PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Mapping the Structures of Transition States and Intermediates in Folding: Delineation of Pathways at High Resolution

A. R. Fersht

Phil. Trans. R. Soc. Lond. B 1995 **348**, 11-15
doi: 10.1098/rstb.1995.0040

Mapping the structures of transition states and intermediates in folding: delineation of pathways at high resolution

A. R. FERSHT

Cambridge Centre for Protein Engineering, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

SUMMARY

The structures of all the intermediates and transition states, from the unfolded state to the native structure, are being determined at the level of individual residues in the folding pathways of barnase and chymotrypsin inhibitor 2 (CI2), using a combination of protein engineering and nuclear magnetic resonance methods. Barnase appears to refold according to a classical framework model in which elements of secondary structure are flickeringly present in the denatured state, consolidate as the reaction proceeds and, when nearly fully formed, dock in the rate-determining step. Unlike barnase, CI2 folds without a kinetically significant folding intermediate. The transition state for its formation has no fully formed elements of secondary structure, and the transition state is like an expanded form of the native structure. CI2 probably represents the folding of an individual domain in a larger protein, whereas barnase represents the folding of a multi-domain protein. The protein engineering methods are being extended to map the pathway in the presence of molecular chaperones. There are parallels between the folding of barnase when bound to GroEL and in solution.

1. INTRODUCTION

(*a*) *Pathways of protein folding*

Levinthal pointed out some 25 years ago that, because of the astronomical number of conformations available to an unfolded polypeptide chain, protein folding could not occur by a random, unbiased search of all conformational space (Levinthal 1968). This has led to the idea that there must be defined pathways of folding by which only a small fraction of conformational space is searched. These mechanisms may be divided up into two basic classes (figure 1). The first involves secondary structure being formed before tertiary structure: 'framework models' (Ptitsyn 1973, 1991). It is postulated that small segments of native-like secondary structure form locally and rapidly. These then lead to the formation of tertiary structure. One sub-division of the mechanism is the 'diffusion–collision' model (Karplus & Weaver 1976, 1994): the elements of secondary structure diffuse, collide and coalesce. The second sub-division is 'propagation' (Wetlaufer 1973): an element of secondary structure acts as a nucleation point and structure spreads out from this. The second class of mechanisms involves a general hydrophobic collapse of the protein, from which state rearrangement occurs, leading to the formation of secondary structure.

(*b*) *The strategy for solving pathways*

There is a simple strategy that has been used by chemists and biochemists for elucidating mechanisms of reactions and pathways that can be directly translated into analysing the pathway of protein folding. The pathway of folding will be defined when the structures of all intermediates and transition states have been characterized from the initial unfolded state to the final folded state. Until recently, the experimental means of tackling the problem of protein folding at high resolution have not been available. There are now two procedures that can characterize the necessary structures at the level of individual residues and even atoms. Nuclear magnetic resonance (NMR) is unparalleled as a spectroscopic technique in the level of detail obtainable for structure in solution. The power of this method can be applied to analysing stable states, including the unfolded state, and also to be used for studying transients by H–D exchange. For highly defined structures, information is given about backbone and side chains. For less characterized structures and H–D exchange, information tends to be more about the backbone and secondary structure.

The only way of analysing transition states is by kinetics, and the only way of analysing detail at individual residues is by kinetic measurements on mutant enzymes: the 'protein engineering approach' (Matouschek *et al.* 1989, 1990; Arcus *et al.* 1994). The protein engineering method may also be used to characterize intermediates. This procedure has been described in detail (Fersht *et al.* 1992) and recently in a more general review (Fersht 1993). The principle of the method is very simple. A suitable side chain in the protein is mutated, generally to a smaller one or to that of alanine, and the change in stability of the protein is measured. The effects of mutation on the stability of intermediates and transition states are also measured. If, for example, a transition state is destabilized by

