Characterization of the residual structure in the unfolded state of the Δ131Δ fragment of staphylococcal nuclease

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · OCTOBER 2006

Impact Factor: 2.63 · DOI: 10.1002/prot.21077 · Source: PubMed

4 AUTHORS, INCLUDING:

Robert B Best
National Institutes of Health
126 PUBLICATIONS **5,180** CITATIONS
SEE PROFILE

Michele Vendruscolo
University of Cambridge
378 PUBLICATIONS **12,622** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Bioinformatics 65:145–152 (2006)

Characterization of the Residual Structure in the Unfolded State of the Δ131Δ Fragment of Staphylococcal Nuclease

Christopher J. Francis,[1,2] Kresten Lindorff-Larsen,[1] Robert B. Best,[1] and Michele Vendruscolo[1]*
[1]*Department of Chemistry, University of Cambridge, Cambridge, United Kingdom*
[2]*Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria, Australia*

ABSTRACT **The determination of the conformational preferences in unfolded states of proteins constitutes an important challenge in structural biology. We use inter-residue distances estimated from site-directed spin-labeling NMR experimental measurements as ensemble-averaged restraints in all-atom molecular dynamics simulations to characterise the residual structure of the Δ131Δ fragment of staphylococcal nuclease under physiological conditions. Our findings indicate that Δ131Δ under these conditions shows a tendency to form transiently hydrophobic clusters similar to those present in the native state of wild-type staphylococcal nuclease. Only rarely, however, all these interactions are simultaneously realized to generate conformations with an overall native topology. Proteins 2006;65:145–152.** © **2006 Wiley-Liss, Inc.**

INTRODUCTION

The characterization of the denatured state of proteins is a problem that has recently attracted much attention,[1–6] as the residual structure present in this state may affect the stability,[7] the folding,[8–11] the misfolding and aggregation[12–14] mechanisms, as well as the interpretation of Φ-values for folding.[15,16] Partially folded states of proteins are also associated with cellular processes such as translocation across membranes, signal transduction, and protein synthesis.[17] In addition, several proteins adopt a well-defined structure only after binding to a target molecule,[18,19] and an accurate structural description of these natively unfolded proteins is essential to understand the manner in which they function.[17,20–23]

For globular proteins under native conditions, nonnative states are difficult to study experimentally because they are not significantly populated. Thus, experimental investigations have been mainly carried out by truncation of the polypeptide chain,[24,25] removal of binding partners,[26] mutations of amino acids,[27] reduction of disulfide bridges,[28] lowering of pH, [29–31] and addition of denaturant.[11,28,29,32]

The Δ131Δ fragment of staphylococcal nuclease, which lacks residues 4 to 12 and 141 to 149 from the wild-type 149-residue sequence, has been studied as a model for the denatured structure of the wild-type protein under nondenaturing conditions.[24,33,34] Notably, the Δ131Δ fragment folds and is enzymatically active in the presence of binding partners[33] and its residual structure has been extensively characterized experimentally using a variety of NMR techniques[24,25,32,33,35,36] Structural studies of proteins in nondenaturing conditions are important, because addition of denaturant may alter the structure of the unfolded state by disrupting attractive interactions[26,37,38] as also suggested by single molecule studies that showed a significant compaction of the protein upon dilution of denaturant.[39]

NMR techniques exploiting nuclear Overhauser effects (NOE) or paramagnetic relaxation enhancements (PRE) through spin-labeling provide interatomic distance information;[40–42] PRE experiments are ideally suited probes for structure determination in unfolded states due to their long-range sensitivity.[22,23,34,43–45] In highly heterogeneous ensembles of conformations, such as those characteristic of unfolded states, dipole–dipole interactions depend on the inverse sixth power of the distance, and enforcing all distance restraints on a single molecule may result in a bias towards more compact structures. Therefore, the determination of the structure of non-native states is complicated by the need to interpret experimental measurements as ensemble averages[11,22,46] to avoid making the restraints too restrictive. A similar procedure has been shown recently to be important also in native states.[42,47–49] Several methods have been suggested to take ensemble averaging into account. Two of these include the assignment of weights to a precalculated unrestrained ensemble[50] or molecular dynamics with time-averaged restraints.[51] A third approach (ensemble-averaged refinement) involves the simulation of numerous structures simultaneously, such that the NMR distances are averaged across the set.[52,53] Ensemble-averaged distances are calculated at each step, and the pseudo-energy due to the violation across the ensemble is then evaluated. While any single structure may not satisfy all the restraints, the ensemble as a whole will do so. This method allows an extensive search of conformational space to be carried out, and it has been used to refine native structures.[42,54]

Kresten Lindorff-Larsen's present address is Department of Biochemistry, Institute of Molecular Biology and Physiology, University of Copenhagen, Universitetsparken 13, DK-2100 Copenhagen Ø, Denmark.

Robert B. Best's present address is Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 5 Bethesda, MD 20892-0520.

*Correspondence to: Michele Vendruscolo, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EN, UK. E-mail: mv245@cam.ac.uk

Received 15 June 2005; Revised 14 December 2005; Accepted 24 January 2006

Published online 21 July 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21077

Recently, we applied this approach to characterize at a coarse-grained level the conformational propensities in the denatured state ensemble of an acyl-coenzyme binding protein[11,45] and in the natively unfolded state ensemble of α-synuclein.[22]

In the present work we determine the residual structure of the Δ131Δ fragment of staphylococcal nuclease under native conditions by using a method based on all-atom molecular dynamics simulations with ensemble-averaged restraints from experimentally derived distances, and compare the results with those obtained without ensemble-averaged restraints. Our results suggest that Δ131Δ populates a broad ensemble of structures, which are stabilized by significant long-range hydrophobic interactions. Many of these interactions are clustered in the regions of Δ131Δ that are involved in the binding of the ligand. Although many of these interactions are individually native-like, they appear to form conformations with a global native-like topology only with a low probability.

