Native and nonnative conformational preferences in the urea-unfolded state of barstar

NEEL S. BHAVESH,[1] JUHI JUNEJA,[2] JAYANT B. UDGAONKAR,[2] AND RAMAKRISHNA V. HOSUR[1]

[1]Department of Chemical Sciences, Tata Institute of Fundamental Research, Mumbai 400 005, India
[2]National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560 065, India

(RECEIVED April 8, 2004; FINAL REVISION July 12, 2004; ACCEPTED August 14, 2004)

Abstract

The refolding of barstar from its urea-unfolded state has been studied extensively using various spectroscopic probes and real-time NMR, which provide global and residue-specific information, respectively, about the folding process. Here, a preliminary structural characterization by NMR of barstar in 8 M urea has been carried out at pH 6.5 and 25°C. Complete backbone resonance assignments of the urea-unfolded protein were obtained using the recently developed three-dimensional NMR techniques of HNN and HN(C)N. The conformational propensities of the polypeptide backbone in the presence of 8 M urea have been estimated by examining deviations of secondary chemical shifts from random coil values. For some residues that belong to helices in native barstar, $^{13}C_{\alpha}$ and ^{13}CO secondary shifts show positive deviations in the urea-unfolded state, indicating that these residues have propensities toward helical conformations. These residues are, however, juxtaposed by residues that display negative deviations indicative of propensities toward extended conformations. Thus, segments that are helical in native barstar are unlikely to preferentially populate the helical conformation in the unfolded state. Similarly, residues belonging to β-strands 1 and 2 of native barstar do not appear to show any conformational preferences in the unfolded state. On the other hand, residues belonging to the β-strand 3 segment show weak nonnative helical conformational preferences in the unfolded state, indicating that this segment may possess a weak preference for populating a helical conformation in the unfolded state.

Keywords: unfolded state; HNN; HN(C)N; secondary chemical shifts

Supplemental material: see www.proteinscience.org

A complete understanding of protein folding requires the characterization of all species populated along the folding coordinate, including the unfolded state, partially folded intermediate states, and the fully folded native state. The importance of the unfolded state ensemble in guiding the folding process is now being realized. Unfolded states of proteins, such as in high concentration of chemical denaturant, have been modeled traditionally as statistical random coils in which local interactions dominate conformational behavior (Smith et al. 1996). Molecular dynamic simulations of unfolded polypeptides in the past few years have, however, indicated that steric clashes can prevent a protein chain from adopting any arbitrary conformation, thus constraining the unfolded state ensemble (Pappu et al. 2000; Van Gunsteren et al. 2001), and probably favoring conformations having native-like dihedral angles (Shortle 2002). Recently, it has been suggested that the average interresidue distances in

Reprint requests to: Jayant B. Udgaonkar, National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560 065, India; e-mail: jayant@ncbs.res.in; fax: +91-80-23636662; or Ramakrishna V. Hosur, Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India; e-mail: hosur@tifr.res.in; fax: +91-22-22804610.

Abbreviations: NMR, nuclear magnetic resonance; HSQC, heteronuclear single-quantum coherence.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.04805204.

collapsed unfolded structures agree well with the statistics of the ideal random-flight chain (Zagrovic and Pande 2003).

There is increasing evidence now from high-resolution NMR experiments that unfolded states of proteins are not completely random (Juneja and Udgaonkar 2003). The parameters and time scales of these NMR experiments can provide considerable information about the ensemble-averaged structural properties of unfolded proteins (Shortle 1996; Dyson and Wright 2001, 2002). One type of structure detected by NMR in unfolded proteins is a hydrophobic cluster, usually formed by local side-chain interactions. The first reported instance of a hydrophobic cluster populated in high urea concentration was in the case of the N-terminal domain of the 434-repressor (Neri et al. 1992). More recently, hydrophobic clusters in urea-unfolded states have been reported for rat intestinal fatty acid binding protein (Hodsdon and Frieden 2001) and apomyoglobin (Lietzow et al. 2002; Schwarzinger et al. 2002). Similarly, hydrophobic clusters were also observed in HIV-1 protease tethered dimer unfolded by guanidine hydrochloride (Bhavesh et al. 2003). For these proteins, the initial collapse of the polypeptide chain that occurs when refolding is initiated by dilution of denaturant is likely to be a consequence of transient interactions between hydrophobic clusters that persist in the unfolded state. Another type of structure detected by NMR in unfolded proteins is fluctuating secondary structure, for example in urea-unfolded barnase (Arcus et al. 1995; Wong et al. 2000). It has been suggested that the folding of barnase is initiated around the native-like local structures that persist in the unfolded state until enough stabilizing long-range interactions form a nucleation site onto which further secondary structure formation and final native packing of tertiary structure occur (Wong et al. 2000).