Phil. Trans. R. Soc. Lond. B (1995) **348**, 11–15
Printed in Great Britain

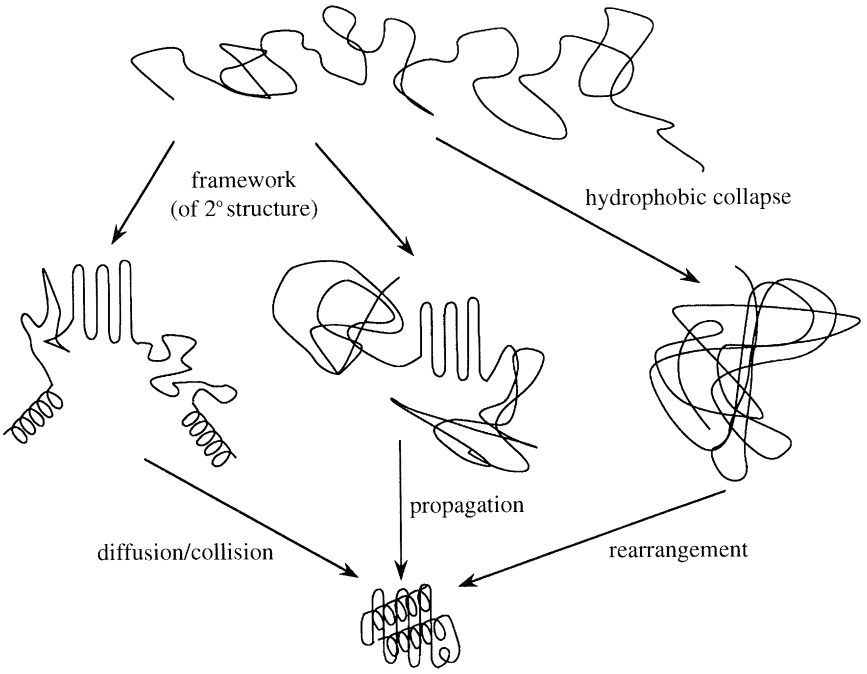

Figure 1. Pathways of folding.

exactly the same amount as the folded state on mutation, then the structure at the site of mutation is most likely the full native structure. If, on the other hand, the mutation does not alter the stability of the transition state at all, then the protein must be fully unfolded at the site of the mutation. These principles may be converted into quantitative terms by defining a parameter ϕ which is the ratio of change in energy of a particular state to the change in energy of the native state on mutation. $\phi = 0$ means structure is not formed; $\phi = 1$ means structure is fully formed.

2. DISSECTION OF THE FOLDING PATHWAY OF BARNASE

Barnase is a small ribonuclease of 110 amino acid residues that is an almost perfect paradigm for studying the folding of a small protein that folds via a distinct folding intermediate. Importantly, it has no disulphide cross-links that prevent its unfolding fully and so it can be used to analyse events that are presumed to follow from peptide biosynthesis onwards. The structure (figure 2) is characterized by a major helix (residues 6–18) docking on to a five-stranded anti-parallel β-sheet. The hydrophobic side chains from each element of secondary structure interdigitate to form the hydrophobic core. We now work backwards from the fully folded structure to the unfolded state showing how those elements of structure persist or become embryonic as the structure unfolds.

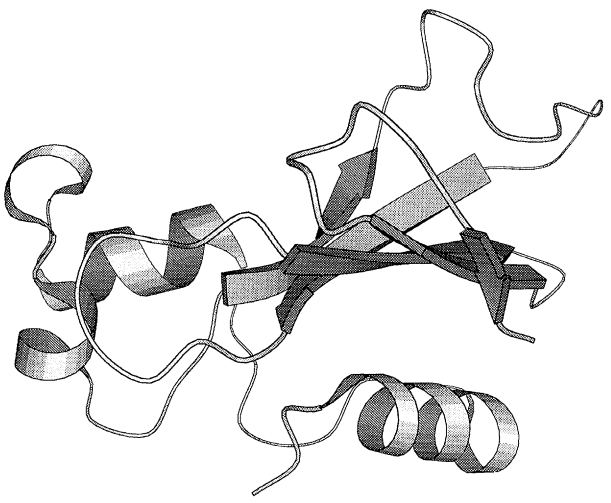

Figure 2. Structure of barnase. The major α-helix is at the bottom.

(*a*) *The unfolding transition state*

The protein engineering method has been applied systematically around the enzyme to map out the structure of the transition state for unfolding and the subsequent folding intermediate. Most ϕ values are either 1 or 0, indicating that structure is close to being fully formed or fully unformed in the transition state. The major helix is formed from residues 8–18. The sheet is basically formed but is weaker at the edges. The hydrophobic core is in the process of being formed. Individual details have been summarized by Fersht (1993).

(*b*) *The folding intermediate*

This is similar to the transition state but slightly more weakened at the edges of the sheet and throughout the core.

