
Journal of Biological Physics **27:** 99–117, 2001.

A Comparison of Experimental and Computational Methods for Mapping the Interactions Present in the Transition State for Folding of FKBP12

E.R.G. MAIN[1], K.F. FULTON[1], V. DAGGETT[2] and S.E. JACKSON[1]
[1]*Cambridge University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, U.K.*
[2]*Department of Medicinal Chemistry, University of Washington, Seattle, WA 98195-7610*

Abstract. The folding pathway of FKBP12, a 107 residue α/β protein, has been characterised in detail using a combination of experimental and computational techniques. FKBP12 follows a two-state model of folding in which only the denatured and native states are significantly populated; no intermediate states are detected. The refolding rate constant in water is 4 s^{-1} at 25 °C. Two different experimental strategies were employed for studying the transition state for folding. In the first case, a non-mutagenic approach was used and the unfolding and refolding of the wild-type protein measured as a function of experimental conditions such as temperature, denaturant, ligand and trifluoroethanol (TFE) concentration. These data suggest a compact transition state relative to the unfolded state with some 70% of the surface area buried. The ligand-binding site, which is mainly formed by two long loops, is largely unstructured in the transition state. TFE experiments suggest that the α-helix may be formed in the transition state. The second experimental approach involved using protein engineering techniques with ϕ-value analysis. Residue-specific information on the structure and energetics of the transition state can be obtained by this method. 34 mutations were made at sites throughout the protein to probe the extent of secondary and tertiary structure in the transition state. In contrast to some other proteins of this size, no element of structure is fully formed in the transition state, instead, the transition state is similar to that found for smaller, single-domain proteins, such as chymotrypsin inhibitor 2 and the SH3 domain from α-spectrin. For FKBP12, the central three strands of the β-sheet (2, 4 and 5), comprise the most structured region of the transition state. In particular Val 101, which is one of the most highly buried residues and located in the middle of the central β-strand, makes approximately 60% of its native interactions. The outer β-strands, and the ends of the central β-strands are formed to a lesser degree. The short α-helix is largely unstructured in the transition state as are the loops. The data are consistent with a nucleation-condensation model of folding, the nucleus of which is formed by side chains within β-strands 2, 4 and 5 and the C-terminus of the α-helix. These residues are distant in the primary sequence, demonstrating the importance of tertiary interactions in the transition state. High-temperature molecular dynamic simulations on the unfolding pathway of FKBP12 are in good agreement with the experimental results.

Key words: FKBP12, immunophilin, protein folding, transition state, two-state folding

1. Introduction

Over the past ten years a variety of experimental and computational approaches have been used to study the folding pathways of small proteins. A number of different experimental approaches have been used to study the structure and energetics of folding transition states. Two main strategies have been adopted – the first is to study wild-type protein and measure the rates of unfolding and refolding as experimental conditions such as temperature, denaturant concentration, ligand concentration, etc. are varied. Temperature-dependence studies provide information on the thermodynamic nature of the transition state, as well as its compactness, as measured by burial of hydrophobic side chains. Studies on the folding and unfolding of wild-type proteins in the presence and absence of ligand have yielded information on the structure of ligand-binding sites during folding [1]. Recently, this type of approach was extended by studying the effect of sugars and alcohols on the rate of folding and unfolding [2, 3]. These experiments yielded information on the extent of hydration and secondary structure formation, particularly α-helices, in the transition state.

The second approach is to use protein engineering techniques, and measure the relative stabilities and rates of folding of wild-type and mutant proteins. ϕ-value analysis can then used to quantitatively measure the energetics of structure formation in the transition state [4]. This experimental approach, which provides detailed, residue-specific information on the energetics of the transition state for folding, has been complemented by molecular dynamic simulations which have provided structural information on the transition state [5–7, 8].

In this paper, we present a comprehensive characterisation of the pathway and transition state for folding of FKBP12. We have employed both experimental approaches outlined above, as well as performing molecular dynamic simulations, to give a very detailed picture of both the energetics and structure of the transition state for folding of this protein. The results obtained using the different experimental strategies are compared, as are the results obtained from theoretical simulations and experiment. This is the first example of the use of all three approaches to characterise the transition state for folding of a single protein.

2. Structure of FKBP12

Human FKBP12 is a 12 kDa protein, 107 residues in length. It contains no disulphide bridges, and all of its seven prolines are in a *trans* conformation in the native state, thus folding is not limited by disulphide bond formation or rearrangement, or proline isomerisation. In addition, it has been shown that FKBP12 undergoes a reversible two-state unfolding under equilibrium conditions which can be induced using urea or guanidinium chloride (GdnHCl) [9]. The urea and GdnHCI-induced denatured states have been characterised using NMR spectroscopy which has shown that, although there is some evidence for fleeting residual structure, there is no evidence for extensive structure in the unfolded state [10]. The structure

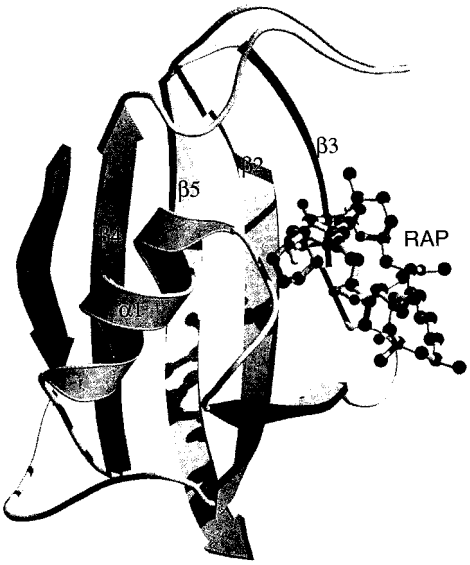

Figure 1. Structure of the human FKBP12-rapamycin complex. The major secondary structural features of FKBP12 are a large, amphiphilic, 5-stranded antiparallel β-sheet with a +3, +1, –3, –1 topology and a small, amphiphilic α-helix. The α-helix, from Ile 56 to Val 63, packs against the hydrophobic face of the β-sheet to form the main hydrophobic core of the protein. The β-sheet comprises of β-strand 1 (Val 2 to Ser 8), β-strand 4 (Arg 71 to Ile 76), β-strand 5 (Leu 97 to Leu 106), β-strand 2 (Thr 21 to Leu 30), β-strand 3, which splits into two (Lys 35 to Ser 38 and Phe 46 to Met 49). The rapamycin-binding site comprises residues Tyr 26, Phe 46, Val 55, Ile 56, Trp 59 and Phe 99. Rapamycin (RAP) is shown using a ball-and-stick representation. The secondary structure representation was produced using *Molscript* [45] and rendered using Raster3D [46].

of native FKBP12, which has been solved by x-ray crystallography and NMR spectroscopy [11] is outlined below and shows some interesting features which may affect the folding pathway.

