Characterization of Conformational Preferences in a Partly Folded Protein by Heteronuclear NMR Spectroscopy: Assignment and Secondary Structure Analysis of Hen Egg-White Lysozyme in Trifluoroethanol†

Matthias Buck,‡ Harald Schwalbe, and Christopher M. Dobson*

Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, England

Received June 14, 1995⊗

ABSTRACT: 2D and 3D heteronuclear NMR methods have been used to characterize the structure of hen egg-white lysozyme in a partially folded state using uniformly ^{15}N-labeled protein. This state is formed by the denaturation of the protein in 70% trifluoroethanol (TFE)/30% water (v/v) and is characterized by substantial helical secondary structure in the absence of extensive tertiary interactions. ^{15}N-filtered 3D NOESY and TOCSY experiments have allowed the sequential assignment of resonances for all but 2 of the 126 main chain amide nitrogen atoms and of the majority of main and side chain proton resonances. The conformation of the polypeptide chain was characterized by analysis of the pattern of NOEs, H_α chemical shift perturbations, $^3J(H^N,H_\alpha)$-coupling constants, and hydrogen exchange protection. These NMR parameters are highly complementary and are consistent with a model for the TFE state in which six regions of the polypeptide chain are substantially ordered in helical conformations. The structure in different regions however, shows different levels of persistency. Five of the helices exhibit significant protection of amide hydrogens against exchange with solvent and are located in regions of the polypeptide which are helical in the native state. By contrast, helical structures of greater flexibility are observed both as extensions to the native-like helices and as a nonnative structure in the region of the molecule which forms the C-terminal part of the β-sheet in the native state. No specific structural preferences are detected in regions corresponding to the long loop and to the N-terminal part of the β-sheet of native lysozyme. A combination of local features of the polypeptide chain, including the predicted propensities of residues for helix formation and for their participation in N- and C-terminal helix capping interactions, allows the conformational behavior of the polypeptide chain of hen lysozyme to be rationalized for this partially folded state. The analysis implies that the nonnative structures are a result of interactions which are local to the polypeptide chain. These, and the highly persistent native-like structures, give insight into species which form early during folding.

It is generally accepted that the refolding of small monomeric proteins occurs along pathways which are defined by partially folded states. The nature and the dynamics of the structure in these states, intermediate between the fully folded and fully unfolded conformations of the polypeptide chain, are of critical importance for the elucidation of the mechanisms by which proteins fold (Barrick & Baldwin, 1993; Fersht, 1993; Dobson, 1991, 1994). Detailed studies of such partially ordered states are likely to provide an understanding of the competing interactions which determine the stability of the structures as well as the manner of their interconversion during the folding process.

In practice, the characterization of intermediates sampled during kinetic refolding is often complicated by their transient nature, the lack of appropriate structural probes, the possible conformational heterogeneity of the species involved, and the parallel nature of some of the folding pathways. Despite these difficulties, the folding pathways of a number of proteins have been analyzed in some detail using a wide variety of different techniques (Radford et al., 1992; Jennings & Wright, 1993; Sosnick et al., 1994; Dobson et al., 1994; Otzen et al., 1994). Equilibrium studies are made difficult by the high degree of cooperativity of the folding process observed for many proteins under denaturing conditions. Studies of protein fragments (Dyson & Wright, 1991; Dill & Shortle, 1991) and of peptide models (Peng & Kim, 1994) provide an alternative strategy, allowing the propensity of a given amino acid sequence to form local structures to be delineated (Dyson et al., 1992; Kemmink & Creighton, 1993; Chakrabartty et al., 1994).

Another strategy is to overcome difficulties in the study of folding intermediates that arise from the high cooperativity of the folding processes by examination of intact proteins under carefully chosen equilibrium conditions. While in some studies low concentrations of GuHCl or extremes of pH have been used to promote local unfolding of the most labile regions of the native structures (Mayo & Baldwin, 1993; Redfield et al., 1994), other studies have used denaturants at high concentrations in order to probe the most persistent regions of protein structure (Neri et al., 1992; Buck et al., 1993, 1994; Logan et al., 1994). It is clear that,

† Contribution from the Oxford Centre for Molecular Sciences, which is supported by the UK EPSRC, BBSRC, and MRC. H.S. was supported by a Human Capital and Mobility grant from the European Commission. The research of C.M.D. is supported in part by an International Research Scholars award from the Howard Hughes Medical Foundation.

* To whom correspondence should be addressed at the New Chemistry Laboratory, University of Oxford.

‡ Present address: Department of Chemistry, Harvard University, 12 Oxford St., Cambridge, MA 02138.

⊗ Abstract published in *Advance ACS Abstracts,* September 1, 1995.

0006-2960/95/0434-13219$09.00/0 © 1995 American Chemical Society

depending on the protein and the solution conditions, structure in these partially ordered states, including the species often referred to as molten globules, may range from very local conformational preferences in otherwise unfolded states to structure in states which are substantially ordered and involve larger segments of the polypeptide chain (Dill & Shortle, 1991; Ptitsyn, 1992). In the latter case the ordered regions of a partially folded protein may encompass significant fractions of structural domains which comprise the overall conformation of the native protein (Dobson, 1994).

Several approaches have been adopted to make possible NMR studies of proteins in nonnative conformations at high resolution. Protection of amide hydrogens against exchange with solvent, for example, is afforded only by structures which are persistently ordered (Englander & Kallenbach, 1984) and can be detected also by indirect methods, such as pH or denaturant concentration pulse quench techniques [e.g., Baum et al. (1989), Roder (1989), Buck et al. (1993), Woodward (1994)], which extract information on the folding from well resolved NMR[1] spectra of the native state. The NMR spectrum of native proteins has also been important for direct studies of denatured states under equilibrium conditions by magnetization transfer experiments (Jeener et al., 1979). Correlations of resonances between native and nonnative states are possible if the two states can be brought into exchange with kinetics of typically greater than 1 s^{-1}, and using homonuclear techniques this has allowed the identification of resolved resonances in denatured states of a number of proteins (Dobson et al., 1984; Baum et al., 1989; Evans et al., 1991; Harding et al., 1991; Alexandrescu et al., 1993). Direct study of denatured states of proteins by homonuclear methods is, however, limited by overlap of resonances in the NMR spectra, which severely complicates their complete assignment and the extraction of NMR parameters.

The problem of resonance overlap in denatured states can be substantially overcome by editing the proton resonances with the chemical shifts of their attached heteronuclei, such as ^{15}N and ^{13}C in isotopically enriched samples. Assignments are then possible either in conjunction with heteronuclear edited transfer of magnetization between the unfolded and the folded states (Wider et al., 1991; Zhang et al., 1994) or by through-space or through-bond correlation techniques within the unfolded state itself (Kay et al., 1990; Stockman et al., 1993; Logan et al., 1994). Site-specific characterization of the conformational and the dynamic behavior in nonnative states of proteins may now in favorable circumstances be possible to an extent comparable to that achieved for native proteins and small peptides (Wüthrich, 1994). Although several such heteronuclear studies have been reported on denatured states, these involve proteins either in states which are substantially structured, possessing at least part of an ordered core (van Mierlo et al., 1993; Feng et al., 1994; Redfield et al., 1994), or states which are largely disordered, containing few, if any, local clusters of structure (Neri et al., 1992; Logan et al., 1994; Alexandrescu et al., 1994b; Arcus et al., 1994). Studying proteins in states with characteristics between these two extremes, such as those which have become known as classical molten globule states, is demanding by NMR techniques because the structures in them may experience conformational averaging on time scales that give rise to broadening of resonances and to a reduction in the resolution and the sensitivity of the NMR signals [e.g., Alexandrescu et al. (1993)]. Thus it is of considerable interest to examine recently developed heteronuclear NMR methods in their application to states of proteins with structural and dynamic features intermediate between largely structured and unstructured states.

Recently, we have described a partially folded state of hen lysozyme obtained by denaturing the protein in a trifluoroethanol (TFE)/water mixture. This partially folded conformation, referred to as the TFE state, differs substantially from both the native and highly denatured states of the protein (Buck et al., 1993). Far-UV CD spectroscopy suggests that the helical structure content exceeds that of the native state. However, hydrogen exchange measurements demonstrate that amides located in the most stable structures in the TFE state are situated in regions which are α-helical in the native state. Extensive and persistent tertiary contacts are absent in this denatured state, although low-angle X-ray scattering data imply that the protein is at least partially compact in the presence of the native disulphides which persist in this denatured state (Shiraki et al., 1995). An important question, therefore, concerns the origin of the additional helical structure, its location, and whether effects of TFE on polypeptide chain interactions promote native-like or induce novel structures.

In this paper we have applied the established sequential assignment method (Wüthrich, 1986) to ^{15}N-filtered 3D spectra of TFE denatured hen lysozyme. The conformational preferences of hen lysozyme in this partially ordered state are then assessed by a number of NMR parameters such as perturbation of H_α chemical shifts from random coil values, $^3J(H^N,H_\alpha)$ coupling constants, the pattern of NOEs, and the degree of protection of amide hydrogens against solvent exchange. Key features of the model for the structure of the TFE state resulting from these data are rationalized on the basis of the propensities of different regions of the sequence of hen lysozyme for the formation of secondary structure.

MATERIALS AND METHODS

Collection of NMR Data. Hen egg-white lysozyme was expressed in *Aspergillus niger* using $^{15}NH_4Cl$ as the sole nitrogen source and purified from filtered culture medium (Roberts et al., 1992; Buck et al., 1995). NMR samples were prepared to contain ca. 3 mM protein in 70% TFE-d_2-OH/30% H_2O (v/v), and the pH was adjusted to 2.0 with phosphoric acid. NMR experiments were performed at 37 and 47 °C on NMR spectrometers belonging to the Oxford Centre for Molecular Sciences with ^1H operating frequencies of 500.2, 600.2, and 750.2 MHz. The ^1H and ^{15}N assignments were made from two sets of data recorded at 37 and 47 °C. 3D NOESY-HSQC, HSQC-NOESY-HSQC, and

[1] Abbreviations: 2D and 3D, two and three dimensional; CD, circular dichroism; CSI, chemical shift index; HEWL, hen egg-white lysozyme (EC 3.2.1.17); HMQC, heteronuclear multiple-quantum correlation; HSQC, heteronuclear single-quantum correlation; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; TFE, 2,2,2,-trifluoroethanol; TOCSY, total correlation spectroscopy; $^3J(H^N,H_\alpha)$ coupling between H^N and H_α; $d_{i,i+1}(\alpha,N)$ distance between, e.g., H_α and H^N of residue i and $i+1$.

TOCSY-HSQC experiments [Driscoll et al. (1990), Frenkiel et al. (1990), Ikura et al. (1990), respectively] were recorded with mixing times of 200 ms in the NOESY and 62 ms in the TOCSY experiments and a recycle delay of typically 1.3 s. The data set comprised 128, 32, and 512 complex points in t_1, t_2, and t_3 for the NOESY-HSQC and TOCSY-HSQC experiments recorded with a digital resolution of 38.3, 25.9, and 13.2 Hz/complex point, respectively, at 600 MHz. The data set for the HSQC-NOESY-HSQC experiment comprised 64, 32, and 512 complex points in t_1, t_2, and t_3 recorded with a digital resolution of 13.0, 25.9, and 13.2 Hz/complex point, respectively, at 600 MHz. Forward linear prediction as implemented in Felix2.3 (Hare Research & Biosym Inc.) was used to extend the data in the ^{15}N dimension from 32 to 48 complex points, and all 3D data sets were zero-filled to yield a final matrix size of 256 × 128 × 2048 real data points.

