The Effect of Increasing the Stability of Non-native Interactions on the Folding Landscape of the Bacterial Immunity Protein Im9

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · SEPTEMBER 2007

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2007.05.010 · Source: PubMed

5 AUTHORS, INCLUDING:

Claire Friel
University of Nottingham
23 PUBLICATIONS **691** CITATIONS
SEE PROFILE

Emanuele Paci
University of Leeds
123 PUBLICATIONS **6,318** CITATIONS
SEE PROFILE

Sheena E Radford
University of Leeds
268 PUBLICATIONS **15,794** CITATIONS
SEE PROFILE

doi:10.1016/j.jmb.2007.05.010 J. Mol. Biol. (2007) 371, 554–568

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

The Effect of Increasing the Stability of Non-native Interactions on the Folding Landscape of the Bacterial Immunity Protein Im9

Victoria L. Morton[1,2], Claire T. Friel[1,2], Lucy R. Allen[1,3] Emanuele Paci[1,3] and Sheena E. Radford[1,2]*

[1]*Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK*

[2]*Institute of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT, UK*

[3]*Institute of Molecular Biophysics, University of Leeds, Leeds, LS2 9JT, UK*

**Corresponding author*

How stabilising non-native interactions influence protein folding energy landscapes is currently not well understood: such interactions could speed folding by reducing the conformational search to the native state, or could slow folding by increasing ruggedness. Here, we examine the influence of non-native interactions in the folding process of the bacterial immunity protein Im9, by exploiting our ability to manipulate the stability of the intermediate and rate-limiting transition state (TS) in the folding of this protein by minor alteration of its sequence or changes in solvent conditions. By analysing the properties of these species using Φ-value analysis, and exploration of the structural properties of the TS ensemble using molecular dynamics simulations, we demonstrate the importance of non-native interactions in immunity protein folding and demonstrate that the rate-limiting step involves partial reorganisation of these interactions as the TS ensemble is traversed. Moreover, we show that increasing the contribution to stability made by non-native interactions results in an increase in Φ-values of the TS ensemble without altering its structural properties or solvent-accessible surface area. The data suggest that the immunity proteins fold on multiple, but closely related, micropathways, resulting in a heterogeneous TS ensemble that responds subtly to mutation or changes in the solvent conditions. Thus, altering the relative strength of native and non-native interactions influences the search to the native state by restricting the pathways through the folding energy landscape.

Keywords: protein folding; intermediate; transition state; Φ−value; molecular dynamics

Introduction

An understanding of the folding energy landscape of a protein sequence requires all species formed to be characterised structurally and thermodynamically, from the initial unfolded state to the final native structure. [1] To understand how proteins fold, therefore, studies have focused on the structural characterisation of accessible regions of the folding landscape, including unfolded, partially folded and rate-limiting transition state (TS) ensembles, using a variety of biophysical methods.[2,3] Commonly, details about the structure of intermediates and TS ensembles have been inferred indirectly from the study of the kinetics of folding and unfolding of wild-type and mutated protein sequences using Φ-value analysis.[4] By employing experimental Φ-values as restraints in molecular dynamics (MD) simulations, direct structural models of these states can then be generated.[5–8] Together, these techniques

Present address: C. T. Friel, Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauer Str. 108, Dresden 01307, Germany.

Abbreviations used: TS, transition state; SIm9, Im9 containing the mutations V37L and V71I; U, unfolded state; I, intermediate state; N, native state; β_X, the β-Tanford value for species X; *SASA*, the solvent accessible surface area.

E-mail address of the corresponding author: s.e.radford@leeds.ac.uk

are beginning to shed light on the conformational properties of non-native species formed during folding, including the number and identity of native and non-native interactions at different stages of folding, as well as the conformational heterogeneity of different states.[9–12]

Whilst Φ-value analysis and other approaches[13–15] provide structural information about intermediate and TS ensembles, several important questions remain unresolved. These include the influence of non-native interactions on the folding landscape,[9–12] and the role of populated intermediates in defining the search to the native state.[16,17] In addition, the heterogeneity of both the intermediate and TS ensembles is difficult to discern: the partial Φ-values that usually characterise these states are most commonly interpreted as resulting from the heterogeneity of conformations within a single predominant folding route,[18,19] although folding on multiple micropathways to the native structure would also give rise to partial Φ-values.[18,20] These scenarios may be distinguished by measuring the sensitivity of the magnitude of the Φ-values to changes in the amino acid sequence or folding conditions used: folding on parallel pathways should give rise to Φ-values that change in response to alterations in the flux of molecules along different folding routes.[21,22] Such behaviour has been observed in the folding of proteins with symmetric topologies with the result that the apparent location of the folding TS can be switched in different homologues.[23,24] For other proteins, the apparent movement of the folding TS away from the native state upon mutation (anti-Hammond behaviour), or the upwards curvature in the unfolding branch of the chevron plot, have been interpreted as resulting from the presence of co-populated folding/unfolding routes.[25–27] However, proteins that do not display these obvious characteristics may also fold *via* multiple, but closely related, pathways (defined here as micropathways). In such cases, the presence of multiple folding routes is more difficult to discern, but may be manifested in subtle changes in Φ-values upon alteration of the sequence or folding conditions.

The experimental tractability of the bacterial immunity proteins, Im7 and Im9, provides a powerful opportunity to investigate these questions. These proteins have 60% sequence identity and adopt a simple four-helix structure in which three long helices (I, II and IV) pack against a short third helix (III) (Figure 1). Extensive studies using ultra-rapid mixing and stopped-flow fluorescence (at pH 7.0 and 10 °C) have shown that Im7 folds through an on-pathway intermediate that contains three of the four native helices (I, II and IV) stabilised by both native and non-native interactions.[28,29] By contrast, under identical conditions the more stable Im9 folds with a two-state transition without a detectable intermediate state.[30] The folding of Im9 can be readily switched to a three-state mechanism involving a highly populated on-pathway intermediate by decreasing the pH (pH ≤ 6.0),[31] or by introducing only two mutations into its sequence (V37L/V71I) (this variant is called SIm9) (Figure 1), strategically

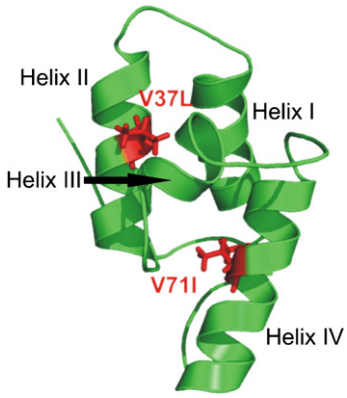

Figure 1. Ribbon diagram showing the structure of Im9. The side-chains altered to create the variant SIm9 are highlighted in red. The Figure was drawn using the program MacPyMol [http://www.pymol.org], using the coordinates 1IMQ.[61]

placed so as to stabilise the intermediate by increasing the stability of non-native interactions involving helices II and IV without altering the stability of the native state substantially ($\Delta\Delta G_{un} = 0.9$ kJ mol^{-1} at pH 7.0, 10 °C). A previously constructed triple mutant V37L/E41V/V71I of Im9 shows similar behaviour.[32] These designed variants of Im9 fold with clear three-state kinetics at neutral pH, SIm9 populating a stable intermediate with a ΔG_{ui} of -11.0 ± 0.1 kJ mol^{-1} (pH 7.0, 10 °C).

The TS ensembles of Im7 and Im9 have been examined using Φ-value analysis, the resulting data indicating that these species have similar structures that contain helices I, II and IV, whilst helix III is absent.[28,33] Interestingly, however, despite their structural similarity, the Φ_{TS}-values for many residues in Im7 exceed significantly the values found for the same residues in Im9 under conditions where the proteins fold with three-state and two-state kinetics, respectively.[28,33] As a consequence, the TS ensemble of Im7 appears less heterogeneous than the TS of Im9, as has been observed in molecular dynamics simulations of these ensembles using the experimental Φ-values as restraints.[34] These data cannot be explained by differences in the placement of the TS ensembles on the folding reaction coordinate relative to the native state, since the solvent-accessible surface area (*SASA*) of both TS (measured by $\beta_{TS} = m_{u\text{-}ts} / M_{un}$, where $m_{u\text{-}ts}$ is the denaturant-dependence of the free energy of folding and M_{un} the denaturant-dependence of the equilibrium stability of the native state) are similar (average β_{TS} for the wild-type proteins and all variants studied at pH 7.0 = 0.94 ± 0.01 and 0.89 ± 0.04 for Im9 and Im7, respectively).[28,33] One possible explanation of the difference in the Φ_{TS} values between Im7 and Im9 is that these proteins fold on multiple closely related micropathways, the flux through which is determined by one or more of the sequence differences in these two proteins.

Here, we investigate the role of non-native interactions in tailoring the energy landscape of immunity protein folding using three different approaches.

Firstly we characterise the thermodynamic properties of the TS ensembles of Im9, SIm9 and Im7 by analysis of the temperature-dependence of the folding and unfolding kinetics. Secondly we carry out a detailed Φ-value analysis of SIm9 at pH 7.0, conditions under which this Im9 variant folds *via* a stable on-pathway intermediate, but the wild-type protein folds with a two-state transition. Finally, using our knowledge of the effect of pH and Na_2SO_4 on the folding properties of Im9,[31,32] we titrate the stabilising contribution of non-native interactions in Im9 and SIm9 folding and determine the Φ-values of the rate-limiting TS ensemble under each condition. By combining the tools of varying the solvent conditions and mutation of the protein sequence, the stability of the intermediate and TS ensembles in Im9 folding can be increased by ~15 kJ $mol^{-1}$and 11 kJ mol^{-1}, respectively, whilst the stability of the native state remains relatively unperturbed. Analysis of the properties of the TS ensemble under each condition using experimental Φ_{TS}-values and restrained MD show that increasing the contribution to stability made by non-native interactions results in a more homogeneous TS ensemble which is characterised by increased Φ_{TS} values. Together with previous analysis of Im7 folding,[28,34] the results suggest that the immunity proteins fold by a common mechanism involving multiple micro-pathways through closely related TS ensembles.

