Toward a Description of the Conformations of Denatured States of Proteins. Comparison of a Random Coil Model with NMR Measurements

Klaus M. Fiebig, Harald Schwalbe, Matthias Buck,† Lorna J. Smith, and Christopher M. Dobson*

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

Received: September 18, 1995; In Final Form: December 11, 1995[⊗]

A strategy is proposed to describe the backbone conformations sampled in denatured states of proteins. Main chain dihedral angle distributions are extracted from the protein data base and used to predict NMR parameters such as coupling constants and NOE intensities. A simple model in which each residue samples its ϕ,ψ distribution noncooperatively has been found to reproduce many of the features of experimental NMR data for hen egg-white lysozyme denatured in 8 M urea at low pH. This model provides a framework which allows identification of residual structure inherent in experimental data of nonnative states of proteins. The effects of introducing local conformational cooperativity on NMR parameters are discussed and analyzed in light of the experimental data for lysozyme.

Introduction

Over the past few years interest in characterizing denatured states of proteins has grown rapidly, particularly because of their importance in studies of protein folding and protein stability;[1−3] such states are also believed to play a role during transport of proteins across membranes and in intracellular protein turnover.[4] Nuclear magnetic resonance (NMR) techniques have been instrumental in the study of the structural properties of denatured states at a residue specific level. Studies using conventional homonuclear 2D methods are, however, hampered by the limited chemical shift dispersion which leads to extreme overlap of the NMR resonances.[5] Despite this, in some cases at least partial resonance assignment has been possible using magnetization transfer techniques.[6,7] Insight has also come from hydrogen exchange studies[8−10] and 2D NMR structural characterizations of smaller peptide fragments excised from the sequences of proteins.[11−17] The development of heteronuclear 3D NMR techniques has, however, had a major impact enabling extensive site-specific characterization of denatured states of proteins.[3,18−25] Near-complete resonance assignments can be obtained which allow the extraction of conformational sensitive parameters such as coupling constants, chemical shifts, and homonuclear (NOE) and heteronuclear relaxation data. For native states such NMR parameters are readily interpretable in terms of specific features of protein structures.[26] Interpretation of data derived from denatured states, however, is often more complex due to the conformational heterogeneity of these states. If the interconversion between conformers is fast on the NMR time scale then the observed NMR parameters will be averaged;[27] even in the case of highly disordered states, however, this does not lead to the complete absence of NOE's.[18,21,24,28] One strategy used to overcome the complexity of data analysis is to interpret chemical shifts in terms of their deviation from so-called "random coil" chemical shifts measured experimentally in small unstructured tetrapeptides.[26,29−31] Patterns of deviations from random coil chemical shifts are taken to be indicative of regions of residual structure, which may be involved in preferential interactions with neighboring residues. This strategy, however, does not provide a detailed description of the conformations sampled in these areas of residual structure.

Here a different approach is pursued in which NMR parameters are predicted by explicitly modeling denatured state conformations and their relative energies (populations). Weights for the sampling of conformational space could be defined by a number of methods but here are derived from main chain dihedral angle distributions extracted from the protein data base. Experimental NMR data for hen egg-white lysozyme denatured in urea at low pH and for peptide fragments taken from the lysozyme sequence allow evaluation of these predictions and the underlying models. We find significant agreement between experimental and predicted NMR parameters for a model based on the derived weights which does not assume any cooperativity in the structure. The model provides a powerful framework for distinguishing random from nonrandom (residual) structure in proteins under denaturing conditions. Residual structure observed experimentally in denatured lysozyme and its peptide fragments may be due to local changes in dihedral propensities which may for instance arise from a change in the ionization state of the residue. Alternatively, residual structure may be due to cooperative effects between residues. We put forward a model which introduces cooperativity between sequential residues in the chain and evaluate its effects on predicted NMR parameters.

Methods

(a) ϕ,ψ Distributions. In recent years, a number of models[32−35] have utilized main chain ϕ,ψ dihedral distributions extracted from the protein data bank[36,37] to describe local conformational preferences in the polypeptide chain. In this context the term "local" refers to interactions between neighboring residues in the sequence. Most of these models assume that averaging over a large database of ϕ,ψ dihedral conformations will eliminate specific nonlocal interactions present in each of the individual protein structures.[38] Experimental support for this assumption has been recently reported.[39,40] Figure 1 demonstrates that ϕ,ψ populations differ significantly for individual amino acid types which indicates that the latter exhibit specific local conformational preferences.[41,42] The majority of these differences can be rationalized in terms of local and intraresidue steric or electrostatic interactions. For example, a

† Present address: Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138.

* To whom correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* January 15, 1996.

