Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 3683–3686, April 1995
Biochemistry

Folding of a nascent polypeptide chain *in vitro*: Cooperative formation of structure in a protein module

(chymotrypsin inhibitor 2/protein fragments)

GONZALO DE PRAT GAY, JAVIER RUIZ-SANZ, JOSE L. NEIRA, LAURA S. ITZHAKI, 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, January 18, 1995

ABSTRACT **We have prepared a family of peptide fragments of the 64-residue chymotrypsin inhibitor 2, corresponding to its progressive elongation from the N terminus. The growing polypeptide chain has little tendency to form stable structure until it is largely synthesized, and what structures are formed are nonnative and lack, in particular, the native secondary structural elements of α-helix and β-sheet. These elements then develop as sufficient tertiary interactions are made in the nearly full-length chain. The growth of structure in the small module is highly cooperative and does not result from the hierarchical accretion of substructures.**

Proteins are synthesized from the N terminus *in vivo*. The information contained in the amino acid sequence encodes the three-dimensional structure of the biologically active protein (1) so that many proteins spontaneously fold to their correct conformations. Accessory proteins are found *in vivo*, however, which appear to prevent off-pathway reactions, such as aggregation (2–4), but the final folded structure of the protein is still dictated by the amino acid sequence. Certain accessory proteins interact with nascent incompletely folded polypeptides. The interactions may possibly be dictated by the structural characteristics of the growing chain as it emerges from the ribosome (5–7).

Most of our current detailed knowledge about protein folding is derived from studies *in vitro*, usually from the refolding of denatured mature polypeptides (8, 9). Studies *in vivo* on the development of structure in a growing polypeptide chain during biosynthesis are too difficult at the molecular level, although pioneering characterization has been attempted by using the binding of antibodies raised against the native structure (10). A parallel approach to the problem is to study the polypeptide chain as it is synthesized *in vitro* under controlled conditions in the absence of complicating factors. Such a study could answer the following questions. What happens to a polypeptide chain as it is synthesized from the N terminus *in vitro*? Do stable structures form early in synthesis? When does recognizable native secondary structure form? Do subsets of structures form and progressively assemble in a hierarchical manner?

Here we examine the development of structure in a small protein, whose refolding pathway *in vitro* has been described in great detail (11–13), by physically dissecting it into N-terminal fragments of increasing length. The protein chosen is the 64-residue chymotrypsin inhibitor 2 (CI-2) from barley seeds because it approaches the simplest system possible and may represent a minimal folding unit comparable to a single domain (module) in a larger protein: it is a very small monomer that consists of a single domain (module) without disulfide crosslinks that constrain the unfolded state, and it has no complications from the presence of *cis* peptidyl-prolyl bonds in its native state. It folds spontaneously in solution via a two-state mechanism, without the accumulation of folding intermediates, and with only one kinetically significant transition state (11, 12). Both crystal (14) and NMR (15) solution structures reveal a single α-helix (residues 12–24) and, from NMR, six strands of β-sheet (residues 3–5, 11–13, 28–34, 45–51, 55–58, and 60–64). There is a type III reverse turn at residues 5–8, a type II reverse turn at residues 8–11, and a type I reverse turn at residues 25–28. The reactive site loop is at residues 35–42, and residues 52–54 form a turn. The transition state for folding is like an expanded form of the folded structure in which most interactions are largely broken. The dominant structural feature in the transition state is the presence of a weakened, but largely formed, α-helix (13, 16). We have obtained a series of N-terminal peptides of CI-2 and have studied their conformation by NMR and circular dichroism in aqueous environment under physiologically compatible conditions of pH and temperature. The structures will be presented in detail elsewhere. Here, we outline sufficient structural information to answer several of the posed questions.