Results

Thermodynamic analysis of the TS ensembles of Im9, SIm9 and Im7

The TS ensembles of Im7 and Im9 have similar structural properties (they are both three-helix species) and similar exposure to the solvent (β_{TS} values ~0.9).[28,33] To analyse the properties of the TS ensembles of these proteins in more detail and to investigate specifically whether the entropy, enthalpy and heat capacity of the TS ensembles of these proteins are similar, the equilibrium stability and folding/unfolding kinetics of these proteins were measured as a function of denaturant concentration and temperature using far-UV CD and stopped-flow fluorescence, respectively. In parallel, the folding and unfolding kinetics, and the equilibrium stability of SIm9 were measured, allowing comparison of all three proteins (Figure 2).

The folding kinetics of Im9 as a function of denaturant concentration, measured from 5–10 °C, is shown in Figure 2(d). The logarithm of the rate constants of folding of this protein depend linearly on the concentration of urea at every temperature analysed, indicating that the protein folds with a two-state transition over the entire temperature range studied. For Im7 and SIm9, by contrast, a distinct roll-over is evident in the denaturant-dependence of the logarithm of the folding rate constants between 10 °C and 20 °C, consistent with these proteins folding *via* a populated intermediate (Figure 2(e) and (f)). At higher temperatures (25–40 °C), the rate of refolding of these proteins becomes too rapid to detect with stopped-flow methods whether an intermediate is still populated. The folding and unfolding kinetics of Im9 *versus* temperature were fit globally to a two-state transition, whilst those for Im7 and SIm9 were fit globally to a three-state transition (see Materials and Methods) and the resulting unfolding rate constants were then used in an Eyring plot to determine the changes in entropy, enthalpy and heat capacity of the unfolding TS barrier relative to the native state (Figure 2(g)–(i); Table 1). The entropy, enthalpy and heat capacity changes for global unfolding were determined for each protein using equilibrium denaturation from 2–88 °C (Figure 2(a)–(c)) and the resulting thermodynamic parameters determined by global fitting of each dataset, allowing accurate determination of the equilibrium denaturation thermodynamics (see Materials and Methods). Equivalence between these parameters and those derived independently from the folding and unfolding kinetics for Im9, which could be accurately measured at all temperatures, confirmed the validity of the fits (data not shown).

The unfolding activation parameters for Im9, SIm9 and Im7 at 10 °C are shown in Figure 3. The results revealed that despite having different global stabilities ($\Delta G_{un} = -27.0 \pm 7.3$ for Im9, $\Delta G_{un} = -25.9 \pm 3.4$ for SIm9, and $\Delta G_{un} = -20.7 \pm 3.7$ for Im7 calculated from the thermodynamic parameters (Table 1)), the rate-limiting TS ensemble of the three proteins has similar free energy (Figure 3(c)), which is comprised both of unfavourable enthalpic and entropic contributions relative to the unfolded state (Figure 3(a) and (b), respectively). Most importantly in the context of the Φ-value analysis described below, the entropy of the TS for Im9 and SIm9 is identical, whilst the entropy of the TS of Im7 is smaller (Figure 3(b)). The heat capacity for all three proteins decreases progressively from the unfolded state through the TS ensemble to the native state (Figure 3(d)). The relative heat capacities ($\beta_{TS}^{Cp} = \Delta Cp_{u\text{-}ts}/\Delta Cp_{un}$) of the TS of Im9, SIm9 and Im7 are 0.6±0.1, 0.4±0.1 and 0.4±0.03, respectively. The data suggest, therefore, that the packing of side-chains in the hydrophobic core of the TS ensembles of all three proteins is far from native, consistent with the low partial Φ_{TS}-values observed for these proteins, and despite their highly compact structures as judged using *m*-value analysis ($\beta_{TS} = 0.91 \pm 0.01$, 0.99±0.01 and 0.89±0.02 for Im9, SIm9 and Im7, respectively). Since β_{TS}^{Cp} and β_{TS} both reflect the burial of non-polar groups, the difference in magnitude of these parameters is surprising, but not unprecedented.[35,36] Good agreement between β_{TS} and β_{TS}^{Cp} has been found for cspB, CI2, HPr, and I27.[37–40,25] However, significant differences in the magnitude of β_{TS} and β_{TS}^{Cp} ($\beta_{TS} - \beta_{TS}^{Cp} > 0.2$) have been found for ubiquitin, FKBP12, protein L[35,41–44] and, as we show here, for the bacterial immunity proteins Im7, Im9 and SIm9. The correlation coefficient between β_{TS} and β_{TS}^{Cp} for all of the proteins in this set is therefore not statistically significant ($r = 0.27$), consistent with

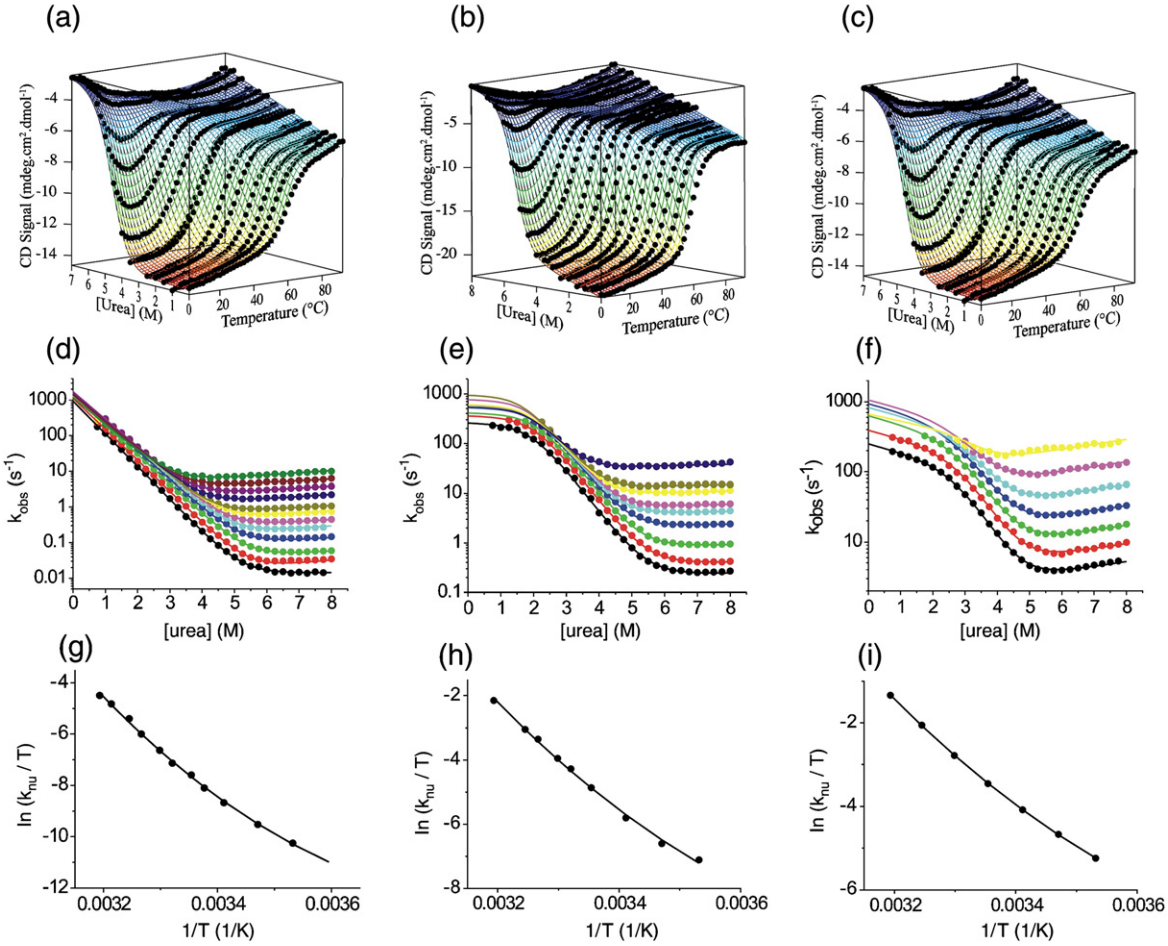

Figure 2. Top: Equilibrium unfolding of (a) Im9, (b) SIm9 and (c) Im7 as a function of temperature and the concentration of urea. The native and denatured baseplanes were defined by the data sets for which there was sufficient data available to give a high degree of confidence in the fit. The data surface was then fitted globally yielding the parameters given in Table 1. Middle: Folding and unfolding kinetics of (d) Im9 at pH 7.0 at (O, 5 °C; O, 10 °C; O, 15 °C; O, 20 °C; O 23 °C; O, 25 °C; O 28 °C; O, 30 °C; O, 33 °C; O, 35 °C; O 38 °C; and O, 40 °C). (e) SIm9 at pH 7.0 at (O, 10 °C; O, 15 °C; O, 20 °C; O, 25 °C; O, 28 °C; O, 30 °C; O, 33 °C; O, 35 °C; and O, 40 °C). (f) Im7 at pH 7.0 with 0.4 M Na_2SO_4 at (O, 10 °C; O, 15 °C; O, 20 °C; O, 25 °C; O, 30 °C; O, 35 °C and O, 40 °C). The continuous line shows the best fit of the data to a model describing a two-state transition for Im9 and a three-state transition for SIm9 and Im7. Bottom: Eyring plots of the temperature-dependence of the rate constant for unfolding of (g) Im9; (h) SIm9 and (i) Im7. The curves were calculated from the fits to the kinetic unfolding data shown in (d) (e) and (f), extrapolated to 0 M urea. The continuous curves show the best fit of the data to equation (7).

previous observations.[36] Although it is difficult to interpret the origins of the differences in β_{TS} and β_{TS}^{Cp} for the immunity proteins in precise molecular detail, the results demonstrate that the activation enthalpy, entropy and heat capacity of the TS ensembles of these proteins are very similar and insensitive to whether folding is two or three-state. Finally, and importantly for the discussions below, $S°$ is identical for Im9 and SIm9, the small differences in activation free energy of these proteins being entirely enthalpic in origin.