For some proteins, multiple types of structures have been detected in the unfolded state. For example, for the drkN SH3 domain, both the unfolded state under folding conditions (U_{exch}) and the unfolded state in guanidine hydrochloride are reasonably compact (Farrow et al. 1995, 1997; Choy et al. 2002). Detailed NMR studies of U_{exch} (Mok et al. 1999; Crowhurst et al. 2002; Crowhurst and Forman-Kay 2003) indicate the presence of multiple structures ranging from conformers with nonnative structure possessing long-range contacts to those with more compact structures maintaining native-like secondary structure. Consequently, it has been suggested that drkN SH3 might fold via multiple folding pathways, conformers with nonnative hydrophobic clusters folding via a hydrophobic collapse mechanism and those with native-like structures folding via a hierarchical condensation mechanism (Crowhurst and Forman-Kay 2003). On the other hand, when the unfolded state appears to be largely unfolded, as determined to be so for the chymotrypsin inhibitor 2 (Kazmirski et al. 2001), the protein is thought to fold in a more cooperative manner with the establishment of long-range contacts linked to collapse and secondary-structure formation (Kazmirski et al. 2001). Clearly, identification of any residual structure and hydrophobic clusters in the unfolded state of a protein plays a crucial role toward obtaining insights into its folding mechanism (Fersht and Daggett 2002; Daggett and Fersht 2003; Gianni et al. 2003).

NMR studies appear to indicate the persistence of native-like structure in the unfolded states of hen lysozyme (Klein-Seetharaman et al. 2002) and the Δ131Δ fragment of staphylococcal nuclease (Shortle and Ackerman 2001; Ackerman and Shortle 2002). The native-like core in the unfolded state of hen lysozyme is stabilized by the involvement of a tryptophan residue in nonnative long-range tertiary interactions (Schwalbe et al. 1997; Klein-Seetharaman et al. 2002). In the case of the Δ131Δ fragment of staphylococcal nuclease, which has been characterized extensively as a model for the unfolded state of the protein under nondenaturing conditions (Alexandrescu et al. 1994), the values observed for the dipolar couplings have been interpreted to indicate a native-like topology that might arise from local steric interactions between side chains and the protein backbone (Srinivasan and Rose 1999; Pappu et al. 2000; Ohnishi and Shortle 2003). More recently, however, this interpretation has been questioned because it has been shown that nonzero residual dipolar couplings of denatured states can also arise from chain-like molecules without a defined three-dimensional (3D) spatial organization (Louhivuori et al. 2003). The topology of the native state is a major determinant of the folding rate (Clementi et al. 2000; Makarov et al. 2002; Miller et al. 2002), and if native-state topologies are indeed present in the unfolded states of these two proteins, it would imply that the entropic cost of bringing two regions of the polypeptide chain together so that they can interact is borne by the unfolded state itself.

Local structural features such as propensity of the polypeptide backbone to populate preferentially the α or β region of φ, ψ space, or to exist in a random coil conformation can be derived from chemical shifts (Wishart and Sykes 1994). Here, we report NMR data on backbone conformational propensities in the unfolded state of barstar in 8 M urea (pH 6.5) at 25°C. Barstar is an 89-amino-acid residue, single-domain α/β protein. Its native structure has been solved by NMR, and consists of three parallel helices packed against a three-stranded parallel β-sheet and a fourth more poorly defined helix that connects the second β-strand and third major helix (Fig. 1; Lubienski et al. 1994). The refolding of barstar from high denaturant concentrations has been characterized extensively, and exhibits multistate kinetics (Schreiber and Fersht 1993; Agashe et al. 1995; Shastry and Udgaonkar 1995; Bhuyan and Udgaonkar 1999; Killick et al. 1999; Sridevi et al. 2000). It is therefore important to get structural information about the unfolded state of the protein to understand the early events of folding. In

this study of the unfolded state of barstar in 8 M urea, secondary chemical shifts indicate that segments that are helical in the native protein do not show any preference for helical structure in the unfolded state. Residues in β-strand 3, however, show nonnative helical conformational preferences. The implications of these observations for the folding of barstar are discussed.

Results and Discussion

Backbone resonance assignments of barstar in 8 M urea

Structural characterization by NMR is often hampered in unfolded and partially folded states of proteins because of the poor dispersion of amide and carbon resonances in these forms. Although this problem is overcome in favorable cases by carrying out standard triple resonance and NOESY experiments on a ^{15}N- and ^{13}C-labeled protein sample (Meekhof and Freund 1999; Schwarzinger et al. 2002), these experiments were not successful in obtaining sequential assignments of guanidine-unfolded HIV-protease (Bhavesh et al. 2001). The novel 3D NMR experimental procedures of HNN and HN(C)N provide a new protocol for the sequential assignment of both folded and unfolded forms of proteins (Panchal et al. 2001). These methods use better ^{15}N dispersion along two of the three dimensions in the 3D triple resonance experiments and have been used successfully to assign backbone amide resonances in unfolded HIV-protease (Bhavesh et al. 2001) and to study the aggregated molten globule state of barstar at pH 3 (Juneja et al. 2002).

