Article No. jmbi.1999.2942 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 291, 445–461

JMB

Mapping the Interactions Present in the Transition State for Unfolding/Folding of FKBP12

Kate F. Fulton[1], Ewan R. G. Main[1], Valerie Daggett[2]* and Sophie E. Jackson[1]*

[1]*Cambridge University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW UK*

[2]*Department of Medicinal Chemistry, University of Washington, Seattle, WA 98195-7610, USA*

**Corresponding authors*

The structure of the transition state for folding/unfolding of the immunophilin FKBP12 has been characterised using a combination of protein engineering techniques, unfolding kinetics, and molecular dynamics simulations. A total of 34 mutations were made at sites throughout the protein to probe the extent of secondary and tertiary structure in the transition state. The transition state for folding is compact compared with the unfolded state, with an approximately 30 % increase in the native solvent-accessible surface area. All of the interactions are substantially weaker in the transition state, as probed by both experiment and molecular dynamics simulations. 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, β-strand 2, β-strand 4 and β-strand 5, comprise the most structured region of the transition state. In particular Val101, 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. The precise residues involved in the nucleus differ in the two simulated transition state ensembles, but the interacting regions of the protein are conserved. These residues are distant in the primary sequence, demonstrating the importance of tertiary interactions in the transition state. The two independently derived transition state ensembles are structurally similar, which is consistent with a Brønsted analysis confirming that the transition state is an ensemble of states close in structure.

Keywords: FKBP12; protein engineering; Φ-value analysis; molecular dynamics; nucleation-condensation

Introduction

In order to understand how a protein folds it is important to characterise all the states on the folding pathway, i.e. the unfolded state (U), intermediate states (I) (if present), the transition state (‡), and the native or folded state (F). X-ray crystallography and NMR spectroscopy have been used extensively in the determination of the three-dimensional structure of native proteins. More recently, heteronuclear multidimensional NMR techniques have been used to characterise the unfolded states of proteins (Logan *et al.*, 1993, 1994; Arcus *et al.*, 1994, 1995; Fiebig *et al.*, 1996; Smith *et al.*, 1996; Wong *et al.*, 1996; Schwalbe *et al.*,

E-mail addresses of the corresponding authors: sej13@cam.ac.uk and daggett@u.washington.edu

Abbreviations used: FKBP12, FK506 binding protein; MD, molecular dynamics; TS, transition state; NOE, nuclear Overhauser enhancement.

0022-2836/99/320445–17 $30.00/0

1997; Zhang & Forman-Kay, 1995, 1997; Neri *et al*., 1992; Alexandrescu *et al*., 1994; Frank *et al*., 1995; Pan *et al*., 1995; Wang & Shortle, 1995, 1996; Freund *et al*., 1996; Gillespie & Shortle, 1997a,b; Farrow *et al*., 1997). From these and other studies it has become clear that unfolded proteins are rarely true ''random coils'', but can contain regions of residual structure (Shortle, 1993). Although folding intermediates (Kim & Baldwin, 1982; Evans & Radford, 1994) and transition states have been characterised using a variety of biophysical methods and probes, it was not until the advent of H/^{2}H pulsed-quenched flow techniques (Baldwin, 1993) and protein engineering methods (Matouschek *et al*., 1989) that high-resolution structural information became available. Intermediates were thought to be essential to the folding process by helping to restrict the conformational space and direct the protein to its folded conformation. However, well-populated intermediates are not required for the fast efficient folding of a protein (Jackson & Fersht, 1991). Many small proteins have now been shown to fold with simple two-state kinetics (Jackson, 1998). There are still relatively few proteins for which the transition state has been characterised in detail. To date the transition states of seven proteins have been determined in detail using the protein engineering method of analysis. These are chymotrypsin inhibitor 2 (CI2), the SH3 domain of α-spectrin and src, the activation domain of procarboxypeptidase A2 (ADAh2) and monomeric λ repressor, which fold with two-state kinetics (Itzhaki *et al*., 1995; Viguera *et al*., 1996; Grantcharova *et al*., 1998; Villegas *et al*., 1998; Burton *et al*., 1997), and barnase and cheY (Serrano *et al*., 1992; Lopez-Hernandez & Serrano, 1996), which display three-state kinetics. The protein engineering method has also been used to characterise the transition state for folding of the dimeric arc repressor (Milla *et al*., 1995), and to probe helix formation in the transition state for folding of protein L (Kim *et al*., 1998). Here we present a detailed characterisation of the transition state for folding/unfolding of FKBP12 using a combination of protein engineering techniques, unfolding kinetics, and molecular dynamics simulations.

The structure of FKBP12 is characterised by a large, amphiphilic, antiparallel five-stranded β-sheet with $+3, +1, -3, -1$ topology. This sheet packs against a small α-helix to form the hydrophobic core (Main *et al*., 1999). FKBP12 undergoes reversible two-state denaturant-induced unfolding under equilibrium conditions (Egan *et al*., 1993). It also folds with two-state kinetics, i.e. no intermediate states are significantly populated on the folding pathway (Main *et al*., 1999). Thus, to define the pathway of folding of FKBP12 we need to characterise the unfolded (U), transition (‡) and native (F) states. The urea and guanidinium chloride (GdnHCl)-induced denatured states have been characterised using NMR spectroscopy, which indicates that, although there is some evidence for fleeting residual structure, there is no evidence for extensive structure in the unfolded state (Logan *et al*., 1994). Here we present the characterization of 34 mutants of FKBP12 to probe the interactions present in the transition state (Table 1). In addition, molecular dynamics simulations of temperature-induced denaturation were performed to obtain atomic-resolution models for the transition state ensemble. We note that the simulations and experiments were done in parallel and not compared until after completion, i.e. the simulated transition state ensembles were predictions and not fit to experiment. Characterisation of the behaviour of these mutants and the molecular models allow us to construct a detailed picture of the structure of the transition state, to investigate the correlation between structure in the transition state and residual structure in the unfolded state, and to probe the importance of conserved residues to stabilisation of the transition state.

Results

Equilibrium experiments

The results of equilibrium experiments on wild-type and mutant FKBP12 and a detailed description of the method of data analysis have been described elsewhere (Main *et al*., 1998).

Unfolding kinetics

Plots of the natural logarithm of the rate constants of unfolding against the final urea concentration for wild-type and mutants are shown in Figure 1. Most of the mutants studied remove favourable interactions and therefore destabilise the protein (Table 2, and Main *et al*., 1998). Correspondingly, the rate of unfolding of these mutants is faster than for the wild-type protein. Data were fit to a linear equation and a second-order polynomial (see Main *et al*. (1999) for equations and discussion). Values of $\ln k_U$ in water and in 3.9 M urea from the linear fit are used in the subsequent φ-value analysis. For a detailed analysis of the results obtained from different fitting methods in water and in 3.9 M urea see the Supplementary Material and Jackson *et al*. (1993).

Calculation of $\Delta\Delta G_{‡\text{-F}}$ and Φ-values

The stability of the transition state of the mutant relative to that of the wild-type protein was calculated from the unfolding kinetics as follows:

$$\Delta\Delta G_{‡\text{-F}} = -RT \ln(k_U/k'_U) \qquad (1)$$

where $\Delta\Delta G_{‡\text{-F}}$ is the difference in energy of the

Table 1. Summary of the positions and interactions of mutated side-chains

Mutant	Position	Primary interactions	% Solvent-accessible surface area of side-chain	Φ_F
Val → Ala2	β1	Interaction between β1 and β4	3	0.4
Val → Ala4	β1	Interaction between β1 and β4, loop5	4	0.3
Ile → Val7	β1	Interaction between β1 and β4	41	0.1
Thr → Ser21	β2	Interaction between β2 and β3/β5	17	0.5
Val → Ala21[a]	β2	Interaction between β2 and β3/β5 and C-terminus	17	0.5
Val → Ala23	β2	Interaction between β2 and β3/β5	5	0.5
Val → Ala24	β2	Interaction between β2 and β2/β3/β5 and α	0	0.4
Thr → Ser27	β2	Interaction between β2 and β3	28	0.4
Val → Ala27[b]	β2	Interaction between β2 and β3/β5 and β2/loop3	28	0.3
Phe → Ala36	β3	Mainly interaction between β3 and loop6, to a lesser degree with β5 and to a lesser degree β3	10	0.1
Leu → Ala50	β3	Mainly β3-α but also β3, C-cap α-helix and loop 1	0	0.4
Val → Ala55	Loop	β3 and α-helix	9	0.1
Ile → Ala56	N-cap	α, loop6, β4, β5	10	0.2
Ala → Gly57[c]	α-helix	α-Helix formation	19	0
Ala → Gly60[d]	α-helix	α-Helix formation	17	0
Ala → Gly61[d]	α-helix	α-Helix formation	4	0.3
Val → Ala63	α-helix	α-Helix formation and tertiary interactions between α-helix and loops 1&6, β2, β4, β5	0	0.5
Val → Ala75[f]	β4	Interactions within β4 and between β4 and β1/β5	18	0.1
Ile → Val76	β4	Mainly between β4 and α-helix, also β4/β5	0	0.4
Ile → Ala76	β4	Similar, also β4-β1	0	0.3
Val → Ala76[g]	β4	β4 and β1, β4, α-helix and loop6	0	0.3
Ile → Val91	Loop	Probes mainly the structure of extended loop6, also loop6-β3 interactions	1	0.3
Ile → Ala91	Loop	Probes mainly the structure of extended loop6, also loop6-β3 interactions	1	0
Val → Ala91[h]	Loop	Mainly probes interactions within loop6	1	0
Leu → Ala97	β5	Mainly the interaction within β5 and between β5-loop6	0	0.2
Val → Ala98	β5	Interaction between β5 and β2/β4	8	0.3
Val → Ala101	β5	Interaction between β5 and β2 and α-helix	0	0.6
Leu → Ala106	β5	Interaction between β5 and loop1	6	0.35

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 (Fersht *et al.*, 1992). The following substitutions were calculated using a fine-structure analysis (Jackson *et al.*, 1993), see the Supplementary Material.

