Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10426–10429, October 1994
Biochemistry

Single versus parallel pathways of protein folding and fractional formation of structure in the transition state

(barnase/chymotrypsin inhibitor 2/linear free-energy relationships)

ALAN R. FERSHT, LAURA S. ITZHAKI, NADIA F. ELMASRY, JACQUELINE M. MATTHEWS, AND DANIEL E. OTZEN

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 12, 1994

ABSTRACT **Protein engineering and kinetic experiments indicate that some regions of proteins have partially formed structure in the transition state for protein folding. A crucial question is whether there is a genuine single transition state that has interactions that are weakened in those regions or there are parallel pathways involving many transition states, some with the interactions fully formed and others with the structural elements fully unfolded. We describe a kinetic test to distinguish between these possibilities. The kinetics rule out those mechanisms that involve a mixture of fully formed or fully unfolded structures for regions of the barley chymotrypsin inhibitor 2 and barnase, and so those regions are genuinely only partially folded in the transition state. The implications for modeling of protein folding pathways are discussed.**

A fundamental question in protein folding is whether there are single or parallel pathways (1, 2). This problem has been recently raised from results of computer simulations based on a highly simplified model of protein folding that suggest there are many parallel transition states (3). These transition states are characterized by having different parts of their structures energetically identical to those elements in the folded protein: some transition states have one particular set of secondary structural elements fully formed, whereas others have different ones. According to Baldwin (2), there is an immediate challenge arising from these findings for experimentalists to determine whether or not the folding of real proteins has a unique transition state.

Our laboratory has been developing protein engineering methods for analyzing the transition states of protein folding and unfolding that can potentially detect such behavior and answer the question (4–6). Site-directed mutagenesis is used to remove parts of a particular side chain and hence the interactions made by them. The change in the Gibbs free energy of folding, $\Delta\Delta G_{F-U}$, is measured from equilibrium denaturation experiments (F, folded; U, unfolded). The change in the free energy of any other state X with respect to the unfolded state, $\Delta\Delta G_{X-U}$, is measured by kinetics. A quantity Φ_F is obtained, defined by $\Phi_F = \Delta\Delta G_{X-U}/\Delta\Delta G_{F-U}$. Φ_F can be interpreted quantitatively for two extreme situations: $\Phi_F = 1$ occurs when the state X is as destabilized by mutation as is the fully folded state, and so the mutated side chain may be assumed to be in its native environment in state X; $\Phi_F = 0$ means the state X is as affected by the mutation as is the fully denatured state, and so the mutated side chain is in a fully denatured environment in state X (5).

Fractional values of Φ are difficult to interpret because it is not obvious *a priori* whether these imply that the structure at the site of mutation is simply weakened or there is a mixture of fully unfolded and fully folded states arising from parallel pathways (4, 5, 7), as implied by the calculations of Sali *et al.* (3). Further, there is not, in general, a linear relationship between Φ and the extent of bond making or breaking, because the terms $\Delta\Delta G_{F-U}$ and $\Delta\Delta G_{X-U}$ can be composed of many different individual energetic terms (5). However, we can analyze the behavior of Φ for certain types of parallel pathways as follows.

Simple Model of Two Parallel Pathways

Suppose, as sketched in Fig. 1, a protein that consists of an α-helix and a β-sheet has two pathways of folding. In one hypothetical pathway (the upper in Fig. 1), the transition state has the helix fully formed and the sheet unfolded. In the other pathway, the helix is fully unfolded and the sheet is fully formed. The average structure of the ensemble of transition states will thus appear to have partially formed helix and sheet. There will thus be apparent fractional values of Φ. The observed rate constant for folding, k_F, is given by

$$k_F = k_1 + k_2, \quad [1]$$

and the observed rate constant for unfolding, k_U, by

$$k_U = k_{-1} + k_{-2}. \quad [2]$$

Suppose a residue A in the helix that does not make contact with the sheet is mutated and destabilizes the helix, and hence the protein, by a change of $\Delta\Delta G_{F-U}$ in its free energy of folding (where $\Delta\Delta G_{F-U}$ is positive). We can calculate how the rate constants for folding and unfolding change on mutation in this simple model. In the direction of folding, the rate constant k_2 is unaffected, since A is unfolded in both the ground state and the transition state. However, k_1 is lowered by a factor of $\exp(\Delta\Delta G_{F-U}/RT)$, since the equilibrium constant for formation of the helix is lowered by that factor. Thus,

