and better than ±3.3‰ (1σ) between different ion microprobe targets. Each analysis is the weighted mean of a set of cycles (usually 15 cycles) ± the standard error on the mean for the set of cycles. This calculated precision is appropriate for the comparison of analyses made during the depth profiling of a single spot, but it does not account for spot-to-spot reproducibility between different cell aggregates, which was determined to be better than ±3.3‰ (1 SD) based on the reproducibility of the graphite powder analyses. Analyses were required to have count rates of greater than 500,000 counts per second on ^{12}C because of the background of the Faraday cup detector.
13. C. H. House *et al.*, *Geology* **28**, 707 (2000).
14. K. D. McKeegan, R. M. Walker, E. Zinner, *Geochim. Cosmochim. Acta* **49**, 1971 (1985).
15. S. J. Mojzsis *et al.*, *Nature* **384**, 55 (1996).
16. C. Paull, personal communication (2001). The mean $\delta^{13}C$ value of methane is –50.4‰ (n = 28).
17. V. Brüchert, L. M. Pratt, T. F. Anderson, S. R. Hoffmann, in *Proceedings of the Ocean Drilling Program, Scientific Research*, J. P. Kennett, J. Baldauf, M. Lyle, Eds. (Ocean Drilling Program, College Station, TX, 1995), pp. 219–229.
18. Archaeol and *sn*-2-hydroxyarchaeol were present in the oil seep sample in quantities too low for determination of their isotopic composition, but a careful examination of the m/z 45/44 ratio trace indicated relatively high δ-values for both components.
19. Some of the variation in these analyses may be attributed to analytical difficulties resulting from the probing of single cells from culture, not encountered during the analysis of cell aggregates. However, the variation observed is substantially greater than the reproducibility found during the repeated analysis of the graphite powder standard, reflecting some degree of organism-specific heterogeneity, probably magnified by culture conditions. The similarity of these mean carbon isotopic compositions to the bulk $\delta^{13}C$ values of each cell culture (–17.8 and –23.6‰, respectively) demonstrates the accuracy of our reported ion microprobe analyses once matrix effects are appropriately corrected.
20. The $\delta^{13}C$ ion microprobe analysis is an average carbon isotopic composition for all carbon exposed to the sputtering process. However, typically organic carbon has a sputter ionization yield about 100 times that of carbonate carbon.
21. R. E. Summons, P. D. Franzmann, P. D. Nichols, *Org. Geochem.* **28**, 465 (1998).
22. T. M. Hoehler, M. J. Alperin, D. B. Albert, C. S. Martens, *Glob. Biogeochem. Cycles* **8**, 451 (1994).
23. D. L. Valentine, W. S. Reeburgh, *Environ. Microbiol.* **2**, 477 (2000).
24. K.-U. Hinrichs, R. E. Summons, V. J. Orphan, S. P. Sylva, J. M. Hayes, *Org. Geochem.* **31**, 1685 (2000).
25. K.-U. Hinrichs, A. Boetius, in *Ocean Margin Systems*, G. W. Wefer *et al.*, Eds. (Springer-Verlag, Heidelberg, in press).
26. M. Elvert, E. Suess, J. Greinert, M. J. Whiticar, *Org. Geochem.* **31**, 1175 (2000).
27. We thank C. Paull (MBARI) and W. Ussler (MBARI) for graciously supplying the $\delta^{13}C$ of CH_4 data used in this study and M. Harrison (UCLA) for support during the development of this new ion microprobe application. We would also like to thank K. Buck (MBARI), S. Goffredi (MBARI), and the crew of the R/V *Western Flyer* and ROV *Tiburon* for their invaluable assistance with this research and C. Coath (UCLA), W. Bach (WHOI), K. Freeman (PSU), P. Girguis (MBARI), O. Beja (MBARI), for stimulating discussions and helpful advice during this project. Funding for this project was provided by the David and Lucile Packard Foundation, the Penn State Astrobiology Research Center and the University of California, Los Angeles, Center for Astrobiology, NASA National Astrobiology Institute. The UCLA ion microprobe is partially supported by a grant from the National Science Foundation Instrumentation and Facilities Program.