Φ-Value analysis of SIm9

Im9 and SIm9 differ by only two amino acid substitutions (Figure 1), yet fold with kinetic mechanisms of different complexity to a native state that has similar global stability ($\Delta\Delta G_{un} = 0.9$ kJ mol^{-1} at pH 7.0, 10 °C). These proteins, therefore, provide an ideal system to investigate the role of increasing the stability of non-native interactions on the properties of the folding landscape. To achieve this, the folding and unfolding kinetics of SIm9 and 16 variants of this protein were measured at pH 7.0 and 10 °C, and the properties of the populated intermediate and the subsequent TS ensemble were determined using Φ-value analysis. The data were then used as restraints in MD simulations to provide structural insights into the TS ensembles. In all, 12 of the mutations involved truncation of buried hydrophobic residues (Figure 4(a); Table 2). Three additional variants involved the substitution of a solvent-exposed alanine to glycine to provide a probe for helix formation at these sites. The rate constants for folding/unfolding (Supplementary Data Figure 1) and the initial and final fluorescence signals for all variants (data not shown) were measured using stopped-flow fluorescence and the resulting data

Table 1. Activation parameters for Im9, SIm9 and Im7 derived from the fit of the unfolding rate constants measured from 10–40 °C at a concentration of urea of 3.0–8.0 M

	Im9			SIm9			Im7		
	ΔC_P^{X-Y} (kJ mol^{-1} K^{-1})	ΔH^{X-Y} (kJ mol^{-1})	ΔS^{X-Y} (kJ mol^{-1} K^{-1})	ΔC_P^{X-Y} (kJ mol^{-1} K^{-1})	ΔH^{X-Y} (kJ mol^{-1})	ΔS^{X-Y} (kJ mol^{-1} K^{-1})	ΔC_P^{X-Y} (kJ mol^{-1} K^{-1})	ΔH^{X-Y} (kJ mol^{-1})	ΔS^{X-Y} (kJ mol^{-1} K^{-1})
Unfolding (N→TS)	3.2±0.4	98.8±6.9	0.07±0.02	2.3±0.6	91.9±8.3	0.07±0.03	1.7±0.07	71.1±1.1	0.01±0.004
Refolding (U→TS)	−1.9±0.2	26.6±2.3	−0.09±0.01	−2.9±0.6	36.6±9.2	−0.03±0.03	−2.1±0.2	29.7±4.1	−0.05±0.01
Equilibrium (N↔U)	5.1±0.5	72.2±7.3	0.16±0.02	5.2±0.1	55.3±3.9	0.1±0.01	3.8±0.1	41.4±3.9	0.06±0.01

The refolding activation parameters for Im7 and SIm9 were computed from the equilibrium and unfolding data. For Im9, accurate refolding kinetics could be determined over the entire temperature range using stopped-flow fluorescence. The resulting directly determined thermodynamic parameters are shown and agreed well with the equilibrium stability determined directly. All measurements were carried out in 50 mM sodium phosphate (pH 7.0), 2 mM DTT, 1 mM EDTA, for Im9 and SIm9; and 50 mM sodium phosphate (pH 7.0), and 0.4 M Na_2SO_4 for Im7. Parameters are cited at 0 M urea and 10 °C.

were fitted globally to a three-state on-pathway model in order to determine Φ_I and Φ_{TS} (see Materials and Methods; and Table 2). In all cases reported, $\Delta\Delta G_{un}$ exceeded 2.5 kJ mol^{-1} (Table 2) providing confidence in the Φ-values determined in this study.[45]

The Φ_I and Φ_{TS} values for SIm9 are shown in Figure 4(a). The substitutions L52A and I53V, which involve residues from helix III that dock into the hydrophobic core of the native protein, result in Φ_I and Φ_{TS} values close to zero. By contrast, with the single exception of the mutation F40L, all other residues resulted in partial Φ_I and Φ_{TS} values, suggesting that helices I, II and IV are formed and docked in both the intermediate and TS ensemble of SIm9 (Table 2). Consistent with this high F_I and F_{TS} values were obtained for Ala to Gly substitutions at solvent exposed sites (A13G, A76G, A77G), indicating the formation of significant helical structure. Like Im7 at pH 7.0 and Im9 at low pH,[28,33] therefore, SIm9 also folds *via* a three-helical intermediate and rate-limiting TS in which helices I, II and IV are formed, but helix III is not. Consistent with this, mutation of F40 to Leu, a residue that lies in helix II and contacts the side-chains of residues from helix III in the core of the native protein, also results in Φ_I and Φ_{TS} values that are close to zero (Figure 4(a); Table 2), indicating that these contacts are not well formed in the intermediate and TS ensembles. Finally, and importantly, the substitutions L18A (helix I), L33A (helix II) and L36A (helix II), result in Φ_I values that exceed Φ_{TS} ($\Delta\Phi \geq 0.3$) (Table 2), consistent with the view that stabilisation of the intermediate involves the formation of non-native interactions involving these residues.[28,46] The structural folding mechanisms of SIm9, Im9 and Im7, therefore, are highly conserved, as expected for such close homologues.[22,47]

The Φ_{TS} values of Im9 and SIm9 are compared in Figure 4(b). The data are striking, revealing that, with the single exception of A13G, every residue mutated that probes interactions in or between helices I, II and

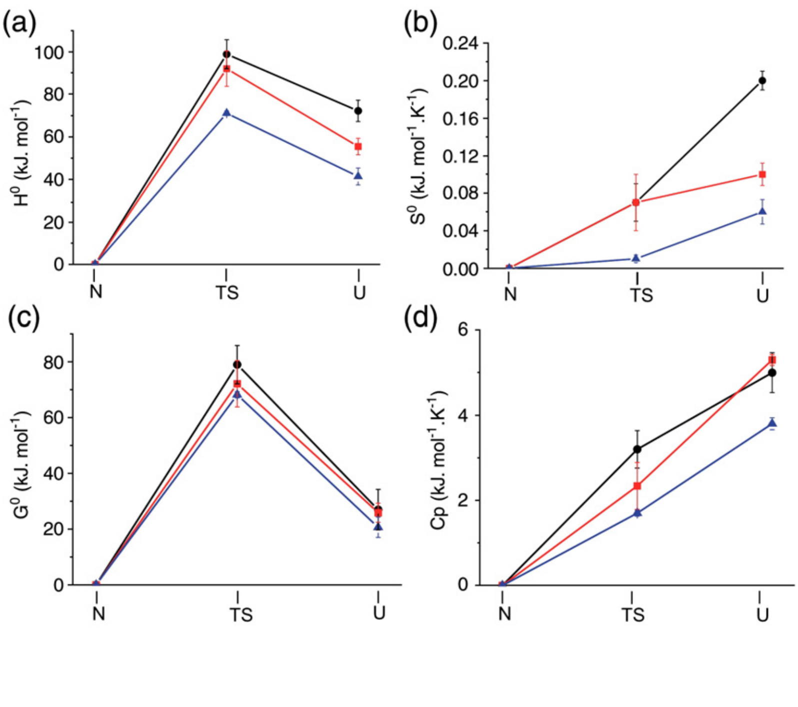

Figure 3. Activation thermodynamic parameters for Im9 (black circles), SIm9 (red squares) and Im7 (blue triangles) in water (pH 7.0) at 10 °C. (a) Enthalpy, $H°$, profile; (b) entropy, $S°$, profile; (c) free energy, $G°$, profile; and (d) heat capacity, C_p, profile. The plots were obtained from the unfolding kinetics and equilibrium stability data shown in Figure 2.

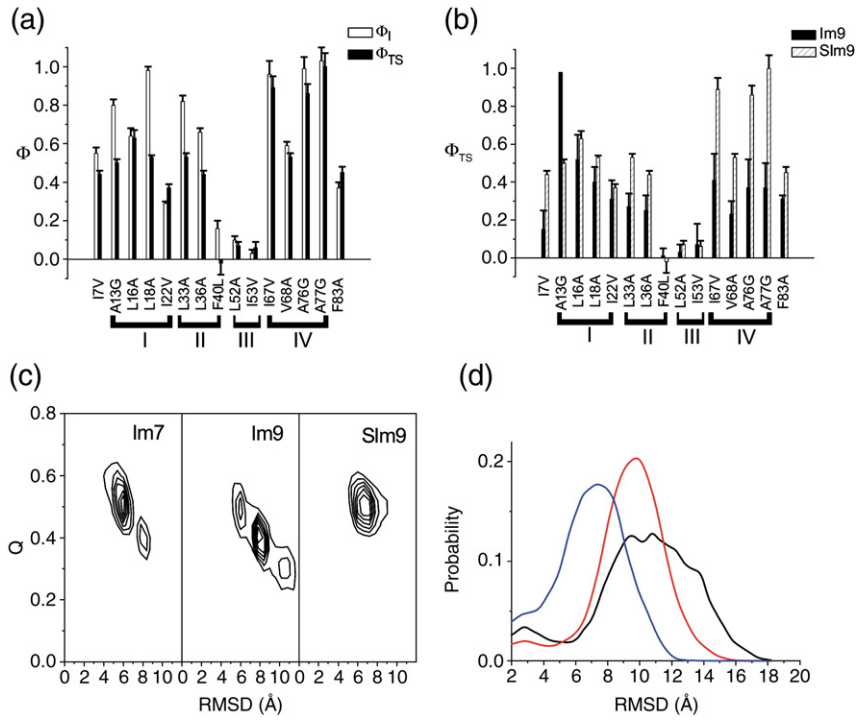

Figure 4. Φ-Value analysis of SIm9. (a) Φ_I-values (open bars) and Φ_{TS}-values (filled bars) for SIm9 at pH 7.0, 10 °C. (b) Comparison of Φ_{TS}-values for SIm9 (hatched bars) with those for Im9 (filled bars) at pH 7.0, 10 °C. Values for Im9 are taken from Friel *et al.*[33] In all cases, the Φ-values are calculated with respect to the unfolded state (see Materials and Methods). (c) Contour plot showing the positions of the sets of structures calculated for Im7, Im9 and SIm9 in Q/RMSD space. RMSD is taken from the energy-minimized native structure, and Q is the fraction of native contacts formed, where a contact is defined as being present if two C^{α} atoms are separated by less than 12 Å and five or more residues apart in the sequence. (d) Distribution of pairwise RMSDs for ensembles of structures generated using sets of between 14 and 18 Φ_{TS}-values for Im7 (blue), Im9 (black) and SIm9 (red) as restraints in MD simulations.

IV in SIm9 results in values for Φ_{TS} that are greater than those for Im9, even though these proteins differ in sequence by only two conservative amino acid substitutions. The differences in Φ_{TS}-values are particularly striking for residues in helix IV, for which $\Delta\Phi_{TS} \geq 0.3$, whilst $\Delta\Phi_{TS} \geq 0.1$ for residues in helices I and II (Table 2). The data suggest that despite having similar *SASA* (β_T and β_T^{Cp}), similar structural properties and similar, if not identical, thermodynamic parameters (activation free energy, enthalpy and entropy), the TS ensembles of Im9 and SIm9 differ significantly in their response to mutation.

