Switching Two-state to Three-state Kinetics in the Helical Protein Im9 via the Optimisation of Stabilising Non-native Interactions by Design

Claire T. Friel1, Godfrey S. Beddard2 and Sheena E. Radford*1

1School of Biochemistry and Microbiology, University of Leeds, Leeds LS2 9JT, UK
2School of Chemistry, University of Leeds, Leeds LS2 9JT, UK

The four-helix protein Im7 folds through an on-pathway intermediate at pH 7.0 and 10°C. By contrast, under these conditions there is no evidence for a populated intermediate in the folding of its more stable homologue, Im9, even in the presence of 0.4 M sodium sulphate. Previous studies using Φ-value analysis have shown that the Im7 intermediate is misfolded, in that three of its four native helices are formed, but are docked in a non-native manner. Using knowledge of the structure of the intermediate of Im7, we have used rational design to stabilise an intermediate formed during the folding of Im9 by the introduction of specific stabilising interactions at positions known to stabilise the Im7 folding intermediate through non-native interactions. We show that the redesigned Im9 sequence folds with three-state kinetics at pH 7.0 and have used Φ-value analysis to demonstrate that this species resembles the misfolded intermediate populated during Im7 folding. The redesigned Im9 sequence folds 20-fold faster than the wild-type protein under conditions in which folding is two-state. The data show that intermediate formation is an important feature of folding, even for small proteins such as Im9 for which these partially folded states do not become significantly populated. In addition, they show that the introduction of stabilising interactions can lead to rapid refolding, even when the contacts introduced are non-native.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: immunity protein; folding; intermediate; rational design; non-native

Introduction

Many small (<100 residue), single domain proteins fold in a two-state manner, without populating detectable intermediate states. Larger proteins, by contrast, usually fold via populated intermediates that increase the ruggedness of the folding energy landscape and may slow folding by creating kinetic traps. A number of small single domain proteins, however, have now been shown to form intermediates during folding. For some of these proteins, the rapid formation or instability of partially folded species has required changes in solvent conditions, rapid techniques or re-analysis of protein unfolding kinetics to detect such states. These data suggest the possibility that unstable (high energy) intermediates may be generically important in the folding of two-state proteins and raise the question as to the structural nature of these species compared with intermediates that are highly populated during folding.

The E. coli binding immunity proteins (Im7 and Im9) provide an ideal system for investigating the role of intermediates in the folding of small, single domain proteins. These four-helix proteins are, respectively, 87 and 86 residues in length, and have 60% sequence identity, yet can fold to the same native structure by different kinetic mechanisms. At pH 7.0 and 10°C, Im9 folds with two-state kinetics. By contrast, under the same conditions, the less stable homologue, Im7, folds with three-state kinetics, involving the population of a compact intermediate. Analysis of the complete folding kinetics of Im7, measured using both continuous and stopped-flow methods, has shown
that this intermediate is productive for folding, i.e. is on-pathway to the native state, whilst a model in which the intermediate is off-pathway cannot describe the data.12 Using Φ-value analysis we have shown that the intermediate formed during Im7 folding contains three of the four native α-helices (helices I, II and IV) and is stabilised by both native and non-native interactions involving residues in the hydrophobic core.13 The kinetic complexity of the folding of Im7 and Im9 (i.e. three-state versus two-state kinetics) is retained at pH 7.0 in the presence of 0.4 M Na2SO4,5,6 demonstrating that the population of an intermediate during the folding of Im9 is not induced by these stabilising conditions. Moreover, since Im9 is more stable than Im7, population of an intermediate in these proteins cannot be rationalised simply on the basis of native state stability.5 Interestingly, at increasingly acidic pH, the native states of Im7 and Im9 (hexahistidine-tagged versions of these proteins) are destabilised, whilst a folding intermediate becomes progressively stabilised in both proteins.6 This suggests that Im7 and Im9 may fold via a common structural mechanism involving formation of a similar three-helical species. Whilst the intermediate formed during Im7 folding has been characterised in detail using Φ-value analysis,13 the intermediate populated during Im9 folding at acidic pH has not been structurally characterised. These species differ, however, in their fluorescence properties; the intermediate formed during Im7 folding is hyperfluorescent, whilst that in Im9 has the same fluorescence intensity as the denatured state.

Using Φ-value analysis we have recently shown that the rate-limiting transition states in Im7 and Im9 folding are structurally related, in that they contain the native helices I, II and IV docked around a specific hydrophobic core, but lack helix III.13,14 These data reinforce the view that Im7 and Im9 may fold with similar structural mechanisms and suggest that the differences in their folding kinetics at pH 7.0 may result from differences in the stability of a common intermediate formed during the folding of both proteins. Subtle sequence differences between the two proteins, therefore, may remove key stabilising interactions from the intermediate formed during the folding of Im9, such that this species is not populated during folding at neutral pH. To test this hypothesis, we have used protein engineering to stabilise an Im9 folding intermediate by rational design, by optimising stabilising non-native interactions based on knowledge of the structure of the on-pathway, misfolded intermediate populated during the folding of Im7.13 The data show that subtle sequence changes in specific locations switch the folding of Im9 from two-state to three-state, highlighting the importance of both partially folded states and non-native interactions in tailoring the folding energy landscape of these small helical proteins.

Results

Mutations predicted to switch the folding kinetics of Im9

Previous studies using Φ-value analysis have shown that the intermediate populated during Im7 folding is a three-helix bundle, containing the native helices I, II and IV, but lacking helix III.13 Most importantly, based on the observation that ten residues (five of which are in helix II) have higher Φ-values in the intermediate than in the rate-limiting transition state (ΦI−ΦTS=0.1–0.9), we have shown that the intermediate is misfolded, in that helices I, II and IV pack non-natively so as to sequester hydrophobic surface area which is buried in the native state by the presence of helix III.13 The single tryptophan (Trp75), which lies in helix IV (Figure 1), is predicted to be buried in a non-native hydrophobic environment in the intermediate, resulting in its characteristic hyperfluorescence; the intermediate is more fluorescent than both the unfolded and native states.12,13 The intermediate is thus stabilised by both native and non-native interactions involving specific hydrophobic residues.