4 April 2001; accepted 5 June 2001

Persistence of Native-Like Topology in a Denatured Protein in 8 M Urea

David Shortle* and Michael S. Ackerman

Experimental methods have demonstrated that when a protein unfolds, not all of its structure is lost. Here we report measurement of residual dipolar couplings in denatured forms of the small protein staphylococcal nuclease oriented in strained polyacrylamide gels. A highly significant correlation among the dipolar couplings for individual residues suggests that a native-like spatial positioning and orientation of chain segments (topology) persists to concentrations of at least 8 molar urea. These data demonstrate that long-range ordering can occur well before a folding protein attains a compact conformation, a conclusion not anticipated by any of the standard models of protein folding.

In recent years, attention has turned to structural characterization of proteins whose native state has broken down in a major conformational transition termed unfolding or denaturation. When nuclear magnetic resonance (NMR) data for several small proteins (*1*, *2*) is combined with hydrodynamic and small-angle x-ray scattering data, the picture emerges that protein chains are relatively compact under mildly denaturing conditions, forming intermediates sometimes referred to as molten globules. As conditions are made less favorable for structure formation, most commonly by adding the chemical denaturants urea or guanidine hydrochloride, the denatured state gradually loses its residual structure and increases in size (*3*). At the highest concentrations of denaturants, the ensemble of conformations is expected to converge toward those of a statistical random coil.

One protein whose denatured states have been extensively studied is staphylococcal nuclease, a small α+β protein of 149 amino acids that lacks disulfide bonds or structural cofactors. Its relatively low stability allows the folded state to be broken down by a variety of perturbations. In the presence of 5 M urea, small-angle x-ray scattering has shown that the chain expands to a radius of gyration of almost 35 Å, more than twice the native state value of 16 Å (*4*). Nuclease can also be denatured by removing a few amino acid residues from both ends of the chain. The Δ131Δ fragment system, consisting of residues 10 to 140, refolds only in the presence of tight-binding ligands (*5*). In buffer at 32°C, it forms a somewhat expanded denatured state, with a radius of gyration approximately 1.3 to 1.5 times that of the folded state. A de novo structure determination of Δ131Δ using paramagnetic relaxation from 14 extrinsic spin labels revealed that many features of the folded arrangement of segment positions and orientations persist in this model denatured state (*6*). Application of this same method to a less structured form of nuclease, the low-salt, acid-denatured form, failed because the measured distance restraints were insufficient to constrain the ensemble of allowed conformations (*7*).

An alternate NMR approach to structure determination involves imposing a slight orientation on a macromolecule in solution by forcing it to tumble in an asymmetric environment (*8*). The small resulting alignment leads to incomplete cancellation of the dipolar coupling between magnetic nuclei close in space. The residual dipolar coupling D_{AB} contains information on the relative orientation of the vector between nuclei A and B with respect to one unique molecular axis determined by the molecular alignment tensor. Alternatively, the information can be interpreted in terms of angular relations between pairs of bond vectors that are independent of the intervening distance (*9*, *10*). With information from residual dipolar couplings alone, relatively high-resolution structures of the backbone can be calculated (*11*). Although physical theory suggests there would be complexities in converting dipolar couplings measured in a denatured protein into sets of orientational restraints (*12*), this approach is especially attractive because the structural information is distance independent.

