Protein Engineering, Design & Selection vol. 18 no. 5 pp. 229–237, 2005
Published online April 27, 2005 doi:10.1093/protein/gzi025

Ubiquitin folds through a highly polarized transition state

Heather M.Went and Sophie E.Jackson[1]

Chemistry Department and Centre for Protein Engineering, Lensfield Road, Cambridge CB2 1EW, UK

[1]To whom correspondence should be addressed.
E-mail: sej13@cam.ac.uk

The small α/β protein ubiquitin has been used as a model system for experimental and computational studies on protein folding for many years. Here, we present a comprehensive ϕ-value analysis and characterize the structure and energetics of the transition state ensemble (TSE). Twenty-seven non-disruptive mutations are made throughout the structure and a range of ϕ-values from zero to one are observed. The values cluster such that medium and high values and found only in the N-terminal region of the protein, whilst the C-terminal region has consistently low ϕ-values. In the TSE, the main α-helix appears to be fully formed (two ϕ-values which specifically probe helical structure are one) and the helix is stabilized by packing against the first β-turn, which is partially structured. In striking comparison, the ϕ-values in the C-terminal region are all very low, suggesting that this region of the protein is largely unstructured in the TSE. Data are consistent with a nucleation–condensation mechanism in which there is a highly polarized folding nucleus comprising the first β-hairpin and the α-helix. Data presented from the protein engineering study and ϕ-value analysis are compared with results from other experimental studies and also computational studies.

Keywords: folding nucleus/nucleation–condensation/transition-state ensemble/two-state

Introduction

Ubiquitin has been used extensively as a model system for protein folding studies. In the early 1990s, work from the Roder group using NMR spectroscopy in conjunction with H/D exchange, and also stopped-flow fluorescence measurements (Briggs *et al*., 1992; Khoransanizadeh *et al*., 1993), established ubiquitin as an excellent model for folding. At the same time, the Williams group used a complementary approach studying peptides and partially structured states of ubiquitin formed in organic solvents (Harding *et al*., 1991; Cox *et al*., 1993; Woolfson *et al*., 1993). Since then, there have been a large number of experimental studies on the protein using a wide variety of techniques to probe structure and folding in both full-length ubiquitin and in peptides. In addition to the many experimental studies, ubiquitin has also been used extensively as a model system in computational studies of folding. The computational approaches adopted have been varied, ranging from evolutionary-based methods using sequence conservation (Michnick and Shakhnovich, 1998), the use of coarse-grained models (Gilis and Rooman, 2001; Fernandez *et al*., 2002) to all-atom molecular dynamic simulations of unfolding (Alonso and Daggett, 1998). Despite the clear interest in the folding pathway of ubiquitin from both experimental and computational protein folding communities, no comprehensive ϕ-value analysis has so far been reported [Roder and co-workers and Sosnick and co-workers, however, have both published the results of limited protein engineering experiments (Khoransandideh *et al*., 1996; Krantz *et al*., 2004; Sosnick *et al*., 2004)].

Here, extensive protein engineering was used, in conjunction with ϕ-value analysis (Matouschek *et al*., 1989; Serrano *et al*., 1992, Jackson *et al*., 1993), to map out in detail the energetics and structure of the transition state ensemble (TSE) for the folding of ubiquitin. Ubiquitin is an ideal system for protein engineering studies: it is a stable protein and much is known about the factors that contribute to its stability (Loladze *et al*., 1999, 2002; Ermolenkoe *et al*., 2002, 2003; Loladze and Makhatadze, 2002; Makhatadze *et al*., 2003). This facilitates extensive protein engineering studies which can be undertaken in all regions of the structure to generate a large number of mutants suitable for analysis. It is a small, monomeric protein of 76 residues which has, under the conditions used in this study, simple two-state folding kinetics (Went *et al*., 2004). This allows the measurement of ϕ-values from unfolding data (measured at high concentrations of denaturant), and also from folding rate constants measured in water. The number of mutants studied, combined with the fact that ϕ-values are measured independently from unfolding and folding rates, in addition to the analysis of multiple mutations at single sites, produces a very robust, highly consistent set of results. Here, 27 mutations have been made throughout the protein which probe both secondary and tertiary structure formation. The effect of these mutations on the stability, unfolding and folding kinetics were measured and ϕ-values for the TSE determined from both unfolding and folding rate constants. With such a comprehensive set of ϕ-values, the structure of the transition state ensemble was mapped out in detail and compared with results from other experimental and computational studies.

Materials and methods

Materials

Isopropyl-thio-β-D-galactoside (IPTG) and analytical-grade guanidinium chloride (GdnHCl) were purchased from Melford Laboratories. All other materials were of analytical grade and purchased from Sigma.

Ubiquitin

A tryptophan-containing mutant of ubiquitin, F45W, was used as a pseudo-wild-type throughout this study. The construction and properties of F45W are described elsewhere (Khoransideh *et al*., 1993; Went *et al*., 2004). Standard mutagenesis

techniques (Stratagene) were used to create single point mutants. All vectors were fully sequenced. Wild-type and mutant protein were expressed and purified as described elsewhere (Went *et al*., 2004). All purified proteins were concentrated, flash-frozen and stored at −80°C. The purity of the protein was determined by SDS–PAGE and mass spectrometry. The protein concentration was determined spectrophotometrically using a molar extinction coefficient, ε, of 6970 M^{-1} cm^{-1}.

Equipment and general procedures

In all experiments, final buffer concentrations were 50 mM Tris–HCl pH 7.4. Buffer and denaturant solutions were prepared as described elsewhere (Main *et al*., 1999).

Equilibrium experiments

Experiments performed under equilibrium conditions have been described in detail elsewhere (Main *et al*., 1999). Final concentrations of ubiquitin were 2 μM. The fluorescence was measured at 353 nm.

Kinetic experiments

Unfolding and refolding measurements were performed by [GdnHCl]-jump experiments using equipment and procedures described elsewhere (Main *et al*., 1999). Final concentrations of ubiquitin were typically 2 μM. Unfolding data were fitted to a single exponential process which included a linear drift when measurements were made over longer time-scales to account for baseline instability. Refolding data were fitted to either a double or triple exponential process. The multiple refolding phases result from heterogeneity in the unfolded state resulting from proline isomerization. Only the major, fast refolding phase is considered in this analysis.

Data analysis

Kinetic data for unfolding and refolding were fit to a two-state model using Equation (1) (Jackson and Fersht, 1991):

$$\ln k = \left(k_F^{H_2O}\exp(-m_{k_F}[D]) + k_U^{H_2O}\exp(+m_{k_U}[D])\right) \quad (1)$$

where $k_F^{H_2O}$ and $k_F^{H_2O}$ are the rate constants for folding and unfolding in water, respectively and m_{k_F} and m_{k_U} are the slopes of the Chevron plot for folding and unfolding, respectively.

Calculation of $\Delta\Delta G_{‡\text{-}N}$, $\Delta\Delta G_{‡\text{-}D}$ and Φ_F values

From the unfolding kinetic data, the effect of the mutation on the stability of the transition state relative to the folded state can be calculated using the equation

$$\Delta\Delta G_{‡\text{-}N} = -RT\ln(k_U/k_U') \quad (2)$$

where $\Delta\Delta G_{‡\text{-}N}$ is the difference in the energy of the transition state ensemble 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. Together with the value of $\Delta\Delta G_{D\text{-}N}$ (the difference in the free energy between the unfolded and native states between wild-type and mutant), obtained from equilibrium experiments, $\Delta\Delta G_{‡\text{-}N}$ can be used to calculate a Φ_F-value according to the equation

$$\Phi_F = 1 - \Delta\Delta G_{‡\text{-}N}/\Delta\Delta G_{D\text{-}N} \quad (3)$$

where the Φ_F-value is a measure of the degree to which the favourable interactions removed on mutation have been made in the transition state ensemble.