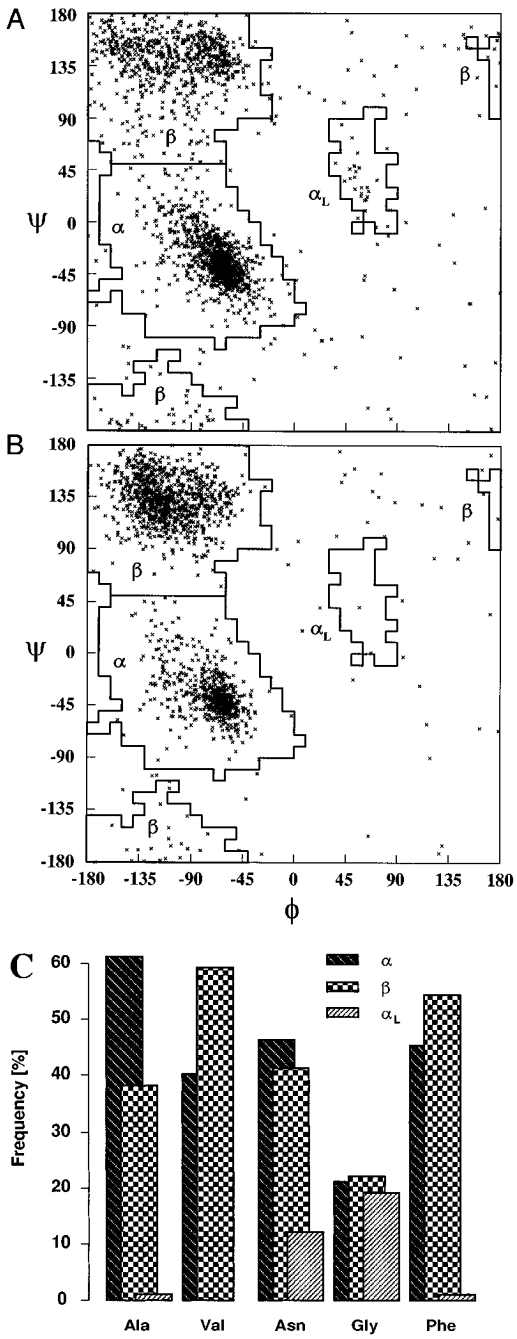

Figure 1. ϕ,ψ Ramachandran plot for alanine (A) and valine (B). Each point represents one of the 2096 and 1728 entries found for alanine and valine, respectively. α, left handed α (α_L), and β regions as defined by Morris et al.[58] are indicated in the figures. The overall number of residues populating the β region is higher for valine (B) than for alanine (A). (C) Histogram of α, α_L, and β populations.

comparison of the ϕ,ψ distributions for alanine and valine (Figure 1, A and B, respectively) shows a marked increase in the population of the less sterically crowded β region for valine, reflecting its bulky β branched side chain. In addition, as Figure 1C demonstrates, there is a significant population of the allowed left-handed α (α_L) region for glycine and asparagine which is greatly reduced or absent for the rest of the amino acids. The large (17%) population of α_L for glycine has been rationalized by it being the least sterically restricted amino acid. For asparagine (7% α_L) it has been suggested that the α_L conformation is stabilized by H bonding of the side chain NH with the backbone carbonyl of the adjacent residues.[38]

The random coil model proposed in the present study is based on the distributions of ϕ,ψ angles obtained from 85 high-resolution crystal structures[43] in the protein data bank (PDB).[36,37]

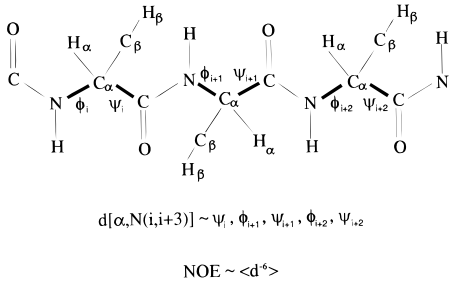

Figure 2. The peptide model. In order to derive distances as a function of the intervening ϕ,ψ angles, a pentapeptide model with explicit backbone atoms was used. Bonds with torsion angles that have rotational freedom are indicated by thick lines. Side chain H_β atoms (except for glycine) are represented by an united atom positioned at the center of mass of the hydrogens. Repulsive van der Waals interactions are taken into account using standard cutoffs.[59] The distance between the H_α of residue i and the H^N of residue $i + 3$ is a function of the dihedrals ψ_i, ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2} as indicated. The NOE intensity is proportional to the inverse sixth power of the proton−proton distance, d.

Swindells *et al.*[38] have used an identical set of protein structures which have a resolution of 2.0 Å or better, R factors of 20% or better, and sequence similarities of 30% or less. In order to calculate NMR observables, the main-chain dihedral distributions found in these 85 proteins are used directly in the form of amino acid specific libraries of ϕ,ψ dihedral angle pairs, each pair depicted as a point in Ramachandran plots of the type seen in Figure 1, A and B. A model which samples dihedral angles at random may thus approximate a maximally disordered polypeptide while retaining as a statistical average some of the conformational biases due to the local interactions of the chain. It is assumed that interconversion between the different ϕ,ψ conformations is fast on the NMR time scale so that NMR parameters are averaged.

(b) Calculation of Coupling Constants. $^3J_{HN\alpha}$ coupling constants were predicted directly from the ϕ distribution for each amino acid using the Karplus equation:[44−46]

$$^3J_{HN\alpha} = 6.4\cos^2(\phi - 60°) - 1.4\cos(\phi - 60°) + 1.9 \quad (1)$$

First, individual coupling constants $^3J_{HN\alpha}$ were calculated for each ϕ,ψ pair of the distribution. Then, in the second step, the population average $\langle ^3J_{HN\alpha}\rangle$ was obtained by averaging the individual $^3J_{HN\alpha}$ values.