The structure of FKBP12 is characterised by a large, amphiphilic, antiparallel five-stranded β-sheet with +3,+1,–3,–1 topology, see Figure 1. An amphiphilic α-helix packs against the hydrophobic face of the β-sheet at an angle of 60° with respect to the long axis. The β-sheet has a right-handed twist and wraps around the helix to form a well-ordered hydrophobic core. The immunosuppressant-binding site, which contains many aromatic side chains, forms a large, shallow hydrophobic pocket between the α-helix and β-sheet. A notable feature of FKBP12 resulting from the +3,+1,–3,–1 topology of the β-sheet is a topological crossing of loop-1 (Pro9 – Gln20 which connects β-strand 1 and β-strand 2), and loop-4 (Ala64 – Gln70 which connects the α-helix to β-strand). Although crossing topologies have been observed in proteins containing parallel β-sheets they are rare in antiparallel β-sheet structures. It has been suggested that the topological crossing of loops may

result in complex folding pathways which would slow folding, and may even result in misfolded species which would form kinetic traps on the folding pathway [12].

3. Two-State Nature of the Folding Transition

Experimentally, it has long been established that stable intermediates are not essential for the fast, efficient folding of a protein [13, 14]. In these simple systems only the denatured and native states are significantly populated, even in the absence of denaturant, and the folding of these proteins is therefore referred to as two-state. Such proteins represent the simplest experimental models for studying folding. More than twenty small proteins have now been shown to fold with simple two-state kinetics [15] and it is likely that many other small proteins, or domains of larger proteins, will also fold in this way. Experimentally, it is not possible to observe intermediates if they are higher in energy than the denatured state. However, computational studies suggest that, in some cases, high-energy intermediate states are not present on folding pathways [16].

By determining the unfolding and refolding rate constants over a wide range of denaturant concentrations we have shown that FKBP12 fits to a two-state model of folding, Figure 2a [17]. These results are supported by stopped-flow circular dichroism experiments in which no fast 'burst' phase, attributable to the accumulation of an intermediate on the folding pathway, is observed [17]. As discussed above, experimentally it is difficult to observe intermediate states if they are higher in energy than the denatured state. In some cases, however, such states have been stabilised, and therefore become observable, by the addition of a chaotropic salt such as sodium sulphate [18]. Even in the presence of 0.5 M Na_2SO_4 the folding kinetics of FKBP12 remain two-state, Figure 2a [17]. Thus, folding intermediates, if present on the pathway of folding of FKBP12, must be high in energy and very unstable. We are currently undertaking further experiments to probe the two-state nature of the folding transition for FKBP12 in detail. Computational studies [16] suggest that the folding of FKBP12 is truly two-state (Bill Eaton, personal communication).

4. Experimental Approaches to Studying Folding Transition States

STRATEGY 1: NON-MUTAGENIC APPROACHES

The unfolding and refolding rate constants for wild-type FKBP12 have been determined over a range of conditions to yield information on the transition state for folding. The dependence of the rate constants on denaturant concentration is shown in Figure 2a. From this the average solvent accessibility of main chain and side chain groups in the transition state relative to the native state, β_T, can be determined. The values calculated using rate constants measured in either urea, Figure 2a, or guanidinium chloride, GdnHCl, Figure 2c, are within experimental at 0.7. This indicates that 70% of the solvent accessible surface area buried in the

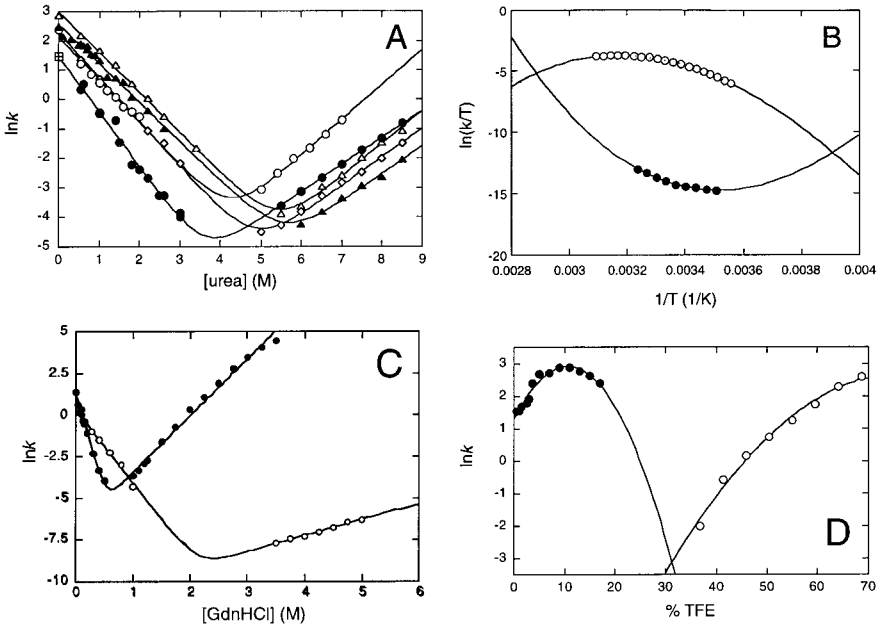

Figure 2. Kinetic unfolding and refolding experiments. (A) [urea]-dependence of the natural logarithm of the rate constants for unfolding and refolding of FKBP12. Rate constants are measured at 25 °C, in 50 mM TrisHCl, 1 mM DTT, pH 7.5. Points between 0.54–8 M were obtained by [urea]-jump experiments (closed circles), the point at 0 M urea was obtained by acid-jump, or alkali-jump experiment (hatched square). Each point is the average of at least five separate experiments. Also shown are the rates measured in 0.5 M Na_2SO_4 (closed triangles), 3.6% TFE (open diamonds), 9.6% TFE (open triangles), and 17% TFE (open circles). The solid curves shows the best fit of the data to a two-state model [13]. (B) Eyring plot of the temperature dependence of the unfolding rate constant at a final urea concentration of 7 M, in 50 mM phosphate, 1 mM DTT, pH 7.5. The solid curve shows the best fit of the data, for equations see [19]. The data shown has been offset by a constant in order to illustrate the unfolding and refolding curves under identical conditions, i.e., in water. The data was not offset during fitting. Eyring plot of the temperature dependence of the fast refolding rate constant at 0 M urea, in 50 mM phosphate, 1 mM DTT, pH 7.5. The solid curve shows the best fit of the data, for equations see [19]. The intersects of the two curves indicate the midpoint of heat or cold denaturation, and are +73 °C and –17 °C, respectively. (C) [GdnHCl]-dependence of the natural logarithm of the refolding rate constant for FKBP12 (closed circles) and the complex formed between FKBP12 and rapamycin (open circles). Other experimental details as Figure 2a. The solid curves shows the best fit of the data to a two-state model, for equation see [13]. (D) [TFE]-dependence of the natural logarithm of the rate constants for unfolding and refolding of FKBP12. Rate constants are measured at 25 °C, in 50 mM TrisHCl, 1 mM DTT, pH 7.5. Points between 0–3% TFE were obtained by pH-jump experiment (closed circles), between 3.6–20% TFE by [TFE]-jump refolding experiment (open circles), and between 36–60% TFE by [TFE]-jump unfolding experiment (closed triangles). The solid curve shows the best fit of the unfolding and refolding data to a second order polynomial.