To orient the denatured fragment Δ131Δ of staphylococcal nuclease, we used strained polyacrylamide gels (*13*, *14*). After diffusing a concentrated solution of protein into a gel cylinder and sliding it to the bottom of an NMR tube, the gel cavities were distorted from their initial spherical symmetry to ellipsoidal symmetries by mechanical compression (*15*). Given the chemical inertness of polyacrylamide, it is assumed that proteins interact with the gel matrix only through

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

*To whom correspondence should be addressed. E-mail: shortle@welchlink.welch.jhu.edu

steric overlap (*13*). Thus, as the molecule tumbles, its long axis (axes) will on average align with the long axis (axes) of the gel cavities, conferring a small net alignment. The line widths observed in NMR spectra demonstrated that the rate of tumbling of Δ131Δ was not greatly slowed by the surrounding gel matrix, as has been demonstrated quantitatively for folded proteins (*14*). The absence of significant differences in chemical shifts between spectra in gels and spectra in free solution suggested that the gel matrix does not greatly perturb the ensemble of conformations, although more expanded conformations will be disfavored relative to compact ones by an unknown amount.

Using protein samples uniformly labeled with ^{2}H and ^{15}N, NMR spectra permitting accurate measurements of the ^{1}H-^{15}N residual dipolar coupling D_{NH} (*16*) were collected on a 600-MHz spectrometer. Values of D_{NH} obtained by compressing and by stretching the same gel sample are shown in Fig. 1. When the values for individual residues in stretch mode are plotted against those in compression mode, a negative correlation is seen (*14*). Because the alignment mechanism is based on steric interactions between the protein and the gel matrix, one axial system of the protein, namely the inertia tensor, plays the same role in both environments. Gel compression with a plunger aligns the long axis of the molecule perpendicular to the magnetic field, whereas stretching aligns it along the field, leading to a reversal in sign of the observed couplings. These dipolar couplings demonstrate the presence of long-range structure that is not effaced by local dynamic motions. When averaged over the conformations assumed during tens of milliseconds, many residues maintain a relatively fixed orientation with respect to a single set of reference axes defined by the time-averaged molecular structure.

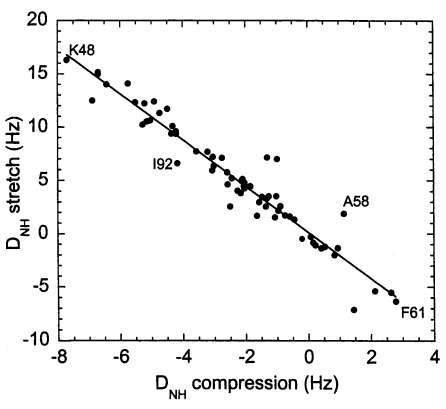

Fig. 1. A scatter plot of the N-H residual dipolar couplings D_{NH} for Δ131Δ in H_2O. After diffusion of protein (0.6 mM, at pH 5.2, 32°C) into a 12% polyacrylamide gel cylinder, the gel was either compressed with a plunger (*x* axis) or stretched by insertion into a smaller glass tube (*y* axis).

To demonstrate that these dipolar couplings arise from persistent structure in Δ131Δ, we attempted to eliminate them by addition of urea. As shown in Fig. 2, relatively large D_{NH} remain even in 8 M urea. Thus, long-range structure must persist under conditions that significantly reduce the forces stabilizing protein structure and that lead to large overall expansion. When the D_{NH} in urea are plotted against those in H_2O, statistically significant correlations ($P < 0.0001$) are found, with the correlation coefficient decreasing with increasing urea concentration. The results presented in Fig. 2 have been reproduced in every important detail in a second, solution-based alignment media (*17*) with the use of liquid crystals [see Web table 1 (*18*)]. Therefore, we conclude that the long-range structure of Δ131Δ in 8 M urea must be similar to that present in water alone, with chain segments retaining the same orientations relative to one another. Because structural studies described above established that Δ131Δ in H_2O has approximately the same topology as folded nuclease, it follows that the overall spatial positioning and orientation of different chain segments in 8 M urea must also be similar to that in the folded state.