In order to investigate the structural properties of the TS ensembles and the effect of a variation in the Φ-values on these, restrained MD simulations were used to determine putative TS ensembles for Im9, Im7 and SIm9 using sets of between 14 and 18 Φ_{TS} values as restraints (see Materials and Methods).[34] Sets of structures for which the experimental Φ_{TS}-values are satisfied were computed and compared. Although this method does not necessarily produce structures that are exactly on the top of the free energy barrier between the native and denatured state,[5] it does provide an indication of the range and type of structures consistent with the Φ-values interpreted as the degree of native structure formation in the TS ensemble. This allows insights into the heterogeneity of structures compatible with a set of partial Φ-values of different magnitude, and how changes in Φ-values are manifested in alterations of the width of these ensembles.[6] The results of this analysis are shown in Figure 4(c) and (d). The data shown in Figure 4(c) demonstrate that the TS ensembles of Im7 and SIm9 have about 40–60% native contacts and an RMSD from the native state between 4 Å and 9 Å, while the TS ensemble of Im9 includes structures that are more heterogeneous and have fewer native contacts. The data presented in Figure 4(d) show that the difference in Φ_{TS}-values between the three proteins (Table 2) [28,33] results in a significant change in the pairwise RMSD distributions. Thus, the TS ensemble of Im7 is relatively well-defined (average pairwise RMSD ~7 Å), while that of Im9 is more heterogeneous (average pairwise RMSD ~11 Å) (Figure 4(d), in agreement with previous results.[34] By contrast, the TS ensemble of SIm9 is intermediate in width between these two proteins (average pairwise RMSD ~9 Å). The average secondary structure content for the TS of the three proteins is also very similar, showing good conservation of helices I, II and IV, and complete absence of helix III (data not shown). The *SASA* was calculated for the three computed ensembles, using the approximate relation:[48]

$$\beta_{TS} = \frac{SASA_U - SASA_{TS}}{SASA_U - SASA_N}$$

β_{TS} was calculated to be 1.02±0.03 for Im9, 1.00±0.03 for SIm9 and 0.95±0.03 for Im7, values very close to the experimental results (β_{TS} = ~0.9 for Im9, SIm9 and Im7). The increased Φ-values for SIm9 are therefore consistent with a significant reduction in the conformational heterogeneity of the rate-limiting TS ensemble for this protein.

The effect of titrating the stability of non-native interactions on the properties of TS

To further analyse the effect of non-native interactions on the properties of the Im9 folding landscape, we made use of the fact that the stability of the intermediate in Im9 folding can be altered systematically by decreasing the pH,[31] adding

Table 2. Parameters determined from the best fit of the folding/unfolding kinetics of SIm9 and its variants to a three-state model at pH 7.0, 10 °C

Variant of SIm9 (position in native structure)	K_{ui} (M_{ui})	k_{in} (m_{in})	k_{ni} (m_{ni})	ΔG_{ui}[a] (kJ mol^{-1})	ΔG_{un}[b] (kJ mol^{-1})	Φ_I	Φ_{TS}
SIm9	105.3±4.4 (4.55±0.07)	279.4±6.8 (0.41±0.04)	0.27±0.03 (0.04±0.04)	−10.9±0.1	−27.3±0.3	–	–
I7V (buried, N terminus)	22.0±0.9 (4.55±0.07)	389.6±10.1 (0.41±0.04)	1.32±0.06 (0.20±0.02)	−7.3±0.1	−20.7±0.2	0.6±0.03	0.4±0.02
A13G (solvent-exposed, helix I)	22.1±0.8 (4.55±0.07)	278.5±7.0 (0.41±0.04)	0.16±0.01 (0.16± 0.02)	−7.3±0.1	−24.9±0.2	0.8±0.03	0.5±0.02
L16A (buried, helix I)	17.9±1.0 (4.55±0.07)	300.3±10.4 (0.41±0.04)	0.22±0.02 (0.34±0.03)	−6.8±0.1	−23.8±0.3	0.6±0.04	0.6±0.04
L18A (buried, helix I)	0.12±0.002 (4.55±0.07)	6050.7±148.2 (0.41±0.04)	6.75±0.15 (0.07±0.01)	5.0±0.1	−10.9±0.1	0.9±0.02	0.5±0.01
I22V (buried, helix I)	35.4±1.4 (4.55±0.07)	206.9±5.1 (0.41±0.04)	2.88±0.12 (0.03±0.02)	−8.4±0.1	−18.5±0.2	0.3±0.01	0.4±0.02
L33A (buried, helix II)	4.1±0.3 (4.55±0.07)	885.6±49.2 (0.41±0.04)	1.74±0.06 (0.15±0.01)	−3.3±0.2	−17.9±0.2	0.8±0.03	0.5±0.02
V34A (buried, helix II)	149.5±9.3 (4.55±0.07)	367.9±9.1 (0.41±0.04)	0.39±0.06 (−0.05±0.04)	−11.8±0.2	−27.9±0.4	[c]	[c]
L36A (buried, helix II)	7.2±0.4 (4.55±0.07)	695.2±28.0 (0.41±0.04)	2.71±0.10 (0.23±0.01)	−4.6±0.1	−17.7±0.2	0.7±0.02	0.4±0.02
F40L (buried, helix II)	46.90±10.1 (4.55±0.07)	710.49±125.58 (0.41±0.04)	49.49±2.74 (0.05±0.01)	−9.05±0.51	−15.32±0.67	0.2±0.04	−0.02±0.06
L52A (buried, helix III)	58.57±5.51 (4.55±0.07)	336.11±20.09 (0.41±0.04)	66.02±2.31 (0.17±0.01)	−9.58±0.22	−13.41±0.27	0.1±0.02	0.1±0.02
I53V (buried, helix III)	99.03±3.91 (4.55±0.07)	256.19±6.26 (0.41±0.04)	2.29±0.15 (0.07±0.02)	−10.81±0.09	−21.91±0.19	0.03±0.02	0.06±0.03
I67V (buried, helix IV)	17.46±0.69 (4.55±0.07)	319.88±8.52 (0.41±0.04)	0.33±0.02 (0.27±0.02)	−6.73±0.09	−22.91±0.18	1.0±0.07	0.9±0.06
V68A (buried, helix IV)	6.52±0.39 (4.55±0.07)	384.11±16.32 (0.41±0.04)	2.48±0.08 (0.35±0.01)	−4.41±0.14	−16.28±0.19	0.6±0.02	0.5±0.02
A76G (exposed, helix IV)	13.33±0.56 (4.55±0.07)	363.99±10.45 (0.41±0.04)	0.36±0.02 (0.25±0.02)	−6.09±0.10	−22.37±0.18	1.0±0.06	0.9±0.05
A77G (exposed, helix IV)	17.80±0.70 (4.55±0.07)	294.72±7.81 (0.41±0.04)	0.27±0.01 (0.39±0.02)	−6.77±0.09	−23.23±0.14	1.0±0.07	1.0±0.07
F83A (buried, C terminus)	3.11±1.24 (4.55±0.07)	133.17±33.77 (0.41±0.04)	46.41±0.84 (0.43±0.01)	−2.73±0.91	−5.21±1.09	0.4±0.03	0.5±0.03

Values for Φ_I and Φ_{TS} are calculated at 0 M urea using SIm9 as the pseudo wild-type. The parameters listed were determined by fitting the data for these variants globally to a three-state on-pathway model involving the rapid formation of an intermediate state (see Materials and Methods). Errors were propagated mathematically. Units of k are s^{-1}; units of M/m are kJ mol^{-1} M^{-1}.

[a] $\Delta G_{ui} = -RT \ln(K_{ui})$.

[b] $\Delta G_{un} = -RT \ln(K_{ui}(k_{in}/k_{ni}))$.

[c] Φ_{TS} could not be determined accurately, as the $\Delta\Delta G_{un}$ upon mutation was too small (< 2.5 kJ mol^{-1}).

Na_2SO_4,[32,49] and, as shown here and previously, by mutation.[32] The folding and unfolding kinetics of Im9 and SIm9 at pH 7.0, pH 6.0 and pH 6.0 + 0.4 M Na_2SO_4 are shown in Figure 5(a) and (b). The data show the switch in kinetic folding mechanism for Im9 from two-state to three-state as the pH is lowered, consistent with previous results,[31] and further stabilisation of the intermediate at pH 6.0 by the addition of Na_2SO_4 (Figure 5(a)). For SIm9, folding is three-state under all three conditions, with the populated intermediate being further stabilised as the pH is decreased and Na_2SO_4 added (Figure 5(b); Table 3). Similarly, the TS for both Im9 and SIm9 are stabilised at pH 6.0 with 0.4 M Na_2SO_4, since both Im9 and SIm9 fold and unfold more rapidly under these conditions compared with pH 7.0. Alteration of the sequence and changes in the folding conditions thus result in an increase in stability of the intermediate ($-\Delta G_{ui}$ is increased from <1 kJ mol^{-1} (the population of an intermediate cannot be detected) to −16 kJ mol^{-1}). In parallel, the TS ensemble is also stabilised under these conditions ($\Delta\Delta G_{u\text{-}ts} = -11$ kJ mol^{-1}, assuming a pre-exponential factor of 1 μs),[50] whilst ΔG_{un} is relatively unperturbed ($\Delta G_{un} = +4$ kJ mol^{-1}) (Figure 5(c) and (d); Table 3). The data thus reinforce the view that the intermediate in Im9 folding is stabilised, at least in part, by non-native interactions,[28,32] and show that these interactions persist, although their contribution to stability is decreased in the rate-limiting TS ensemble. By altering the solution conditions in this manner, therefore, the effect of increasing the contribution of non-native interactions on the properties of the TS ensemble (Φ_{TS}) can be analysed independently of a change in the stability of the native state. Importantly, no significant change in the kinetic or equilibrium m-values was observed under any of the conditions employed, ruling out changes in the *SASA* of the unfolded, intermediate, TS, or native states as the conditions are altered (Table 3). Any effect of varying the conditions on Φ_{TS}, therefore, cannot be attributed to Hammond behaviour, as has been found for other proteins.[51,52]

To determine the effect of ΔG_{ui} on Φ_{TS}, seven individual mutations were made in Im9 and SIm9,