(*c*) *The unfolded state*

One could be excused from thinking that the unfolded state is the least interesting on the pathway. But, in many ways it is the most interesting because it could contain the clues as to whether framework or collapsed models hold. Until recently, unfolded states could not be studied at high resolution: NMR spectroscopy was

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

obscured by poor dispersion of the signals. However, the laboratory of Wüthrich has introduced the application of NMR magnetization transfer to assign an unfolded state (Neri *et al.* 1992; Wüthrich 1994), and the laboratory of Fesik (Logan *et al.* 1994) has used multinuclear methods for assigning another protein. We have combined both approaches to assign fully the unfolded state of barnase (Arcus *et al.* 1994). Although the structure has not yet been solved, there is sufficient information from chemical shifts and sequential nuclear overhauser effects (NOEs) to implicate regions of structure. First, there is a tendency for residues 6–18 to be in their helical formation. Second, the centre of the β-sheet has a collapsed structure. This appears to be non-native but can probably rearrange to native-like structure easily.

(*d*) *Summary of folding pathway*

Barnase thus appears to fold by a framework model. Weak structure in the major α-helix and the β-sheet is present in the unfolded state. These consolidate in a folding intermediate, tighten up in the transition state as the hydrophobic core consolidates to give the final folded structure.

3. FOLDING PATHWAY OF THE BARLEY CI2 INHIBITOR

The barley CI2 inhibitor is a small (64 residue) polypeptide that specifically inhibits chymotrypsin and subtilisin. It is an excellent paradigm for studying the folding of a very small protein that folds without there being a kinetically significant intermediate (Jackson & Fersht 1991). There is just a single rate-determining state linking the unfolded and folded state. This transition state has been analysed by the protein engineering methods both in the direction of folding and unfolding (Jackson *et al.* 1993; Otzen *et al.* 1994). The structure is found to be the same when measured in both directions. The transition state for the folding of CI2 is quite different in nature from that for barnase. Most of the ϕ values are fractional, there being no values of 1. The best formed part of the structure is the N-terminal region of the single α-helix. The rest of the protein is like the folded structure that has been expanded by 30%.

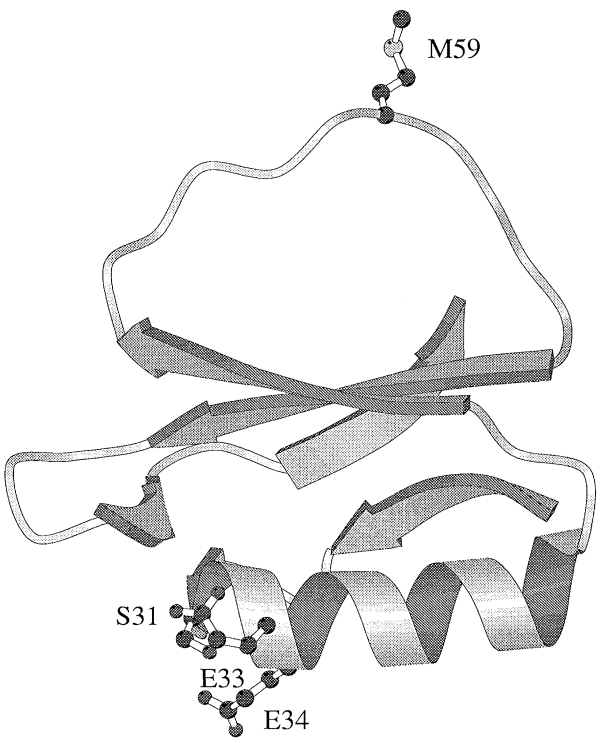

Figure 3. Structure of CI2. The single methionine is illustrated as are the residues at the N-terminal region of the helix which constitute the best-formed region in the transition state for refolding.

(*a*) *Molecular dynamics simulation of the transition state*

Li & Daggett (1994) have simulated the unfolding of CI2 using molecular dynamics. The picture they obtained is in remarkable agreement with that found from the protein engineering method. Their calculations flesh out in structural terms what the fractional ϕ values mean. Their structure shows a molecule with a distorted helix, the core coming apart and the sheet disintegrating. Caflisch & Karplus (1994) have simulated the transition state for the unfolding of barnase, again in agreement with experiments.