There are several mechanisms that might reduce or eliminate the correlation between couplings seen in Fig. 2. Reduction in local structure in 6 M urea significantly reduces the order parameter for N-H bond vectors as measured by ^{15}N relaxation (*19*), a phenomenon that decreases the net projection of each bond vector along its average orientation. In addition, changes in overall shape could modify the inertia tensor of the molecule, altering the way the protein aligns within the gel cavity and thereby reducing the correlation between D_{NH}. From the correlations in Fig. 2, we conclude that the urea-induced expansion of the denatured state is approximately symmetric and does not greatly distort the ratios of axis lengths of the inertia tensor. Furthermore, the rapid local motions of the N-H bond vectors, which increase in angular amplitude at higher urea concentrations, must be sufficiently symmetric that the mean direction of the bond vector is not dramatically changed.

The residual dipolar couplings of full length, wild-type nuclease unfolded in 4 M urea correlate well with those of Δ131Δ under the same conditions (Pearson correlation coefficient, r = 0.91), but no correlation is found with the couplings of folded nuclease (data not shown). Because correlation between D_{NH} measured under two conditions occurs when the same molecular axes are involved in both alignment tensors, this finding demonstrates a significant change in the overall shape of nuclease on denaturation, as reflected in the inertia tensor.

The physical chemical basis for long-

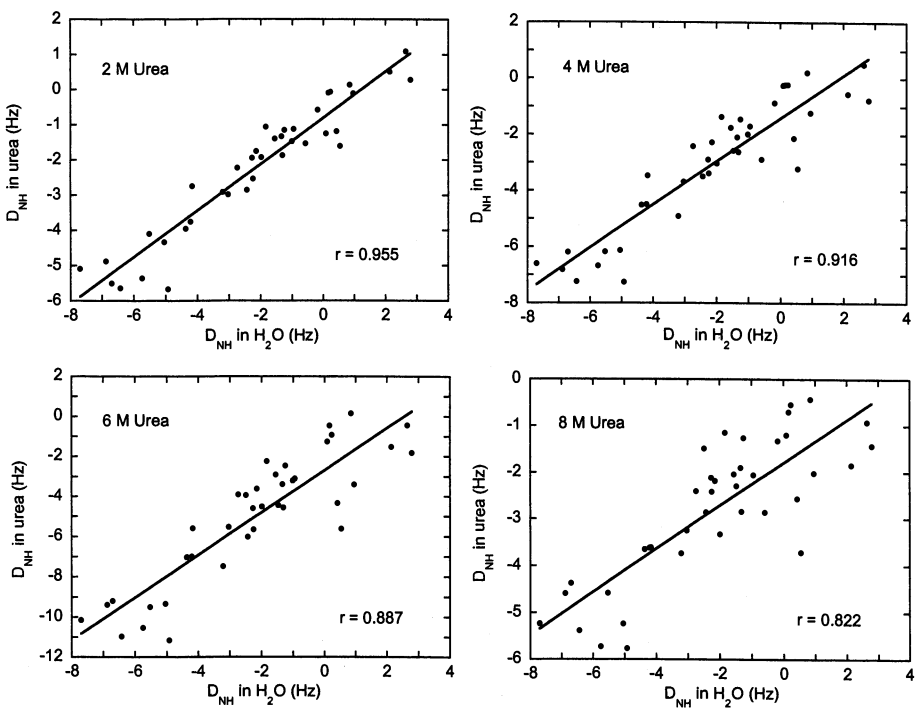

Fig. 2. Scatter plots of the N-H dipolar couplings D_{NH} for Δ131Δ in urea. The values measured at the indicated urea concentration are plotted along the *y* axis, whereas the values in H_2O are plotted along the *x* axis. Gel cylinders were strained by compression, and NMR data were collected as in Fig. 1. r is the Pearson correlation coefficient. All N-H peaks that displayed Lorentzian line shapes and could be assigned with confidence are included. Data are available at *Science* Online (*18*).

range structure in the presence of relatively little short-range structure is not clear. Because hydrophobic interactions are not completely eliminated in 8 M urea (*3*), one explanation is that these weakened, long-range interactions continue to place chain segments, on average, in the correct positions and orientations. Considering that the ensemble of denatured conformations in 5 M urea distributes the nuclease chain in approximately 10 times the volume of the native state, it is difficult to see how short-range attractive forces, such as dispersion forces or hydrogen bonds, could be responsible for this long-range structure.

