J. Mol. Biol. (1997) **268**, 170–184

JMB

Characterization of Long-range Structure in the Denatured State of Staphylococcal Nuclease. II. Distance Restraints from Paramagnetic Relaxation and Calculation of an Ensemble of Structures

Joel R. Gillespie and David Shortle*

Department of Biological Chemistry, The Johns Hopkins University School of Medicine Baltimore, Maryland 21205, USA

**Corresponding author*

Structural analysis of Δ131Δ, a fragment model of the denatured state of staphylococcal nuclease, has been extended by obtaining long-range distance restraints between chain segments by paramagnetic relaxation enhancement. Fourteen unique PROXYL spin labels were introduced at sites that are solvent-exposed in the native state, and the resulting enhancements of T_2 for the amide protons were measured by NMR spectroscopy. When these data were combined with either measured or estimated correlation times τ_c, the r^{-6}-weighted, time and ensemble-averaged distance between the spin label and 30 to 60 amide protons could be calculated for each spin-labeled protein. On the basis of approximately 700 such loose distance restraints, ensembles of compatible structures were generated by a combined distance geometry/molecular dynamics approach. Because of the large uncertainty in the physical basis of these distance restraints, a number of calculations were carried out to establish the sensitivity of the calculated structures to systematic errors in these restraints. Overall, the structural features reflected in the paramagnetic relaxation data were robust; large variations in τ_c, in the bounds window of allowed distances, or in the number of restraint distances used had small effects on the general features common to all calculated structures. The global topology of this denatured form of staphylococcal nuclease, as described by an ensemble of conformations consistent with the data, is strikingly similar to that of the native state, the major difference being the segregation of two hydrophobic segments that form a beta hairpin in the native state. These findings suggest that the topology of a protein's fold is established in the denatured state in the absence of cooperative interactions involving tight packing or stable hydrogen bonding. Hydrophobic interactions alone may encode global topology.

Keywords: protein folding; folding intermediates; spin labeling; hydrophobic interactions

Introduction

In an effort to gain insights into the interactions and processes that drive the folding of proteins, a number of laboratories have undertaken structural studies of proteins in the denatured state; the ensemble of random and partially folded conformations formed when the native state breaks down (for reviews, see Dill & Shortle, 1991; Dobson, 1994; Shortle, 1996a). Initial results indicate that NMR spectroscopy can readily characterize the segments of a denatured polypeptide chain that retain some level of residual secondary structure, especially alpha helices and turns. However, for the parts of the chain that are dynamic and lack the high chemical shift dispersion associated with well-packed native-like structure, few long-range NOEs can be detected and assigned, making

Abbreviations used: NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; PRE, paramagnetic relaxation enhancement; RMSD, root-mean-square deviation.

0022–2836/97/160170–15 $25.00/0/mb970953

the identification of long-range structure very problematic.

Clearly the dipolar interaction between protons, with its restriction to distances under 5.5 Å and its high sensitivity to correlation times, is not an ideal probe for identifying transient interactions between mobile chain segments over long distances. The dipolar interaction between a free electron and a proton, however, may be a more suitable probe of the global structure of partially folded proteins. Since a free electron has a magnetic moment that is 657 times larger than that of a proton, the range of this interaction extends out as far as 20 to 25 Å. Consequently, the attachment of spin labels at specific sites on the surface of proteins and measurement of the resulting effects on nearby protons may provide a promising strategy for structural characterization of denatured proteins.

In the accompanying paper (Gillespie & Shortle, 1997), we describe the application of paramagnetic relaxation enhancement by spin label probes to the characterization of a Δ131Δ, a large fragment of staphylococcal nuclease. This fragment, which lacks several amino acid residues from the amino and carboxy termini, serves as a tractable model of the physiologic denatured state, in that it remains denatured at physiologic temperatures and pH in the absence of denaturants. Extensive characterization by NMR and CD spectroscopy has revealed residual structure that depends sensitively on the conditions of solution (Alexandrescu *et al.*, 1994; Alexandrescu & Shortle, 1994; Wang & Shortle, 1995, 1996).

To attach single spin labels to Δ131Δ, 14 unique cysteine mutations were introduced by mutagenesis at sites that were predicted to be on the surface of the denatured state, and these cysteine residues were alkylated with the spin label PROXYL-iodoacetamide. To quantify the distance between the free electron of the PROXYL group and amide protons in the protein, a series of ^{1}H-^{15}N correlation spectra were collected and analyzed before and after reduction of the nitroxide free radical to the hydroxylamine. From their variation along the chain as a function of distance from the site of the spin label, the R_2 enhancements revealed several distinctly non-random patterns (Gillespie & Shortle, 1997).

Here, we turn to the challenge of extracting structural insights about the denatured state of staphylococcal nuclease from this large set of data, which contains information on the distance separating 690 residue-residue pairs. While computational methods based on distance geometry and simulated annealing are used routinely by NMR spectroscopists to convert a set of distance restraints into a family of ''compatible'' structures, three features of this experimental system make such an analysis less than straightforward. (1) The object of study is not a single three-dimensional conformation but rather an ensemble of many conformations that are rapidly interconverting. Thus the goal is to describe an ensemble, not one unique structure. (2) Calculated distance restraints will be averaged over the time-scale probed by the method (10 to 100 ms). An average structure may not correspond well with any single conformation or set of conformations within the ensemble. (3) The distance dependence of the dipolar interaction between electron -proton pairs leads to an inverse r^{-6}-dependent weighting of this ensemble. Consequently, the distance restraints are certain to describe a more compact structure than the true average for the ensemble. In the face of these problematic issues, we proceed in a conservative fashion.

Results

Calculation of distance restraints from paramagnetic relaxation enhancement data

As described briefly in the Theory section of the accompanying paper, the Solomon-Bloembergen equation describes how a free electron enhances the T_2 relaxation of a proton at the end of a fixed vector in the same molecule, which is free to rotate isotropically in solution with a global tumbling time of τ_c:

$$\Delta(1/T_2) = \Delta R_2 = K/r^6\left((4\tau_c + 3\tau c/(1+\omega_H^2\tau_c^2)\right) \quad (1)$$

The theory has been expanded to deal with the more complex case of motion along the electron-proton vector, and it can be shown that when these motions are high-frequency, the net effect is a statistical averaging of the distance r, not a change in the apparent correlation time τ_c (Krugh, 1979).

From the measured values of ΔR_2 and either the measured or estimated value of τ_c, a distance restraint d_R is calculated by solving equation (1) for r. From the sixth-power relationship between r and ΔR_2, it is clear that shorter distances will contribute more extensively to the value of d_R than longer distances. Consequently, use of this equation is certain to give an underestimate of the linearly weighted, ensemble average distance. With the available data, there is no obvious strategy for reliably correcting for this bias.

As discussed in the previous section, measurements of τ_c are available for only some of the 690 vectors and even these values are subject to large errors. Given that the measured values of τ_c ranged from 1 to 8 ns, it was felt that the best approach was to use 4 ns, the average calculated value of τ_c for all vectors. Because the distance depends on the sixth-root of the correlation time, the error in d_R introduced by even large errors in τ_c is relatively small. For instance, a relative error of 50% in the correlation time for a given vector would produce an error of about 9% in the measured distance, or approximately ±1.5 Å for a d_R of 16 Å. Since these large uncertainties in the value of d_R cannot be reliably estimated, they are dealt with at the next

stage, in which an ensemble averaged structure is computed using the set of all d_R.

Structure calculations

In the first set of structure calculations, the set of d_R values for 12 of the labeled proteins were calculated with τ_c set to 4 ns. The two proteins D19C* and K28C* (the asterisk designates the PROXYL label) were omitted from consideration because other NMR evidence (Wang & Shortle, 1996) suggests that the segment from 13 to 44 is involved in a separate structure, a dynamic hydrophobic bundle, that is more or less autonomous from the rest of Δ131Δ. The distance geometry/simulated annealing approach of Nilges *et al.* (1988) was applied to a full atom representation of Δ131Δ using the XPLOR 3.8 software.

Each distance restraint was imposed as a harmonic square well potential centered on d_R with bounds of $\pm X$ Å, where X is a variable. Structural energy terms from steric repulsion, bond length, bond angles and dihedral angles were included. Since Coulombic and van der Waals attractive terms were not included in the potential function, the empirical distance restraints served as the only ''attractive'' force to drive the chain from random coil conformations to more structured ones. Because the data do not provide for constraints of the first 34 amino acid residues (10 to 43), these residues were not included in the chain model.

