Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 10600–10603, October 1996
Biochemistry

Initiation sites of protein folding by NMR analysis

(heteronuclear NMR/barnase/nucleation/denatured/cross relaxation)

STEFAN M. V. FREUND, KAM-BO WONG, AND ALAN R. FERSHT

Medical Research Council Unit for Protein Function and Design and Cambridge Centre for Protein Engineering, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

Contributed by Alan R. Fersht, July 16, 1996

ABSTRACT **Detailed characterization of denatured states of proteins is necessary to understand the interactions that funnel the large number of possible conformations along fast routes for folding. Nuclear magnetic resonance experiments based on the nuclear Overhauser effect (NOE) detect hydrogen atoms close in space and provide information about local structure. Here we present an NMR procedure that detects almost all sequential NOEs between amide hydrogen atoms (H_N–H_N NOE), including those in random coil regions in a protein, barnase, in urea solutions. A semi-quantitative analysis of these H_N–H_N NOEs identified partly structured regions that are in remarkable agreement with those found to form early on the reaction pathway. Our results strongly suggest that the folding of barnase initiates at the first helix and the β-turn between the third and the fourth strands. This strategy of defining residual structure has also worked for cold-denatured barstar and guanidinium hydrochloride-denatured chymotrypsin inhibitor 2 and so should be generally applicable.**

Small proteins can fold very rapidly from denatured states that could have an astronomical number of conformations. This has inspired theoreticians to search for processes that could narrow the search for the native state (1–4) and experimentalists to search for residual structure in denatured states of polypeptides (5, 6). The most specific spectroscopic technique is nuclear magnetic resonance (NMR), which generally involves assessing short range interresidual contacts, mainly H_N–H_N NOEs, and analyzing other parameters, such as backbone $^3J_{HNH\alpha}$ coupling constants and chemical shift deviations from random coil values, which depend on conformation. In contrast to studies on native proteins, NMR spectroscopy of denatured states is complicated because of the poor resolution of signals and because the denatured state is a rapidly interconverting mixture of different populations. Just the average environment of a residue is measured, which tends to represent a small proportion of structured conformation and a large fraction of random conformations. The NMR experimental evidence is therefore heavily biased toward random structure and hence close to random coil values. Nuclear Overhauser effect (NOE) measurements with their r^{-6} distance dependency have the highest sensitivity to detect minor populations of nonrandom behavior. However, observation of NOEs in denatured proteins is often hampered by severe overlapping of signals in the ^{1}H dimensions. The high dispersion found in the ^{15}N dimensions for denatured proteins may allow the separation of otherwise unresolved signals. Here we demonstrate that a gradient selected sensitivity enhanced heteronuclear single quantum coherence (HSQC)–nuclear Overhauser effect spectroscopy (NOESY)–HSQC experiment (7, 8), which offers optimal resolution for the analysis of short ranging sequential amide-to-amide distances by dispersing H_N–H_N NOE cross-peaks into two nitrogen dimensions, allows the detection of almost all sequential H_N–H_N NOEs for urea-denatured barnase. Estimated cross relaxation rates R_cs based on cross-peak-to-diagonal peak ratios were used to define residual structure in the urea-denatured state of barnase. We discuss the results in context of the known folding pathway of barnase.

MATERIALS AND METHODS

NMR Spectroscopy. An NMR sample of 2.0 mM uniformly ^{15}N-labeled barnase in 5.5 M urea solution at pH 4.5 was prepared as described (9). ^{15}N gradient selected sensitivity enhanced HSQC–NOESY–HSQC experiments were acquired on a Bruker AMX 500 spectrometer, equipped with an inverse triple-resonance single-axis gradient probe. All experiments were acquired as matrices of 1024 × 64 × 64 complex points (F3, F2, F1) with 8 transients per t_1 and t_2 increment. The sweep widths were set to 8064, 1572, and 1572 Hz (F3, F2, F1). The acquisition domain was zero-filled to 4096 complex points followed by application of a Gaussian window function. The F2 dimension was extended to 128 complex points by linear prediction (10). F1 and F2 dimensions were then zero-filled to 256 complex points and weighted with sine and Gaussian window functions respectively. All data were processed and analyzed in FELIX 2.3 (Biosym Technologies, San Diego).