METHODS

NMR-Derived Distance Restraints

Inter-residue distances for the Δ131Δ fragment of staphylococcal nuclease were determined experimentally at pH 5.3 and 32°C using paramagnetic spin labels by Gillespie and Shortle.[33,34] In that study, 14 PROXYL spin labels were introduced in Δ131Δ at sites with high solvent exposure in the native state of staphylococcal nuclease and that resulted in only small changes in stability and m values in the wild-type protein. The resulting 787 distances are used as ensemble-averaged restraints in the simulations presented in this work.

Restrained Monte Carlo C_α Simulations

As a first step in the modeling procedure, we obtained an ensemble of structures representing the state populated by the Δ131Δ fragment under physiological conditions by using a coarse-grained model and a Monte Carlo approach.[11] The protein was represented by a chain of C_α pseudo-atoms separated by 3.8 Å and having a hard-sphere diameter of 5 Å. We have recently shown that ensemble averaging can efficiently be taken into account by simulating in parallel an ensemble consisting of N_{rep} replicas of the protein.[11] In these simulations each Monte Carlo step involves the random selection of one residue from each of the N_{rep} replicas and the application of a crankshaft move.[55] Pairwise distances (r_{ij}) are averaged at each step according to

$$d_{ij}^{calc} = (N_{rep}^{-1} \sum_{k=1}^{N_{rep}} r_{ij,k}^{-6})^{-1/6} \quad (1)$$

and compared to PRE-derived distance restraints using a penalty function of the form

$$E = \sum_{i,j} (d_{ij}^{exp} - d_{ij}^{calc})^2 \quad (2)$$

In the calculation of E, upper and lower bounds on distance restraints were set to $d_{ij} \pm 5$ Å with a square-well potential, unless the corresponding d_{ij}^{exp} was estimated to be ≥20 Å, in which case the upper bound was left unrestrained. The acceptance of each step was evaluated according to the Metropolis criterion.[56]

Restrained Molecular Dynamics All-Atom Simulations

We first carried out a series of unrestrained molecular dynamics simulations the Δ131Δ fragment using CHARMM[57] with an implicit model for the solvent (EEF1)[58] at increasing temperatures (400 K, 500 K, 550 K, 600 K and 700 K). In each case, the simulation was started from a highly unfolded conformation and heated for 2 ns from 10 K to the desired temperature; the simulations were then continued for further 10 ns.

In order to obtain more detailed structural information about the unfolded state, we then performed all-atom ensemble molecular dynamics simulations with PRE-derived distance restraints.[42,52,53] Starting structures for the all-atom ensemble simulations were generated from randomly chosen structures obtained from the C_α simulations using the program MAXSPROUT.[59] Each structure was then minimized for 400 steps by steepest descent in the EEF1 force field.

Two sets of restrained simulations were performed: (1) by applying the distance restraints to a single molecule,[34] and (2) by following the procedure described above (Eqs. 1 and 2) for an ensemble of 25 replicas. In both cases, we used a biased molecular dynamics method[60] to obtain structures that satisfy the restraints. For this purpose, we defined a reaction coordinate

$$\rho(t) = \frac{1}{N_{PRE}} \sum_{i,j} (d_{ij}^{exp} - d_{ij}^{calc})^2 \quad (3)$$

where the sum is over the N_{PRE} experimentally measured distances d_{ij}^{exp} and d_{ij}^{calc}, the ensemble-averaged distances between C_α atoms (see Eq. (1), are allowed to vary freely between their experimental upper and lower bounds. The restraint energy E defined in Eq. 4 is then added to the standard force field

$$E(\rho,t) = \begin{cases} \frac{M\alpha}{2}(\rho - \rho_0)^2 & \text{if } \rho(t) \geq \rho_0 \\ 0 & \text{if } \rho(t) < \rho_0 \end{cases} \quad (4)$$

In this expression, M is the number of replicas and $\rho_0(t)$ is defined as

$$\rho_0(t) = \min_{0 \leq \tau \leq t} \rho(\tau) \quad (5)$$

This energy term is non-zero only when ρ is larger than the lowest value reached since the start of the simulation, so the system moves by spontaneous fluctuations in the direction of better agreement with experiments. The force constant α controls the relative weight of the restraints with respect to the force field. The structures were initially heated from 50 K to 540 K over 100 ps and α was set to 100; next, α was increased every 1 ns by a factor 10 (i.e., α = 10^3, 10^4, 10^5, 10^6). Sampling of the denatured state was

obtained in two 10 ns simulations, at 300 K and at 540 K, respectively, with $\alpha = 10^6$. These temperatures do not correspond to physical ones, but they control the balance between energy and entropy on the energy surface modified by the introduction of the experimentally derived restraints.

Cross-validation was conducted by removing 20% of the distance restraints chosen at random, and then back-calculating these distances for both 25 replica and single molecule simulations at 540 K for 10 ns. In order not to bias the cross-validated results towards the structures obtained by using all the available restraints, the structural ensemble was not started from the existing solution, but recalculated each time using the same procedure described above.

