Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10943–10946, November 1994
Biochemistry

The structure of the transition state for the association of two fragments of the barley chymotrypsin inhibitor 2 to generate native-like protein: Implications for mechanisms of protein folding

(protein engineering/protein fragments)

GONZALO DE PRAT GAY, JAVIER RUIZ-SANZ, BEN DAVIS, 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 22, 1994

ABSTRACT **Possible early events in protein folding may be studied by dissecting proteins into complementary fragments. Two fragments of chymotrypsin inhibitor 2 [CI2-(20–59) and CI2-(60–83)] associate to form a native-like structure in a second-order reaction that combines collision and rearrangement. The transition state of the reaction, analyzed by the protein engineering method on 17 mutants, is remarkably similar to that for the folding of the intact protein—a structure that resembles an expanded version of the folded structure with most interactions significantly weakened. The exception is that the N-terminal region of the single α-helix (the N-capping box) is completely formed in the transition state for association of the fragments, whereas it is reasonably well formed for the intact protein. Preliminary evidence on the structures of the individual fragments indicates that both are mainly nonnative, lacking native secondary structure and having regions of nonnative buried hydrophobic clusters. The association reaction does not result from the collision of a subpopulation of two fully native-like fragments but involves a considerable rearrangement of structure.**

The pathway of protein folding will be solved when the structures of all the stable, metastable, and transition states in the reaction are characterized (1). The currently most obscure and controversial area is that of the earliest events because they are the most inaccessible to experiment. Progress is being made here in two directions: the direct solution of unfolded structures by NMR (2–4) and the structure of individual fragments of proteins (5–8). The study of fragments has two attractions, especially for those systems in which native-like complexes may be reconstituted from the individual fragments. First, information is given about the more local events. Second, it is easier to obtain detailed information about the structure of small fragments in solution by NMR and other spectroscopic methods.

The truncated form of chymotrypsin inhibitor 2 (CI2, residues 20–83) provides an excellent system for studying protein folding (Fig. 1). It is a single module protein that folds and unfolds by a two-step mechanism (10–13). The transition state for folding and unfolding has been characterized in detail by the protein engineering method (14). The method is based on a parameter, ϕ_F, that is an index of structure formation. ϕ_F is derived from a comparison of kinetic and equilibrium measurements of folding of wild-type and mutant proteins and is defined as the ratio of the changes of activation energy and equilibrium free energy of folding on mutation. $\phi_F = 0$ means that the structure containing the mutation is unformed in the transition state, $\phi_F = 1$ means that the structure is most likely to be fully formed, and fractional values mean that the structure is present in a weakened form (15, 16). The transition state is like an expanded form of the folded protein where most of the elements of secondary and tertiary structure are weakened by >50% (14). The single α-helix of CI2 is the best formed structure, especially at its N terminus. Independent computational studies give a remarkably similar description of the transition state (17). CI2 has been cleaved into two fragments, CI2-(20–59) and CI2-(60–83), by use of CNBr at the single methionine, Met-59 (Fig. 1) (18). Peptide CI2-(20–59) contains the α-helix (residues 31–43). CI2-(60–83) contains three of the major strands of the β-sheet. The two fragments associate to form a noncovalent, but fully folded, complex similar to uncleaved CI2 [and now shown by NMR to be virtually identical, apart from immediately around the cleaved bond and the final turn of the α-helix (B.D., unpublished data)]. Furthermore, they associate by simple second-order kinetics (19). The observed second-order rate constant for the association contains contributions from a true bimolecular association reaction and the contributions from a protein folding step. We now analyze the transition state for the major phase of association and folding by the protein engineering method using mutant peptides generated by the action of CNBr on the mutants used for characterizing the transition state for folding of the intact protein. (The minor phases from the isomerization of cis → trans peptidyl prolyl bonds are not considered here.) We also report preliminary observations from NMR and circular dichroism spectroscopy on the structure of the isolated fragments of the wild-type truncated CI2 in water that are sufficient to show that the structures are predominantly nonnative.