[a] Val → Ala21, the ϕ-value is calculated from data for Thr → Val21 and Thr → Ala21.

[b] Val → Ala27, the ϕ-value is calculated from data for Thr → Val27 and Thr → Ala27.

[c] Ala → Gly57, the ϕ-value is calculated from data for Arg → Ala57 and Arg → Gly57.

[d] Ala → Gly60, the ϕ-value is calculated from data for Glu → Ala60 and Glu → Gly60.

[e] Ala → Gly61, the ϕ-value is calculated from data for Glu → Ala61 and Glu → Gly61.

[f] Val → Ala75, the ϕ-value is calculated from data for Thr → Val75 and Thr → Ala75. Thr75 makes some electrostatic interactions in the transition state involving the OH group, thus leading to a high *S* and Φ-value for Thr → Val75, see Table 2. The Val → Ala Φ value shown here is low as the methyl group cannot make the same interactions as the OH group. The Thr → Val75 Φ value is therefore used in the comparison between *S* and Φ-values.

[g] Val → Ala76, the ϕ-value is calculated from data for Ile → Val76 and Ile → Ala76.

[h] Val → Ala91, the ϕ-value is calculated from data for Ile → Val91 and Ile → Ala91.

transition state of unfolding relative to the folded state between the wild-type and mutant, and k_U and k'_U are the rate constants of unfolding for the wild-type and mutant, respectively. The value for $\Delta\Delta G_{‡\text{-}F}$ is normalised against the value for $\Delta\Delta G_{U\text{-}F}$, the difference in energy between the unfolded and native states for wild-type and mutant. The ratio of these values is called the Φ-value. In particular, we focus on the Φ-value for folding, Φ_F:

$$\Phi_F = \Delta\Delta G_{‡\text{-}U}/\Delta\Delta G_{U\text{-}F} \quad (2)$$

where $\Delta\Delta G_{‡\text{-}U}$ is the difference in energy of the transition state relative to the unfolded state between the wild-type and mutant. For proteins that fold with two-state kinetics, $\Delta\Delta G_{‡\text{-}F} + \Delta\Delta G_{‡\text{-}U} = \Delta\Delta G_{U\text{-}F}$, so that $\Phi_F = 1 - \Delta\Delta G_{‡\text{-}F}/\Delta\Delta G_{U\text{-}F}$. Here, Φ_F is calculated using equation (2) and the unfolding data. Values of Φ_F calculated using values of $\Delta\Delta G_{‡\text{-}F}$ calculated from a linear fit of the unfolding data are shown in Table 2 (see Supplementary Material for the results for the second-order polynomial fit).

Interpretation of Φ-values

An extensive description of the interpretation of Φ-values and the assumptions and limitations of this method have been presented (Fersht *et al.*,

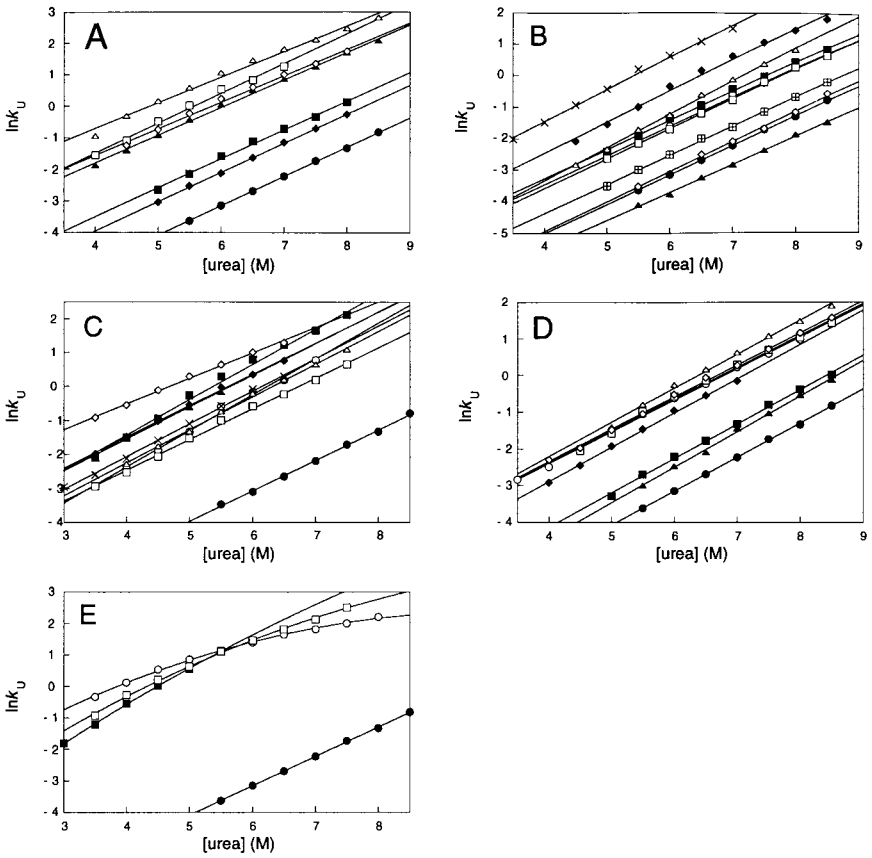

Figure 1. Unfolding kinetics of wild-type and mutant FKBP12. Unfolding rate constant as a function of [urea], in 50 mM Tris-HCl (pH 7.5), 1 mM DTT. (a) The α-helix mutants: wild-type (filled circle), EA61 (filled square), EG61 (open square), EA60 (filled triangle), EG60 (open triangle), RA57 (filled diamond), RG57 (open diamond). (b) The β-sheet mutants: wild-type (filled circle), IV7 (open circle), TA21 (filled square), TS21 (open square), TV21 (filled triangle), TS27 (open triangle), TA27 (filled diamond), TV27 (open diamond), TA75 (cross), TV75 (crossed square). (c) Hydrophobic core mutants: wild-type (filled circles), VA2 (open circle), VA4 (filled square), VA23 (open square), VA24 (filled triangle), LA50 (open triangle), IA56 (filled diamond), ID56 (open diamond), IT56 (cross). (d) Hydrophobic core mutants: wild-type (filled circles), VA63 (open circle), IV76 (filled square), IA91 (open square), IV91 (filled triangle), VA98 (open triangle), VA101 (filled diamond), LA106 (open diamond). Continuous line indicates the best fit of the data to a linear equation. (e) Very destabilising hydrophobic core mutants: wild-type (filled circles), FA36 (open circles), IA76 (filled square), LA97 (open square). The continuous line shows the best fit of the data to a second-order polynomial equation.

1992). $\Phi_F = 0$ when $\Delta\Delta G_{‡\text{-}U} = 0$, i.e. when the mutation has no effect upon the energy of the transition state relative to the unfolded state. In this case, the transition state is unstructured in the region of the mutation. Conversely, $\Phi_F = 1$ when $\Delta\Delta G_{‡\text{-}U} = \Delta\Delta G_{U\text{-}F}$, i.e. the interaction energy lost upon mutation is the same in the native and transition states. This result suggests that the transition state is highly structured in the region of the mutation. Fractional Φ-values are more difficult to interpret.

A number of different situations can give rise to fractional Φ-values. One example would be a protein that folds by parallel pathways in which portions of the protein are native-like in the transition state of one pathway (i.e. $\Phi_F = 1$), but unfolded in the transition state of another pathway (i.e. $\Phi_F = 0$). Brønsted analysis of the unfolding and equilibrium data for FKBP12, however, suggests that the transition state is an ensemble of states close in structure. (See Results; single *versus* parallel pathways). 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 non-polar side-chain is replaced by another non-polar 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 non-polar contacts in the transition state relative to the native state, as has been observed for CI2 (Jackson & Fersht, 1993; Li & Daggett, 1996).