RESULTS

Ensemble-Averaged Simulations of the Δ131Δ Fragment

Restrained Monte Carlo C_α simulations

In the simulations by Gillespie and Shortle,[34] all the available distance restraints from PRE experiments were imposed on a single molecule. This procedure generates an ensemble of relatively compact structures, which satisfies a cross-validation test fairly well.[34] In this test, a fraction of the experimentally derived distances are used as restraints, and the remaining ones are predicted from the structures. It is important, however, to investigate in more detail the effect of the type of average given in Eq. 1. We therefore performed Monte Carlo replica simulations for values of N_{rep} ranging from 1 to 50. We found that the average radius of gyration (R_g) increased with the number of replicas from values corresponding to compact native-like structures (16 Å, for $N_{rep} = 1$) to values of about 23 Å (for $N_{rep} \geq 20$), as shown in Figure 1. The latter value was consistent with the finding that Δ131Δ is 1.3 to 1.5 times more expanded in the unfolded state than in the folded state.[32] Unrestrained simulations resulted in much more expanded structures with ($R_g = 37.0$ Å). In the simulations with $N_{rep} = 20$ only 7 of the 787 distances were violated and in no case by more than 1 Å.

Restrained all-atom molecular dynamics simulations

We carried out molecular dynamics simulations with ensemble-averaged restraints using $N_{rep} = 25$, consistent the Monte Carlo simulations described above and with previous studies of α-synuclein[22] and ACBP.[11,45] The simulations were performed at a pseudo-temperature of 540 K in order to obtain an ensemble with an average radius of gyration compatible with that estimated experimentally [Fig. 1(c)] All restrained distances satisfied the restraints within the bounds. The internal consistency of the method that we used of structure determination was tested by carrying out a cross-validation procedure, in which we left out 20% of the experimental restraints.[11,61] We observed that $R_{free} = \Sigma(d_{ij}^{calc} - d_{ij}^{exp})^2/\Sigma d_{ij}^{exp11}$ decreased from 0.69 in the one-replica simulations to 0.44 in the 25-replica simulations.

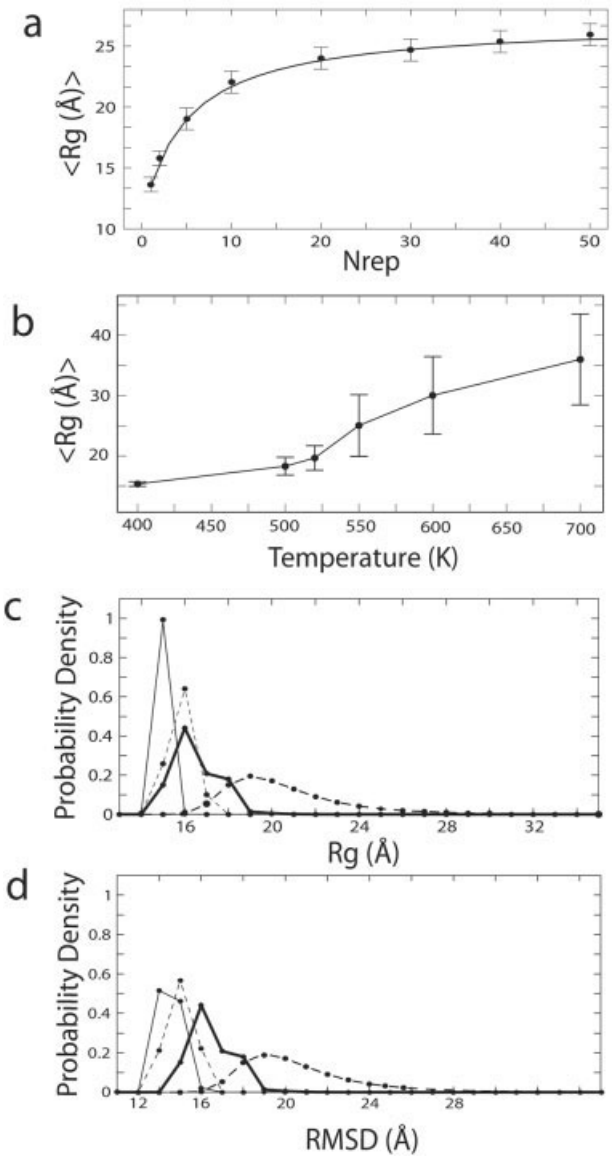
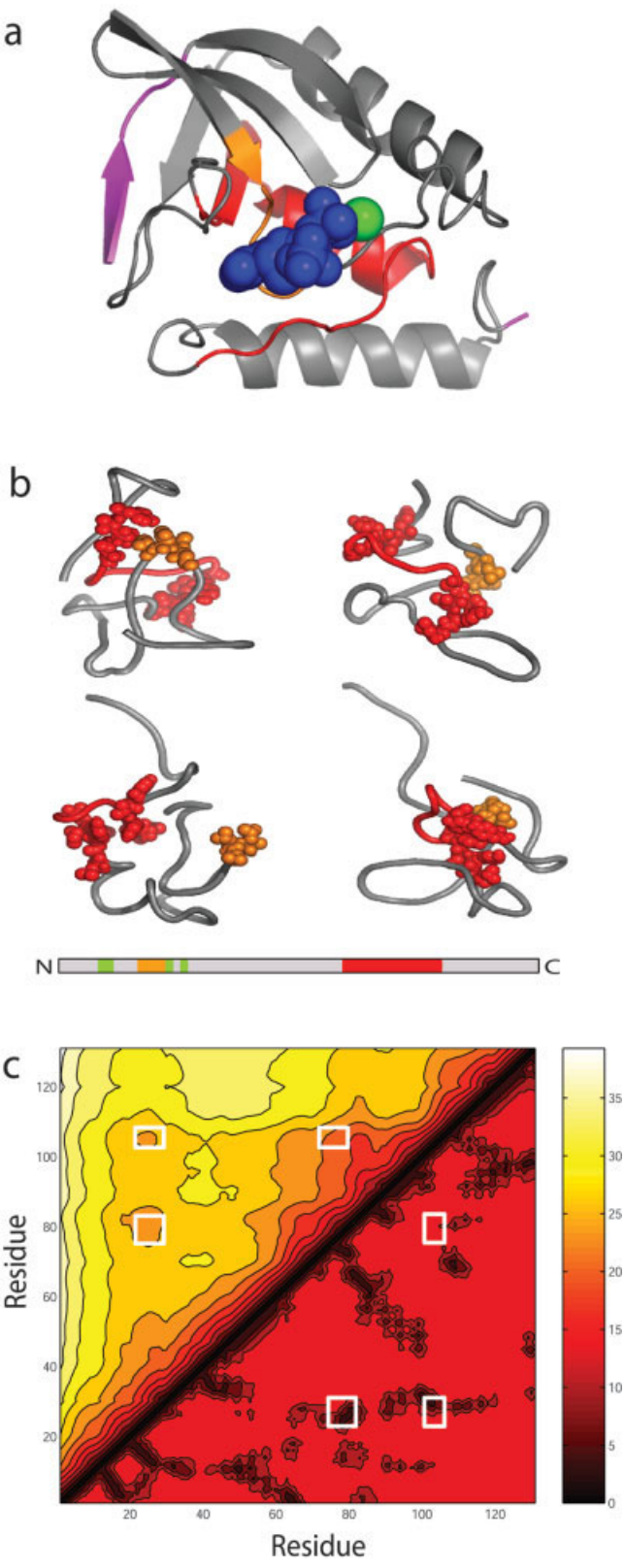