The complete H^N and ^{15}N backbone resonance assignments of barstar unfolded in 8 M urea at pH 6.5 and 25°C were obtained by 3D HNN and HN(C)N experiments. Figure 2 shows an illustrative sequential walk through the $F_1(^{15}N)$–$F_3(H^N)$ planes for the stretch Gly 81 to Glu 76 in the HNN spectrum of the unfolded protein. All the backbone amide assignments obtained for unfolded barstar in 8 M urea at pH 6.5 are indicated in the ^{1}H–^{15}N HSQC spectrum in Figure 3. It is satisfying to note that despite the small chemical shift dispersion (the chemical shifts of H^N are within a range of 0.7 ppm), as is expected for a disordered state, we were able to obtain complete assignments. Following these assignments, $^{13}C_\alpha$, $^{13}C_\beta$, and ^{13}CO resonances were assigned from CBCANH (Grzesiek and Bax 1992a), CBCA(CO)NH (Grzesiek and Bax 1992b), and HNCO (Kay et al. 1990) triple resonance experiments in a straightforward manner. All the assigned chemical shifts have been listed in Supplemental Table 1. They have also been deposited in the BMRB database (accession no. 6227).

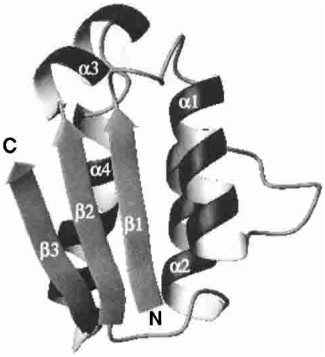

Figure 1. NMR structure of barstar (PDB ID 1BTA; Lubienski et al. 1994).

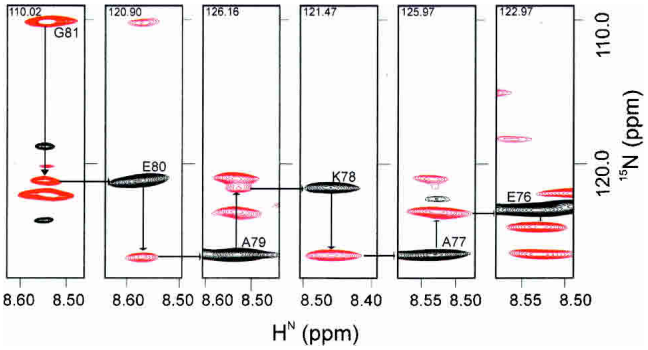

Figure 2. Illustrative sequential walk through F_1–F_3 planes in the HNN spectrum of barstar in 8 M urea at pH 6.5 and 25°C. Sequential connectivities are indicated for the Gly 81 to Glu 76 stretch in the protein primary sequence. Strips from the spectra at the appropriate H^N chemical shifts are shown. Black and red contours are positive and negative peaks, respectively. The distinct Gly 81 plane in the HNN spectrum served as a starting triplet fixed point to obtain sequential connectivities. The F_2 (^{15}N) chemical shift is indicated at the *top* of each strip of the HNN spectrum (for details, see Bhavesh et al. 2001 and Panchal et al. 2001).

Secondary chemical shifts

In native protein structures, the $^{13}C_\alpha$ resonances in α-helices are shifted downfield typically by an average of 3.1 ± 1.0 ppm, and the $^{13}C_\alpha$ resonances in β-sheets are shifted upfield typically by −1.5 ± 1.2 ppm (Dyson and Wright 2001). In contrast, in unfolded states of proteins, which are highly dynamic, the observed NMR parameters are a population-weighted average of all structures in the conformational ensemble; consequently, the observed secondary shifts are considerably smaller. Nevertheless, deviations of chemical shifts from random coil values reflect the relative population of dihedral angles in the α and β regions of conformational space (Dyson and Wright 2001). Chemical shifts are the only NMR parameters that can detect even small conformational propensities of the protein backbone in disordered states of proteins, which cannot be detected by NOEs or coupling constants.

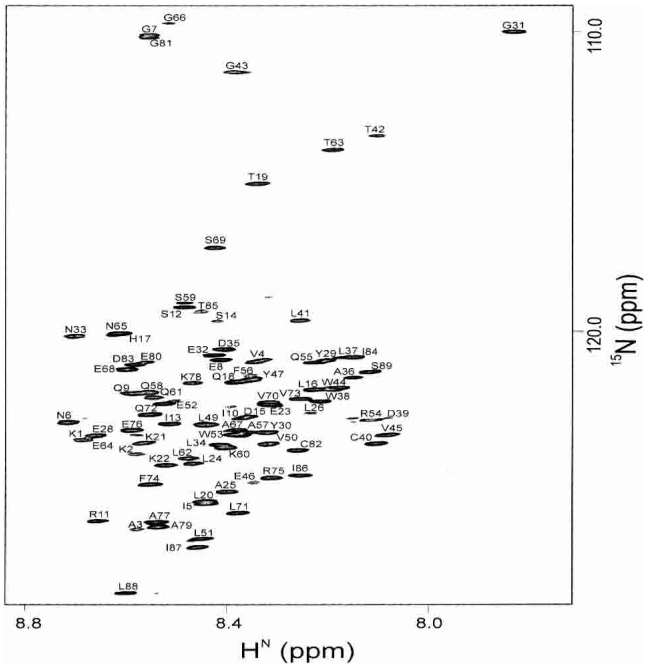

Figure 3. ^{1}H–^{15}N HSQC spectrum of barstar in 8 M urea at pH 6.5 and 25°C. The H^N and ^{15}N backbone assignments are indicated.