In an attempt to switch the folding of Im9 to three-state kinetics at pH 7.0 by rational redesign, eight residues were considered, each of which is known to be important in stabilising the misfolded intermediate of Im7 via non-native interactions (Figure 1). Five of these residues are identical in Im7 and Im9, whilst three differ between the two proteins (residues 19, 37 and 71 in Im9). In each case, the residue in Im9 is less hydrophobic (Val) than the equivalent residue (Leu/Ile) in Im7 (Figure 1). The decrease in hydrophobicity of these residues, at positions known to be important in stabilising non-native helix interfaces in the Im7 intermediate, may explain why a similar structure, if formed during the folding of Im9, would be less stable than the intermediate populated in the folding of Im7. Each of these residues in Im9 was therefore replaced individually with the equivalent residue from Im7 in the mutations V19L, V37L and V71I (Figures 1 and 2). In addition, three mutations

![Image](303x153 to 529x223)

Figure 1. Alignment of the sequences of Im9 and Im7. Note that in Im7 there is an additional residue in the loop linking helices I and II. The position of the four helices in native Im9 is shown by boxes.40 Residues in Im9 mutated as part of this study are marked (*). Residues in Im7 for which side-chain truncations destabilise the intermediate by at least 4 kJ/mol and for which ΦI−ΦTS are marked (#). Residues are numbered according to the Im9 sequence. Colours correspond to those used in Figure 2."
coordinates from the solution structure for Im9 (1IMQ).40 

Im9* mutants described above were initially exam-

ined at pH 6.0, 10°C in 0.4 M Na2SO4. Under these con-

ditions the wild-type protein folds with three-

state kinetics,6 permitting accurate quantitative 
determination of the effect of mutations on the 

stability of the intermediate. In all cases the rate 

constants for folding/unfolding were collected as 

well as the initial and final fluorescence signals

from the refolding transients (see Materials and

Methods). The data were fitted to a model describ-

ing a three-state transition using three different

procedures (see Materials and Methods). Para-

meters determined from each fitting method were 

compared (Tables 1 and 2) and resulted in very 
similar calculated values of ΔG_{ui} and ΔG_{un}, within 

1–2 kJ/mol (see Materials and Methods). Only the 

parameters obtained from method 1 are quoted in

the text.

Increasing the hydrophobicity of residues

involved in non-native inter-helical interactions

stabilises an intermediate in the folding of Im9

The folding and unfolding kinetics of the Im9' 

variants V37L and V71I are shown in Figure 3(a)–

d. These mutations significantly stabilise the 

intermediate (ΔΔG_{ui} = –6.2 and –4.9 kJ/mol, 

respectively), but have little effect on the stability of

the native state (ΔΔG_{un} = +1.1 and –0.3 kJ/mol, 

respectively) (Table 1). These data suggest, there-

fore, that Val37 and Val71 form transient non-native 

interactions that stabilise an intermediate during the 

folding of Im9'.

Interestingly, increasing the hydrophobicity of

Val19 in helix I of Im9 does not significantly 

stabilise the intermediate (ΔΔG_{ui} for the mutant 

V19L = –0.4 kJ/mol (Table 1)), despite the fact that 

truncating the equivalent residue (Leu19) to Ala in 

Im7 destabilises the intermediate by >6 kJ/mol.13 

This mutation, however, destabilises native Im9' by 

7.6 kJ/mol, presumably as a result of over-packing 

of the core of the native protein. Despite the 

inability of this substitution to stabilise the Im9' 

intermediate, the observation that the intermediate 

and native states respond very differently to a 

change in the side-chain at this position suggests 

that non-native structure could be formed in the 

vicinity of this mutation in the intermediate. 

Alternatively, the core of the intermediate may be 

less well ordered than in the native state and 

therefore less stabilised by over-packing.

The stabilised intermediate is not a randomly 

collapsed species

To test if the observed effects are specific to 

substitutions at the sites chosen a number of 

controls were performed. Firstly, Val34 was 

replaced with Leu. This residue lies on the 

solvent-exposed face of helix II (Figure 2). Its 

substitution by Leu increases hydrophobicity at a 

position predicted, by analogy with the intermedi-

ate formed during Im7* folding,13 to be remote from 

the helix interfaces in the Im9' intermediate. This 

substitution also increases the helical propensity of 

helix II to a similar extent as V37L (from 5% to 

~11%).13) Substituting Val34 with Leu, however, 

stabilises the Im9' folding intermediate by only

were included as controls in which the hydropho-

bicity of residues at positions known not to be 

involved in stabilising the Im7' folding intermedi-

ate was increased. Two such mutants (E31L and 

V34L) are on the solvent-exposed face of helix II in 

native Im9 and are also predicted to be solvent-

exposed in the intermediate formed during Im9' 

folding. The third (V68I) is buried in the core of

native Im9 but, by analogy with mutation of the 

equivalent residue in Im7* (V69A, 

native Im9 but, by analogy with mutation of the 

intermediate in the folding of Im9'.

Differences in loop flexibility were also considered as a possible reason for the different folding kinetics of Im7 and Im9. Increasing the flexibility of the loop linking helices I and II (loop I–II) by the substitution A28G stabilises the Im7' intermediate, but not the native state (ΔΔG_{ui} = –1.9 kJ/mol, 

ΔΔG_{un} = +0.3 kJ/mol),13 suggesting that the formation 
of non-native helix-interfaces in the Im7' 

folding intermediate creates strain in this short turn. 