(c) Calculation of Random Coil NOE Intensities. The prediction of NOE intensities requires an atomic level description of the peptide residues. We have used a simplified peptide model shown in Figure 2 in the present study. It consists of an all atom peptide backbone representation (atoms N, H^N, C_α, H_α, C, O) on to which C_β and H_β atoms are added for all amino acids but glycine, for which both H_α atoms are modeled explicitly. All H_β atoms are treated as united atoms positioned collinearly with the $C_\alpha-C_\beta$ bond placed at the center of mass of the H_β atoms, in order to account approximately for χ_1 averaging. Repulsive van der Waals interactions between N, H^N, C_α, H_α, C, O, C_β, and H_β atoms are modeled explicitly by disallowing steric overlap between these atoms, while interactions involving other side chain atoms (C_γ, C_δ, etc.) are neglected and are not included in the current model. We are presently extending the sophistication of the model to include an all-atom side chain representation which will provide a more accurate description of side chain protons.

Ensembles of conformations are generated by a Monte Carlo procedure which for each conformation generates a new set of ϕ,ψ backbone angles. This procedure selects ϕ,ψ angles at random from the amino acid specific dihedral angle libraries.

All other bond and torsion angles are kept at their ideal (average) values. Conformers with steric overlap are excluded. Predicted values of NMR observable parameters, especially NOE intensities, converge at sampling sizes of 10^5 conformers.

NOE intensities were calculated using the two-spin approximation[47] by population-weighted averaging of the inverse sixth power of the respective proton−proton distance $\langle 1/d^6 \rangle$ over an ensemble of 10^5 random coil structures generated by the Monte Carlo sampling procedure. NOE intensities for all combinations of H^N, H_α, and H_β protons of $(i, i+1)$, $(i, i+2)$, and $(i, i+3)$ residue pairs were thus obtained. To display the results graphically, standard NOE maps with intensity cutoffs equivalent to proton−proton distances of 2.5, 3.5, and 4.3 Å for strong, medium, and weak NOE's, respectively, were used.[48] For NOE's that involve glycine and residues having methyl groups, a 0.4 Å larger cutoff was used for the Hα protons (for glycine) or Hβ protons (for the other residues) to account for the increased NOE intensity arising from multiple protons, the neglect of the longer side chains of valine, isoleucine, and theronine in the simplified model, and the specific relaxation properties of methyl groups.[49]

(d) Calculation of Cooperative NOE Intensities. To model the effects of cooperativity, the random coil sampling procedure was modified to incorporate the cooperativity inherent in the ϕ,ψ distributions of tripeptides found in the protein data base. Statistics on tripeptide segments from the PDB (data not shown) indicate that 27% are in the "all-α" (αααα) conformation, 18% are in the "all-β" ($\beta\beta\beta\beta$) conformation, and the rest are in mixed conformations (α$\beta\beta\beta$, αβαα, etc.). α and β regions are defined in Figure 1A,B. In the modified Monte Carlo sampling procedure we introduce a cooperativity factor c^α. For the random noncooperative case ($c^\alpha = 0$) the population of all-α conformers is equal to the product of the α propensities (P_i^α) of the individual residues ($i = 1, 2, 3$),

$$P_{\min}^{\alpha\alpha\alpha} \equiv P^{\alpha\alpha\alpha}(c^\alpha = 0) = P_1^\alpha P_2^\alpha P_3^\alpha \tag{2}$$

The most cooperative case is defined as $c^\alpha = 1$. Here the likelihood of the all-α conformation is equal to the minimum of the individual α probabilities P_i^α

$$P_{\max}^{\alpha\alpha\alpha} \equiv P^{\alpha\alpha\alpha}(c^\alpha = 1) = \min(P_1^\alpha, P_2^\alpha, P_3^\alpha) \tag{3}$$

This is to ensure that the intrinsic α population of each individual residue is not exceeded. For example, in a tripeptide with 30, 40, and 50% α-propensity for residues 1, 2, and 3, respectively, the highest possible all-α population $P_{\max}^{\alpha\alpha\alpha}$ at cooperativity $c^\alpha = 1$ is 30%. The minimum all-α population $P_{\min}^{\alpha\alpha\alpha}$ (at $c^\alpha = 0$), however, is 6%. The all-α population is defined as a function of the cooperativity factor c^α as

$$P^{\alpha\alpha\alpha}(c^\alpha) = P_{\min}^{\alpha\alpha\alpha} + c^\alpha(P_{\max}^{\alpha\alpha\alpha} - P_{\min}^{\alpha\alpha\alpha}) \tag{4}$$

From the above equations we can derive a cooperativity weight $w^{\alpha\alpha\alpha}$ (≥ 1.0) which is used to weight each all-α conformation generated by the Monte Carlo procedure,

$$w^{\alpha\alpha\alpha} = 1 + c^\alpha(P_{\max}^{\alpha\alpha\alpha}/P_{\min}^{\alpha\alpha\alpha} - 1) \tag{5}$$

When introducing cooperativity weighting factors care must be taken not to perturb the intrinsic α populations P_i^α of each residue i. Removal of this bias is facilitated by an additional weighting factor w_i^α (≤ 1.0) for each amino acid i to rescale its overall α population P_i^α. For a tripeptide consisting of three different amino acids, the following equation must hold:

$$P_i^\alpha = (P_i^\alpha - \sum_{j\neq i} P_j^\alpha P_i^\alpha - P_{\min}^{\alpha\alpha\alpha})w_i^\alpha + (P_i^\alpha \sum_{j\neq i} P_j^\alpha w_j^\alpha)w_i^\alpha + P^{\alpha\alpha\alpha}(c^\alpha); \quad i = 1, 2, 3 \tag{6}$$