Calculation of the R_c. R_c was determined as $1/t_m \ln[(1 + x)/(1 - x)]$, where t_m is the mixing time and x is the experimental signal intensity ratio between NOE and diagonal peaks (11, 12). Signal intensities of well-resolved cross-peaks were quantified as integrated peak areas by a lineshape fitting of one dimensional vectors (F2) of the ^{15}N HSQC–NOESY–HSQC experiments.

RESULTS AND DISCUSSION

Detection of H_N–H_N NOEs for Almost All Residues in Denatured Proteins. The distance d_{NN}, and hence the strength of the H_N–H_N NOE, is a function of the backbone Φ, Ψ angles (13): $d_{NN} = (13.6 + 2.69 \cos\Phi - 7.26 \cos\Psi - 1.26 \cos\Phi \cos\Psi - 3.73 \sin\Phi \sin\Psi)^{1/2}$. In theory, the distance d_{NN} between two adjacent amide protons can vary between the van der Waals limit of 2.0 Å and a maximum distance of about 4.7 Å. For example, in the sterically allowed α-region of the (Φ,Ψ) space of the Ramachandran plot (14), d_{NN} is in the region of 2.8 Å for a regular helix, whereas for β-sheet structures, which occupy the β-region of the (Φ,Ψ) space, d_{NN} is around 4.3 Å. A random coil mainly occupies the β-region, which results in an averaged distance of between 4.0 and 4.5 Å. In theory, all of these values are within the limits of NOE detection. However, in practice, limitations are caused by intrinsic sensitivity and severe overlapping of signals.

We have used a three-dimensional ^{1}H, ^{15}N, ^{15}N–HSQC–NOESY–HSQC experiment (7, 8), which offers optimal separation of signals by dispersing NOE cross-peaks into two nitrogen dimensions. The observed H_N–H_N NOEs are fre-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: NOE, nuclear Overhauser effect; HSQC, heteronuclear single quantum coherence; NOESY, nuclear Overhauser effect spectroscopy; R_c, cross relaxation rate.