MATERIALS AND METHODS

Materials. Peptides CI-2-(1–5), -(1–13), and -(1–25) were prepared by using a solid-phase Applied Biosystems Synergy personal peptide synthesizer and purified by reverse-phase HPLC on a C_8 Dynamax 300A Rainin column, with a linear gradient of acetonitrile/water in 0.1% trifluoroacetic acid. After lyophilization, the peptides were dissolved in water, flash-frozen in liquid nitrogen, and kept at −20°C. The naturally occurring Met-40 (position 59 in original numbering) was mutated to Leu to remove the CNBr cleavage site and the mutant M59L CI-2 was used as template for subsequent mutations. Fragments CI-2-(1–28), -(1–50), -(1–53), and -(1–60) were obtained by CNBr cleavage of the mutants M59L/Q47M, M59L/F69M, M59L/K72M, and M59L/V79M and purified by HPLC as indicated for the small peptides. The fragment CI-2-(1–63) was obtained by introducing a stop codon at position 83 in wild-type CI-2 (L.S.I., unpublished data). All proteins were purified as described (17) except for CI-2-(1–63), which formed mainly inclusion bodies. The bacterial pellet was subjected to two rounds of sonication in 100 mM Tris·HCl (pH 7.9). The sonication pellet was dissolved in 6 M guanidinium chloride and dialyzed against 50 mM sodium acetate (pH 4.5); a vast precipitate was discarded and the soluble fraction containing >90% CI-2-(1–63) was dialyzed against water, concentrated by lyophilization, dissolved in 25 mM Tris·HCl, pH 7.9/0.2 M NaCl, and applied to a Superdex 75 (Pharmacia) gel filtration column. The peak was dialyzed against water and stored at −70°C after flash freezing in liquid nitrogen. Large fragments starting from residues 1–50 were

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: CI-2, chymotrypsin inhibitor 2; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; 1D, one dimensional.

*To whom reprint requests should be addressed.

treated with 6 M guanidinium chloride, dialyzed against 7.75 mM NaH_2PO_4/2.25 mM Na_2HPO_4, pH 6.3, and stored. Concentrations of fragments were determined from the extinction coefficient of model compounds (18). The homogeneity of all fragments was checked by HPLC and capillary electrophoresis, and their molecular masses were confirmed by electrospray mass spectroscopy.

Methods. Far-UV CD spectra of fragments (20 μM) were measured in 10 mM sodium phosphate (pH 6.3) at various temperatures. The details on CD experiments are described elsewhere (17). In the range of concentration studied (0–100 μM) at 25°C, the ellipticity $[\Theta]_{MRW}$ at 192 and 220 nm do not change, indicating the absence of aggregation at this temperature.

NMR spectra were recorded on a Bruker AMX-500 spectrometer equipped with an X32 computer. Sample concentration was typically 0.4–0.7 mM, in 20 mM sodium phosphate (pH 6.5), with 90% H_2O/10% ^2H_2O. 3,3,3-Trimethylsilypropionate was used as a reference for 0 ppm chemical shift. Fresh samples for NMR experiments were prepared by dissolving the lyophilized protein in 10% ^2H_2O/90% H_2O for experiments in water or 98% ^2H_2O/2% H_2O for exchange experiments, in both cases, in 20 mM sodium phosphate (pH 7.0). The temperature was 15.0 ± 0.1°C for all experiments. Two-dimensional homonuclear nuclear Overhauser effect spectroscopy (NOESY) spectra (19) and total correlation spectroscopy (TOCSY) spectra (20) were recorded at 500 MHz on a Bruker AMX-500. The experiments were acquired in all cases with 2048 real data points by using the TPPI method (21) with 256 $t1$ increments for NOESY spectra and 128 $t1$ increments in TOCSY experiments. Full details on the NMR experiments and assignment will be published elsewhere. The two-dimensional NMR spectra were processed by using BRUKER-UXNMR software on an X32 computer. Before Fourier transformation, the data matrix was zero-filled to get a final matrix size of 2 K × 1 K (1 K = 1×2^{10}). The chemical shifts were referenced relative to an external solution of 3,3,3-trimethylsilylpropionate.