Table 2. Unfolding data and ϕ-value analysis

Mutant	Equilibrium data $\Delta\Delta G_{U\text{-}F}$ (kcal mol^{-1})	$\ln k_U^{H_2O}$	$m_{\ddagger\text{-}F}$ (M^{-1})	$\ln k_U^{3.9\ M\ urea}$	Φ_F (0 M urea)	Φ_F (3.9 M urea)
Wild-type		-8.46 ± 0.16	0.89 ± 0.02	-4.97 ± 0.08		
Val → Ala2	2.43 ± 0.12	-6.62 ± 0.22	1.06 ± 0.04	-2.47 ± 0.07	0.55 ± 0.09	0.39 ± 0.03
Val → Ala4	2.78 ± 0.10	-5.60 ± 0.16	1.04 ± 0.03	-1.54 ± 0.06	0.39 ± 0.03	0.27 ± 0.01
Ile → Val7	0.92 ± 0.11	-7.15 ± 0.14	0.92 ± 0.02	-3.55 ± 0.06	0.16 ± 0.03	0.09 ± 0.01
Thr → Ala21	1.60 ± 0.11	-6.94 ± 0.10	0.92 ± 0.02	-3.36 ± 0.04	0.44 ± 0.06	0.40 ± 0.04
Thr → Ser21	1.44 ± 0.11	-7.37 ± 0.02	0.95 ± 0.003	-3.67 ± 0.04	0.55 ± 0.09	0.47 ± 0.05
Thr → Val21	-0.86 ± 0.13	-9.03 ± 0.08	0.89 ± 0.01	-5.56 ± 0.05	0.61 ± 0.2	0.59 ± 0.1
Val → Ala23	2.97 ± 0.10	-6.19 ± 0.03	0.92 ± 0.01	-2.56 ± 0.04	0.55 ± 0.04	0.52 ± 0.03
Val → Ala24	3.19 ± 0.11	-5.43 ± 0.04	0.97 ± 0.01	-1.64 ± 0.04	0.44 ± 0.03	0.38 ± 0.02
Thr → Ala27	1.97 ± 0.17	-6.40 ± 0.20	0.99 ± 0.03	-2.56 ± 0.09	0.38 ± 0.06	0.28 ± 0.03
Thr → Ser27	$1.49 + 0.11$	-7.54 ± 0.10	1.05 ± 0.02	-3.44 ± 0.04	0.63 ± 0.1	0.39 ± 0.04
Thr → Val27	-0.23 ± 0.12	-8.73 ± 0.19	0.95 ± 0.03	-5.03 ± 0.09	0.31 ± 0.3	0.85 ± 2.00
Phe → Ala36	3.54 ± 0.15	-2.01 ± 0.19	0.55 ± 0.03	0.13 ± 0.07	-0.08 ± 0.01	0.15 ± 0.01
Leu → Ala50	2.57 ± 0.11	-6.12 ± 0.10	0.97 ± 0.02	-2.34 ± 0.04	0.46 ± 0.04	0.39 ± 0.02
Val → Ala55	2.13 ± 0.13	-5.30 ± 0.22	0.93 ± 0.04	-1.67 ± 0.09	0.12 ± 0.01	0.08 ± 0.01
Ile → Ala56	2.48 ± 0.12	-5.17 ± 0.09	0.92 ± 0.02	-1.58 ± 0.03	0.21 ± 0.02	0.19 ± 0.01
Ile → Thr56	1.81 ± 0.13	-5.92 ± 0.08	0.96 ± 0.02	-2.16 ± 0.02	0.17 ± 0.02	0.08 ± 0.01
Ile → Asp56	3.16 ± 0.20	-3.54 ± 0.10	0.76 ± 0.02	-0.60 ± 0.03	0.08 ± 0.01	0.18 ± 0.01
Arg → Ala57	0.81 ± 0.11	-7.66 ± 0.06	0.93 ± 0.01	-4.05 ± 0.02	0.42 ± 0.1	0.33 ± 0.05
Arg → Gly57	2.29 ± 0.10	-4.94 ± 0.16	0.84 ± 0.03	-1.65 ± 0.07	0.09 ± 0.01	0.14 ± 0.01
Glu → Ala60	2.13 ± 0.11	-5.32 ± 0.10	0.88 ± 0.02	-1.89 ± 0.04	0.13 ± 0.01	0.14 ± 0.01
Glu → Gly60	2.84 ± 0.12	-3.97 ± 0.19	0.82 ± 0.03	-0.79 ± 0.08	0.06 ± 0.004	0.13 ± 0.01
Glu → Ala61	0.84 ± 0.11	-7.17 ± 0.17	0.92 ± 0.03	-3.59 ± 0.07	0.091 ± 0.02	0.03 ± 0.004
Glu → Gly61	2.49 ± 0.10	-5.29 ± 0.21	0.95 ± 0.04	-1.59 ± 0.07	0.25 ± 0.02	0.20 ± 0.01
Val → Ala63	2.97 ± 0.10	-5.88 ± 0.07	0.87 ± 0.01	-2.50 ± 0.03	0.49 ± 0.04	0.51 ± 0.02
Thr → Ala75	2.6 ± 0.11	-5.51 ± 0.12	1.01 ± 0.02	-1.55 ± 0.04	0.34 ± 0.03	0.24 ± 0.01
Thr → Val75	0.81 ± 0.13	-8.05 ± 0.08	0.92 ± 0.01	-4.47 ± 0.04	0.70 ± 0.30	0.63 ± 0.20
Ile → Val76	0.76 ± 0.12	-7.90 ± 0.12	0.94 ± 0.02	-4.24 ± 0.05	0.56 ± 0.2	0.43 ± 0.09
Ile → Ala76	3.81 ± 0.18	-5.30 ± 0.09	1.14 ± 0.02	-0.73 ± 0.02	0.51 ± 0.04	0.34 ± 0.02
Ile → Val91	0.38 ± 0.11	-8.33 ± 0.16	0.97 ± 0.02	-4.54 ± 0.07	0.80 ± 1	0.33 ± 0.1
Ile → Ala91	1.54 ± 0.10	-5.86 ± 0.15	0.87 ± 0.02	-2.47 ± 0.07	0.001 ± 0.00	0.04 ± 0.003
Leu → Ala97	3.56 ± 0.11	-3.38 ± 0.20	0.78 ± 0.03	-0.32 ± 0.08	0.16 ± 0.01	0.23 ± 0.01
Val → Ala98	2.16 ± 0.13	-5.94 ± 0.17	0.93 ± 0.02	-2.31 ± 0.08	0.31 ± 0.03	0.27 ± 0.02
Val → Ala101	2.75 ± 0.10	-6.64 ± 0.10	0.94 ± 0.02	-2.99 ± 0.03	0.61 ± 0.07	0.57 ± 0.03
Leu → Ala106	2.32 ± 0.11	-5.86 ± 0.07	0.88 ± 0.01	-2.43 ± 0.03	0.34 ± 0.03	0.35 ± 0.02

In some cases mutations delete several interactions in the native state and therefore may report on the integrity of different regions of structure. In other cases, the mutation may involve the substitution of a polar or charged side-chain for a non-polar side-chain, or *vice versa*. In this case, one must take into account the difference in solvation energies of the wild-type and mutant side-chains in the unfolded, transition and folded states, which complicates interpretation of fractional Φ-values. These problems can be overcome by making a series of mutations at a single position and using a fine structure analysis to calculate Φ-values of composite mutations for which Φ-values can be more easily interpreted. For example, Thr → Val and Thr → Ala mutations at position 21, 27 and 75. Fractional Φ-values are obtained for all these mutations, and as the mutations involve the substitution of polar for non-polar residues, these values are difficult to interpret. Using a thermodynamic cycle a Φ-value for the composite Val → Ala mutation can be calculated. In this case, the solvation energies of the two non-polar side-chains are very similar and the Φ-value can be interpreted in a more straightforward manner. In addition, fine structure analysis can be used to study the interactions of individual moieties within a side-chain. For example, the series of mutations Ile → Val → Ala → Gly allows us to attribute structure formation to each $C^{\delta 1}$, C^{γ}, and C^{β} methyl(ene) group within the side-chain. This analysis is important if different methyl(ene) groups in the side-chain interact with different elements of structure in the native state. The Φ-values calculated for composite mutations at positions 21, 27, 75, 76, and 91 using the fine structure analysis are shown in Table 1.

Molecular dynamics simulations of unfolding and identification of the transition state

Because of the many complications and assumptions associated with the interpretation of structural attributes from energetics and the desire for detailed structural models, molecular dynamics

Table 3. Comparison of experiment and simulation

Residue	Mutation	Φ_F-value	Side-chains within 4.5 Å in the crystal structure	Calculated *S* value	Side-chains within 4.5 Å in TS structures
Val24	Val→Ala	0.4	Cys22, Tyr26, Phe46, Trp59, Val101, Leu103,	0.5	Cys22, Tyr26, Phe48, Met66, Val101, Leu103
Leu50	Leu→Ala	0.4	Pro16, Phe48, Glu60 Ala64	0.2	Thr21, Phe48, Gln53
Val63	Val→Ala	0.5	Pro16, Cys22, Phe48, Trp59, Met66, Val101	0.4	Phe15, Pro16, Arg57, Trp59, Glu61
Ile76	Ile → Val	0.4	Val2, Ile56, Gly58,	0.3	Val2, Leu74, Tyr80,
	Ile → Ala	0.3	Trp59, Leu74, Tyr89,		Ile91, Pro93, Thr96,
	Val → Ala	0.3	Ala81, Phe99		Leu97, Phe99
Val101	Val→Ala	0.6	Val24, Tyr26, Val63, Trp59, Met66, Leu74, Phe99	0.5	Val24, Tyr26, Leu74, Met66, Phe99

(MD) simulations of the unfolding of FKBP12 were performed. Two high-temperature (498 K) denaturation simulations were performed in water, beginning from the crystal (D-1, van Duyne *et al.*, 1991) and NMR (D-2, Michnick *et al.*, 1991) structures. In addition, a control simulation at 298 K was performed beginning with the crystal structure. At high temperature, the protein expanded rapidly and reached a C^{α} 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 (Li & Daggett, 1994, 1996; Daggett *et al.*, 1998). Briefly, this method is used to compare all structures against all other structures to determine clusters of similar conformations. Then, we define the transition state region as the structures immediately at and preceding the first major conformational change (i.e. when we leave the first cluster near the native state). The clustering for the D-2 simulation and the position of the transition state are shown in Figure 2. The rationale for using the departure from the first main cluster to represent the transition state has been presented by Li & Daggett (1994, 1996). The transition state ensembles used were 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). Representative structures from these time periods are given in Figure 3. The two transition state ensembles are structurally similar with a C^{α} RMS deviation of 5 Å. The transition state structures also differ from the starting structures to the same extent (Figure 3).

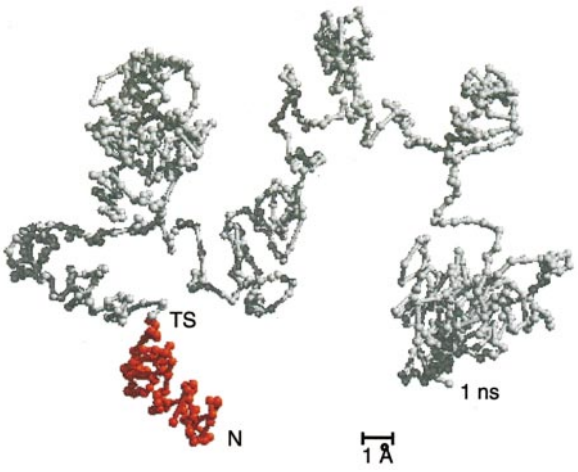

Figure 2. Projection of conformational clustering of D-2 simulation. The first cluster near the native state is coloured red. The transition state region is defined as when the protein leaves this cluster and 25 structures preceding this point are collected and used as the transition state ensemble.