Table I. Activation parameters for the refolding and unfolding of FKBP12 calculated from the Eyring plots

	$\Delta C_p^{\ddagger}$ (kcal mol^{-1} K^{-1})	$\Delta G^{\ddagger}$ (298K) (kcal mol^{-1})	$\Delta S^{\ddagger}$ (kcal mol^{-1} K^{-1})	T$\Delta S^{\ddagger}$ (298K) (kcal mol^{-1})	$\Delta H^{\ddagger}$ (298K) (kcal mol^{-1})
unfolding	1.04 ± 0.08	18.7 ± 0.01	-16 ± 1	-4.7 ± 0.3	14 ± 0.3
refolding	-0.67 ± 0.01	16.6 ± 0.01	-17 ± 0.2	-5.0 ± 0.07	11.6 ± 0.07

native state is already buried in the transition state. These data suggest that, on average, the structure in the transition state is compact relative to the unfolded state. This value is typical of small monomeric proteins [15].

The temperature dependence of the unfolding and refolding rate constants is shown in Figure 2b in the form of Eyring plots. Both unfolding and refolding Eyring plots are clearly non-linear. Non-linearity is observed when there is a significant difference in heat capacity between the initial state and the transition state. In this case $\Delta H^{\ddagger}$, the activation enthalpy, and thus $\Delta G^{\ddagger}$, the activation energy, are dependent on temperature. These data can be analysed [19] to yield values for $\Delta C_p^{\ddagger}$, the change in heat capacity between the initial and transition state, $\Delta S^{\ddagger}$, the change in entropy between initial and transition state, $\Delta H^{\ddagger}$, the change in enthalpy between initial and transition state, and $\Delta G^{\ddagger}$, the change in free energy between initial and transition state. The values are shown in Table I. Subscript U is used to denote the activation parameters for the unfolding reaction and subscript F is used to denote the activation parameters for the refolding reaction.

There is a large enthalpic contribution to the unfolding and refolding activation energy at 298 K, and a much smaller entropic contribution. In general it is difficult to interpret changes in enthalpy and entropy as the factors contributing to these terms are complex and include solvation effects. The activation enthalpy for unfolding is large and positive, as has been observed for other proteins and is attributed to the loss of favourable interactions present in the native state. The large, positive activation enthalpy for folding can be attributed to the formation of hydrophobic interactions in the transition state. It is interesting to note that the activation entropies for unfolding and refolding are both small and negative. Values for $\Delta S_F^{\ddagger}$ are frequently negative, a result of the decrease in chain entropy. Negative values for $\Delta S_U^{\ddagger}$, however, are less common, and indicate a decrease in entropy between native and transition state. For FKBP12 this may result from the increase in solvent accessibility of hydrophobic side chains in the transition state, forcing water molecules to order around exposed hydrophobic groups.

The kinetics of unfolding and folding were determined as a function of denaturant concentration in the presence of the tight-binding ligand rapamycin, Figure 2c. As expected the FKBP12-rapamycin complex is more stable than free FKBP12 and

the complex unfolds many orders of magnitude more slowly. Although the exact difference in unfolding rate between free and complexed FKBP12 varies slightly with denaturant concentration, at 3.5 M GdnHCl the difference is greater than 2×10^5. In comparison, there is relatively little effect on the folding rate, Figure 2c. These data suggest that the ligand-binding site, formed by a number of long loops, is only very weakly structured in the transition state.

Trifluoroethanol and other alcohols have recently been used to probe the extent of secondary structure formation in the transition state for folding of acyl phosphatase (AcP) [2, 3]. We performed similar experiments on FKBP12 in order to study the extent of secondary formation in the transition state. The [urea]-dependence of the unfolding and refolding rate constants was measured at three different TFE concentrations: 3.6%, 9.6% and 17% (v/v), Figure 2a. For all three TFE concentrations the data fit well to a two-state model. In 3.6 and 9.6% TFE the protein folds faster and unfolds more slowly than in the absence of TFE, consistent with the equilibrium results which show that low concentrations of TFE stabilise the native state relative to the unfolded state [17]. In 17% TFE the protein folds faster but also unfolds faster than in the absence of TFE, consistent with the equilibrium results which show a slight destabilisation of the native state relative to the unfolded state under these conditions [17]. Figure 2d shows the variation in the rate of unfolding and refolding with final TFE concentration. The folding rate increases with increasing concentrations of TFE up to a maximum rate enhancement of 4-fold at 9.6% TFE, the refolding rate then decreases with increasing TFE concentration. This effect has also been observed for acylphosphatase (AcP), an α/β protein which folds with two-state kinetics [2, 3]. In this case, the rate enhancements were attributed to TFE's ability to stabilise a transition state with a relatively hydrated and disorganised core but having a significant level of secondary, particularly α-helical, structure. The data on FKBP12 suggest that either the α-helix or β-hairpin may have significant structure in the transition state.

STRATEGY 2: PROTEIN ENGINEERING METHODS AND ϕ-VALUE ANALYSIS

The protein engineering methods and ϕ-value analysis used to characterise folding transition states have been described extensively elsewhere [4]. First, a non-disruptive single amino-acid substitution is introduced into the protein to produce a mutant protein. The thermodynamic stability, unfolding and refolding kinetics of the mutant are then measured and compared to those of the wild-type protein. The effect of the mutation on the folding pathway is assessed quantitatively using ϕ-value analysis: the difference in the change in free energy between the transition state and denatured state between mutant and wild type ($\Delta\Delta G_{\ddagger-U}$) can be calculated, and is compared to the difference in the change in free energy between the native and denatured state, between mutant and wild-type ($\Delta\Delta G_{F-U}$). The ϕ-value is simply the ratio of these two differences in free energy and is therefore a measure of the extent to which interactions that the mutated side chain makes in the