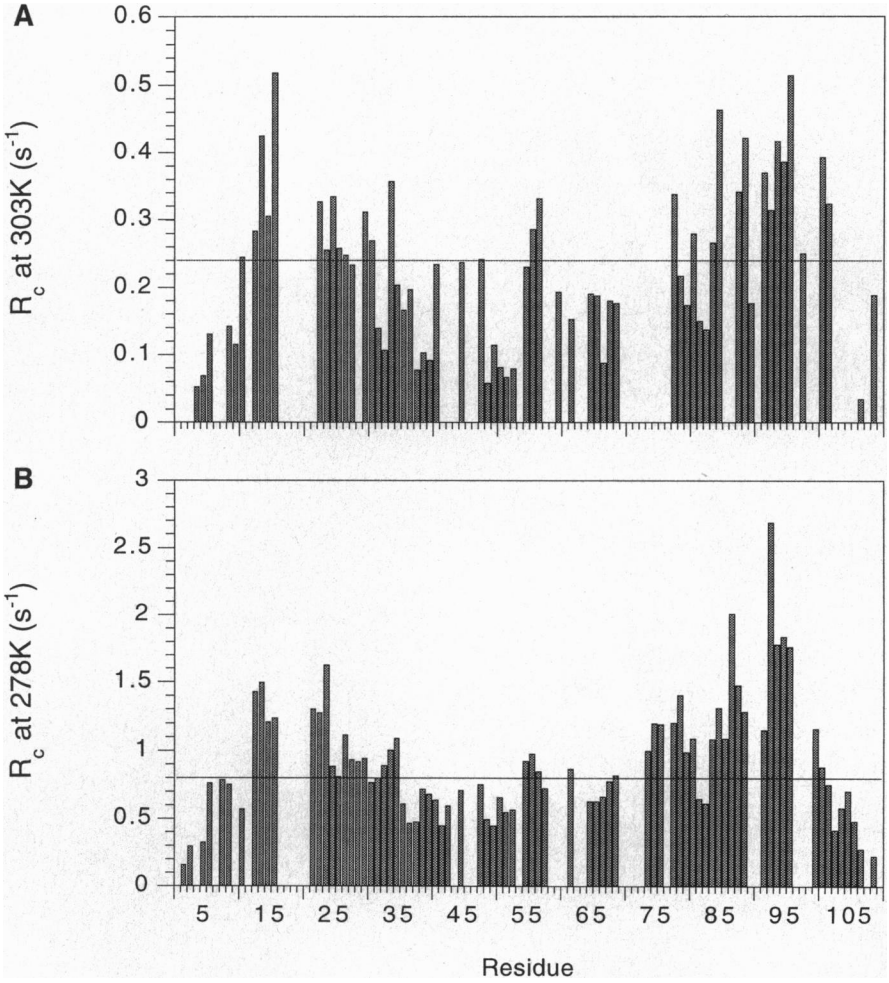

FIG. 1. R_cs obtained for urea-denatured barnase at 303 K (*A*) and 278 K (*B*).

quency labeled with two ^{15}N frequencies and so are well separated. Ninety-three of 106 H_N–H_N NOEs of barnase denatured in 5.5 M urea (pH 4.5) were observed at a mixing time of 250 ms. A parallel study on cold-denatured barstar (15) and guanidinium hydrochloride-denatured chymotrypsin inhibitor 2 (J. Tan and A.R.F., unpublished results) revealed virtually complete sets of H_N–H_N NOEs (79 of 86 and 56 of 59, respectively). That H_N–H_N NOEs were obtained for almost all residues suggests that sequential amide-to-amide contacts can indeed be observed for all regions of denatured proteins, including random coil. Previously, the simple qualitative observation of H_N–H_N NOEs in denatured proteins has been used as evidence for the presence of residual helical structure because of their shorter H_N–H_N distances (5, 6). But, as we observe so many H_N–H_N NOE cross-peaks, a semi-quantitative analysis of intensities is clearly necessary.

Estimation of R_cs. The quantitative description of H_N–H_N NOEs is complicated by leakage mechanisms that attenuate the signal intensities. Observed large intensity variations in the ^1H, ^{15}N correlation experiments indicate that these leakage mechanisms vary for different residues and, therefore, distort the analysis of NOE intensity measurements. To remove this contribution we have adopted an approach (11, 12) that estimates R_cs based on the cross-peak-to-diagonal peak ratio. The "population weighted" R_cs defining the dipolar interaction between two amide signals is given by (11, 12):

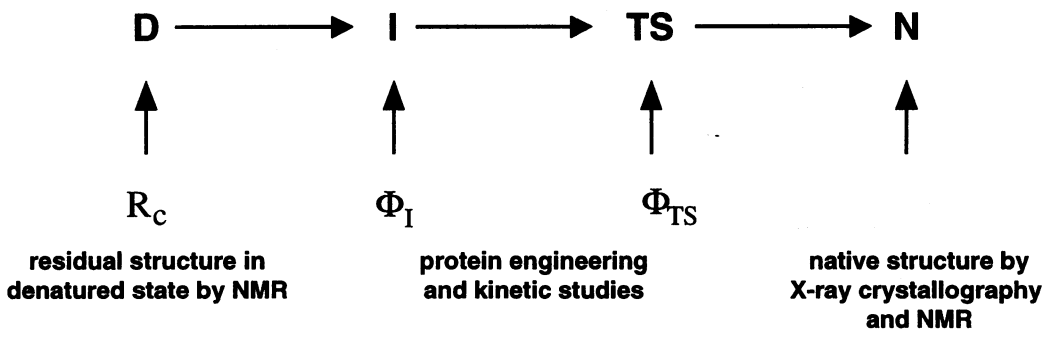

FIG. 2. Schematic diagram of the folding pathway of barnase. Starting from the denatured state (D), barnase folds via a rapid formation of an intermediate (I) followed by a rate-determining transition state (TS), which leads to the native state (N). The structure of different states can be analyzed by NMR (D, N), protein engineering (I, TS), and x-ray crystallography (N).