Fig. 1. (a) R_g obtained by Monte Carlo simulations as a function of the number of replicas (N_{rep}). The statistical errors were calculated from the probability distribution of R_g for each value of the number M of replicas. (b) Average R_g for unrestrained Δ131Δ molecular dynamics simulations as a function of the temperature. Error bars represent 1 standard deviation in the R_g values. Distributions of R_g (c) and of RMSD from the native state (d) for the ensemble of structures determined by restrained molecular dynamics simulations. The distributions corresponding to one-replica simulations at 300 K (thin solid line) and 540 K (dotted line) are shifted towards more compact values than the 25-replica simulations at 300 K (thick solid line) and at 540 K (dashed line). The latter simulations provide our model for the unfolded state of Δ131Δ and show that very rarely this protein populates conformations with overall native-like structural features under physiological conditions.

In order to analyze the structural properties of the ensemble of conformations that we determined, we calculated the average residue–residue distances over all the 25,000 structures generated during the sampling (Fig. 2). The interactions most consistently observed in the simulations occured between the three regions of the sequence formed by residues 25 to 30, 76 to 82 and 102 to 106,

respectively [boxed regions in Fig. 2(c)]. These regions correspond to peaks in the hydropathy profile[62] (see below) and are important for the binding to the ligand, as



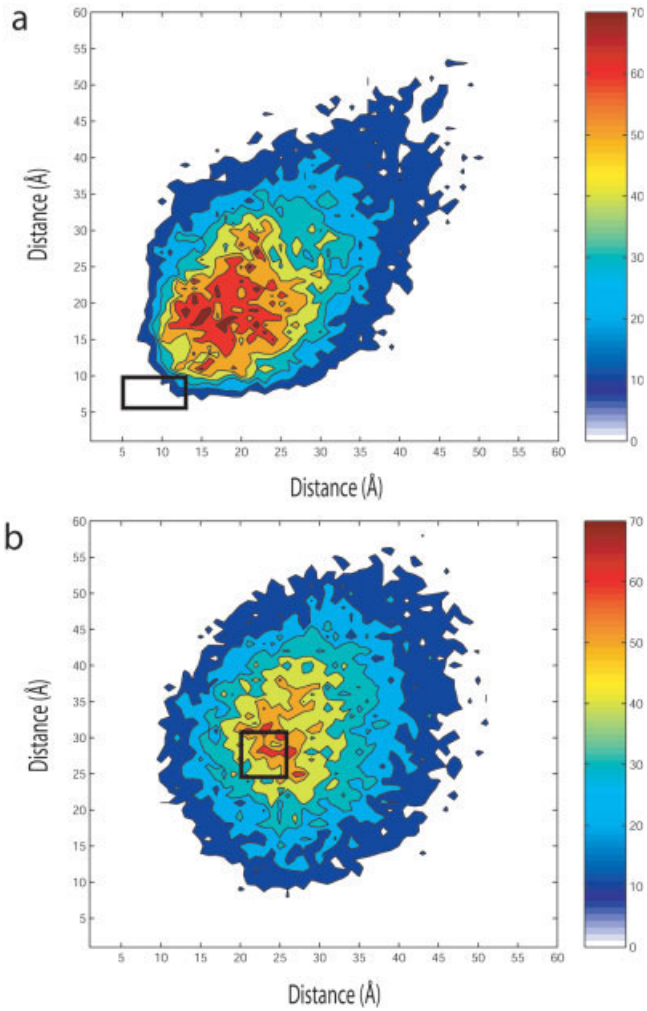


Fig. 3. (a) Two-dimensional histogram of the distances between three groups of residues in the main hydrophobic regions of Δ131Δ; (*x*-axis) distance between G_1 (residues 27–29) and G_2 (residues 80–82), and (*y*-axis) distance between G_1 and G_3 (residues 103–105). Boxed areas represent the average distances observed in the native structure. Although G1 and G2 (or G1 and G3) have a significant probability to be closer than 10 Å, it is unlikely for G1, G2, and G3 to form a single hydrophobic cluster, as they do in the native state of staphylococcal nuclease. (b) For comparison, we show the histogram of the distances between three regions that do not have a significant tendency to interact in Δ131Δ; (*x*-axis) distance between G_4 (residues 37–39) and G_5 (residues 90–92), and (*y*-axis) distance between G_4 and G_6 (residues 113–115). G4, G5, and G6 are never observed to form a single, non-native cluster. Distances were sorted into 1 × 1 Å bins.