$$k_F = k_1\exp(-\Delta\Delta G_{F-U}/RT) + k_2. \quad [3]$$

Conversely, for unfolding, k_{-1} is unaffected on mutation, since the helix is fully formed in both the ground state and the transition state, but k_{-2} is increased by a factor of $\exp(\Delta\Delta G_{F-U}/RT)$, since the helix becomes fully unfolded as the reaction proceeds to the transition state, and the equilibrium constant for unfolding increases by that factor. Thus,

$$k_U = k_{-1} + k_{-2}\exp(\Delta\Delta G_{F-U}/RT). \quad [4]$$

Qualitatively, as the mutation becomes progressively more destabilizing as $\Delta\Delta G_{F-U}$ increases, there is a switch from the

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.

Abbreviation: CI2, chymotrypsin inhibitor 2.

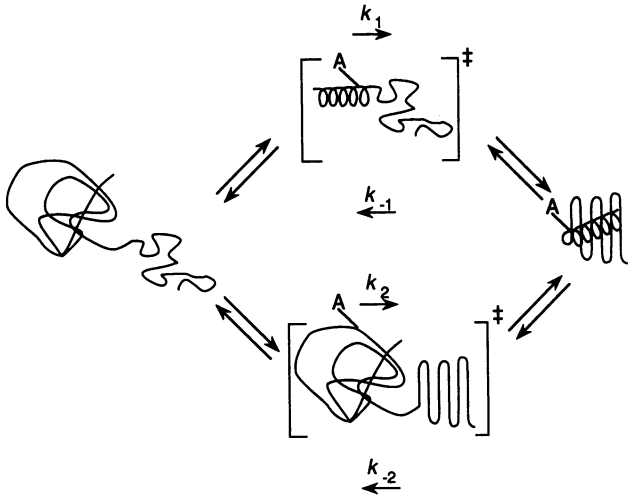

FIG. 1. Parallel versus single pathways. The protein folds between two states with rate-determining transition states (‡). The states could be the unfolded and folded states of the barley chymotrypsin inhibitor 2 (CI2) or the major folding intermediate and the final folded state of barnase. The transition state in the upper pathway has, for example, a helix formed and the regions containing, for example, the sequence of a β-sheet unfolded. The converse holds for the lower. The residue A is a target for mutagenesis.

upper pathway in Fig. 1, which has the helix being fully formed, to the lower pathway, which has the helix unfolded in the transition state. We can test for this type of behavior by examining the folding or unfolding of mutants of proteins that have a series of mutations in a single element of secondary structure. For unfolding, for example, suppose $k_{-1} \gg k_{-2}$ for the wild-type protein. Then, a plot of ln k_U versus $\Delta\Delta G_{F-U}/RT$ will initially have a slope of close to 0 in the region $k_{-1} \gg k_{-2}\exp(\Delta\Delta G_{F-U}/RT)$. As $\Delta\Delta G_{F-U}/RT$ increases so that $k_{-2}\exp(\Delta\Delta G_{F-U}/RT)$ becomes $\gg k_{-1}$, the slope will tend to 1.0. There will be an intermediate region of variable and fractional slope.

This model can be described by simple equations because it is based on the two extreme Φ values, 0 and 1.0. This type of behavior can be contrasted with a reaction having a single pathway with a transition state that has partial structure formation and which may follow the Brønsted equation.

Brønsted Behavior. If there are simple relationships between the rate constants and bond energy changes and all mutations probe the same degree of structure formation, then the observed rate constant for folding could follow the Brønsted equation,

$$\ln k_F = \ln k_F^o - \beta_F \Delta\Delta G_{F-U}/RT, \quad [5]$$

and that for unfolding,

$$\ln k_U = \ln k_U^o + (1 - \beta_F)\Delta\Delta G_{F-U}/RT, \quad [6]$$

where k_F^o and k_U^o are the rate constants for folding and unfolding of the parent molecule and β_F is a constant that is related to the degree of structure formation.