For proteins that fold with two-state kinetics, Φ_F can also be calculated from refolding data:

$$\Delta\Delta G_{‡\text{-}D} = RT\ln(k_F/k_F') \quad (4)$$

where $\Delta\Delta G_{‡\text{-}D}$ is the difference in the energy of the transition state ensemble relative to the denatured state between the wild-type and mutant and k_F and k_F' are the rate constants of folding for the wild-type and mutant, respectively. In this case, the Φ_F-value can then be calculated from the equation

$$\Phi_F = \Delta\Delta G_{‡-D}/\Delta\Delta G_{D-N} \quad (5)$$

Results

Equilibrium measurements

The pseudo-wild-type ubiquitin (F45W) and all the mutants unfolded cooperatively under equilibrium conditions following a two-state transition (data not shown). Data were analysed as described elsewhere (Jackson and Fersht, 1991) to yield a value for $\Delta\Delta G_{D\text{-}N}$, the difference in the free energy of unfolding between wild-type and mutant. Values for $\Delta\Delta G_{D\text{-}N}$ are shown in Table I. Further thermodynamic parameters obtained from the equilibrium unfolding experiments are given in Table 1 of the Supplementary data, available at *PEDS* Online. The mutant proteins are mainly destabilized with respect to the wild-type ubiquitin, typically between 0 and 4 kcal/mol.

Kinetic measurements

The pseudo-wild-type ubiquitin (F45W) and all the mutants unfold with a single exponential decay. Almost all the mutants unfold at the same rate or faster than the wild-type as expected for destabilizing mutants. The rates of unfolding vary exponentially as a function of denaturant concentration (Figure 1), as has been widely observed (Tanford, 1968, 1970; Matthews, 1987). The refolding of the wild-type and mutant proteins is a multiphasic process: the fast major phases corresponds to the folding of the protein whilst the slower minor phases correspond to the population of molecules whose folding rate is limited by proline isomerisation (data not shown). The analysis presented here uses only the rate of the major refolding phase. Under the conditions used, the rate of folding is also observed to vary exponentially as a function of denaturant concentration (Figure 1). Together, the unfolding and refolding kinetics give rise to the typical V-shaped or chevron plot (Figure 1). This is typical for small proteins which fold with simple two-state kinetics, i.e. no intermediate is significantly populated during folding (Jackson and Fersht, 1991). In this case, data are fitted to Equation 1 to yield values for $k_U^{H_2O}$ and $k_F^{H_2O}$ the rate constants for folding and unfolding in water, respectively and m_{k_U} and m_{k_F}, the slopes of the unfolding and folding arm of the chevron plot, respectively. Values are summarised in Table I. The kinetic data can also be used to calculate the thermodynamic parameters $\Delta G_{D\text{-}N}$ and $m_{D\text{-}N}$ and values compared with those obtained directly from equilibrium experiments. For wild-type and mutant proteins the values obtained from kinetic and equilibrium experiments are within error (data not shown), showing that the proteins all show two-state behaviour.

In general, the slopes of the unfolding and folding limb of the chevron plots for the mutants are very similar to wild-type (see Table I and Figure 1), suggesting that the mutations have little effect on the position of the transition state ensemble

Table I. Kinetic and thermodynamic parameters for the unfolding and folding of wild-type[a] and mutant ubiquitins

Mutant	$\Delta\Delta G^{H_2O}_{D\text{-}N}$ (kcal/mol)	$\ln k_U^{H_20}$	$\ln k_U^{4MGdnHCl}$	$m_{\ddagger\text{-}N}$ (M^{-1})	$\ln k_F^{H_2O}$	$m_{\ddagger\text{-}D}$ (M^{-1})
Wild-type		−6.43 ± 0.15	−0.98 ± 0.05	1.36 ± 0.03	5.85 ± 0.07	2.65 ± 0.05
Ile→Val3	1.18 ± 0.04	−4.57 ± 0.11	0.57 ± 0.04	1.28 ± 0.02	5.06 ± 0.05	2.57 ± 0.04
Ile→Ala3	3.18 ± 0.04	−2.85 ± 0.05	2.59 ± 0.02	1.36 ± 0.01	3.78 ± 0.12	2.90 ± 0.14
Val→Ala5	2.41 ± 0.04	−3.65 ± 0.05	0.99 ± 0.03	1.16 ± 0.01	3.51 ± 0.17	2.53 ± 0.14
Thr→Ala7	1.37 ± 0.04	−5.66 ± 0.12	−0.50 ± 0.03	1.29 ± 0.02	4.39 ± 0.04	2.89 ± 0.04
Thr→Ala9	−0.06 ± 0.04	−4.30 ± 0.10	0.55 ± 0.05	1.21 ± 0.02	5.69 ± 0.05	2.78 ± 0.04
Ile→Val13	1.16 ± 0.04	−4.28 ± 0.15	0.36 ± 0.02	1.16 ± 0.03	5.03 ± 0.05	2.66 ± 0.04
Ile→Ala13	3.40 ± 0.04	−2.03 ± 0.10	1.94 ± 0.01	0.99 ± 0.02	3.40 ± 0.12	2.90 ± 0.15
Leu→Ala15	3.85 ± 0.04	−2.16 ± 0.06	2.26 ± 0.03	1.10 ± 0.01	3.20 ± 0.09	2.80 ± 0.11
Val→Ala17	1.65 ± 0.04	−4.55 ± 0.08	0.32 ± 0.04	1.22 ± 0.01	4.26 ± 0.09	2.76 ± 0.08
Asp→Asn21	0.85 ± 0.04	−6.01 ± 0.09	−1.18 ± 0.03	1.21 ± 0.02	4.17 ± 0.06	2.59 ± 0.05
Thr→Ala22	1.74 ± 0.04	−4.46 ± 0.07	0.50 ± 0.03	1.24 ± 0.01	4.22 ± 0.07	2.74 ± 0.07
Ile→Val23	0.45 ± 0.04	−5.54 ± 0.10	−0.57 ± 0.04	1.24 ± 0.02	5.18 ± 0.08	2.74 ± 0.07
Ile→Ala23	2.78 ± 0.04	−2.62 ± 0.07	1.86 ± 0.02	1.12 ± 0.02	4.60 ± 0.09	3.17 ± 0.11
Ile→Gly23	3.58 ± 0.04	−2.41 ± 0.07	2.17 ± 0.04	1.14 ± 0.02	3.06 ± 0.06	2.34 ± 0.07
Val→Ala26	3.37 ± 0.04	−2.44 ± 0.07	3.15 ± 0.03	1.40 ± 0.02	3.61 ± 0.07	2.96 ± 0.08
Lys→Ala27	2.56 ± 0.04	−1.14 ± 0.07	2.97 ± 0.02	1.03 ± 0.02	6.15 ± 0.16	3.47 ± 0.21
Ala→Gly28	0.52 ± 0.04	−6.58 ± 0.12	−1.06 ± 0.06	1.38 ± 0.02	4.91 ± 0.04	2.78 ± 0.03
Ile→Val30	1.21 ± 0.04	−4.71 ± 0.10	0.08 ± 0.04	1.20 ± 0.02	4.78 ± 0.04	2.71 ± 0.04
Ile→Ala30	3.22 ± 0.04	−2.36 ± 0.10	1.60 ± 0.04	0.99 ± 0.02	4.31 ± 0.10	3.37 ± 0.14
Gln→Ala41	1.46 ± 0.04	−3.12 ± 0.05	1.49 ± 0.01	1.15 ± 0.01	5.56 ± 0.05	2.41 ± 0.04
Leu→Ala43	4.36 ± 0.05	1.62 ± 0.04	4.36 ± 0.01	0.68 ± 0.01	6.11 ± 0.18	3.73 ± 0.30
Leu→Ala50	2.73 ± 0.04	−1.10 ± 0.12	3.47 ± 0.05	1.14 ± 0.03	5.85 ± 0.08	3.10 ± 0.10
Leu→Ala56	4.05 ± 0.04	−0.16 ± 0.09	4.85 ± 0.04	1.25 ± 0.02	5.60 ± 0.17	3.58 ± 0.25
Ile→Val61	1.12 ± 0.04	−4.38 ± 0.09	0.66 ± 0.03	1.26 ± 0.02	5.73 ± 0.11	2.66 ± 0.09
Ile→Ala61	3.23 ± 0.04	−0.60 ± 0.18	4.63 ± 0.08	1.31 ± 0.06	5.23 ± 0.09	2.21 ± 0.10
Leu→Ala67	2.57 ± 0.04	−1.74 ± 0.11	3.54 ± 0.01	1.32 ± 0.03	5.45 ± 0.07	2.86 ± 0.08
Leu→Ala69	3.05 ± 0.04	−1.30 ± 0.06	2.49 ± 0.03	0.95 ± 0.01	5.32 ± 0.11	2.98 ± 0.12