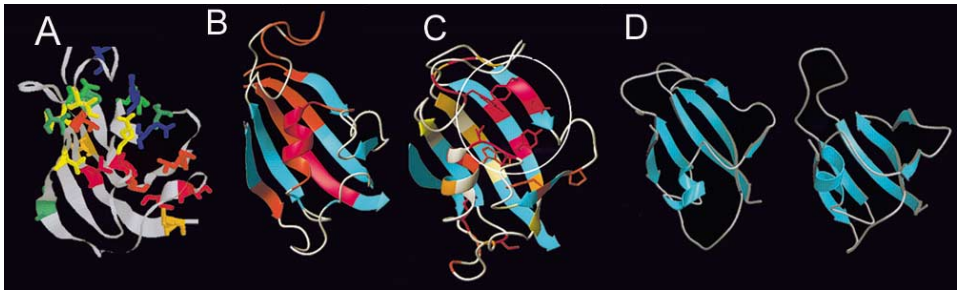

Figure 3. Results of the ϕ-value analysis. (A) ϕ-values are colour coded from dark blue (ϕ_F = 0), green (ϕ_F = 0.2), yellow (ϕ_F = 0.3), orange (ϕ_F = 0.4), to red (ϕ_F = 0.6). (B) Residual structure in the urea- and guanidinium chloride-denatured state of FKBP12 as determined by heteronuclear NMR NOE data [10]. αN (i, i+2) and αN (i, i+3) NOEs are shown in red. αN (i, i+2) are shown in orange. (C) Conserved residues: red >90%, red-orange >80%, dark-orange >70%, orange >60%, gold >50%, yellow >40%, khaki >30%, pale yellow >20%, beige >10%, white >0%. (D) Comparison of results from high-temperature molecular dynamic simulations and experimentally-determined ϕ-values. Snapshot of TS21 at 150 ps, snapshot of TS2 at 118 ps.

native state are present in the transition state. For example, $\phi = 0$ when $\Delta\Delta G_{\ddagger-U} = 0$, i.e., the mutation has no effect upon the energy of the transition state relative to the unfolded state. In this case, it can be inferred that the transition state is unstructured in the region of the mutation. Conversely, $\phi = 1$ when $\Delta\Delta G_{\ddagger-U} = \Delta\Delta G_{F-U}$, i.e., the interaction energy lost upon mutation is the same in the native and transition states. This suggests that the transition state is highly structured in the region of the mutation. Fractional ϕ-values are more difficult to interpret and can arise from a number of different situations. For example, if a protein folds by parallel pathways in which regions of the protein are native-like in the transition state of one pathway (i.e., $\phi = 1$) but unfolded in the transition state of another pathway (i.e., $\phi = 0$). Brønsted analysis of the unfolding and equilibrium data for FKBP12, however, suggests that the transition state is an ensemble of states close in structure [20]. Even given a single dominant folding pathway, the analysis of fractional ϕ-values can be complicated when there are changes in solvation energy. However, in the case where a nonpolar side chain is replaced by another nonpolar side chain and where water does not enter the site of mutation, there should be an approximately linear relationship between the ϕ-value and the extent of formation of nonpolar contacts in the transition state relative to the native state [6, 21].

CHARACTERISATION OF THE TRANSITION STATE FOR FOLDING BY PROTEIN ENGINEERING METHODS

Thirty four mutations were made at positions throughout the protein to probe the extent of secondary and tertiary structure formation in the transition state for folding of FKBP12. Equilibrium experiments measuring $\Delta\Delta G_{F-U}$ have been pub-

Table II. Φ-value analysis: Summary of results, positions and interactions of mutated side chains

Mutant	Position	Primary interactions	% Solvent accessible surface area of side chain
Val→Ala2	β1	interaction between β1 and β4	3
Val→Ala4	β1	interaction between β1 and β4, loop5	4
Ile→Val7	β1	interaction between β1 and β4	41
Thr→Ser21	β1	interaction between β2 and β3/β5	17
Val→Ala21	β2	interaction between β2 and β3/β5 and C-terminus	17
Val→Ala23	β2	interaction between β2 and β3/β5	5
Val→Ala24	β2	interaction between β2 and β2/β3/β5 and α	0
Thr→Ser27	β2	interaction between β2 and β3	28
Val→Ala27	β2	interaction between β2 and β3/β5 and β2/loop3	28
Phe→Ala36	β2	mainly interaction between β2 and loop6, to a lesser degree with β and to a lesser degree β3	10
Leu→Ala50	β3	mainly β3-α but also β3, C-cap α-helix and loop 1	0
Val→Ala55	loop	β3 and α-helix	9
Ile→Ala56	N-cap	α, loop6, β4, β5	10
Ala→Gly57	α-helix	α-helix formation	19
Ala→G1y60	α-helix	α-helix formation	17
Ala→G1y61	α-helix	α-helix formation	4
Val→Ala63	α-helix	α-helix formation and tertiary interactions between α-helix and loops 1&6, β2, β4, β5	0
Val→Ala75	β4	interactions within β4 and between β4 and β1/β5	18
Ile→Va176	β4	mainly between β4 and α-helix, also β4/β5	0
Ile→Ala76	β4	similar, also β4-β1	0
Val→Ala76	β4	β4 and β1, β4, α-helix and loop6	0
Ile→Val91	loop	probes mainly the structure of extended loop6, also loop6-β3 interactions	1
Ile→Ala91	loop	probes mainly the structure of extended loop6, also loop6-β3 interactions	1
Val→Ala91	loop	mainly probes interactions within loop6	1
Leu→Ala97	β5	mainly the interaction within β5 and between β5-loop6	0
Val→Ala98	β5	interaction between β5 and β2/β4	8
Val→Ala101	β5	interaction between β5 and β2 and α-helix	0
Leu→Ala106	β5	interaction between β5 and loop1	6

Data taken from [20, 22]. Only ϕ-values for non-polar to non-polar substitutions are reported here. These ϕ-values report quantitatively on the interactions formed in the transition state [4].

lished elsewhere [22]. The unfolding and refolding rate constants for wild-type and mutant FKBP12 were determined over a range of denaturant concentrations and these data used to calculate $\Delta\Delta G_{\ddagger-U}$ [20]. ϕ-values for the mutants can be calculated from these data and are summarised in Table II. The results are shown graphically in Figure 3a.

The ϕ-values vary between 0 and 0.6. Thus, no mutated residue is in a region that is completely structured in the transition state, where one would expect the side chain to make the same interactions as in the native state and the ϕ-value to be 1. Conversely, some of the mutated residues are in regions which must be largely unstructured in the transition state as, in these cases, the ϕ-values are 0. As discussed above, for those mutated residues which have fractional values, this indicates that they are in regions of the protein which are partially structured in the transition state. For the types of mutation made in this study the ϕ-value really reflects the extent to which favourable interactions present in the native state have formed in the transition state. Thus, for Val 101 which has the highest value at 0.6, 60% of the interaction energy this side chain makes in the native state is present in the transition state. Therefore, the ϕ-values can be used directly as a guide of the extent of structure formation in the transition state.