Measurement of Coupling Constants. $^3J(H^N,H_\alpha)$-coupling constants were measured using HMQCJ and HNHA experiments and were determined from line shape simulations and integration of cross peak intensities [Kay and Bax (1989), Vuister and Bax (1994), respectively]. The HMQCJ experiment was performed on a 600 MHz spectrometer. The data set comprised 432 and 1024 complex points in t_1 (^{15}N) and t_2 (H^N) with a resolution of 1.9 and 5.5 Hz/complex point, respectively. Shifted (72°) squared sine-bell apodization was used in t_1, and the data matrix was zero-filled to 1024 real data points. Coupling constants were fitted to doublets in ω_1, after summing over the cross peak along ω_2 and four times zero-filling, by optimization of the coupling constant and line widths for each of the doublet components using home-written software (C. Redfield). The HNHA experiment was performed on a 500 MHz spectrometer equipped with a home-built triple resonance probe with shielded x,y,z-gradients. The data set comprised 48, 30, and 1024 complex points in t_1 (^{15}N), t_2 (H_α), and t_3 (H^N) with a resolution of 28.8, 117.2, and 5.5 Hz/complex point, respectively. The constant time delay to defocus the $^3J(H^N,H_\alpha)$-coupling constant was set to 26.1 ms. The row data were mirror image linear predicted (Zhu & Bax, 1990) in the ^{15}N dimension to 96 complex points, and the final matrix size was zero filled to 128, 128, and 512 real data points in ω_1, ω_2, and ω_3, respectively. The ratios of cross peak to diagonal peak intensities were determined using the volume integration routine in Felix2.3. Values extracted from both sets of experiments were found to yield coupling constants that were identical within ±0.4 Hz. No correction factors have been applied to account for differential relaxation (Harbison, 1993).

Direct Measurement of Amide Hydrogen Exchange. Lyophilized protein was dissolved in 70% TFE-d_3/30% D_2O (v/v) to a final concentration of 3 mM and the pH rapidly adjusted to 2.0. A series of 12 fast 2D $^{15}N-^1H$ HSQC spectra, each taking 23 min, were recorded at 37 °C. Single-exponential decays were fitted to the residue amide occupancy, measured by peak heights over the time course of spectra. The resulting rates (k_{ex}) were compared to intrinsic exchange rates (k_{int}) predicted for amides in an unstructured polypeptide chain, in order to account for near neighbor inductive effects on acid and water catalysis (Bai et al., 1993). The effect of TFE on the intrinsic exchange rates was found to be small (within a factor of 3) for model peptides at pH 2 (Buck et al., 1993), and protection factors (k_{int}/k_{ex}) are reported uncorrected for TFE cosolvent effects.

RESULTS

Resonance Assignment. Figure 1 shows the $^{15}N-^1H$ HSQC spectrum at a ^1H frequency of 750 MHz of hen lysozyme denatured in a 70% TFE/30% H_2O (v/v) mixture at 37 °C. Critical features of this spectrum are typical of those obtained for denatured proteins. Most significant is that the ^1H chemical shift dispersion characteristic of native states is dramatically reduced. Thus, while main chain amide protons resonate over a range of 2.6 ppm in the native state of hen lysozyme (Redfield & Dobson, 1988), their dispersion is reduced to around 1 ppm in the TFE denatured state. The same observation holds true for the side chain protons, for which only small deviations are observed from random coil chemical shift values (Wüthrich, 1986) ($\Delta\delta \leq 0.2$ ppm for the great majority). ^1H resonance overlap therefore severely complicates the resonance assignment of lysozyme in TFE using homonuclear spectra. The residual dispersion of the ^{15}N chemical shifts is, however, significant in this and other denatured states of hen lysozyme (Schwalbe et al., in preparation) as well as in other proteins (Logan et al., 1994; Alexandrescu et al., 1994b; Arcus et al., 1994). Consequently, the amide resonances are well dispersed through their correlation with the attached ^{15}N nuclei. Thus, it is apparent that in excess of 85% of the cross peaks are fully resolved in the $^{15}N-^1H$ HSQC spectrum of this denatured state.

In addition to well resolved $^{15}N-^1H$ cross peaks, NOESY spectra of the TFE denatured state of hen lysozyme were found to contain a large number of intense interproton NOEs. This made possible the application of the sequential assignment approach based on the identification of spin systems and their connection through NOEs between residues adjacent in sequence (Wüthrich, 1986). TOCSY transfer was, however, found to be poor for residues in helices because $^3J(H^N,H_\alpha)$ coupling constants are small. Furthermore, in the case of the TFE state, resonances arising from residues in helices were often observed to have larger than average line widths (see below). Nevertheless, at a preliminary level, around 30% of resonances could be assigned to spin system types in TOCSY spectra of the TFE state, serving as starting points for further assignment. Resonances could be assigned to certain amino acid types by the proximity of their ^{15}N, H^N, H_α, and particularly side chain ^1H chemical shifts to those expected in a random coil (Wüthrich, 1986; Wishart et al., 1992a,b, 1995). Some of the resonances show particularly clearly such amino acid specific clustering into a narrow region of ^{15}N chemical shifts. Thus, for example, all 12 glycine resonances are found in the region between 106 and 112 ppm.

Analysis of the NOE data confirms that the great majority of NOEs from the amide protons throughout the protein are to protons belonging to residues within a distance of +3/−2 residues in sequence. The intensity of NOEs between an amide proton and other protons of the residue itself, or those of the preceding residue, is generally higher than to others, suggesting that possible assignments of the most intense NOEs should first be sought from protons within the same residue and immediately neighboring residues. Furthermore, since the intensities of the sequential NN($i,i+1$) cross peaks in the NOESY spectra are generally high, the 3D HSQC-NOESY-HSQC experiment often provides unambiguous links between adjacent amides even when their H^N shifts overlap.

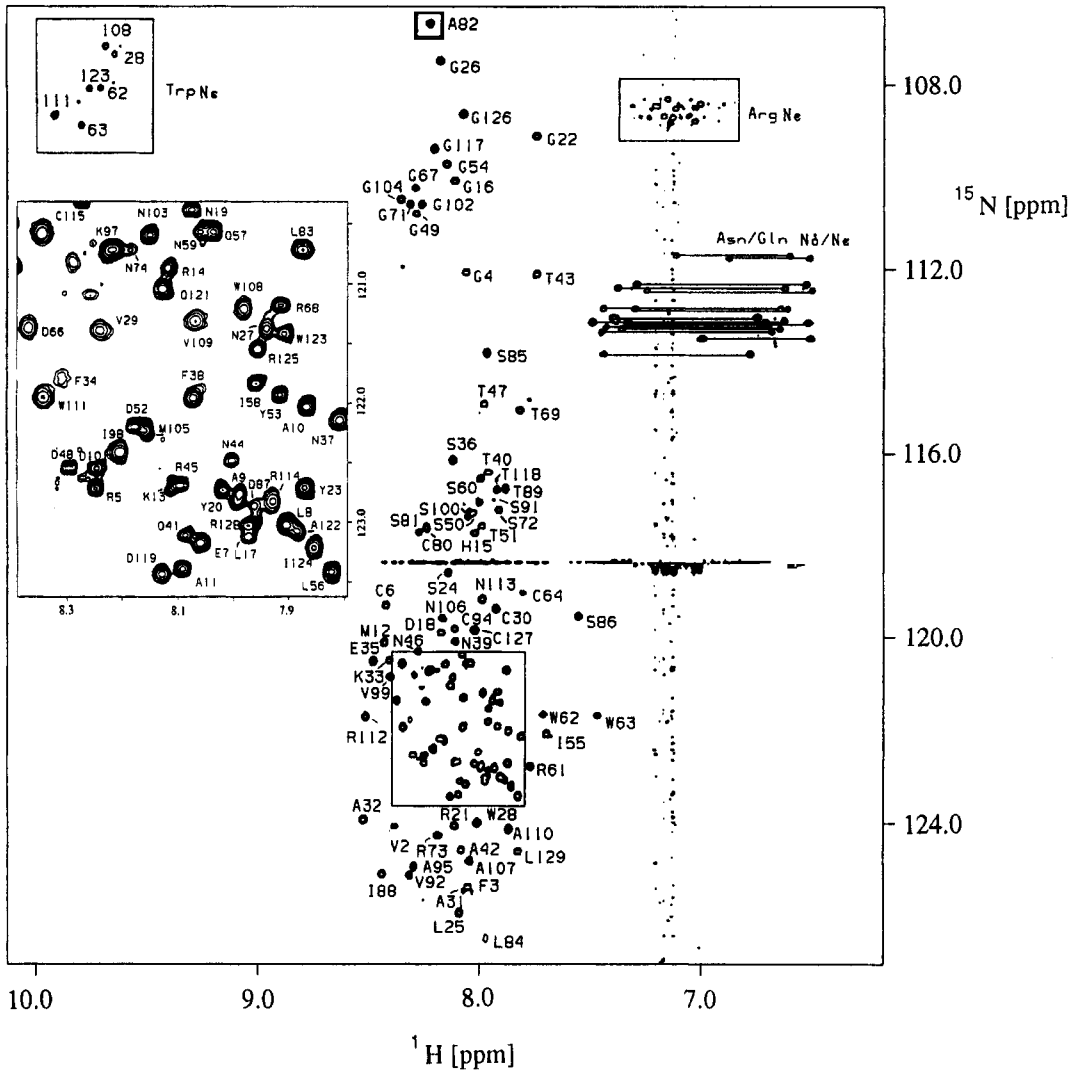

FIGURE 1: $^{15}N-^{1}H$ HSQC spectrum of the TFE-denatured state at 37 °C at a ^{1}H frequency of 750 MHz. A total of 117 out of the 126 expected main chain $^{15}N-^{1}H$ correlations could be assigned at this temperature. Cross-peaks in small boxes are folded in from outside the ^{15}N frequency region displayed. The most crowded region of the spectrum is enlarged.

From these experiments, it was possible to assign sequentially the 63 N-terminal and the 12 C-terminal residues, with the exception of Phe34 and Val120, in the set of 3D spectra of the TFE denatured state recorded at 37 °C. Assignment of resonances for residues in the C-terminal half of the molecule was complicated by the reduced number of NOEs observed in the 3D NOESY experiments and their apparently weaker intensity. These resonances correspond to $^{15}N-^{1}H$ correlations which have increased line widths in the HSQC spectrum (Figure 1). In order to attempt to enhance the rates of the processes which give rise to the line broadening, such as the interconversion of distinct conformations at a rate comparable to their difference in chemical shifts, and to attempt to resolve the overlap of certain resonances, the NMR data were also acquired at 47 °C under otherwise identical experimental conditions. In this set of spectra many of the weak $^{15}N-^{1}H$ correlations are found to be more intense, and additional resonances can be detected. Overlap of several $^{15}N-^{1}H$ correlations is resolved, as the chemical shifts of the amide protons in particular are often highly sensitive to temperature (Meruka et al., 1995, and references cited therein). By contrast, the chemical shifts of side chain protons bonded to carbons were found to be nearly identical at the two temperatures. It was possible to assign resonances for all but two of the expected 126 main chain amides in the spectra acquired at 47 °C. The $^{15}N-^{1}H$ correlations for Lys97 and Lys116 remained unassigned, probably because their line widths are still too large for their detection. However, NOEs from the H_α and some of the side chain protons of these two residues to neighboring amide protons were observed. The assignments made at the higher temperature permitted the assignment of spectra at the lower temperature for all but a few resonances of residues 65, 75–78, 90, and 120, which, although being assignable at 47 °C, are presumably too broad to be seen at 37 °C. It is interesting to note that several of these residues are located in close sequence proximity to cysteines 64, 80, 94, and 115 which are involved in three of the four disulfide cross-bridges which remain intact in this denatured state.