This strain may be exacerbated in Im9 as the loop 

between helices I and II is one residue shorter than 

that in Im7 (Figure 1). To test the possibility that 

increased strain in loop I–II in Im9 destabilises its 

putative folding intermediate, Ala25 was mutated 

to Gly. In addition, loop I–II was lengthened by 

insertion of a Gly between residues 27 and 28 (in 

the mutant 27G28) (Figure 2).

The folding and unfolding kinetics of all of the 

Im9' mutants described above were initially exam-

ined at pH 6.0, 10°C in 0.4 M Na2SO4. Under these con-

ditions the wild-type protein folds with three-

state kinetics,6 permitting accurate quantitative 
determination of the effect of mutations on the

Figure 2. Ribbon diagram of the structure of native Im9 
ilustrating the positions of the mutations made in this 

study and the single tryptophan (W74). The Figure was 

prepared using Molscript41 and Raster3D42 using the 

coordinates from the solution structure for Im9 (1IMQ).40
Table 1. The stability of the intermediate ($\Delta G_{ui}$) and native ($\Delta G_{un}$) states of Im9* and its variants calculated from the parameters determined from the best fit of the refolding/unfolding data to a model describing a three-state folding transition

<table>
<thead>
<tr>
<th>Method</th>
<th>$\Delta G_{ui}$</th>
<th>$\Delta G_{un}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>pHi 6.0+0.4 M Na2SO4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Im9*</td>
<td>-6.6</td>
<td>-24.7</td>
</tr>
<tr>
<td>V19L († hydrophobicity)</td>
<td>-7.0</td>
<td>-17.1</td>
</tr>
<tr>
<td>V37L († hydrophobicity)</td>
<td>-12.8</td>
<td>-12.1</td>
</tr>
<tr>
<td>V71H († hydrophobicity)</td>
<td>-11.5</td>
<td>-11.6</td>
</tr>
<tr>
<td>E41V (fluorescence)</td>
<td>-9.7</td>
<td>-28.4</td>
</tr>
<tr>
<td>A25G († loop flexibility)</td>
<td>-8.7</td>
<td>-24.1</td>
</tr>
<tr>
<td>27G28 († loop flexibility)</td>
<td>-8.5</td>
<td>-18.7</td>
</tr>
<tr>
<td>V37L/E41V/V71I</td>
<td>-16.8</td>
<td>-26.9</td>
</tr>
<tr>
<td>E31L (control)</td>
<td>-7.2</td>
<td>-21.9</td>
</tr>
<tr>
<td>V34L (control)</td>
<td>-8.7</td>
<td>-26.1</td>
</tr>
<tr>
<td>V68I (control)</td>
<td>-6.6</td>
<td>-21.3</td>
</tr>
</tbody>
</table>

pH 7.0 (0 M Na2SO4)

<table>
<thead>
<tr>
<th>Method</th>
<th>$\Delta G_{ui}$</th>
<th>$\Delta G_{un}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Im9*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>V37L/E41V/V71I</td>
<td>-10.5</td>
<td>-27.1</td>
</tr>
</tbody>
</table>

The data were fitted using three different procedures (see Materials and Methods) and the parameters determined by each method are listed. Comparison of the values calculated from parameters determined by each of these methods shows $\Delta G_{ui}$ and $\Delta G_{un}$ to be within 1–2 kJ/mol (see Materials and Methods). nd, not determined.

$\Delta G_{ui} = -RT\ln(k_{ui})$.

$\Delta G_{un} = -RT\ln(k_{un}/k_{ui})$. The units of $\Delta G_{ui}$ and $\Delta G_{un}$ are kJ mol$^{-1}$.

* Im9* data collected at pH 7.0 were fitted to a model describing a two-state folding transition.

---

**Figure 3.** Folding and unfolding kinetics of the Im9* mutants V37L, V71I and E41V at 10°C at pH 6.0, 0.4 M Na2SO4. The upper panels show the observed folding and unfolding rate constants as a function of the concentration of urea. The lower panels show the initial (squares) and final (diamonds) fluorescence signals for the refolding transitions. Continuous black lines show the best fit of the rate constants and fluorescence signals to a model describing a three-state transition (method 1). The best fits to the data for wild-type Im9* obtained under the same conditions are also shown (continuous red lines).
Table 2. Folding and unfolding parameters determined for wild-type Im9\(^*\) and its variants

<table>
<thead>
<tr>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(K_{1}) ((M_{1}))</td>
<td>(k_{m1} (m_{1}))</td>
</tr>
<tr>
<td>pH 6.0 + 0.4 M Na(_2)SO(_4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Im9(^*)</td>
<td>16.7 (4.20)</td>
<td>660 (0.40)</td>
</tr>
<tr>
<td>V19L (↑ hydrophobicity)</td>
<td>20.0 (4.20)</td>
<td>260 (0.50)</td>
</tr>
<tr>
<td>V37L (↑ hydrophobicity)</td>
<td>235.3 (4.20)</td>
<td>155 (0.40)</td>
</tr>
<tr>
<td>V71I (↑ hydrophobicity)</td>
<td>133.3 (4.20)</td>
<td>340 (0.40)</td>
</tr>
<tr>
<td>E41V (fluorescence)</td>
<td>62.5 (4.30)</td>
<td>325 (0.40)</td>
</tr>
<tr>
<td>A25G (↑ loop flexibility)</td>
<td>40.0 (4.20)</td>
<td>390 (0.40)</td>
</tr>
<tr>
<td>27G28 (↑ loop flexibility)</td>
<td>37.7 (4.30)</td>
<td>215 (0.40)</td>
</tr>
<tr>
<td>V37L/E41V/V71I</td>
<td>1250 (4.20)</td>
<td>120 (0.40)</td>
</tr>
<tr>
<td>E31L (control)</td>
<td>21.1 (4.20)</td>
<td>400 (0.40)</td>
</tr>
<tr>
<td>V34L (control)</td>
<td>40.0 (4.20)</td>
<td>500 (0.40)</td>
</tr>
<tr>
<td>V68I (control)</td>
<td>16.7 (4.20)</td>
<td>380 (0.40)</td>
</tr>
<tr>
<td>pH 7.0 (0 M Na(_2)SO(_4))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Im9(^*)</td>
<td>–</td>
<td>1200 (4.80)</td>
</tr>
<tr>
<td>V37L/E41V/V71I</td>
<td>87.0 (4.50)</td>
<td>320 (0.20)</td>
</tr>
</tbody>
</table>

The data were fitted to a model describing a three-state folding transition, except for wild-type Im9\(^*\) at pH 7.0, which was fitted to a two-state model. Rate constants and \(m\)-values have units of s\(^{-1}\) and kJ mol\(^{-1}\) M\(^{-1}\), respectively. nd, not determined.