An alternative explanation is that the basic topology of the polypeptide chain is not encoded by attractive interactions, but rather by steric repulsion between residues (*20*). If the correct topology permits a much larger number of conformations through greater variation of phi and psi angles for many residues, the protein chain will move into the "native-like" region of conformation space because of the high entropy it acquires there. Attempts to predict protein structure from sequence lend support to this explanation, through the finding of a larger number of structurally similar neighbors around the native conformation than around topologically incorrect conformations (*21*, *22*).

References and Notes

1. D. Shortle, *Curr. Opin. Struct. Biol.* **6**, 24 (1996).
2. Y. K. Mok, C. M. Kay, L. E. Kay, J. Forman-Kay, *J. Mol. Biol.* **289**, 619 (1999).
3. K. A. Dill, D. Shortle, *Annu. Rev. Biochem.* **60**, 795 (1991).
4. J. M. Flanagan, M. Kataoka, D. Shortle, D. M. Engelman, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 748 (1992).
5. A. T. Alexandrescu, C. Abeygunawardana, D. Shortle, *Biochemistry* **33**, 1063 (1994).
6. J. R. Gillespie, D. Shortle, *J. Mol. Biol.* **268**, 170 (1997).
7. _____, unpublished data.
8. N. Tjandra, *Struct. Fold. Des.* **9**, R205 (1999).
9. J. Meiler, N. Blomberg, M. Nilges, C. Griesinger, *J. Biomol. NMR* **16**, 245 (2000)
10. N. R. Skrynnikov, L. E. Kay, *J. Biomol. NMR* **18**, 239 (2000)
11. F. Delaglio, G. Kontaxis, A. Bax, *J. Am. Chem. Soc.* **122**, 2142 (2000).
12. The value of the dipolar coupling between ^{15}N and its covalently bonded ^{1}H is given by the equation (*8*)

$$D_{NH}=S\gamma_N\gamma_H/r^3_{NH}\{A_a(3\cos^2\theta - 1) + 3/2A_r(\sin^2\theta\cos2\phi)\}$$

where S is the generalized order parameter for the N-H bond vector, γ_N and γ_H are the gyromagnetic ratios of ^{15}N and ^{1}H, A_a and A_r are the axial and rhombic components of the alignment tensor, and θ and ϕ are the angles relating the orientation of the N-H bond vector to the alignment tensor. Although in folded proteins S can usually be approximated by 1.0, in $\Delta131\Delta$ it assumes a range of values from 0.4 to 0.9 (*23*). These measured values are model-dependent estimates, so their use will introduce considerable error into the angular restraints derived from the data.
13. R. Tycko, F. J. Blanco, Y. Ishii, *J. Am. Chem. Soc.* **122**, 9340 (2000)
14. H. J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, S. Grzesiek, *J. Biomol. NMR* **18**, 303 (2000).
15. Cylindrical 12% polyacrylamide gels (1:20 bis-acrylamide) were cast in segments of 5/32-inch ID Tygon tubing covered at one end with parafilm. After polymerization with 0.015% tetramethyl-ethylenediamine and 0.015% ammonium persulfate, each gel was removed, washed for 3 to 6 hours in distilled water, cut to a length of 21 mm, and placed in a microfuge tube containing 300 to 450 μl of protein solution (1.0 to 1.5 mM protein, 1 mM sodium azide, 10% D_2O, pH 5.1 to 5.3). Both the gel and the protein solution were made to the same urea concentration before incubation. After 4 to 18 hours at room temperature, the gel was pushed into an NMR tube. For compression, the 5-mm NMR tube was a Shigemi microcell with 4.5-mm ID, and the gel was compressed with a Shigemi glass plunger to a length of approximately 17 mm. To stretch the gel to approximately 30 mm, the NMR tube was a 35-mm segment of Wilmad 4-mm NMR tube (3.2-mm ID), which was then positioned inside a 5-mm Shigemi microcell with 4.2-mm ID.
16. M. Ottiger, F. Delaglio, A. Bax, *J. Magn. Reson.* **131**, 373 (1998).
17. M. Ruckert, G. Otting, *J. Am. Chem. Soc.* **122**, 7793 (2000).
18. Web table 1 is available at *Science* Online at www.sciencemag.org/cgi/content/full/293/5529/487/DC1.
19. J. F. Sinclair, D. Shortle, *Protein Sci.* **8**, 991 (1999).
20. R. Srinivasan, G. D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14258 (1999).
21. D. Shortle, K. T. Simons, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11158 (1998).
22. R. Bonneau, C. E. M. Strauss, D. Baker, *Proteins Struct. Func. Genet.* **43**, 1 (2001).
23. A. T. Alexandrescu, D. Shortle, *J. Mol. Biol.* **242**, 527 (1994).