In addition to correlations involving main chain resonances, the $^{15}N-^{1}H$ HSQC spectrum contains $^{15}N-^{1}H$ correlations arising from side chain nitrogens of the 11 arginine, 6 tryptophan, 14 asparagine, and 3 glutamine residues of the protein. No appreciable NOEs were detected in the 3D NOESY spectrum involving protons bonded to the arginine $^{15}N^{\epsilon}$ side chain nuclei. However, several NOEs involving tryptophan indole-ring protons and asparagine/glutamine side chain protons are seen in the 3D NOESY-

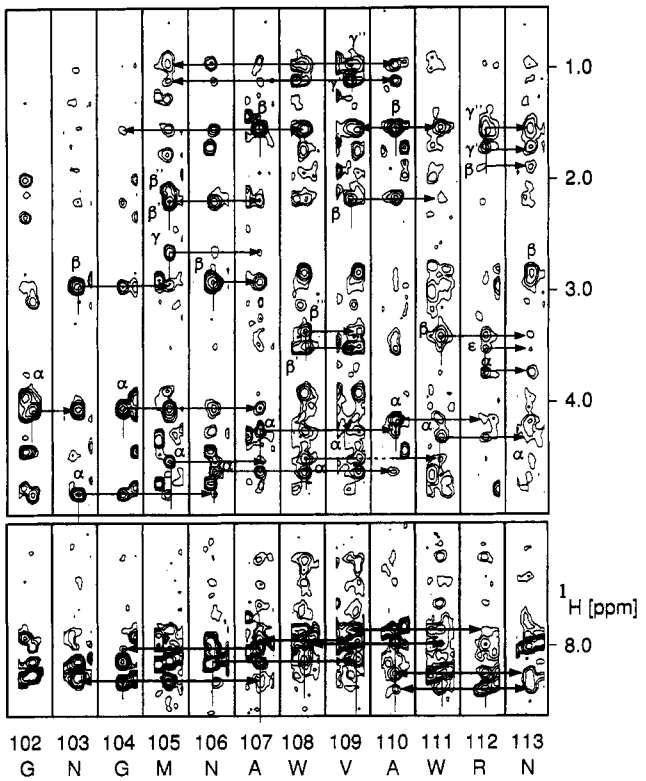

FIGURE 2: 3D NOESY-HSQC spectrum cut into ω_1, ω_3 strips at specific ^{15}N frequencies in ω_2 showing residues 102−113 in the TFE-state, recorded at 47 °C. Assignments and medium range NOEs of interest are labeled.

HSQC spectrum. Assignment of these NOEs to the side chain ^{15}N-bound atoms was possible for all 6 tryptophan indole and 13 of the 17 asparagine/glutamine NH_2 groups. In order to assign the remaining protons of the tryptophan rings, not all of which were detected in ^{15}N-edited spectra, high-resolution 2D ^{1}H−^{1}H NOESY and TOCSY experiments were recorded in 70% TFE-d_2-OH/30% H_2O and 70% TFE-d_2-OD/30% D_2O on an unlabeled sample. As with the main chain, the great majority of NOEs involving not only the side chain NH or NH_2 groups but also aromatic ring protons are to protons of residues located within +3/−2 residues in sequence. A significant number of NOEs could not be assigned within this range, however, suggesting that they may arise from longer range contacts in the TFE state, such as those made as a consequence of the disulfide bridges. However, these NOE assignments could not be made unambiguously due to the overlap of ^{1}H resonances. The resonance assignments made in this study are given in the Supporting Information.

Analysis and Interpretation of NMR Data. Figure 2 shows ω_1, ω_3 strips at the respective ^{15}N frequency in ω_2 from the 3D NOESY-HSQC spectrum for a representative helical region of the TFE state between residues 102 and 113. In the lower part of this spectrum NN(i,i+3) cross peaks from Gly104 to Ala107 and from Asn106 to Val109 are observed together with NN(i,i−3) cross peaks from Ala107 to Gly104 and NN(i,i+2) cross peaks from Val109 to Trp111 and from Ala110 to Arg112. Asn106 shows an αN(i,i+4) cross peak to Ala110, while Asn103, Gly104, and Trp108 show αN-(i,i+3) cross peaks. At the same time, the two nondegenerate resonances of the two γCH_3 groups of Val109 show γN-(i,i−4) cross peaks to Met105. The fact that the chemical shifts of these methyl groups differ significantly suggests that this residue is involved in a level of structure sufficient to generate a distinct environment for the side chain.

Spin diffusion, however, can influence NOE intensities and may be anticipated in a molecule for which the ^{15}N NMR relaxation behavior of the main chain indicates an effective global correlation time of approximately 10 ns at 37 °C (Buck, Schwalbe, and Dobson, manuscript in preparation). In order to assess whether the apparently high cross peak density in the NOESY spectrum recorded with a mixing time of 200 ms is due to extensive spin diffusion, 2D homonuclear NOESY spectra were recorded with mixing times of 50, 100, 200, and 400 ms. Several resolved NOEs from aromatic side chains to main chain and side chain protons were monitored. The data suggest that the initial rate approximation holds true for many, but not all, of the NOEs. By comparison with NOESY spectra of the native state at longer mixing times (Smith et al., 1993), an upper limit of 6.5 Å may be put on distances indicated by the weak cross peaks in the present study. Thus, while spin diffusion may influence intensities of some of the NOEs, it does not provide an explanation for the observation of a large number of medium-range αN(i,i+3) and αN(i,i+4) NOEs. Furthermore, the coexistence of αN(i,i+3) and αN(i,i+4) NOEs with αN(i,i+2) and NN(i,i+2) NOEs is anticipated from the distances of <4.4 Å for these protons in α-helical regions ($\phi = -57°$ and $\psi = -47°$) (Wüthrich et al., 1984) and is consistent with the distances found in the native helices of hen lysozyme. Similarly, the distances associated with NN-(i,i+3) NOEs is 4.8 Å in an ideal helix, and these cross peaks are indeed detected in 3D NOESY spectra of the native state. The pattern of medium-range NOEs therefore provides a clear indication of helical structure. That these effects are limited to specific regions of the polypeptide chain implies that in the remaining regions helical structure is either absent or experiences substantially different dynamic behavior.

Apart from small changes in the dynamic behavior which improve the quality of spectra at the higher temperature, there is no evidence for significant structural changes in the TFE state over this 10 °C range of temperature from either CD or NMR spectroscopy; the data at the two temperatures were therefore combined to give an overall NOE pattern for the TFE state. Figure 3 shows the pattern and intensity of the medium-range NOEs as well as the short-range NOEs used as the most prominent connections in sequential resonance assignment. To take into account the variation of cross-peak intensities found in the ^{15}N−^{1}H correlations, NOE intensities to a given amide proton were scaled to the intensity of the autocorrelation peak in the NOESY spectra. Intrinsic to the study of proteins under denaturating conditions is the incompleteness of the analysis of short- and medium-range NOEs, which arises at least in part from the limited chemical shift perturbations in 3D NOESY spectra of these states, as noted above. So, for example, in the region shown in Figure 2, αN(i,i+3,4) cross peaks from Met105 to Trp108 or Val109 could not be confirmed because of overlap of the H_α resonances of Met105 and Trp108. A similar situation occurs for many other resonances examined in this analysis.

The NOE data can be used, however, to identify approximate regions which have particular structural preferences in the TFE state. Figure 3 shows that several regions of the polypeptide chain, for example, residues 50−58 and

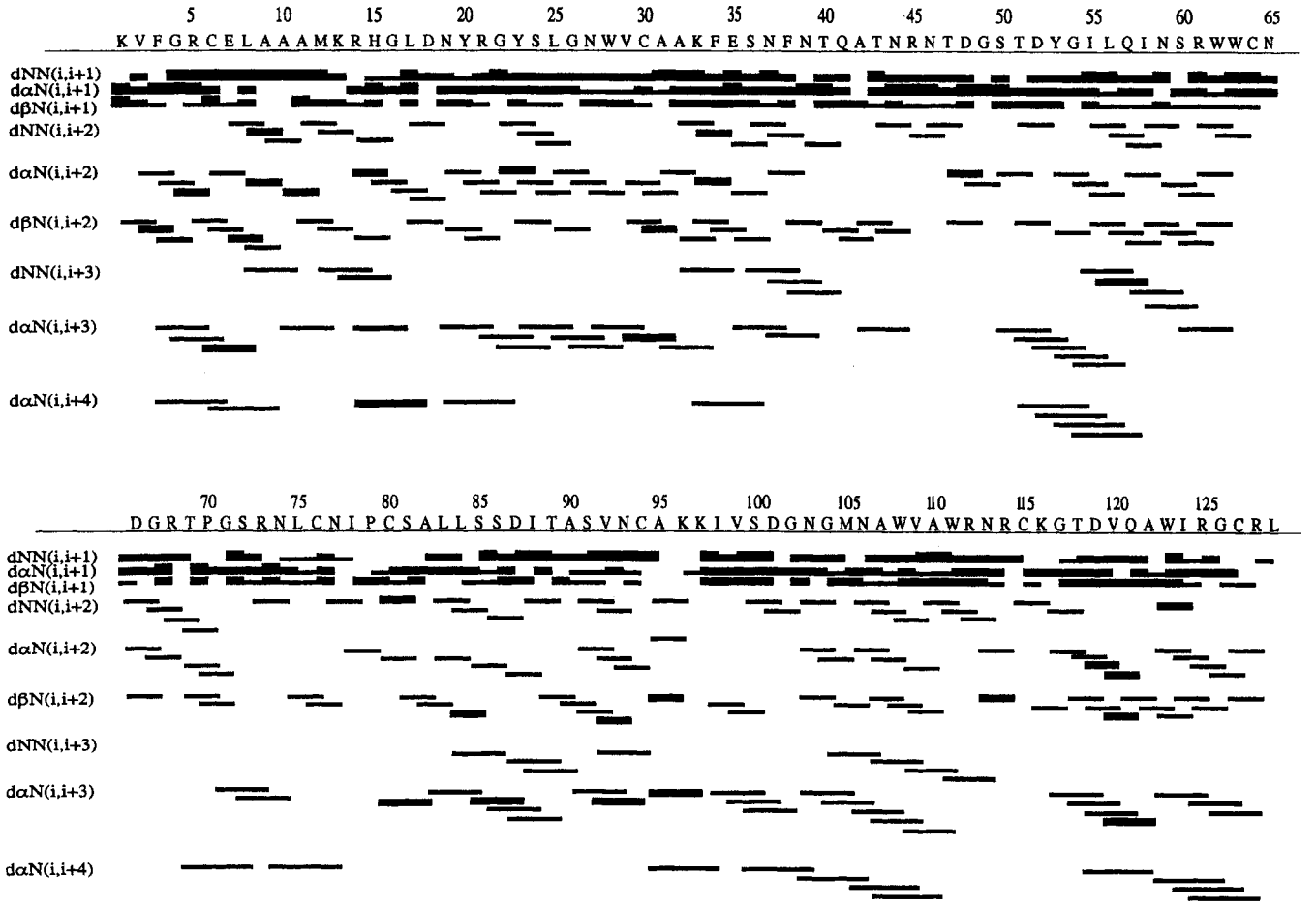

FIGURE 3: Summary of short- and medium-range NOEs in the TFE state. The data are the sum of the NOEs seen in spectra at 37 and 47 °C. Relative intensities are indicated as strong, medium, and weak.