Instead of substituting the wild-type side-chain of all 12 mutant residues with the cysteine-PROXYL group to correctly model the position of the electron, d_R was defined as the distance from the most terminal atom of the wild-type side-chain to residue-specific H_N atoms. Thus the ε nitrogen atom of lysine (45, 64, 71, 78, 84, 97 and 116), C-4 of the phenolic ring of tyrosine (113), the ζ carbon atom of arginine (105), and the carboxy O of glutamate (52 and 57) and glutamine (123) were defined as the site of the free electron. While the number of rotatable bonds and the H_N to electron distance are not precisely modeled by this convention, this source of error is small in comparison to other assumptions discussed below. The technical details of the structure calculations are described in Methods.

There are many sources of uncertainty in the values of d_R. The three most signficant are: (1) d_R is derived from a heterogeneous ensemble of partially folded conformations undergoing dynamic inter-conversion, not a single unique structure. Because of the r^{-6} dependence of the paramagnetic enhancement of T_2 relaxation (ΔR_2), the conformations in the ensemble are not weighted in proportion to their frequency. Conformations with short values of r will contribute disproportionately to d_R. However, simulations of realistic motions involving both radial and angular displacements of the vector separating two dipoles suggest that biasing to shorter effective distances may be quite small or even negligible (LeMaster *et al.*, 1988). (2) Errors in the use of one correlation time τ_c for all electron proton vectors. (3) Errors in the accuracy with which ΔR_2 has been measured (Gillespie & Shortle, 1997).

Since there is no reliable way of estimating the uncertainty in each d_R, the approach taken was to assume that the errors from these and other sources can best be dealt with by adjusting the uncertainty of d_R through the distance bounds used to convert d_R into an energy term. For the distance range from $d_R -$ bounds to $d_R +$ bounds, the energy is constant, and no preference was given to any value of r within this range during the structure calculation.

Using the initial values of d_R, the effect of varying the distance bounds over a wide range on the internal consistency of the calculated structures was examined. In effect, ten structures were calculated with each of the following bounds: ±2, ±3, ±4, ±5, ±7, ±9, ±16, ±20, ±30, ±40, and ±50 Å, and the RMSD of the distances between the C^α atoms on comparing the mean structure and each of the ten individual structures was calculated as:

$$\mathrm{RMSD}(C^{\alpha}_{\mathrm{distances}}) = \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} 2(r_{ij} - \langle r_{ij} \rangle)^2 / N(N-1) \right)^{1/2}$$

where N is the number of residues, the sum is taken over all i and j, and $\langle r_{ij} \rangle$ is the average distance between the C^α atoms of residues i and j for ten structures calculated with $\tau_c = 4$ ns and bounds $= \pm 5$ Å. As shown in Figure 1A, there is a very shallow minimum at 5 Å, suggesting that the set of d_R do not tightly restrain the polypeptide chain and therefore rather large errors in d_R should not grossly alter the very large ensemble of structures consistent with the set of distance restraints. In other words, errors as large as ±4 Å to ±7 Å in each distance restraint can be tolerated without significantly altering the structural features defined by the full set of distance restraints. The bounds range chosen for subsequent calculations is ±5 Å.

A second indication that the chain remains well constrained when the bounds are increased comes from the average radius of gyration R_g for the ten structures calculated for each bound (Figure 1B). R_g remains between 14 and 15 Å when the bounds vary from ±4 Å to ±9 Å. Not until the bounds are expanded above 10 Å do the calculated conformations of Δ131Δ start to expand and become unconstrained.

To characterize the effects of errors in τ_c on the values of d_R used for structure calculation, values of d_R obtained by solving equation (1) for fixed values of τ_c are plotted in Figure 2A. d_R increases gradually as the rotational correlation time is varied from 1 ns to 10 ns, the most likely range of values in Δ131Δ. This slow increase in d_R is a consequence of the fact that paramagnetic broaden-

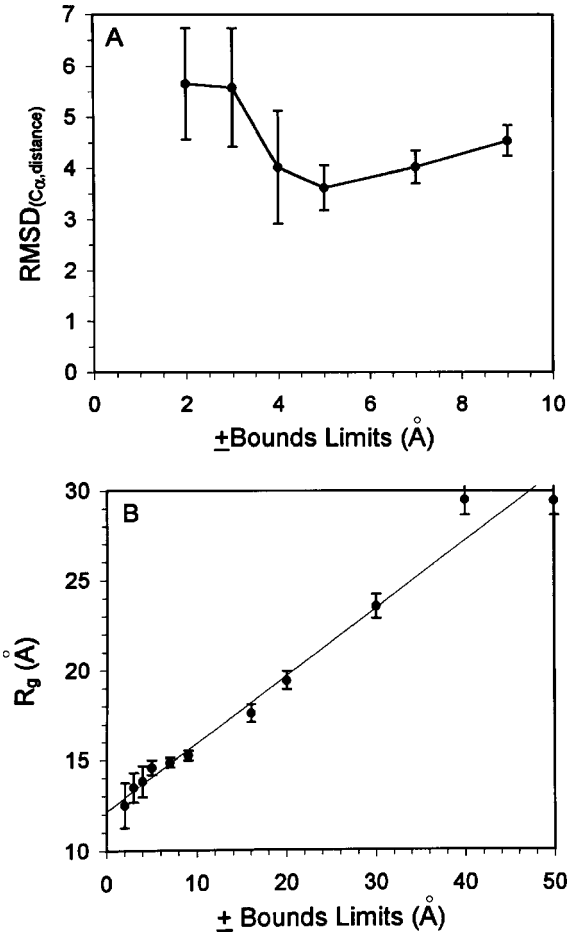

Figure 1. A, RMSD (C^{α} distances) as a function of distance bounds on d_R, in ±Å, (i.e. the allowed variation in d_R without an energy penalty). At each value of bounds, ten structures were calculated and their C^{α}–C^{α} distances were compared to the average C^{α}–C^{α} distances for ten structures calculated assuming $\tau_c = 4$ ns with ±5Å bounds. For bounds of ±5 Å and above, there was no restraint violation within a set of ten calculated structures. At shorter distances, the average number of restraint violations were: 1 ± 2 for the 4 Å bounds; 39 ± 9 for the 3 Å bounds; and 116 ± 7 for the 2 Å bounds. B, Average radius of gyration R_g for ten calculated structures as a function of distance bounds, in ±Å.

ing is more efficient for long correlation times. Therefore, in order to maintain a constant ΔR_2, the distance between the electron and proton must be increased, albeit in accord with its $(\tau_c)^{1/6}$ dependency.

The effects of systematic errors in τ_c on the structures calculated were approximated by using values of d_R obtained for τ_c from 1 ns to 8 ns to calculate ten structures, with the bounds set at ±5 Å. These ensembles of structures were then compared with respect to the average radius of gyration R_g and the RMSD of C^{α} distances from their average value when a τ_c of 4 ns was used. As can be seen in Figure 2B, the average R_g for sets of ten structures based on correlation times longer than 4 ns are not significantly increased, whereas for shorter correlation times, the set of calculated structures undergoes a modest compaction from 14.5 to 13.0 Å. When τ_c is varied from 2 ns to 8 ns, there is little or no change in the similarity among sets of calculated structures (Figure 2C). These results suggest that the principal structural features of Δ131Δ do not change dramatically with systematic variations in d_R resulting from errors in τ_c, again presumably because the set of d_R do not tightly restrain the polypeptide chain.

To make certain that the distance geometry/simulated annealing was eliminating all information contained in the starting template structure, a fully extended chain, one other conformation was used as a starting point. Using d_R based on $\tau_c = 4$ ns (the average value for all electron/proton vectors; Gillespie & Shortle, 1997) and ±5 Å bounds, ten structures were calculated beginning from the coordinates of native nuclease instead of a fully extended chain. Again, the RMSD($C^{\alpha}_{distances}$) between this set of structures and the average of ten structures initiated from a fully extended state were the same within standard error, as were the RMSDs among members within each of these two sets.

To determine the robustness of the calculated structure to the number of the distance restraints, two different cross-validation strategies were used. First, the d_R data for each one of the 12 labeled proteins were removed, leaving the restraints derived from the other 11 proteins for calculation of a family of ten structures. On comparing these structures to the average distance matrix calculated using all 12 restraints, only small differences in RMSD($C^{\alpha}_{distances}$) were found. The protein whose omission produced the largest change was K78C*, suggesting that this particular label contributes the largest number of significant constraints. Most significantly, omission of the distance restraints derived from labeled protein R105C* (Gillespie & Shortle, 1977), the one mutant most likely to have significantly perturbed the structure of Δ131Δ, did not significantly change the ensemble of calculated structures.