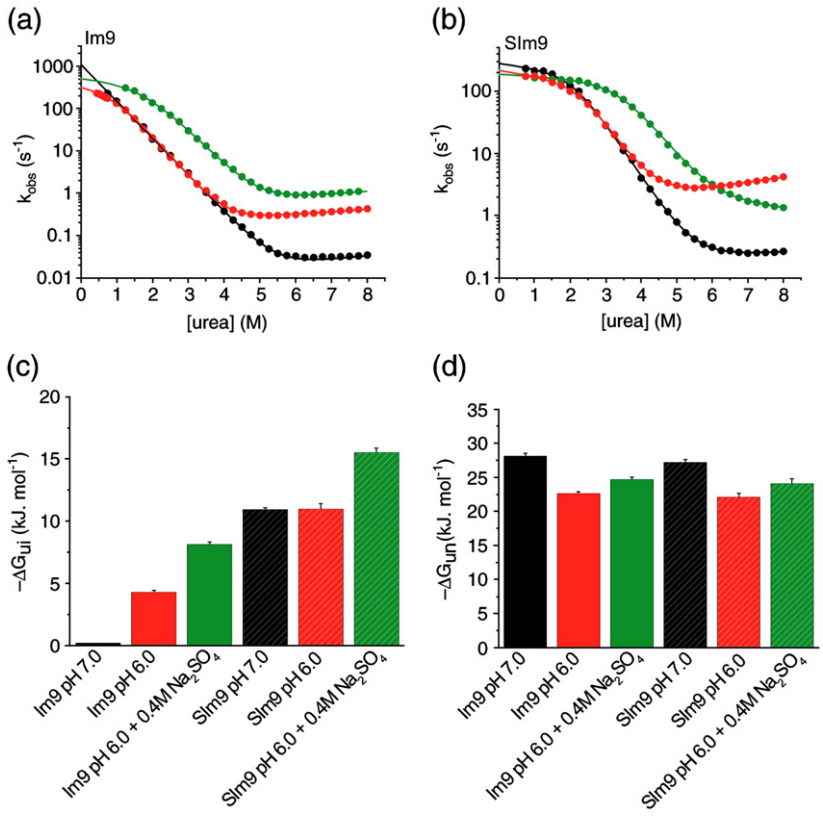

Figure 5. Titrating the stability of non-native interactions in (a) Im9 and (b) SIm9 folding at pH 7.0 (black); pH 6.0 (red) and pH 6.0 + 0.4 M Na_2SO_4 (green). The data were acquired at 10 °C. The data are fitted to a model describing a three-state transition; except for Im9 at pH 7.0 which is fitted to a two-state model (see Materials and Methods, and Table 3). Bar charts showing ΔG_{ui} and ΔG_{un} under each condition employed are shown in (c) and (d), respectively.

targeting residues in helix I (L16A and L18A), helix II (L33A and L36A), helix III (I53V) or helix IV (I67V and A76G). These residues were chosen because their substitution results in a large change in ΔG_{un} ($\Delta\Delta G_{un} \geq 5$ kJ mol^{-1}), allowing accurate determination of Φ_{TS},[45] and a large difference in the Φ_{TS}-value for the residue in Im9 and SIm9 at pH 7.0 (Figure 4(b)), thereby providing a sensitive probe. The Φ_I and Φ_{TS}-values were calculated for each variant under each solution condition employed (using Im9 at the appropriate condition as the wild-type for the Im9 variants, and SIm9 at the appropriate condition as the wild-type for the SIm9 variants) (see Materials and Methods). The results of this analysis were striking: for all six residues probed in helices I, II, and IV, under every condition employed in which folding is three-state, Φ_{TS}-values that exceed those obtained when folding is two-state were obtained (Figure 6(a), (b) and (d); and data not shown). Plotting Φ_{TS} *versus* ΔG_{ui} revealed a weak, but significant correlation between these parameters (Supplementary data Figure 2). Only for I53V in helix III was this trend not observed, Φ_{TS} for this residue remaining close to zero under all conditions employed, as expected for a TS in which this helix is not formed (Figure 6(c)).

Table 3. Parameters determined from the best fit of the folding/unfolding kinetics of Im9 at pH 7.0 to a two-state model and Im9 and SIm9 under other conditions to a three-state model at 10 °C

Protein (condition)	K_{ui} (M_{ui})	k_{in} (m_{in})	k_{ni} (m_{ni})	ΔG_{ui}[a] (kJ mol^{-1})	ΔG_{un}[b] (kJ mol^{-1})	β_I	β_{TS}
Im9 (pH 7.0)	–	1106.5±20.4 (4.71±0.15)	0.007±0.001 (0.46±0.03)	–	−28.2±0.3	–	0.91±0.01
Im9 (pH 6.0)	6.3±0.27 (4.78±0.0.29)	365.9±8.8 (0.55±0.04)	0.15±0.01 (0.30±0.02)	−4.3±0.1	−22.7±0.2	0.85±0.1	0.95±0.05
Im9 (pH 6.0 + 0.4M Na_2SO_4)	24.0±1.6 (4.20±0.40)	527.7±17.9 (0.44±0.03)	0.35±0.05 (0.34±0.05)	−7.5±0.2	−24.7±0.4	0.85±0.1	0.95±0.10
SIm9 (pH 7.0)	105.3±4.4 (4.55±0.07)	279.4±6.8 (0.41±0.04)	0.27±0.03 (0.04±0.04)	−10.9±0.1	−27.3±0.3	0.91±0.01	0.99±0.01
SIm9 (pH 6.0)	107.49±20.16 (4.74±0.13)	213.70±7.27 (0.47±0.07)	1.91±0.18 (0.37)±0.04)	−11.0±0.4	−22.1±0.5	0.85±0.02	0.93±0.02
SIm9 (pH 6.0 + 0.4M Na_2SO_4)	742.67±81.63 (4.26±0.16)	182.5±4.35 (0.21±0.03)	4.75±1.11 (−0.37±0.08)	−15.6± 0.3	−24.1±0.6	1.0±0.02	1.1±0.02

The parameters listed were determined by fitting the data for these variants globally together with the fluorescence signals (see Materials and Methods). Errors were propagated mathematically. Units of k are s^{-1}; units of M/m are kJ mol^{-1} M^{-1}.

[a] $\Delta G_{ui} = -RT \ln(K_{ui})$.

[b] $\Delta G_{un} = -RT \ln(K_{ui}(k_{in}/k_{ni}))$.

To understand better the effect of increasing the stability of non-native interactions on the properties of the TS ensemble, we again used restrained MD to calculate structures compatible with the sets of Φ_{TS}-values obtained for Im9 and SIm9 under each condition examined. The simulations were performed using the seven Φ_{TS}-values obtained under each condition as restraints. Calculating the results using fewer restraints yields conformational distributions that are slightly broader than those shown in Figure 4(d), which utilised between 14 and 18 restraints. Qualitatively, however, the results of the two simulations are similar, in that the TS ensemble for Im9 is significantly more heterogeneous than that for SIm9, presumably since the seven sites chosen for mutation include key residues that make large numbers of long-range contacts in the native state and TS ensemble.[8] Most importantly, since each simulation in Figure 6(e) and (f) was performed using the same number of restraints, the properties of the different ensembles can be compared. The results show a marked decrease in the pairwise RMSD distributions for the TS ensemble under all conditions in which folding is three-state (Figure 6(e)). By contrast, the *SASA* distributions are very similar to each other (Figure 6(f)), in accord with the experimental β_{TS} values, which are approximately the same for all six conditions (Table 3).

In addition to providing information about the relationship between the stability of non-native interactions and the apparent heterogeneity of the TS ensemble, these experiments allow the Φ_I-value to be compared with Φ_{TS} over the range of conditions employed. The data show that increasing $-\Delta G_{ui}$ results in a small, but significant increase in Φ_I for residues in helices I, II and IV (Figure 6(a), (b), (d)). Most strikingly, whilst the Φ_I and Φ_{TS}-values for residues in helices I and IV are similar under each individual condition, the Φ_I-values for residues in helix II consistently exceed Φ_{TS}, and this difference increases with increasing ΔG_{ui}. For example, while $\Phi_I - \Phi_{TS} = 0.27$ for L33A in Im9 at pH 6.0 (Figure 6(b), condition 2), $\Phi_I - \Phi_{TS} = 0.6$ for the same mutation in SIm9 at pH 6.0 + 0.4 M Na_2SO_4 (Figure 6(b), condition 6). A similar trend is observed for the variant L36A, also in helix II (data not shown). These results highlight the important role of residues in helix II in forming stabilising non-native interactions in the intermediate ensemble and demonstrate that