\(^{a}\) Relative fluorescence signals of the unfolded and intermediate species in the absence of denaturant which best fit the observed data (see Materials and Methods). Fluorescence signals are not used in the fitting procedure using method 2. In all cases, an adequate fit to the fluorescence signals was obtained with method 3 using the signals of U and I determined using method 1 listed above.

\(^{b}\) \(k_{m1} (m_{1})\).

\(^{c}\) \(k_{ni} (m_{ni})\).
2.1 kJ/mol (Table 1), indicating that the increase in helical propensity caused by the substitution of a Val for a Leu does not account for the large increase in stability of the intermediate caused by the substitution V37L (ΔΔC\text{un} = −0.6 kJ/mol). In a second control, Glu31 was substituted by Leu (Figure 2). This mutation at a solvent-exposed site increases hydrophobicity, but does not change the predicted helical propensity of helix II. Importantly, the mutation E31L has little effect on the stability of the Im9\* folding intermediate (ΔΔC\text{un} = −6.2 kJ/mol). In a third control, the substitution of an A25G (ΔΔC\text{un} = 6.2 kJ/mol) increases hydrophobicity, but does not change the predicted helical propensity of helix II. The interplay of entropic effects with changes in hydrophobicity of natively solvent-exposed residues does not increase the stability of the Im9\* folding intermediate. Finally, mutation of the natively buried residue, Val68, to Ile (Figure 2) has no effect on the stability of the Im9\* folding intermediate (Table 1). Taken together, these results demonstrate that a general increase in hydrophobicity does not result in an increased population of the intermediate in Im9\* folding. By contrast, increasing the hydrophobicity of residues predicted to be involved in forming non-native inter-helical contacts results in substantial stabilisation of the intermediate formed during Im9\* folding.

Increasing flexibility in the helix I–helix II loop

The mutants A25G and 27G28 are predicted to stabilise the intermediate in Im9\* folding by reducing strain created in the loop between helices I and II by the non-native docking of these helices, despite the predicted stabilisation of the denatured state by insertion of a glycine (ΔΔC\text{un} = −3.0 kJ/mol) and the entropic cost (0.5–0.8 kJ/mol) of closing the longer loop.17 Interestingly, both of these mutations stabilise the intermediate populated in Im9\* folding by ΔΔC\text{un} = −2 kJ/mol (Table 1), consistent with the hypothesis that rotation of helices I and II is important in stabilising this state. However, whilst A25G has little effect on the stability of the native state (ΔΔC\text{un} = +0.6 kJ/mol), increasing the length of the loop by insertion of a glycine destabilises the native state substantially (ΔΔC\text{un} = +6.0 kJ/mol). Loop I–II has a very different sequence in Im7\* and Im9\* (Figure 1) and residues in this region make numerous contacts with side-chains in helices I and II. The interplay of entropic effects with changes in stability arising from other sequence effects makes interpretation of mutations in this loop difficult. Therefore, although successful in stabilising the intermediate, these mutations were not pursued further.

Mutations predicted to alter the fluorescence properties of the Im9\* folding intermediate

One of the most characteristic features of the Im7\* folding intermediate is that it is hyperfluorescent.5,12 Interestingly, although Im9\* populates an intermediate during folding at pH 6.0 in 0.4 M Na\textsubscript{2}SO\textsubscript{4}, this species has a fluorescence quantum yield similar to that of the denatured state (Figure 3 and Table 2). This suggests that the environment of the conserved tryptophan, which is largely responsible for the fluorescence signal of the immunity proteins, differs significantly in the intermediates of Im7\* and Im9\*. The mutation V42A decreases the fluorescence of the Im7\* intermediate by 30%, but has no effect on the fluorescence of the native state,13 suggesting that this residue, although distant (>12 Å) from Trp75 in the native state of Im7, contributes to the environment of a single tryptophan residue in the misfolded intermediate. In Im9 the residue equivalent to V424 is a Glu (E41) (Figure 1). This residue could play a role in determining the fluorescence of the Im9\* intermediate, either directly by quenching tryptophan fluorescence18 or indirectly by preventing helix II in Im9\* from adopting the position that buries V42 in the Im7\* intermediate, due to the energetic penalty of burying a charged residue. To investigate the role of Glu41 in Im9\* folding, this residue was mutated to Val (Figure 2). The mutation E41V stabilises the Im9\* folding intermediate by 3.1 kJ/mol and causes a significant increase in its fluorescence (Figure 3 and Table 2). By contrast, the control mutation, E31L, which lies in the N-terminal turn of the native helix II (Figure 2), has little effect on the stability of the intermediate (ΔΔC\text{un} = −0.6 kJ/mol) and has no effect on its fluorescence signal (Table 2). Interestingly, the mutation V37L, which is located in the centre of helix II and lies 1.5 turns away from E41 (Figure 2), also causes an increase in fluorescence of the Im9\* intermediate similar to that observed for E41V (Figure 3(b) and Table 2). Fitting the data to a model describing a three-state transition (see Materials and Methods) suggests that the fluorescence signals of U, I and N for E41V in the absence of denaturant are 0.65, 0.84 and 0.20, respectively. The fluorescence signals of U, I and N for wild-type Im7\*, fitted in an identical manner, are 0.72, 1.30 and 0.23, respectively.13 The data indicate, therefore, that residues in the C-terminal half of helix II contribute to the non-native environment of the tryptophan in the Im9\* folding intermediate. Nonetheless, the fluorescence signal of the intermediate populated during the folding of Im9\* E41V is lower than that of the intermediate formed during Im7\* folding.