119−129, are characterized by an extensive pattern of αN-(*i*,*i*+3) and αN(*i*,*i*+4) NOEs, indicative of α-helical structure. By contrast, other regions of the polypeptide chain are characterized by NOEs which are few in number and weak. Regions of the polypeptide encompassing residues 42−49 and 60−84, for example, appear not to adopt well defined structures for a substantial fraction of time, a conclusion which is consistent with the observation from CD spectroscopy that TFE is unable to stabilize the entire polypeptide chain in a fully helical conformation. Given the large number of medium-range NOEs in combination with their often incomplete pattern, as seen for example, for residues 3−15, it is not possible to define the location of helices (and in particularly helix termini) unambiguously in this denatured state from NOE data alone. Thus, a number of other complementary NMR parameters were analyzed.

Chemical shift perturbations of H_α resonances from random coil values ($H_\alpha^{obs} - H_\alpha^{random\ coil}$) (Wishart et al., 1995) and the chemical shift index as defined by Wishart et al. (1992a) are shown in Figure 4. The higher than average perturbation of the chemical shifts of several protons can be associated with local features of the polypeptide sequence (Figure 4a). Thus two of the most perturbed resonances of H_α protons, Thr69 and Thr80, are adjacent to proline residues 70 and 79. The H_α chemical shift of Thr69 also experiences a similar perturbation in oxidized and reduced lysozyme in the presence of 8 M urea (Schwalbe et al., in preparation), suggesting that it is caused by the proximity of Pro70 and does not reflect conformational preferences particular to the TFE state. It is interesting that other highly perturbed H_α chemical shifts (beyond 0.3 ppm) are of residues close in sequence to some of the aromatic residues in lysozyme; for example, the H_α chemical shifts of residues 29, 33, and 38 are likely to be affected by ring current shifts from the side chains of Trp28, Phe34, and Phe38. Similarly, the H_α shifts of residues 109 and 112 can be attributed to their proximity to residues Trp108 and Trp111. In other regions of the protein, however, aromatic residues such as Trp62 and Trp63, or Tyr20 and Tyr23, appear not to give rise to shifts of more than 0.2 ppm in resonances of nearby residues. Stretches of residues whose H_α resonances experience significant upfield perturbations (above −0.1 ppm) are clustered in six regions of the protein sequence (Figure 4). Using recently revised values for random chemical shifts (Wishart et al., 1995), the chemical shift index has been used to define these regions as helical segments denoted I−VI here according to the rules of Wishart et al. (1992a). Interestingly, all 12 glycine H_α resonances experience no, or only slightly downfield, shifts in the TFE state, suggesting that these residues are not part of highly populated helical or turn-like structures.

$^3J(H^N,H_\alpha)$ coupling constants measured at 58 main chain sites are shown in Figure 5a. The data set has been obtained from two types of experiments, the HMQCJ and the HNHA experiment [Kay et al. (1990) and Vuister and Bax (1994), respectively]. The data are incomplete, since both experiments rely on well resolved peaks in the 2D $^{15}N-^1H$ correlation plane. It is apparent that the majority of $^3J(H^N,H_\alpha)$ coupling constants are characteristic of helical structures. A coupling constant of 3.3 Hz, inferred from the

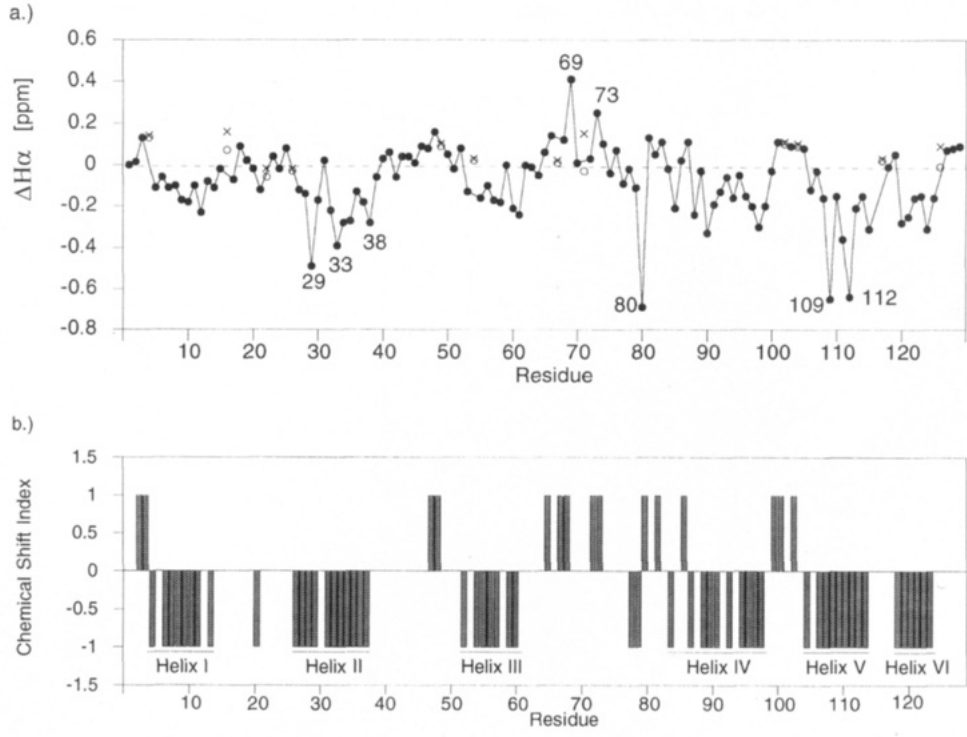

FIGURE 4: (a) H_α chemical shift perturbations from random coil values of Wishart et al. (1995) and (b) CSI (chemical shift index) of Wishart et al. (1992a,b) plotted as a function of protein sequence. The data for glycines are plotted as (○) and if nondegenerate, the second resonance as (×) in panel a to indicate their position. Residues of interest are labeled by sequence number. Regions I−VI, assigned according to CSI rules to be in helical secondary structure, are marked in panel b.

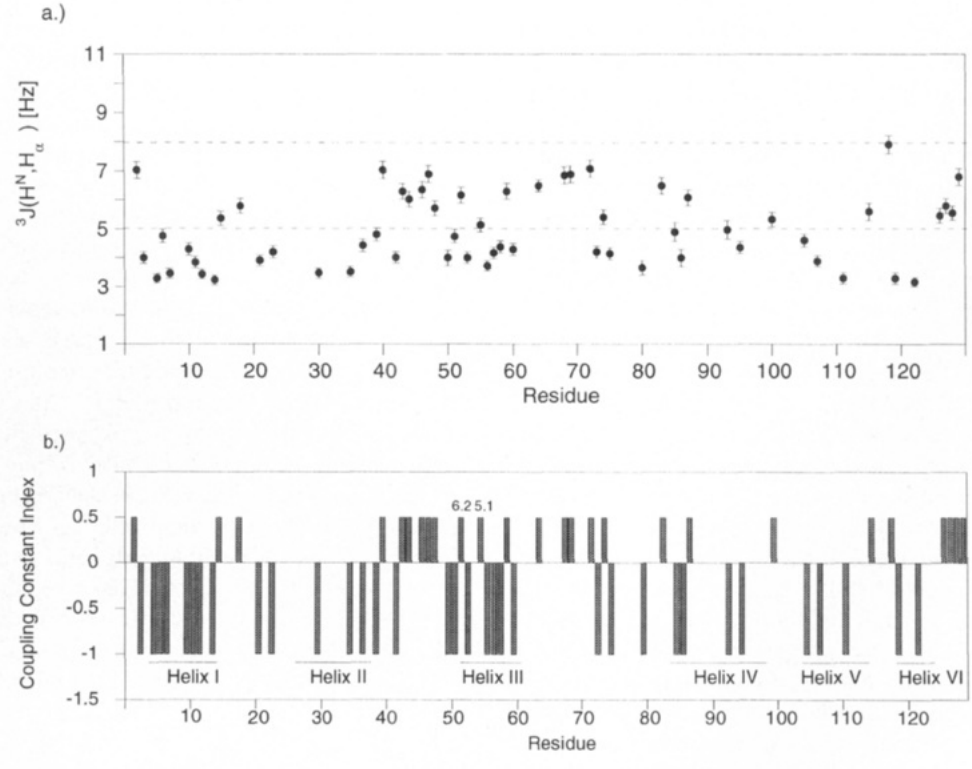

FIGURE 5: (a) 58 $^3J(H^N,H_\alpha)$ coupling constants measured in HMQCJ (Kay et al., 1989) and 3D-HNHA (Vuister et al., 1993) experiments. Coupling constants with values below 5 Hz suggest that the conformational preferences of the residue lie effectively entirely in the helical region of ϕ space. (b) Coupling constant index as described in the main text; the coupling constants for residues 51 and 55 are indicated. Regions I−VI, assigned to be in helical secondary structure using the chemical shift index (Figure 4), are marked for comparison in panel b.

conformation of an ideal helix (Wüthrich, 1986), can be used to estimate the population of the helical region of ϕ space, on a simple two-state model taking 8.5 Hz as a value of $^3J(H^N,H_\alpha)$ for a residue in an extended conformation. The estimated populations are in excess of 80% in the regions for which NOEs and shift perturbations characteristic of helical structure are detected and 30−40% elsewhere. However, this is undoubtedly a conservative estimate, as