Although the non-mutagenic studies show that the transition state is quite compact, the protein engineering results establish that there are a number of regions which are largely unstructured in the transition state. These include the N-terminus of the α-helix and many of the loops. Although no region of structure is fully formed in the transition state, regions of the β-sheet are comparatively structured, the structure centering around β-strands 2, 4, and 5. In particular Val 101, which is one of the most highly buried residues and located in the middle of the central β-strand (β-strand 5), makes approximately 60% of its native interactions. The outer β-strands, and the ends of the central β-strands are formed to a lesser degree. The data are consistent with a nucleation-condensation mechanism of protein folding in which the nucleus, which is partially formed in the transition state, comprises side chains within β-strands 2, 4 and 5 and the C-terminus of the α-helix. These residues are distant in the primary sequence, demonstrating the importance of tertiary interactions in forming and stabilising the transition state.

COMPARISON OF RESIDUAL STRUCTURE IN THE DENATURED STATE AND STRUCTURE IN THE TRANSITION STATE

The structure of the urea-and GdnHCl-denatured state of FKBP12 has been studied by NMR spectroscopy [10]. Although there is extensive conformational averaging in the unfolded state, as shown by chemical shift data, vicinal coupling constants, relaxation times and amide proton exchange rates, all of which were as expected for a random coil, a small number of medium-range ($i \rightarrow i + 2$) and ($i \rightarrow i + 3$) Nuclear Overhauser Enhancement crosspeaks (NOEs) were also observed. These are indicative of the transient formation of secondary structure in the unfolded

state. In particular, there is some weak helical structure in the region of the α-helix and β-strand 5, see Figure 3b [10]. Residual structure in denatured states may play an important role in the early stages of folding by directing the protein along a particular pathway. However, we see little correlation between the residues that are in partially structured regions in the transition state, as determined from the ϕ-value analysis, and residues involved in regions with transient structure in the denatured state. In particular, although there is evidence for the transient formation of native-like structure in the α-helix in the denatured state, the helix appears to be largely unstructured in the transition state for folding.

COMPARISON OF STRUCTURE IN THE TRANSITION STATE AND SEQUENCE CONSERVATION

It has been suggested that the folding nuclei of proteins are likely to be highly conserved. A comparison of primary sequences shows that 23 out of the 107 residues are invariant in all known FKBPs. Sequence conservation in the FKBP family is shown graphically in Figure 3c. Comparing sequence conservation with ϕ-values we see little correlation. For example, Thr 21, Val 23 and Val 24 on β-strand 2 have some of the highest ϕ-values for FKBP12, yet are not conserved. Phe 36, Val 55, Ile 56 and Ile 91, on the other hand, which all have very low ϕ-values, are completely conserved (probably due to functional requirements – they all bind to the immunosuppressant ligands FK506 and rapamycin). In contrast, Val 101, which has one of the highest ϕ-values, is invariant and Val 63, whose ϕ-value is also moderately high, is also highly conserved. This suggests that it may be difficult to predict the folding nucleus on the basis of sequence homology.

5. Computational Approaches to Studying Folding Transition States

COMPARISON OF THE TRANSITION STATE DETERMINED EXPERIMENTALLY AND HIGH-TEMPERATURE MOLECULAR DYNAMIC SIMULATIONS

Two high temperature (498 K) denaturation simulations were performed in water, beginning from the crystal (D-1 [11]) and NMR (D-2 [12]) structures. In addition, a control simulation at 298 K was performed beginning with the crystal structure. Details of the simulations are given elsewhere [20]. At high temperature, the protein expanded rapidly and reached a Cα root-mean-square (RMS) deviation of 8 and 12 Å within 2 ns, for the D-1 and D-2 simulations, respectively. In comparison, the protein remained within 2.2 Å of the starting structure in the control simulation. Putative transition state structures were identified in the denaturation simulations using a conformational clustering method described previously [6, 8, 23]. The transition state ensembles comprised of 25 structures each from 150–155 ps in the simulation beginning from the crystal structure (TS1) and 115–120 ps beginning with the NMR structure (TS2). The average structures from these time

periods are shown in Figure 3d. The two transition state ensembles are structurally similar with a Cα RMS deviation of 5 Å.

Comparison of the computational results, where only two unfolding simulations are performed, with experimental data, in which approximately 10^{15} molecules are studied, is obviously non-ideal. However, even for a relatively small protein such as FKBP12, the simulations are so computationally-demanding it is only realistic to perform a small number. Despite this, there is good agreement between experimental and computational results.

The simulated transition state structures contain partial structure. The central core of the β-sheet is fairly structured with fraying of the sheet at the edges, even for the more well-ordered strands. The α-helix is only weakly structured with some retention of structure at the C-terminus, particularly in TS2. The loops are all disordered. The simulations are in good agreement with the experimental results. For a more quantitative comparison see [20].

6. Comparison of Experimental Approaches: An Investigation of the Effect of TFE on the Folding Pathway of FKBP12

In general, there is good agreement between results obtained using the two experimental approaches. However, there is a discrepancy between results from protein engineering experiments, which show that the α-helix is largely unstructured in the transition state and no β-hairpins are fully formed, and results obtained with TFE that suggest that the α-helix or β-hairpin may have significant structure in the transition state. These conflicting results imply that either there has been a change in the folding pathway on addition of TFE or that TFE does not act by stabilising secondary structure in the transition state. In order to investigate this apparent discrepancy further, we undertook a ϕ-value analysis in 9.6% TFE. Details of these experiments have been published elsewhere [24]. ϕ-values calculated for mutants in 0% and 9.6% TFE are shown in Figure 4a and 4b, respectively. Remarkably, within experimental error, there is no difference in the ϕ-values measured. These results conclusively show that TFE does not change the folding pathway of FKBP12; the transition state in the presence of 9.6% TFE still has a largely unstructured α-helix and has no fully structured β-hairpins. Thus, the rate enhancement observed for the folding of FKBP12 in 9.6% TFE cannot be attributed to stabilisation of the α-helix

Figures 4 and 5. →

Figure 4. Comparison of ϕ-values in water and 9.6% TFE. (A) Transition state for folding of FKBP12 in water (ϕ-values between 0–0.1 are shown in green, 0.2–0.4 yellow, 0.4–0.5 orange and 0.5 and above red). (B) Transition state for folding of FKBP12 in 9.6% TFE (ϕ-values between 0–0.1 are shown in green, 0.2–0.4 yellow, 0.4–0.5 orange and 0.5 and above red).

Figure 5. Free energy diagrams for the folding of FKBP12. 0% (black lines) and 9.6% (red lines) TFE. The unfolded state, transition state and native state are designated U, ‡ and F respectively.