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 (Figure 3). The loops are all disordered (Figure 3). This description is in qualitative agreement with the experimental results, but a semi-quantitative comparison is needed. To this end, we calculate a structure index along the sequence that is a composite of the tertiary contacts for a residue and its extent of native secondary structure (see Materials and Methods and Daggett *et al.*, 1996, 1998; Li & Daggett, 1998). This structure index, or *S* value, is typically between 0 and 1, as with Φ-values. We can then compare the experimental Φ-value with the structure index calculated from the simulated transition state structures (Figure 4(a)). Both the *S* and Φ-values vary along the sequence, with the highest values in the core of the β-sheet and the C terminus of the helix. The *S* and Φ-values are generally in good agreement, predicting similar extents of structure, as illustrated by the

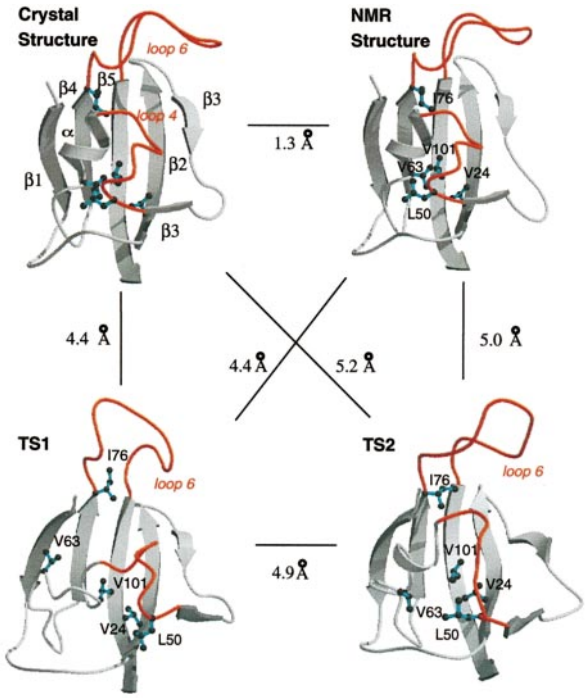

Figure 3. Main-chain traces of the native and transition state structures of FKBP12. Representative transition state structures (TS1, 153 ps; TS2, 118 ps) are displayed and the C^{α} RMS deviations are given between the transition state and experimentally determined native structures. The secondary structure is labelled in the first structure and all other structures are given in the same orientation. Some residues of interest in the hydrophobic core are displayed in cyan and labelled (the residues are not labelled in the crystal structure because of the secondary structure labels, but they are in the same orientation as in the NMR structure). Loops 4 and 6 are coloured red. The diagram was produced using the programs MOLSCRIPT (Kraulis, 1991) and Raster3D (Bacon & Anderson, 1988; Merritt & Murphy, 1994).

their placement within the horizontal lines in Figure 4(a). There is a fair degree of variance between the *S* values for TS1 and TS2 for some residues (Figure 4(a)), resulting in a correlation coefficient between the average and experimental values of 0.62, or 0.72 if the two most problematic residues (Leu50 and Glu60) are excluded. But, if TS1 and TS2 are considered interchangeable, all but Leu50 and Glu60 have values for TS1 or TS2 within 0.2 and usually <0.1 of the Φ-value (S-TS1/TS2 line in Figure 4(a)), leading to a correlation coefficient between simulation and experiment of 0.82 for all residues and 0.90 if Leu50 and Glu60 are removed from the comparison (Figure 4(b)). Representative TS1 and TS2 structures are coloured by the *S* and Φ-values in Figure 5, ranging from highly structured in blue (above the top horizontal line in Figure 4(a)) to unstructured in red (below the lower horizontal line in Figure 4(a)). In addition, the crystal structure is included for comparison and coloured by the experimental Φ-values. It is evident from the coloring of the transition state structures that despite site-specific variations in the *S* values, their overall features are similar. The transition state models are discussed in greater depth below in an attempt to help explain the experimental observables.

Structure of the transition state

The α-helix

The folding of the α-helix of FKBP12 was probed by mutation of the side-chains of Arg57, Glu60 and Glu61 to Ala and Gly. All of the Φ-values for these residues are close to zero, suggesting that the α-helix is primarily unstructured in the transition state (see Tables 1 and 2). Direct interpretation of the Φ-values of these mutants is complicated, as each mutation results in the loss of many different interactions. For example, deletion of Arg57 removes a salt bridge with Glu61 and many packing interactions as well as changing the solvation energy of the side-chain. Calculation of the composite Ala → Gly mutations at these positions circumvents some of these problems and reveals that the N terminus of the α-helix is unstructured moving towards weak structure at the C terminus. The structure of the N-cap of the α-helix, as defined by Richardson & Richardson (1988), in the transition state was also probed by mutation of Ile56. The Φ-value for Ile → Ala56 is 0.2, consistent with a largely unstructured N terminus of the helix. Val63 is the C-terminal residue of the α-helix. Mutation to Ala disrupts interactions both within the helix and between the α-helix and the β-sheet. It is therefore not possible *a priori* to use Val → Ala63 as a strict probe of helical structure. Mutation of Val63 to Gly would allow a Φ-value for the composite Ala → Gly63 to be calculated but, unfortunately, the Gly63 mutant was too destabilised to analyse (data not shown).

The calculated *S* values and experimentally derived Φ-values are in good agreement for the α-helix with the exception of Glu60 (Figure 4). Both TS1 and TS2 predict large *S* values due to retention of helical geometry at this residue, although the neighbouring residues are not as helical, and it participates in non-native electrostatic interactions. Analogous to the Ala → Gly mutation, such complications can be alleviated by calculation of a modified *S* value that excludes the side-chain atoms past C^{β}, resulting in a lower value of 0.3. Inspection of the transition state structures from the molecular dynamics simulations shows that the helix is only partially structured (Figure 3). In TS1, the main-chain of the helix is virtually unstructured with a couple of unconnected and distorted turns stabilised by packing interactions between Val63 and the non-polar portion of Arg57 and a

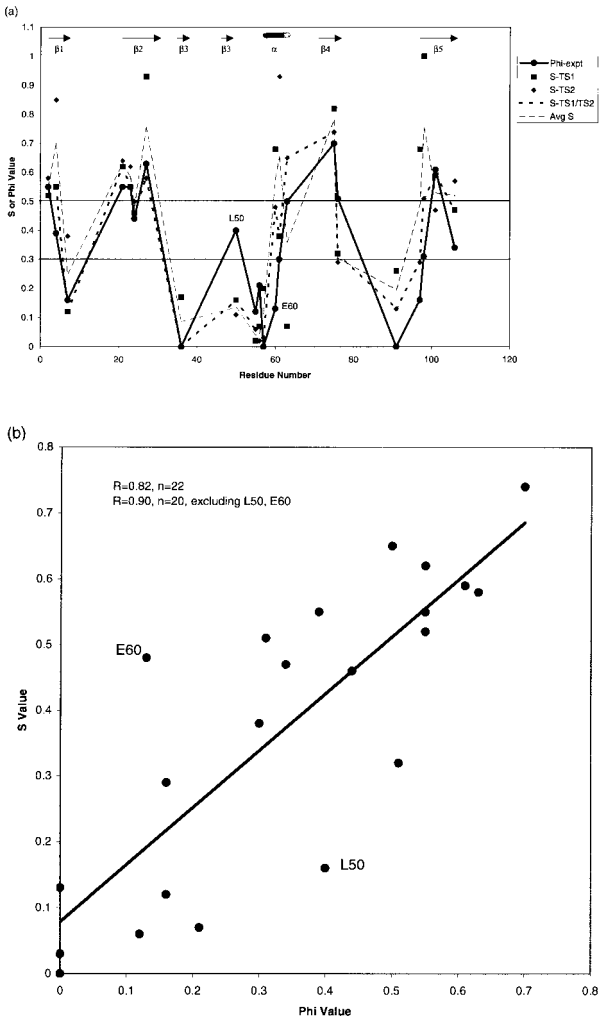

Figure 4. Comparison of structure in the transition state as probed by experiment and simulation. (a) Distribution of structure in the transition state along the sequence. The Φ_F (the 0 M urea values in Table 2 or the Ala → Gly values in Table 1) and *S* values for TS1, TS2, their average and a combination of the best values, TS1/TS2 are plotted for each mutated residue. *S* values are calculated over the simulated transition state ensembles, 25 structures from each simulation. The horizontal lines delimit regions of little to no structure in the transition state, values <0.3; partial structure, $0.3 < \Phi$ or $S < 0.5$; and more highly structured, values $\geqslant 0.5$. The Φ_F value for Thr → Val75 is used, see the footnote to Table 1. (b) Scatter plot showing the correlation between the Φ_F and *S* values.

native salt bridge between Arg57 and Glu61. In the case of TS2, the helix is more native-like, but the N terminus is extended. The C-terminal turn of the helix is intact by virtue of native packing between Trp59 and Val63 and a non-native salt bridge between Glu60 and Arg57.

The β-sheet

The structure of the β-sheet in the transition state was probed by mutation of Val2, Val4, Val23, Ile7, Thr21, Val24, Thr27, Thr75, Leu97 and Val98. These residues interact almost exclusively with other residues in the β-sheet and make few contacts with the α-helix or loops in the native state (see Table 1). The Φ-values for these residues range from 0.1-0.5 with many around 0.3-0.4 (Figure 4), suggesting that the β-sheet is weakly formed in the transition state. The residues in β-strands 2, 4 and 5, are more structured on average than the other strands, with Φ-values of ∼0.5-0.6. In comparison, the C-terminal portion of β-strand 1 and the N terminus of β-strand 5 are less structured in the transition state (Φ-values are ∼ 0.3, see Table 1 and Figure 4).