(*b*) *Fragments of CI2*

Peptide fragments of proteins provide a convenient and attractive way of analysing some characteristics of the unfolded states. Small fragments are more tractable to analysis by NMR, and can be studied in simple aqueous solution in water in the absence of denaturants. They also provide the means of looking for local effects in structure since residues far removed in sequence may be physically removed in the fragment. CI2 may be conveniently cleaved into two fragments using CNBr on the single methionine residue at position 59. A whole family of fragments of CI2 has been generated by cleaving all mutants made so far in this laboratory at their unique methionine 59 (Prat Gay & Fersht 1994; Prat Gay *et al.* 1994*a*, *b*). These two fragments recombine to give native-like structure. NMR studies on the complex show that it is almost identical to the native structure apart from the loop in which cleavage has occurred (B. Davies & A. R. Fersht, unpublished). The individual fragments have collapsed structure, but have non-native hydrophobic clusters being formed. The protein engineering method has been used to analyse the transition state for the association of the two fragments (Prat Gay *et al.* 1994*b*). It is remarkably like the transition state for the formation of an intact protein. Thus, two collapsed non-native structures can associate to give a folded protein via a transition state that is similar to that of the formation of the native protein.

(*c*) *Summary of folding pathway of CI2*

The folding pathway of CI2 thus appears to involve the rearrangement of a non-active collapsed structure to form the folded structure by a single transition state.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

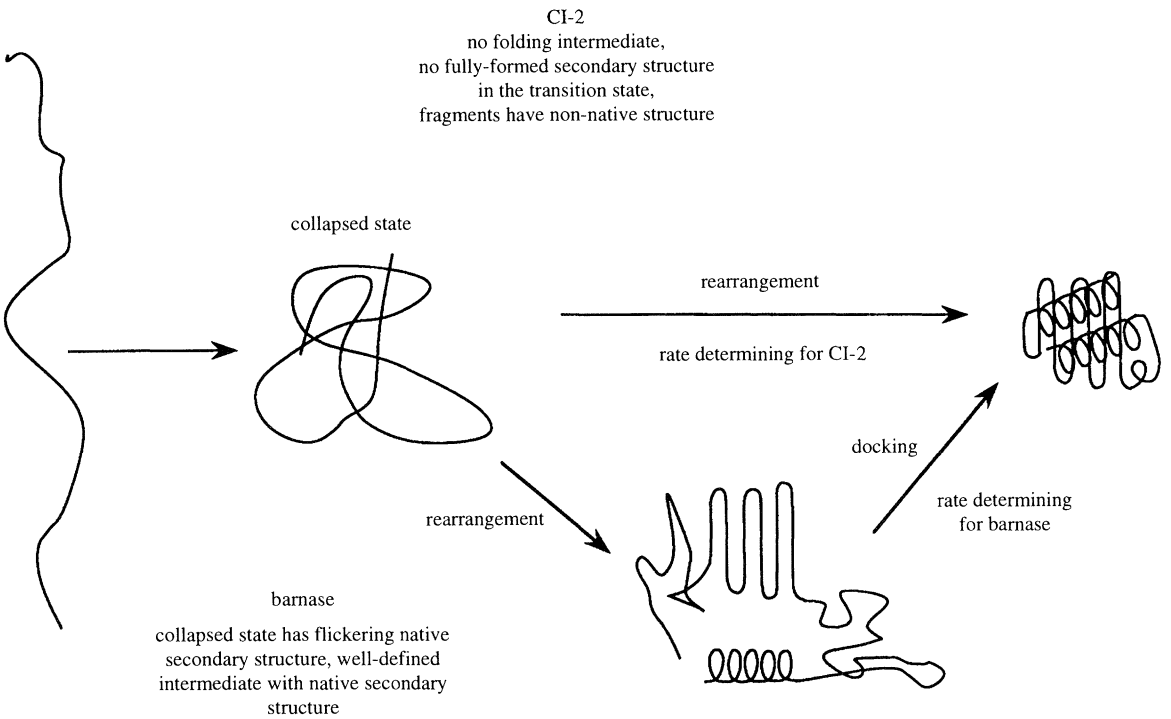

Figure 4. Unified scheme for the folding of the multi-domain (multi-module) protein barnase and the single-domain (single module) CI2.