We do not expect *a priori* that the Brønsted equation should be followed with precision for a series of mutations, even when there is an element of structure that changes in a concerted manner during the folding reaction, because of the aforementioned complications in relating changes in rate constants to the energies of individual bonds (5). Despite those possible complications, however, it has been found that changes in reactivity on the mutation of proteins can follow the Brønsted equation quite precisely, as demonstrated in catalysis by the tyrosyl-tRNA synthetase (8, 9) and other enzymes (10), and this has now been rationalized on theoretical grounds (11). We have sufficient data on certain elements in two proteins to test the simple model. The larger the range of rate constants being examined, the more the sensitivity of the analysis.

Example 1. The Minicore of CI2. The 64-residue fragment of the barley serine protease inhibitor CI2 is a beautiful paradigm for such an analysis because it folds and unfolds according to a simple two-state model with a single rate-determining transition state (or single family of transition states) that can be examined in both the direction of folding and the direction of unfolding (12, 13). The protein has two major elements of classical secondary structure, an α-helix and a β-sheet, that dock on each other to form the major hydrophobic core. There is a second hydrophobic cluster, which we term the minicore. This consists of just three residues, Leu-51, Val-57, and Phe-69, which have been extensively probed by mutation to other hydrophobic residues. A plot of ln k_U versus $\Delta\Delta G_{F-U}/RT$ (Fig. 2) fits nicely to a Brønsted equation with $1 - \beta_F = 0.64 \pm 0.02$ over a range of nearly 5 kcal/mol in $\Delta\Delta G_{F-U}$ and a factor of 300 in rate constants (correlation coefficient, $r = 0.99$). The corresponding plot of ln k_F versus $\Delta\Delta G_{F-U}/RT$ (Fig. 2) fits a Brønsted equation with $\beta_F = 0.3 \pm 0.02$ ($r = 0.95$) over the same range of $\Delta\Delta G_{F-U}$ and a factor of 10 in rate constant [a smaller range because $\beta_F < (1 - \beta_F)$]. The data clearly do not fit the simple two-pathway model of Fig. 1, which is described by Eqs. 3 and 4, but do fit a model of partial structure formation.

The surprisingly good fit to a simple Brønsted equation must be due to two factors: all the mutations probe the same region of structure, which changes in a concerted manner during the reaction, and changes in activation energies parallel changes in equilibrium stability. Fersht *et al.* (5) have pointed out that the situation of hydrophobic-to-hydrophobic mutations is the one where it is most likely that changes in reaction rate constants mirror the changes in actual bond energies of the bonds that are in the process of being made or broken.

Example 2. The α-Helix of CI2. There are many probes along the helix but there are smaller ranges of destabilization energies and rate constants on mutation than for the minicore (Fig. 3). Nevertheless, there are acceptable Brønsted plots for unfolding ($1 - \beta_F = 0.65 \pm 0.08$; $r = 0.87$) and for folding ($\beta_F = 0.38 \pm 0.06$, $r = 0.80$). These occur over a range of values of $\Delta\Delta G_{F-U}$ of 2.5 kcal/mol, which is ample enough to detect the curvature expected for the parallel-pathway mechanism. The Brønsted plots are more scattered than those for

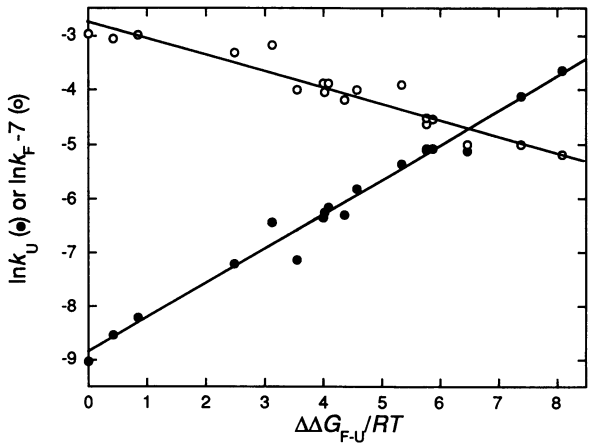

FIG. 2. Brønsted plot of rate constants for folding and unfolding for mutations in the minicore of CI2 (wild type, L51A, L51I, L51V, L51V/F69L, L51V/F69A, L51A/F69L, L51A/F69A, L51A/V57A, L51V/V57A, L51A/V57A/F69L, L51V/V57A/F69L, V57A, V57A/F69L, V57A/F69A, F69L, F69V, and F69A, where L51A indicates that wild-type residue Leu-51 is mutated to alanine, etc.).