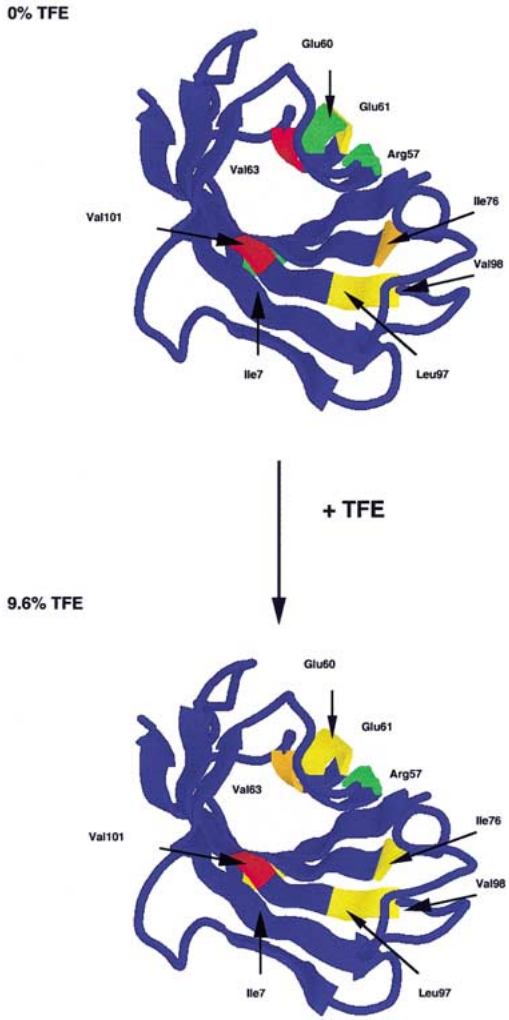

Figure 4.

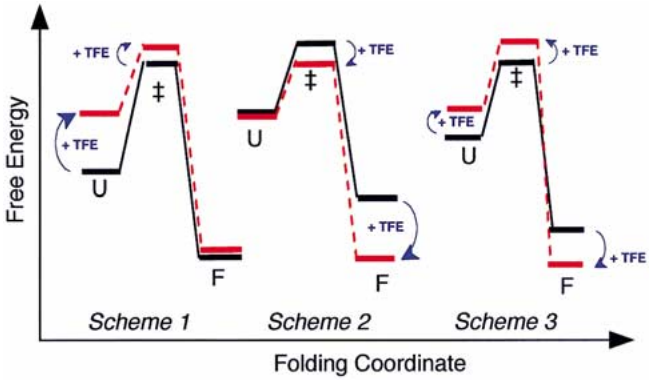

Figure 5.

or β-hairpin as has been reported elsewhere [2, 3]. Instead, these results suggest that TFE has a global effect on the energetics of the protein structure rather than specifically stabilising any individual element of secondary structure. In a recent paper by Kentsis et al. [25] TFE was reported to increase the refolding rate of an α-helical protein. Their analysis of two mutants suggested that TFE affects the unfolded ensemble, destabilising it relative to the transition state and native state. We have now tested the generality of this conclusion in a structurally very different protein, FKBP12, which contains both α and β structural elements. By analysing mutations throughout different regions of secondary and tertiary structure we have shown that TFE does not affect any single element of structure in the transition state.

Ideally, we would like to be able to assign the differences in rates observed in TFE to changes in the energy of a particular state on the folding pathway. Unfortunately, this is not possible as TFE affects many different structural and physical properties of the system, as recently pointed by Ionescu and Matthews [26]. For example, TFE may weaken hydrophobic interactions, strengthen intramolecular hydrogen bonds, change the dielectric constant of the solvent, and may act as an osmolyte. It is not possible to distinguish how these factors affect the energy of the unfolded, transition and native states, Figure 5. For example, one interpretation of our data is that TFE mainly acts as an osmolyte, disrupting the water structure surrounding the protein. In this case, we expect the unfolded state to be most destabilised, the transition state to a lesser extent and the native state the least (scheme 1, Figure 5). This agrees well with recent work on the mechanism of helix induction by TFE, where it is proposed that, at [TFE] $<$20%, TFE stabilises helical structure by disordering the local hydration shell around the polypeptide in a chaotropic manner [27]. Furthermore, there is a growing body of evidence that supports the conclusion that low concentrations of TFE affect the solvation of the protein [25, 28].

7. Discussion

There are still relatively few proteins for which ϕ-value analysis has been used to analyse the structure of the transition state for folding. The best characterised are barnase [29–31], CI2 [32], the SH3 domain from α-spectrin [33] and src [34] cheY [35] and the activation domain of procarboxypeptidase (ADAh2) [36]. The folding pathway of FKBP12 shares many similarities with that of CI2. Both fold with two-state kinetics according to a nucleation-condensation mechanism in which there is only one nucleus formed in the transition state [20, 32]. It is interesting to compare FKBP12 and CI2 with the pathways proposed for barnase and cheY which show three-state folding kinetics. The latter have been proposed to fold according to a nucleation-condensation mechanism with multiple nucleation sites [32, 35]. FKBP12 is 107 residues, compared with 64 for CI2, 110 for barnase and 129 for cheY. One might therefore expect the folding of FKBP12 to fold with mul-

tiple nucleation sites in a manner similar to barnase and cheY. However, whereas barnase and cheY contain subdomains (regions of structure which have more close contacts within that region than with the rest of the protein), FKBP12 and CI2 have a single-domain structure. This suggests that folding nuclei result from the subdomain structure, not the overall tertiary structure.

Both experimental and computational results indicate that the α-helix is largely unstructured in the transition state. In particular, the N-terminus of the helix is not formed at all, however, there is some evidence for weak structure at the C-terminus. This is in contrast to the results obtained for many other α/β or $\alpha+\beta$ proteins for which the helices are often partially or completely structured in the transition state. For example, it is known from ϕ-value analysis and MD that one and possibly both of the α-helices in barnase are structured in the transition state [8, 31, 37], the α-helix in CI2 is partially structured [5, 6, 32] in the transition state, and H/D experiments have shown that amide protons are protected from exchange early on the pathway of folding in the α-domain of lysozyme [38, 39]. A peptide corresponding to the helical region of FKBP12 has been shown to have no structure in water, however, this peptide does adopt a helical conformation in TFE [40]. Results comparing the structure of this peptide with peptides corresponding to other regions of the protein indicate that it has a higher tendency towards adopting nonrandom conformations [40]. These results have been interpreted as suggesting that the helix forms early during folding, however, the results presented here do not support this view. This discrepancy highlights the dangers of searching for initiation sites of folding in short peptides.

As might be expected, and has been observed for other proteins, loops are generally unstructured in the transition state and only form very late on the folding pathway after the rate-limiting transition state.

The data for FKBP12 are consistent with the nucleation-condensation mechanism of protein folding, as proposed initially for C12 [14, 32, 41]. In this model there is a search of conformations in the unfolded state of the protein until sufficient tertiary interactions are formed to stabilise certain elements of structure. When sufficient interactions have been made, the transition state is reached and there follows a rapid formation of the final structure. The nucleus of this folding event does not form stable structure in the absence of other interactions but requires the presence of various long-range interactions to form, i.e., the formation of the folding nucleus is coupled with more general formation of structure. The folding nucleus is simply the best formed part of the structure in the transition state. For FKBP12 the folding nucleus appears to involve residues within β-strands 2, 4 and 5 and the C-terminus of the α-helix. These residues make approximately 50–60% of the interactions present in the native state in the transition state. Thus, as with CI2, the folding nucleus is in the process of being formed as the transition state is reached.