Fig. 2. (a) Native structure of wild-type staphylococcal nuclease (PDB: 1SNC [63]). The two segments that are truncated in the Δ131Δ fragment (4–12 and 141–149) are colored in purple. The two main hydrophobic regions of the sequence (residues 25–30 and residues 76–106) are shown in gold and in red, respectively. The charged residues that chelate Ca^{+} ions (in green) and the pdTp inhibitor (in blue) are also indicated. (b) Selected structures of Δ131Δ that illustrate how the principal long-range interactions, in the cases when they are realized, involve residues in the two main hydrophobic regions (in gold and in red, respectively). Most of the structures that we generated are less compact than those shown here. (c) Comparison between the inter-residue distance map of Δ131Δ, as determined in this work (upper triangle), and that of the corresponding region in the native state of staphylococcal nuclease (lower triangle). This comparison shows that the closest long-range average distances (about 20 Å) in Δ131Δ involve the residues (shown as boxes) that interact with the ligand and that are located within the two main hydrophobic regions.

indicated by an X-ray crystallography study of staphylococcal nuclease in the presence of an inhibitor (pdTp) and Ca^{2+} ions.[63] Residues 26 to 32 form a combination of hydrogen bonds, electrostatic and hydrophobic interactions with pdTp. Further, the polar groups of residues K75, Y76, and R78 are involved in hydrogen bonds with the nucleotide phosphates while the aliphatic components of their side-chains, along with D74 and L80, provide a hydrophobic cavity hosting the remainder of the nucleotide. Such a hydrophobic cavity is completed by interactions Y104 and Y106.

In the simulations we included distances derived from the first two N-terminal spin labels, which were omitted in the studies of Gillespie and Shortle on the basis that the N-terminus should form an autonomously folding structure.[33,34] Our findings suggest that this region is important in making long range contacts; however we do not observe significant local structure within this region itself (Fig. 2).

We further studied whether the preferential interactions between the three main hydrophobic regions, which are in contact in the folded state, occur cooperatively. We thus considered three groups of three residues, G_1, G_2, and G_3, formed by (L27, L28, L29), (L80, A81, Y82) and (A103, Y104, V105), respectively. We present in Figure 3(a) a two-dimensional histogram of the average distance between G_1 and G_2 (x-axis) and the average distance between G_1 and G_3 (y-axis). The plot illustrates how the average distances between G_1 and G_2 and between G_1 and G_3 range from native-like values to rather more unfolded values. Only rarely, however, the three regions are in contact simultaneously and therefore only in a small fraction of the structures both distances are native-like at the same time. These results indicate a tendency of Δ131Δ to form transient native-like long-range interactions. Such interactions, however, often involve only a portion of the molecule, and in most cases the protein is not found in an overall native-like topology. For comparison, we also considered three other groups of residues, G_4, G_5, and G_6, formed by (H37, P38, K39), (V90, N91, E92) and (E113, N114, H115), respectively, which have a similar separation along the sequence as G_1, G_2, and G_3, but are not in contact in the native state. Our results indicate that these regions do not form preferential interactions in the unfolded state of Δ131Δ [Fig. 3(b)].

Disorder preferences

We compared the properties of the ensemble of structures of Δ131Δ that we determined with those obtained through methods to predict disordered regions in proteins. To characterize our structures we used the quantity

$$D_i = \frac{1}{N-3} \sum_{j \neq i, i \pm 1} r_{ij} \qquad (6)$$

where the sum is extended to all the C_α atoms in a sequence of N residues, excluding the nearest neighbors of residue i. The value of D_i can be considered as a measure of the overall coordination of residue i. High values of D_i indicate a high degree of expansion and of conformational heterogeneity in the region of residue i. A significant correlation was found between the D_i distances and the disorder probabilities obtained by using disEMBL[64] and PONDR VLS1;[65] the coefficients of correlation were 0.7 and 0.6, respectively (Fig. 4). These disorder probabilities are based on the extent to which amino acid sequences occur in regions of poorly defined electron densities in structures determined by X-ray crystallography. The correlation shown in Figure 4 indicates that the regions of the sequence that have the highest propensity to be unstructured are actually those that are most disordered in the unfolded state of Δ131Δ.

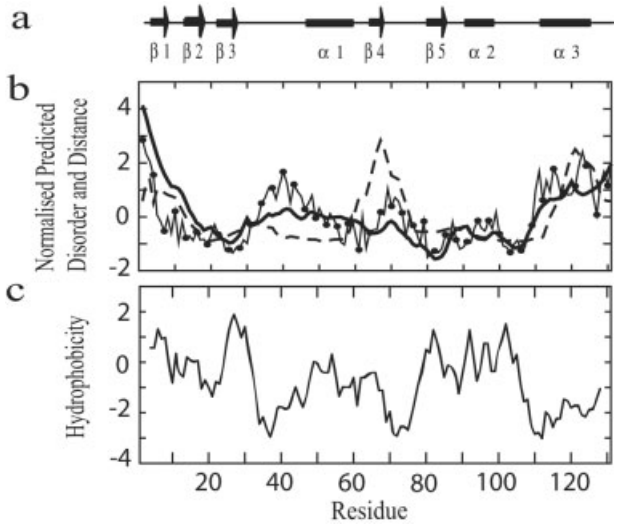


Fig. 4. (a) Schematic illustration of the native secondary structure as defined in the crystal structure.[63] (b) Comparison between the inter-residue distances (thick solid line) averaged over all residues nonadjacent in sequence and the chain disorder predicted using disEMBL[64] (thin solid line) and PONDR VLS1[65] (dashed line). Datasets were normalized to have a mean of 0 and a standard deviation of 1. (c) Hydrophobicity plot[62] obtained by using a window of five residues; large values correspond to more hydrophobic regions.