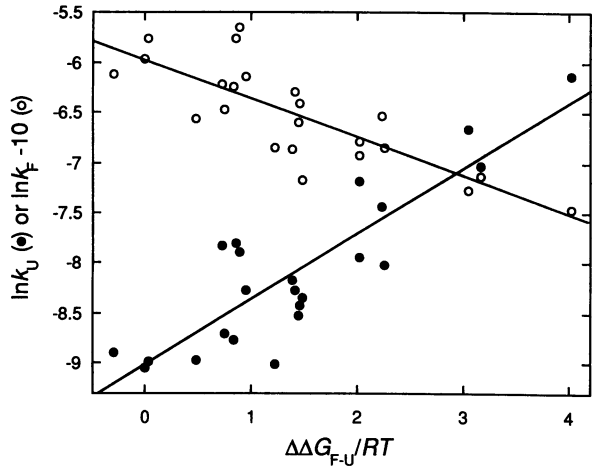

FIG. 3. Brønsted plot of rate constants for folding and unfolding for mutations in the α-helix of CI2 (wild type, S31G, S31A, E33Q, E33D, E33N, E34Q, E34D, E34N, E33A/E34A, S31G/E33A/E34A, S31A/E33A/E34A, K36A, K36G, K37A, K37G, V38A, I39V, L40A, L40G, Q41A, Q41G, and K43A). The mutations are of predominantly solvent-exposed residues.

the minicore, perhaps because the mutations are of polar residues—i.e., types of changes that are not expected to fit such simple binding relationships well—and also because the value of Φ probably varies along the helix (5).

Other regions of the structure either do not have enough probes for the above analysis or show, as in the major hydrophobic core (12), a gradation of Φ values within the structure so that neither simple Brønsted plots nor Eqs. 3 and 4 are followed.

Example 3. The Major α-Helix of Barnase. The small ribonuclease barnase folds via a distinct folding intermediate with multiphasic kinetics (6). All existing data support a mechanism of folding of wild-type enzyme in which the major α-helix (residues 6–18) is predominantly folded in the major folding intermediate and residues 12–18 are nearly fully folded in the subsequent transition state, which is the rate-determining transition state for unfolding. The data for analysis of barnase (Fig. 4) are restricted to the monophasic unfolding reaction, which represents the folded state proceeding to the major folding intermediate. Plots of ln k_U against $\Delta\Delta G_{F-U}/RT$ are biphasic for the reaction either in 7.25 M urea or in water. In water, the value of $1 - \beta_F$ tends to 0.33 ± 0.04 for low values of $\Delta\Delta G_{F-U}$ ($-1.6 < \Delta\Delta G_{F-U}/RT < 2.8$, $r = 0.92$) and to 0.78 ± 0.03 at high values of $\Delta\Delta G_{F-U}$ ($2.8 < \Delta\Delta G_{F-U}/RT < 8.5$, $r = 0.99$). In 7.25 M urea, these decrease to 0.14 ± 0.03 ($r = 0.80$) and 0.58 ± 0.04 ($r = 0.98$), respectively. [It is expected that the transition state for unfolding moves closer to the structure of the folded state at high concentrations of urea because of the Hammond effect (15).] Although this biphasic behavior may appear qualitatively consistent with Eqs. 3 and 4, it must be noted that the slopes of the curves at high destabilization of the helix tend to values significantly less than 1.0. Quantitatively, the data do not fit the simple parallel-pathway model involving competition between either fully unfolded or fully folded helices. They support, instead, either a single transition state in which the helix becomes progressively more unfolded in the transition state as it is destabilized or two competing pathways where one has the helix nearly fully folded and the other has it severely weakened, but not fully unfolded.