In the second cross-validation, sets of 50 randomly chosen d_R were successively deleted from the collection of 690 restraints. Figure 3 shows the RMSD($C^{\alpha}_{distances}$) as a function of number of d_R used in the calculation, again with all comparisons made to the average structure obtained with all restraints using $\tau_c = 4$ ns and ±5 Å bounds. As can be seen for three independent runs, the increase in RMSD($C^{\alpha}_{distances}$) is very gradual initially and then rises steeply when fewer than 100 restraints are included.

Approximately 200 proton-proton NOEs have been assigned in Δ131Δ and, of these, 42 are between residues pairs with spacing between $i,i+2$ and $i,i+6$ (Alexandrescu *et al*, 1994; Zhang *et al*., 1997). The single NOE between residues separated by six involved Thr82 to Gly88. Several additional NOEs between residues separated by five and four

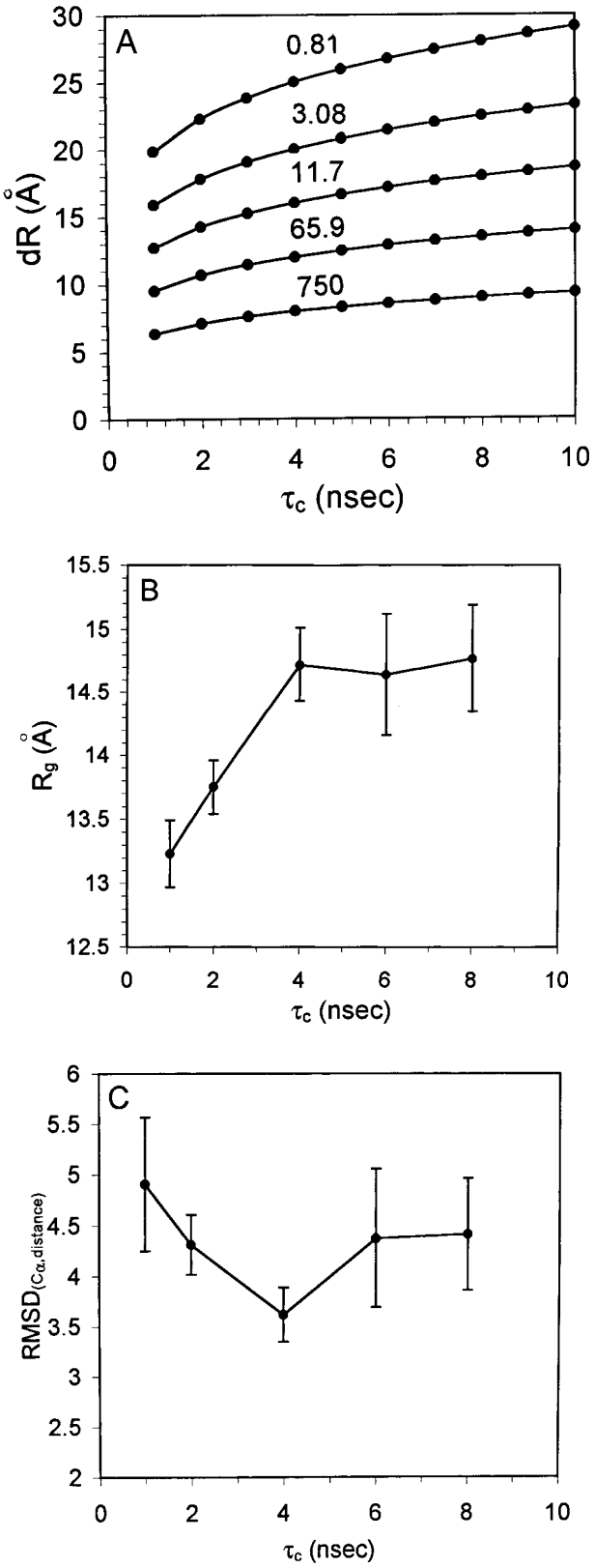

Figure 2. A, A plot of the relationship between a distance restraint d_R derived from a measured value of ΔR_2 (indicated above each curve) and the rotational correlation time τ_c (ns). B, Average radius of gyration R_g, ± the standard deviation, for ten structures of Δ131Δ as a function of correlation time τ_c used to calculate d_R. C, RMSD (C^α distances) of ten structures of Δ131Δ calculated on the basis of different correlation times when compared to the average C^α–C^α distance of ten structures with $\tau_c = 4$ ns. Bounds of ±5 Å were used for all calculations. The filled circle represents the mean value of the RMSD and the error bars represent the standard deviation.

residues occur in the loop defined by residues 82 to 88, whereas none further than three residues apart has been assigned outside this short sequence. These medium-range NOEs define two highly populated tight turns (τ83-86, τ94-97) plus transient helical structure in two helices (α1, residues 56 to 68; α2, residues 98 to 106).

When these 42 NOEs are included as restraints along with the set of d_R from paramagnetic relaxation enhancement, the calculated structures now acquire these turns and broken helices but overall there are no major changes in the C^α–C^α distance map (Figure 4A *versus* B). The RMSD($C^\alpha_{distances}$) and the radius of gyration (for residues 44 to 140) are essentially the same with or without these short and medium-range NOEs. The RMSD($C^\alpha_{distances}$) of ten structures calculated with the NOE restraints was 3.6(±0.3) Å (with respect to the set mean) with a mean radius of gyration of 14.8(±0.3) Å, while that calculated without NOEs was 3.6(±0.4) Å with mean radius of gyration of 14.7(±0.3) Å. Calculations using only NOE restraints (no PRE data included) gave an RMSD($C^\alpha_{distances}$) of 47.9(±1.5) Å with a radius of gyration of 46.8(±1.0) Å. This last observation indicates that the structural collapse of the chain is solely due to the long-range paramagnetic relaxation enhancement data and that short-range restraints in the form of NOEs make minor adjustments to atomic coordinates within a larger framework defined by the paramagnetic relaxation enhancements.

When the C^α–C^α distance maps were averaged over four or 16 independently calculated conformations (Figure 4 C and D), structural details that are ''random noise'' (i.e. unique to one or a small fraction of conformations) are lost in the process of averaging, whereas details common to the majority of conformations are retained. This allows the structure encoded by the distance restraints to be unequivocally defined. As can be seen by comparing Figure 4 C and D, long-range structure in the second half of Δ131Δ involves pairs of segments running both antiparallel (patterns perpendicular to the diagonal) and parallel (patterns parallel with the diagonal) with each other, plus interactions between residues separated by as many as 40 intervening residues (e.g. residue 100 to residue 140). These features are most easily appreciated in three-dimensional space, as described below.

An attempt was made to define the relative position of the chain segment 10-43 using the paramagnetic data from labeled proteins D19C* and K28C*. Inspection of the values of ΔR_2 in Figure 7

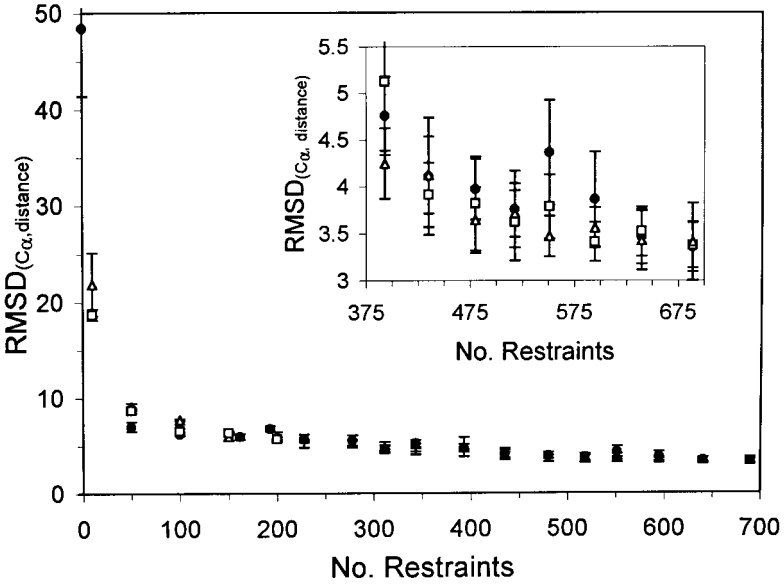

Figure 3. A plot of the dependence of RMSD (C^{α} distances) on the number of restraints d_R used to calculate the structure of Δ131Δ. In three independent runs (different symbols) beginning with all 690 restraints, 50 restraints were removed at random and ten structures calculated using the remaining restraints. This was repeated until no or only a few restraints were left. Comparison is made between each set of ten structures with a fixed number of d_R restraints to the average C^{α}–C^{α} distances (τ_c, 4 ns; ±5 Å bounds). The insert shows the variation in RMSD as the number of restraints is reduced by half.

of the accompanying paper reveal relatively uniform broadening occuring across Δ131Δ. These values of ΔR_2 were converted to distance restraints and an attempt was made to ''dock'' the D19C* and K28C* labels with the calculated structures of Δ131Δ (residues 44 to 140) by carrying out a grid search of the volume surrounding Δ131Δ. In neither case could a reasonable energy minimum with few restraint violations be identified after searching a box 40 Å by 40 Å by 40 Å centered on Δ131Δ in 5 Å steps, with positions on opposite sides of Δ131Δ often being nearly equivalent in energy. This observation suggests that the amino-terminal quarter of Δ131Δ is largely disjointed from the remainder of the protein and may act as an independently folding sub-domain (Wang & Shortle, 1996).