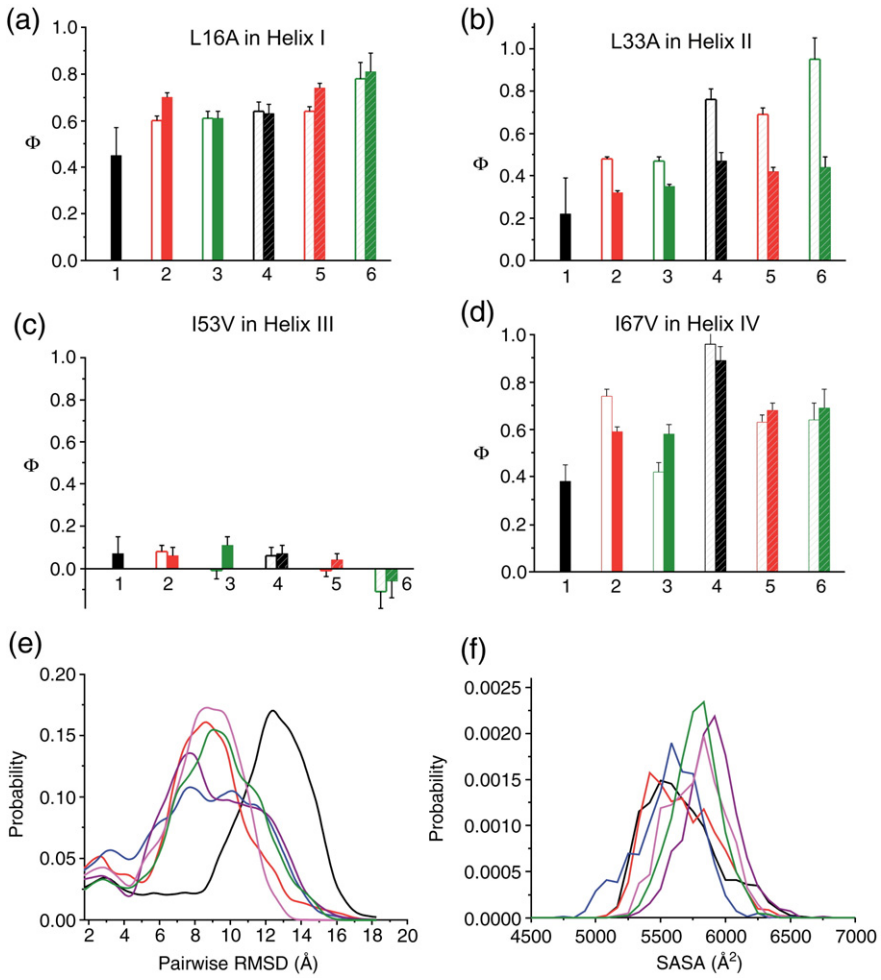

Figure 6. Comparison of Φ_I (open bars) and Φ_{TS} (filled bars) for (a) L16A (helix I); (b) L33A (helix II); (c) I53V (helix III) and (d) I67V (helix IV) under different solution conditions. Im9: pH 7.0, pH 6.0, pH 6.0 + 0.4 M Na_2SO_4, 1–3, respectively. SIm9: pH 7.0, pH 6.0, pH 6.0 + 0.4 M Na_2SO_4, 4–6, respectively. The data are ordered in terms of increasing $-\Delta G_{ui}$, and colour-coded as in Figure 5(c) and (d). At pH 7.0 the folding of Im9 is two-state. (e) Distribution of pairwise RMSDs of structures generated using MD simulations restrained with sets of Φ_{TS}-values (L16A, L18A, L33A, L36A, I53V, I67V and A76G) for Im9 and SIm9 under different conditions. Im9 pH 7.0 (black), Im9 pH 6.0 (red), Im9 pH 6.0 + 0.4 M Na_2SO_4 (blue), SIm9 pH 7.0 (purple), SIm9 pH 6.0 (pink), and SIm9 pH 6.0 + 0.4 M Na_2SO_4 (green) (f) Distribution of solvent-accessible surface area (*SASA*) for the ensembles shown in (e).

these interactions are partially, but not completely, disrupted in the TS ensemble, consistent with previous results obtained for Im7.[28]

Discussion

Previous studies have revealed that the folding intermediate of Im7 is stabilised, at least in part, by non-native interactions.[28,46] Here, we show that an intermediate with similar structural properties becomes populated during Im9 folding by lowering the pH, adding Na_2SO_4, or by the introduction of only two mutations that increase the hydrophobicity of residues in the native helices II and IV. Detailed Φ-value analysis combined with restrained MD simulations have shown that like the TS ensemble formed during Im7 and Im9 folding,[32–34] the TS ensemble of SIm9 also contains native-like helices I, II and IV, but lacks helix III. As well as possessing similar structural properties and relative compactness, these TS ensembles have similar thermodynamic properties ($\Delta H°$, $\Delta G°$, $\Delta Cp°$ and $\Delta S°$). Thus, despite folding with mechanisms of different kinetic complexity at neutral pH, the structural folding mechanism of the immunity proteins is highly conserved, as expected for proteins within the same fold family.[22] Folding of these proteins thus involves a ubiquitous partially folded intermediate and a common three-helix TS ensemble. These results highlight the importance of topology in governing the mechanism of folding, and illustrate the role that the sequence and solution conditions has in tailoring the ruggedness of the energy landscape, such that by manipulating these parameters, partially folded species that previously were undetected can be revealed and characterised.

The results presented also highlight the importance of non-native interactions in immunity protein folding. Thus, by increasing systematically the contribution of non-native interactions to the stability of the intermediate formed transiently during Im9 and SIm9 folding, a parallel increase in the magnitude of Φ_I-values is observed for residues in the three helices structured in this state. Comparison of the magnitudes of the Φ_I and Φ_{TS}-values, however, revealed a significant decrease in Φ_{TS}-values for residues in helix II, with $\Phi_I - \Phi_{TS} \geq 0.2$ (Figure 6(b)). This suggests that interactions that are formed in the intermediate must be disrupted, at least in part, on traversing the rate-limiting TS ensemble. Interestingly, at concentrations of urea close to the midpoint where the intermediate is no longer visibly populated and folding is two-state, SIm9 folds more rapidly at pH 6.0 with 0.4 M Na_2SO_4 compared with pH 7.0, whilst the protein also unfolds more rapidly under the former conditions (green and black lines in Figure 5(b)). Im9 shows similar behaviour (green and black lines in Figure 5(a)). Together, the results suggest that remnants of the non-native interactions formed in the intermediate must persist in the rate-limiting TS, and that these interactions stabilise the TS ensemble even under conditions where the intermediate is destabilised such that it no longer accumulates during folding.

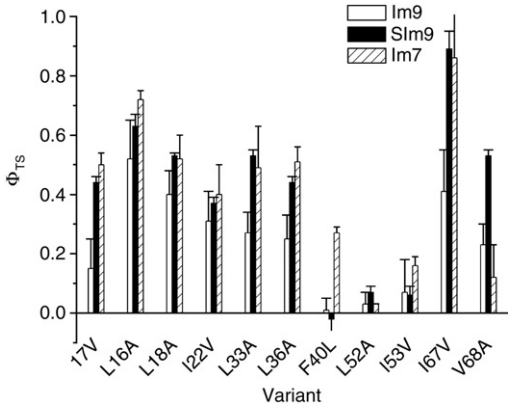

Figure 7. Bar chart highlighting the difference in Φ_{TS}-values observed for the equivalent residues mutated in Im9 (open bar), SIm9 (filled bar) and Im7 (hatched bar). Data were acquired at pH 7.0, 10 °C, in the absence (Im9 and SIm9) or in the presence (Im7) of 0.4 M Na_2SO_4. Values are shown for buried hydrophobic deletion mutations. The values cited for Im7 are from Capaldi *et al.*,[28] and those for Im9 are from Friel *et al.*[33] Note that Im7 is one residue longer than Im9 due to the insertion of a single residue in the loop linking helices I and II. The residues shown here are numbered according to the Im9 sequence.

Comparison of the Φ_{TS}-values for the common set of 11 residues that have been mutated in Im9, SIm9 and Im7 is shown in Figure 7. The results highlight the dramatic increase in Φ_{TS}-values that occurs in these proteins when folding is three-state i.e. under conditions in which the contribution of stabilising non-native interactions is high. Thus, despite differing by only two amino acid residues, the Φ_{TS}-values for SIm9 are significantly and consistently larger than those of Im9 when the two proteins are studied under identical conditions. In addition, the Φ_{TS}-values of SIm9 are very similar in magnitude to those of Im7 (even though Im7 and SIm9 have different native state stability and were studied under different conditions (with/without 0.4 M Na_2SO_4)).[28] By contrast, studies of other homologous proteins (α-spectrin and src homology SH3 domains, or acylphosphatase and the activation domain of human procarboxypeptidase), which are 36% and 13% identical in sequence, respectively, and each fold with a conserved two-state mechanism, giving rise to Φ_{TS}-values that differ by less than 0.2.[53–55] Similarly, the Φ_{TS}-values of α-spectrin SH3 domains studied under conditions in which the stability of the native protein is titrated from −2.9 to −16.3 kJ mol^{-1} result in Φ_{TS}-values that differ by ≤0.2.[54] The differences in magnitude of the Φ_{TS}-values for Im9 and SIm9, which are 98% identical in sequence and have native state stabilities that differ by only 0.9 kJ mol^{-1}, are therefore dramatic. Since this difference cannot be attributed to movement of the TS on the reaction coordinate (Hammond or anti-Hammond behaviour), we suggest that the increase in Φ_{TS}-values between Im9 and SIm9

folding correlates with the observed switch in kinetic folding mechanism from two-state to three-state, and is caused by increasing the stability of non-native interactions formed early during folding.

Fractional Φ_{TS}-values are commonly interpreted as indicative of partly formed interactions within a TS ensemble formed on a single dominant pathway.[19] Alternatively, such Φ-values may reflect folding on multiple pathways. In such a situation, the experimentally determined Φ-values represent the average over all TS structures encountered.[18] Whilst the presence of multiple pathways for small single-domain proteins can be difficult to discern experimentally, features such as upwards curvature in the unfolding branch of the chevron plot,[25] different structural location of the folding nucleus in individual members of a protein family,[23,24] or movements of the TS ensemble in response to mutation (anti-Hammond behaviour),[20,21] are suggestive of multiple folding routes. Despite analysis of more than 80 different site-directed mutations and of folding under a range of Na_2SO_4 concentrations, none of these features has been observed during immunity protein folding.[28,32,33,49] However, by specifically altering the stability of non-native *versus* native interactions by changes in solution conditions and/or specific redesign of the sequence of helices II and IV, a sensitivity of the Φ_{TS}-values to mutation and folding conditions is revealed. Rather than folding on a single predominant route, the data are more consistent with a model for folding involving multiple micropathways to the native state.[21,22] Such a scenario is consistent with the low partial Φ_{TS}-values observed for these proteins despite their very high β_{TS} values and the differences in the magnitude of their β_{TS} and β_{TS}^{Cp} values. In this situation, altering the strength of stabilising non-native interactions results in a change in flux through different micropathways such that certain routes become favoured over others. Assuming that the TS ensembles on the different pathways have similar structural properties, the result is a TS ensemble characterised by higher Φ-values (i.e. an increased homogeneity); without gross alteration of its structural properties (i.e. it retains a common three-helix structure).

Taken together, the data presented show that tipping the balance between the contribution of stabilising native and non-native interactions has a critical role in shaping the energy landscape of immunity protein folding. Over-stabilisation of non-native interactions results in a rugged landscape involving kinetic traps and a restriction in the number of micropathways available for folding, resulting in a relatively homogeneous TS ensemble. By contrast, when the strength of native and non-native contacts is balanced judiciously, as occurs in the sequence of Im9 at pH 7.0, intermediates are no longer visibly populated, folding occurs *via* multiple micropathways and a conformationally diverse TS ensemble ensues. The ability to enhance specifically the stability of non-native interactions in Im9 folding, together with careful and quantitative analysis of the folding mechanisms of this family of proteins using Φ-values has thus revealed new insights into the folding of these small single-domain proteins that would not have been possible by analysis of a single family member under a single set of solution conditions alone.