[a]All mutants are made on a pseudo-wild-type which has a tryptophan engineered at position 45.

relative to the denatured and native states. However, two mutants (L43A and L69A) have significantly lower unfolding slopes than wild-type. Both of these mutations are highly destabilizing (Table I), the lower slope indicating that, for these mutants, the TSE is more native-like than for wild-type and the other mutants. It is interesting that other, equally destabilizing mutants such as L67A do not appear to affect the position of the TSE. Similar results have been observed for other proteins and attributed to the effect of mutations on the local maxima and minima on a broad transition state (Oliveberg, 2001). In comparison, only one mutant (I23G) shows a significant difference in the slope of the refolding arm of the chevron compared with wild-type. Again, I23G is a highly destabilized mutant and, in this case, the change in m_{k_F} may be due either to movement of the position of the TSE or to an effect on the denatured state on the introduction of a highly flexible glycine residue.

ϕ-Value analysis

The rate constants for unfolding and refolding can both be used to calculate a ϕ_F-value; see Materials and methods. For unfolding, the ϕ_F-value is calculated at 4 M GdnHCl, thereby minimizing the large errors associated with extrapolating kinetic data measured at high concentrations of denaturant to water. This approach has been validated for other proteins (Jackson *et al*., 1993); ϕ_F-values can also be calculated directly from refolding data in water. Both sets of values are shown in Table II. Individual errors are not given as a normal analysis of compound errors underestimates the errors as ϕ_F approaches zero. Typically, ϕ_F-values can be reproduced to within 0.05–0.1. In most cases, the values calculated at 4 M GdnHCl and in water are within error, as expected for a protein which shows simple two-state kinetics and for which there is no significant movement in the transition state with mutation or denaturant concentration. For all mutants, the values vary from zero to one, there are no values significantly above one (which would indicate non-native interactions in the transition state) or below zero (which may indicate the presence of significant interactions in the denatured state).

In addition to the ϕ_F-values calculated for the mutations made, a fine-structure analysis can be used at positions where multiple mutations have been made. This is useful for several reasons. First, to probe more specifically certain interactions. For example, Ile→Val23 or Ile→Ala23 probe mainly tertiary interactions formed between the α-helix and the β-sheet, the additional mutation Ile→Gly23, allows a ϕ_F-value to be calculated for the composite mutation Ala→Gly23. This 'hypothetical' mutant probes formation of the α-helix alone. Details of the fine-structure analysis are given in Table 2 in the Supplementary data. The values calculated at five positions are shown in Table II. In general, there is excellent agreement between ϕ-values calculated for different mutations of a single residue.

Discussion

Structure in the transition state ensemble

An extensive description of the interpretation of Φ_F-values and the assumptions and limitations of this method has been presented elsewhere (Fersht *et al*., 1992) and will not be discussed here. The method has now been rigorously tested over the past 10 years on a number of different proteins. Essentially, $\Phi_F = 0$ when the mutation has no effect on the energy of the