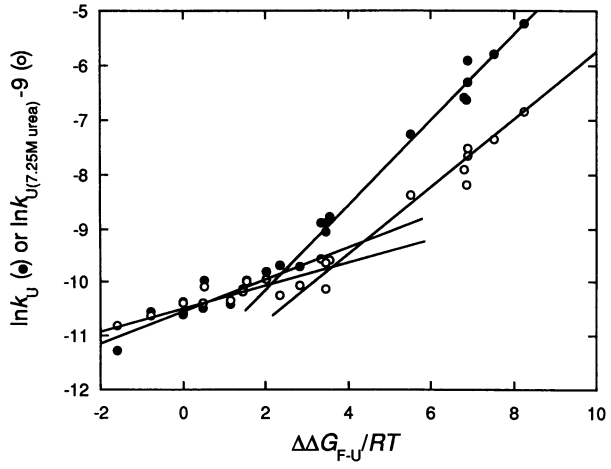

FIG. 4. Brønsted plot of rate constants for unfolding in water (○) and in 7.25 M urea (●) for mutations in the major α-helix of barnase (wild type, D8A, D12G, D12A, Y13A, Q15I, T16G, T16A, T16S, T16R, Y17G, Y17A, H18K, H18Q, H18A, H18G, D12G/Y17G, D12A/Y17G, Y13A/T16S, Y13A/Y17A, T16S/Y17A, and Y13A/T16S/Y17A). Data are from ref. 14 or are unpublished data obtained by identical procedures.

Extension of the Simple Model to Multiple Pathways

Suppose there are many pathways of folding and unfolding that have, for example, a helix either completely formed or completely unfolded in the transition state. Then there will be a family of rate constants for folding for those with the helix formed, $\Sigma_n k_{1,n}^{helix}$, and a family for those with the helix unfolded, $\Sigma_n k_{2,n}^{coil}$. If a residue in the helix is mutated as before to destabilize it by $\Delta\Delta G_{F-U}$, then each member of the family with the helix being formed in the transition state will be attenuated by a factor of $\exp(\Delta\Delta G_{F-U}/RT)$, as before, and each of the other family will be unaffected. The observed rate constant for folding will be, on mutation,

$$k_F = \sum_n k_{1,n}^{helix} \exp(-\Delta\Delta G_{F-U}/RT) + \sum_n k_{2,n}^{coil}. \quad [7]$$

This equation behaves with increasing values of $\Delta\Delta G_{F-U}$ in exactly the same manner as for the simple model of just two pathways. The same is true for the unfolding reactions. Thus, the above analysis also eliminates multiple pathways with elements of structure either fully formed or fully unfolded.

Discussion

The kinetic procedures presented here allow the detection of parallel pathways of protein folding and unfolding that involve transition states of either fully formed or fully unfolded elements of structure. We have examined two cases where there appears to be partial formation of elements of structure in the transition state and one where mutation leads to apparent partial formation. Our analysis of the data eliminates those mechanisms that attribute the partial formation of structure to its arising from a mixture of parallel pathways involving either fully formed or fully unfolded structural elements. Instead, the data are consistent with either a single pathway of folding that involves the particular elements of structure being partially formed or a mixture of pathways where at least some have partial formation of structure.

The nature of evidence from kinetics is that it tends to eliminate alternatives rather than prove particular mechanisms. The experiments described here do not prove conclusively that there is a single pathway of folding for barnase or CI2. But we have eliminated a whole class of mechanisms, and this is sufficient to establish that there is partial formation of structure in the transition state of folding or proteins. This has implications for theoretical analyses of protein folding. The Monte Carlo simulations of Sali *et al.* (3) that proposed parallel pathways with a mixture of fully formed and fully

unfolded structural elements used an empirical potential function that assumed that the pairwise interaction energies between residues are the same in transition states as in folded structures (2). This assumption must be modified in order to refine their model.