RESULTS

Generation of Fragments. A family of nine peptide fragments of CI-2 (Fig. 1) was created by using chemical synthesis for small peptides and fragmentation of full-length structures for larger peptides; we introduced Met residues by protein engineering to be used as cleavage sites for CNBr (22). A discrete number of fragments was chosen, each of which contains an additional element of the regular secondary structural elements in the sequence. Wild-type CI-2 contains just one Met at position 40. We use a truncated form of CI-2 in which the first 19 amino acids that are disordered in both the crystal and solution structures have been deleted (14, 23). (This is the form that is the most studied by us, and its folding properties are identical to those of wild type.) We have renumbered the sequence of the truncated protein by denoting its first residue as number 1: Met-59 is thus Met-40 in this paper. Mutations were introduced that seemed likely to minimize structural changes. Fragments were obtained from the following mutants: M40L/Q28M, M40L/F50M, M50L/K43M, and M40L/V60. Glu-28 interacts with Asp-45 and Val-47 in wild-type CI-2, which are not present in the CI-2-(1–28) (13). In a similar manner, mutations of Phe-50 and Lys-53 to Met mainly affect interactions with residues between the cleavage site and the C terminus, and Val-60 also does not appear to have critical interactions. The stability and kinetics of folding of these mutants are similar to those of wild-type CI-2 (D. Otzen and A.R.F., unpublished data).

Structural Characterization. The fragments were first screened by CD spectroscopy, then by one-dimensional (1D) ^{1}H-NMR, and where appropriate, by multidimensional NMR methods. CI-2 is an α/β protein with 20% α-helix and 45% β-sheet. Its far-UV CD spectrum does not show distinct contributions from any particular secondary structure motif (17) but is comparable to that of several other α/β proteins (24): an ellipticity maximum at around 192 nm, a minimum at around 205 nm, a negative contribution at around 220 nm that is less pronounced for CI-2 than usual because of its relatively low content of α-helix, and a negative band at 233 nm that has been attributed to Trp-5 in its folded conformation (17). The one simple unequivocal assignment from CD studies is that for the α-helix. The signals from β-turns and β-sheet structures vary considerably from protein to protein and other complications arise involving aromatic residues and charged groups in asymmetric environments (24–27). For these reasons, we just focus here on significant overall qualitative and quantitative changes between the spectra of the fragments as the chain grows from the N terminus and the recovery of the full-length CI-2 spectrum, indicative of a mature folded polypeptide.

Peptides CI-2-(1–5) to -(1–28). CI-2-(1–5) and CI-2-(1–13) have a broad positive band in the far-UV CD spectra between 210 and 230 nm, which suggests a residual β-turn based on data from several proteins (28). This positive band decreases gradually with increasing temperature, suggesting the melting of weak structure. No evidence of structure in either peptide was found by NMR. CI-2-(1–25) and CI-2-(1–28) contain the amino acid sequence of the α-helix of native CI-2. The characteristic helix signal around 220 nm is not seen, and there is a minimum at 198 nm in the far-UV CD spectra, usually regarded as indicative of disordered structure. A helical signal is induced by the addition of trifluoroethanol but quantitative analysis of trifluoroethanol titration experiments (29) predicts <5% helix in the absence of cosolvent. ^{1}H-NMR data indicate that fragments CI-2-(1–5) to -(1–28) are largely disordered in aqueous solution with weak helical signals being induced by trifluoroethanol. The spectral data thus suggest largely disordered conformations with some weak local interactions for the four small fragments.

Peptide CI-2-(1–40). Initial CD and 1D ^{1}H-NMR studies indicated a predominantly disordered structure (17). More detailed studies using heteronuclear NMR showed the presence of nonnative hydrophobic clustering stabilized by local interactions (30). This fragment buries a hydrophobic region of 12 residues that is highly insoluble as an isolated peptide: this is the most significant hydrophobic region in the whole CI-2 molecule as shown by a hydrophilicity plot (30). We obtained 1D ^{1}H-NMR spectra of fragments from CI-2-(1–40) to full-length CI-2, at pH 6.5 and 25°C. The lack of chemical shift dispersion for CI-2-(1–40) under these conditions shows that it lacks any significant tertiary fold, as is also the case at pH 4.5 and 5°C (17). CI-2-(1–40) thus has nonnative structure with locally driven burial of hydrophobic surfaces.