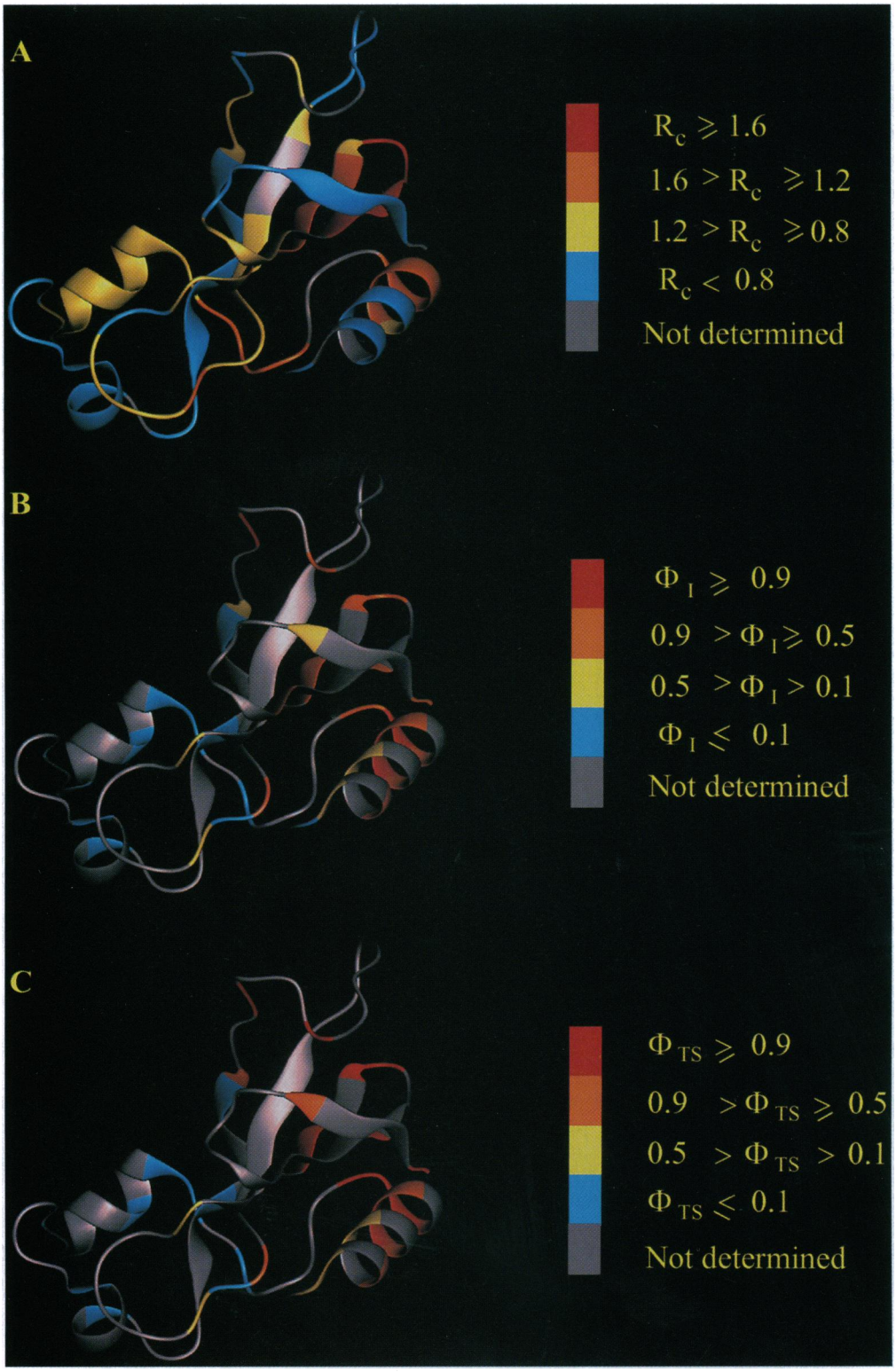

FIG. 3. Comparison of R_c and Φ-value analysis. (*A*) Residual structure found in the denatured state (D) is quantified based on R_cs. The trend of R_c observed at 278 K (see Fig. 2*B*) was enhanced by applying a 3-point smoothing function (KALEIDAGRAPH 3.0, Synergy Software). Residues with $R_c \geq 1.6$, $1.6 > R_c \geq 1.2$, $1.2 > R_c \geq 0.8$ and $R_c < 0.8$ are colored red, orange, yellow, and blue, respectively. (*B* and *C*) Results obtained from Φ_I and Φ_{TS} value analyses, respectively (17). Residues with $\Phi \geq 0.9$, $0.9 > \Phi \geq 0.5$, $0.5 > \Phi > 0.1$, and $\Phi \leq 0.1$ are colored red, orange, yellow, and blue, respectively. For example, a Φ_I value of zero indicates that the sequence position in the intermediate (I) is as unfolded as in the denatured state (D), whereas a Φ_I value of 1 indicates a fully formed interaction at this position as in the native state (N). Partial Φ values ($0 < \Phi < 1$) correspond to either partly formed interactions or a mixed state. Grey shading indicates that the R_c, Φ_I or Φ_{TS} values were not determined for these residues.