24. We thank M. Markus and D. Torchia for helpful discussions, L. Kay for providing us with the HSQC-IPAP pulse sequence, F. Delaglio for providing NMRPipes, and M. Massiah for assistance with NMR processing software. Supported by NIH grant GM34171 (D.S.).

6 March 2001; accepted 5 June 2001

Haplotype Variation and Linkage Disequilibrium in 313 Human Genes

J. Claiborne Stephens,* Julie A. Schneider, Debra A. Tanguay, Julie Choi, Tara Acharya, Scott E. Stanley, Ruhong Jiang, Chad J. Messer, Anne Chew, Jin-Hua Han, Jicheng Duan, Janet L. Carr, Min Seob Lee, Beena Koshy, A. Madan Kumar, Ge Zhang, William R. Newell, Andreas Windemuth, Chuanbo Xu, Theodore S. Kalbfleisch, Sandra L. Shaner, Kevin Arnold, Vincent Schulz, Connie M. Drysdale, Krishnan Nandabalan, Richard S. Judson, Gualberto Ruaño, Gerald F. Vovis

Variation within genes has important implications for all biological traits. We identified 3899 single nucleotide polymorphisms (SNPs) that were present within 313 genes from 82 unrelated individuals of diverse ancestry, and we organized the SNPs into 4304 different haplotypes. Each gene had several variable SNPs and haplotypes that were present in all populations, as well as a number that were population-specific. Pairs of SNPs exhibited variability in the degree of linkage disequilibrium that was a function of their location within a gene, distance from each other, population distribution, and population frequency. Haplotypes generally had more information content (heterozygosity) than did individual SNPs. Our analysis of the pattern of variation strongly supports the recent expansion of the human population.

Large-scale investigations of sequence variation within the human species have only just begun (*1*–*3*). Initial estimates are that sequence differences between an individual's maternal and paternal genomes occur on average at about every 500 to 2000 bases (*2*–*4*), with the most frequently cited value being one difference approximately every kilobase (*5*). However, relatively little is known about the pattern of DNA sequence variation among humans, within a population and between different populations. In particular, the pattern of linkage disequilibrium among closely spaced SNPs, for example, those that are less than 20 kb apart, is known only for a few well-studied genes, and the results from these studies are highly discordant (*6*–*9*).

We have undertaken a systematic discovery of gene-based sequence variation in 82 unrelated individuals, whose ancestors were from various geographical origins. The sample size and composition were sufficient to detect, with high certainty, globally distributed variants present at a frequency of at least 2% and population-specific variants present at a frequency of at least 5%. Our population sample, using the definitions of the U.S. Cen-

Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511, USA.

*To whom correspondence should be addressed. E-mail: c.stephens@genaissance.com