Description of the ensemble of Δ131Δ structures

Δ131Δ is an ensemble of many dynamically interconverting conformations. Because its structure is calculated from a large number of imprecise, ''average'' restraints to backbone protons, we have concluded that the best representation of these structures in Euclidean space is not a single aver-

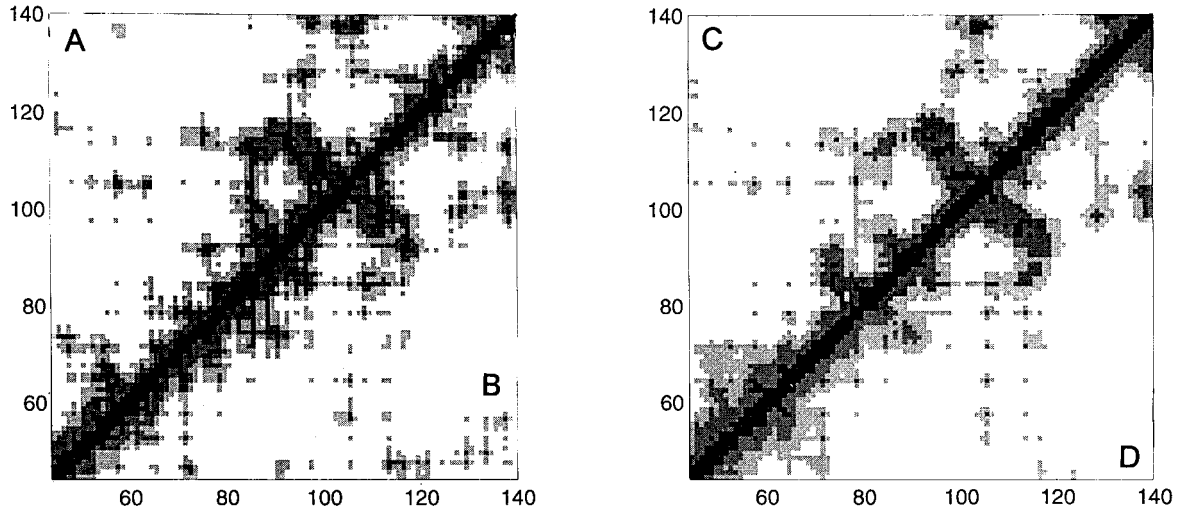

Figure 4. Four C^{α} distance matrices of calculated structures of Δ131Δ. Distances between C^{α} atoms are shaded in five categories: 0 to 5 Å, black; 5 to 10 Å, dark gray; 10 to 15 Å, medium gray; 15 to 20 Å, light gray; >20 Å, white. A, One structure, where $\tau_c = 4$ ns, ±5 Å bounds, no NOE. B, One structure as in A but with 42 medium-range NOE restraints included. C, Average distance matrix of four structures calculated as in B. D, Average distance matrix of 16 structures calculated as in B.

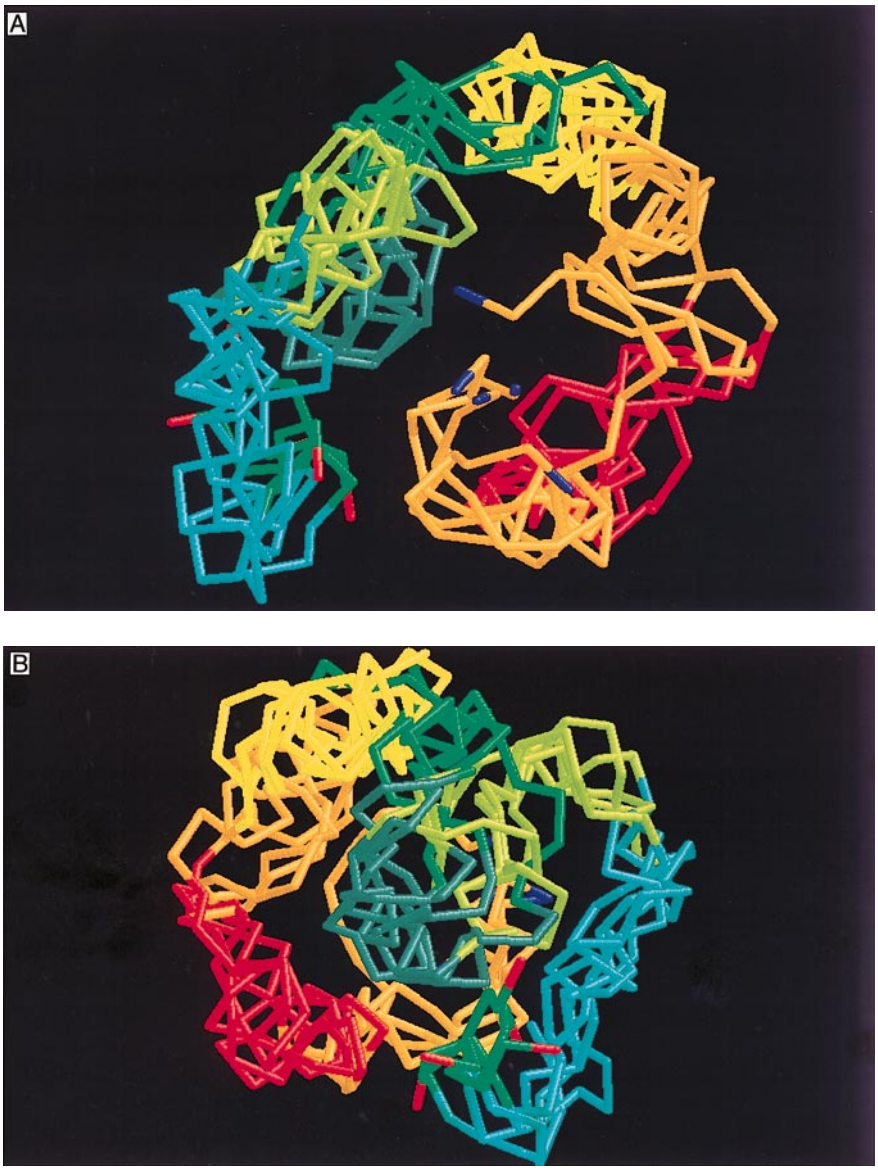

Figure 5. Ensemble of five virtual C^{α} traces corresponding to Δ131Δ structures calculated with all paramagnetic relaxation enhancement restraints (τ_c, 4 ns; ±5 Å bounds) plus 42 NOE restraints. Residue 44, dark blue; segment 45-55, gold/orange; segment 56-67, red; segment 68-77, gold/orange; segment 78-89, bright yellow; segment 90-98, dark green; segment 99-108, slate blue; segment 109-121, yellow-green; segment 122-136, light gray-blue; segment 137-139, dark green; residue 140, red. Drawn with the RasMol software (Sayle, 1995). A, Perspective from above, demonstrating the overall U-shape, with the chain termini at the opening of the U and the hydrophilic 78-89 segment forming the bottom. B, Perspective from below and slightly behind the bottom of the U-shape.

age structure, but rather a set of superimposed backbone traces. Five structures displayed as virtual C^{α} backbone traces, in our opinion, fairly convey the heterogeneity in allowed backbone positions while permitting each individual chain to be traced for most of its course. The five structures shown in Figures 5 and 6 were picked at random from a large set calculated using a τ_c of 4 ns plus a bounds window of ±5 Å and incorporating the 42 medium-range NOEs.