Three-state folding at neutral pH

In order to stabilise the intermediate of Im9\* further, several mutations were combined in the triple mutant V37L/E41V/V71I. V37L and V71I were included as they cause the greatest increase in stability of the intermediate, but do not change the stability of the native state (Table 1). E41V was also included since this mutation stabilises the intermediate and causes an increase in its fluorescence signal. By contrast with V37L and V71I, however, this mutation also stabilises the native state of Im9\* (Table 1). The folding/unfolding kinetics of V37L/E41V/V71I at pH 6.0 in 0.4 M Na\textsubscript{2}SO\textsubscript{4} are shown in Figure 4(a) and (b). Under these conditions the
native state of the triple mutant is stabilised by only 2.2 kJ/mol relative to wild-type Im9*, whilst the intermediate is stabilised by 10.2 kJ/mol (Table 1).

At pH 7.0 (in the absence of Na₂SO₄) the triple mutant also folds with clear three-state kinetics, involving the population of a stable (ΔG = 10.5 kJ/mol), hyperfluorescent intermediate (Figure 4(c) and (d)). Detailed analysis shows that the rate constants and amplitudes of folding of the triple mutant are independent of protein concentration at both pH 7.0 in the absence of sodium sulphate (protein concentrations ranged from 0.3–12.5 µM) (Figure 5(a)) and at pH 6.0 in 0.4 M sodium sulphate (protein concentrations 0.1–5.0 µM (data not shown)). The data indicate, therefore, that stabilisation of the intermediate in Im9* folding cannot be attributed to intermolecular association events. Similarly, the refolding kinetics of Im7* are also independent of protein concentration over a wide range (Figure 5 (b) and Ferguson et al.5). The combination of these substitutions, therefore, switches the folding kinetics of Im9* from two-state to three-state at pH 7.0, by the incorporation of residues with increased hydrophobicity at specific positions that selectively stabilise the folding intermediate relative to both the native and unfolded states (Figure 4(c)).

The intermediates of Im7* and Im9* are structurally related

The intermediate populated during the folding of the triple mutant of Im9* resembles the folding intermediate of Im7* in that both species are specifically stabilised/destabilised by similar substitutions at equivalent positions. In addition, substitutions at equivalent locations in both proteins cause a change in the fluorescence signal of their respective intermediates. The intermediates of Im7* and the Im9* triple mutant differ, however, in their degree of compactness relative to the native state (∆G = 0.8 and 0.9, respectively, at pH 7.0 in the absence of Na₂SO₄) and in the extent of their hyperfluorescence (see above). To investigate the structural similarity of the intermediates of Im7* and V37L/E41V/V71I Im9* in more detail, six residues were mutated and the effects of these mutations on the stability of the folding intermediate and rate-limiting transition state were determined using Φ-value analysis (Table 3). Firstly, Ile53 was mutated to Val. This residue lies in helix III of native Im9* and the equivalent residue in Im7* makes few stabilising interactions in the intermediate or rate-limiting transition state (Φ-values = 0.10 and 0.16, respectively)13 (Figure 6). Similarly, the
mutation I53V in Im9 results in a $\Phi_{TS}$-value of 0.07, indicating that the transition state of wild-type Im9 also lacks helix III.\textsuperscript{14} The folding and unfolding kinetics of Im9\textsuperscript{V37L/E41V/V71I} are shown in Figure 4(e) and (f). Only the unfolding rate constants are affected by the mutation ($\Phi_{TS}=0.04$) (Table 3 and Figure 6), demonstrating that the intermediate and rate-limiting transition state ensembles of V37L/E41V/V71 Im9 also lack helix III.

Selected mutations were also made in helices I, II and IV of V37L/E41V/V71 Im9 (Figure 6 and Table 3). The $\Phi$-values for the mutations L16A (helix I), L36A (helix II), I67V and A76G (helix IV) of 0.63, 0.44, 0.59 and 0.61, respectively, suggest that akin to the folding intermediate of Im7*, these helices are also significantly formed and docked in the folding intermediate of V37L/E41V/V71 Im9. Finally, the mutation F40L (helix II) was created. This mutation is particularly informative since in Im7* it reports on the formation of non-native interactions in the folding intermediate. In native Im7 the side-chain of Phe41 is buried, with the $\varepsilon 1$-, $\varepsilon 2$- and the $\zeta$CH interacting almost exclusively with the side-chain of Leu53 from helix III.\textsuperscript{19} These interactions are removed in the mutation F41L. Despite the absence of helix III in the intermediate of Im7, however, the $\Phi$-value for the mutant F41L is 1.2,\textsuperscript{13} demonstrating that Phe41 forms non-native stabilising interactions in the Im7* intermediate. The stabilising effect of these interactions is then reduced in magnitude in the rate-limiting transition state ensemble ($\Phi_{TS}=0.3$) (Figure 6). The mutation F40L in V37L/E41V/V71 Im9* also results in a significant $\Phi$-value of 0.3 (Table 3 and Figure 6), suggesting that the intermediate of V37L/E41V/V71 Im9* is also stabilised in part by non-native interactions. Together with the results presented above, the data suggest that the folding intermediates of V37L/E41V/V71 Im9* and Im7* are structurally related in that both species comprise helices I, II and IV but lack helix III and are stabilised by both native and non-native interactions.