DISCUSSION

The conformational properties of the Δ131Δ fragment of staphylococcal nuclease have been extensively studied experimentally by a range of biophysical methods.[24,25,32–36] In particular, Gillespie and Shortle used PRE-derived distance information, distance geometry, and simulated annealing methods to model the C-terminal portion of 96 residues of Δ131Δ.[33,34] In order to sample expanded conformations, they carried out a series of simulations in which the upper-bound of the restraints was increased from 2 to 50 Å, and 10 structures were generated in each case. They noted that there are three significant limitations with that type of approach: (1) the distances are measured experimentally for an ensemble of molecules, rather than for a single one; (2) the experimental distances are averaged, with a r^{-6} weight; and (3) the time scale of the motions reported (10–100 ms) is considerably greater than the time scale accessible to simulations.

In this work we determined an ensemble of structures in order to characterize the state populated by Δ131Δ under

physiological conditions by using ensemble-averaged restrained molecular dynamics simulations. In this way, we aim at reducing significantly the problems arising from the three points described above. Our results indicate that the three main hydrophobic regions in the sequence, which are in spatial proximity in the folded state, have also the tendency to interact in the unfolded state (Fig. 2), although in very few cases these interactions take place simultaneously to give rise to an overall native-like topology (Fig. 3) of the type seen, for example, in computational studies of unfolded states of all α proteins.[66] The hydrophobic regions that appear to be most structured are also important for the function of staphylococcal nuclease, which hydrolyses double-or single-stranded DNA. The ligand binding site has been identified by X-ray crystallography studies in which a nucleotide-derived inhibitor (pdTp) and Ca^{2+} ions were shown to be bound through hydrophobic interactions with the deoxythymidine and chelating interactions with a small number of negatively charged N-terminal residues and the calcium.[63] An analysis of the average inter-residues distances in the ensemble of conformations that we generated reveals how long-range hydrophobic interactions are present, although transiently, between regions corresponding to residues 25 to 30 and 76 to 82 (Fig. 2). These results indicate the presence of partial formation of contacts corresponding to the N-terminal β-barrel in staphylococcal nuclease, principally between strands β3 and β5 (see Fig. 4 for a list of secondary structure elements), in addition to interactions extending over seventy residues from β3 to the unstructured region of residues 102 to 106 between α2 and α3. Figure 2 also illustrates how the hydrophobic region spanning residues 76 to 106, and corresponding to the principal ligand binding loop, is characterized by the presence of significant residual structure. This loop is also among the few regions in the Δ131Δ fragment in which nonsequential NOEs have been detected.[35]

It has been noted[67] that natively unfolded proteins are characterized by an average hydrophobicity (0.39 ± 0.05) lower than that of globular proteins (0.48 ± 0.03). In addition, when normalized by the sequence length, globular proteins tend to have a lower average net charge (0.04 ± 0.04) than natively unfolded proteins (0.12 ± 0.09). In the latter case, these differences may have evolved to avoid aggregation.[13,68] We analyzed Δ131Δ in the view of these observations about natively unfolded proteins. The truncation of wild-type staphylococcal nuclease into the Δ131Δ fragment increases the net positive charge from +8 to +9, owing to the loss of three positively charges residues from the N-terminus, and of four negatively charged residues, three of which are from the C-terminus. The normalized net charge at pH 7.0 of wild-type staphylococcal nuclease is 0.06, and is 0.07 for Δ131Δ, suggesting a charge profile consistent with both structural classes. Moreover, studies of the Δ131Δ fragment suggest that conformational dynamics are insensitive to mutations resulting in charge neutralizations[69] and that pKa perturbations in the unfolded state can be accounted for by a model that assumes nonspecific perturbations of a random coil.[70] Wild-type staphylococcal nuclease and Δ131Δ have mean hydrophobicities of 0.40 and 0.42, respectively. The presence of hydrophobic interactions in Δ131Δ was suggested from observations in which slow local dynamics, followed patterns of hydrophobicity.[69] The residual structure observed at pH 5.0 was dramatically reduced at pH 3.0 where the protein adopts a very high net positive charge.[69] So, while the major source of structural ordering in Δ131Δ appears to be due to the hydrophobic effect, neither the net hydrophobicity or the charge differ considerably from those of wild-type staphylococcal nuclease.