It has been argued that the concomitant formation of the folding nucleus and transition state is preferable to the early formation of a stable nucleus which may

lead to a decrease in the overall rate of folding [14]. In contrast to the nucleation model of folding proposed by Wetlaufer [42, 43] formation of the folding nucleus in the nucleation-condensation mechanism need not be rate-limiting since a significant fraction of the overall structure must be in place in order to provide the favourable long-range interactions necessary to form the nucleus, as depicted in Figure 3a.

It is now common practice to search for initiation sites in the folding of proteins by searching for structure in isolated peptides. Our results on FKBP12 suggest that, at least for this structure, this is a fruitless task as the residues most important in stabilising the transition state are distant in primary sequence. For FKBP12, therefore, there is little correlation between transient structure in peptides and the folding pathway. For other proteins, however, particularly those which have more significant structure in isolated peptides, there may prove to be a correlation. The extent to which residual structure either in the unfolded state, or in peptides, plays a role in folding is therefore likely to vary from protein to protein and structure to structure.

8. Conclusions

We have shown that the 107-residue FKBP12 folds with simple two-state kinetics in water [17]. It folds efficiently without populating intermediate states or misfolded species which could act as kinetic traps. Thus, both its size and structure, which includes the topological crossing of two loops, do not prevent it from folding with a simple two-state mechanism. The transition state has been characterised using three different experimental and computational approaches which are in good agreement [17, 20]. Although no region of structure is fully formed in the transition state, regions of the β-sheet are comparatively structured, with the structure centering around β-strands 2, 4, and 5. The data are consistent with a nucleation-condensation mechanism of protein folding in which the nucleus, which is partially formed in the transition state, comprising side chains within β-strands 2, 4 and 5 and the C-terminus of the α-helix. These residues are distant in the primary sequence, demonstrating the importance of tertiary interactions in the transition state.

In addition, we have shown that TFE does not significantly change the folding pathway of FKBP12 [24]. However, although TFE affects the folding and unfolding rate of FKBP12, we have established that this effect is not caused by the stabilisation of a single element of structure [24]. Instead, the effect appears to be global in nature affecting all regions of the protein structure equally. There is evidence that other co-solvents may exert their effects by similar mechanisms [44]. Thus, an accurate interpretation of studies using TFE and other co-solvents can only be achieved by combining them with other techniques such as protein engineering.

Acknowledgements

We would like to thank Professor Alan Fersht for his support and Tom Rippin for help with the Ile → Ala7 mutant. SEJ is a Royal Society University Research Fellow, ERGM is supported by a BBSRC studentship, KFF is supported by an Elmore Scholarship, Gonville & Caius College, Cambridge. Financial support for the computational studies was provided by the National Institutes of Health (GM 50789 to VD). We are also grateful to Prof Steve Ley for kindly providing rapamycin.

References

1. Sancho, J., Meiering, E. and Fersht, A.R.: Mapping Transition States of Protein Unfolding by Protein Engineering of Ligand-Binding Sites. *J. Mol. Biol.* **221** (1991), 1007–1014.
2. Chiti, F., Taddei, N., van Nuland, N.A.J., Magherini, F., Stefani, M., Ramponi, G. and Dobson, C.M.: Structural Characterisation of the Transition State for Folding of Muscle Acylphophatase. *J. Mol. Biol.* **283** (1998), 893–903.
3. Chiti, F., Taddei, N., Webster, P., Hamada, D., Fiaschi, T., Ramponi, G. and Dobson, C.M.: Acceleration of the Folding of Acylphosphatase by Stabilisation of Local Secondary Structure. *Nat. Struct. Biol.* **6** (1999), 380–387.
4. Fersht, A.R., Matouschek, A. and Serrano, L.: The Folding of an Enzyme. 1. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding. *J. Mol. Biol.* **224** (1992), 771–782.
5. Daggett, V., Li, A.J., Itzhaki, L.S., Otzen, D.E. and Fersht, A.R.: Structure of the Transition-State for Folding of a Protein-Derived from Experiment and Simulation. *J. Mol. Biol.* **257** (1996), 430–440.
6. Li, A.J. and Daggett, V.: Identification and Characterization of the Unfolding Transition-State of Chymotrypsin Inhibitor 2 by Molecular-Dynamics Simulations. *J. Mol. Biol.* **257** (1996), 412–429.
7. Ladurner, A.G., Itzhaki, L.S., Daggett, V. and Fersht, A.R.: Synergy between Simulation and Experiment in Describing the Energy Landscape of Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **95** (1998), 8473–8478.
8. Daggett, V., Li, A.J. and Fersht, A.R.: Combined Molecular Dynamics and Phi-Value Analysis of Structure-Rreactivity Relationships in the Transition State and Unfolding Pathway of Barnase: Structural Basis of Hammond and Anti-Hammond Effects. *J. Amer. Chem. Soc.* **120** (1998), 12740–12754.
9. Egan, D.A., Logan, T.M., Liang, H., Matayoshi, E., Fesik, S.W. and Holzman, T.F.: Equilibrium Denaturation of Recombinant Human FK506 Binding Protein in Urea. *Biochemistry* **32** (1993), 1920–1927.
10. Logan, T., Theriault, Y. and Fesik, S.: Structural Characterization of the FK506 Binding-Protein Unfolded in Urea and Guanidine-Hydrochloride. *J. Mol. Biol.* **236** (1993) 637–648.
11. van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J.: Atomic-Structure of FKBP-FK506, An Immunophilin-Immunosuppressant Complex. *Science* **252** (1991), 839–842.
12. Michnick, S.W., Rosen, M.K., Wandless, T.J., Karplus, M. and Schreiber, S.L.: Solution Structure of FKBP, a Rotamase Enzyme and Receptor for FK506 and Rapamycin. *Science* **252** (1991), 836–842.
13. Jackson, S.E. and Fersht, A.R.: Folding of Chymotrypsin Inhibitor-2. 1. Evidence for a Two-State Transition. *Biochemistry* **30** (1991), 10428–10435.