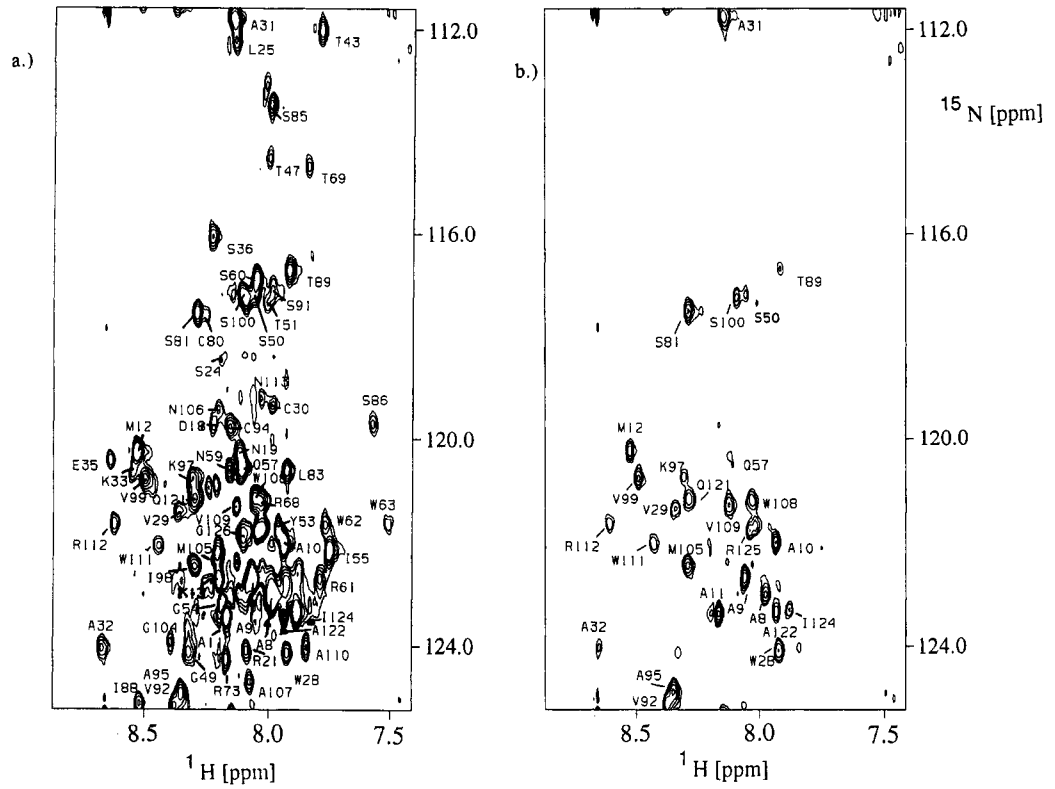

FIGURE 6: ^{15}N−^{1}H HSQC spectra recorded (a) 10 and (b) 100 min after dissolution of the protein in 70% TFE-d_2-OD/30%D_2O (v/v), pH 2.0, 37 °C. Amides with significant residual proton occupancy are labeled.

coupling constants calculated on the basis of helical regions in crystal structures are 4.8 Hz on average (Smith et al., manuscript submitted); using this value, the population of the helical region of ϕ space on the simple two-state model is calculated to be 114 ± 20% for these residues. Several residues have coupling constants larger than 5 Hz suggesting they are not part of well populated regular helical structures. Indeed, coupling constants of up to 7 Hz are detected for residues 40−48 and 64−72 indicating that these regions adopt significant nonhelical conformations or helical structures which have substantially averaged conformations in the presence of TFE. Figure 5b shows a coupling constant index, in which residues with coupling constants below 5 Hz are assigned a value of −1 and residues with averaged coupling constants of 5−8 Hz and with coupling constants above 8Hz are assigned values of +0.5 and +1.0, respectively. Good agreement is apparent when the regions with low coupling constants are compared with those defined as helical using the chemical shift index. It is interesting that the helical regions defined in this way are separated by residues with coupling constants greater than 5 Hz, largely indicative of nonhelical conformational preferences. No coupling constants greater than 8 Hz or patterns characteristic of well populated β-strand or β-turn conformations were observed.

Direct Detection of Hydrogen Exchange from the TFE Denatured State. The assignment of the spectrum of the TFE state has allowed hydrogen exchange kinetics to be measured directly. Two examples of spectra recorded 10 and 100 min after the initiation of exchange are shown in Figure 6, panels a and b. The intrinsic exchange rate for an alanine residue in an unstructured model peptide is 0.84 min^{-1} under these conditions (Bai et al., 1993). Around 60 amides have almost completely exchanged in the spectrum recorded after 10 min and are therefore likely to have protection factors less than 10. It is noticeable that the great majority of these amides are associated with resonances which have among the most intense cross peaks in the ^{15}N−^{1}H HSQC spectrum in TFE/H_2O solution (Figure 1). This indicates that they are located in regions of the polypeptide which are characterized by extensive flexibility. Indeed, this is confirmed by the analysis of the main chain ^{15}N relaxation data (Buck, Schwalbe, and Dobson, manuscript in preparation). Exchange rates for the more slowly exchanging amides were determined by fitting the time dependence of the intensity data, and protection factors were calculated by comparison of these rates to those predicted for the amides in an unstructured model peptide, taking into account the temperature dependence and side chain inductive effects (Bai et al., 1993). The protection factors of 53 amides are shown in Figure 7 and range from 2 to 250. Amides which have amongst the highest protection factors in the TFE-denatured state (P > 10) are clustered in several distinct regions of the polypeptide chain whose location is in good agreement with those identified as helical by the other NMR parameters above.

DISCUSSION

Collation of NMR Results. The combination of significant residual chemical shift dispersion of resonances, particularly of ^{15}N nuclei, and of resonance line widths sufficiently narrow to yield intense signals has been central to this study using ^{15}N-labeled hen lysozyme. Thus it has been possible to assign nearly completely the resonances of a 129-residue protein in a partially structured state and to exploit the wealth

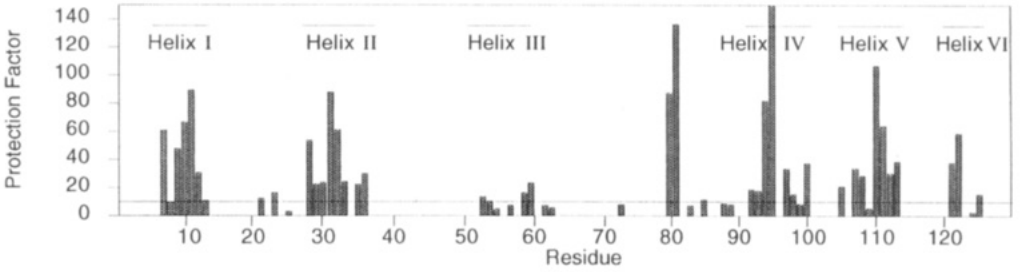

FIGURE 7: Protection factors against solvent exchange of amides in the TFE state. The location of helices, as identified by the chemical shift index (Figure 4), is indicated. Several breaks are introduced in the pattern of residues with protection factors greater than 10, because some amide resonances are very weak (residues 109, 123, and 125) or are not observed (residues 34 and 96) even in the HSQC spectrum of the fully protonated protein under these conditions. The data agree well with those reported previously for lysozyme in 50% TFE/50% H_2O, 27 °C at pH 2.0 (Buck et al., 1993), both in magnitude and in the overall pattern. Amides of residues 105, 107, and 121 were not monitored previously because they are only weakly protected in the native state. This also applies to several amides close to cysteines 6, 30, 80, and 94, while others, for example, amides of residues 80 and 81, have higher protection than observed previously. The amides of residues 64 and 65 were too weak to be observed in these experiments, and it is therefore not clear whether the disulfide 64−80 is responsible for this protection.

of detail that heteronuclear NMR methods can provide. The NMR parameters which could be obtained for the TFE denatured state of lysozyme by studying ^{15}N-labeled protein comprise the pattern of the NOEs and their intensities, H_α chemical shift perturbations, $^3J(H^N,H_\alpha)$ coupling constants, and amide hydrogen exchange rates. The interpretation of both chemical shift values and coupling constants in denatured states of proteins differs in subtle respects from that possible in native states. The measured $^3J(N^H,H_\alpha)$ coupling constants are relatively uniform in many segments of the sequence, whereas, in native proteins, a considerable local variation in their value is often observed, even if the main chain is involved in certain secondary structures (Smith et al., manuscript submitted). This arises because dihedral angles may be fixed in nonstandard regions of ϕ-space by tertiary contacts, whereas, in denatured states coupling constants of 5−8 Hz are more likely to represent the effects of conformational averaging. Similarly, while the chemical shifts in native proteins are affected by both local and long-range interactions, the H_α chemical shift pertubations observed in TFE are likely to be dominated by local effects. Upfield perturbations to the chemical shifts of H_α protons, therefore, represent primarily the contributions of helical structure. Definition of helix limits in peptides and denatured states of proteins, in particular, is complicated by the increased dynamics observed at helix termini, often leading to transient formation of additional helical turns. This is indicated by the presence of medium-range NOEs, in the absence of a complete pattern of amide protection, low coupling constants, or significant chemical shift perturbations, as seen, for example for residues 37−41, 58−61, and 126−129 in Figure 8.

A further complication arises because the majority of the data sets are incomplete in certain regions of the sequence due to the severe overlap of resonances or their low intensity due to large line widths. Thus, for example, the NOE patterns expected for helices appear to be incomplete, notably towards the latter part of the segments comprising residues 5 to 15 and residues 20 to 38. Resonance overlap of the H_α resonances for these residues (Table 1, Supporting Information) accounts in part for this observation. However, the chemical shift, hydrogen exchange protection and coupling constant data define better the C-terminal parts of these putative helices and provide complementary information that regions comprising residues 33−38 and 111−115, which are poorly defined by $\alpha N(i,i+3)$ and $\alpha N(i,i+4)$ NOEs, are indeed in helical conformations. Comparison of several NMR parameters, shown in Figure 8, is necessary to derive with confidence the location of highly populated helical structure. Perturbations of H_α chemical shifts from random coil values provide the most complete data which can be used to describe the secondary structure for the TFE denatured state, while the presence of medium-range NOEs can be used most convincingly to confirm the conformational preferences.

Figure 8 shows that there is a remarkable correspondence of the different NMR parameters which allows us to infer a consistent picture of the location of the helical structure in the TFE-denatured state (see figure legend for details). The helical structures, denoted I−VI in Figures 8 and 9, are described as having considerable perisistency when defined by a majority of the different NMR parameters and when, in addition, amide protection indicates conformational stability. Parts of the polypeptide chain which the NMR parameters partially define as helical are commonly found as extensions to these highly ordered regions of the polypeptide and presumably correspond to dynamically averaged structures at the helix termini; this is seen, for example, at the C-terminus, residues 126−129, of the protein. This correspondence of the interpretation of the NMR parameters extends to other regions of the polypeptide chain. For example, NOEs, H_α chemical shift perturbations, and hydrogen exchange protection factors as well as coupling constants all indicate that the regions encompassing residues 41−50 and 71−80 are predominantly unstructured in the TFE state.

The Nature of Structure in the TFE-Denatured State. Comparison of the six consensus regions (I−VI in Figure 8) of the lysozyme sequence possessing extensive helical character in 70% TFE/30% H_2O with the conformation of these regions of the protein in the native state (Figure 9) shows that all residues which are in helical structures in the native state are also involved in helical structure in the TFE-denatured state. Significantly, the hydrogen exchange data demonstrate that these native-like helices, and in particular their central regions, have a high degree of structural persistency, meaning that the population of any nonhelical conformational states is very low for these regions.