**Table 3.** Parameters determined from the best fit of the folding/unfolding kinetics for V37L/E41V/V71 Im9* and its variants to a three-state model (see Materials and Methods)

<table>
<thead>
<tr>
<th>Protein</th>
<th>$K_{ui}$ (M$\textsubscript{a}$)</th>
<th>$K_{in}$ (m$\textsubscript{a}$)</th>
<th>$K_{ui}$ (m$\textsubscript{a}$)</th>
<th>$\Delta G_{ui}^*$ (kJ/mol)</th>
<th>$\Delta G_{un}^*$ (kJ/mol)</th>
<th>$\Phi_I$</th>
<th>$\Phi_{TS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>V37L/E41V/V71</td>
<td>87.0 (4.5)</td>
<td>320 (0.20)</td>
<td>0.043 (0.30)</td>
<td>-10.5</td>
<td>-31.5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>V37L/E41V/V71+L16A$^c$</td>
<td>10.0 (4.5)</td>
<td>375 (0.20)</td>
<td>0.18 (0.41)</td>
<td>-5.4</td>
<td>-23.5</td>
<td>0.63</td>
<td>0.58</td>
</tr>
<tr>
<td>V37L/E41V/V71+L36A$^c$</td>
<td>16.7 (4.6)</td>
<td>1200 (0.30)</td>
<td>0.68 (0.35)</td>
<td>-6.6</td>
<td>-22.6</td>
<td>0.44</td>
<td>0.27</td>
</tr>
<tr>
<td>V37L/E41V/V71+I67V$^e$</td>
<td>83.3 (4.5)</td>
<td>305 (0.20)</td>
<td>0.35 (0.30)</td>
<td>-10.4</td>
<td>-26.3</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>V37L/E41V/V71+F40L$^d$</td>
<td>20.0 (4.5)</td>
<td>1800 (0.20)</td>
<td>4.8 (0.25)</td>
<td>-7.0</td>
<td>-21.0</td>
<td>0.33</td>
<td>-0.06</td>
</tr>
<tr>
<td>V37L/E41V/V71+I53V$^e$</td>
<td>26.3 (4.5)</td>
<td>320 (0.30)</td>
<td>0.10 (0.40)</td>
<td>-7.7</td>
<td>-26.7</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>V37L/E41V/V71+A76G$^e$</td>
<td>20.0 (4.5)</td>
<td>380 (0.30)</td>
<td>0.13 (0.35)</td>
<td>-7.0</td>
<td>-25.8</td>
<td>0.61</td>
<td>0.54</td>
</tr>
</tbody>
</table>

The parameters listed were determined by fitting the data for these variants using method 1 (see Materials and Methods). Values for $\Phi_I$ and $\Phi_{TS}$ are calculated at 0 M urea using V37L/E41V/V71 Im9* as the pseudo wild-type (see Materials and Methods).

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

$b$ $\Delta G_{un}^* = -RT\ln(K_{ui}/K_{in})$.

c Parameters determined when these data are fitted using methods 2 and 3 give $\Phi$-values that differ by less than 0.10.

d Parameters determined when these data are fitted using methods 2 and 3 give $\Phi$-values that differ by less than 0.15.

e Parameters determined when these data are fitted using methods 2 and 3 give $\Phi$-values that differ by less than 0.05.
Discussion

Misfolded intermediates: a common theme in immunity protein folding

The ability to redesign folding systems by systematic mutation is the ultimate test of our understanding of protein folding. Previous studies have used both experimental and computational methods to achieve this based on knowledge of the native structure and using the principle that folding involves the formation of increasing numbers of stabilising native interactions. Alterations of folding pathways, thus, have been accomplished by changing the intrinsic stability of different β-hairpins in protein L and protein G by optimising native interactions of the main-chain and side-chains. In other cases, a change in folding pathway has resulted from altering contact order by circular permutation. Here we have taken a different approach in which a switch from two- to three-state kinetics is achieved without significantly changing the folding pathway by the introduction of interactions that specifically stabilise a misfolded intermediate. By contrast with other studies, the switch in the folding kinetics of Im9 is not achieved simply by non-specific stabilisation or destabilisation of all compact states. Instead, the switch from two-state to three-state kinetics in Im9 folding is accomplished by rational redesign, based on knowledge of the structural properties of the intermediate of Im7. As a consequence, a folding intermediate that hitherto was too unstable to be detected during the folding of Im9 at pH 7.0 is stabilised substantially relative to both the native and denatured states.

Several of the changes incorporated in the sequence of redesigned Im9 involve residues known to be important in determining the binding specificity of the protein for its cognate colicin toxin. Val37 and Glu41 in Im9 form part of the binding surface for the E9 colicin, whilst residues in the same region of Im7 are not involved in colicin binding. The switch in folding mechanism caused by substitution of these residues fulfils our previous prediction that portions of the sequence selected for colicin specificity influence folding by altering the strength of non-native interactions formed transiently during folding. Functional constraints thus tailor the ruggedness of the folding energy landscape for these proteins.

Tuning the folding landscape by native and non-native interactions

One of the most fascinating features about the folding of Im7 is the observation that the intermediate is stabilised, at least in part, by significant non-native interactions between helices I, II and IV. This results in the population of a hyperfluorescent, stable intermediate that is easily detectable using stopped-flow kinetics. Here we show that a similar intermediate is formed during Im9 folding but, for this protein, residue-specific variations in sequence result in this species being too unstable or too transient to detect at pH 7.0. Stabilisation of all compact states, for example by the addition of sodium sulphate, is not sufficient to populate the intermediate during Im9 folding. By contrast, stabilisation of the intermediate relative to the native and unfolded states is achieved by lowering the pH or, as we show here, by the optimisation of stabilising non-native interactions by site-specific substitutions (Table 1), such that the intermediate becomes highly populated during folding. Further experiments using Φ-analysis will be needed to characterise the folding intermediate of V37L/E41V/V71I Im9 in more detail. Nonetheless, our results show that
misfolding is a common feature of the folding of Im7 and Im9 and suggest that non-native interactions may contribute significantly to the folding of other proteins, including those that fold without significantly populating partially folded states. In accord with this view, a number of small, single domain proteins have recently been shown to form transient non-native interactions during folding, both experimentally and in simulations.28–30

Transient intermediates have been observed in simulations of two-state folding13 and have been suggested to enhance the rate of folding, as long as these species do not become so stable that they form kinetic traps.32 Interestingly, we show here that V37L/E41V/V71I Im9* folds more rapidly than the can even enhance the rate of folding.