3. UNIFIED SCHEME FOR THE FOLDING OF CI2 AND BARNASE

Barnase, although being small, does appear to have sub-structures that Go calls modules (Yanagawa *et al.* 1993), i.e. regions that make most of their interactions within themselves. CI2, on the other hand, appears to be like a single module, there being no substructures. We suggest that the folding of CI2 is representative of the folding of single modules of structure, whereas barnase represents the folding of larger structures that contain separate modules. Thus, we postulate, as in figure 4, that the folding pathway of CI2 involves the rearrangement of its collapsed state to that of the fully folded module (Otzen *et al.* 1994). Barnase first forms a state that has collection of regions of secondary structure in some of its individual modules which then coalesce in the rate-determining step to form the final folded structure. Thus, barnase folds by a clear framework model because it is large enough to do so. However, the framework does have an initial collapse step.

4. FOLDING *IN VITRO* VERSUS FOLDING *IN VIVO*

Protein folding *in vivo* is now known to involve a series of accessory proteins, which will be discussed by Hartl, Sigler, Tokatlidis *et al.* and Jaenicke in this volume. Members of the hsp70 class of protein (DnaK) coat the nascent polypeptide chain as it is synthesized on the ribosome. The polypeptide chain is then released under the influence of ATP and other proteins to be transferred to GroE whose structure is detailed by Sigler (this volume). This very large protein consists of two doughnut-shaped rings of 7-mers stacked one upon the other to give a large central cavity. There is currently controversy as to where the polypeptide binds and as to whether or not the polypeptide can fold while being bound to GroE or it has to be released into solution to do so.

We can tackle the folding of barnase in the presence of GroE using the protein engineering method. The use of mutants is one of the most powerful procedures for studying proteins *in vivo* and so our refined method adapts readily to provide important information about the pathway of folding in the presence of accessory proteins.

(*a*) *Barnase folds while binding to GroE*

Unfolded barnase is found to bind to GroE: when the two are mixed together, the rate of refolding of barnase is slowed down as measured by the regain in its catalytic activity (Gray & Fersht 1993). Further, the rate constant for the regain of activity levels off to a constant value in the presence of saturating concentration of GroE. This means that a folding event must occur while the peptide is bound to GroE because otherwise, by the principle of mass action, any unfolded barnase that is released would be mopped up by saturating concentrations of GroE. Further, the rate of refolding of barnase in the presence of saturating GroE is sensitive to mutations in barnase that are known to affect the rate of folding (Gray *et al.* 1993).

(*b*) *Folding when bound to GroE parallels folding in solution*

We have measured the rate constant for the refolding of barnase mutants in solution for the step of the intermediate progressing to the folded state. We found

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

that the rate constants for the refolding of the same mutants when bound to GroE are all scaled down by approximately the same factor (Gray *et al.* 1993). This shows that the refolding rate constant in the presence of GroE responds to mutation in the same way as does the transition from intermediate to folded state in solution. This is consistent with the pathway of folding in the presence of GroE being essentially the same as that in solution. The most likely explanation is that barnase refolds within the central cavity of GroE, consistent with the idea of GroE acting as a cage in which molecules can fold without competition from aggregation (Ellis 1994). Although there is evidence that larger proteins require dissociation from GroE for folding to occur (Weissman *et al.* 1994), it is possible that barnase is a model for the folding of small parts of larger molecules in the presence of GroE. Given that a protein as small as barnase can fold within the central cavity, it is thus quite likely that larger proteins could fold by different parts of their structure entering the cavity during various collisions and associations.

Studies on the spontaneous refolding of proteins *in vitro* and in the presence of molecular chaperones are thus converging. It is likely that within the next few years, the pathways of folding will be known at high resolution.

5. REFERENCES

Arcus, V.L., Vuilleumier, S., Freund, S.V., Bycroft, M. & Fersht, A.R. 1994 Towards solving the folding pathway of barnase: the complete backbone ^{13}C, ^{15}N and ^{1}H NMR assignments of its pH-denatured state. *Proc. natn. Acad. Sci. U.S.A.* **91**, 9412–9416.

Caflisch, A. & Karplus, M. 1994 Molecular-dynamics simulation of protein denaturation – solvation of the hydrophobic cores and secondary structure of barnase. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1746–1750.

Ellis, R.J. 1994 Roles of molecular chaperones in protein folding. *Curr. Opin. struct. Biol.* **4**, 117–122.

Fersht, A.R. 1993 Protein folding and stability – the pathway of folding of barnase. *FEBS Lett.* **325**, 5–16.