This equation ensures that the sum of the different types of populations in which residue i is in the α state is equal to its total α population P_i^α. The first term in the above equation is the rescaled population of tripeptides with residue i in the α Ramachandran region; the second term corresponds to conformations where two residues (i and $j \neq i$) are in the α state, and the final term accounts for the cooperative c^α dependent all-α population. With the above equation the weighting factors can be defined iteratively as

$$w_i^\alpha = \frac{P_i^\alpha - P^{\alpha\alpha\alpha}(c^\alpha)}{P_i^\alpha - P_i^\alpha \sum_{j\neq i}(1 - \omega_j^\alpha)P_j^\alpha - P_{\min}^{\alpha\alpha\alpha}}; \quad i=1,2,3 \tag{7}$$

Noncooperative conformations (those that are not all-α) are weighted by the product of the individual residues weighting factors w_i^α given that the ϕ,ψ angles of the residue are in the α region of the Ramachandran map. Weighting factors for residues that are not in the α region are unity. Although only α cooperativity is discussed in the present paper, equations similar to the ones above have been also derived for β cooperativity.

Results and Discussion

(a) Characteristics of the NMR Parameters of a Random Coil. Using the proposed model for the random coil state described in the previous sections we can predict the NMR parameters expected for a maximally disordered peptide or denatured protein with a given amino acid sequence. Figure 3 shows $^3J_{HN\alpha}$ values, NOE ratios, and NOE pattern predicted for three segments of lysozyme (residues 19−24, 50−55, and 95−119); these have been chosen as illustrations as they contain very different structures (turn, β-sheet, and α-helix, respectively) in the native state of lysozyme, but are also regions with reduced spectral overlap and complete and unambiguous spectral assignments. The full set of lysozyme data will be discussed in a subsequent publication.[50] It is immediately apparent that parameters such as $^3J_{HN\alpha}$ coupling constants and the αH−NH/NH−NH intensity ratios for NOE's between sequential $(i, i+1)$ residues are predicted to be far from uniform along the sequence. The observed differences directly reflect the variations in ϕ and ϕ,ψ torsion angle preferences of the different amino acid types. For example, $^3J_{HN\alpha}$ coupling constants have the smallest values for residues with high α propensities (Figure 3A), particularly alanine (5.8 Hz), or a high population of positive ϕ angles as is seen for glycine (5.8, 6.2 Hz), and largest values for residues which favor β-conformations (7.2−7.5 Hz for threonine, valine, and isoleucine). These differences are significant compared to an uncertainty of less than 0.5 Hz in the experimental measurement.

Figure 3C shows that the random coil model predicts sequential $(i, i+1)$ NOE's for all residues. The intensity ratio between $(i, i+1)$ αH−NH and NH−NH NOE's is often used as a measure for secondary structure since it depends on both ϕ and ψ dihedrals. For the lysozyme sequence these random coil NOE intensity ratios are predicted to vary between 2.7 for residues with higher than average β-propensity and approximately 1.0 for residues with increased α-propensity (see Figure 3B). Glycines show the lowest αH−NH/NH−NH intensity ratios (0.7). Predicted ratios vary by a factor of 3, which is

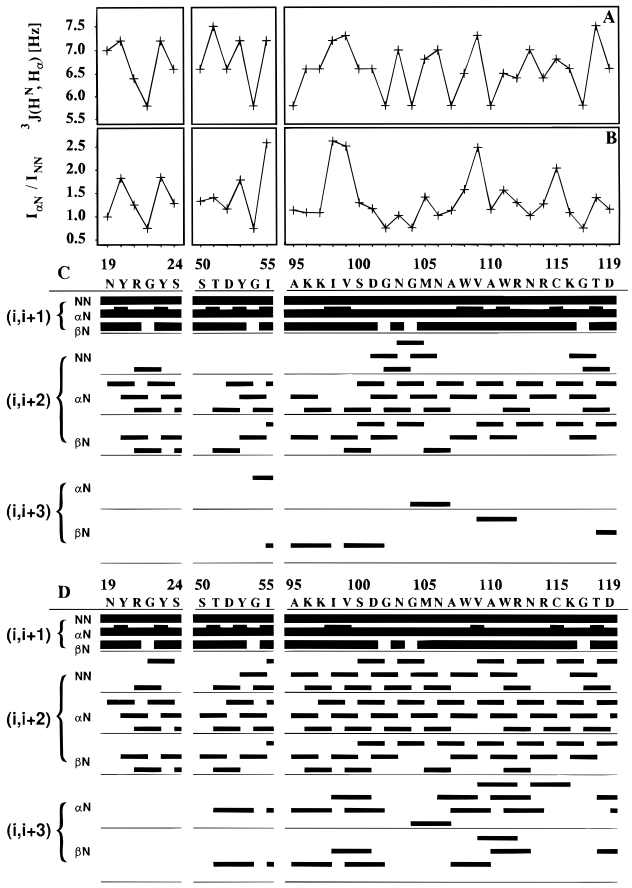

Figure 3. Predictions of NMR parameters for residues 19−24, 50−55, and 95−119 of hen egg-white lysozyme. (A) $^3J_{HNH\alpha}$ coupling constants [Hz]. (B) Ratios of sequential (i, $i + 1$)αN/NN cross peak intensities. (C) NOE map for the random coil model. Each bar represents a predicted NOE between two protons. NOE's are classified as strong, medium, and weak as indicated by the height of the bar. Weak NOE cutoffs correspond to a distance of 4.3 Å for all proton−proton distances excluding NOE's involving methyl groups (cutoff 4.7 Å) and are determined from the intraresidue reference NOE between glycine H_α and NH for 10 (of 12) nonoverlapping cross peaks in the experimental NOESY-HSQC spectrum. (D) NOE map for the cooperative model at a cooperativity level of $c^\alpha = 0.2$.

small compared to the overall range of intensity ratios of up to a factor of 10^3 observed in NOESY spectra of native state proteins.