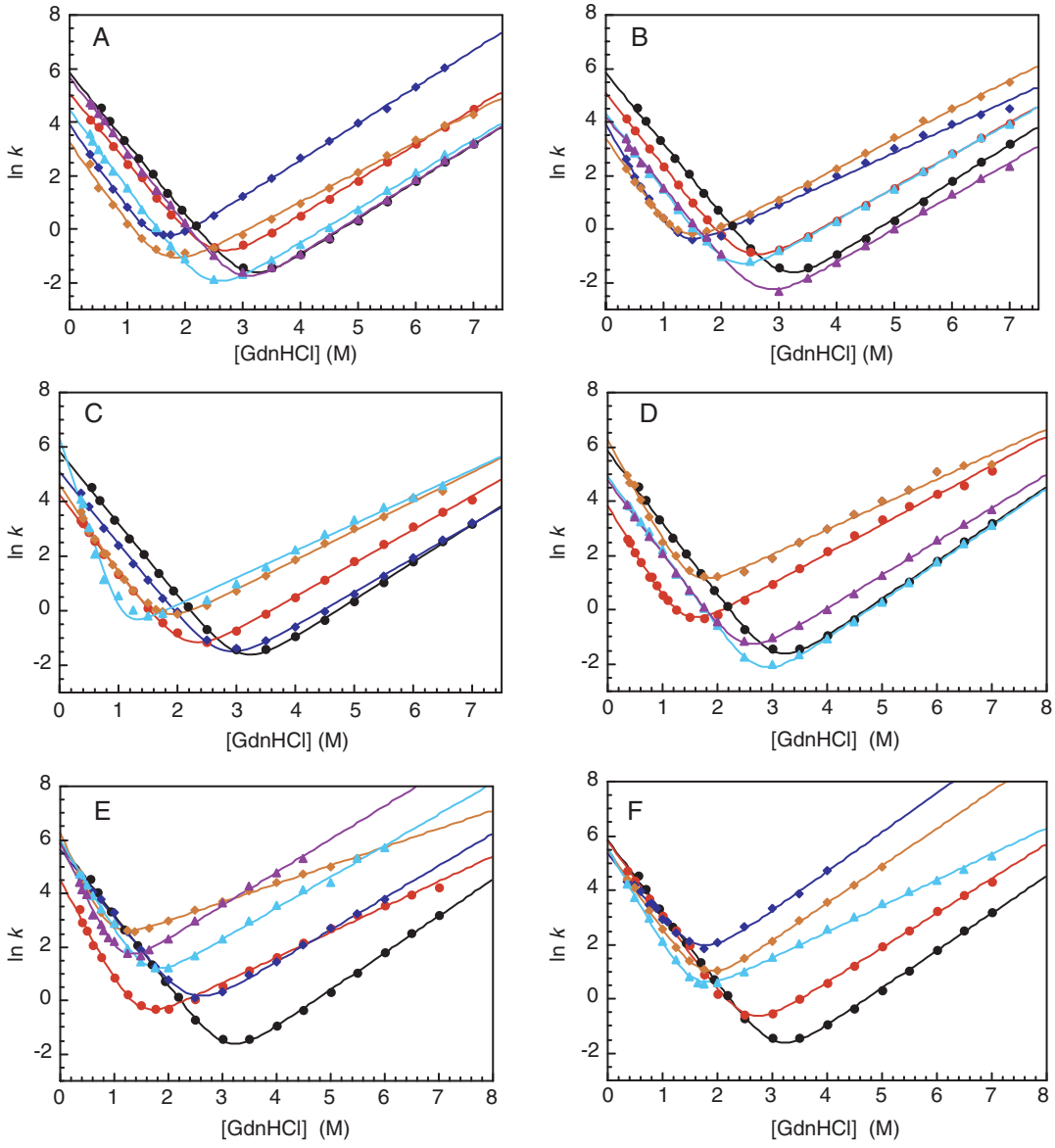

Fig. 1. Chevron plots of unfolding and refolding data for wild-type and mutant ubiquitin. Unfolding and refolding rate constants were measured as a function of [GdnHCl] in 50 mM Tris, pH 7.4 at 25°C. (**A**) Wild-type (black circles), I3V (red circles), I3A (blue diamonds), V5A (yellow diamonds), T7A (turquoise triangles), T9A (purple triangles); (**B**) wild-type (black circles), I13V (red circles), I13A (blue diamonds), L15A (yellow diamonds), V17A (turquoise triangles), D21N (purple triangles); (**C**) wild-type (black circles), T22A (red circles), I23V (blue diamonds), I23A (yellow diamonds), I23G (turquoise triangles); (**D**) wild-type (black circles), V26A (red circles), K27A (yellow diamonds), A28G (turquoise triangles), I30V (purple triangles); (**E**) wild-type (black circles), I30A (red circles), Q41A (blue diamonds), L43A (yellow diamonds), L50A (turquoise triangles), L56A (purple triangles); (**F**) wild-type (black circles), I61V (red circles), I61A (blue diamonds), L67A (yellow diamonds), L69A (turquoise triangles). Solid lines indicate the best fit of the data Equation 1.

transition state relative to the denatured state. In this case, the transition state can be said to be largely unstructured in the region of the mutation. Conversely, when $\Phi_F = 1$, the interaction energy lost upon mutation is the same in the native and transition states suggesting that the transition state is highly structured in the region of the mutation. Fractional ϕ-values are, in general, more difficult to interpret and may result from a number of different situations. A simple interpretation is that they result from partial structure formation in the region of mutation in the transition state; however, they may also be due to parallel or multiple pathways in which the interaction is fully formed in some but not others. Brønsted analysis has been used in some cases to distinguish between these cases (Itzhaki *et al*., 1995; Fulton *et al*., 1999) and, in conjunction with computational studies, there is strong evidence that distinct folding pathways are rare and that, for small proteins, the transition state ensemble measured experimentally reflects a number of closely related pathways. In the simplest cases, where mutations have been made from non-polar to non-polar residues fractional, Φ_F-values are a measure of the degree to which interactions have formed in the transition state (Itzhaki *et al*., 1995).

For ubiquitin, we obtain many Φ_F-values of zero or one which are straightforward to interpret (Table II). In addition, we obtain a very consistent pattern of values throughout the structure (Table II). Table III summarizes the positions of the mutated residues, the primary interactions that are disrupted upon mutation, the burial of the residue and the measured

Table II. Φ-values calculated from unfolding and refolding data

Mutant	Φ	
	Unfolding data	Refolding data
Ile→Val3	0.2	0.4
Ile→Ala3	0.3	0.4
Val→Ala3[a]	0.4	0.4
Val→Ala5	0.5	0.6
Thr→Ala7	0.8	0.6
Thr→Ala9[b]	–	–
Ile→Val13	0.3	0.4
Ile→Ala13	0.5	0.4
Val→Ala13[a]	0.6	0.4
Leu→Ala15	0.5	0.4
Val→Ala17	0.5	0.6
Asp→Asn21	1.1	1.2
Thr→Ala22	0.5	0.6
Ile→Val23	0.5	0.9
Ile→Ala23	0.4	0.3
Ile→Gly23	0.5	0.5
Val→Ala23[a]	0.4	0.2
Ala→Gly23[a]	0.8	1.2
Val→Ala26	0.3	0.4
Lys→Ala27	0.1	–0.1
Ala→Gly28	1.1	1.1
Ile→Val30	0.5	0.5
Ile→Ala30	0.5	0.3
Val→Ala30[a]	0.6	0.2
Gln→Ala41	0.0	0.1
Leu→Ala43	0.3	0.0
Leu→Ala50	0.0	0.0
Leu→Ala56	0.1	0.0
Ile→Val61	0.1	0.1
Ile→Ala61	0.0	0.1
Val→Ala61[a]	–0.1	0.1
Leu→Ala67	0.0	0.1
Leu→Ala69	0.3	0.1

[a]Φ-values have been calculated for a 'virtual' mutation from data from two mutants of the same residue using a fine-structure analysis.
[b]Φ not calculated for this mutant as $\Delta\Delta G_{D\text{-}N}$ is close to 0.

Φ_F-values. Essentially, the Φ_F-values can be split into three main groups, those that are close to one, close to zero and fractional values. This is illustrated in Figure 2.

Φ_F = 1. Three mutants have Φ_F-values of 1, indicating a significant amount of structure in these regions in the transition state ensemble. These are residues Asp21, whose side chain forms a hydrogen bond within the turn region going into the α-helix, Ala→Gly23 (a composite mutation, values for which are calculated from data from the two mutants Ile→Gly23 and Ile→Ala23) and Ala→Gly28 (Figure 2). Together, these results indicate that the turn leading into the α-helix, and the α-helix itself, are largely formed in the transition state ensemble.

Fractional Φ_F-values. Many residues within the N-terminal region of the protein (residues 1–36) have fractional Φ_F-values (Table III and Figure 2). These all probe a combination of secondary and tertiary structure formation and the extent to which the N-terminal region of the hydrophobic core is formed. Values range from 0.3 to 0.7, averaging 0.5, indicating that ~50% of the favourable interaction energy found in the native state is present in the transition state ensemble. It is difficult to definitively obtain structures from these results but the data strongly suggest a TSE with a relatively well-formed helix and possibly β-hairpin, but in which packing of the two is significantly weaker than in the native state.

Φ_F = 0. There are many residues, all lying within the C-terminal region (residues 37–76) of the protein, which have values of zero or close to zero (Figure 2). Hence the C-terminal region of ubiquitin appears to be highly disordered in the transition state ensemble.

Together, these results present a very clear picture of the transition state ensemble of ubiquitin which is highly polarized. Our results show that the central α-helix is largely structured in the transition state, with the first β-hairpin being slightly less so. Although these elements of secondary structure are formed, the packing between them is somewhat weaker than in the native state. However, this packing is likely to be important in stabilizing the secondary structural elements in the N-terminal region as they form. In contrast to the N-terminal region, the C-terminal part of the protein is largely unstructured in the transition state ensemble and appears to fold only after the protein has reached the top of its energy barrier.

With such a detailed picture of the transition state ensemble generated from experimental data, we can begin to compare our results with those from other experimental and computational studies.