14. Fersht, A.R.: Optimization of Rates of Protein Folding: The Nucleation-Collapse Mechanism for the Folding of Chymotrypsin Inhibitor 2 (CI2) and Its Consequences. *Proc. Natl. Acad. Sci. USA* **92** (1995), 10869–10873.
15. Jackson, S.E.: How Do Small Single-Domain Proteins Fold? *Folding & Design* **3** (1998), R81–R90.
16. Munoz, V. and Eaton, W.A.: A Simple Model for Calculating the Kinetics of Protein Folding from Three-Dimensional Structures. *Proc. Natl. Acad. Sci. USA* **96** (1999), 11311–11316.
17. Main, E.R.G., Fulton, K.F. and Jackson, S.E.: Folding of FKBP12: Pathway of Folding and Characterisation of the Transition State. *J. Mol. Biol.* **291** (1999), 429–444.
18. Khorasanizadeh, S., Peters, I.D. and Roder, H.: Evidence for a Three-State Model of Protein Folding from Kinetic Analysis of Ubiquitin Variants with Altered Core Residues. *Nature Struct. Biol.* **3** (1996), 193–205.
19. Chen, B., Baase, W.A. and Schellman, J.A.: Low Temperature Unfolding of a Mutant of Phage T4 Lysozyme. 2. Kinetic investigations. *Biochemistry* **26** (1989), 691–699.
20. Fulton, K.F., Main, E.R.G., Daggett, V. and Jackson, S.E.: Mapping the Interactions Present in the Transition State for Folding/Unfolding of FKBP12. *J. Mol. Biol.* **291** (1999), 445–461.
21. Jackson, S.E., el Masry, N. and Fersht, A.R.: 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** (1993), 11270–11278.
22. Main, E.R.G., Fulton, K.F. and Jackson, S.E.: The Context-Dependent Nature of Destabilising Mutations on The Stability of FKBP12. *Biochemistry* **37** (1998), 6145–6153.
23. Li, A.J. and Daggett, V.: Characterization Of The Transition-State Of Protein Unfolding By Use Of Molecular-Dynamics – Chymotrypsin Inhibitor-2. *Proc. Natl. Acad. Sci. U.S.A.* **91** (1994), 10430–10434.
24. Main, E.R.G. and Jackson, S.E.: Does Trifluoroethanol Affect Folding Pathways and Can It be Used as a Probe of Structure in Transition States? *Nature Struct. Biol.* **6** (1999), 831–835.
25. Kentsis, A. and Sosnick, T.R.: Trifluoroethanol Promotes Helix Formation by Destabilizing Backbone Exposure: Desolvation Rather than Native Hydrogen Bonding Defines the Kinetic Pathway of Dimeric Coiled Coil Folding. *Biochemistry* **37** (1998), 14613–14622.
26. Ionescu, R.M. and Matthews, C.R.: Folding under the Influence. *Nature Struct. Biol.* **6** (1999), 304–307.
27. Walgers, R., Lee, T.C. and CammersGoodwin, A.: An Indirect Chaotropic Mechanism for the Sstabilization of Helix Cconformation of Peptides in Aqueous Trifluoroethanol and Hexafluoro-2-Propanol. *J. Amer. Chem. Soc.* **120** (1998), 5073–5079.
28. Storrs, R.W., Truckses, D. and Wemmer, D.E.: Helix Propagation in Trifluoroethanol Solutions. *Biopolymers* **32** (1992), 1695–1702.
29. Matouschek, A., Kellis, J.T. Jr., Serrano, L. and Fersht, A.R.: Mapping the Transition State and Pathway of Protein Folding by Protein Engineering. *Nature* **342** (1989), 122–126.
30. Matouschek, A., Kellis, J.T. Jr., Serrano, L., Bycroft, M. and Fersht, A.R.: Transient Folding Intermediates Characterized by Protein Engineering. *Nature* **346** (1990), 440–445.
31. Serrano, L., Matouschek, A. and Fersht, A.R.: The Folding of an Enzyme. 3. Structure of the Transition State for Unfolding of Barnase Analysed by a Protein Engineering Procedure. *J. Mol. Biol.* **224** (1992), 805–818.
32. Itzhaki, L.S., Otzen, D.E. and Fersht, A.R.: The Structure of the Transition State for Folding of Chymotrypsin Inhibitor 2 Analysed by Protein Engineering Methods: Evidence for a Nucleation-Collapse Mechanism for Protein Folding. *J. Mol. Biol.* **254** (1995), 260–288.
33. Viguera, A.R., Serrano, L. and Wilmanns, M.: Different Folding Transition-States May Result In The Same Native Structure. *Nat. Struct. Biol.* **3** (1996), 874–880.
34. Grantcharova, V.P., Riddle, D.S., Santiago, J.V. and Baker, D.: Important Role of Hydrogen Bonds in the Structurally Polarized Transition State for Folding of the src SH3 Domain. *Nat. Struct. Biol.* **5** (1998), 714–720.

35. Lopez-Hernandez, E. and Serrano, L.: Structure of the Transition State for Folding of the 129 aa Protein CheY Resembles that of a Smaller Protein, CI-2. *Folding & Design* **1** (1996), 43–55.
36. Villegas, V., Martinez, J.C., Aviles, F.X. and Serrano, L.: Structure of the Transition State in the Folding Process of Human Procarboxypeptidase A2 Activation Domain. *J. Mol. Biol.* **283** (1998), 1027–1036.
37. Matthews, J.M. and Fersht, A.R.: Exploring the Energy Surface of Protein Folding by Structure-Reactivity Relationships and Engineered proteins: Observation of Hammond Behaviour for the Gross Structure of the Transition State and Anti-Hammond Behaviour for Structural Elements for Unfolding/Folding of Barnase. *Biochemistry* **34** (1995), 6805–6814.
38. Miranker, A., Radford, S.E., Karplus, M. and Dobson, C.M.: Demonstration by NMR of Folding Domains in Lysozyme. *Nature* **349** (1991), 633–636.
39. Radford, S.E., Dobson, C.M. and Evans, P.A.: The Folding of Hen Lysozyme Involves Partially Structured Intermediates and Multiple Pathways. *Nature* **358** (1992), 302–307.
40. Callihan, D. and Logan, T.M.: Conformations of Peptide Fragments from the FK506 Binding Protein: Comparison with the Native and Urea-Unfolded States. *J. Mol. Biol.* **285** (1999), 2161–2175.
41. Fersht, A.R.: Nucleation Mechanisms in Protein Folding. *Curr. Opin. Struct. Biol.* **7** (1997), 3–9.
42. Wetlaufer, D.B.: Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **70** (1973), 697–701.
43. Wetlaufer, D.B.: Nucleation in Protein Folding – Confusion of Structure and Process. *Trends Biochem. Sci.* **15** (1990), 414–415.
44. Wang, A. and Bolen, D.W.: A Naturally Occurring Protective System in Urea-Rich Cells: Mechanism of Osmolyte Protection of Proteins Against Urea Denaturation. *Biochemistry* **36** (1997), 9101–9108.
45. Kraulis, P.: Molscript, a Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J. Appl. Crystallogr.* **24** (1991), 946–950.
46. Merritt, E.A. and Murphy, M.E.P.: Raster3D Version-2.0 – A Program For Photorealistic Molecular Graphics. *Acta Crystallogr. D.* **50** (1994), 869–873.