$$R_c = 1/t_m \ln[(1 + x)/(1 - x)] = AJr^{-6}, \quad [1]$$

where x is the experimental intensity ratio of H_N–H_N NOE cross-peak-to-diagonal peak; t_m is the mixing time; constant $A = 0.2\gamma^4(h/2\pi)^2(\mu_0/4\pi)^2$, where γ is the magnetogyric ratio for hydrogen, h is Planck's constant, μ_0 is the permeability of free space; the term $J = |6J(2\omega_H) - J(0)|$, where $J(0)$ and $J(2\omega_H)$ are the spectral density functions at frequencies 0 and

$2\omega_H$, respectively; r is the population averaged H_N–H_N distance.

The R_c depends both on interproton distances and on dynamic behavior (16). According to Eq. **1**, large R_cs correspond to a decrease in the average H_N–H_N distance r and/or an increase in the term J. Our results from ^{15}N relaxation studies reveal that $J(0)$ is at least two orders of magnitude larger than $J(2\omega_H)$ (unpublished results). To a first approximation, J can be simplified, therefore, to $J(0)$. To illustrate the physical meaning of $J(0)$, $J(0) = 2/5\tau_c$ (where τ_c is the overall correlation time) in the simplified case of an isotropic sphere. In summary, an increase in R_c indicates either shorter averaged H_N–H_N distances and/or an increase in the effective correlation time, which reflects local restriction of backbone mobility. For residual structure populating the sterically allowed α-region of the (Φ, Ψ) space, shorter averaged H_N–H_N distances and increased effective correlation times both contribute to increased R_c. For residual structure occupying the β-region of the (Φ, Ψ) space, the averaged H_N–H_N distance is similar to random coil values, but the local restriction in mobility increases the effective correlation time. The R_c is therefore a general parameter defining residual structure in denatured proteins. It has to be emphasized that (*i*) R_c has a higher sensitivity to detect residual structure in the α-region of (Φ, Ψ) space but does not exclude the detection of residual structure in the β-region and (*ii*) the R_c, and hence NOE, reflects variations in interproton distances, as well as dynamic properties of the denatured state. To separate the contribution from interproton distances and effective correlation times, ^{15}N relaxation time measurements are in progress to estimate the values of $J(0)$.

Residual Structure in Urea-Denatured Barnase Defined by R_c. We have used R_cs to quantify residual structure in urea-denatured barnase, the 110-residue ribonuclease whose folding pathway has been the subject of intense investigation (17). Fig. 1 summarizes the R_cs obtained for barnase at 303 K (*A*) and at 278 K (*B*). Lowering the temperature increases the R_c in general and enhances the trends obtained at 303 K. Large R_cs (R_c at 303 K $> 0.24\ s^{-1}$, R_c at 278 K $> 0.80\ s^{-1}$) were obtained for the regions 13–16, 22–35, 56–57, 74–81, 84–89, 92–96, and 100–102. An earlier study of acid- and urea-denatured barnase (9, 18) found a limited set of NOEs within these regions.

Implications for the Folding of Barnase. Folding of barnase involves the rapid formation of an intermediate (I) followed by the rate determining transition state (TS) (Fig. 2). The denatured state (D) can be described as an ensemble of interconverting conformers with transient structured regions, which are defined via R_cs. On the other hand, the structure of the intermediate (I) and the transition state (TS) has been characterized by Φ-value analysis, which measures the relative degree of formation of structure by analyzing the kinetics of folding and unfolding of mutants (17, 19). This combined approach of using R_c and Φ-value analysis thus provides a detailed structural study of all major species along the folding coordinate.