Comparison with other experimental studies

Ubiquitin has long been a popular model for folding studies: the Roder group published extensively in the early 1990s with the first studies on the folding kinetics (Briggs and Roder, 1992; Khorasanizadeh *et al*., 1993, 1996). Originally, ubiquitin was thought to be a classic example of a small protein which populated an intermediate state during folding (Khorasanizadeh *et al*., 1996). However, more recent studies have suggested this may not be the case (Krantz and Sosnick, 2000). Recent work from our laboratory has shown that the pseudo-wild-type ubiquitin used in this study folds with two-state kinetics which can only be switched to three-state in the presence of a stabilizing salt such as sodium sulphate (Went *et al*., 2004). In this study, the folding of ubiquitin is two-state and analysed as such. In addition to the many experimental studies on the folding of full-length ubiquitin, there is also a wealth of information on the tendency of isolated peptides of ubiquitin to form structure in solution (Searle *et al*., 1995; Zerella *et al*., 1999; 2000; Jourdan and Searle, 2000; Jourdan *et al*., 2000), and also on the structure of the 'A' state, a partially structured state observed in methanol (Harding *et al*., 1991; Cox *et al*., 1993; Woolfson *et al*., 1993; Jourdan and Searle, 2001; Prompers *et al*., 2001). For ubiquitin, there is very good agreement on the elements we find structured in the transition state ensemble from the φ-value analysis and those peptides which have a tendency to form structure in isolation, — namely the first β-hairpin and the α-helix. In addition, there is good agreement between the Φ_F-values and the most structured region found within the 'A' state (Cox *et al*., 1993). Molecular dynamic simulations of this state show that there are highly collective motions on the nanosecond time-scale corresponding to large-scale movements between segments (Prompers *et al*., 2001). As the transition state ensemble has a similar structure to this 'A' state, it is likely that it also has such collective motions and is significantly more dynamic than the native state.

Table III. Summary of the positions and interactions of mutated side chains

Mutant	Position	Primary interactions	% side chain solvent-accessible surface area	Average Φ_F
Ile→Val3	β2	Interaction between β2 and β1/β3	0	0.3
Val→Ala3	β2	Interaction between β2 and β1, turn 8/9	0	0.4
Val→Ala5	β2	Interaction between β2 and β1/β3/β4, α-helix	0	0.5
Thr→Ala7	β2	Interaction between β2, turn 1 and β1/β3	16.5	0.7
Thr→Ala9	Turn 1	Interaction between β1 and β2 in β-hairpin	92.9	-
Ile→Val13	β1	Probes interactions within β1 and with C-terminus of α-helix	3.9	0.4
Val→Ala13		Interactions between β1 and β2 and C-terminus of α-helix	3.9	0.5
Leu→Ala15	β1	Interactions within β1, between β1 and β2 and with α-helix	4.0	0.5
Val→Ala17	β1	Interaction with β2, α-helix, 3_{10} helix and turn 2	2.8	0.6
Asp→Asn21	Turn 2	Interactions within turn and with β1 and with a-helix	30.6	1
Thr→Ala22	Turn 2	Interactions with N-terminal end of α-helix and turn 6	47.5	0.5
Ile→Val23	α-Helix	Interaction with β5 and 3_{10} helix	0	0.7
Val→Ala23	α-Helix	Interactions within α-helix, with β4/β5 and turn 5	0	0.3
Ala→Gly23	α-Helix	α-Helix formation and interaction with turn 5	0	1
Val→Ala26	α-Helix	Interactions within α-helix, with β1, β4 and 3_{10} helix	0	0.3
Lys→Ala27	α-Helix	Interaction within α-helix, with β4 and turn 5	9.7	0
Ala→Gly28	α-Helix	α-Helix formation	62.9	1
Ile→Val30	α-Helix	Interactions within helix and with β1/β2/β4	0.1	0.5
Val→Ala30	α-Helix	Interactions between α-helix and β1/β3/β4	0.1	0.3
Gln→Ala41	β4	Interactions with a-helix, β3 and turn 3	0	0.1
Leu→Ala43	β4	Interactions with α-helix and β2 and β3	0	0.1
Leu→Ala50	β5	Interactions with β3 and β4, 3_{10} helix and N-terminal of α-helix	0	0
Leu→Ala56	3_{10} helix	Interactions with α-helix, β1/β2, turns 2 and 8/9	0	0.1
Ile→Val61	Turn 8/9	Interactions with 3_{10} helix and β3/β4/β5	0	0.1
Val→Ala61	Turn 8/9	Interactions with β2/β3/β4 and 3_{10} helix	0	0
Leu→Ala67	β3	Interactions with β2/β4/β5 and turn 8/9	0	0
Leu→Ala69	β3	Interactions within β3, with β2 and β4 and with α-helix	0.7	0.2

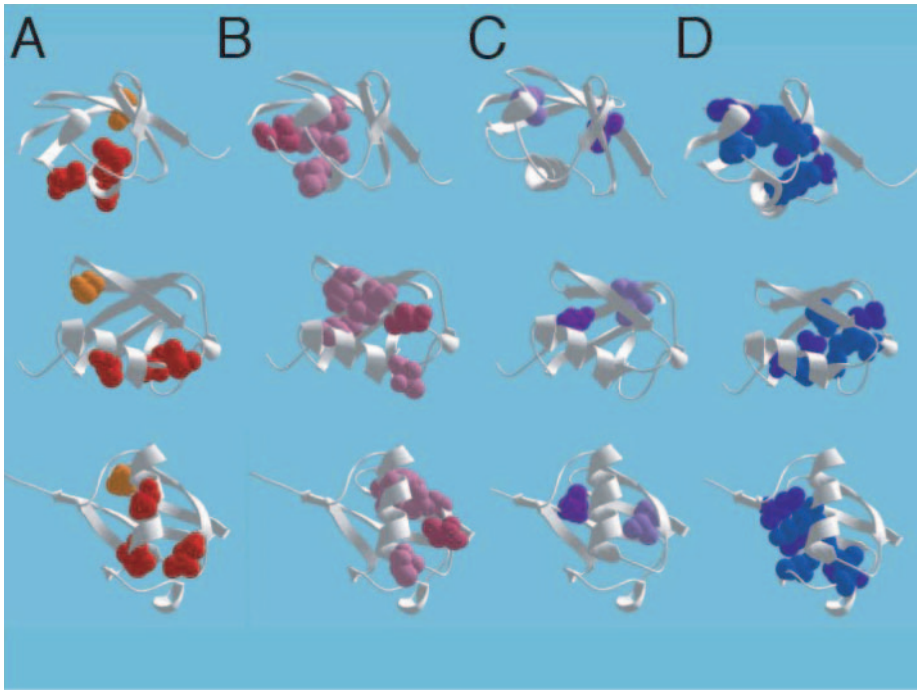

Fig. 2. Results from the ϕ-value analysis of the transition state ensemble for folding of ubiquitin. The structure of ubiquitin is shown in three different orientations. (**A**) Residues with high ϕ-values are shown in red ($\phi \approx 1$) and orange ($\phi = 0.7$) and probe α-helical formation and structure in the first β-turn, respectively; (**B**) residues with medium ϕ-values are shown in light pink ($\phi = 0.5$) and dark pink ($\phi = 0.6$) and probe tertiary interactions in the N-terminal region of the hydrophobic core; (**C**) residues with medium/low ϕ-values are shown in light purple ($\phi = 0.3$) and dark purple ($\phi = 0$) and probe tertiary interactions in the N-terminal region of the hydrophobic core; (D) residues with low ϕ-values are shown in dark blue ($\phi \leqslant 0.2$) and probe tertiary interactions in the C-terminal region of the hydrophobic core.