FIG. 1. Amino acid sequence, secondary structural elements (13), and positions of cleavage to generate N-terminal fragments of CI-2. Underlines indicate β-sheet and dashed underlines indicate α-helix. Arrows indicate the position of cleavage for each fragment.

Peptide CI-2-(1–50). The ^{1}H-NMR spectrum of CI-2-(1–50) shows little chemical shift dispersion in the amide region, suggesting the absence of stable tertiary interactions.

Peptide CI-2-(1–53). The CD spectrum of this fragment shows increases in the maximum at 190 nm and the minimum at around 205 nm, with respect to CI-2-(1–50), moving closer to the spectrum of the full-length protein. In contrast to the effects on the smaller fragments, there are practically no changes in the CD spectrum on increasing the temperature. The 1D ^{1}H-NMR spectrum of this fragment has a large spread of chemical shifts in the amide region, with Trp-5 shifted to 11.6 ppm, Glu-7 shifted to 10.8 ppm, and a distinctive peak at 5.7 ppm, presumably corresponding to Phe-50 in the β-sheet region, all indicating the presence of folded structure. Further evidence for folding is provided by the appearance of methyl peaks of Ile-20 at 0.1 and −0.1 ppm, a strong indication of native-like tertiary fold (15); these are absent in CI-2-(1–40) and in CI-2-(1–50), but they are clearly present in CI-2-(1–53).

The combination of the dispersion in the 1D ^{1}H-NMR spectra and the comparison with full-length CI-2 of both the shape of the CD spectrum of CI-2-(1–53) and its temperature dependence clearly indicates it has tertiary interactions and it is rather compact.

Peptide CI-2-(1–60). The elongation from CI-2-(1–53) to -(1–60) leads to a small increase of ellipticity at 190 nm and no change at 205 nm in the CD spectra; the Trp-5 band at 233 nm is absent at 25°C. The temperature dependence of the spectrum of fragment CI-2-(1–60) shows a gradual decrease of ellipticity in the 220-nm region, reaching a broad minimum around 208–215 nm at 80°C, possibly caused by aggregation. There is also a large chemical shift dispersion in the amide signals in the 1D ^{1}H-NMR spectrum of CI-2-(1–60). At high field, the two Ile-20 peaks, which are distinctive of folded structure, appear at 0.15 and −0.15 ppm, plus a third peak at −0.95, corresponding to γ-CH_3 in native CI-2 (data not shown).

An almost complete assignment of the spectra was obtained to define the structure of this peptide more thoroughly. An extensive study of the chemical shifts, nuclear Overhauser effect (NOE) cross peaks, and amide proton protection rates provides enough evidence to describe stable structural elements. The unique α-helix of CI-2 (residues 12–24) is completely formed, as the characteristic NOEs and solvent protection show (Fig. 2). Residues 51–57 form a native-like β-hairpin, as indicated by the protection of the amide protons of residues Phe-50, Val-51, Ile-57, Ala-58, and Gln-59, and a weak NOE cross peak between Hα protons of residues 51 and 57. In addition, we observe slow exchange rates for amide protons of Leu-32 and Val-34, which interact with the aromatic moiety of Phe-50. This triad forms the so-called hydrophobic minicore of folded CI-2 (13). Fig. 3 is a ribbon representation of full-length CI-2, indicating the elements in the fragment that have native-like structure. Other regions of the structure are disordered but show native-like chemical shifts, which suggest that these regions fluctuate between folded and unfolded conformations. Further, Trp-5 is probably in a different environment from that in CI-2-(1–53) or native CI-2, as seen from the absence of the CD band at 233 nm and NOE evidence for different conformations of Trp-5. Fragment CI-2-(1–60) is partially folded, with the unique α-helix and a β-hairpin well defined.

Fragment CI-2-(1–63) is a 1-residue deletion of the full-length protein. It has compact native-like structure as indicated by its far-UV CD spectrum and 1D ^{1}H-NMR. However, it is much less stable to denaturation by temperature and denaturing agents (unpublished results). Full NMR assignments for this fragment have been obtained.