The *S* values for the simulated transition state structures are in good agreement overall for the β-strands with the exception of Val98 in β-strand 5, whose *S* value ranges from 0.5-1.0 compared with a Φ-value of 0.3. The β-strands 2, 4, and 5 represent the core of the β-sheet in the TS models (Figures 3 and 5). There is significant fraying of the structure away from this core and at the edges of even the structured strands.

The loops

There are several long loops in FKBP12 that are involved in binding the immunosuppressant ligands FK506 and rapamycin (see Figure 1 in Main *et al*. (1999)). A number of mutations probe the structure of these loops in the transition state. Phe → Ala36 probes the interactions between β-strand 3 and loop 6, and to a lesser extent β-strand 2. Val → Ala55 probes the interactions between loop 4 and β-strand 3, and the α-helix, whilst mutation of Ile91 probes the integrity of loop 6 and interactions between loop 6 and β-strand 3. In all cases, the Φ-values are close to zero, indicating that there is little to no structure. (The Φ-value for Ile → Val91 is 0.3 compared with zero for Ile → Ala91 and Val → Ala91, however, the error associated with this value is large because $\Delta\Delta G_{U\text{-}F}$ is small. Thus, the Φ-value of zero for IA91 and VA91 is more accurate.) This is consistent with the ligand-binding studies (Main *et al.*, 1999).

The MD-generated TS structures support the Φ-value analysis. The calculated S values are low for the three residues 36, 55 and 91 (Figure 4). Both loops 4 and 6, which these mutations probe, are distorted in the transition state structures and no longer tightly cap the active site (Figures 3 and 5).

The hydrophobic core

The main hydrophobic core of FKBP12 is formed by the packing of the amphipathic α-helix against the five-stranded β-sheet. The side-chains of Leu50, Val63, Ile76 and Val101 are used to probe this region. The close contacts (<4.5 Å) that these side-chains make in the native and transition states are given in Table 3. The Φ and average *S* values are

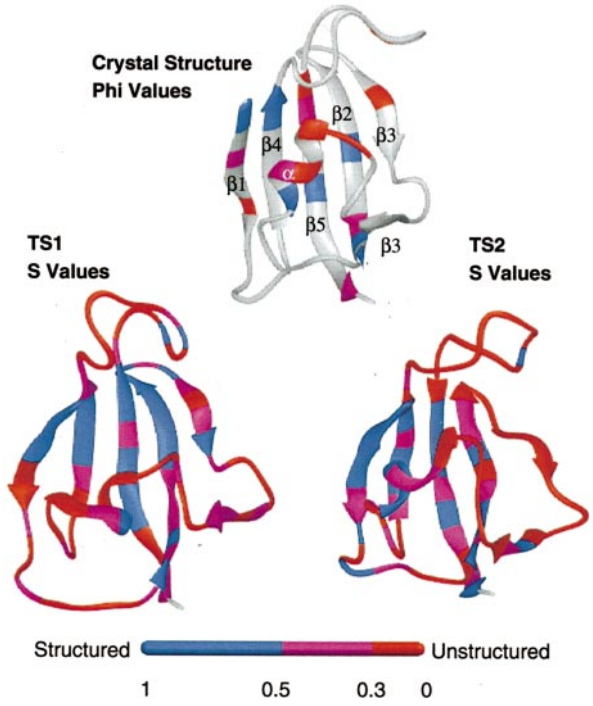

Figure 5. Graphical representation of the structure of the transition state as probed by experiment and simulation. A representative structure from the TS2 ensemble (at 118 ps) is coloured by Φ_F and S values, $0 < \Phi_F$, $S < 0.3$, red; $0.3 < \Phi_F < 0.5$, magenta; $\Phi_F > 0.5$, blue. For undetermined Φ_F-values the ribbon is coloured white. Note, due to the large number of low values for FKBP12, a different cutoff for depicting the extent of structure is being used compared with our other recent studies (see, for example, Bond *et al*., 1997).

in good agreement for these residues (Figure 4 and Table 3), with the exception of Leu50 which has a low S value.

Leu50 is in a loop connecting β-strand 3 with the α-helix. It is completely buried and makes interactions with Pro16 in loop 1, Phe48 in β-strand 3, Glu60 in the α-helix, and Ala64 which is at the C-terminal end of the α-helix in the native state. Truncation of Leu50 to Ala yields a Φ-value of 0.4 for the Leu → Ala50 mutant, which indicates that approximately 40 % of the interaction energy of the native state is maintained in the transition state. The loop containing Leu50 is disrupted in the TS structures (Figure 3) but some tertiary contacts remain, yielding a S value of 0.2. Interestingly, this residue only makes tight packing contacts with three residues in the TS structures and only one of which is native (see Table 3).

The mutation of Val → Ala63 probes the interaction between the α-helix and the β-sheet in the transition state. Truncation of Val to Ala removes favourable van der Waals' interactions with Pro16 in loop 1, Cys22 in β-strand 2, Phe48 in β-strand 3, Trp59 in the α-helix, Met66 in the loop following the α-helix, and Val101 in β-strand 5. A Φ-value of 0.5 suggests that this region of the α-helix, and in particular its interactions with the β-sheet, are partially formed in the transition state. The average S value for Val63 (0.4) is in good agreement with experiment (0.5), but the individual values for TS1 and TS2 range from 0.1 to 0.65 (Figure 4). Val63 makes primarily local interactions with other residues in the helix (see Table 3). In the case of TS1, Val63 points towards the core and β-structure but close contacts are not made (Figure 3). In TS2, Val63 rotates and points away from the core but maintains contacts with loop 1 (Figure 3).

Mutation of Ile76 probes both the interactions within the β-sheet, especially between β-strand 4 and β-strands 1 and 5, as well as the interactions between β-strand 4 and the α-helix. Most of the interactions with the α-helix, however, are with residues at its N terminus, which, as discussed above, is largely unstructured in the transition state. The Φ-value of 0.4 for Ile → Val76, and Φ-values of 0.3 for Ile → Ala76 and the composite mutation Val → Ala76 may reflect the number of interactions between β-strand 4 and β-strands 1 and 5 that are formed in the transition state in this region. This suggests that this region of the β-sheet is partially structured in the transition state. The average S value for Ile76 is 0.3. Ile76 participates in a variety of packing interactions in the two transition state ensembles; the majority of these interactions are with β-strand 5 and are both native and non-native, the other interactions with β-strand 1 and local loop residues were only native (see Table 3 and Figure 3).

The mutation Val → Ala101 also probes the interactions between the β-sheet and the α-helix by removing interactions between β-strand 5 and β-strand 2, and between β-strand 4 and Trp59 in the α-helix. Val → Ala101 has a Φ-value of 0.6, one of the highest values obtained. This residue also interacts with Val24 and Val63 in the native state and these residues have moderately high Φ-values (0.4-0.5). Val101 is in the middle of the central strand (β-strand 5) and is therefore poised in the very centre of the β-sheet and hydrophobic core. The calculated S value for Val101 is also high at 0.5-0.6. Unlike the other core residues discussed above, Val101 made extensive tertiary contacts in the TS ensembles and all were native interactions (see Table 3 and Figure 3).

Interpretation of fractional Φ-values: single *versus* parallel pathways

Whereas Φ-values of 0 or 1 can be interpreted in a straightforward manner (Fersht *et al*., 1992), fractional values of Φ are more problematic. Fractional values may result from changes in the solvation energy of the mutated side-chain in either the unfolded, transition or native state and be unrelated to structure *per se*. Consequently, we try to

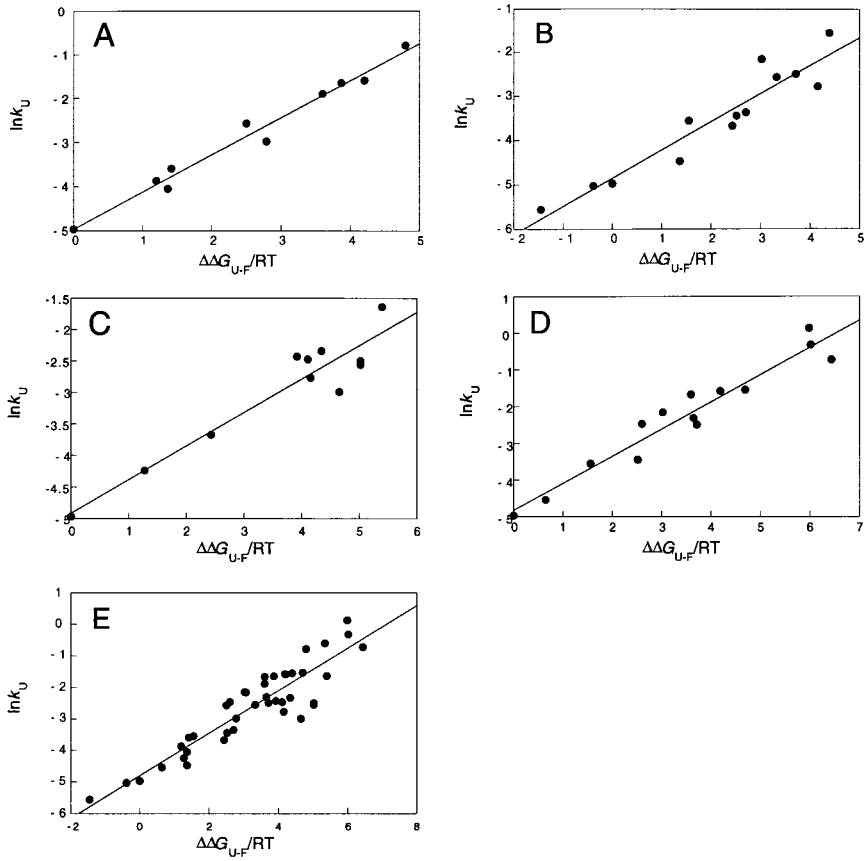

Figure 6. Brønsted analysis. Plot of $\ln k_U$ at 3.9 M urea *versus* $\Delta\Delta G_{U\text{-}F}/RT$ for (a) α-helix mutants (slope = 0.84, correlation coefficient = 0.99), (b) β-sheet mutants (slope = 0.64, correlation coefficient = 0.95), (c) hydrophobic mutations with high Φ-values (slope = 0.53, correlation coefficient = 0.95), (d) hydrophobic mutations with low Φ-values (slope = 0.74, correlation coefficient = 0.97), (e) all mutants (slope = 0.68, correlation coefficient = 0.91).