In rigid molecules, NOE's may be observed between protons that are up to 5 Å apart, depending on the overall correlation time of the molecule and the signal-to-noise level of the data. In the case of multiple conformations which interconvert rapidly, two protons may show an NOE even for a sparsely populated conformation if the two protons are sufficiently close in space. For example, conformers with interproton distances of 2.5, 3.5, or 4.5 Å will have similar detectable NOE intensities if they are populated 2, 12, or 53% of the time, respectively. Figure 3C demonstrates that a number of medium range NOE's are expected to be observed even for a highly disordered state and that the pattern of predicted NOE's is nonuniform. For the three fragments considered here 59 (i, $i + 2$) and 7 (i, $i + 3$) NOE's are predicted. The αH−NH(i, $i + 2$) NOE's are the most frequently predicted medium range NOE's. These NOE's would not be observed for a β-strand (interproton distance > 5.5 Å) and would be very weak for a regular α-helix (interproton distance 4.3 Å). However, in a mixed $\beta\alpha$-conformation the αH−NH(i, $i + 2$) distance can be as short as 3.7 Å which therefore contributes strongly to the d^{-6} averaged intensity since $\beta\alpha$-conformers are populated approximately 25% of the time. Seven NH−NH(i, $i + 2$) NOE's are predicted; these distances

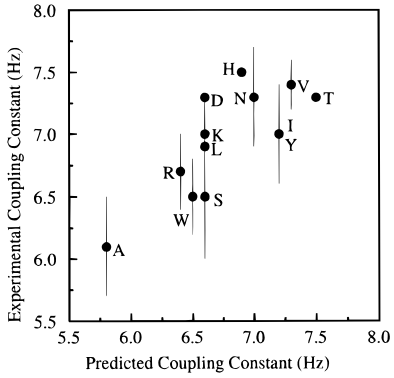

Figure 4. A comparison of residue specific mean experimental $^3J_{HNH\alpha}$ coupling constants with those predicted from the random coil model. The experimental values are derived from three peptide fragments of hen lysozyme (residues 13−33, 86−102, and 105−115) each of which appears, from the NOE and chemical shift data, to be largely disordered. The coupling constants were extracted from cross peaks in DQF-COSY spectra using a fitting procedure.[60] The vertical lines show ±1 standard deviation. The correlation coefficient for the data is $R = 0.76$.

are short for segments with high $\beta\alpha_L$ propensities and are thus predicted for glycines and asparagines in the $i + 1$ position. Two αH−NH(i, $i + 3$) NOE's in the segments 54−57 and 104−107 are predicted; the αH−NH(i, $i + 3$) distance is short for peptides in ααα or $\beta\alpha\alpha$-conformations (3.4 and 4.2 Å, respectively). Similarly the five predicted βH−NH(i, $i + 3$) NOE's should be observed in ααα and $\beta\alpha\alpha$ conformations (3.6 and 3.2 Å, respectively).

(b) Comparison of the Cooperative and Noncooperative Models. Although in many cases a random coil model with only local interactions seems to provide a reasonable description of the conformational populations adopted in disordered peptides and denatured proteins, the effects of cooperativity, particularly with regard to helix formation, have also been considered. For example, consider the two extremes where a peptide of three amino acids is either (A) fully random or (B) fully cooperative for a situation in which the individual α space populations (ϕ,ψ's in the α region of the Ramachandran map; see Figure 1, A and B) are given as 50% for each residue. The species where all residues are in the α conformation (ααα) is then populated by 12.5% of the molecules for the random case A and by 50% for the cooperative case B. These two extreme states would give identical values for NMR parameters which probe the conformation of a single residue such as $^3J_{HNH\alpha}$ coupling constants, but significantly different intensities for medium range NOE's such as αH−NH(i, $i + 3$). Figure 3D shows the predicted NOE's assuming a cooperativity of $c^\alpha = 0.2$ for the three segments. For the cooperative case significantly larger numbers of NH−NH(i, $i + 2$), αH−NH(i, $i + 3$), and βH−NH(i, $i + 3$) NOE's are predicted, whereas the number of predicted αH−NH(i, $i + 2$) and βH−NH(i, $i + 2$) NOE's increases only slightly. Thus, as anticipated, clusters of (i, $i + 3$) NOE's are predicted when α space is populated by three adjacent residues in a cooperative manner.

(c) Comparisons with NMR Data for Denatured Lysozyme. The predicted NMR parameters for a random coil state can be used as a framework for interpreting the experimental data obtained from denatured proteins and isolated peptides. This is demonstrated here by a comparison of the predictions for hen lysozyme with experimental NMR data for urea denatured lysozyme and disordered peptide fragments taken from the lysozyme sequence.