At low resolution, the structure of Δ131Δ has a rough U shape (Figure 5A), with a cleft dividing two arms or sub-domains of unequal length. The short arm, which is colored in orange and red, is formed from amino-terminal residues 44 through 77. The long arm, colored in shades of green, consists of the carboxy-terminal residues 90 through 140, and the bottom or bend of the U is formed by a hydrophilic segment, residues 78 through 89. The short arm or A sub-domain is simple in shape, forming a rough cylinder curved upward at its amino terminus (Figure 6A). The carboxy terminal or C sub-domain, however, is both larger and more complex in shape. As can be seen in

Figure 6C, the chain traces the pattern of an N or sideways S that is approximately perpendicular to the plane of the U.

Since there are reasons to anticipate that many features of the denatured state will be native-like, at least transiently, we describe the details of the structure of Δ131Δ in terms of the spatial arrangements of chain segments that correspond to major local structures in the native state; alpha helices, beta strands and loops/turns.

The A sub-domain (residues 44 to 77)

As mentioned above, the chain segment corresponding to residues 10 to 43 is not seen by NMR, so the description of the structures that satisfy the paramagnetic relaxation distance restraints begins with residue 44. The hydrophilic segment from 44 to 55 (gold/orange in Figure 6A) forms an extended loop in the native state that plays an important role in substrate binding and catalysis. In Δ131Δ, this chain segment does not participate in any-long range interaction and is the least constrained component of Δ131Δ. As can be seen in Figure 5A, the first few residues are placed above the plane of the U. Since the missing β1-β2-β3-h segment is attached to residue 44, this placement of residues 44 to 48 suggests that this unseen segment may lie above Δ131Δ as viewed in Figure 5A.

The next segment, 56-67 (red, Figure 6A), forms the first alpha helix (α1) in native nuclease. The five very weak i, $i+3$ NOEs identified in this seg-

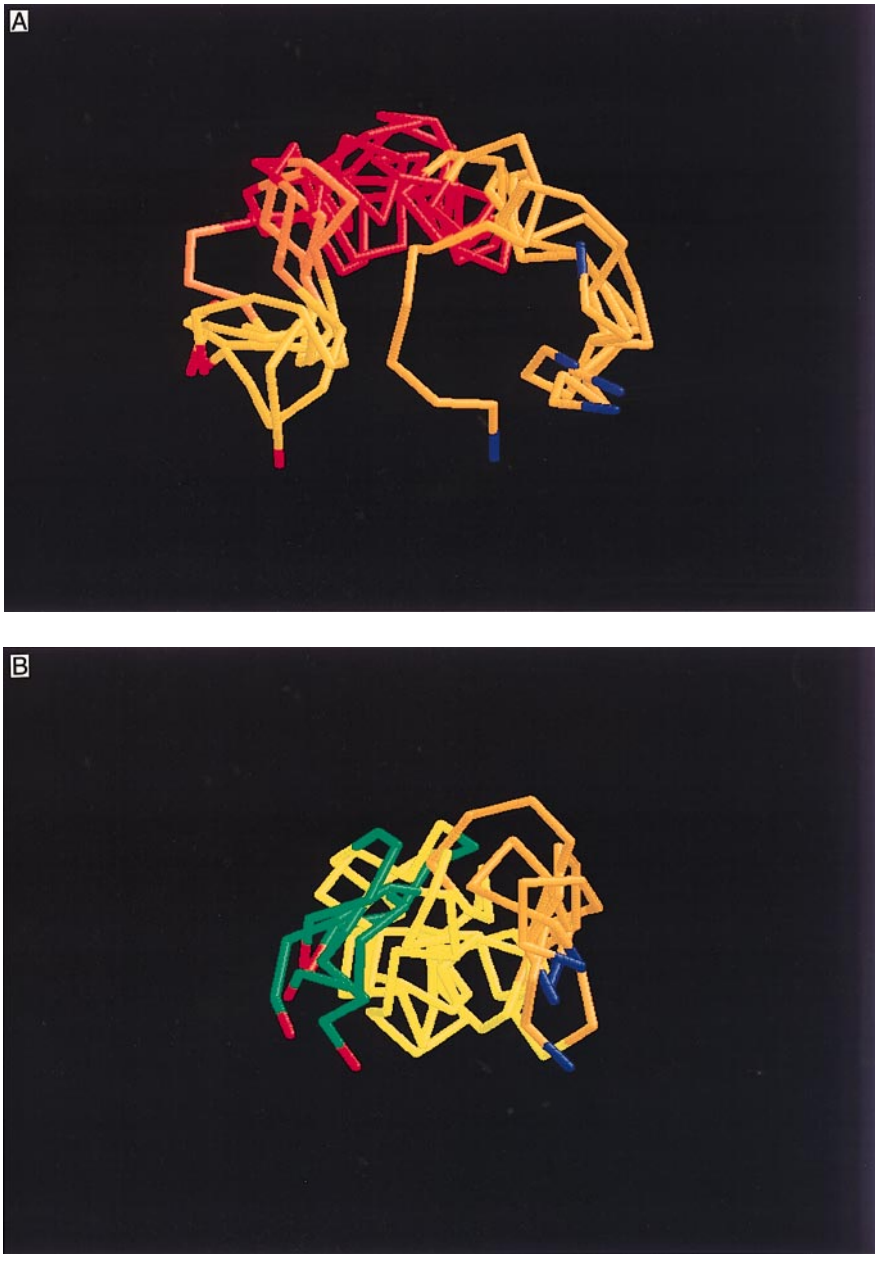

Figure 6(a,b) (*legend on page* 178)

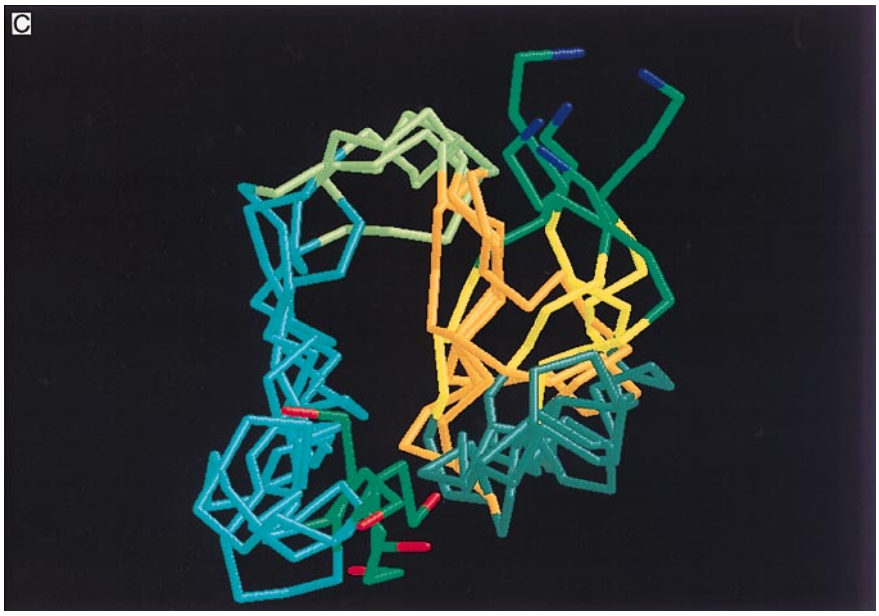

Figure 6. Ensemble of five virtual C^{α} traces of the same Δ131Δ structures as shown in Figure 5, but broken into three segments and with slightly different coloring schemes. Drawn with the RasMol software (Sayle, 1995). A, Residues 44 to 77. Residue 44, dark blue; segment 45-55, gold/orange; segment 56-67 (α1), red; segment 68-70, salmon/pink; segment 71-76 (β4), yellow/gold; residue 77, bright red. B, Residues 71 to 94. Residue 71, dark blue; segment 72-77 (β4), gold/orange; segment 78-89 (hydrophilic loop), bright yellow; segment 90-93 (β5), green; residue 94, bright red. C, Residues 90 to 140. Residue 90, dark blue; segment 91-94 (β5), green; segment 95-98, yellow; segment 99-108 (α2), dark gray; segment 109-116 (β6*), gold/orange; segment 117-121, pale yellow-green; segment 123-136 (α3), light gray-blue; segment 137-139, green; residue 140, bright red.

ment have been incorporated into the list of distance restraints used to calculate the structures shown in Figure 6A. Although this segment never forms a true alpha helix, it is relatively straight and contains a variety of loops and tight turns that are confined within a cylinder approximately 15 Å in length and 10 Å in diameter.

In the native structure, the α1 segment is terminated by three residues, 68 to 70 (salmon/pink, Figure 6A), that form a 90° turn or bend separating β4 (residues 71 to 77, Figure 6A) from the alpha helix. Similarly, in Δ131Δ, the polypeptide chain undergoes a change in direction of approximately 90° over the course of these same three residues (salmon color) plus one or a few of their neighbors.