Materials and Methods

Mutagenesis and protein purification

Mutagenesis was carried out using the QuikChange site-directed mutagenesis kit (Stratagene) with the Im9 gene in pTrc99a as the template.[31] All proteins were expressed as His-tagged versions and purified to homogeneity as described.[31]

Data collection and analysis

Temperature-dependence of the equilibrium stability of Im9, SIm9 and Im7

The equilibrium stability of Im9, SIm9 and Im7 were measured at different temperatures and concentrations of urea using far-UV CD (Jasco J-715 spectropolarimeter; Jasco, UK), as described.[56] Samples were made from filtered stock solutions of the appropriate buffers (50 mM sodium phosphate (pH 7.0), 2 mM DTT, 1 mM EDTA for Im9 and SIm9, or 50 mM sodium phosphate (pH 7.0), 0.4 M Na_2SO_4, 1 mM EDTA, for Im7) and 0–9 M urea in 0.5 M increments. Temperature ramps were performed for each sample with a 2 deg.C increment from 2 °C to 88 °C using a protein concentration of 20 μM. The CD signal at 222 nm at each temperature was recorded and the resulting data set fit to a two-dimensional function:

$$\Theta_{222}(T,C) = \left[\frac{\Theta_N(T,C) + K_{eq}(T,C)\Theta_D(T,C)}{1 + K_{eq}(T,C)}\right] \quad (1)$$

where K_{eq} is the equilibrium constant, Θ_{222} (T,C) is the CD signal in the two dimensions of temperature and denaturant concentration, and Θ_N (T,C) and Θ_D (T,C) are the native and denatured baseplanes as a function of temperature (T) and concentration of urea (C), respectively.[56] Using:

$$\Delta G_{un}(T,C) = \Delta G_{un}(T) - M_{un}(C) \quad (2)$$

where $\Delta G_{un}(T)$ is the temperature-dependence of the free energy in the absence of urea, and M_{un} is the denaturant-dependence of ΔG_{un} at a specific temperature, and taking into consideration the fact that M_{un} may also have a temperature-dependence, $M_{un}(T)$, and an associated temperature-dependent enthalpy $h(T)$ and entropy $s(T)$ per mole denaturant, which will themselves have a dependence c on heat capacity, $\Delta G_{un}(T,C)$ was determined from:

$$\Delta G_{un}(T,C) = h(C - C_{hs}) + Ts(C - C_{hs}) + \left\{\left[\Delta C_p - c(C - C_{hs})\right].\left[(T - T_{hs}) - T\ln\left(\frac{T}{T_{hs}}\right)\right]\right\} \quad (3)$$

and used to calculate M_{un}, ΔH_{un} and ΔS_{un} using equations (4) to (6):

$$M_{un} = h + c(T - T_{hs}) - T\left[s + c\ln\left(\frac{T}{T_{hs}}\right)\right] \quad (4)$$

$$\Delta H_{un}(T,C) = \Delta G_{un}(T,C) + T\Delta S_{un}(T,C) = -h(C - C_{hs}) + [\Delta C_p - c(C - C_{hs})](T - T_{hs}) \quad (5)$$

$$\Delta S_{un}(T,C) = -\frac{d\Delta G(T,C)}{dT} = -s(C - C_{hs}) + \left[\Delta C_p - c(C - C_{hs})\right]\ln\left(\frac{T}{T_{hs}}\right)\Bigg] \quad (6)$$

The data were fit globally to equation (3) using Igor Pro 5.0 (Wavemetrics) as described.[56]

Analysis of the temperature-dependence of the folding and unfolding rate constants

All kinetic experiments were performed using an Applied Photophysics SX18.MV stopped-flow fluorimeter as described.[33] Buffers used were as described above. All buffers containing sodium sulphate were prepared gravimetrically.[32] The kinetics of folding and unfolding were measured between 0.75 M and 8.00 M final concentration of urea in 0.25 M increments, from 5–40 °C using a final protein concentration of 5 μM. The resulting transients were fit to a single-exponential (Im7) or a double-exponential (Im9 and SIm9) function, the slower of which has been attributed to proline isomerisation.[31,33] The rate constants determined were used to fit the resulting data to a two-state (Scheme 1) or three-state (Scheme 2) mechanism, as appropriate.

$$U \underset{k_{nu}}{\overset{k_{un}}{\rightleftharpoons}} N$$

Scheme 1.

$$U \underset{k_{iu}}{\overset{k_{ui}}{\rightleftharpoons}} I \underset{k_{ni}}{\overset{k_{in}}{\rightleftharpoons}} N$$

Scheme 2.

where U, I and N are the unfolded, intermediate and native states, respectively.

The data for each protein measured at all temperatures were fit globally using IgorPro 5.0 (Wavemetrics). When fitting data to Scheme 2, the formation of the intermediate was assumed to occur rapidly as a pre-equilibrium step. k_{ui} was fixed at 20,000 s^{-1} for Im9 and SIm9, and at 3000 s^{-1} for Im7,[29,32] and the stability of the intermediate was determined by allowing k_{iu} to vary. The *m*-values m_{ui}, m_{iu}, m_{in} were allowed to vary, but were assumed to be identical for each individual protein at all temperatures, and m_{ni} was allowed to vary freely. In all cases where the folding and unfolding branches of the chevron plot were well defined, the global fits were within error of those obtained by fitting each curve individually.

The temperature-dependence of the rate constants for unfolding from the native state (k_{nu} or k_{ni}) for each protein determined from global analysis of the chevron plots were fitted to:

$$\Delta G^{n-ts}(T) = \Delta H^{n-ts}(T_m) + \Delta C_p^{n-ts}(T - T_m) - T[\Delta S^{n-ts}(T_m) + \Delta C_p^{n-ts}\ln(T/T_m)] \quad (7)$$

where ΔG^{n-ts}, ΔH^{n-ts}, ΔS^{n-ts} and ΔC_p^{n-ts} are the free energy, enthalpy, entropy, and heat capacity differences between the TS and the native state at the reference temperature, T_m.

Φ-Value analysis

Φ_I and Φ_{TS}-values for SIm9 at pH 7.0, 10 °C were determined by analysis of the folding and unfolding kinetics as a function of urea concentration for the wild-type protein and 16 site-directed mutants. For each variant analysed, folding was three-state, as demonstrated by the kinetic roll-over in the logarithm of the folding rate constant at low concentrations of denaturant and by the presence of a significant increase in the fluorescence signal in the dead-time of the stopped-flow experiment (3 ms). Previous studies have demonstrated that all the observed folding and unfolding data for Im9 from pH 8.0–5.0, in the presence of Na_2SO_4, are best described by a three-state transition with an on-pathway intermediate.[31] The chevron plots for SIm9 and each variant, together with the initial and final fluorescence signals were fitted globally to Scheme 2, assuming that mutation altered the equilibrium stability of the intermediate and/or the rate constants of folding/unfolding, while M_{ui} and m_{in} were assumed to be invariant upon mutation. In this manner, accurate fits to the kinetic data for all variants could be obtained, even in situations where the intermediate was populated only marginally. In all cases, where the kinetic roll-over is well defined the global fits were within error of those obtained by fitting each curve individually. Analysis of the kinetics of folding/unfolding showed no dependence on protein concentration (from 0.5–50 μM).

Analysis of the chevron plots for folding/unfolding of Im9 and SIm9 at different pH values and/or different concentrations of Na_2SO_4 were fitted individually to the appropriate kinetic model (two-state for Im9 at pH 7.0 and three-state for all other conditions). For SIm9, the initial and final fluorescence signals were also used to constrain the fits.

Φ_{TS}-values were then calculated according to

$$\Phi_{TS} = \frac{\Delta\Delta G_{u-ts}^{wt-mut}}{\Delta\Delta G_{un}^{wt-mut}} \quad (8)$$

where $\Delta\Delta G_{u-ts}^{wt-mut}$ is the change in free energy of the TS referenced to the unfolded state and $\Delta\Delta G_{un}^{wt-mut}$ is the difference in free energy between the native and unfolded state of the wild-type (wt) and mutant (mut) proteins, respectively.

Similarly, Φ_I was calculated according to

$$\Phi_I = \frac{\Delta\Delta G_{u-i}^{wt-mut}}{\Delta\Delta G_{u-n}^{wt-mut}} \quad (9)$$

Errors on all parameters were propagated mathematically.

Restrained molecular dynamics simulations

The Φ-value restrained MD method for calculating putative TS structures has been described in detail.[6] In brief, the trajectory is forced towards structures that satisfy $\Phi_i^{calc} = \Phi_i^{exp}$ for the available set of Φ_{TS}-values for each protein, where Φ_i^{exp} is the measured Φ_{TS}-value of residue *i*, and Φ_i^{calc} is the fraction of native side-chain contacts formed by residue *i*. This is achieved by adding a harmonic term to the Hamiltonian, which forces ρ, the average square deviation between Φ_i^{calc} and Φ_i^{exp}, towards zero.

Simulations were carried out using all-atom models of the proteins,[57] and the EEF1 implicit solvent model[58]

within the CHARMM program.[59] Langevin dynamics, with a timestep of 2 fs and a friction coefficient of 0.1 ps^{-1} were used. In the cases of Im9 (PDB code 1IMQ) and Im7 (PDB code 1AYI), the minimized PDB structures were used as starting and reference structures. For SIm9, the PDB of Im9 was modified to contain the mutations V37L and V71I and the resulting structure was minimized. Replica exchange MD (REMD)[60] was used together with the restrained MD method with the aim of improving the sampling of conformational space. Ten replicas, in the temperature range 300–420 K were used, and swaps attempted every 500 steps (1 ps). The starting structures for REMD were generated by heating the minimized PDB structure to the replica temperature, reducing ρ towards zero using biased MD, and then maintaining ρ close to zero using a harmonic bias. REMD was run for 150 ns, and the 300 K trajectory used for analysis. ρ was maintained at ~0.01 during the REMD run.

The initial set of simulations (Figure 4(c) and (d)) were done with sets of 18 Φ_{TS}-values as restraints for Im7 and Im9,[28,30] and 14 Φ_{TS}-values for SIm9. The set of simulations depicted in Figure 6(e) and (f) used the sets of seven Φ_{TS}-values reported here as restraints.