MATERIALS AND METHODS

The parent proteins have been described (12, 14, 20). The methods of preparation and characterization of the fragments and measurement of the kinetics and equilibria of their association have been described in detail (18, 19). Kinetics and equilibria were measured at 25°C in 10 mM phosphate buffer (pH 6.3). The titration with 2,2,2-trifluoroethanol (TFE) of the formation of helix in CI2-(20–59) was monitored by the change in ellipticity at 222 nm as described by Jasanoff and Fersht (21). All rate constants for association were measured by monitoring both circular dichroism and fluorescence. Both techniques were found to give, within experimental error, the same results, as previously demonstrated for the wild-type fragments. Second-order kinetics was found for generation of native-like structure for all mutants.

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: CI2, chymotrypsin inhibitor 2; TFE, 2,2,2-trifluoroethanol; TOCSY, total correlated spectroscopy; HMQC, heteronuclear multiple-quantum correlation; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy.

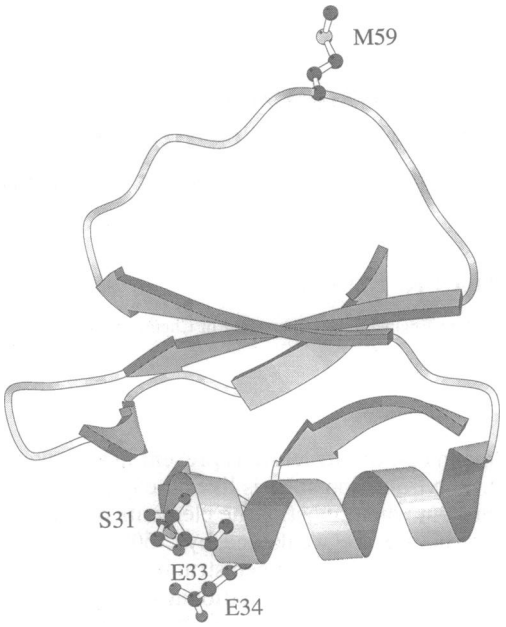

FIG. 1. Diagram of the barley CI2 inhibitor showing its secondary structure, the unique methionine at which CNBr cleavage is targeted, and the residues constituting the N-capping box [prepared with the program MOLSCRIPT (9)].

NMR Experiments. Uniformly ^{15}N-labeled CI2 was prepared, purified, and cleaved with CNBr (18, 19) to produce uniformly ^{15}N-labeled CI2-(20–59) and CI2-(60–83). Samples containing 1–1.5 mM protein were dissolved in 50 mM acetate-d_3 buffer (pH 4.6) in either 90% H_2O/10% ^2H_2O or 100% ^2H_2O. All spectra were acquired at 5°C on a Bruker AMX500 spectrometer. Three-dimensional {^{15}N} ^1H total correlated spectroscopy–heteronuclear multiple-quantum correlation (TOCSY-HMQC), {^{15}N} ^1H nuclear Overhauser effect spectroscopy (NOESY)-HMQC (22), two-dimensional {^{15}N} ^1H heteronuclear single-quantum correlation (23) and {^{15}N} ^1H TOCSY-HMQC, as well as homonuclear double quantum filtered correlated spectroscopy, NOESY, and TOCSY were acquired, and the details will be described elsewhere. Spectra were processed with the FELIX software package (Biosym).

RESULTS

Association of Mutant Fragments: Probing the Folding/Association Pathway by the Protein Engineering Method. We can analyze the transition state for formation of the complex by the protein engineering method. For this purpose, we introduce a modified ϕ value. Previously, ϕ_F was related to the transition state and unfolded and folded structures of an intact protein by Eq. 1,

$$\phi_F = \Delta\Delta G_{\ddagger-U}/\Delta\Delta G_{F-U}, \quad [1]$$

where $\Delta\Delta G_{F-U}$ is the change in free energy of folding on mutation and $\Delta\Delta G_{\ddagger-U}$ is the change in energy of the relevant transition state for folding (where F = folded state, U = unfolded state and ‡ = transition state) (15, 16). For fragments, we have to substitute for $\Delta\Delta G_{\ddagger-U}$ a term that represents the change on mutation of the free energy of the transition state of the association reaction relative to an unfolded state, which is the free fragments in water. This may be measured from the values of the second-order rate constants for association, k_{on}, for wild-type and mutant fragments and Eq. 2,

$$\Delta\Delta G_{\ddagger-U(FRAG)} = RT\ln(k_{on}^{wt}/k_{on}^{mut}). \quad [2]$$