The final segment of the A sub-domain is residues 71 to 77 (yellow/gold, Figure 6A), which form a beta strand (β4) in the native state. While this segment runs roughly perpendicular to the preceding α1 segment, it is not consistently extended or beta-like. Instead, it displays, on average, a slight curvature with the concave face turned inward toward the rest of the A domain, and the convex face turned out towards solution. None of the NMR parameters (chemical shifts, short-range NOEs, coupling constants) for this segment is indicative of an extended structure (Alexandrescu *et al*., 1994).

The hydrophilic loop

The segment from residues 78 to 89 (bright yellow in Figure 6B) is largely hydrophilic and forms a long loop in the native structure with a type I beta turn positioned near its center (residues 83 to 86). In Δ131Δ, this tight turn is the most clearly defined native-like structure detected by NMR spectroscopy. Many medium-range NOEs are observed between residues 82 and 88, including three NOEs from T82 to G88. Since these three NOEs were used as distance restraints for calculating the structures in Figure 6B, this turn is present in all of the structures. However, instead of a long loop with τ83-86 defining the point of chain reversal, this hydrophilic segment forms a more compact, S-shaped arrangement, often with a second turn with close approach between Gly79 and Asp83.

As seen in Figure 5A, the detailed placement of this segment between the A and C sub-domains is not well defined; although five chain traces display the close approach of 82 and 88, they are otherwise quite random. However, this Figure does show the most striking difference in chain segment interactions between native nuclease and Δ131Δ: β4 (gold/orange, Figure 5) and β5 (dark green, Figure 5) do not interact closely with each other. Instead, they diverge in order to interact with other segments in the A and the B sub-domains, re-

spectively, and the average distance between the C^{α} atoms of their center residues, Val74 in β4 to Ile 92 in β5, is 13.5(±3.7) Å in Δ131Δ versus 4.7 Å in the native conformation.

The C sub-domain (residues 90 to 140)

The first segment after the hydrophilic loop, residues 90 to 94 (green in Figure 6C), forms the fifth beta strand, β5, in the native structure. Although not fully extended, this chain segment in Δ131Δ follows a consistently straighter path than the β4 segment, as it descends through the plane of the U (Figure 5B and 6C). The next segment, residues 95 to 98 (yellow. Figure 6C), which terminates β5 by forming a type I′ tight turn in the native state, shows no consistent chain reversal in Δ131Δ. Instead, this segment is relatively extended, more or less continuing in the direction of β5.

As seen in Figure 6C, the five traces of β5 do not superimpose well. While the amino termini at 90 are somewhat clustered in space, the chain traces diverge, leading to much poorer superpositioning of the carboxy terminus at 94 and at 98. The result, or perhaps the cause, of this divergence is the heterogeneity in the orientation and shape of the next segment, alpha helix 2 or α2, residues 99 to 106, with residues 107 and 108 forming a helix cap (dark gray). Although five $i,i+3$ NOEs seen in several NOESY experiments were included as restraints on this chain segment, the end result is an imperfect helix that lacks a consistent orientation with respect to the remainder of the structure.

The chain segment after α2, residues 109 to 116 (gold/orange, Fig. 6A.), forms a long extended strand in native nuclease. While it is not conventionally considered to be a beta strand because it lacks regular hydrogen bonding with other chain segments, we designate it here as β6* to draw attention to its highly extended conformation. In Δ131Δ, the β6* segment is also relatively extended. Because of the variability in the position of α2 and thus of its carboxy terminus, the first few residues of β6* are very poorly clustered. Yet, as seen in Figure 6C, these segments (colored gold) show remarkable convergence at their carboxy termini, K116, which lie at a point high above the plane of the U.

Following residue 116 is proline 117 and residues 118 to 121 (pale yellow-green, Figure 6C), which form a bend in the native structure before the start of the last, long alpha helix α3, residues 122 to 136 (light gray-blue, Figure 6C). Similarly, in Δ131Δ, the chain reverses its direction over the span of residues 117 to 121, and the α3 segment follows a course down through the plane of the U as a well-clustered bundle to terminate in the last segment of Δ131Δ, residues 137 to 140 (green, Fig 6C). In the native structure these four residues form a large expanded loop of hydrophobic residues at the end of α3, positioning the carboxy terminus (colored red) of the chain near the carboxy end of helix 2. Although relatively poorly constrained in Δ131Δ, these last few residues assume a loosely clustered set of positions near α2.

Discussion and Conclusions

Paramagnetic relaxation enhancement has, in the past, been applied as a probe of physical distance only in cases where the distance to be measured is presumed to be constant and the correlation time for the electron-proton vector is assumed to depend on isotropic rotation (Schmidt & Kuntz, 1984). The reasons for these two restrictions are theoretical; the equations relating the T_2 enhancement to the distance between the electron and a proton cannot be rigorously solved unless these two conditions are met.

Neither of these these conditions is met in the current application. The distances between chain segments in a denatured protein are not constant, and on a timescale of nanoseconds, rigid body rotation is unlikely to describe all of the motion of any two chain segments with respect to each other. Consequently, there is no prospect for converting the measured values of ΔR_2 into precise distance restraints.

However, it must be noted that, on a smaller scale, essentially the same situation holds for the interpretation of proton-proton NOEs in folded proteins. At physiological temperatures, the native conformation can be viewed as an ensemble of (very) similar conformations; thermal motion causes some variation in the proton-proton distance plus local, anisotropic motions that are faster than molecular tumbling. Nevertheless, NOEs have been proven to be a reliable source of structural information when used in a ''semi-quantitative'' way to restrain interproton distances between 2 and 5 Å Therefore, the same spirit of a semi-quantitative restraint is used in the structural interpretation of the paramagnetic relaxation enhancement data presented in the accompanying paper (Gillespie & Shortle, 1997). The distance restraint d_R calculated from ΔR_2 and τ_c is interpreted as an average distance weighted for different conformational features in proportion to their frequency in the ensemble, and biased toward shorter distances by the r^{-6} term. As a result, it is clear that calculated structures will combine structural features that may not necessarily be present simultaneously in any one conformation. In addition, calculated structures will undoubtedly be significantly more compact than the true ensemble mean conformation. These major limitations, plus the fact that a denatured protein is likely to be a very heterogeneous ensemble, must be kept in mind when interpreting the results of structural calculations. To this end, it makes more sense to interpret the calculated structures as an ensemble, not as individual or average structures.

The validity of this semi-quantitative approach is supported by the observation that the ensemble of

Figure 7. Superpositioning of residues 56 to 140 of folded, wild-type staphylococcal nuclease (wide C^{α} trace) with respect to an ensemble of five structures calculated for Δ131Δ (narrow C^{α} traces). The three alpha helices are shaded in red. The three beta strands are yellow (β4), yellow green (β5) and orange (β6*). Drawn with the RasMol software (Sayle, 1995).

structures consistent with the full set of d_R is quite robust to changes in the parameters that determine the precise values of d_R. Changes in τ_c from 2 ns to 8 ns have small effects on the calculated structures. Varying the bounds window from ±4 Å to ±7 Å does not eliminate any of the salient features of the calculated structures. As a consequence, neither the uncertainty in τ_c nor the error introduced by the simplified method of measuring ΔR_2 (Gillespie & Shortle, 1977) appear to significantly affect the qualitative features of the structures shown in Figures 5 and 6.

Similarly, the ensemble of structures is robust to removing different subsets of d_R from the calculation. The fact that as many as 75% of restraints can be randomly dropped with little net effect suggests that the structural features in Δ131Δ, poorly defined as they may be, are overdetermined by the data. Oshiro *et al.* (1991) demonstrated that approximately one restraint per degree of freedom is required to obtain the correct overall structure. Since there are relatively few constrained degrees of freedom in the calculated structures, perhaps it is not surprising that they are over-determined by the large number of imprecise restraints obtained by paramagnetic relaxation enhancement.

The ensemble of calculated structures show several interesting topological features: (1) the relatively tight clustering of chain segments within well-demarcated volumes, as seen by the minimal mixing of chain colors in Figures 5 and 6; (2) the approximate U-shape assumed by the backbone, which divides the chain into two subdomains. In space-filling models, an aqueous cleft extends down a variable distance between these sub-domains. (3) The presence of bends or turns at positions that correspond very closely to bends/turns seen in the native state (residues 68 to 70, 82 to 88, 95 to 98, 108 to 110, 117 to 122). (4) The local topology of the four chain segments β5, α2, β6* and α3, which form a sideways ''S'' arrangement, is generally similar to that seen in the native state.