SASA was calculated using the program NACCESS†. $SASA_N$ and $SASA_U$ were estimated from unrestrained simulations of 1 ns at 300 K and 575 K, respectively.

Acknowledgements

We thank Keith Ainley for technical assistance and Alison Ashcroft for mass spectrometry. We thank Graham Spence, Eva Sanchez-Cobos and Stuart Knowling for helpful discussions. We thank Chris Gell for all his help with the use of the Igor software for global fitting. This work was supported by the BBSRC, EPSRC, INTAS, the University of Leeds and The Wellcome Trust.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2007.05.010

References

1. Fersht, A. R. (1995). Mapping the structures of transition states and intermediates in folding: delineation of pathways at high resolution. *Phil. Trans. Roy. Soc. ser. B*, **348**, 11–15.
2. Mittermaier, A. & Kay, L. E. (2006). New tools provide new insights in NMR studies of protein dynamics. *Science*, **312**, 224–228.
3. Jahn, T. R. & Radford, S. E. (2005). The Yin and Yang of protein folding. *FEBS. J.* **272**, 5962–5970.
4. Fersht, A. R. & Daggett, V. (2002). Protein folding and unfolding at atomic resolution. *Cell*, **108**, 573–582.
5. Allen, L.R. & Paci, E. (2007). Transition state for protein folding using molecular dynamics and experimental restraints. *J. Phys. Condens. Matter* In the press.
6. Paci, E., Vendruscolo, M., Dobson, C. M. & Karplus, M. (2002). Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol.* **324**, 151–163.
7. Paci, E., Clarke, J., Steward, A., Vendruscolo, M. & Karplus, M. (2003). Self-consistent determination of the transition state for protein folding: application to a fibronectin type III domain. *Proc. Natl Acad. Sci. USA*, **100**, 394–399.
8. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature*, **409**, 641–645.
9. Di Nardo, A. A., Korzhnev, D. M., Stogios, P. J., Zarrine-Afsar, A., Kay, L. E. & Davidson, A. R. (2004). Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation. *Proc. Natl Acad. Sci. USA*, **101**, 7954–7959.
10. Feng, H., Zhou, Z. & Bai, Y. (2005). A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc. Natl Acad. Sci. USA*, **102**, 5026–5031.
11. Mason, J. M., Cliff, M. J., Sessions, R. B. & Clarke, A. R. (2005). Low energy pathways and non-native interactions: the influence of artificial disulfide bridges on the mechanism of folding. *J. Biol. Chem.* **280**, 40494–40499.
12. Nishimura, C., Dyson, H. J. & Wright, P. E. (2006). Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J. Mol. Biol.* **355**, 139–156.
13. Neuweiler, H., Doose, S. & Sauer, M. (2005). A microscopic view of miniprotein folding: enhanced folding efficiency through formation of an intermediate. *Proc. Natl Acad. Sci. USA*, **102**, 16650–16655.
14. Palmer, A. G., 3rd, Kroenke, C. D. & Loria, J. P. (2001). Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339**, 204–238.
15. Vugmeyster, L., Kroenke, C. D., Picart, F., Palmer, A. G. & Raleigh, D. P. (2000). 15 r-1 rho measurements allow the determination of ultrafast protein folding rates. *J. Am. Chem. Soc.* **122**, 5387–5388.
16. McLeish, T. C. (2005). Protein folding in high-dimensional spaces: hypergutters and the role of nonnative interactions. *Biophys. J.* **88**, 172–183.
17. Fersht, A. R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl Acad. Sci. USA*, **92**, 10869–10873.
18. Fersht, A. R., Itzhaki, L. S., ElMasry, N. F., 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.
19. Fersht, A. R. (1995). Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.* **5**, 79–84.
20. 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-condensation mechanism for protein folding. *J. Mol. Biol.* **254**, 260–288.
21. Lindberg, M. O. & Oliveberg, M. (2007). Malleability of protein folding pathways: a simple reason for complex behaviour. *Curr. Opin. Struct. Biol.* **17**, 21–29.
22. Zarrine-Afsar, A., Larson, S. M. & Davidson, A. R. (2005). The family feud: do proteins with similar

† http://wolf.bms.umist.ac.uk/naccess

structures fold via the same pathway? *Curr. Opin. Struct. Biol.* **15**, 42–49.
23. Lindberg, M. O., Haglund, E., Hubner, I. A., Shakhnovich, E. I. & Oliveberg, M. (2006). Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proc. Natl Acad. Sci. USA*, **103**, 4083–4088.
24. Olofsson, M., Hansson, S., Hedberg, L., Logan, D. T. & Oliveberg, M. (2007). Folding of S6 structures with divergent amino acid composition: pathway flexibility within partly overlapping foldons. *J. Mol. Biol.* **365**, 237–248.
25. Wright, C. F., Lindorff-Larsen, K., Randles, L. G. & Clarke, J. (2003). Parallel protein-unfolding pathways revealed and mapped. *Nature Struct. Biol.* **10**, 658–662.
26. 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 structural elements for unfolding/folding of barnase. *Biochemistry*, **34**, 6805–6814.
27. Chavez, L. L., Gosavi, S., Jennings, P. A. & Onuchic, J. N. (2006). Multiple routes lead to the native state in the energy landscape of the beta-trefoil family. *Proc. Natl Acad. Sci. USA*, **103**, 10254–10258.
28. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Struct. Biol.* **9**, 209–216.
29. Capaldi, A. P., Shastry, M. C. R., Kleanthous, C., Roder, H. & Radford, S. E. (2001). Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nature Struct. Biol.* **8**, 68–72.
30. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. & Radford, S. E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* **286**, 1597–1608.
31. Gorski, S. A., Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2001). Acidic conditions stabilise imtermediates populated during the folding of Im7 and Im9. *J. Mol. Biol.* **312**, 849–863.
32. Friel, C. T., Beddard, G. S. & Radford, S. E. (2004). Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J. Mol. Biol.* **342**, 261–273.
33. Friel, C. T., Capaldi, A. P. & Radford, S. E. (2003). Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. *J. Mol. Biol.* **326**, 293–305.
34. Paci, E., Friel, C. T., Lindorff-Larsen, K., Radford, S. E., Karplus, M. & Vendruscolo, M. (2004). Comparison of the transition state ensembles for folding of Im7 and Im9 determined using all-atom molecular dynamics simulations with phi value restraints. *Proteins: Struct. Funct. Genet.* **54**, 513–525.
35. Main, E. R., Fulton, K. F. & Jackson, S. E. (1999). Folding pathway of FKBP12 and characterisation of the transition state. *J. Mol. Biol.* **291**, 429–444.
36. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
37. Schindler, T. & Schmid, F. X. (1996). Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry*, **35**, 16833–16842.
38. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428–10435.
39. Tan, Y. J., Oliveberg, M. & Fersht, A. R. (1996). Titration properties and thermodynamics of the transition state for folding: comparison of two-state and multi-state folding pathways. *J. Mol. Biol.* **264**, 377–389.
40. Van Nul, N. A., Meijberg, W., Warner, J., Forge, V., Scheek, R. M., Robillard, G. T. & Dobson, C. M. (1998). Slow cooperative folding of a small globular protein HPr. *Biochemistry*, **37**, 622–637.
41. Crespo, M. D., Platt, G. W., Bofill, R. & Searle, M. S. (2004). Context-dependent effects of proline residues on the stability and folding pathway of ubiquitin. *Eur. J. Biochem.* **271**, 4474–4484.
42. Bofill, R., Simpson, E. R., Platt, G. W., Crespo, M. D. & Searle, M. S. (2005). Extending the folding nucleus of ubiquitin with an independently folding beta-hairpin finger: hurdles to rapid folding arising from the stabilisation of local interactions. *J. Mol. Biol.* **349**, 205–221.
43. Scalley, M. L. & Baker, D. (1997). Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. *Proc. Natl Acad. Sci. USA*, **94**, 10636–10640.
44. Scalley, M. L., Yi, Q., Gu, H., McCormack, A., Yates, J. R., III & Baker, D. (1997). Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry*, **36**, 3373–3382.
45. Fersht, A. R. & Sato, S. (2004). Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl Acad. Sci. USA*, **101**, 7976–7981.
46. Gsponer, J., Hopearuoho, H., Whittaker, S. B., Spence, G. R., Moore, G. R., Paci, E. *et al.* (2006). Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl Acad. Sci. USA*, **103**, 99–104.
47. Baker, D. (2000). A Surprising simplicity to protein folding. *Nature*, **405**, 39–42.
48. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein. Sci.* **4**, 2138–2148.
49. Cobos, E. S. & Radford, S. E. (2006). Sulfate-induced effects in the on-pathway intermediate of the bacterial immunity protein Im7. *Biochemistry*, **45**, 2274–2282.
50. Schuler, B., Lipman, E. A. & Eaton, W. A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, **419**, 743–747.
51. Ternstrom, T., Mayor, U., Akke, M. & Oliveberg, M. (1999). From snapshot to movie: Φ analysis of protein folding transition states taken one step further. *Proc. Natl Acad. Sci. USA*, **96**, 14854–14859.
52. Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E. & Fersht, A. R. (1995). Movement of the position of the transition state in protein folding. *Biochemistry*, **34**, 13656–13662.
53. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005–1009.
54. Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010–1016.

55. Villegas, V., Martinez, J. C., Aviles, 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.
56. Dimitriadis, G., Drysdale, A., Myers, J. K., Arora, P., Radford, S. E., Oas, T. G. & Smith, D. A. (2004). Microsecond folding dynamics of the F13W G29A mutant of the B domain of staphylococcal protein A by laser-induced temperature jump. *Proc. Natl Acad. Sci. USA*, **101**, 3809–3814.
57. Neria, E., Fischer, S. & Karplus, M. (1996). Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **105**, 1902–1921.
58. Lazaridis, T. & Karplus, M. (2000). Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* **10**, 139–145.
59. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). Charmm - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
60. Sugita, Y. & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Letters*, **314**, 141–151.
61. Osborne, M. J., Breeze, A. L., Lian, L., Reilly, A., James, R., Kleanthous, C. & Moore, G. R. (1996). Three-dimensional solution structure and ^{13}C nuclear magnetic resonance assignments of the colicin E9 immunity protein Im9. *Biochemistry*, **35**, 9505–9512.

Edited by K. Kuwajima

(Received 12 February 2007; received in revised form 2 May 2007; accepted 6 May 2007)
Available online 10 May 2007