Among the various secondary chemical shifts, those of H_α, $^{13}C_\alpha$, and ^{13}CO are diagnostic of structural information in the polypeptide chain (Dyson and Wright 2001, 2002). Here we have used $^{13}C_\alpha$ and ^{13}CO secondary shifts for barstar in 8 M urea at pH 6.5 and 25°C, and these are shown in Figure 4. Positive secondary shifts for $^{13}C_\alpha$ and ^{13}CO indicate a preference for φ, ψ angles in the helical conformation, whereas negative secondary shifts indicate a preference for φ, ψ angles in the β-sheet conformation. The secondary chemical shifts of barstar unfolded in 8 M urea are small because of rapid conformational averaging, as expected for a disordered state of a protein (Dyson and Wright 2001, 2002). Several residues in the urea-unfolded state of barstar that form helices in the native structure (Fig. 1), however, show definite positive deviations of the $^{13}C_\alpha$ chemical shifts (Fig. 4A), indicating a propensity of the polypeptide backbone toward populating the helical region of φ, ψ space. In fact, some residues show positive deviations of ≥0.6 ppm. These include Ile 13, Ser 14, His 17, Lys 21, and Lys 22, which belong to helix 1 in native barstar; Cys 40, which belongs to helix 2; Gln 58 to helix 3; and Val 70, which belongs to helix 4 (Lubienski et al. 1994). Ala 25 at the C terminus of helix 1 and Glu 57 and Gly 81, which are at the N and C termini of helices 3 and 4, respectively, also show ≥0.6-ppm secondary shifts. On the other hand, there are a few residues in the segments that are helical in the native structure, which show negative secondary shifts indicating a propensity toward an extended or β-strand conformation in the urea-unfolded protein. In helix 1 (Ile 13–Leu 24), five residues have negative $^{13}C_\alpha$ secondary shifts; Asp 15 has an upfield shift of 0.5 ppm, whereas the other four residues have ≤0.3-ppm upfield chemical shifts. Most of the residues with positive deviations of $^{13}C_\alpha$ resonances in helix 1, however, show ≥0.6-ppm downfield shifts. In helix 2 (Leu 34–Leu 41), four residues show negative and four residues show upward $^{13}C_\alpha$ secondary shifts between 0.3 and 0.5 ppm except Cys 40, which has a positive secondary shift of >0.6 ppm. In helix 3 (Gln 58–Gln 61), two residues show very small negative $^{13}C_\alpha$ secondary shifts. In helix 4 (Ala 67–Ala 79), five residues have negative $^{13}C_\alpha$ secondary shifts; Glu 68 shows a 0.5-ppm negative shift, and the other four residues show ≤0.4-ppm negative shifts. Most of the residues in helix 4 have positive $^{13}C_\alpha$ secondary shifts of ≥0.4 ppm.

The deviations from random coil values for the ^{13}CO resonances of barstar in 8 M urea are small. These secondary shifts, in general, follow the pattern of $^{13}C_\alpha$ secondary shifts, as seen in Figure 4B. There appears to be a clustering of positive secondary shifts in the regions of the polypeptide that form helices 1 and 2 in the native state, again indicating a propensity of the main-chain backbone of these residues for a helical conformation. Residue Leu 37 in helix 2, however, shows a negative ^{13}CO secondary shift of ~0.5 ppm. Residues corresponding to helices 3 and 4 in native barstar show a distribution of positive and negative shifts suggesting conformational averaging in these regions in the unfolded state.

The analysis of the chemical shifts of sequence-corrected $^{13}C_\alpha$ and ^{13}CO resonances in urea-unfolded barstar suggests that many residues that form helices in the native protein preferentially adopt φ, ψ angles corresponding to the helical region of conformational space even in the unfolded state. These residues are, however, separated by residues showing negative $^{13}C_\alpha$ and ^{13}CO secondary shifts. In a helix, hydrogen bonds are formed between residues *i* and *i* + 4, and it is the enthalpy of the hydrogen bonds that compensates for the loss of entropy that occurs when residues are fixed into a helical conformation. In the absence of the stabilizing H-bonds, as in the case of a single amino acid or two to three consecutive amino acids showing positive $^{13}C_\alpha$ and ^{13}CO secondary shifts, it is too costly, entropy-wise, to form the first turn of a helix. Thus, it is unlikely that any residual native-like helical structure exists in the urea-unfolded state, in those segments that form helices in native barstar. For the acid-unfolded state of apomyoglobin and the low pH state of acyl-coenzyme A binding protein, residual helical structure was indicated by stretches of consecutive residues showing positive $^{13}C_\alpha$ and ^{13}CO secondary shifts in regions corresponding to some of the native helices (Yao et al. 2001; Thomsen et al. 2002).