Most significantly, the global topology of Δ131Δ is nearly identical with that of wild-type nuclease, as is shown in Figure 7 by optimally superimposing the folded state and an ensemble of five calculated Δ131Δ structures. (A similar conclusion about the topology of a different denatured fragment of staphylococcal nuclease was reached by Ermacora *et al.* (1996) on the basis of a smaller number of inferred distances.) The binary relationships between pairs of chain segments are essentially preserved, except for the striking divergence of the β4-β5 pair. In the native state, these hydrophobic segments pair to form an antiparallel beta ribbon, whereas in Δ131Δ they are separated by a considerable distance (which in some structures involves a solvent-filled cleft). It would appear that local hydrophobic contacts take precedence over the long-range interaction between these two segments. The structure of Δ131Δ is generally consistent with what would be expected from a hydrophobic collapse dominated by interactions between residues that are nearby in the sequence.

Two arguments suggest that this native-like topology cannot be attributed to a small fraction of Δ131Δ molecules that transiently fold to the native state. (1) Previous studies under these conditions

of temperature and pH suggest that less than 1% of molecules will be in the native state. The time-constant for refolding of wild-type nuclease has been measured under similar conditions to be several hundred milliseconds (Epstein *et al.*, 1971); this is a lower limit for Δ131Δ, with experimental studies suggesting that it is several orders of magnitude longer (Wang & Shortle, 1996). Since paramagnetic relaxation enhancement has been measured over the much shorter interval that proton magnetization is in the transverse plane (50 ms or less), the native and denatured states are in very slow exchange relative to this relaxation. Consequently, there is no mechanism by which folded species could contribute more to the measured relaxation enhancement than in proportion to their fraction of the total population, i.e. less than 1%. (2) The large distance separating the β4 and β5 segments is not consistent with such a hypothesis. This structural feature is amply supported by ΔR_2 data from at least three labeling sites: K71C* (at the beginning of β4), K78C* (just after the end of β4) and K97C* (just after the end of β5).

Before addressing the question of why Δ131Δ forms the structure shown in Figures 5 and 6, the point needs to be re-emphasized that these calculated conformations may not correspond to any one conformation within the ensemble. Because of the r^{-6} weighting of ΔR_2, they will be more compact than the true ensemble mean conformation, although the magnitude of this artifactual compaction may be relatively small (LeMaster *et al.*, 1988). And because a denatured protein is likely to be a very heterogeneous ensemble of conformations, some of the structural features may not be present simultaneously in any one conformation. While the calculated ensemble conveys an impression of being static with conformations more or less equally well superimposed throughout the molecule, ^{15}N relaxation data suggest that the dynamic behavior of the chain is far from uniform (Alexandrescu & Shortle, 1994). The residues in the segment from 90 to 115 have high order parameter ($S^2 = 0.4$ to 0.8), suggesting relatively small amplitude local motions. However, the segment from 120 to 140 is quite dynamic ($S^2 = 0.1$ to 0.2), suggesting behavior approximating that of a small, unstructured peptide. And the segment from 44 to 80 is intermediate between these two extremes ($S^2 = 0.1$ to 0.5). Our position is that the ensemble of calculated structures is best viewed as a kind of Platonic ideal toward which the chain transiently converges in its constantly frustrated attempts to reach the native state.

Even though a segment of 34 residues in Δ131Δ cannot be seen by the methods employed, it may be playing a central role in determining the structure that can be seen. Work in progress has shown that single amino acid substitutions within this ''missing'' segment of Δ131Δ (residues 10 to 43) will in some cases increase or decrease the ^{1}H NMR line widths of amide protons in strands β4 and β5 (Y. Wang & D. S., unpublished results). These observations hint at a dynamic interaction between these two components, suggesting something like a resonance between alternative arrangements. The presence of the central cleft in Δ131Δ, which appears not to be filled by side-chains, suggests that its U-shape may not be a consequence of interactions between the A and B subdomains. Perhaps the missing chain segment 10-43 serves to bridge this cleft, an arrangement of chain segments that would be fully consistent with the native-like topology of the ''visible part'' of Δ131Δ.

Amide hydrogen exchange kinetics by deuterium uptake (Wang & Shortle, 1996) and by magnetization transfer (Mori *et al.*, 1997) demonstrate the lack of significant protection of the amide protons (protection factors <3 to 4) in Δ131Δ. Any hydrogen bonding involving the peptide backbone thus appears not to be sufficiently stable to be present more than 75 to 80% of the time. As mentioned above, the small order parameters determined by ^{15}N relaxation suggest that, for many segments of the chain, there is little stable packing of residues on a time-scale of a nanosecond or longer. The interactions that stabilize a particular structural feature are likely to be local, not long-range. In view of these conclusions, it is difficult to make a case for either stable hydrogen bonding or tight packing with strong dispersion forces playing a significant role in shaping the structure of Δ131Δ.

With all of the data taken together, Δ131Δ appears to be a highly dynamic ensemble with little cooperativity in the local structures that form transiently. In simple terms, the interactions among its chain segments more closely approximate those of a liquid than a solid. In the absence of stable hydrogen bonding networks or strong van der Waals attraction among large numbers of well-packed, cooperatively interacting residues, the residual structure found in Δ131Δ is probably stabilized by hydrophobic interactions. If this conclusion is more or less correct, then it appears that hydrophobic interactions alone may encode many, perhaps most, of the topological features of the global fold of nuclease.

One extension of this line of reasoning is that the hydrophobic collapse so popular in the literature of folding kinetics (Miranker & Dobson, 1996) could result in a liquid state in which the polypeptide chain, shaped by hydrophobic interactions, is constrained to a volume and shape consistent with the global fold of the native state. The completion of folding in a second step could be viewed as analogous to a fusion reaction, in which the freezing of the liquid-like denatured state is driven by van der Waals interactions into the solid-like native state.

This simple model of protein folding as a phase transition between a hydrophobic liquid and a van der Waals solid has several attractive features. (1) It emphasizes the fact that hydrophobic interactions exist only in the denatured

state. Once the native state state is attained, the interaction between non-polar atoms is dominated by dispersion forces; interaction with water is precluded. (2) The model attributes a clearly defined role in folding to the denatured state and its residual structure (Shortle, 1996a). In this state, chain entropy has been greatly reduced and the protein chain, though dynamic, is constrained to lie within a volume and "shape" that makes the transition to the native state analogous to the freezing of an organic liquid (with both polar and non-polar functional groups). In addition to reducing the strength of interactions in the native state, destabilizing mutations could exert their effect by distorting the volume/shape of the liquid denatured state. (3) The model naturally accounts for some of the unexpected plasticity of the native state, such as its frequent accommodation of insertions of one or two residues or deletions with fairly large local changes in structure. (4) It emphazises the equilibrium physical chemistry underlying protein folding, not transition states and kinetic pathways (Shortle *et al.*, 1996).

If it is generally the case that the global topologies of proteins are established in the denatured state by hydrophobic interactions, then the prediction of protein structure from amino acid sequence might more profitably proceed in two independent steps. In the first, the ensemble of conformations formed in the denatured state could be modeled on the basis of a free energy target function for the hydrophobic interaction, such as burial of non-polar surface. Then in the second step, the search for a well-packed, optimally hydrogen bonded, low-energy native state could be confined to the much smaller volume of conformation space demarcated by the global topology.

Methods

Calculation of distances from paramagnetic relaxation rate enhancements

Distances between the N-O bond in the nitroxide spin-label and the backbone amide protons were calculated using equation (1) and the experimental measurements of ΔR_2 and τ_c (Gillespie & Shortle, 1997). Because of the large relative uncertainty in the values of τ_c calculated from the data, a constant value of 4 ns (the average value over all measured electron-proton vectors) was used in the principal restraint calculations unless otherwise noted.

Structure calculation

The structure of Δ131Δ was calculated using the hybrid distance geometry/simulated annealing algorithms of XPLOR-3.1 (Brunger, 1992) on a Silicon Graphics Power Challenge 10000 GR (eight processors). The distance geometry and simulated annealing parameters used were essentially those of Nilges *et al.* (1988) with a few minor changes in the simulated annealing temperatures and cooling steps. An all atom (including hydrogen) representation of residues 44 to 140 was used in all cases.