Figure 4 shows a comparison of predicted $^3J_{HNH\alpha}$ coupling constants with $^3J_{HNH\alpha}$ data derived from several peptide fragments of lysozyme.[39,51,52] The good correlation ($R = 0.76$) suggests that local conformational preferences of the peptide

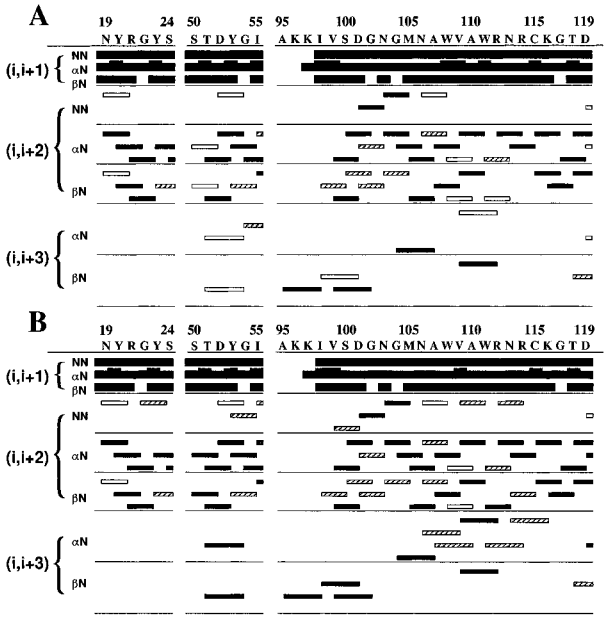

Figure 5. Comparison of predicted and experimental NOE maps for segments 19−24, 50−55, and 95−119 of lysozyme for the random coil (A) and the cooperative (B) models. In order to compare with the simulation, all experimental NOE's involving side chains are designated as β. Predicted and observed NOE's are shown in black bars, NOE's observed but not predicted are shown in open bars, and NOE's predicted but not observed are shown in dashed bars. Strong, medium, and weak NOE's are indicated by the height of the bar. Experimental NOE's were extracted from 3D-^{15}N-filtered NOESY-HSQC experiments using a 200 ms NOE mixing time performed on a 3 mM sample of lysozyme in 8 M urea at pH 2. Further experimental details will be published elsewhere.[50]

backbone in denatured states are similar to the ones found in native states. Moreover, Serrano et al.[40] have shown that the observed H_α chemical shifts for unstructured peptides correlate well with predicted chemical shifts, using similar database derived ϕ,ψ distributions and a parametrization for the H_α chemical shift deviation from shifts measured in short peptides as previously introduced by Wishart et al.[53] These two findings provide evidence which supports the use of PDB derived populations of ϕ,ψ angles to model denatured states.

A comparison of the random coil NOE predictions with NOE's identified in a ^{1}H-^{15}N NOESY-HSQC spectrum of lysozyme in 8M urea shows that for several regions of the sequence we see significant agreement between prediction and experiment. This is the case for residues 95−119, shown in Figure 5A, where four of the six experimentally observed (i, i + 3) NOE's are predicted by the model (black bars). This result suggests that for this region of the chain conformational preferences are determined predominantly by local interactions. For other parts of the sequence, such as residues 19−24 and 50−55, there is a significantly poorer agreement for the random coil model which may indicate residual nonrandom structure. In some cases the nature of such localized structure has been characterized. For example, for residues 19−24 (Figure 5A) an interaction between the NH of glycine 22 and the aromatic ring of tyrosine 20 is likely to be responsible for the observation of two (i, i + 2) NOE's which are not predicted by the model. Such a tyrosine−glycine interaction has been observed previously in denatured BPTI.[54] The existence of this interaction is further substantiated by chemical shift analysis of the H^N and H_α protons of residues 20−24, which show among the largest deviations from random coil chemical shifts for both the urea denatured state[50] at pH 2 and lysozyme peptide fragments.[52,55]

For residues 50−55 a much higher agreement between prediction and experiment is achieved with the cooperative model. In this region, four of the five NOE's which are not predicted by the random coil model (white bars in Figure 5A) are predicted correctly in Figure 5B (black bars). A model of localized cooperative interactions in this segment is found to be consistent with the experimental NOE data. In fact H^N and H_α chemical shifts for residues 50−55 of urea denatured lysozyme are significantly perturbed from random coil chemical shifts.[50] Threonine 51 has been previously identified as one of the most perturbed residue side chains of the lysozyme denatured state spectra due to interactions with tyrosine 53.[56] Additionally, experiments on alcohol-denatured lysozyme show that residues 50−55 adopt a helical conformation in 70% (v/v) trifluoroethanol (TFE).[57]

TABLE 1: *R* Values for NOE Predictions

	noncooperative			cooperative	
NOE type	R^a	R^b	$\langle R_{rand}\rangle^c$	R^a	R^b
NN(i, i + 2)	1.0	0.5	1.22 ± 0.16	3.1	1.67
αN(i, i + 2)	0.29	0.29	0.46 ± 0.07	0.33	0.21
βN(i, i + 2)	1.08	0.67	1.21 ± 0.12	0.92	0.67
αN(i, i + 3)	1.0	1.0	2.26 ± 0.51	3.25	1.0
βN(i, i + 3)	1.0	0.6	2.19 ± 0.47	0.75	0.2

[a] Full sequence. [b] Residues 19−24, 50−55, 95−119. [c] Average and standard deviation of 20 randomized lysozyme sequences.