Certain regions of the sequence are, however, found to possess helical structures in TFE that have no counterpart in the native state. These structures are formed as extensions to the native-like helices B, C, and D and the C-terminal

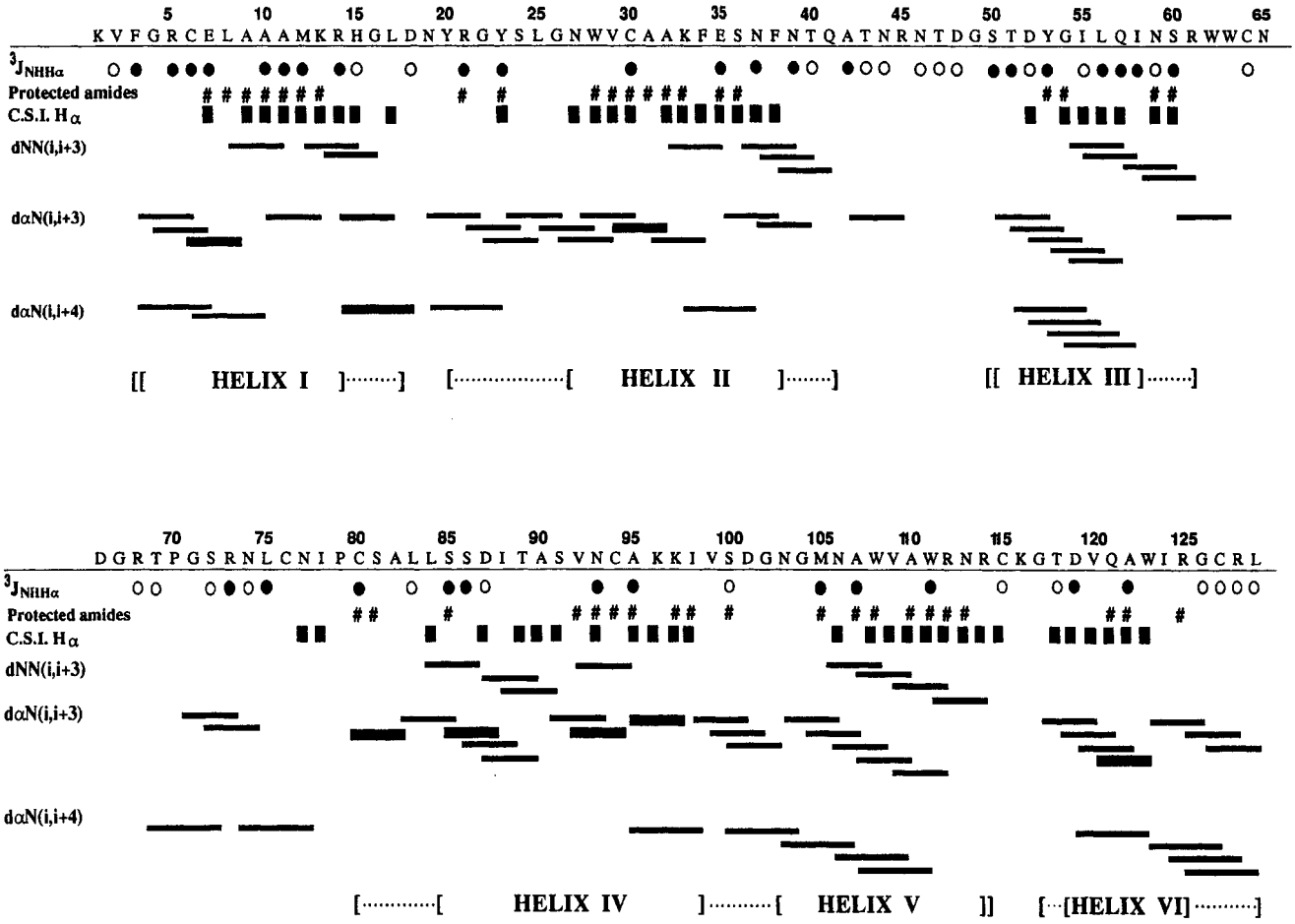

FIGURE 8: Summary of the principal NMR parameters defining helical structure in the TFE state. Medium-range NOEs, characteristic of helical structure, are indicated by horizontal lines, hydrogen exchange protection ($P > 10$) is marked (#), and significant $^1H_\alpha$ chemical shift perturbation from random coil values (CSI $\Delta\delta < -0.1$ ppm) is indicated by (■). $^3J[H^N,H_\alpha]$ coupling constants <5.0 Hz are shown as (●), and coupling constants >5.0 Hz as (○). The data indicative of helical structure are clustered in six regions of the polypeptide, denoted helix I−VI, and are collated (inner brackets) to describe the approximate limits of persistent helical secondary structure as the consensus of the majority of NMR parameters [CSI and protection, coupling constant and/or αN(i,i+3,4) NOE]. The six regions I−VI encompass residues 3−14, 27−38, 50−58, 85−98, 103−114, and 119−125. Note that the first four amides in an α-helix are usually not hydrogen bonded to other main chain hydrogen bond acceptors and that amide exchange protection may therefore be absent. Theoretical calculations suggest also that the extent of H_α chemical shift perturbation is considerably diminshed for the N-terminal four residues in an ideal helix (Osapay & Case, 1994). Helix limits due to transient structural preferences are also indicated (outer brackets).

3_{10}-helix, by up to one turn at either end. In addition, residues 50−58, which form part of the β-sheet in the native state, are also stabilized in a helical conformation in TFE. Figure 10 illustrates the structural differences as manifested in the pattern of NOEs in the NOESY-HSQC spectra for this region of the protein in the native and TFE states. Whereas in the native state spectrum there is clear evidence for medium- and long-range NOEs typical of β-sheet structure, αN(i,i+3) and NN(i,i+3) NOEs which are characteristic of helical structure are observed in the TFE state. However, this helix is not as persistent as the native-like helices as judged by the main chain amide hydrogen exchange (protection factors < 20) and ^{15}N relaxation behavior (Buck, Schwalbe, and Dobson, manuscript in preparation). Conformational preferences, other than those toward helical structures, can, furthermore, be inferred for short regions of the polypeptide chain in TFE; for example, residues 17−20 and 60−64 show extensive i,i+2 NOEs. The data suggest that these regions could populate partial turn-like structures. Residues 40−49 and 65−75, corresponding to the N-terminal part of the β-sheet and part of the long loop in the native protein, respectively, appear to be highly disordered in the TFE state.

Approximately 73 residues are involved in highly populated helical structure in the TFE-denatured state, with up to a further 30 residues likely to be involved in transient helical structures. This number is in reasonable agreement with the 80 residues inferred to be in helical structure from the far-UV CD signal under similar conditions (Buck et al., 1993). By comparison, 44 residues are part of α-helical structure in the native state. This suggests that an increase in helix regularity or changed contributions from disulphides and aromatic residues (Woody, 1985) are not the principal components responsible for the increase of the far-UV CD ellipticity in the TFE state over that of the native state. Indeed, the NMR data confirm the stabilization of several regions of the polypeptide in helical structures which have no counterpart in the native state. For example, the structure corresponding to the C-terminal 3_{10}-helix in the native state (residues 120−124) has been extended in the TFE-state right up to the C-terminus itself (residue 129), with αN(i,i+4) NOEs indicating specifically the formation of α-helical conformers (Wüthrich et al., 1984). By contrast, the 3_{10}-helices of the native protein, contribute little, if any, to the far-UV CD ellipticity given that they are short (Gans et al., 1991). A further major contribution to the CD signal of the

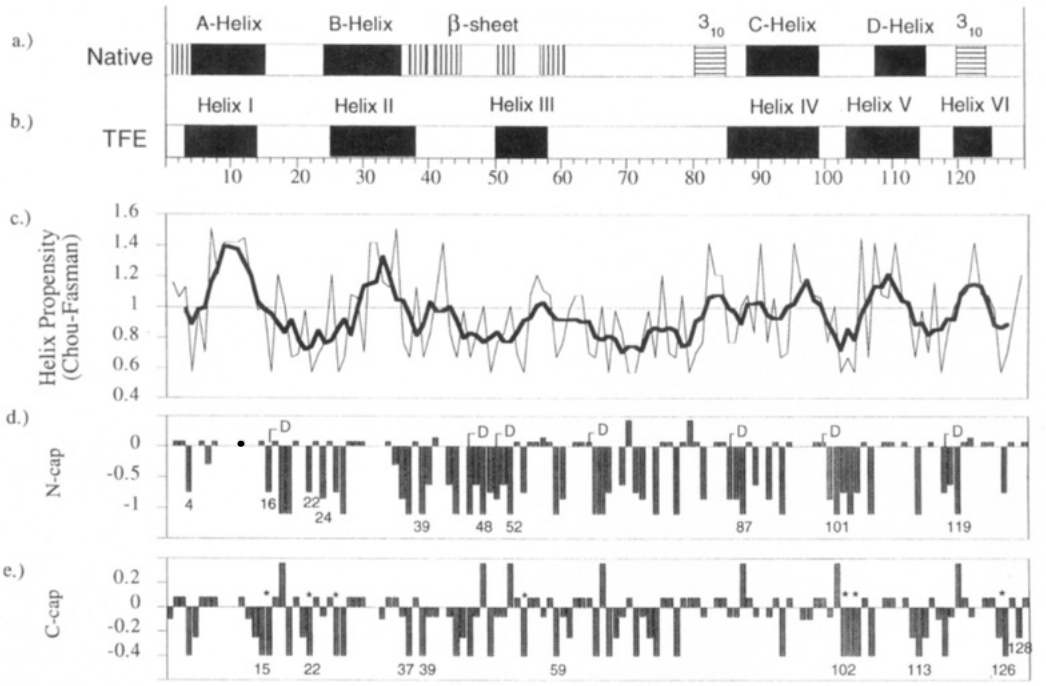

FIGURE 9: (a) Secondary structural units as defined in the native state crystal structure (Handoll, 1985) and (b) in the TFE denatured state of hen lysozyme assigned by consensus of individual NMR parameters (Figure 8). (c) Helix propensity derived using the scale of Chou and Fasman (1978) shown as individual points and averaged over a sliding window of five residues. (d) N- and (e) C-capping potential, respectively, of amino acids according to the scale of Munoz and Serrano (1994).

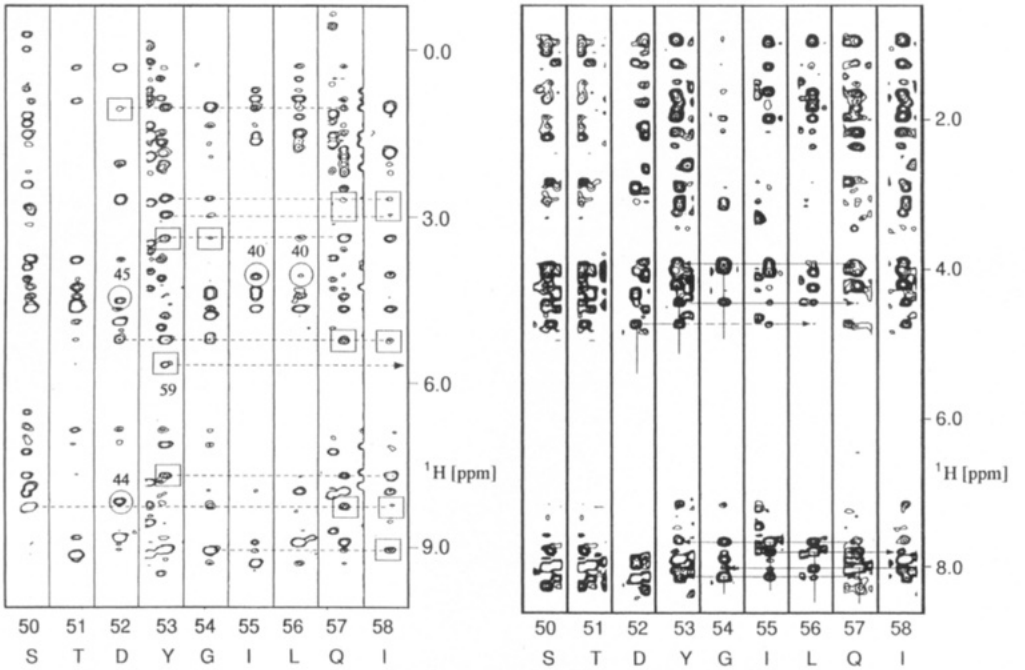

FIGURE 10: ω_1, ω_3 strips at the respective ^{15}N-frequency in ω_2 showing the region of residues 50−58 in the 3D NOESY-HSQC spectrum, (right) for the TFE state recorded at 47 °C and (b) for the native state recorded at 35 °C, pH 3.8 (left). Assignments and NOEs of interest are labeled.