Fersht, A.R., Matouschek, A. & Serrano, L. 1992 The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. *J. molec. Biol.* **224**, 771–782.

Gray, T.E., Eder, J., Bycroft, M., Day, A.G. & Fersht, A.R. 1993 Refolding of barnase mutants and pro-barnase in the presence and absence of GroEL. *EMBO J.* **12**, 4145–4150.

Gray, T.E. & Fersht, A.R. 1993 Refolding of barnase in the presence of GroE. *J. molec. Biol.* **232**, 1197–1207.

Jackson, S.E., elMasry, N. & Fersht, A.R. 1993 Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor-2 – a critical test of the protein engineering method of analysis. *Biochemistry* **32**, 11270–11278.

Jackson, S.E. & Fersht, A.R. 1991 Folding of chymotrypsin inhibitor-2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428–10435.

Karplus, M. & Weaver, D.C. 1976 Protein folding dynamics. *Nature, Lond.* **260**, 404–406.

Karplus, M. & Weaver, D.L. 1994 Protein folding dynamics – the diffusion – collision model and experimental data. *Protein Sci.* **3**, 650–668.

Levinthal, C. 1968 Are there pathways for protein folding? *J. chem. Phys.* **85**, 44–45.

Li, A. & Daggett, V. 1994 Characterization of the transition state of protein unfolding using molecular dynamics: chymotrypsin inhibitor 2. *Proc. natn. Acad. Sci. U.S.A.* **91**, 10430–10434.

Logan, T.M., Theriault, Y. & Fesik, S.W. 1994 Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J. molec. Biol.* **236**, 637–648.

Matouschek, A., Kellis, J.T. Jr, Serrano, L., Bycroft, M. & Fersht, A.R. 1990 Transient folding intermediates characterized by protein engineering. *Nature, Lond.* **346**, 440–445.

Matouschek, A., Kellis, J.T. Jr, Serrano, L. & Fersht, A.R. 1989 Mapping the transition state and pathway of protein folding by protein engineering. *Nature, Lond.* **342**, 122–126.

Neri, D., Billeter, M., Wider, G. & Wuthrich, K. 1992 NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science, Wash.* **257**, 1559–1563.

Otzen, D.S.I.L., ElMasry, N., Jackson, S.E. & Fersht, A.R. 1994 The structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc. natn. Acad. Sci. U.S.A.* **91**, 10422–10425.

Prat Gay, G. de & Fersht, A.R. 1994 Generation of a family of protein fragments for structure-folding studies. I: Folding complementation of two fragments of chymotrypsin inhibitor-2 formed by cleavage at its unique methionine residue. *Biochemistry* **33**, 7957–7963.

Prat Gay, G. de, Ruiz-Sanz, J., Davis, B. & Fersht, A.R. 1994*a* The structure of the transition state for the association of two fragments of barley chymotrypsin inhibitor 2 to generate native-like protein: Implications for mechanisms of protein folding. *Proc. natn. Acad. Sci. U.S.A.* **91**, 10943–10946.

Prat Gay, G. de, Ruiz-Sanz, J. & Fersht, A.R. 1994*b* Generation of a family of protein fragments for structure-folding studies. II. Kinetics of association of the two chymotrypsin inhibitor-2 fragments. *Biochemistry* **33**, 7964–7970.

Ptitsyn, O.B. 1973 Stage mechanism of the self-organization of protein molecules. *Dokl. Acad. Nauk. SSR* **210**, 1213–1215.

Ptitsyn, O.B. 1991 How does protein synthesis give rise to the 3D-structure. *FEBS Lett.* **285**, 176–181.

Weissman, J.S., Kashi, Y., Fenton, W.A. & Horwich, A.L. 1994 Groel-mediated protein-folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* **78**, 693–702.

Wetlaufer, D.B. 1973 Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. natn. Acad. Sci. U.S.A.* **70**, 697–701.

Wuthrich, K. 1994 NMR assignments as a basis for structural characterization of denatured states of globular proteins. *Curr. Opin. struct. Biol.* **4**, 93–99.

Yanagawa, H., Yoshida, K., Torigoe, C., Park, J.S., Sato, K., Shirai, T. & Go, M. 1993 Protein anatomy – functional roles of barnase module. *J. biol. Chem.* **268**, 5861–5865.