DISCUSSION

Three stages are observed in the folding of the nascent chain of CI-2 *in vitro*. The early stage produces small peptides [CI-2-(1–5) to -(1–28)] that are highly disordered. Changes in their CD spectra with temperature suggest, however, the presence of marginally stable structures or equilibrium mixtures that are not observable by NMR. A second stage involves the formation of compact structures, as seen on elongation from residues 40 to 53. Fragment CI-2-(1–40) was previously studied by heteronuclear NMR and was shown to contain local hydrophobic clusters in water but is nonnative (30). A largely hydrophobic sequence from residues 27 to 40 is buried in this peptide. Hydrophobic clustering, albeit nonnative, is the route

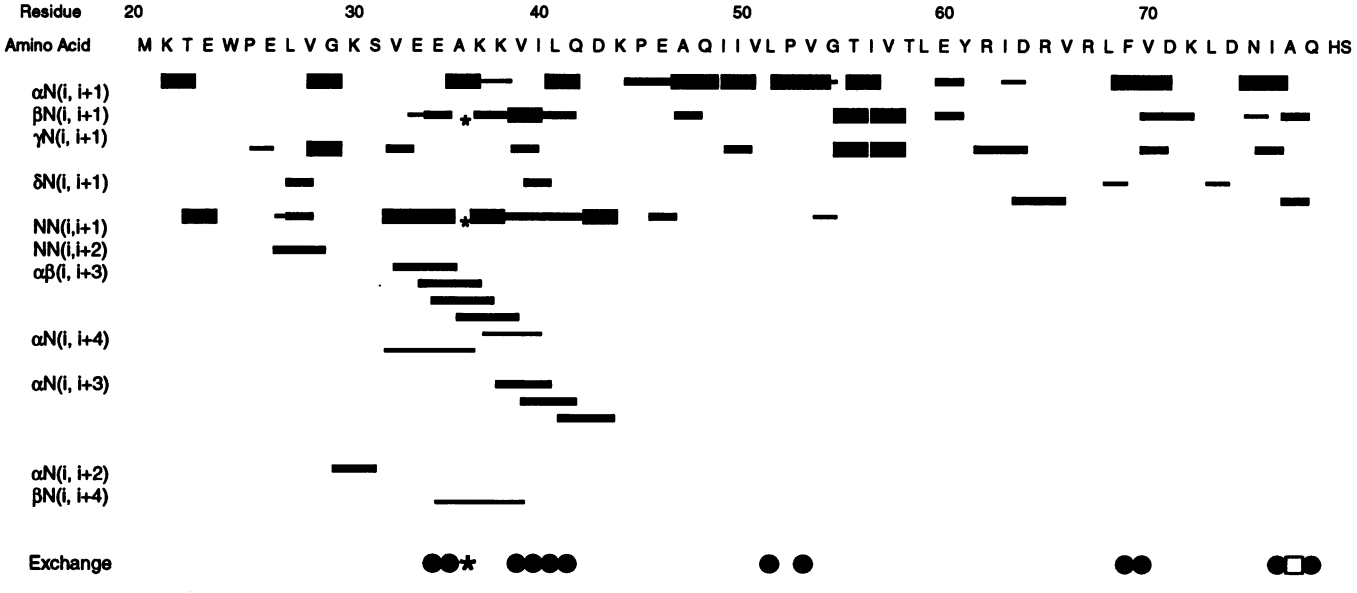

FIG. 2. Pattern of NOE cross peaks found in NOESY spectra of the ^{1}H-NMR spectrum of a 2 mM sample of fragment CI-2-(1–60) at 288 K, in 20 mM sodium phosphate (pH 7.0). The intensities of the sequential and intraresidue NOEs were classified into three categories (strong, medium, and weak) based on the visual inspection of the cross peak and indicated by the height of the bars. Solid circles represent positions that are protected against hydrogen/deuterium exchange; asterisks mark signals that could not be unambiguously assigned because of overlapping signals. The open square indicates a protected proton that can be observed only in a 1D spectrum.