In the regions of urea-unfolded barstar corresponding to β-strands 1 and 2, $^{13}C_\alpha$ and ^{13}CO secondary shifts indicate rapid conformational averaging between the α and β re-

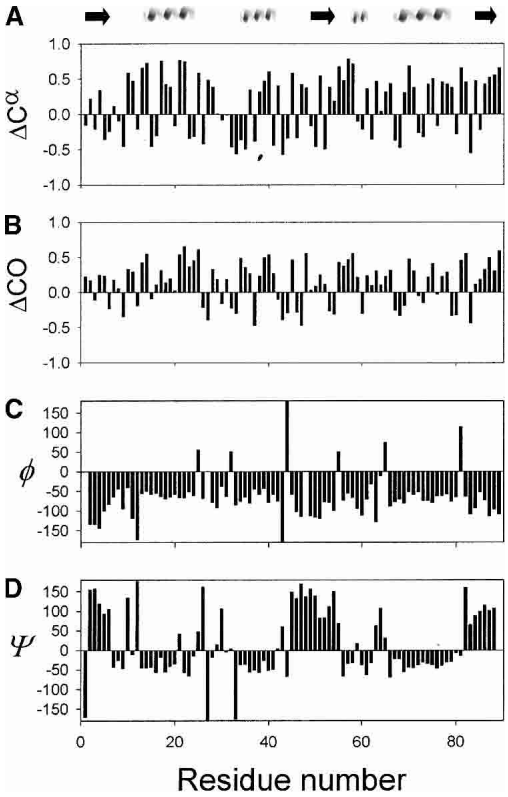

Figure 4. Secondary chemical shifts for (*A*) C_α and (*B*) CO resonances in barstar in 8 M urea at pH 6.5 and 25°C. The chemical shifts have been corrected for contributions from the local amino acid sequence (Schwarzinger et al. 2001). Reference random coil shifts used are those determined for the peptide Ac-GGXGG-NH_2 in 8 M urea at pH 2.3 (Schwarzinger et al. 2000). The φ, ψ backbone dihedral angles for native barstar are plotted against the primary sequence in *C* and *D*, respectively. The native secondary structural elements are indicated at the *top* in the figure.

gions of φ, ψ space so that the polypeptide chain in these stretches of the protein does not appear to preferentially populate either region in the unfolded state. Most residues corresponding to β-strands 1 and 2 in the native protein have secondary shifts close to random coil values, as well as a distribution of positive and negative shifts. In β-strand 2, Leu 51 has a positive $^{13}C_\alpha$ shift of >0.5 ppm, but Val 50 and Glu 52 show negative $^{13}C_\alpha$ shifts of ~0.5 ppm. Interestingly, most of the residues that belong to β-strand 3 (Ile 84–Ser 89) in native barstar have positive $^{13}C_\alpha$ secondary shifts of ≥0.5 ppm, suggesting that these show nonnative helical conformational propensities in the unfolded state. This is confirmed by the observation of positive ^{13}CO secondary shifts for all residues corresponding to β-strand 3, with Ile 87 and Ser 89 showing downfield shifts of ≥0.5 ppm.

Comparison of 8 M urea-unfolded barstar and cold denatured barstar

In an earlier NMR study of the cold denatured state of the C40/82A mutant form of barstar in 3 M urea and 5°C, chemical shift analysis supported by NOE data provided evidence for native-like residual helical structure in the segments corresponding to helices 1 and 2 in the native structure (Wong et al. 1996). Although our results also indicate that several residues in the helical regions of native barstar show helical conformational propensities in its urea-unfolded state, these are juxtaposed by residues showing propensities toward an extended conformation. As discussed above, it is therefore very unlikely that these segments form helices in the urea-unfolded state. It may be noted that as shown in Figure 4, C and D, there is some degree of variation observed in the ψ torsion angles in native barstar as well, in that some residues in helices 1 and 3 show positive values. Residues that form helices are, in general, characterized by negative values of ψ and those that form β-strands by positive values of ψ.

Wong et al. (1996) also observed nonnative helical residual structure in part of the second β-strand together with the loop connecting it to helix 3, and suggested that this may exist because of the formation of a hydrophobic cluster involving Trp 53 and Phe 56. In this study of the unfolded form in 8 M urea, residues Leu 51 and Trp 53–Glu 57 in this region along with Gln 58 in helix 3 show helical conformational propensities: Leu 51, Gln 55, Glu 57, and Gln 58 have positive $^{13}C_\alpha$ shifts of ≥0.6 ppm. This suggests that nonnative residual helical structure may exist in this region even in urea-unfolded barstar. Rearrangement of this stretch of the polypeptide chain may occur during folding to enable spatial positioning of the native topology.

