded states of myoglobin and implications for protein folding, *Nat. Struct. Biol. 5*, 148−155.
10. Eliezer, D., Chung, J., Dyson, H. J., and Wright, P. E. (2000) Native and non-native structure and dynamics in the pH 4 intermediate of apomyoglobin, *Biochemistry 39*, 2894−2901.
11. Meekhof, A. E., and Freund, S. M. V. (1999) Probing residual structure and backbone dynamics on the milli- to picosecond timescale in a urea-denatured fibronection type III domain, *J. Mol. Biol. 286*, 579−592.
12. Frank, M. K., Clore, G. M., and Gronenborn, A. M. (1995) Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy, *Protein Sci. 4*, 2605−2615.
13. Peti, W., Smith, L. J., Redfield, C., and Schwalbe, H. (2001) Chemical shifts in denatured proteins: Resonance assignments for denatured ubiquitin and comparisons with other denatured proteins, *J. Biomol. NMR 19*, 153−165.
14. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., Mcbay, D. L., Mossakowska, D. E., Smith, R. A. G., Dobson, C. M., and Smith, L. J. (1997) NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein, *J. Mol. Biol. 274*, 152−159.
15. Buevich, A. V., Shinde, U. P., Inouye, M., and Baum, J. (2001) Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies, *J. Biomol. NMR 20*, 233−249.
16. Zhang, O., and Forman-Kay, J. D. (1995) Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer, *Biochemistry 34*, 6784−6794.
17. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1995) Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in solution in aqueous buffer, *Biochemistry 34*, 868−878.
18. Jantzen, H. M., Admon, A., Bell, S. P., and Tjian, R. (1990) Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins, *Nature 344*, 830−836.
19. Jantzen, H. M., Chow, A. M., King, D. S., and Tjian, R. (1992) Multiple domains of the RNA polymerase I activator hUBF interact with the TATA-binding protein complex hSL-1 to mediate transcription, *Genes Dev. 6*, 1950−1963.
20. Zomerdijk, J. C. B. M., and Tjian, R. (1998) Initiation of transcription on human rRNA genes, in *Transcription of Ribosomal RNA Genes by Eukaryotic RNA Polymerase I* (Paule, M. R., Ed.) pp 121−134, Springer-Verlag, New York.

21. Xu, Y., Yang, W., Wu, J., and Shi, Y. (2002) Solution structure of the first HMG box domain in human upstream binding factor, *Biochemistry 41*, 5415−5420.
22. Zhang, X., Zhang, J., Li, X., Xu, J., Huang, H., Chen, Q., Wu, J., and Shi, Y. (2005) Compact molten globule-like state of hUBF HMG box 1 at extremely low pH, *Biochim. Biophys. Acta 1748*, 66−73.
23. Yang, W., Zeng, W., Zhou, D., and Shi, Y. (2002) Cloning, expression, secondary structure characterization of HMG box 1 of hUBF from *E. coli* and its binding to DNA, *Biochim. Biophys. Acta 1598*, 147−155.
24. Panchal, S. C., Bhavesh, N. S., and Hosur, R. V. (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for H^N and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins: Application to unfolded proteins, *J. Biomol. NMR 20*, 135−147.
25. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Wright, P. E., and Dyson, H. J. (2000) Random coil chemical shifts in acidic 8 M urea: Implementation of random coil chemical shift data in NMRView, *J. Biomol. NMR 18*, 43−48.
26. 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.
27. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^1H, ^{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.
28. Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, H. 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.
29. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) Spectral density function mapping using ^{15}N relaxation data exclusively, *J. Biomol. NMR 6*, 153−162.
30. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure, *J. Mol. Biol. 222*, 311−333.
31. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: A fast and simple method for the assignment of proetein secondary structure through NMR spectroscopy, *Biochemistry 31*, 1647−1651.
32. Wishart, D. S., and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination, *Methods Enzymol. 239*, 363−392.
33. Spera, S., and Bax, A. (1991) Empirical correlation between protein backbone conformation and Cα and Cβ ^{13}C nuclear magnetic resonance chemical shifts, *J. Am. Chem. Soc. 113*, 5490−5492.
34. Yao, J., Dyson, H. J., and Wright, P. E. (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins, *FEBS Lett. 419*, 285−289.
35. Wuthrich, K. (1986) NOE-observable ^1H-^1H distances in proteins, in *NMR of Proteins and Nucleic Acids*, pp 117−129, John Wiley and Sons, New York.
36. Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J., and Wright, P. E. (1993) Peptide models of protein folding initiation sites: Secondary structure formation by peptides corresponding to G- and H-helices of myoglobin, *Biochemistry 32*, 6337−6347.
37. Wong, K. B., Freund, S. M. V., and Fersht, A. R. (1996) Cold denaturation of barst: ^1H, ^{15}N, and ^{13}C NMR assignment and characterization of residual structures, *J. Mol. Biol. 259*, 805−818.
38. Freund, S. M. V., Wong, K. B., and Fersht, A. R. (1996) Initiation sites of protein folding by NMR analysis, *Proc. Natl. Acad. Sci. U.S.A. 93*, 10600−10603.
39. Dyson, H. J., and Wright, P. E. (1991) Defining solution conformations of small linear peptides, *Annu. Rev. Biophys. Biophys. Chem. 20*, 519−538.
40. Alexandrescu, A. T., Abeygunawardana, C., and Shortle, D. (1994) Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: A heteronuclear NMR study, *Biochemistry 33*, 1063−1072.
41. Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M., and Dobson, C. M. (1996) Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations, *J. Mol. Biol. 255*, 494−506.
42. Andersen, N. H., Neidigh, J. W., Harris, S. M., Lee, G. M., Liu, Z., and Tong, H. (1997) Extracting information from the temperature gradients of peptide NH chemical shifts. 1. The importance of conformational averaging, *J. Am. Chem. Soc. 119*, 8547−8561.
43. Baxter, N. J., and Williamson, M. P. (1997) Temperature dependence of ^1H chemical shifts in proteins, *J. Biomol. NMR 9*, 359−369.
44. Baxter, N. J., Hosszu, L. L. P., Waltho, J. P., and Williamson, M. P. (1998) Characterisation of low-energy exited states of folded proteins, *J. Mol. Biol. 284*, 1625−1639.
45. Merutka, G., Dyson, H. J., and Wright, P. E. (1995) "Random coil" ^1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series, GGXGG, *J. Biomol. NMR 5*, 14−24.
46. Key, L. E., Torchia, D. A., and Bax. A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease, *Biochemistry 28*, 8972−8979.
47. Bai, Y., Chung, J., Dyson, H. J., and Wright, P. E. (2001) Structural and dynamic characterization of an unfolded state of poplar apoplastocyanin formed under nondenaturing conditions, *Protein Sci. 10*, 1056−1066.
48. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1997) Characterization of the backbone dynamics of folded and denatured states of an SH3 domain, *Biochemistry 36*, 2390−2492.
49. Lipari, G., and Szabo, A. (1982) Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J. Am. Chem. Soc. 104*, 4546−4559.
50. Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) Conformation of peptide fragments of proteins in aqueous solution: Implication for initiation of protein folding, *Biochemistry 27*, 7167−7175.
51. Finkelstein, A. V., and Galzitskaya, O. V. (2004) Physics of protein folding, *Phys. Life Rev. 1*, 23−56.

BI0501939