Trapped, the generation of non-native interactions when such species do not become energetically mediates are an important feature of folding and, energy landscapes, our data suggest that inter-states. Rather than simply complicating folding stabilising non-native interactions in high energy states. This results accord with previous data, which suggest that remnants of the non-native contacts found in the Im7 intermediate persist in the rate-limiting transition state.14 Also, recent simulations of the rate-limiting transition state ensembles of Im7 and Im9 using molecular dynamics calculations have highlighted the importance of both native and non-native interactions in stabilising the transition states of these proteins.28

The rate of folding of naturally evolved protein sequences can therefore be increased in a number of ways, including decreasing contact order,33 optimising local interactions,34,35 altering hydrophobic packing36 or, as shown here, by introducing stabilising non-native interactions in high energy states. Rather than simply complicating folding energy landscapes, our data suggest that intermediates are an important feature of folding and, when such species do not become energetically trapped, the generation of non-native interactions can even enhance the rate of folding.

Whilst two-state folding transitions have been a major focus for experimentalists and theoreticians over the last few years, it is now becoming apparent that many proteins fold through intermediate states. For example, the presence of high-energy (unstable) intermediates that are on-pathway has been inferred recently for 16 proteins designated hitherto as two-state, by re-analysis of non-linearity in the logarithm of the folding or unfolding rate constants versus denaturant concentration.30 In accord with this view, an unstable intermediate has recently been proposed to be formed transiently during the folding of Im9 using an improved Go-like model.37 Unstable, but transiently populated intermediates are now increasingly being observed in the folding of small, single domain proteins as techniques capable of their detection become more widely employed.6–8 Rather than being unique to larger proteins, therefore, folding via distinct intermediates is a feature of the folding of many proteins, including small, single domain proteins that fold with apparent two-state kinetics.

Materials and Methods

Mutagenesis and protein purification

Mutagenesis was carried out using the QuikChange site-directed mutagenesis kit. All mutants were sequenced to ensure that the gene contained the desired change and no others. Mutants were expressed, purified and shown to be of the correct mass as described.14 Im7 and Im9 are hexahistidine-tagged versions of the wild-type proteins. Residue numbers referred to are those of the untagged proteins.

Stopped-flow fluorescence

All kinetic experiments were performed using an Applied Photophysics SX18.MV stopped-flow fluorimeter at 10 °C, as described.14 The dead-time of the instrument is 3 ms as determined using the quenching of N-acetyl tryptophanamide with N-bromosuccinimide. Folding was initiated by mixing 50 μM protein in 50 mM sodium phosphate buffer (either at pH 6.0 with 0.4 M Na2SO4 or at pH 7.0 without Na2SO4), containing 2 mM DTT, 1 mM EDTA and 8 M urea in a 1 : 10 ratio with the same buffer lacking protein, containing various concentrations of urea. Buffers containing sodium sulphate were prepared gravimetrically, to circumvent errors involved in measuring pH in the presence of high concentrations of sodium ions.38 Unfolding was measured in the same way, but the initial protein solution did not contain urea. All the refolding time-courses were well described by a double exponential function. The slower, second phase, which in all cases was <10% of the total amplitude, has previously been attributed to proline-limited folding events5 and is not considered further here. All unfolding data were well described by a single exponential function.

The initial (extrapolated to time=0) and final fluorescence signals determined from the fit to the raw data at each urea concentration were normalised using equation (1):

\[
F_{\text{norm}} = \frac{F_{\text{obs}} - F_{\text{buf}}}{F_{7.75} - F_{\text{buf}}}
\]

where \(F_{\text{obs}}\) is the observed fluorescence, \(F_{\text{buf}}\) is the fluorescence of the buffer alone and \(F_{7.75}\) is the signal of the unfolded state at 7.75 M urea.

Concentration dependence

To allow comparison of kinetic traces at a single denaturant concentration but different protein concentrations each trace was normalised to the signal of the unfolded state according to equation (2):

\[
\text{Normalised Fluorescence} = \frac{(\text{Refolding Signal})_{\text{M protein}} - (\text{Buffer Signal})_{\text{M protein}}}{(\text{Denatured Signal})_{\text{M protein}} - (\text{Buffer Signal})_{\text{M protein}}}
\]

In this manner, account is taken of the scattering effect of urea containing solutions, which can make a significant contribution to the measured signal at protein concentrations less than 1 μM.

Analysis of kinetic data

The kinetic data for all proteins studied were fitted to
either a two-state model (Scheme 1) or an on-pathway three-state model (Scheme 2):

where U, I and N represent the denatured, intermediate and native states, respectively and $k_{ui}$ is the microscopic rate constant for the conversion of U to I. Previous analysis using continuous-flow and stopped-flow fluorescence has shown that the intermediate formed during Im9* folding is on-pathway.\textsuperscript{12} Also, previous analysis of the folding of Im9*, as a function of pH, has shown that a model involving the movement of a single transition state cannot describe the complete data set.\textsuperscript{2} A three-state model in which the intermediate is on the pathway for folding was therefore used to fit the data for all Im9* variants showing three-state kinetics.