The equivalent term for $\Delta\Delta G_{F-U}$ is obtained from the dissociation constants of the mutant and wild-type complexes:

$$\Delta\Delta G_{F-U(FRAG)} = -RT\ln(K_d^{wt}/K_d^{mut}), \quad [3]$$

where K_d^{wt} is the dissociation constant of the wild-type complex, etc. The value of ϕ for association, ϕ_A, is given by

$$\phi_A = [\ln(k_{on}^{wt}/k_{on}^{mut})]/[\ln(K_d^{mut}/K_d^{wt})]. \quad [4]$$

There is a close similarity between the values of ϕ_A for formation of the complex and ϕ_F for the folding of intact CI2 (Table 1). A plot of ϕ_A versus ϕ_F has a slope of 1.1 ± 0.2, an intercept of −0.08 ± 0.07, and a correlation coefficient of 0.86. The most significant difference is at the N terminus of

Table 1. ϕ values for formation of transition state for association of fragments

Mutation(s)	$\Delta\Delta G_{F-U}$,* kcal·mol^{-1} (intact)	$\Delta\Delta G_{F-U}$,† kcal·mol^{-1} (fragments)	Location	Alters interactions with	ϕ_F§ (intact)	ϕ_A¶ (fragments)
T22A	0.85	0.43	β-strand 1	V82	0.26	0.04
S31A	0.89	0.98	N-cap of helix	Helix	0.42	0.90
E33N	0.70	0.43	N-cap, +2	Helix	0.75	1.13
E33A/E34A	0.83	0.95	N-cap, +2, +3	Helix	0.70	0.65
S31A/E33A/E34A	1.67	1.39	N-cap, +2, +3	Helix	0.41	1.01
K36A	0.48	0.86	Helix	Helix, I48	0.32	0.08
K37G	0.97	1.4	Helix	Helix	0.73	0.47
V38A	0.48	0.49	Helix	Helix, β_1, core	−0.26	−0.14
I39V	1.30	1.35	Helix	Core, β_1, β_4, β_6	0.31	0.40
K43A	0.65	0.58	C-cap of helix	24, 39, 42, 44–46, 63–66, 82	−0.29	−0.64
I48V	1.11	0.49	β-strand 3	36, 39, 40, 46, 50, 66, 68	0.15	0.52
I49A	2.12	1.12	β-strand 3	47, 48, 50, 51, 65	0.31	0.47
L51A	2.37	1.13	β-strand 3	49, 50, 55, 57, 67, 69	0.22	0.17
V53T	1.03	0.82	β-strand 2	52, 54, 69, 71	0.27	0.00
V53A	0.63	0.00		78	0.11	0.00
F69L	2.11	1.17	β-strand 4	51–53, 55, 57, 67, 68, 70, 78	0.29	0.25
V70A	1.95	1.56	β-strand 4	Helix, β_3, β_4, β_5	0.26	0.34

*Change in energy of intact protein on mutation (unpublished data).
†Change in free energy of association of fragments.
§ϕ value for folding of intact protein.
¶ϕ value for association of fragments.

the α-helix. This is the best formed region in the transition state for folding of the intact protein and is even more formed in the assembly of fragments, with values of ϕ_A tending to 1.0.

Evidence for Nonnative Structure in CI2-(20–59) and CI2-(60–83). Fragments CI2-(20–59) and CI2-(60–83) showed no evidence of compact tertiary structure from fluorescence, circular dichroism, and one-dimensional NMR. We have produced ^{15}N isotopically labeled fragments and fully assigned their two-dimensional NMR spectra; the sequence-specific ^1H and ^{15}N assignments of CI2-(20–59) and CI2-(60–83) were obtained by standard heteronuclear experiments (22, 23). Deviations from random coil chemical shifts (24) and unambiguous NOE data were used to identify possible regions of residual structure in the isolated fragments. Deviations from random coil chemical shift ($\Delta\delta_{ran}$) of ≥0.1 ppm for a nonlabile proton and ≥0.3 ppm for a labile proton were taken to be significant.