In addition, it is observed in this study that residues in the native β-strand 3 segment appear to show a preference for populating a nonnative helical conformation in the 8 M urea-unfolded form. In contrast, the C-terminal segment from Asn 65 to Ser 89, which corresponds to helix 4 and β-strand 3 in folded barstar, shows a random coil conformation in the cold-denatured protein (Wong et al. 1996). The differences in random versus nonrandom behavior of some segments of the backbone between the cold-denatured state and the urea-unfolded state of barstar probably arise because of the different methods of denaturation. Whereas cold denaturation has an effect on the enthalpy and entropy of the protein (Privalov 1990), denaturation by urea occurs as a consequence of altered solvent properties (Schiffer and Dotsch 1996). Urea changes properties such as viscosity, dielectric constant, and surface tension of the solution, which would have an effect on the folding and unfolding free energies (Graziano 2001). In future work, it will be necessary to obtain NOE data to confirm the analysis of secondary shifts, and the existence of residual native or nonnative structure in the urea-unfolded state of the protein.

Implications for the folding of barstar

Helices 1 and 2 have been suggested to be the initiation sites for barstar folding from the cold-denatured state (Wong et

al. 1996). In contrast, it appears unlikely that these helices serve as initiation sites for barstar folding from the 8 M urea-unfolded state, because none of the helical regions of native barstar appear to have a preference for helical conformations in the 8 M urea-unfolded state. It is not surprising, therefore, that during the folding of barstar, the formation of specific secondary structure is seen to occur only after the initial collapse of the polypeptide chain (Agashe et al. 1995; Rami and Udgaonkar 2002; Pradeep and Udgaonkar 2004). The NMR results presented here suggest that the segment that forms β-strand 3 in the folded protein has some preference for nonnative helical structure in the unfolded state, but it is not known, at present, whether any of the intermediates on the folding pathway of barstar possess nonnative structure. The folding of barstar has been shown to occur via multiple competing pathways (Shastry and Udgaonkar 1995; Bhuyan and Udgaonkar 1999; Sridevi et al. 2004), and it remains to be seen if conformers with different native and nonnative structures in the unfolded state fold via different routes.

Materials and methods

Protein purification and NMR sample preparation

The barstar gene in pMT316 plasmid was expressed in MM294 *Escherichia coli* cells and purified as described earlier (Khurana and Udgaonkar 1994). For uniform labeling of the protein either singly with ^{15}N or doubly with ^{15}N and ^{13}C, bacteria were grown in M9 minimal media containing $^{15}NH_4Cl$ and ^{13}C-glucose as the only sources of nitrogen and carbon, respectively, as described (Juneja et al. 2002). The protein used for all experiments was >95% pure.

The sample for NMR experiments was prepared by dissolving lyophilized protein in 20 mM sodium phosphate buffer containing 8 M urea (pH 6.5) to a final concentration of 1.1 mM.

NMR spectroscopy

All NMR spectra were collected on a 600 MHz Varian Unity Plus spectrometer at 25°C. Typical spectral widths for all 2D and 3D experiments recorded were 7002.8 Hz for ^{1}H, 1460 Hz for ^{15}N, 13,500 Hz for $^{13}C_{\alpha\beta}$, and 2500 Hz for ^{13}CO. The ^{1}H and ^{15}N carrier frequencies were set at 4.71 ppm (water) and 120 ppm, respectively. The ^{13}C carrier frequency was set at 56.0 ppm for HNN, HN(C)N, 45.0 ppm for CBCANH and CBCA(CO)NH, and 174.0 ppm for HNCO. In all spectra, ^{1}H, ^{13}C, and ^{15}N chemical shifts were referenced, respectively, to HDO (4.71 ppm at 32°C), indirectly to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), and to trimethylsilyl sodium propionate (TSP).

The 2D ^{1}H–^{15}N HSQC spectrum was recorded with 512 complex t_1 increments, 8192 t_2 points, and four scans for each fid. The 3D HNN spectrum was recorded with the following parameters: 32 complex points along t_1 (^{15}N) and t_2 (^{15}N) and 1024 complex points along t_3 (H^N), 16 scans for each fid, and $T_N = T_C = 14$ msec. The 3D HN(C)N spectrum of the same sample was recorded using parameters identical to those used for the HNN experiment. The T_{CC} delay was set to 4.5 msec. The acquisition time for the two experiments was ~26 h each. CBCANH and CBCA(CO)NH spectra were acquired with 1024 (H^N), 32 (^{15}N), and 80 (^{13}C) complex points, and an HNCO spectrum was acquired with 1024 (H^N), 32 (^{15}N), and 50 (^{13}C) complex points.

Secondary chemical shifts

The deviations of specific chemical shifts (secondary shifts) from their random coil values are highly sensitive to conformational preferences of the protein backbone (Wishart and Sykes 1994). Because the chemical shifts are influenced both by neighboring amino acids and local backbone structure, it was important to correct these for contributions from the local amino acid sequence (Schwarzinger et al. 2000). The random coil values used in the secondary shift analysis (Schwarzinger et al. 2000) were corrected using sequence-dependent correction factors determined for a set of Ac-GGXGG-NH_2 peptides in 8 M urea at pH 2.3 (Schwarzinger et al. 2001). Deviations in specific chemical shifts were then calculated by subtracting the corrected random coil values from the measured chemical shifts for all the residues in urea-unfolded barstar.

Acknowledgments

All NMR spectra have been recorded in the National High-Field NMR facility at the Tata Institute of Fundamental Research, Mumbai. This work was funded by TIFR and by the Wellcome Trust.