Taken together, these results indicate that the structural properties of Δ131Δ are intermediate between those of globular proteins and of natively unfolded ones, thus suggesting why the truncation of wild-type staphylococcal nuclease gives rise to a sequence with a weak but significant tendency to form preferential interactions, rather than a random-coil–like state. These results are also consistent with a recent study in which residual dipolar couplings were predicted with high accuracy for Δ131Δ for an ensemble of structurally heterogeneous structures.[6]

CONCLUSIONS

We used distances obtained by paramagnetic relaxation enhancement measurements as ensemble-averaged restraints in molecular dynamics simulations to determine an ensemble of conformations representing the state populated by the Δ131Δ fragment of staphylococcal nuclease under physiological conditions. Analysis of this ensemble revealed a tendency for Δ131Δ to explore transiently native-like interactions. In particular, a tendency to form residual structure was detected in the hydrophobic regions that are in proximity in the native structure and that are known to bind the pdTp ligand. Only rarely, however, these interactions create cooperatively conformations with an overall native-like topology. In other proteins, significant non-native interactions can also be present in the unfolded state, and they may help to avoid misfolding by burying hydrophobic groups in the early stages of folding to reduce the tendency to aggregate of partially folded conformations.[1] Therefore, the characterization of the residual structure in the unfolded states of specific proteins may reveal how weak conformational preferences may lead them towards conformations involved in the initial events leading to folding and, in less benign cases, to misfolding.

REFERENCES

1. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Durchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H. Long-range interactions within a nonnative protein. Science 2002;295:1719–1722.
2. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci U S A 2004;101:12491–12496.
3. Fitzkee NC, Rose GD. Reassessing random-coil statistics in unfolded proteins. Proc Natl Acad Sci U S A 2005;101:12497–12502.
4. Chattopadhyay K, Elson EL, Frieden C. The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods. Proc Natl Acad Sci U S A 2005;102:2385–2389.

5. Jha AK, Colubri A, Freed KF, Sosnick TR. Statistical coil model of the unfolded state: resolving the reconciliation problem. Proc Natl Acad Sci U S A 2005;102:13099–13104.
6. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci U S A 2005;102:17002–17007.
7. Ding KY, Louis JM, Gronenborn AM. Insights into conformation and dynamics of protein gbl during folding and unfolding by nmr. J Mol Biol 2004;335:1299–1307.
8. Capaldi AP, Kleanthous C, Radford SE. Im7 folding mechanism; misfolding on a path to the native state. Nature Struct Biol 2002;9:209–216.
9. Daggett V, Fersht AR. The present view of the mechanism of protein folding. Nature Rev Mol Cell Biol 2003;4:497–502.
10. Crowhurst KA, Forman-Kay JD. Aromatic and methyl NOES highlight hydrophobic clustering in the unfolded state of an SH3 domain. Biochemistry 2003;42:8687–8695.
11. Lindorff-Larsen K, Kristjansdottir S, Teilum K, Fieber W, Dobson CM, Poulsen FM, Vendruscolo M. Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. J Am Chem Soc 2004;126:3291–3299.
12. Ueda T, Yamada H, Aoki H, Imoto T. Effect of chemical modifications of tryptophan residues on the folding of reduced hen egg-white lysozyme. J Biochem 1990;108:886–892.
13. Uversky VN, Fink AL. Conformational constraints for amyloid fibrillation: the importance of being unfolded. Biochim Biophys Acta 2004;1698:131–153.
14. Platt GW, McParland VJ, Kalverda AP, Homans SW, Radford SE. Dynamics in the unfolded state of beta2-microglobulin studied by NMR. J Mol Biol 2005;346:279–294.
15. Wong KB, Clarke J, Bond CJ, Neira JL, Freund SM, Fersht AR, Daggett V. Towards a complete description of the structural and dynamic properties denatured state of barnase and the role of residual structure in folding. J Mol Biol 2000;296:1257–1282.
16. Salvatella X, Dobson CM, Fersht AR, Vendruscolo M. Determination of the folding transition states of barnase by using Φ_1-value-restrained simulations validated by double mutant $\Phi_{1,1}$-values. Proc Natl Acad Sci U S A 2005;102:12389–12394.
17. Uversky VN. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding corss-roads: which way to go? Cell Mol Life Sci 2003;60:1852–1871.
18. Dunker AK, Obradovic Z. The protein trinity — linking function and disorder. Nature Biotech 2001;19:805–806.
19. Fink AL. Natively unfolded proteins. Curr Opin Struct Biol 2005;15:35–41.
20. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nature Rev Mol Cell Biol 2005;6:197–208.
21. Vendruscolo M, Dobson CM. Towards complete descriptions of the free-energy landscapes of proteins. Phil Trans R Soc A 2005;363:433–450.
22. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM. Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 2005;127:476–477.
23. Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, M MZ. Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. Proc Natl Acad Sci U S A 2005;102:1430–1435.
24. Alexandrescu AT, Abeygunawardana C, and Shortle D. Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: a heteronuclear NMR study. Biochemistry 1994;33:1063–1072.
25. Choy WY, Shortle D, Kay LE. Side chain dynamics in unfolded protein states: an NMR based ^2H spin relaxation study of Δ131Δ. J Am Chem Soc 2003;125:1748–1758.
26. Zhang OW, Forman-Kay JD. NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. Biochemistry 1997;36:3959–3970.
27. Mayor U, Guydosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, Freund SMV, Alonso DOV, Daggett V, Fersht AR. The complete folding pathway of a protein from nanoseconds to microseconds. Nature 2003;421:863–867.
28. Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM. Structural and dynamical properties of a denatured protein. heteronuclear 3d nmr experiments and theoretical simulations of lysozyme in 8 m urea. Biochemistry 1997;36:8977–8991.
29. Arcus VL, Vuilleumier S, Freund SM, Bycroft M, Fersht AR. A comparison of the pH, urea and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. J Mol Biol 1995;254:305–321.
30. Yao J, Chung J, Eliezer D, Wright PE, Dyson HJ. NMR structural and dynamic characterization of the acid-unfolded state of apomyglobin provides insights into the early events in protein folding. Biochemistry 2001;40:3561–3571.
31. Thomsen JK, Kragelund BB, Teilum K, Knudsen J, Poulsen FM. Transient intermediary states with high and low folding probabilities in the apparent two-state folding equilibrium of ACBP at low pH. J Mol Biol 2002;318:1191–1199.
32. Shortle D, Ackerman MS. Persistence of Native-like topology in a denatured protein in 8M urea. Science 2001;293:487–489.
33. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. i. paramagnetic relaxation enhancement by nitroxide spin labels. J Mol Biol 1997;268:158–169.
34. Gillespie JR, Shortle D. 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. J Mol Biol 1997;268:170–184.
35. Zhang OW, Kay LE, Shortle D, Forman-Kay JD. Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease Delta 131 Delta, using NMR methods with improved resolution. J Mol Biol 1997;272:9–20.
36. Sallum CO, Martel DM, Fournier RS, Matousek WM, Alexandrescu AT. Sensitivity of NMR residual dipolar couplings to perturbations in folded and denatured staphylococcal nuclease. Biochemistry 2005;44:6392–6403.
37. Crowhurst KA, Tollinger M, Forman-Kay JD. Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. J Mol Biol 2002;322:163–178.
38. Mayor U, Grossmann JG, Foster NW, Freund SMV, Fersht AR. The denatured state of engrailed homeodomain under denaturing and native conditions. J Mol Biol 2003;333:977–991.
39. Schuler B, Lipman EA, Eaton WA. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature 2002;419:743–747.
40. Wüthrich K. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. Science 1989;243:45–50.
41. Donaldson LW, Skrynnikov NR, Choy WY, Muhandiram DR, Sarkar B, Forman-Kay JD, Kay LE. Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy. J Am Chem Soc 2001;123:9843–9847.
42. Iwahara J, Schwieters CD, Clore GM. Ensemble approach for NMR structure refinement against H-1 paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. J Am Chem Soc 2004;126:5879–5896.
43. Kragelund KTBB, Poulsen FM. Transient structure formation in unfolded acylcoenzyme A binding protein observed by site-directed spin labelling. J Mol Biol 2002;324:349–357.
44. Lietzow MA, Jamin M, Dyson HJ, Wright PE. Mapping long-range contacts in highly unfolded proteins. J Mol Biol 2002;322:655–662.
45. Kristjansdottir S, Lindorff-Larsen K, Fieber W, Dobson CM, Vendruscolo M, Poulsen FM. Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. J Mol Biol 2005;347:1053–1062.
46. Torda AE, Scheek RM, van Gunsteren WF. Time averaged distance restraints in molecular dynamics simulations. Chem Phys Lett 1989;157:289–294.
47. Best RB, Vendruscolo M. Determination of protein structures consistent with nmr order parameters. J Am Chem Soc 2004;126:8090–8091.
48. Clore GM, Schwieters CD. Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. Biochemistry 2004;43:10678–10691.
49. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendrus-