An alternative approach for theoreticians for the simulation of the structure of transition states for protein folding is to use an old trick used by kineticists—study the kinetics of the reverse reaction that goes through the same transition state or set of transition states (ref. 16, p. 89). This greatly simplifies the problem in protein folding because the starting point is the best-defined structure on the protein folding pathway, the folded state. This is particularly appropriate for CI2 since we have directly demonstrated that the same transition state can be studied by both folding and unfolding kinetics (12, 13, 17). Simulations by Li and Daggett (18) suggest a transition state that is astonishingly similar to the structure analyzed experimentally by us. Simulations of the unfolding of barnase (19) also indicate a single transition state similar to that found in our laboratory (6). Available evidence does suggest, therefore, that there are basically single transition states for the final stages of folding of CI2 and barnase, subject to the following caveats.

Our opinion is that the very first stages of folding do have multiple pathways, but these funnel into a single pathway in the later stages for many proteins (20, 21). Sali *et al.* (3) suggest that the state whose formation they may be modeling is a "molten globule," and this would appear to be more consistent with the experimental data. It must be emphasized that we are not suggesting that there are never parallel pathways of protein folding in the later transition states. For example, there are minor parallel paths for the folding of barnase and CI2 resulting from small subpopulations in the unfolded state having cis peptidyl-prolyl bonds (7, 22). The folding of lysozyme from an unfolded state that has four disulfide bridges also appears to have parallel phases (23). We assumed in our initial protein engineering studies on mapping a transition state for protein folding that it would be an ensemble of similar structures because of the low energies of many noncovalent bonds (4). There could thus be a fluctuation around a basic pathway which has a relatively broad minimum in its energy surface and a wide and long saddle point at its transition state. The disorganized elements of structure in the transition state, such as the loops that are not folded in the transition state of barnase, could have a wide range of conformational possibilities, but the better-formed structures would be better defined (20). Such an ensemble of structures constitutes a different situation from that where there are the distinct competing pathways, illustrated in Fig. 1.

D.E.O. was supported by a fellowship from the Danish Natural Science Research Council.

1. Harrison, S. C. & Durbin, R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4028–4030.
2. Baldwin, R. L. (1994) *Nature (London)* **369**, 183–184.
3. Sali, A., Shakhnovich, E. & Karplus, M. (1994) *J. Mol. Biol.* **235**, 1614–1636.
4. Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. (1989) *Nature (London)* **342**, 122–126.
5. Fersht, A. R., Matouschek, A. & Serrano, L. (1992) *J. Mol. Biol.* **224**, 771–782.
6. Fersht, A. R. (1993) *FEBS Lett.* **325**, 5–16.
7. Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M. & Fersht, A. R. (1990) *Nature (London)* **346**, 440–445.
8. Fersht, A. R., Leatherbarrow, R. & Wells, T. N. C. (1987) *Biochemistry* **26**, 6030–6038.
9. Avis, J. M. & Fersht, A. R. (1993) *Biochemistry* **32**, 5321–5326.
10. Toney, M. D. & Kirsch, J. F. (1992) *Protein Sci.* **1**, 107–119.
11. Warshel, A., Schweins, T. & Fothergill, M. (1994) *J. Am. Chem. Soc.*, in press.
12. Jackson, S. E., ElMasry, N. & Fersht, A. R. (1993) *Biochemistry* **32**, 11270–11278.
13. Jackson, S. E. & Fersht, A. R. (1991) *Biochemistry* **30**, 10428–10435.
14. Matouschek, A., Serrano, L. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 819–835.
15. Matouschek, A. & Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7814–7818.
16. Fersht, A. R. (1985) *Enzyme Structure and Mechanism* (Freeman, New York), 2nd Ed.
17. Otzen, D. E., Itzhaki, L. S., El Masry, N. F., Jackson, S. E. & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 10422–10425.
18. Li, A. & Daggett, V. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 10430–10434.
19. Caflisch, A. & Karplus, M. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 1746–1750.
20. Matouschek, A. & Fersht, A. R. (1991) *Methods Enzymol.* **202**, 82–112.
21. Fersht, A. R. (1994) *Biochem. Soc. Trans.* **22**, 267–273.
22. Jackson, S. E. & Fersht, A. R. (1991) *Biochemistry* **30**, 10436–10443.
23. Radford, S. E., Dobson, C. M. & Evans, P. A. (1992) *Nature (London)* **358**, 302–307.