H/D exchange experiments on ubiquitin have been used to detect and characterize possible folding intermediates (Briggs *et al*., 1992; Gladwin and Evans, 1996), and also to measure unfolding rate constants in native-like conditions (Sivaraman *et al*., 2001). Little protection is seen within the deadtime of these experiments, suggesting that, for ubiquitin, secondary structural elements can only be populated marginally ahead of or at the same time as tertiary structure formation (Gladwin and Evans, 1996). Our results suggest that significant secondary structure formation only occurs within the N-terminal

region as it reaches its transition state and that this secondary structure is, in part, stabilized by tertiary interactions between the helix and first β-hairpin.

Sosnick and co-workers have used solvent isotope effects to determine the extent of backbone hydrogen bond formation in the transition state ensemble (TSE) (Krantz *et al*., 2000). They conclude that few hydrogen bonds are formed in the TSE. Although we do not probe backbone hydrogen bond formation directly, we have a number of probes of secondary structure formation. In our studies, these indicate that the first β-hairpin and α-helix are largely formed in the TSE. It is hard to imagine the formation of secondary structure without the concomitant formation of backbone hydrogen bonds. In contrast to the Sosnick study, we therefore conclude that a significant number of backbone hydrogen bonds are formed in the TSE.

The temperature dependence of hydrogen bonds in ubiquitin has been monitored by NMR spectroscopy and used to establish the regions of the protein that are stable towards thermal expansion/unfolding (Cordier and Grzesiek, 2002). Results show that β-strand 5 has a very low thermal stability whereas β1, β2 and the α-helix all appear to be much more resilient towards temperature. This is in good agreement with our results that show that the region encompassing β1–β2 and the α-helix is the most structured in the TSE. Hence the folding nucleus appears to be formed from the most inherently stable region of substructure within the protein.

Recently, the equilibrium pressure-induced unfolding and cold denaturation of ubiquitin have been probed with NMR spectroscopy (Jonas, 2002; Kitahara and Akasaka, 2003; Babu *et al*., 2004). The pressure-induced denaturation experiments reveal an intermediate state in which there has been a local unfolding event in the region of residues 33–42 and 70–76. We only have one φ-value which specifically probes this region, Gln41, which is close to zero, consistent with this region unfolding early. A novel method for measuring the cold unfolding of proteins in water at sub-zero temperatures has recently been published by Wand and co-workers (Babu *et al*., 2004). They used this method to study the cold unfolding of ubiquitin and found a stable core, based around the N-terminal region of the helix and the β-sheet that packs against it, that is resistant to denaturation. In general, this is in good agreement with our experimental results, although there are some slight discrepancies,— such as the stability of the 3_{10}-helix and the β-turn between strands 1 and 2. Hence the regions that are found to be stable against cold denaturation are, in general, the same as the regions which we find constitute the folding nucleus and which fold early and, therefore, unfold late on the folding pathway.

Recent work from the Fernandez group has unfolded ubiquitin using mechanical force and followed unfolding/folding transitions by atomic force microscopy (Carrion-Vazquez *et al*., 2003; Fernandez and Li, 2004). Although these studies have yet to shed much light on the folding pathway, they have been very instructive in illustrating that the mechanical strength of ubiquitin depends strongly on the linkage between domains in the polyprotein (Carrion-Vazquez *et al*., 2003).

Perhaps the most relevant experimental work to the study we present here is a recent study from the Sosnick group characterizing the folding pathway of ubiquitin using a novel protein engineering method and ψ-value analysis (Krantz *et al*., 2004). In this approach, bi-histidine sites are engineered throughout the protein and metal binding to these sites is used to perturb the energy levels in a manner similar to mutations perturbing energy levels in φ-value analysis. From their data, Sosnick and co-workers conclude that ubiquitin folds with a single folding nucleus, the TSE having a common obligate core comprising the α-helix and four out of the five β-strands. In this case, they conclude that there is considerably more structure in the transition state, which has a very native-like topology, than is apparent from either their φ-value analysis on seven residues (3, 15, 30, 43, 50, 56 and 67) (Sosnick *et al*., 2004) or from our data presented here (note: our φ-values are in excellent agreement with their φ-values, but here we have significantly extended the φ-value analysis). From this, they conclude that φ-values underestimate the amount of structure in the TSE, resulting from the fact that fractional φ-values are difficult to interpret because of heterogeneity in the folding pathways and the possible change in flux through pathways on destabilization of a specific pathway by mutation, relaxation effects in the TSE to accommodate the mutation and Hammond and anti-Hammond behaviour (Sosnick *et al*., 2004). Recently, however, Fersht has reassessed the original ψ-value analysis published by Sosnick and co-workers and shown that some of the equations are incorrect due to omission of saturation terms for metal binding (Fersht, 2004). This has led to the publication of an erratum on the original ubiquitin ψ-value analysis paper (Krantz *et al*., 2005). In addition, Fersht has shown that 'ψ-value analysis does not give information *per se* on transition state heterogeneity and ψ does not respond linearly with φ or the degree of formation of structure'. A straight comparison of φ- and ψ-values is, therefore, not valid. In addition, Fersht points out that if there is any plasticity in the transition state which can accommodate the crosslink formed by the metal and bi-histidines, then ψ-values would be significantly greater than zero even for an unstructured region, leading to an overestimation of structure in the transition state (Fersht, 2004). Because of this and the fact that our φ-values agree well with most of the other experimental and computational data available, we are confident in our conclusion that ubiquitin folds with a highly polarized transition state with little structure in the C-terminal region of the protein.

Comparison with computational studies

The large amount of experimental data on the folding pathway of ubiquitin, in conjunction with its relatively small size, has made it the focus of many different computational studies simulating folding pathways. A wide variety of approaches have been employed and, in some cases, the region(s) of ubiquitin that are predicted to fold early or form the folding nucleus, on the basis of the simulations are known allowing a direct comparison with our experimental φ-values. Although quantitative comparisons between experimental and computational results have been undertaken for some studies (Li and Daggett, 1994, 1996), here we compare results on a qualitative basis only.

All-atom molecular dynamic (MD) simulations of ubiqiutin have been undertaken by the Daggett group (Alonso and Daggett, 1995, 1998). In the first study, MD was used to probe the structure and dynamics of the partially structured state observed experimentally in 60% MeOH. In this case, the simulations agreed well with experimental results and correctly predicted the H/D exchange protection pattern observed. In the subsequent study, the unfolding of ubiquitin was induced by high temperatures, the MD simulations were then quenched at various times down to folding temperatures and the collapse of

the polypeptide chain back towards the native structure followed. The aim of these simulations was not to characterize a transition state ensemble but to establish the degree to which the protein could be refolded after unfolding to states with different degrees of structure. Qualitatively, however, the simulations can be compared to our experimental φ-values. In the simulations, one of the first events observed is the loss in a number of tertiary interactions including the docking between the α-helix and the β-sheet. This is in good agreement with our results, which show that although the helix remains largely intact, the tight packing of the core—in both the N- and C-terminal regions of the protein—has been lost. In addition, the simulations show that the interactions between parallel strands β1 and β5 are also lost early, as is the interaction between β1 and β2. Although, the latter observation does not agree with our results, it is interesting to note that, in the simulations, whereas the interaction between β1 and β5 once lost is never regained, the interaction between β1 and β2 fluctuates over time.