(d) Comparison of Experimental NOE Data with Predictions for Random Sequences. To evaluate more rigorously the predictive power of the random coil model, we have predicted NOE patterns for randomized lysozyme sequences, generated by randomly permuting the order of the 129 amino acid residues. After calculating the NOE pattern maps for these sequences, the agreement between experimental NOE's and NOE's predicted for the random sequences was measured utilizing a residual index R defined as

$$R = (N^-_{exp} + N^-_{theory})/N_{exp} \quad (8)$$

In this equation N_{exp} is the number of observed NOE's, N^-_{exp} the number of observed NOE's which are not predicted, and N^-_{theory} the number of predicted NOE's which are not observed. We exclude predicted NOE's which cannot be compared to the experimental data due to spectral overlap. Table 1 shows R values for the predicted NOE map of the full wild type lysozyme sequence and its peptide segments for the noncooperative and cooperative models. Additionally, predictions were undertaken on 20 randomized lysozyme sequences. The average "random" R values $\langle R_{rand}\rangle$ of these predictions and their standard deviations are also listed in the table. It is found that in particular (i, i + 3) NOE predictions differ significantly from predictions for random sequences by more than 2 standard deviations.

Conclusions

Characterization of the conformations sampled in denatured states of proteins is of major importance in attempting to gain insight into the relative stabilities and structural biases that exist within local regions of polypeptide sequences. These are of considerable interest because of their likely significance in the early stages of protein folding. The random coil model described in this paper is based upon the use of specific ϕ, ψ distributions for amino acid types determined from analysis of structures in the protein data bank. The present method provides the means for characterizing the conformations in denatured proteins and individual peptides. It is known that the same proteins denatured in different ways can have different NMR parameters and other characteristics.[56,57] By analyzing these insight into the role of specific denaturants can be gained.

The good correlation between NMR coupling constants predicted from the model and experimental data for peptide

fragments of hen lysozyme provides a firm basis for the approach. The pattern of NOE's predicted for the sequence of lysozyme correlates remarkably well, given the simplicity of the model, with experimental data obtained from heteronuclear NMR studies of the intact protein denatured in 8 M urea at low pH. In certain regions, however, marked deviations between the predicted and the experimental data suggest that additional conformational preferences exist in the denatured state. Two cases of this have been examined in more detail here. In one of them the NMR data are indicative of specific side chain interactions between a glycine and a tyrosine residue, of a type found previously in denatured states of other proteins. In the other, a number of medium range NOE's that are observed experimentally but not predicted in the simple model, could be predicted very successfully by introducing a degree of cooperativity into the model. Improvement of NMR techniques and new pulse sequences tailored for the investigation of denatured states should enable more rigorous testing of the predictions and may lead to a framework for the detailed analysis of NMR data for a variety of nonnative states of proteins.

Acknowledgment. This is a contribution from the Oxford Centre for Molecular Sciences which is supported by the U.K. Engineering and Physical Sciences Research Council, the Biotechnology and Biological Sciences Research Council, the Medical Research Council. The research of C.M.D. is supported in part by an International Research Scholars award from the Howard Hughes Medical Institute. K.M.F. acknowledges support from the Wellcome Trust by a Hitchings-Elion Fellowship. H.S. is supported by the Human Capital and Mobility program of the European Community, and L.J.S. is a Royal Society University Research Fellow. We are grateful to Janet Thornton and Malcolm MacArthur for valuable discussions and their expertise in analysis of the protein data base. We also acknowledge stimulating discussions with Christian Griesinger.