Structural calculations were performed in four stages: substructural embedding, template fitting with simulated annealing regularization, slow-cooling simulated annealing refinement, and restrained minimization. Substructure embedding and regularization were performed using the bounds matrix defined from the experimental and holonomic restraints. At all subsequent steps, the following potential function was used:

$$E(\text{total}) = E(\text{bonds}) + E(\text{improp.}) + E(\text{angles}) + E(\text{vdW, repulsive}) + E(d_R),$$

where each energy term on the right is multiplied by a coefficient that changes during the course of the simulated annealing (Nilges *et al.*, 1988). In other words, the distance restraints d_R were the only attractive force that could act to increase the structure of the polypeptide chain. For restraints established from both paramagnetic relaxation enhancement data and from NOESY data, $E(d_R)$ was modeled as a harmonic square well:

$$E(d_R) = 0, \quad \text{for } d_R - \text{bounds} < r < d_R + \text{bounds}$$

$$E(d_R) = K[(d_R - \text{bounds} - r]^2 \quad \text{for} \quad r < d_R - \text{bounds}$$

$$E(d_R) = K[r - (d_R + \text{bounds}]^2 \quad \text{for} \quad r > d_R + \text{bounds}$$

where r is the distance between atoms in the computer model and K is a constant (XPLOR 3.8).

All annealing protocols were performed with Berendsen temperature coupling to an external bath (Berendsen *et al.*, 1984) initially at 3000 K. Chirality was introduced prior to regularization; following regularization, 58 cycles of Verlet dynamics (each consisting of 34 steps of 5 fs) were performed with temperature coupling and a cooling rate of −50 K/cycle. Further slow-cooling refinement was carried out with another 54 cycles of Verlet dynamics (each consisting of 74 steps of 5 fs, with a −50 K/cycle cooling rate). Finally, the annealed structures were minimized by up to 10,000 rounds of restrained conjugate gradient minimization with equivalent weights for all potential terms. Minimization was halted prior to the step limit when convergence was achieved. The program PROCHECK v.3.0 (Laskowski *et al.*, 1993) was used to test the chirality and geometry of all structures generated.

Choice of bounds

Identical distance geometry calculations were performed using sets of calculated d_R, with distance bounds of ±2, ±3, ±4, ±5, ±7, ±9 ±16, ±20, ±30, ±40 and ±50 Å. Ten structures were calculated for each bounds set, with d_R held constant. The structures were compared on the basis of the number of restraint violations, the total conformational energy, and the effective restraint contribution to the conformational energy. In addition, the structures were compared as described in detail in Results.

Choice of templates

Identical distance geometry calculations were carried out using distance bounds limits of ±5 Å and using starting templates with either a fully extended geometry (all *trans* peptide bonds, ϕ and $\psi = 180°$), or using the coordinates from the crystal structure determined at 1.65 Å (Loll & Lattmann, 1989). In the latter case, hydrogen atoms were explicitly built into the coordinate list,

and the structure was minimized by 200 rounds of conjugate gradient minimization using a potential containing bond angle and van der Waals repulsion terms only. Ten independent structures were calculated using each template.

Comparison of structures

The similarity within a family of structures and between structural families calculated with different parameters was determined using three characteristics: the radius of gyration, the solvent-accessible surface area with a probe radius of 1.4 Å (Lee & Richards, 1971), and the root-mean square of the C^{α} difference distance matrix.

Structural comparisons within and between ensembles of calculated structures were made by comparing the RMSD of equivalent C^{α} - C^{α} distances as:

$$\mathrm{RMSD}(C^{\alpha}_{\mathrm{distances}}) = \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} 2(r_{ij} - \langle r_{ij} \rangle)^2 / N(N-1) \right)^{1/2}$$

where N is the number of residues, r_{ij} is the distance in Å from C^{α} (residue i) to C^{α} (residue j), and $\langle r_{ij} \rangle$ is the same distance averaged over an ensemble of ten structures.

Structural validation

The validity and degree of determination of the structures generated for Δ131Δ (τ_c 4 ns, bounds ±5 Å) were assessed in two ways. In the first approach, randomly selected sets of 50 distance restraints were sucessively deleted from the list of 690 distance restraints derived from the entire experimental data set involving 12 spin labels, and an ensemble of ten structures was generated from the modified list. In the second approach, all of the restraints obtained from a single labeled protein were deleted from the restraint list, giving 12 separate data sets corresponding to each of the possible sets of 11 labeled proteins. Ten structures were generated and then compared to the average structure derived using all 690 restraints.

Acknowledgements

We thank the Biophysics Department for use of their Silicon Graphics Power Challenge 10000 GR for all of the computations using X-PLOR, and J. Wrabl, J. Sinclair and Y. Wang for their comments and suggestions on the manuscript. This research was supported by NIH grant GM34171 to D. S.

References

Alexandrescu, A. T. & Shortle, D. (1994). Backbone dynamics of a highly disordered 131-residue fragment of staphylococcal nuclease. *J. Mol. Biol.* **242**, 527–546.

Alexandrescu, A. T., Abeygunawardana, C. & Shortle, D. (1994). Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease, a heteronuclear NMR study. *Biochemistry,* **33**, 1063–1072.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, N. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–90.

Brunger, A. T. (1992). *X-PLOR 3.1 Manual,* Yale University Press, New Haven.

Dill, K. A. & Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825.

Dobson, C. M. (1994). Solid evidence for molten globules. *Curr. Opin. Struct. Biol.* **4**, 636–640.

Epstein, H. F., Schecter, A. N., Chen, R. F. & Anfinsen, C. B. (1971). Folding of staphylococcal nuclease: kinetic studies of two processes in acid renaturation. *J. Mol. Biol.* **60**, 499–508.

Ermacora, M. R., Ledman, D. W. & Fox, R. O. (1996). Mapping the structure of a non-native state of staphylococcal nuclease. *Nature Struct. Biol.* **3**, 59–66.

Gillespie, J. R. & Shortle, D. (1997). Characterization of long range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**, 158–169.

Krugh, T. R. (1979). Spin-label-induced nuclear magnetic resonance relation studies of enzymes. In *Spin Labeling II: Theory and Applications* (Berliner, L, ed.), pp. 339–372. Academic Press, New York.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the steroechemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–91.

Lee, B. K. & Richards, F. M. (1971). The interpretation of protein structure: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.

LeMaster, D. M., Kay, L. E., Brunger, A. T. & Prestegard, J. H. (1988). Protein dynamics and distance determination by NOE measurements. *FEBS Letters,* **236**, 71–76.

Loll, P. J. & Lattmann, E. E. (1989). The crystal structure of the ternary complex of staphylococcal nuclease, Ca^{2+} and the inhibitor pdTp, refined at 1.65 A. *Proteins: Struct. Funct. Genet.* **5**, 183–201.

Miranker, A. D. & Dobson, C. M. (1996). Collapse and cooperativity in protein folding. *Curr. Opin. Struct. Biol.* **6**, 31–42.

Mori, S., van Zijl, P. C. M. & Shortle, D. (1997). Measurement of water-amide proton exchange rates in the denatured state of staphylococcal nuclease by a magnetization transfer technique. *Proteins: Struct. Funct. Genet.* In the press.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Letters,* **229**, 317–324.

Oshiro, C. M., Thomason, J. & Kuntz, I. D. (1991). Effects of limited input distance constraints upon the distance geometry algorithm. *Biopolymers,* **31**, 1049–1064.

Sayle, R. (1995). In *RasMol V2.6, Molecular visualization program.* Published electronically on the World Wide Web. (ftp://ftp.dcs.ed.ac,uk/pub/rasmol).

Schmidt, P. G. & Kuntz, I. D. (1984). Distance measurements in spin-labeled lysozyme. *Biochemistry,* **23**, 4261–4266.
Shortle, D. (1996a). The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10**, 27–34.
Shortle, D., Wang, Yi, Gillespie, J. & Wrabl, J. O. (1996). Protein folding for realists, a timeless phenomenon. *Protein Sci.* **5**, 991–1000.
Wang, Y. & Shortle, D. (1995). The equilibrium folding pathway of staphylococcal nuclease, Identification of the most stable chain-chain interactions by NMR and CD spectroscopy. *Biochemistry,* **34**, 15895–15905.
Wang, Y. & Shortle, D. (1996). A dynamic bundle of four adjacent hydrophobic segments in the denatured state of staphylococcal nuclease. *Protein Sci.* **5**, 1898–1906.
Zhang, O., Forman-Kay, J. D., Shortle, D. & Kay, L. E. (1997). Triple-resonance NOESY-based experiments with improved spectral resolution: applications to structural characterization of unfolded, partially folded, and folded proteins. *J. Mol. Biol.* In the press.

Edited by P. E. Wright

(Received 12 December 1996; received in revised form 27 January 1997; accepted 31 January 1997)

http://www.hbuk.co.uk/jmb

Supplementary material comprising one Table is available from JMB Online.