We find a striking correlation between residual structure in the urea-denatured state and regions that were found to fold early from protein engineering methods. Residual structure in the denatured state (D) as defined by large R_cs is found in the following regions (Fig. 3*A*): (*i*) near the C terminus of the first helix, (*ii*) in the loop connecting the first and second helices, (*iii*) near the β-turn between the third and the fourth strands, and (*iv*) to a lesser extent, in the second helix. The structure of the intermediate (I) involves tertiary interactions among the first helix, the third and the fourth strands, and the C-terminal residues (Fig. 3*B*). Partial Φ-values for most of these regions suggest either a partially folded structure or a population of different states. Moving to the transition state (TS) (Fig. 3*C*), the structure found in the intermediate (I) consolidates with most of the Φ-values increased to 1, indicating fully formed interactions in these regions. That residual structure is found in only the first helix and the β-turn between the third and the fourth strands, but not in the C terminus, suggests that (*i*) the first helix and the β-turn are initiation sites for folding and (*ii*) these two initiation sites develop into a nucleus for folding, which contains tertiary interaction involving the C-terminal residues. Our findings agree with the nucleation–condensation mechanism for protein folding (20), which proposes that there are embryonic initiation sites that are dominated by local interactions among adjacent residues, which develop into nucleation sites when they are stabilized by making interactions with residues remote in the primary sequence. Residual structure found in the second helix and the loop between the first and the second helices is clearly not, therefore, productive as the initiation site. Our results strongly indicate that helix formation in the first helix is initiated at the C terminus, consistent with the Φ-value analysis. Moreover, the R_c rates are the largest among the residues near the β-turn between the third and the fourth strands. It is tempting to assume that the β-turn is the starting point for β-sheet formation.

Conclusions. We have found that in denatured proteins, H_N–H_N NOEs are observed for regions with populations of structured conformations, as well as for unstructured random coil regions. This implies that qualitative observation of H_N–H_N NOEs alone is inadequate to define residual structure in denatured proteins. An approach that normalizes H_N–H_N NOEs as cross-peak-to-diagonal peak ratio was used to identify residual structure in urea-denatured barnase. Regions with residual structure, defined by R_cs, show a remarkable correlation with regions that fold early. This suggests that these regions are potential initiation sites for protein folding. The conformational preferences of the denatured state are sufficient to narrow the search for the native structure under conditions that favor folding and funnel the reaction to the major transition state. The R_c can be used to define residual structure in denatured proteins and, in combination with protein engineering results, may be able to identify initiation sites of folding.

K.-B.W. was supported by a scholarship from the Croucher Foundation, Hong Kong.

1. Dill, K. A., Bromberg, S., Yue, K. Z., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995) *Protein Sci.* **4,** 561–602.
2. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995) *Science* **267,** 1619–1620.
3. Karplus, M. & Sali, A. (1995) *Curr. Opin. Struct. Biol.* **5,** 58–73.
4. Shakhnovich, E. (1996) *Folding Design* **1,** R50–R54.
5. Shortle, D. (1993) *Curr. Opin. Struct. Biol.* **3,** 66–74.
6. Shortle, D. (1996) *Curr. Opin. Struct. Biol.* **6,** 24–30.
7. Frenkiel, T., Bauer, C., Carr, M. D., Birdsall, B. & Feeney, J. (1990) *J. Magn. Reson.* **90,** 420–425.
8. Kay, L. E., Keifer, P. & Saarinen, T. (1992) *J. Am. Chem. Soc.* **114,** 10663–10665.
9. Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M. & Fersht, A. R. (1995) *J. Mol. Biol.* **254,** 305–321.
10. Zhu, G. & Bax, A. (1992) *J. Magn. Reson.* **100,** 202–207.
11. Macura, S. & Ernst, R. R. (1980) *Mol. Physiol.* **41,** 95–117.
12. Esposito, G. & Pastore, A. (1988) *J. Magn. Reson.* **76,** 331–336.
13. Saulitis, J. & Liepins, E. (1990) *J. Magn. Reson.* **87,** 80–91.
14. Ramachandran, G. N. & Sassiekharan, V. (1968) *Adv. Protein Chem.* **28,** 283–437.
15. Wong, K.-B., Freund, S. M. V. & Fersht, A. R. (1996) *J. Mol. Biol.* **259,** 805–818.
16. Ernst, R. R., Bodenhausen, G. & Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Oxford Univ. Press, Oxford).
17. Fersht, A. R. (1993) *FEBS Lett.* **325,** 5–16.
18. Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M. & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9412–9416.
19. Fersht, A. R. (1994) *Curr. Opin. Struct. Biol.* **5,** 79–84.
20. Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10869–10873.