References

Ackerman, M.S. and Shortle, D. 2002. Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. *Biochemistry* **41:** 13791–13797.

Agashe, V.R., Shastry, M.C., and Udgaonkar, J.B. 1995. Initial hydrophobic collapse in the folding of barstar. *Nature* **377:** 754–757.

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.

Arcus, V.L., Vuilleumier, S., Freund, S.M.V., Bycroft, M., and Fersht, A.R. 1995. A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: Implications for the initiation of protein folding. *J. Mol. Biol.* **254:** 305–321.

Bhavesh, N.S., Panchal, S.C., and Hosur, R.V. 2001. An efficient high-throughput resonance assignment procedure for structural genomics and protein folding research by NMR. *Biochemistry* **40:** 14727–14735.

Bhavesh, N.S., Sinha, R., Krishna Mohan, P.M., and Hosur, R.V. 2003. NMR elucidation of early folding hierarchy in HIV-1 protease. *J. Biol. Chem.* **278:** 19980–19985.

Bhuyan, A.K. and Udgaonkar, J.B. 1999. Observation of multistate kinetics during the slow folding and unfolding of barstar. *Biochemistry* **38:** 9158–9168.

Choy, W.-Y., Mulder, F.A.A., Crowhurst, K.A., Muhandiram, D.R., Millett, I.S., Doniach, S., Forman-Kay, J.D., and Kay, L.E. 2002. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *J. Mol. Biol.* **316:** 101–112.

Clementi, C., Nymeyer, H., and Onuchic, J.N. 2000. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298:** 937–953.

Crowhurst, K.A. and Forman-Kay, J.D. 2003. Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain. *Biochemistry* **42:** 8687–8695.

Crowhurst, K.A., Tollinger, M., and Forman-Kay, J.D. 2002. Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. *J. Mol. Biol.* **322:** 163–178.

Daggett, V. and Fersht, A.R. 2003. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* **4:** 497–502.

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.

———. 2002. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. A new perspective on unfolded proteins. *Adv. Protein. Chem.* **62:** 311–340.

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.

———. 1997. Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* **36:** 2390–2402.

Fersht, A.R. and Daggett, V. 2002. Protein folding and unfolding at atomic resolution. *Cell* **108:** 573–582.

Gianni, S., Guydosh, N.R., Khan, F., Caldas, T.D., Mayor, U., White, G.W., DeMarco, M.L., Daggett, V., and Fersht, A.R. 2003. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci.* **100:** 13286–13291.

Graziano, G. 2001. On the solubility of aliphatic hydrocarbons in 7 M aqueous urea. *J. Phys. Chem. B* **105:** 2632–2637.

Grzesiek, S. and Bax, A. 1992a. An efficient experiment for sequential backbone assignment of medium sized isotopically enriched proteins. *J. Magn. Reson.* **99:** 201–207.

———. 1992b. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114:** 6291–6293.

Hodsdon, M.E. and Frieden, C. 2001. Intestinal fatty acid binding protein: The folding mechanism as determined by NMR studies. *Biochemistry* **40:** 732–742.

Juneja, J. and Udgaonkar, J.B. 2003. NMR studies of protein folding. *Curr. Sci.* **84:** 157–172.

Juneja, J., Bhavesh, N.S., Udgaonkar, J.B., and Hosur, R.V. 2002. NMR identification and characterization of the flexible regions in the 160 kDa molten globule-like aggregate of barstar at low pH. *Biochemistry* **41:** 9885–9899.

Kay, L.E., Ikura, M., Tschudin, R., and Bax, A. 1990. Three-dimensional triple resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89:** 496–514.

Kazmirski, S.L., Wong, K.-B., Freund, S.M.V., Tan, Y.-J., Fersht, A.R., and Daggett, V. 2001. Protein folding from a highly disordered denatured state: The folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci.* **98:** 4349–4354.

Khurana, R. and Udgaonkar, J.B. 1994. Equilibrium unfolding studies of barstar: Evidence for an alternative conformation which resembles a molten globule. *Biochemistry* **33:** 106–115.

Killick, T.R., Freund, S.M., and Fersht, A.R. 1999. Real-time NMR studies on a transient folding intermediate of barstar. *Protein Sci.* **8:** 1286–1291.

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 nonnative protein. *Science* **295:** 1719–1722.

Lietzow, M.A., Jamin, M., Dyson, H.J., and Wright, P.E. 2002. Mapping long-range contacts in a highly unfolded protein. *J. Mol. Biol.* **322:** 655–662.

Louhivuori, M., Pääkkönen, K., Fredriksson, K., Permi, P., Lounila, J., and Annila A. 2003. On the origin of residual dipolar couplings from denatured proteins. *J. Am. Chem. Soc.* **125:** 15647–15650.

Lubienski, M.J., Bycroft, M., Freund, S.M.V., and Fersht, A.R. 1994. Three dimensional solution structure and ^{13}C assignments of barstar using nuclear magnetic resonance spectroscopy. *Biochemistry* **33:** 8866–8877.