References and Notes

(1) Dobson, C. M. *Curr. Opin. Struct. Biol.* **1992**, *2*, 6.
(2) Shortle, D. *Curr. Opin. Struct. Biol.* **1993**, *3*, 66.
(3) Wüthrich, K. *Curr. Opin. Struct. Biol.* **1994**, *4*, 93.
(4) Dill, K. A.; Shortle, D. *Annu. Rev. Biochem.* **1991**, *60*, 795.
(5) Dobson, C. M.; Hanley, C.; Radford, S. E.; Baum, J.; Evans, P. A. In *Conformations and Forces in Protein Folding*; Nall, B. T., Dill, K. A., Eds.; American Association for the Advancement of Science: Washington, DC, 1991; Chapter 12, p 175.
(6) Dobson, C. M.; Evans, P. A.; Williamson, K. L. *FEBS Lett.* **1984**, *168*, 331.
(7) Evans, P. A.; Kautz, R. A.; Fox, R. O.; Dobson, C. M. *Biochemistry* **1989**, *28*, 362.
(8) Robertson, A. D.; Baldwin, R. L. *Biochemistry* **1991**, *30*, 9907.
(9) Radford, S. E.; Buck, M.; Topping, K. D.; Dobson, C. M.; Evans, P. A. *Proteins: Struct., Funct., Genet.* **1992**, *14*, 237.
(10) Buck, M.; Radford, S. E.; Dobson, C. M. *J. Mol. Biol.* **1994**, *237*, 247.
(11) Dyson, H. J.; Merutka, G.; Waltho, J. P.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1992**, *226*, 795.
(12) Dyson, H. J.; Sayre, J. R.; Merutka, G.; Shin, H. C.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1992**, *226*, 819.
(13) Kemmink, J.; Creighton, T. E. *J. Mol. Biol.* **1993**, *234*, 861.
(14) Waltho, J. P.; Feher, V. A.; Merutka, G.; Dyson, H. J.; Wright, P. E. *Biochemistry* **1993**, *32*, 6337.
(15) Lumb, K. J.; Kim, P. S. *J. Mol. Biol.* **1994**, *236*, 412.
(16) Blanco, F. J.; Rivas, G.; Serrano, L. *Nature Struct. Biol.* **1994**, *1*, 584.
(17) Munoz, V.; Serrano, L.; Jimenez, M. A.; Rico, M. *J. Mol. Biol.* **1995**, *247*, 648.
(18) Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K. *Science* **1992**, *257*, 1559.
(19) Neri, D.; Wider, G.; Wüthrich, K. *FEBS Lett.* **1992**, *303*, 129.
(20) Stockman, B. J.; Euvrard, A.; Scahill, T. A. *J. Biomol. NMR* **1993**, *3*, 285.
(21) Alexandrescu, A. T.; Abeygunawardana, C.; Shortle, D. *Biochemistry* **1994**, *33*, 1063.
(22) Arcus, V. L.; Vuilleumier, S.; Freund, S. M. V.; Bycroft, M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9412.
(23) Logan, T. M.; Olejniczak, E. T.; Xu, R. X.; Fesik, S. W. *J. Biomol. NMR* **1993**, *3*, 225.
(24) Logan, T. M.; Theriault, Y.; Fesik, S. W. *J. Mol. Biol.* **1994**, *236*, 637.
(25) Zhang, O. W.; Kay, L. E.; Olivier, J. P.; Forman-Kay, J. D. *J. Biomol. NMR* **1994**, *4*, 845.
(26) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.
(27) Dobson, C. M.; Karplus, M. In *Methods in Enzymology*; Hirs, C. H. W., Timasheff, S. N., Eds.; Academic Press: London, 1986; Chapter 18, p 362.
(28) Buck, M.; Radford, S. E.; Dobson, C. M. *Biochemistry* **1993**, *32*, 669.
(29) Bundi, A.; Wüthrich, K. *Biopolymers* **1979**, *18*, 285.
(30) Merutka, G.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **1995**, *5*, 14.
(31) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 332.
(32) Gregoret, L. M.; Cohen, F. E. *J. Mol. Biol.* **1991**, *219*, 109.
(33) Sun, S. J.; Luo, N.; Ornstein, R. L.; Rein, R. *Biophys. J.* **1992**, *62*, 104.
(34) Sun, S. J. *Protein Sci.* **1993**, *2*, 762.
(35) Kolinski, A.; Skolnick, J. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 338.
(36) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(37) Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. In *Crystallographic Databases-Information Content, Software Systems, Scientific Applications*; Allen, F. H., Bergerhoff, G., Seivers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn, 1978.
(38) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nature Struct. Biol.* **1995**, *2*, 596.
(39) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494.
(40) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322.
(41) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95.
(42) Gibrat, J. F.; Robson, B.; Garnier, J. *Biochemistry* **1991**, *30*, 1578.
(43) The protein data bank file names of the 85 coordinate sets used are 256ba, 7aata, 1abk, 1ads, 1apme, 1arb, 1ayh, 3b5c, 1bbpa, 3bcl, 1bop, 4bp2, 1btc, 2ca2, 2cdv, 3cla, 1cmba, 2cmd, 1cox, 1cpca, 2cpp, 1crn, 1csc, 1csee, 1csei, 4dfra, 1dri, 3ebx, 1end, 4enl, 2er7e, 1ezm, 2fb4h, 4fgf, 1fiaa, 1fkf, 1fxd, 4fxn, 1gpbo, 1gdlo, 1gky, 1gpr, 3grs, 2had, 1hoe, 2hpr, 1hyp, 1ifc, 3il8, 1lmba, 2ltna, 2ltnb, 1ltsa, 1lz1, 2mcm, lovaa, 2ovo, 5p21, 9pap, 1paz, 1pgx, 2pia, 1pii, 1poc, 2por, 5pti, 2rhe, 1rnh, 9rnt, 7rsa, 5ruba, 3rubs, 4rxn, 2sara, 2sici, 1ten, 1tfg, 1thg, 5tima, 2trxa, 2tsca, 1ubq, 9wgaa, 2wrpr, and 1ycc.
(44) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11.
(45) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870.
(46) Pardi, A.; Billeter, M.; Wüthrich, K. *J. Mol. Biol.* **1984**, *180*, 741.
(47) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimension*; Oxford University Press: Oxford, U.K., 1987.
(48) Wüthrich, K.; Billeter, M.; Braun, W. *J. Mol. Biol.* **1984**, *180*, 715.
(49) Werbelow, L. G.; Marshall, A. *J. Magn. Reson.* **1973**, *11*, 299.
(50) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Grimshaw, S.; Dobson, C. M., manuscript in preparation.
(51) Bolin, K. A. D. Phil. thesis, University of Oxford, Oxford, 1994.
(52) Smith, D. K.; Smith, L. J.; Dobson, C. M., unpublished data.
(53) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647.
(54) Kemmink, J.; van Mierlo, C. P. M.; Scheek, R. M.; Creighton, T. E. *J. Mol. Biol.* **1993**, *230*, 312.
(55) Van den Berg, B.; Pitkeathly, M.; Dobson, C. M.; Radford, S. E., manuscript in preparation.
(56) Evans, P. A.; Topping, K. D.; Woolfson, D. N.; Dobson, C. M. *Proteins: Struct., Funct., Genet.* **1991**, *9*, 248.
(57) Buck, M.; Schwalbe, H.; Dobson, C. M. *Biochemistry* **1995**, *34*, 13219.
(58) Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. *Proteins: Struct., Funct., Genet.* **1992**, *12*, 345.
(59) Insight II, Biosym Technologies, San Diego, CA, 1993.
(60) Smith, L. J.; Sutcliffe, M. J.; Redfield, C.; Dobson, C. M. *Biochemistry* **1991**, *30*, 986.

JP952747V