colo M. Simultaneous determination of protein structure and dynamics. Nature 2005;433:128–132.
50. Choy WY, Forman-Kay JD. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. J Mol Biol 2002;308:1011–1032.
51. Torda AE, Scheek RM, van Gunsteren WF. Time averaged distance restraints in NMR based structural refinement. In: *Computational aspects of the study of biological macromolecules by nuclear magnetic resonance Spectroscopy,* Hoch JC, editor, New York: Plenum Press; 1991:219–225.
52. Scheek RM, Torda AE, Kemmink J, van Gunsteren WF. Structure determination by NMR: The modeling of NMR parameters as ensmble averages. In *Computational aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy,* Hoch JC, editor, 209–217 (Plenum Press, New York, USA, 1991).
53. Bonvin AMJJ, Brunger AT. Conformational variability of solution nuclear-magnetic-resonance structures. J Mol Biol 1995;250:80–93.
54. Burling FT, Brunger AT. Thermal motion and conformational disorder in protein crystal-structures: Comparison of multi-conformer and time-averaging models. Isr J Chem 1994;34:165–175.
55. Vendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues form a critical contact network in protein folding transition state. Nature 2001;409:641–645.
56. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087–1092.
57. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. J Comp Chem 1983;4: 187–217.
58. Lazaridis T, Karplus M. Effective energy function for protein dynamics and thermodynamics. Proteins 1999;35:133–152.
59. Holm L, Sander C. Database algorithm for generating protein backbone and sidechain co-ordinates from a C alpha trace; application to model building and detection of co-ordinate errors. J Mol Biol 1991;218:183–194.
60. Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. J Mol Biol 1999;288:441–459.
61. Burling FT, Weiss WI, Flaherty KM, Brunger AT. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. Science 1996;271:72–77.
62. Kyte J, Doolittle RF. A simple method for displaying the hydrophatic character of a protein. J Mol Biol 1982;157:105–132.
63. Loll PJ, Lattman EE. The crystal structure of the ternary complex of staphylococcal nuclease, Ca2+, and the inhibitor pdTp, refined at 1.65 A. Proteins 1989; 5:183–201.
64. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: Implications for structural proteomics. Structure 2003;11:1453–1459.
65. Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 2006;61 S7:176–182.
66. Zagrovic B, Pande V. How does averaging affect protein structure comparison on the ensemble level? Biophys J 2004;87:2240–2246.
67. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.
68. Tompa P. Intrinsically unstructured proteins. Trends Biochem. Sci. 2002;27:527–533.
69. Choy WY, Kay LE. Probing residual interactions in unfolded protein states using NMR spin relaxation techniques: An application to Δ131Δ. J Am Chem Soc 2003;125:11988–11992.
70. Zhou HX. Residual charge interactions in unfolded staphylococcal nuclease can be explained by the Gaussian-chain model. Biophys J 2002;83:2981–2986.