Makarov, D.E., Keller, C.A., Plaxco, K.W., and Metiu, H. 2002. How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc. Natl. Acad. Sci.* **99:** 3535–3539.

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 fibronectin type III domain. *J. Mol. Biol.* **286:** 579–592.

Miller, E.J., Fischer, K.F., and Marqusee, S. 2002. Experimental evaluation of topological parameters determining protein-folding rates. *Proc. Natl. Acad. Sci.* **99:** 10359–10363.

Mok, Y.-K., Kay, C.M., Kay, L.E., and Forman-Kay, J. 1999. NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *J. Mol. Biol.* **289:** 619–638.

Neri, D., Billeter, M., Wider, G., and Wüthrich, K. 1992. NMR determination of residual structure in a urea-denatured protein. *Science* **257:** 1559–1563.

Ohnishi, S. and Shortle, D. 2003. Effects of denaturants and substitutions of hydrophobic residues on backbone dynamics of denatured staphylococcal nuclease. *Protein Sci.* **12:** 1298–1302.

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.

Pappu, R.V., Srinivaan, R., and Rose, G.D. 2000. The Flory isolated-pair hypothesis is not valid for polypeptide chains: Implications for protein folding. *Proc. Natl. Acad. Sci.* **97:** 12565–12570.

Pradeep, L. and Udgaonkar, J.B. 2004. Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J. Biol. Chem.* (in press).

Privalov, P.L. 1990. Cold denaturation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **25:** 281–305.

Rami, B.R. and Udgaonkar, J.B. 2002. Mechanism of formation of a productive molten globule form of barstar. *Biochemistry* **41:** 1710–1716.

Schiffer, C.A. and Dotsche, V. 1996. The role of protein–solvent interactions in protein unfolding *Curr. Opin. Biotechnol.* **7:** 428–432.

Schreiber, G. and Fersht, A.R. 1993. The refolding of *cis* and *trans*-peptidyl-prolyl isomers of barstar. *Biochemistry* **32:** 11195–11203.

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.

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.

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.

Schwarzinger, S., Wright, P.E., and Dyson, H.J. 2002. Molecular hinges in protein folding: The urea-denatured state of apomyoglobin. *Biochemistry* **41:** 12681–12686.

Shastry, M.C. and Udgaonkar, J.B. 1995. The folding mechanism of barstar: Evidence for multiple pathways and multiple intermediates. *J. Mol. Biol.* **247:** 1013–1027.

Shortle, D. 1996. Structural analysis of non-native states of proteins by NMR methods. *Curr. Opin. Struct. Biol.* **6:** 24–30.

———. 2002. Composites of local structure propensities: Evidence for local encoding of long-range structure. *Protein Sci.* **11:** 18–26.

Shortle, D. and Ackerman, M.S. 2001. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293:** 487–489.

Smith, L.J., Fiebig, K., Schwalbe, H., and Dobson, C.M. 1996. The concept of a random coil. Residual structure in peptides and denatured proteins. *Folding Des.* **1:** R95–R106.

Sridevi, K., Juneja, J., Bhuyan, A.K., Krishnamoorthy, G., and Udgaonkar, J.B. 2000. The slow folding reaction of barstar: The core tryptophan region attains tight packing before substantial secondary and tertiary structure formation and final compaction of the polypeptide chain. *J. Mol. Biol.* **302:** 479–495.

Sridevi, K., Lakshmikanth, G., Krishnamoorthy, G., and Udgaonkar, J.B. 2004. Increasing stability decreases conformational heterogeneity in a protein folding intermediate ensemble. *J. Mol. Biol.* (in press).

Srinivasan, R. and Rose, G.D. 1999. A physical basis for protein secondary structure. *Proc. Natl. Acad. Sci.* **96:** 14258–14263.

Thomsen, J.K., Kragelund, B.B., Teilum, K., Knudsen, J., and Poulsen, F.M. 2002. Transient intermediary states with high and low folding probabilities in the apparent two-state folding equilibrium of ACBP at low pH. *J. Mol. Biol.* **318:** 805–814.

Van Gunsteren, W.F., Burgi, R., Peter, C., and Daura, X. 2001. The key to solving the protein-folding problem lies in an accurate description of the denatured state. *Angew. Chem. Int. Ed. Engl.* **40:** 351–355.

Wishart, D.S. and Sykes, B.D. 1994. Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239:** 363–392.

Wong, K.-B., Freund, S.M.V., and Fersht, A.R. 1996. Cold denaturation of barstar: ^1H, ^{15}N and ^{13}C NMR assignment and characterization of residual structure. *J. Mol. Biol.* **259:** 805–818.

Wong, K.-B., Clarke, J., Bond, C.J., Neira, J.L., Freund, S.M.V., Fersht, A.R., and Daggett, V. 2000. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.* **296:** 1257–1282.

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.

Zagrovic, B. and Pande, V.S. 2003. Structural correspondence between the α-helix and the random-flight chain resolves how unfolded proteins can have native-like properties. *Nat. Struct. Biol.* **10:** 955–961.