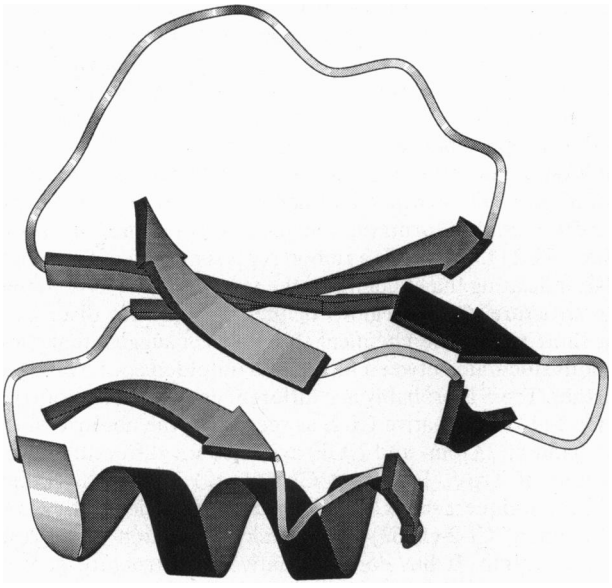

FIG. 3. Ribbon diagram representation of CI-2 [using the program MOLSCRIPT (37)] in which the regions in fragment CI-2-(1–60) that have native secondary structure are solid.

to the most stable conformation in solution for the polypeptide chain at this stage.

The tertiary stage of folding starts from fragment CI-2-(1–53), which appears to be compactly folded according to the CD spectrum and its temperature dependence and 1D NMR. The dramatic change between CI-2-(1–50) and -(1–53) occurs in the presence of two key residues in the stability of the native-like fold of CI-2: homocysteine-50 reverts to the native Phe-50 and Val-51 is added. These are the two last residues in β-strand 4 and make critical interactions linking all the major structural elements, namely, the α-helix and β-strands 3–5, and stabilizing at the same time the main hydrophobic core and minicore (15, 18, 19, 32). Elongation to residue 60 yields a folded fragment in which the α-helix is fully formed as is the β-hairpin between residues 51 and 57. The rest of the structure is fluctuating and there appears to be some nonnative interactions.

There seems to be little drive for the formation of stable structures in the initial stages of synthesis from the N terminus. This will avoid any kinetic traps of forming too stable intermediates. Further, the lack of structure will allow the binding of chaperones should the chain become exposed. A major element of regular secondary structure, the single α-helix, does not form significantly until sufficient tertiary interactions are present to stabilize it. Although kinetic pathways of folding of intact proteins do not need to bear any resemblance to the series of equilibrium intermediate states during elongation, there is a common theme with CI-2: formation of the α-helix in the refolding of intact CI-2 occurs in parallel with the formation of tertiary interactions, and formation of stable α-helix in the stepwise synthesis of CI-2 occurs when sufficient tertiary interactions have accumulated. Will these observations be general? The very small CI-2 is probably representative of an independent folding unit or domain of larger proteins. Barnase is a more complex protein, consisting of different modules (33), and has an important α-helix close to its N-terminal end (residues 6–18). The helix in N-terminal fragments of barnase is also largely disordered (34), and isolated N-terminal helices in general do not form highly stable structures (31).

It has been suggested that the folding behavior of nascent polypeptides is hierarchical, with elements of structure building up sequentially (35, 36). This may well happen for more complex proteins in which their constituent modules or domains may form sequentially. But for the single module of structure of CI-2, the folding is not hierarchical and the onset of native-like structure is concerted and highly cooperative.