From many bioinformatics studies, it is clear that specific residues in proteins are highly conserved, either for function (in active sites), molecular recognition (protein–ligand, protein–protein, protein–substrate binding sites) or stability. It has also been proposed that residues involved in the folding nucleus should also be highly conserved to ensure fast and efficient folding (Shakhnovich *et al*., 1996; Mirny and Shakhnovich, 2001). Using this premise, Shakhnovich and co-workers have developed a computational technique for identifying folding nuclei by sequence comparison of protein superfamilies and a Monte Carlo sequence design strategy. Applied to the ubiquitin superfamily, eight residues have been identified as potential nucleation sites: Ile3, Val5, Leu15, Val17, Ile30, Leu50, Leu67 and Leu69 (Michnick and Shakhnovich, 1998). Five of these are in the N-terminal region and are residues which we are confident are, at least, partially structured in the TSE. The other three, however, are in the C-terminal region of the hydrophobic core. Our φ-values in this region are consistently low, suggesting that this region of the protein is not involved in stabilizing the TSE. It is interesting that two recent papers (Plaxco *et al*., 2000; Larson *et al*., 2002) suggest that there is little correlation between the degree of conservation of residues and their involvement in any folding nucleus. Instead, a significant correlation is observed between the contributions of individual sequence positions and the transition state structure across a set of homologous proteins (Plaxco *et al*., 2000). From this, it is suggested that the structure of the TSE is more conserved than the specific interactions that stabilize it.

Many different semi-empirical methods have been used to probe the folding pathway of ubiquitin, including one recently developed by Fernandez (2001) which takes into account desolvation events during folding. Qualitatively, this approach captures some of the features we observe experimentally—a nucleating state is formed concomitant with a collapse-inducing topology in which the secondary structure present is stabilized by tertiary interactions. The nucleation site predicted by these simulations involves both N- and C-terminal regions of the protein, with β-strands 1 and 5 associating. We see little evidence for the involvement of the C-terminal region of the protein in the TSE, all φ-values reporting on β-strand 5 formation being close to zero. More recent simulations from the Fernandez group, however, show a much better agreement with experimental results (Fernandez, 2002; Fernandez and Colubri, 2002; Fernandez *et al*., 2002). In these simulations, coarse-grained approaches are used which consider the torsional evolution of the polypeptide chain and the geometric constraints that are imposed on backbone motion from steric clashes, in addition to pairwise interactions dependent upon solvation. Again, these simulations capture qualitatively some of the features of the energy landscape that we infer from experimental data. In particular, one of the studies predicts the early formation of the first β-hairpin and part of the α-helix, consistent with our results. The importance of three-body correlations in the folding pathway of ubiquitin has also been shown in a separate simulation (Sosnick *et al*., 2002), which demonstrates that, using this model, there is a nucleating event in the folding of ubiquitin after which the number of three-body correlations and progression to the native state occur rapidly, consistent with the nucleation–condensation model for ubiquitin that we propose.

Other computational approaches have included the prediction of early events in folding by studying the intrinsic stability of fragments of proteins likely to constitute independent folding units (Gilis and Rooman, 2001). In this case, residues 1–35 of ubiquitin, which comprise the β-hairpin and α-helix, are identified, in excellent agreement with results from fragment studies and our φ-value analysis. In addition, off-lattice bead models with simple energy functions give very good agreement with the φ-values presented here (Sorensen and Head-Gordon, 2002). In this case, the major α-helix and β-strands 1–3 are well formed in a putative transition state structure, the first part of the β-sheet also forming early. In excellent agreement with our experimental studies, these simulations also show that the C-terminal region is the last to fold having only a 'spectator' role in folding (Sorensen and Head-Gordon, 2002).

The Baker group has used an alternative approach in which residues are considered simply as ordered or not and in which only contiguous segments can fold. Their algorithm takes into account the entropic cost of ordering residues which is compensated for by the formation of favourable native interactions and hydrogen bonds as well as incorporating parameters describing backbone torsional strain. The results for ubiquitin are available on the Baker group website (http://tools.bakerlab.org/~predictions/cgi-bin/test.cgi; PDB code: 1ubq). In general, this algorithm predicts that many residues in the C-terminal half of the protein are involved in the formation of the folding nucleus, in contrast with our experimental φ-values.

Conclusions

A comprehensive φ-value analysis on the folding of ubiquitin generates a clear picture of the structure and energetics of the transition state ensemble. We find that the protein folds in accordance with a nucleation–condensation mechanism in which there is a highly polarized folding nucleus. This nucleus comprises several elements of secondary structure including the major α-helix, which appears to be fully formed in the TSE, in addition to the N-terminal β-hairpin. There is evidence to suggest that, although the elements of secondary structure are highly structured, the packing between them is somewhat weakened in the TSE compared with the native state. In contrast to the N-terminal region of ubiquitin, we find that the C-terminal ca 30 residues contribute little to the TSE, in which they are largely unstructured. The folding of ubiquitin also contains some features consistent with a diffusion-collision model, in which regions of the protein with high inherent secondary

structure propensity form early and then collide and consolidate in the transition state.

By comparing results from φ-value analysis with results from other experimental studies using complementary approaches, we can obtain an even more detailed picture of the folding pathway. Fragment studies on peptides of ubiquitin have been very useful in determining the intrinsic conformational propensities of different regions of the protein and their stabilities. These studies have shown that the N-terminal β-hairpin and α-helix can form transient native-like structure in the absence of the C-terminal region, whereas the C-terminal region on its own has little propensity for forming structure. Hence the early steps in the folding of ubiquitin are likely to involve the transient formation of native-like secondary structure in the N-terminal half of the protein. We speculate that this transient structure fluctuates until it is stabilized by tertiary contacts, at which point a folding nucleus is formed and the transition state is reached. A comparison of the folding pathway of ubiquitin with results from experiments conducted under equilibrium conditions suggests that regions of the protein which form the folding nucleus also form stable sub-structures which are resistant to some forms of denaturation such as pressure and cold-induced denaturation. In cases where folding becomes less cooperative, therefore, these studies can be used to provide information on the likely nucleation sites for folding.

A comparison was also made of the folding pathway as determined from our experimental φ-values with the myriad of computational studies on the folding pathway of this protein. In general, there is reasonable agreement between the two, the computational studies capturing many of the general features of the folding process, if not always predicting correctly the finer details.

The results presented here, along with the comprehensive comparison, leads us to conclude that many different experimental and computational approaches to studying folding are valid and provide complementary information on the folding pathway and insights into different regions of the complex energy landscape. Combined, these different experimental and computational approaches provide a robust and detailed picture of the folding landscape.

Acknowledgements

The work was supported, in part, by the BBSRC and The Welton Foundation. We thank Professor Sir Alan Fersht and Professor Chris Dobson for access to biophysical equipment and helpful discussions. We also thank Tobin Sosnick for many stimulating discussions.