CI2-(20–59). A hydrophobicity plot of CI2-(20–59) reveals a stretch of hydrophobic residues from 47 to 59 (Fig. 2). Residues 22–25, 38–44, and 49–56 form stretches of residues with nonrandom coil shifts. Although these values of $\Delta\delta_{ran}$ are relatively small (≤0.26 ppm for nonlabile atoms and ≤0.45 ppm for labile atoms), the deviations outside these regions are in general smaller, ≤0.10 and ≤0.20 ppm, respectively. It seems likely that the side chains in the peptide form a hydrophobic cluster.

NOEs from a three-dimensional NOESY-HMQC experiment and a two-dimensional homonuclear NOESY experiment were used to identify regions of residual structure. NOEs involving the aromatic ring of W24 were the only side-chain–side-chain NOEs to be assigned unambiguously. The observed NOEs are primarily clustered around the regions 21–27, 38–45, and 48–56. These regions correspond well to those having nonnative random coil shifts, as discussed above. The NOEs are primarily short-range HN–side-chain interactions; few HN backbone NOEs are observed. No classical elements of secondary structure are defined by these NOEs. NOEs are seen between the W24 indole ring protons and the H_γ of T22; T22H_γ is shifted 0.23 ppm from its random coil value. This is the only unambiguously assigned side-chain–methyl interaction and is nonnative. The side chain of W24 makes further interactions with residues 22–23 and 25–27.

There is a small cluster of NOEs, including two weak $d_{NN}(i, i + 1)$ NOEs, in the region 32–36. The H^α resonances in this region show a small negative deviation from random coil ($\Delta\delta_{ran}$ ranges between −0.05 and −0.16 ppm). This may be a tentative indication of the formation of either a nascent turn or a small population of helix-containing molecules. Quantitative TFE titration (21) of the formation of the helix extrapolates to a helical content in water of 3% at 5°C, rising to 15% at 25°C. This suggests that the NMR data are consistent with the nascent helix (26).

CI2-(60–83). Significant deviations from random coil values are observed between residues 62 and 71. The majority of these deviations are ≤0.2 ppm, lower than those observed in CI2-(20–59). Residues 72–83 have chemical shifts close to those expected for random coil. Preliminary NOE evidence suggests also the formation of a nonnative hydrophobic cluster in this region.

DISCUSSION

There are two key observations from this study. The first is that the transition state for the formation of native-like structure from two fragments of the CI2 inhibitor is remarkably similar to that for the folding of denatured intact CI2 inhibitor. The values of ϕ_F for the folding of intact protein are closely correlated to the equivalent ϕ_A values calculated from the association kinetics. The overall structure of the transi-