1. Anfinsen, C. B. (1973) *Science* **181,** 223–230.
2. Ellis, R. J. (1987) *Nature (London)* **328,** 378–379.
3. Gething, M. J. & Sambrook, J. (1992) *Nature (London)* **355,** 33–45.
4. Freedman, R. B. (1992) in *Protein Folding*, ed. Creighton, T. E. (Freeman, Company, New York), pp. 455–539.
5. Langer, T., Lu, C., Echols, H., Hayer, M. K. & Hartl, F. U. (1992) *Nature (London)* **356,** 683–689.
6. Hartl, F. U., Hlodan, R. & Langer, T. (1994) *Trends Biochem. Sci.* **19,** 20–25.
7. Frydman, J., Nimmesgern, E., Ohtsuka, K. & Hartl, F. U. (1994) *Nature (London)* **370,** 111–117.
8. Fersht, A. R. (1993) *FEBS Lett.* **325,** 5–16.
9. Creighton, T. E. (1992) in *Protein Folding*, ed. Creighton, T. E. (Freeman, New York).
10. Friguet, B., Djavadiohaniance, L., King, J. & Goldberg, M. E. (1994) *J. Biol. Chem.* **269,** 15945–15949.
11. Jackson, S. E. & Fersht, A. R. (1991) *Biochemistry* **30,** 10428–10435.
12. Jackson, S. E. & Fersht, A. R. (1991) *Biochemistry* **30,** 10436–10443.
13. Otzen, D., Itzhaki, L., elMasry, N., Jackson, S. E. & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10422–10425.
14. McPhalen, C. A. & James, M. N. G. (1987) *Biochemistry* **26,** 261–269.
15. Ludvigsen, S., Shen, H., Kjær, M., Madsen, J. C. & Poulsen, F. M. (1991) *J. Mol. Biol.* **222,** 621–635.
16. Li, A. & Daggett, V. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10430–10434.
17. Prat Gay, G. de & Fersht, A. R. (1994) *Biochemistry* **33,** 7957–7963.
18. Gill, S. C. & von Hippel, P. H. (1989) *Anal. Biochem.* **182,** 319–326.
19. Jenner, J., Meier, B. H., Bachman, P. & Ernst, R. R. (1979) *J. Chem. Phys.* **71,** 4546–4553.
20. Braunsweiler, L. & Ernst, R. R. (1983) *J. Magn. Reson.* **53,** 521–528.
21. Marion, D. & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113,** 967–974.
22. Sancho, J. & Fersht, A. R. (1992) *J. Mol. Biol.* **224,** 741–747.
23. Jackson, S. E., elMasry, N. & Fersht, A. R. (1993) *Biochemistry* **32,** 11270–11278.
24. Brahms, S. & Brahms, J. (1980) *J. Mol. Biol.* **138,** 149–178.
25. Sears, D. W. & Beychok, S. (1973) in *Physical Principles and Techniques of Protein Chemistry*, ed. Leach, S. J. (Academic, New York), pp. 445–475.
26. Woody, R. W. (1985) in *Conformation in Biology and Drug Design*, ed. Hruby, V. J. (Academic, New York), pp. 15–32.
27. Johnson, W. C., Jr. (1988) *Annu. Rev. Biophys. Chem.* **17,** 145–166.
28. Chang, C. T., Wu, C. S. & Yang, J. T. (1978) *Anal. Biochem.* **91,** 13–31.
29. Jasanoff, A. & Fersht, A. R. (1994) *Biochemistry* **33,** 2129–2135.
30. Prat Gay, G. de, Ruiz-Sanz, J., Davis, B. & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10943–10946.
31. Dyson, H. J. & Wright, P. E. (1993) *Curr. Opin. Struct. Biol.* **3,** 60–64.
32. Harpaz, Y., elMasry, N., Fersht, A. R. & Henrick, K. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 311–315.
33. Yanagawa, H., Yoshida, K., Torigoe, C., Park, J.-S., Sato, K., Shirai, T. & Go, M. (1993) *J. Biol. Chem.* **268,** 5861–5865.
34. Sancho, J., Neira, J. L. & Fersht, A. R. (1992) *J. Mol. Biol.* **224,** 749–758.
35. Jaenicke, R. (1991) *Biochemistry* **30,** 3147–3161.
36. Tsou, C. L. (1988) *Biochemistry* **27,** 1809–1812.
37. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.