References

Alonso,D.O.V. and Daggett,V. (1995) *J. Mol. Biol.*, **247**, 501–520.
Alonso,D.O.V. and Daggett,V. (1998) *Protein Sci.*, **7**, 860–874.
Babu,C.R., Hilser,V.J. and Wand,A.J. (2004) *Nat. Struct. Mol. Biol.*, **11**, 352–357.
Briggs,M.S. and Roder,H. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2017–2021.
Carrion-Vazquez,M., Li,H., Lu,H., Marszalek,P.E., Oberhauser,A.F. and Fernandez,J.M. (2003) *Nat. Struct. Biol.*, **10**, 738–743.
Cordier,F. and Grzesiek,S. (2002) *J. Mol. Biol.*, **317**, 739–752.
Cox,J.P.L., Evans,P.A., Packman,L.C., Williams,D.H. and Woolfson,D.N. (1993) *J. Mol. Biol.*, **234**, 483–492.
Ermolenko,D.N., Thomas,S.T., Aurora,R., Gronenborn,A.M. and Makhatadze,G.I. (2002) *J. Mol. Biol.*, **322**, 123–135.
Ermolenko,D.N., Richardson,J.M. and Makhatadze,G.I. (2003) *Protein Sci.*, **12**, 1169–1176.
Fernandez,A. (2001) *J. Chem. Phys.*, **114**, 2489–2502.
Fernandez,A. (2002) *Proteins*, **47**, 447–457.
Fernandez,A. and Colubri,A. (2002) *Proteins*, **48**, 293–310.
Fernandez,J.M. and Li,H. (2004) *Science*, **303**, 1674–1678.
Fernandez,A., Colubri,A. and Berry,R.S. (2002) *Physica A*, **307**, 235–259.
Fersht,A.R. (2004) *Proc. Natl Acad. Sci. USA*, **101**, 17327–17328.
Fersht,A.R., Matouschek,A. and Serrano,L. (1992) *J. Mol. Biol.*, **224**, 771–782.
Fulton,K.F., Main,E.R.G., Daggett,V. and Jackson,S.E. (1999) *J. Mol. Biol.*, **291**, 445–461.
Gilis,D. and Rooman,M. (2001) *Proteins*, **42**, 164–176.
Gladwin,S.T. and Evans,P.A. (1996) *Fold. Des.*, **1**, 407–417.
Harding,M.M., Williams,D.H. and Woolfson,D.N. (1991) *Biochemistry*, **30**, 3120–3128.
Itzhaki,L.S., Otzen,D.E. and Fersht,A.R. (1995) *J. Mol. Biol.*, **254**, 260–288.
Jackson,S.E. and Fersht,A.R. (1991) *Biochemistry*, **30**, 10428–10435.
Jackson,S.E., elMasry,N. and Fersht,A.R. (1993) *Biochemistry*, **32**, 11270–11278.
Jonas,J. (2002) *Biochim. Biophys. Acta*, **1595**, 145–159.
Jourdan,M. and Searle,M.S. (2000) *Biochemistry*, **39**, 12355–12364.
Jourdan,M. and Searle,M.S. (2001) *Biochemistry*, **40**, 10317–10325.
Jourdan,M., Griffiths-Jones,S.R. and Searle,M.S. (2000) *Eur. J. Biochem.*, **267**, 3539–3548.
Khorasanizadeh,S., Peters,I.D., Butt,T.R. and Roder,H. (1993) *Biochemistry*, **32**, 7054–7063.
Khorasanizadeh,S., Peters,I.D. and Roder,H. (1996) *Nat. Struct. Biol.*, **3**, 193–205.
Kitahara,R. and Akasaka,K. (2003) *Proc. Natl Acad. Sci. USA*, **100**, 3167–3172.
Krantz,B.A. and Sosnick,T.R. (2000) *Biochemistry*, **39**, 11696–11701.
Krantz,B.A., Moran,L.B., Kentsis,A. and Sosnick,T.R. (2000) *Nat. Struct. Biol.*, **7**, 62–71.
Krantz,B.A., Dothager,R.S. and Sosnick,T.R. (2004) *J. Mol. Biol.*, **337**, 463–475.
Krantz,B.A., Dothager,R.S. and Sosnick,T.R. (2005) *J. Mol. Biol.*, **347**, 1103.
Larson,S.M., Ruczinski,I., Davidson,A.R., Baker,D. and Plaxco,K.W. (2002) *J. Mol. Biol.*, **316**, 225–233.
Li,A.J. and Daggett,V. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 10430–10434.
Li,A.J. and Daggett,V. (1996) *J. Mol. Biol.*, **257**, 412–429.
Loladze,V.V. and Makhatadze,G.I. (2002) *Protein Sci.*, **11**, 174–177.
Loladze,V.V., Ibarra-Molero,B., Sanchez-Ruiz,J.M. and Makhatadze,G.I. (1999) *Biochemistry*, **38**, 16419–16423.
Loladze,V.V., Ermolenko,D.N. and Makhatadze,G.I. (2002) *J. Mol. Biol.*, **320**, 343–357.
Main,E.R.G., Fulton,K.F. and Jackson,S.E. (1999) *J. Mol. Biol.*, **291**, 429–444.
Makhatadze,G.I., Loladze,V.V., Ermolenko,D.N., Chen,X.F. and Thomas,S.T. (2003) *J. Mol. Biol.*, **327**, 1135–1148.
Matouschek,A., Kellis,J.T.,Jr, Serrano,L. and Fersht,A.R. (1989) *Nature*, **342**, 122–126.
Matthews,C.R. (1987) *Methods Enzymol.*, **154**, 489–511.
Michnick,S.W. and Shakhnovich,E. (1998) *Fold. Des.*, **3**, 239–251.
Mirny,L. and Shakhnovich,E. (2001) *J. Mol. Biol.*, **308**, 123–129.
Oliveberg,M. (2001) *Curr. Opin. Struct. Biol.*, **11**, 94–100.
Plaxco,K.W., Larson,S., Ruczinski,I., Riddle,D.S., Thayer,E.C., Buchwitz,B., Davidson,A.R. and Baker,D. (2000) *J. Mol. Biol.*, **298**, 303–312.
Prompers,J.J., Scheurer,C. and Bruschweiler,R. (2001) *J. Mol. Biol.*, **305**, 1085–1097.
Searle,M.S., Williams,D.H. and Packman,L.C. (1995) *Nat. Struct. Biol.*, **2**, 999–1006.
Serrano,L., Matouschek,A. and Fersht,A.R. (1992) *J. Mol. Biol.*, **224**, 805–818.
Shakhnovich,E.I., Abkevich,V. and Ptitsyn,O. (1996) *Nature*, **379**, 96–98.
Sivaraman,T., Arrington,C.B. and Robertson,A.D. (2001) *Nat. Struct. Biol.*, **8**, 331–333.
Sorensen,J.M. and Head-Gordon,T. (2002) *Proteins*, **46**, 368–379.
Sosnick,T.R., Berry,R.S., Colubri,A. and Fernandez,A. (2002) *Proteins*, **49**, 15–23.
Sosnick,T.R., Dothager,R.S. and Krantz,B.A. (2004) *Proc. Natl Acad. Sci. USA*, **101**, 17377–17382.
Tanford,C. (1968) *Adv. Protein Chem.*, **23**, 121–282.
Tanford,C. (1970) *Adv. Protein Chem.*, **24**, 1–95.
Went,H.M., Benitez-Cardoza,C.B. and Jackson,S.E. (2004) *FEBS Lett.*, **567**, 333–338.
Woolfson,D.N., Cooper,A., Harding,M.M., Williams,D.H. and Evans,P.A. (1993) *J. Mol. Biol.*, **229**, 502–511.
Zerella,R., Evans,P.A., Ionides,J.M.C., Packman,L.C., Trotter,B.W., Mackay,J.P. and Williams,D.H. (1999) *Protein Sci.*, **8**, 1320–1331.
Zerella,R., Chen,P.Y., Evans,P.A., Raine,A. and Williams,D.H. (2000) *Protein Sci.*, **9**, 2142–2150.

Received February 11, 2005; accepted March 22, 2005

Edited by Elizabeth Meiering