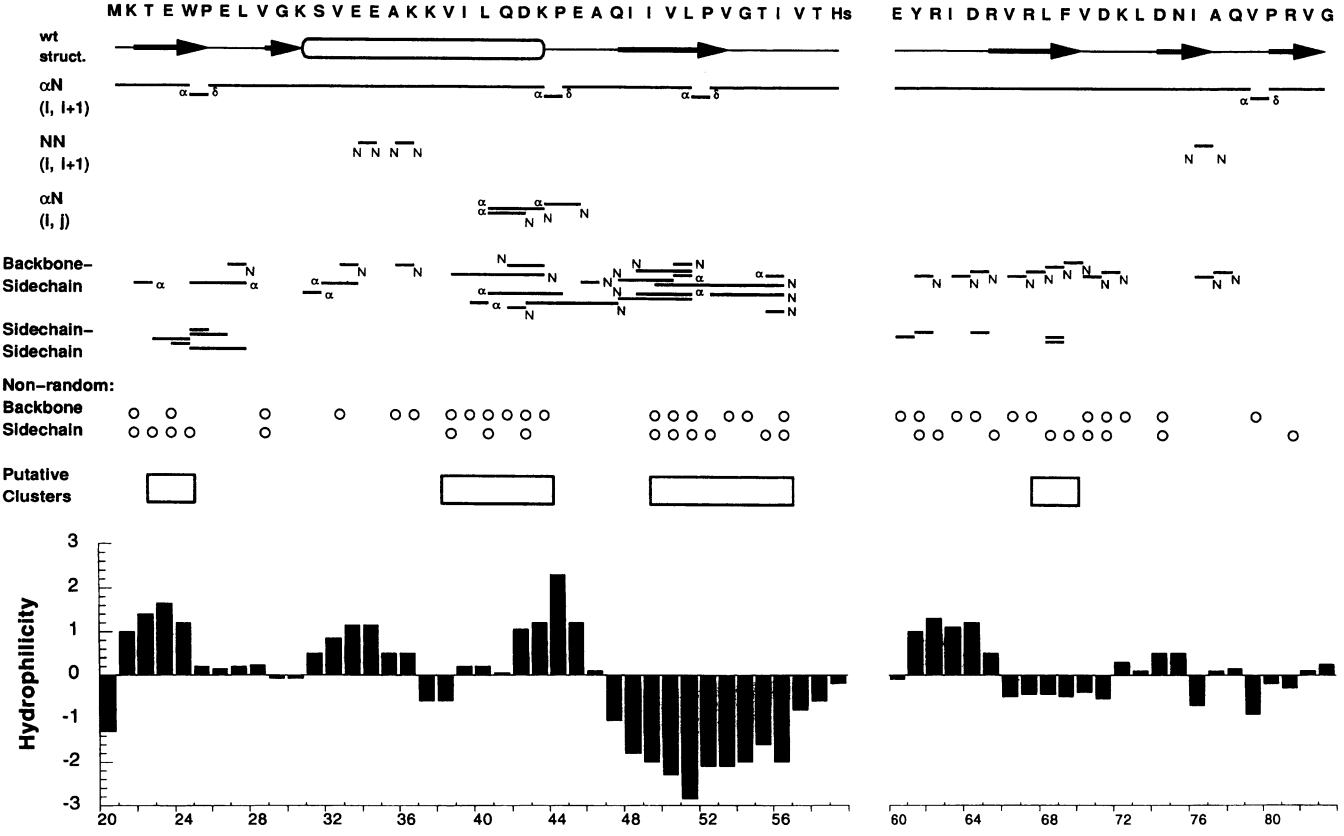

FIG. 2. Sequence of CI2, indicating the secondary structure (arrows, strands of sheet; cylinder, helix), hydrophilicity [determined by the method of Kyte and Doolittle (25)], and the NOE connectivities.

tion state for the association of fragments is thus like that recently described for the folding of the intact protein (14); the structure of the transition state is like an expanded version of the folded structure. The most significant difference concerns the first four residues of the α-helix, 31–34. These constitute what has been described as an N-capping box (27) and has been shown to contribute significantly to the stability of the helix (20). These residues are the most ordered in the transition state for the folding of the intact protein (14) (Table 1). In the transition state for the association/folding reaction for formation of the complex, however, the ϕ_A values are close to 1.0. The second observation is that the preliminary studies on the structures of the two fragments indicate that they have largely nonnative structure with regions of nonnative hydrophobic clusters. In particular, the helical component is very low, although possibly marginally detectable. This leads to a fundamental question of how the two fragments associate.

How Does Docking of Two Fragments Occur? There are two extreme mechanisms for the association of fragments that are predominantly nonnative to give native-like folded structures: (*i*) association of nonnative structures followed by rearrangement and (*ii*) rearrangement of both to having native structure followed by association (association results from small subpopulations of correctly formed conformations, which are in equilibrium with the predominant nonnative structures, binding to each other).

Studies on the ground states alone will not answer these questions, but clues come from the kinetics, especially from the ϕ values as follows. Suppose there is a small proportion of native structure, F_N, in a fragment of total concentration $[F]_0$, that is largely nonnative, F_{NN}. When $[F_{NN}] \gg [F_N]$, $[F_N] = K[F]_0$, where K is the equilibrium constant between the F_N and F_{NN}. Suppose the native conformations of the two fragments, F_i and F_j, associate with a rate constant k_{ij}. Then the observed rate constant for association k_{obs}, is given by $k_{obs} = k_{ij}K_i[F_i]_0K_j[F_j]_0$. If a mutation is made in either fragment such that it lowers the stability of the fragment by $\Delta\Delta G$ but is not directly involved in the interaction between the two fragments, then the equilibrium constant K will change by $\exp(-\Delta\Delta G/RT)$. The value of k_{obs} will thus also change by $\exp(-\Delta\Delta G/RT)$, and so ϕ will be 1.0. Most of the ϕ values in Table 1 are closer to 0 than to 1.0 and so the elements of structure are not fully formed in the transition state. In general, therefore, the fragments that associate cannot be fully preformed, thus ruling out mechanism *ii*. The exception is the N-capping box, residues 31–34. The ϕ values are close to 1 for this region and there is evidence for a small, or nascent (26), proportion of helical content from NMR and TFE titration. This is consistent with two possible mechanisms; either folding is initiated from a small subpopulation of the ground state that has the N-capping box fully formed (but the other regions are not so formed) or folding occurs by the association of two nonnative fragments and the N-cap simply rearranges first in the transition state. In both cases, the other regions of structure examined are clearly in the process of rearranging in the transition state.