The data were fitted to Scheme 2 using three different approaches and the comparison of the resulting data used as a test of the robustness of the thermodynamic parameters determined from the fits. In method 1 the observed rate constants and initial and final fluorescence signals for each data set were fitted simultaneously by manually varying the parameters ($k_{ui}$, $k_{nu}$, $k_{ni}$ and $k_{nu}$) their respective $m$-values and the fluorescence signals of U, I and N) until the predicted data satisfactorily described the observed results, as described.\textsuperscript{5,13,14} The folding/unfolding kinetics of the intermediate could not be measured directly, therefore the formation of the intermediate was assumed to occur as a pre-equilibrium step; $k_{ui}$ was fixed at 20,000 s\textsuperscript{-1} and the stability of the intermediate was altered by changing only $k_{nu}$. Only the equilibrium parameters for the intermediate $K_{ui}=k_{nu}/k_{ni}$ and $M_{ui}=(m_{ui}+m_{iu})$ are listed.

In method 2 the logarithm of the rate constants were fitted to equation (3) using the non-linear regression function in Sigma-Plot (Jandel Scientific) without weighting. In this method the initial and final fluorescence signals were not considered.

\[
\ln k_{obs} = \ln \left( \frac{K_{ui}^0 \exp(-M_{ui}[\text{urea}]/RT)}{1 + K_{ui}^0 \exp(-M_{ui}[\text{urea}]/RT)} \right) k_{in}^0 \times \exp(-m_{ui}[\text{urea}]/RT) + k_{in}^0 \exp(m_{iu}[\text{urea}]/RT) \right) \]

(3)

In method 3 a two-step process was used to fit the rate constants and the initial and final fluorescence signals. The first step was a Monte Carlo method, which allowed estimation of all parameters by simultaneous weighted fitting of the observed rate constants and the initial and final fluorescence signals. Typically 1000 estimates were initially made and the best fit was used as a starting point for a second or third series of fits. Values for the normalised fluorescence signals of the native, intermediate and unfolded states (in water) and the urea-dependence of these fluorescence signals were determined and held constant during the subsequent fitting procedure. When a suitably good fit was obtained a non-linear least squares fit using the Marquardt-Levenburg algorithm (IGOR, WaveMetrics) was used to refine the parameters. The goodness of fit was judged from the chi-squared statistic and by the uniformity of the residuals. A reasonable initial fit is

\[
U = \frac{k_{in}}{k_{nu}} N
\]

Scheme 1.

Equilibrium denaturation

The stability of selected Im9* variants were also determined using equilibrium urea titration. Equilibrium denaturant titrations were monitored using fluorescence spectroscopy ($\lambda_{em}=280$ nm, $\lambda_{em}=360$ nm) at 10 °C, using a Photon Technologies International fluorimeter. Solutions contained 0-8 M urea in 50 mM sodium phosphate buffer (either at pH 6.0 with 0.4 M Na$_2$SO$_4$ or at pH 7.0 without Na$_2$SO$_4$), containing 2 mM DTT and 1 mM EDTA. A protein concentration of ~3 μM was used for each sample. The observed signal as a function of denaturant concentration was fitted to a two-state transition as described by equation (4),\textsuperscript{49} using the non-linear regression function in Sigma-Plot (Jandel Scientific):

\[
\text{Obs Signal} = \frac{[a(\text{urea}) + b] \exp(DG_{\text{un}}^0 - m_{\text{un}}(\text{urea})/RT) + [c(\text{urea}) + d]}{(1 + \exp(DG_{\text{un}}^0 - m_{\text{un}}(\text{urea})/RT))}
\]

(4)

where $a$ is the fluorescence signal of the native protein in water and $b$ the urea dependence of this signal, $c$ and $d$ are the fluorescence signals of the denatured protein and the dependence of this signal on the concentration of urea.

To allow comparison of data from different experiments and for different proteins, both the observed fluorescence signal (Obs. Signal) and the fit to the raw data were converted to relative fluorescence according to equation (5).

\[
\text{relative fluorescence} = \frac{\text{Obs Signal} - \text{Signal}_{\text{im}}}{\text{Signal}_{\text{im}} - \text{Signal}_{\text{im}}}
\]

(5)

The $DG_{\text{un}}$, and $M_{ui}$-values determined from these studies are in good agreement with those determined from kinetic measurements. These results are shown in the Supplementary Data.

$\Phi$-Value analysis

$\Phi_1$ and $\Phi_{\text{TS}}$-values were calculated as described.\textsuperscript{13} The $\Phi_1$-values were calculated according to equation (6):

\[
\Phi_1 = \frac{\Delta DG_{\text{un}}^{\text{wt}}-\text{mut}}{\Delta DG_{\text{un}}^{\text{wt}}-\text{mut}}
\]

(6)

The $\Phi_{\text{TS}}$-values were calculated with respect to the unfolded state, according to equation (7):

\[
\Phi_{\text{TS}} = \frac{\Delta DG_{\text{un}}^{\text{wt}}-\text{mut} + \Delta DG_{\text{un}}^{\text{wt}}-\text{mut}}{\Delta DG_{\text{un}}^{\text{wt}}-\text{mut}}
\]

(7)
Acknowledgements

We thank Keith Ainley for technical assistance, Alison Ashcroft for mass spectrometry and Colin Kleanthous, Anthony Beeble and Tom McLeish for helpful discussions. We acknowledge with thanks Andrew Capaldi and members of the Radford group for helpful contributions. C.T.F. is supported by the Biotechnology and Biological Sciences Research Council (BBSRC). S.E.R. is a BBSRC Professorial Fellow. This work was supported by the BBSRC, the University of Leeds and The Wellcome Trust. The authors are members of the Astbury Centre for Structural Molecular Biology, which is part of the North of England Structural Biology Centre and is funded by the BBSRC.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2004.06.076

References


A. R. Fersht

(Received 8 April 2004; received in revised form 25 June 2004; accepted 25 June 2004)