focus on Φ-values for mutations that substitute a non-polar residue for another non-polar residue, where differences in changes of solvation are small. For these cases, fractional values may result from either: (i) the structure at the site of mutation being weakened in the transition state; or (ii) a mixture of species some with the structure at the site of mutation being fully folded and others fully unfolded in the transition state, arising from parallel pathways. Whether there is a genuine single pathway or parallel pathways, which would result in multiple distinct transition states, is a fundamental question. Fersht *et al*. (1994) have proposed a kinetic test to distinguish between these possibilities. This test is based on Brønsted behaviour observed in physical organic chemistry for simple systems in which a single bond is made/broken in the transition state of the reaction. Brønsted behaviour for protein folding reactions can be analysed using equation (3):

$$\ln k_U = \ln k_U^o + (1 - \beta_F)\Delta\Delta G_{U\text{-}F}/RT \qquad (3)$$

where k_U^o is the rate constant for unfolding of the parent molecule, in this case wild-type FKBP12, k_U is the rate constant for unfolding of mutant FKBP12, and β_F is a constant which is related to the degree of structure formation in the transition state. For parallel pathways involving transition states in which interactions are either fully formed or fully broken, one would expect non-linear Brønsted behaviour. For parallel pathways we expect a change in pathway as one pathway becomes destabilised relative to another, depending on the elements of structure that are destabilised upon mutation. For a protein folding reaction to show Brønsted behaviour, plots of $\ln k_U$ *versus* $\Delta\Delta G_{U\text{-}F}/RT$ should be linear for single elements of structure. Figure 6 shows the relationship between $\ln k_U$ and $\Delta\Delta G_{U\text{-}F}/RT$ for sets of mutations in the α-helix (Figure 6(a)), β-sheet (Figure 6(b)), and the hydrophobic core for those mutations with high Φ-values (Figure 6(c)), low Φ-values (Figure 6(d)) and all mutants (Figure 6(e)). In all five cases there is a strong correlation and no indication of non-linearity, suggesting that FKBP12 folds *via* a single rate-determining transition state. The surprisingly good

fit of the data to a simple Brønsted equation should be due to two factors: (i) all the mutations probe the same element of structure, which changes in a concerted manner during the reaction; and (ii) changes in activation energies parallel changes in equilibrium stability. These results are similar to those obtained for the folding of chymotrypsin inhibitor 2 (Fersht *et al.*, 1994). While only two simulations have been performed, the resulting transition state ensembles are partially structured and similar and wildly divergent structures are not observed (Figure 3). Thus, the data for FKBP12 are consistent with either a single pathway that involves particular elements of structure being partially formed, or a mixture of structurally related pathways where some have partial structure. The data are not consistent with a scheme involving parallel pathways or transition states, some with elements fully formed and others unformed.

β_F is an indication of the extent of structure in the transition state. For mutations in the α-helix and the β-sheet values of β_F are 0.16 and 0.36, respectively. These values confirm the Φ-value results and show that the α-helix is largely unstructured in the transition state, whereas the β-sheet, although very much weakened, is partially formed in the transition state. The values of β_F for mutations in the hydrophobic core with high Φ-values is 0.47 compared with 0.26 for those with low Φ-values, as expected.

Discussion

Structure of the transition state for folding/unfolding of FKBP12

Previous studies have shown that the transition state for folding of FKBP12 is relatively compact with retention of approximately 70 % of the buried surface area of the native state (Main *et al.*, 1999). Using protein engineering methods and Φ-value analysis, complemented with MD simulations, we have now characterised the structure of the transition state in some detail. A total of 34 mutants were constructed to probe specific regions of secondary and tertiary structure. The results from these studies are shown in Figure 5.

The 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 other α/β or α + β 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 (Serrano *et al.*, 1992; Matthews & Fersht, 1995; Daggett *et al.*, 1998), the α-helix in CI2 is partially structured (Itzhaki *et al.*, 1995; Li & Daggett, 1996; Daggett *et al.*, 1996) in the transition state, and H/^{2}H experiments and MD have shown that amide protons are protected from exchange early on the pathway of folding in the α-domain of lysozyme (Miranker *et al.*, 1991; Radford *et al.*, 1992; Kazmirski & Daggett, 1998). 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 trifluoroethanol (Callihan & Logan, 1999). 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 non-random conformations (Callihan & Logan, 1999). These results have been interpreted as suggesting that the helix forms early during folding; however, the results presented here do not support this contention. The discrepancies highlight 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 interactions in the β-sheet of FKBP12 are considerably weakened in the transition state: on average approximately 30-40 % of the interaction energy is present in the transition state. Some regions of the β-sheet, however, are more highly structured than others (Figures 3, 4 and 5). In particular, the N terminus Val2 in β-strand 1, Thr21, Val23, Val24 and Thr27 in β-strand 2, Thr75 and Ile76 in β-strand 4, and Val101 in β-strand 5, have slightly higher than average Φ-values, around 0.5. In addition, Val63 at the C terminus of the helix has a slightly higher than average value. The *S* values are in agreement with this assessment; but also extend the list of important residues to include other nearby residues: Gln3, Val4, Glu5, Lys73, Val98, Asp100 and Glu102 (these residues have *S* values $\geqslant 0.5$ in both TS1 and TS2). It is possible to speculate that these residues participate in a folding ''nucleus'' around which the rest of the structure condenses. Interestingly, these residues fall into two groups; the first has residues pointing into the hydrophobic core (residues 2, 4, 24, 63, 76 and 101) and the other residues are on the back, exposed side of the sheet. These residues are shown in various structures from the D-2 simulation (Figure 7) with residues pointing into the core coloured green and the others in red. While these 16 residues have Φ and *S* values >0.5, by virtue of being in a β-sheet, some may be carried along by the actual residues in the nucleus. For example, if the inner six residues make up the actual nucleus, when they come into close contact the outer residues are almost assured to be roughly in the sheet as well. On the other hand, the exposed residues may be crucial to the process. Because these are neighbouring residues it becomes difficult to determine cause and effect. All of the residues are dispersed throughout the protein in the unfolded state but are in closer proximity in the transition state. Further collapse and

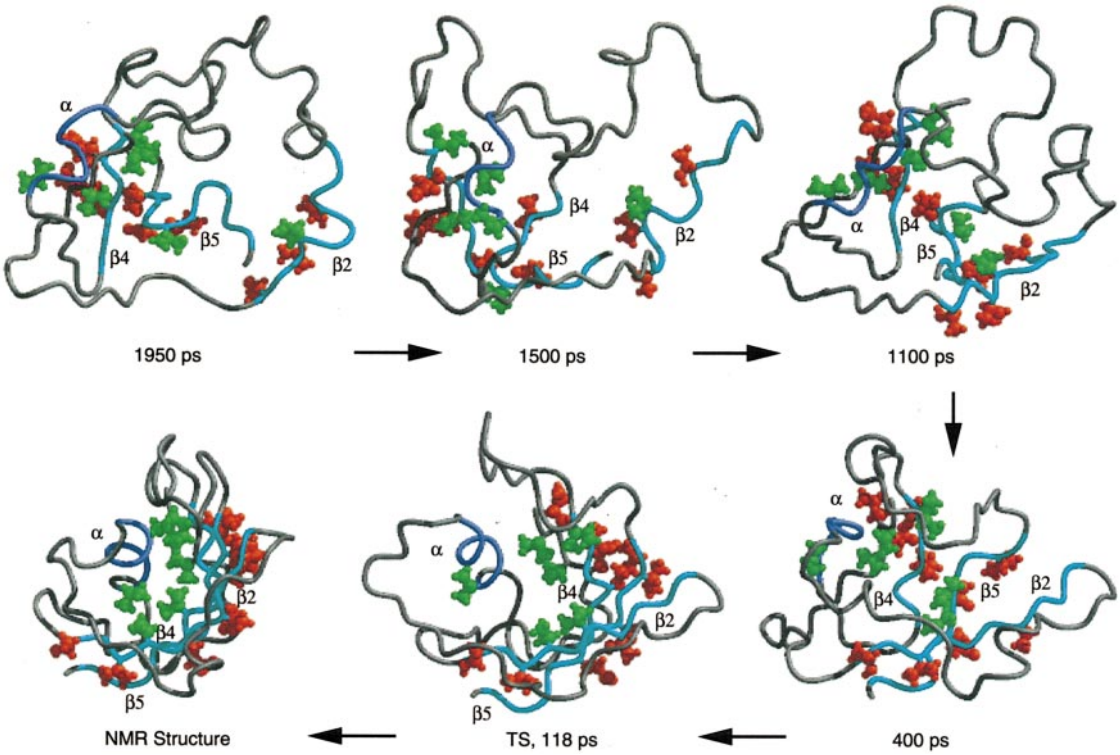

Figure 7. Residues involved in the nucleus along the folding pathway. Snapshots from the simulation beginning from the NMR structure (D-2) over time. Note that the structures are presented in the direction of folding to illustrate the disperse nature of the residues important in the transition state in the denatured state and how the residues do not come together until late in folding. Also, due to the distribution of these residues throughout the protein their packing interactions in the nucleus are linked to β-structure formation. The residues with large *S* values in the TS are depicted explicitly in green or red. The residues pointing into the hydrophobic core in the TS are green: Val2, Val4, Val24, Val63, Ile76, and Val101. The other residues are on the outside, solvent-exposed side of the sheet: Gln3, Glu5, Thr21, Val23, Thr27, Lys73, Thr75, Val98, Asp100, and Glu102. The β-strands 2, 4 and 5 are coloured cyan and the α-helix is shown in blue.

consolidation of the structure then occurs onto the scaffold provided by the nucleus, particularly the six core residues. Some of the residues comprising a more extended nucleus in TS1 and TS2 (e.g. with an *S* value >0.5 in one of the two transition states) differed, but the participating regions of the structure were preserved. These results provide support for lattice simulations that suggest that while the folding nucleus is probably not unique, the number of residues involved is limited (Guo & Thirumalai, 1995, 1997).