TFE state is the formation of α-helical structure in part of the region corresponding to the β-sheet in the native state.

The data on the location and extent of helical structure in the TFE state are highly consistent with a recent study by circular dichroism of four peptide fragments, spanning the sequence of hen lysozyme in segments of 20−46 residues in length (Yang et al., in press). Two peptides comprising residues 1−40 and 84−129, which together correspond to the regions of the protein which form the major α-helices in the native state, were both found to adopt extensive α-helical structure in the presence of TFE. Is was not possible, however, to stabilize α-helical structure in a peptide corresponding to the long loop region of the native protein (residues 61−82) even at high TFE concentrations. By contrast, studies of a peptide corresponding to the β-sheet region of the native protein (residues 41−60) suggested that several residues may adopt α-helical conformations in the presence of 50% TFE (Yang et al., 1994). Remarkably, the sum of the far-UV CD spectra of these four peptides was found to be nearly identical to spectra obtained under identical solution conditions for the intact oxidized protein and to those of a derivative of lysozyme in TFE in which

the disulfide bonds had been reduced and blocked by methylation (Yang et al., in press). These results imply that the determinants which allow the polypeptide chain to adopt stable helical structures must be associated with highly localized features of the amino acid sequence.

Comparison with Predictions for Local Structural Propensities. The profile for intrinsic helical propensity derived using the scale of Chou and Fasman (1978) is shown in Figure 8c; similar profiles are obtained using the helix propensity scales of Richardson and Richardson (1988), O'Neil & Degrado (1990), Wojcik et al. (1990), Horowitz et al. (1992), Chakrabartty et al. (1994), and of Blaber et al. (1994). Regions predicted to have high intrinsic helical propensity are all found to be in helical conformations in both the TFE and native states (Figure 8a,b). There is little difference in the helical propensities for regions which form α- and 3_{10}-helices in the native state, suggesting that their structure in the native fold could be influenced by other interactions. Of particular interest is that a part of the region which forms the triple stranded β-sheet in the native structure, residues 50−58, has higher than average helical propensity on five of the seven scales examined; this is the region of the sheet found to adopt helical structure in the TFE state of the intact protein. Thus the overall correspondence between the regions of the polypeptide chain which have adopted helical conformations in TFE and those with significant intrinsic propensity for helical structure is excellent. A more detailed comparison, however, shows several exceptions to this. For example, the N- and C-terminal extensions to the B-helix are not anticipated on the basis of the predicted propensities of these residues. Furthermore, higher than average helical propensity is seen for residues 41−44 and 61−64 in patterns derived using the majority of the propensity scales but is not accompanied by helix formation in the TFE-denatured state. Helical propensity in the region of residue 80−84 is fulfilled by helix formation in the native state but not in the TFE state where the helix begins near residue 85.

These discrepancies may be due to the fact that the propensity scales have been derived for individual residues located in the central regions of helices and do not explicitly take into account medium-range interactions involving the side chains of a helix (Scholtz & Baldwin, 1992). Interactions between side chains at the termini of helices and the main chain have, for example, been found to confer significant stability to helix formation (Serrano et al., 1992). The details of some of these capping interactions are, however, complicated (Harper & Rose, 1993; Harpaz et al., 1994). A simplified profile of sites with N- and C-capping potentials is plotted as a function of the protein sequence in Figure 8 panels d and e using a scale derived from energetic analysis of site-directed helix capping mutants (Munoz & Serrano, 1994). The patterns constructed using the scales of Richardson and Richardson (1988), Presta and Rose (1988), Dasgupta and Bell (1993) are closely similar. It is noticeable that while the A- and B-helices have helix limits in the native state which coincide with the positions of amino acids suitable for capping, such correspondence is much poorer for the termini of the C- and particularly the D-helix. In these cases alternative interactions, part of the native tertiary fold, must compensate for the loss of these potentially favorable interactions in the native state. Ile98, Met105, and Trp108, for example, are buried in the hydrophobic core of the protein, which is likely to result in conformations unsuitable for helix propagation.

By contrast to these comparisons with the structures in the native conformation of hen lysozyme, the limits of helical structure as identified by the NMR parameters for the TFE state are found to coincide closely with the positions of residues which are predicted to participate favorably in helix capping interactions. Aspartic and glutamic acid residues, for example, are preferred amino acids at the first or second position of a helix, reinforcing the capping effect of side chains at the N-terminus (Dasgupta & Bell, 1993), and the location of aspartic acids at residues 52, 87, 101, and 119 appears to be in excellent agreement with the approximate positions of the N-terminal residues of helices III, IV, V, and VI in the TFE-state. However, as seen for the region corresponding to the long loop of the native state, propensities to form favorable helix capping interactions appear to be insufficient on their own for helix formation in TFE. Overall, the data on hen lysozyme reinforce the notion that TFE stabilizes helical structures only in regions of the polypeptide chain which have considerable local propensity to adopt helical conformations. This conclusion is consistent with previous studies of peptides and protein fragments (Lehrman et al., 1990; Dyson et al., 1992; Sönnichsen et al., 1992). In addition, other studies have demonstrated that helix capping interactions persist in the cosolvent and act as helix stop signals (Nelson & Kallenbach, 1986, 1989; Bruch et al., 1991; Storrs et al., 1992; Shin et al., 1993; Toumandjie & Johnson, 1994).

TFE is thought to denature native states of proteins in a manner analogous to that of other alcohols by the disruption of tertiary interactions, presumably by weakening the hydrophobic effect relative to hydrogen-bonding interactions (Mizuno et al., 1984; Nelson & Kallenbach, 1986; Storrs et al., 1993; Thomas & Dill, 1993). A large number of studies suggest that TFE preferentially stabilizes α-helical structures [e.g., Nelson and Kallenbach (1989) and Lehrman et al. (1990)]. TFE has also been reported to stabilize β-turn and hairpin structures (Sönnichsen et al., 1992; Shin et al., 1993; Blanco et al., 1994) and to generate elements of structure related to those in a molten globule state in water (Smith et al., 1994). A bias toward helical structures arises because these structures involve only local interactions, while β-sheet structures are often stabilized by specific nonlocal hydrophobic interactions between side chains (Thomas & Dill, 1993). The switch from β- to α-structure on addition of TFE for residues 50−58, as well as similar phenomena in other proteins, can be rationalized on the basis of local chain propensities toward α-helical conformations (Sönnichsen et al., 1992; Lui et al., 1994, Shiraki et al., 1995). While it is likely that propensity scales and prediction methods are still limited in a number of respects, and that the structures which can be formed are still incompletely understood, the correspondence between the structures observed in an environment such as TFE with those predicted on the basis of local interactions is remarkable and suggests that these structures may be of relevance in early stages of protein folding before long-range interactions are established.

Significance for Understanding the Nature of Folding Intermediates. The majority of polypeptide sequences do not adopt stable structures when considered as isolated peptides in aqueous solutions (Dyson & Wright, 1991). Extensive secondary structure is formed early in protein

folding in a process thought to be associated with the compaction of the polypeptide chain, for example, as part of a hydrophobic collapse (Dill & Shortle, 1991). Nonspecific interactions generated in such environments appear to exist in early folding intermediates and molten globule states (Ptitsyn, 1992; Jennings & Wright, 1993; Dobson et al., 1994) and are thought to be sufficient to stabilize polypeptides in distinct conformations. Nonspecific effects of TFE can be considered in some respects as analogues of these interactions because the structures which are stabilized share a number of the characteristics of these folding intermediates. For example, a significant feature of the TFE states of hen lysozyme as well as those of α-lactalbumins and other proteins [Buck et al. (1993), and Alexandrescu et al. (1994a), and Shiraki et al. (1995), respectively)] is the high content of α-helical structure. However, preferential stabilization of α-helical structure by TFE may also account for the fact that the protection factors against amide hydrogen exchange with solvent observed in the TFE state are higher than those seen for other partially structured states in aqueous solution (Jeng & Englander, 1991; Jennings & Wright, 1993; Buck et al., 1994) and much higher than those seen in the earliest species sampled during refolding (Radford et al., 1992).

In addition to the formation of largely native-like structure, TFE also stabilizes regions of the hen lysozyme sequence in conformations which are compatible with local structural preferences but have no counterpart in the native state. Similar observations have recently been made in a number of other studies on proteins and peptides under equilibrium conditions (Neri et al., 1992; Sönnichsen et al., 1992; Kemmink & Creighton, 1993; Lui et al., 1994; Alexandrescu et al., 1994a; Arcus et al., 1994). Nonnative structures are likely to arise because the specific interactions which contribute to their formation and stability are entirely local, while additional nonspecific interactions are also provided by the rest of the polypeptide or by cosolvents or denaturants. By contrast, the formation of the corresponding native structures involves a significant proportion of specific, often long-range interactions. Local nonnative structures may, therefore, precede and indeed compete with the higher order structures formed later in folding. In the case of hen lysozyme, for example, it has been shown for the majority of molecules that the formation of the native structure in the β-subdomain of the protein proceeds at a much slower rate than the formation of stable structure in the α-subdomain, when refolding takes place from a variety of denatured states (Radford et al., 1992; Kotik et al., 1995). The possibility that a transient population of nonnative structure in the β-sheet region could be responsible for these observations, and that these structures may be similar to those formed in TFE, is an intriguing one. However, an important result of the present study concerning the nonnative structures detected in TFE is their relative instability compared to the native-like structures. Indeed, the most persistent structures which are generated by local interactions in a cosolvent environment are found to be entirely native-like in this partially ordered state. This supports the view that native, not nonnative or misfolded, structures are likely to be in general the major determinants of the folding pathways and kinetics (Buck et al., 1993; Kemmink & Creighton, 1993; Baldwin 1995).