Implications for the Mechanism of Protein Folding. The general similarity of the transition state for association of fragments with that for folding suggests that the similar mechanisms are involved in both processes. The higher values of ϕ for the N-cap of the helix in the transition state of the intact protein need not be a significant difference since it has been shown that the transition state for protein folding can move with destabilization of the folded protein according to the Hammond postulate (28). The data for both do not fit a model in which small elements of secondary structure are fully formed in the denatured state. Instead, the secondary structure is consolidated as tertiary structure is formed, and tertiary and secondary structure are formed in parallel. That is, much of the formation of structure is concerted rather than stepwise. Although hydrophobic clusters are initially formed, they are nonnative. The protein then has to rearrange from these. The hydrophobic clusters serve to bury local hydrophobic surfaces and restrict the conformational space available to the unfolded state of the protein.

1. Fersht, A. R. (1993) *FEBS Lett.* **325**, 5–16.
2. Neri, D., Billeter, M., Wider, G. & Wuthrich, K. (1992) *Science* **257**, 1559–1563.
3. Logan, T. M., Thériault, Y. & Fesik, S. W. (1994) *J. Mol. Biol.* **236**, 637–648.
4. Arcus, V. L., Vuilleumier, S., Freund, S. V., Bycroft, M. & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 9412–9416.
5. Wright, P. E., Dyson, H. J. & Lerner, R. A. (1988) *Biochemistry* **27**, 7167–7175.
6. Shin, H. C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J. & Wright, P. E. (1993) *Biochemistry* **32**, 6356–6364.
7. Kippen, A. D., Sancho, J. & Fersht, A. R. (1994) *Biochemistry* **33**, 3778–3786.
8. Sancho, J. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 741–747.
9. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950.
10. Jackson, S. E. & Fersht, A. R. (1991) *Biochemistry* **30**, 10428–10435.
11. Jackson, S. E. & Fersht, A. R. (1991) *Biochemistry* **30**, 10436–10443.
12. Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M. & Fersht, A. R. (1993) *Biochemistry* **32**, 11259–11269.
13. Jackson, S. E., elMasry, N. & Fersht, A. R. (1993) *Biochemistry* **32**, 11270–11278.
14. Otzen, D., Itzhaki, L. S., elMasry, N., Jackson, S. E. & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, in press.
15. Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. (1989) *Nature (London)* **342**, 122–126.
16. Fersht, A. R., Matouschek, A. & Serrano, L. (1992) *J. Mol. Biol.* **224**, 771–782.
17. Li, A. & Daggett, V. (1994) *Proc. Natl. Acad. Sci. USA* **91**, in press.
18. Prat Gay, G. d. & Fersht, A. R. (1994) *Biochemistry* **33**, 7957–7963.
19. Prat Gay, G. d., Ruiz-Sanz, J. & Fersht, A. R. (1994) *Biochemistry* **33**, 7964–7970.
20. elMasry, N. F. & Fersht, A. R. (1994) *Protein Eng.* **7**, 777–782.
21. Jasanoff, A. & Fersht, A. R. (1994) *Biochemistry* **33**, 2129–2135.
22. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989) *Biochemistry* **28**, 6150–6156.
23. Bax, A., Ikura, M., Kay, L. E., Torchia, D. A. & Tschudin, R. (1990) *J. Magn. Res.* **86**, 304–318.
24. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
25. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
26. Dyson, H. J., Houghten, R. A., Wison, I. A., Wright, P. E. & Lerner, R. E. (1988) *J. Mol. Biol.* **201**, 201–207.
27. Harper, E. T. & Rose, G. D. (1993) *Biochemistry* **32**, 7605–7609.
28. Matouschek, A. & Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7814–7818.