Comparison of the structure in the transition state with residual structure in the unfolded state

The structure of the urea-unfolded state of FKBP12 has been studied by NMR spectroscopy (Logan *et al.*, 1994). There is extensive conformational averaging in the unfolded state, as shown by C^{α} proton chemical shift data, vicinal coupling constants, $^{3}J_{HN,H\alpha}$, ^{15}N T_1 and T_2 relaxation times, and amide proton exchange rates, all of which were as expected for a random coil. However, a small number of medium-range $(i \rightarrow i+2)$ and $(i \rightarrow i+3)$ nuclear Overhauser enhancement cross-peaks (NOEs) were observed indicative of the transient formation of secondary structure in the unfolded state, in particular some weak helical structure in the region of the α-helix and β-strand 5. The simulations show little evidence of structure in the unfolded state, but there are some hints of helical structure in the region of the α-helix as well as helical turns at the C terminus (Figure 7). Residual structure in denatured states may play an important role in the early stages of the folding pathway by directing the protein along a particular pathway. 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 which may be involved in regions with transient structure in the denatured state. In particular, although there is evidence for

the transient formation of helical structure in the denatured state, the helix appears to be largely unstructured in the transition state.

The nucleation-condensation mechanism for protein folding

The data for FKBP12 are consistent with the nucleation-condensation mechanism of protein folding, as proposed initially for CI2 (Itzhaki *et al.*, 1995; Fersht, 1995, 1997). 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 formed.

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 (Fersht, 1995). In contrast to the nucleation model of folding proposed by Wetlaufer (1973, 1990), 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 7. 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 this may be a fruitless task. The residues of most importance in stabilising the transition state are distant in sequence.

Comparison of folding behavior of FKBP12 with other proteins

There are still relatively few proteins for which the protein engineering method has been used to analyse the structure of the transition state for folding. The best characterised are barnase (Matouschek *et al.*, 1989, 1990; Serrano *et al.*, 1992), CI2 (Itzhaki *et al.*, 1995), the SH3 domain from α-spectrin (Viguera *et al.*, 1996) and src (Grantcharova *et al.*, 1998), cheY (Lopez-Hernandez & Serrano, 1996) and the activation domain of procarboxypeptidase (ADAh2) (Villegas *et al.*, 1998). 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 (Itzhaki *et al.*, 1995). 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 (Itzhaki *et al.*, 1995; Lopez-Hernandez & Serrano, 1996). FKBP12 has 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 multiple 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.

Are the residues involved in stabilising the transition state and forming a folding nucleus highly conserved?

A comparison of the primary sequences of a number of FKBP12-like proteins shows that 23 residues out of 107 are invariant in all known FKBPs. By comparing these residues with Φ-values obtained in this study we see little correlation. For example, Thr21, Val23 and Val24 on β-strand 2 have some of the highest Φ-values for FKBP12, yet they are not conserved. Phe36, Val55, Ile56 and Ile91, 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, Val101, which has one of the highest Φ-values in this protein, is invariant and Val63, whose Φ-value is also moderately high, is highly conserved. This suggests that it may be difficult to predict the folding nucleus on the basis of sequence homology.

Conclusions

Protein engineering techniques combined with molecular dynamics simulations have been used to characterise the structure of the transition state for folding of FKBP12 at high resolution. Although previous studies have shown that the transition state is quite compact with some 70 % of the buried surface area in the native state present in the transition state (Main *et al.*, 1999), present studies have established that there are a number of regions which are largely unstructured in the transition state including the α-helix and loops. Although no

region of structure is fully formed in the transition state, regions of the β-sheet are comparatively structured, with the structure being centred 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, comprises side-chains within β-strands 2, 4 and 5, and the C terminus of the α-helix. The simulated transition state ensembles suggest that particular regions of the protein are retained in the nucleus, but that the transition state is heterogeneous and the precise residues involved can be somewhat variable. These residues are distant in the primary sequence, demonstrating the importance of tertiary interactions in the transition state.

Materials and Methods

Materials

Wild-type and mutant FKBP12 were constructed, expressed and purified as described (Main *et al*., 1998). Protein concentration was determined spectrophotometrically using a molar extinction coefficient, ε, of 9927 M^{-1} cm^{-1} at 278 nm for both wild-type and mutant FKBP12. Ultra-grade urea was used (Fisher Scientific UK Ltd.). All stock solutions of urea were made using volumetric flasks, flash-frozen and stored at −20 °C to prevent degradation. In all experiments the final buffer concentrations were 50 mM Tris-HCl (pH 7.5), 1 mM DTT. All other materials were analytical grade and purchased from Sigma.

Methods

Equilibrium experiments

Equilibrium experiments were performed and data analysed as described (Main *et al*., 1998).

Kinetic experiments

An Applied Photophysics Stopped-flow Reaction Analyser (model SF.17MV) was used and data were acquired and analysed using the Applied Photophysics Kinetic Workstation, version 4.099, supplied. In all studies the temperature was 25(±0.1)°C.

Unfolding studies

Unfolding was initiated by diluting one volume of aqueous protein solution (approximately 22 μM FKBP12 in 50 mM Tris (pH 7.5), 1 mM DTT), into ten volumes of concentrated urea (containing 50 mM Tris-HCl (pH 7.5), 1 mM DTT). The initial urea solutions were such that the final urea concentrations varied between 3.0 and 8.5 M.

Data analysis for unfolding kinetics

Fluorescence traces were analysed using either Kaleidagraph or Applied Photophysics software. The data were fitted to an equation describing a single exponential process with linear drift and offset. The drift, when present, is very small and results from photolysis and baseline instability.

Molecular dynamics simulations

Three simulations of FKBP12 were performed for 2 ns each. Two of the simulations began from the 1.7 Å crystal structure of van Duyne *et al*. (1991, 1fkb). One of these was a control simulation at 298 K, while the other was a simulation of thermal denaturation at 498 K, D-1. Another 498 K simulation (D-2) was performed beginning with the average NMR structure of Michnick *et al*. (1991, 1fks). The potential energy function and protocols for the MD as implemented within the program ENCAD (Levitt, 1990) have been described elsewhere (Levitt *et al*., 1995, 1997). All simulations were performed at neutral pH (Lys and Arg residues were positively charged and Asp and Glu residues were negatively charged). Water molecules were added around the protein to fill a rectangular box, with walls at least 8 Å away from any protein atom, resulting in the addition of 3571, 3470 and 4026 water molecules for the 298 K, D-1, and D-2 simulations, respectively. The density of the solvent was set to the experimental value for 498 K, 0.829 g/ml or 298 K, 0.997 g/ml (Kell, 1967; Haar *et al*., 1984) by adjusting the volume of the box, resulting in at least a 10 Å shell of water molecules at 498 K. Periodic boundary conditions were used to minimize boundary effects. Preparation of the systems followed the protocol described by Daggett *et al*. (1998). The simulations were carried out for 2000 ps, 2 ns, each, which was sufficient to unfold the protein.

For semi-quantitative comparison with the experimental Φ values, structure indices, or *S* values were calculated for each residue averaging over the transition state ensemble identified from the simulation. *S* is defined as the product of S_{2^o} and S_{3^o}, i.e. $S = (S_{2^o})(S_{3^o})$, where S_{2^o} is the fraction of native secondary structure and S_{3^o} is the fraction of tertiary structure retained in this conformational ensemble. The approach is discussed in more detail by Daggett *et al*. (1996).

Acknowledgements

We thank Professor Alan Fersht for his support and Tom Rippin for help with the Ile → Ala7 mutant. S.E.J. is a Royal Society University Research Fellow, E.R.G.M. is supported by a BBSRC studentship, K.F.F. 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 V.D.).

References

Alexandrescu, A. T., Abeygunawardana, C. & Shortle, D. (1994). Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease—a heteronuclear NMR-study. *Biochemistry*, **33**, 1063-1072.

Arcus, V. L., Vuilleumier, S., Freund, S., Bycroft, M. & Fersht, A. (1994). Toward solving the folding pathway of barnase-the complete backbone C-13, N-15, and H-1-NMR assignments of its pH-denatured state. *Proc. Natl Acad. Sci. USA,* **91**, 9412-9416.

Arcus, V. L., Vuilleumier, S., Freund, S. M. V., Bycroft, M. & Fersht, A. R. (1995). A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. *J. Mol. Biol.* **254**, 305-321.

Bacon, D. J. & Anderson, W. F. (1988). A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219-220.

Baldwin, R. L. (1993). Pulsed H/D-exchange studies of folding intermediates. *Curr. Opin. Struct. Biol.* **3**, 84-91.

Bond, C. J., Wong, K., Clarke, J., Fersht, A. R. & Daggett, V. (1997). Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. *Proc. Natl. Acad. Sci. USA,* **94**, 13409-13413.

Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L. & Oas, T. G. (1997). The energy landscape of a fast-folding protein mapped by Ala → Gly substitutions. *Nature Struct. Biol.* **4**, 305-310.

Callihan, D. & Logan, T. M. (1999). Conformations of peptide fragments from the FK506 binding protein: comparison with the native and urea-unfolded states. *J. Mol. Biol.* **285**, 2161-2175.

Daggett, V., Li, A., Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1996). Structure of the transiton state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* **257**, 430-440.

Daggett, V., Li, A. & Fersht, A. R. (1998). A combined molecular dynamics and ϕ-value analysis of structure-reactivity relationships in the transition state and unfolding pathway of barnase: the structural basis of Hammond and anti-Hammond effects. *J. Am. Chem. Soc.* **120**, 12740-12754.

Egan, D. A., Logan, T. M., Liang, H., Matayoshi, E., Fesik, S. W. & Holzman, T. F. (1993). Equilibrium denaturation of recombinant human FK506 binding protein in urea. *Biochemistry,* **32**, 1920-1927.

Evans, P. A. & Radford, S. E. (1994). Probing the structure of folding intermediates. *Curr. Opin. Struct. Biol.* **4**, 100-106.

Farrow, N. A., Zhang, O. W., Forman-Kay, J. D. & Kay, L. E. (1997). Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry,* **36**, 2390-2402.

Fersht, A. R. (1995). 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**, 10869-10873.

Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3-9.

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. Mol. Biol.* **224**, 771-782.

Fersht, A. R., Itzhaki, L. S., El Masry, N., Matthews, J. M. & Otzen, D. E. (1994). Single versus parallel pathways of protein-folding and fractional formation of structure in the transition-state. *Proc. Natl Acad. Sci. USA,* **91**, 10426-10429.

Fiebig, K. M., Schwalbe, H., Buck, M., Smith, L. J. & Dobson, C. M. (1996). Toward a description of the conformations of denatured states of proteins-comparison of a random coil model with NMR measurements. *J. Phys. Chem.* **100**, 2661-2666.

Frank, M. K., Clore, G. M. & Gronenborn, A. M. (1995). Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein-G by multidimensional heteronuclear NMR-spectroscopy. *Protein Sci.* **4**, 2605-2615.

Freund, S. M. V., Wong, K. B. & Fersht, A. R. (1996). Initiation sites of protein folding by NMR analysis. *Proc. Natl Acad. Sci. USA,* **93**, 10600-10603.

Gillespie, J. R. & Shortle, D. (1997a). Characterization of long-range structure in the denatured state of staphylococcal nuclease. 1. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**, 158-169.

Gillespie, J. R. & Shortle, D. (1997b). Characterization of long-range structure in the denatured state of staphylococcal nuclease. 2. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **268**, 170-184.

Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nature Struct. Biol.* **5**, 714-720.

Guo, Z. & Thirumalai, D. T. (1995). Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. *Biopolymers,* **36**, 83-103.

Guo, Z. & Thirumalai, D. T. (1997). The nucleation-collapse mechanism in protein folding: Evidence for the non-uniquemess of the folding nucleus. *Fold. Des.* **2**, 377-391.

Haar, L., Gallagher, J. S. & Kell, G. S. (1984). *NBS/NRC Steam Tables: Thermodyanmic and Transport Properties and Computer Programs for Vapor and Liquid States of Water in SI Units,* Hemisphere Publication Corporation, Washington DC.

Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). 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**, 260-288.

Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81-R90.

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

Jackson, S. E., Moracci, M., El Masry, N., Johnson, C. M. & Fersht, A. R. (1993). Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor-2. *Biochemistry,* **32**, 11259-11269.

Kazmirski, S. & Daggett, V. (1998). Non-native interactions in protein folding intermediates: molecular dynamics simulations of hen lysozyme. *J. Mol. Biol.* **284**, 793-806.

Kell, G. S. (1967). Precise representation of volume properties of water at one atmosphere. *J. Chem. Eng. Data,* **12**, 66.

Kim, D. E., Yi, Q., Gladwin, S. T., Goldberg, J. M. & Baker, D. (1998). The single helix in protein L is largely disrupted at the rate-limiting step in folding. *J. Mol. Biol.* **284**, 807-815.

Kim, P. S. & Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459-489.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Levitt, M. (1990). *ENCAD: Energy Caclulation and Dynamics*, Molecular Applications Group, Stanford, CA.

Levitt, M., Hirshberg, M., Sharon, R. & Daggett, V. (1995). Potenetial energy function and parameters for simulations of the moelcular dynamics of proteins and nucleic acids in solution. *Comp. Phys. Commun.* **91**, 215-231.

Levitt, M., Hirshberg, M., Sharon, R., Laidig, K. E. & Daggett, V. (1997). Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J. Phys. Chem. ser. B,* **101**, 5051-5061.

Li, A. & Daggett, V. (1994). Characterization of the transition-state of protein unfolding by use of molecular-dynamics-chymotrypsin inhibitor-2. *Proc. Natl Acad. Sci. USA,* **91**, 10430-10434.

Li, A. & Daggett, V. (1996). Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J. Mol. Biol.* **257**, 412-429.

Li, A. & Daggett, V. (1998). Molecular dynamics simulation of the unfolding of barnase: Characterization of the major intermediate. *J. Mol. Biol.* **275**, 677-694.

Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. (1993). A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J. Biomolecular NMR,* **3**, 225-231.

Logan, T., Theriault, Y. & Fesik, S. (1994). Structural characterization of the FK506 binding-protein unfolded in urea and guanidine-hydrochloride. *J. Mol. Biol.* **236**, 637-648.

Lopez-Hernandez, E. & Serrano, L. (1996). Structure of the transition state for folding of the 129 aa protein cheY resembles that of a smaller protein, CI-2. *Fold. Design*, **1**, 43-55.

Main, E. R. G., Fulton, K. F. & Jackson, S. E. (1998). Stability of FKBP12: The context-dependent nature of destabilising mutations. *Biochemistry,* **37**, 6145-6153.

Main, E. R. G., Fulton, K. F. & Jackson, S. E. (1999). Folding pathway of FKBP12 and characterization of the transition state for folding. *J. Mol. Biol. ???.*

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,* **342**, 122-126.

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

Matthews, J. M. & Fersht, A. R. (1995). Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for the structural elements for unfolding/folding of barnase. *Biochemistry,* **34**, 6805-6814.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0-a program for photorealistic molecular graphics. *Acta Crystallog. sect. D,* **50**, 869-873.

Michnick, S. W., Rosen, M. K., Wandless, T. J., Karplus, M. & Schreiber, S. L. (1991). Solution structure of FKBP, a rotamase enzyme and receptor for FK506 and rapamycin. *Science,* **252**, 836-842.

Milla, M. E., Brown, B. M., Waldburger, C. D. & Sauer, R. T. (1995). P22 arc repressor-transition-state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry,* **34**, 13914-13919.

Miranker, A., Radford, S. E., Karplus, M. & Dobson, C. M. (1991). Demonstration by NMR of folding domains in lysozyme. *Nature,* **349**, 633-636.

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

Pan, H., Barbar, E., Barany, G. & Woodward, C. (1995). Extensive nonrandom structure in reduced and unfolded bovine pancreatic trypsin-inhibitor. *Biochemistry,* **34**, 13974-13981.

Radford, S. E., Dobson, C. M. & Evans, P. A. (1992). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature,* **358**, 302-307.

Richardson, J. S. & Richardson, D. S. (1988). Amino-acid preferences for specific locations at the ends of alpha-helices. *Science,* **240**, 1648-1652.

Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J. & Dobson, C. M. (1997). Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry,* **36**, 8977-8991.

Serrano, L., Matouschek, A. & Fersht, A. R. (1992). The folding of an enzyme. 3. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **22**, 805-818.

Shortle, D. (1993). Denatured states of proteins and their roles in folding and stability. *Curr. Opin. Struct. Biol.* **3**, 66-74.

Smith, L. J., Fiebig, K. M., Schwalbe, H. & Dobson, C. M. (1996). The concept of a random coil-residual structure in peptides and denatured proteins. *Fold. Des.* **1**, R95-R106.

van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1991). Atomic-structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science,* **252**, 839-842.

Viguera, A. R., Serrano, L. & Wilmanns, M. (1996). Different folding transition-states may result in the same native structure. *Nature Struct. Biol.* **3**, 874-880.

Villegas, V., Martínez, J. C., Avilés, F. X. & Serrano, L. (1998). Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.* **283**, 1027-1036.

Wang, Y. & Shortle, D. (1995). The equilibrium folding pathway of staphylococcal nuclease: identification of the most stable chain-chain interactions by NMR and CD spectroscopy. *Biochemistry,* **34**, 15895-15905.

Wang, Y. & Shortle, D. (1996). A dynamic bundle of four adjacent hydrophobic segments in the

denatured state of staphylococcal nuclease. *Protein Sci.* **5**, 1898-1906.
Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl Acad. Sci. USA,* **70**, 697-701.
Wetlaufer, D. B. (1990). Nucleation in protein folding-confusion of structure and process. *Trends Biochem. Sci.* **15**, 414-415.
Wong, K. B., Freund, S. M. V. & Fersht, A. R. (1996). Cold denaturation of barstar-H-1, N-15 and C-13 NMR assignment and characterization of residual structure. *J. Mol. Biol.* **259**, 805-818.
Zhang, O. W. & Forman-Kay, J. D. (1995). Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. *Biochemistry*, **34**, 6784-6794.
Zhang, O. W. & Forman-Kay, J. D. (1997). NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. *Biochemistry*, **36**, 3959-3970.

Edited by J. Thornton

(Received 26 November 1998; received in revised form 24 May 1999; accepted 10 June 1999)

http://www.academicpress.com/jmb

Supplementary material comprising three Tables is available from JMB Online.