It is interesting to note that the most persistent structures in the TFE state of hen lysozyme appear to be organized as helices in regions corresponding to the α-subdomain of the native protein. This resembles, at least superficially, the structure found in the major intermediate populated during kinetic refolding of the protein (Radford et al., 1992). The later kinetic refolding intermediates, however, protection factors against amide hydrogen exchange exceed 500, and the α-helical subdomain becomes stabilized as a single cooperative event which also involves amides in loops, for example, residues 17 and 23. The lack of persistent structure involving these residues of the polypeptide chain in the presence of TFE, and the smaller magnitude of amide protection overall, indicates that this partially folded state is substantially less ordered and native-like than the major kinetic intermediate. This is consistent with the view that the persistent and more structured states which are populated during the later stages of protein folding involve specific interactions within substantially compact structures, such as those generated by native-like contacts between helices as well as the stable secondary structure such as found for a protein in the presence of TFE (Dobson et al., 1994). Such observations support a model of protein folding which is based predominantly on a hierarchical process of native-like structure formation and stabilization.

ACKNOWLEDGMENT

We thank Donald A. MacKenzie, Andrew Spencer, David J. Jeenes and David B. Archer of the Agricultural and Food Research Council, Institute of Food Research, Norwich, U.K., for the supply of ^{15}N-labeled hen lysozyme and Christina Redfield and Jonathan Boyd for help with NMR spectroscopy, data analysis, and useful discussion.

SUPPORTING INFORMATION AVAILABLE

One table giving assignments of ^{15}N and ^{1}H resonances for the TFE-denatured state in 70% TFE/H_2O (v/v), pH 20, at 37 °C (4 pages). Ordering information is given on any current masthead page.

REFERENCES

Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J., & Dobson, C. M. (1993) *Biochemistry 32*, 1707−1718.
Alexandrescu, A. T., Ng, Y.-L., & Dobson, C. M. (1994a) *J. Mol. Biol. 235*, 587−599.
Alexandrescu, A. T., Abeygunawardana, C., & Shortle, D. (1994b) *Biochemistry 33*, 1063−1072.
Arcus, V., Vuilleumier, S., Freund, S. M. V., Bycroft, M., & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9412−9416.
Bai,Y., Milne, J. S., Mayne, L., & Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86
Baldwin, R. L. (1995) *J. Biomol. NMR. 5*, 103−109.
Barrick, D., & Baldwin, R. L. (1993) *Protein Sci. 2*, 869−876.
Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7−13.
Blaber, M., Zhang, X.-J., Lindstrom, J. D., Pepiot, S. D., Baase, W. A., & Matthews, B. W. (1994) *J. Mol. Biol. 235*, 600−624.
Blanco, F. J., Jimenez, A., Pineda, A., Rico, M., Santoro, J., & Nieto, J. L. (1994) *Biochemistry 33*, 6004−6014.
Bruch, M. D., Dhingra, M. M., & Gierasch, L. M. (1991) *Proteins: Struct., Funct., Genet. 10*, 130−139.
Buck, M., Radford, S. E., & Dobson, C. M. (1993) *Biochemistry 32*, 669−678.
Buck, M., Radford, S. E., & Dobson, C. M. (1994) *J. Mol. Biol. 237*, 247−254.
Buck, M., Boyd, J., Redfield, C., MacKenzie, D. A., Jeenes, D. J., Archer, D. B., & Dobson, C. M. (1995) *Biochemistry 34*, 4041−4055.

Chakrabartty, A., Kortemme, T., & Baldwin, R. L. (1994) *Protein Sci. 3*, 843−852.
Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251−276.
Dagsgupta, S., & Bell, J. (1993) *Int. J. Pept. Protein Res. 41*, 499−511.
Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795−825.
Dobson, C. M. (1991) *Curr. Opin. Struct. Biol. 1*, 22−27.
Dobson, C. M. (1994) *Curr. Biol. 4*, 636−640.
Dobson, C. M., Evans, P. A., & Williamson, K. L. (1984) *FEBS Lett. 168*, 331−334.
Dobson, C. M., Evans, P. A., & Radford, S. E. (1994) *Trends Biochem. Sci. 19*, 31−37.
Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 3542−3556.
Dyson, H. J., & Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem 20*, 519−538.
Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., & Wright, P. E. (1992) *J. Mol. Biol. 226*, 795−817.
Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521−655.
Evans, P. A., Topping, K. D., Woolfson, D. N., & Dobson, C. M. (1991) *Proteins: Struct., Funct., Genet. 9*, 248−266.
Feng, Y., Sligar, S. G., & Wand, A. J. (1994) *Nature: Struct. Biol. 1*, 30−35.
Fersht, A. R. (1993) *FEBS Lett. 325*, 5−16.
Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B., & Feeney, J. (1990) *J. Magn. Reson. 90*, 420−425.
Gans, P. J., Lyu, P. C., Manning, M. C., Woody, R. W., & Kallenbach, N. R. (1991) *Biopolymers 31*, 1605−1614.
Handoll, H. H. G. (1985) D. Phil. Thesis, University of Oxford, Oxford, U.K.
Harbison, G. S. (1993) *J. Am. Chem. Soc. 115*, 3026−3027.
Harding, M. M., Williams, D. H., & Woolfson, D. N. (1991) *Biochemistry 30*, 3120−3128.
Harpaz, Y., Elmasry, N., Fersht, A. R., & Henrick, K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 311−315.
Harper, E. T., & Rose, G. D. (1993) *Biochemistry 30*, 7605−7607.
Horovitz, A., Matthews, J. M., & Fersht, A. R. (1992) *J. Mol. Biol. 227*, 560−568.
Ikura, M., Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 204−209.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546-4553.
Jeng, M.-F., & Englander, S. W. (1991) *J. Mol. Biol. 221*, 1045−1061.
Jennings, P. A., & Wright, P. E. (1993) *Science 262*, 892−896.
Kay, L. E., & Bax, A. (1989) *J. Magn. Res. 86*, 110−126.
Kay, L. E., Ikura, M., Tschudin, R., & Bax, A. (1990) *J. Magn. Reson. 89*, 496−514.
Kemmink, J., & Creighton, T. E. (1993) *J. Mol. Biol. 234*, 861−878.
Kotik, M., Radford, S. E. & Dobson, C. M. (1995) *Biochemistry 34*, 1714−1724.
Lehrman, S. R., Tuls, J. L., & Lund, M. (1990) *Biochemistry 29*, 5590−5596.
Logan, T. M., Theriault, Y., & Fesik, S. W. (1994) *J. Mol. Biol. 236*, 637−648.
Liu, Z.-P., Rizo, J., & Gierasch, L. M. (1994) *Biochemistry 33*, 134−142.
Mayo, S. L., & Baldwin, R. L. (1993) *Science 262*, 873−876.
Merutha, G., Dyson, H. J., & Wright, P. E. (1995) *J. Biomol. NMR 5*, 14−24.
Mizuno, K., Kaido, H., Kimura, K., Miyamoto, K., Yoneda, N., Kawabata, T., Tsurusaki, T., Hashizume, N., & Shindo, Y. (1984) *J. Chem. Soc., Faraday Trans. I 80*, 879−894.
Munoz, V., & Serrano, L. (1994) *Nature: Struct. Biol. 1*, 399−409.
Nelson, J. W., & Kallenbach, N. R. (1986) *Proteins: Struct., Funct., Genet. 1*, 211−217.
Nelson, J. W., & Kallenbach, N. R. (1989) *Biochemistry 28*, 5256−5261.
Neri, D., Billeter, M., Wider, G., & Wüthrich, K. (1992) *Science 257*, 1559−1563.
O'Neil, K. T., & Degrado, W. F. (1990) *Science 250*, 646−651.
Osapay, K., & Case, D. A. (1994) *J. Biomol. NMR 4*, 215−230.
Otzen, D. E., Itzhaki, L. S., ElMasry, N. F., Jackson, S. E., & Fersht, A. F. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10422−10425.
Peng, Z., & Kim, P. S. (1994) *Biochemistry 33*, 2136−2141.
Ptitsyn, O. B. (1992) in *Protein Folding* (Creighton, T. E., Ed.) Freeman, New York.
Presta, L. G., & Rose, G. D. (1988) *Science 240*, 1632−1641.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302−305.
Redfield, C., & Dobson, C. M. (1988) *Biochemistry 27*, 122−136.
Redfield, C., Smith, R. A. G., & Dobson, C. M. (1994) *Nature: Struct. Biol. 1*, 23−29.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648−1652.
Roberts, I. N., MacKenzie, D. A., Jeenes, D. B., Radford, S. E., Robinson, C. V., Aplin, R. T., & Dobson, C. M. (1992) *Biotechnol. Lett. 14*, 897−902.
Roder, H. (1989) *Methods Enzymol. 176*, 446−473.
Scholtz, J. M., & Baldwin, R. L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 95−118.
Serrano, L., Sancho, J., Hirshberg, M., & Fersht, A. R. (1992) *J. Mol. Biol. 227*, 544−559.
Shin, H.-C., Merutka, G., Waltho, J. P., Wright, P. E., & Dyson, H. J. (1993) *Biochemistry 32*, 6348−6355.
Shiraki, K., Nishikawa, K., & Goto, Y. (1995) *J. Mol. Biol. 245*, 180−194.
Smith, L. J., Sutcliffe, M. J., Redfield, C., & Dobson, C. M. (1993) *J. Mol. Biol. 229*, 930−944.
Smith, L. J., Alexandrescu, A. T., Pitkeathly, M., & Dobson, C. M. (1994) *Structure 2*, 703−712.
Sönnichsen, F. D., Van Eyk, J. E., Hodges, R. S., & Sykes, B. D. (1992) *Biochemistry 31*, 8790−8798.
Sosnick, T. R., Mayne, L., Hiller, R., & Englander, S. W. (1994) *Nature: Struct. Biol. 1*, 149−156.
Stockman, B. J., Euvrard, A., & Scahill, T. A. (1993) *J. Biomol. NMR 3*, 285−296.
Storrs, R. W., Truckses, D., & Wemmer, D. E. (1992) *Biopolymers 32*, 1695−1702.
Thomas, P. D., & Dill, K. (1993) *Protein Sci. 2*, 2050−2065.
Toumadje, A., & Johnson, W. C. (1994) *Biopolymers 34*, 969−973.
van Mierlo, C. P. M., Darby, N. J., Keeler, J., Neuhaus, D., & Creighton, T. E. (1993) *J. Mol. Biol. 229*, 1125−1146.
Vuister, G. W., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
Wider, G., Neri, D., & Wüthrich, K. (1991) *J. Biomol. NMR 1*, 93−98.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992a) *Biochemistry 31*, 1647−1651.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992b) *J. Mol. Biol. 222*, 311−333.
Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., & Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
Wojcik, J., Altman, K.-H., & Scheraga, H. A. (1990) *Biopolymers 30*, 121−134.
Woodward, C. (1994) *Curr. Opin. Struct. Biol. 4*, 112−116.
Woody, R. W. (1985) in *The Peptides*, Vol. 7, pp 15−114, Academic Press, New York.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
Wüthrich, K. (1994) *Curr. Opin. Struct. Biol. 4*, 93−99.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715−740.
Yang, J. J., Pitkeathly, M., & Radford, S. E. (1994) *Biochemistry 33*, 7345−7353.
Yang, J. J., Buck, M., Pitkeathly, M., Kotik, M., Haynie, D. T., Dobson, C. M., & Radford, S. E. (1995) *J. Mol. Biol.* (in press).
Zhang, O. W., Kay, L. E., Olivier, J. P., & Forman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.
Zhu, G., & Bax, A. (1990) *J. Magn. Reson. 90*, 405−410.

BI951342Y