Trapping the On-pathway Folding Intermediate of Im7 at Equilibrium

Graham R. Spence, Andrew P. Capaldi and Sheena E. Radford*

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

The four-helical protein Im7 folds via a rapidly formed on-pathway intermediate ($k_{UI} = 3000\, s^{-1}$ at pH 7.0, 10 °C) that contains three (helices I, II and IV) of the four native $\alpha$-helices. The relatively slow ($k_{IN} = 300\, s^{-1}$) conversion of this intermediate into the native structure is driven by the folding and docking of the six residue helix III onto the developing hydrophobic core. Here, we describe the structural properties of four Im7* variants designed to trap the protein in the intermediate state by disrupting the stabilising interactions formed between helix III and the rest of the protein structure. In two of these variants (I54A and L53AI54A), hydrophobic residues within helix III have been mutated to alanine, whilst in the other two mutants the sequence encompassing the native helix III was replaced by a glycine linker, three (H3G3) or six (H3G6) residues in length. All four variants were shown to be monomeric, as judged by analytical ultracentrifugation, and highly helical as measured by far-UV CD. In addition, all the variants denature co-operatively and have a stability ($\Delta G_{uf}$) and buried hydrophobic surface area ($M_{uf}$) similar to those of the on-pathway kinetic intermediate. Structural characterisation of these variants using 1-anilino-8-napthalene sulphonic acid (ANS) binding, near-UV CD and 1D $^1$H NMR demonstrate further that the trapped intermediate ensemble is highly structured with little exposed hydrophobic surface area. Interestingly, however, the structural properties of the variants I54A and L53AI54A differ in detail from those of H3G3 and H3G6. In particular, the single tryptophan residue, located near the end of helix IV, and distant from helix III, is in a distinct environment in the two sets of mutants as judged by fluorescence, near-UV CD and the sensitivity of tryptophan fluorescence to iodide quenching. Overall, the results confirm previous kinetic analysis that demonstrated the hierarchical folding of Im7 via an on-pathway intermediate, and show that this species is a highly helical ensemble with a well-formed hydrophobic core. By contrast with the native state, however, the intermediate ensemble is flexible enough to change in response to mutation, its structural properties being tailored by residues in the sequence encompassing the native helix III.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: immunity protein; folding; equilibrium intermediate; kinetic analysis; rational design

*Corresponding author

Introduction

Many proteins, including a number of small, single domains, populate partially structured intermediates whilst folding into the native state. Despite a long-standing interest in such folding intermediates, the nature and importance of these species remains the subject of debate. Some studies suggest that partially folded species result from non-specific collapse of the unfolded state under native conditions, or arise from transient
intermolecular association events. However, other studies have demonstrated that some intermediates are distinct states separated from the unfolded ensemble by an energetic barrier and/or are monomeric species, the formation of which is independent of protein concentration. There is also disagreement regarding the role that intermediates play in folding. Whilst some intermediates appear to be off-pathway kinetically, other intermediates appear to be productive on-pathway species that serve to narrow the conformational search for the native structure. Irrespective of their role(s) in folding (and this may vary from protein to protein), it is important to understand the structure of intermediate ensembles, the mechanism of their formation and their relationship to the native state. Despite the central importance of these questions for delineating structural models of folding, however, the rapid formation, instability and transient nature of folding intermediates makes answering these questions a challenging task, especially for single-domain proteins that fold rapidly to the native state.

Information about the structural properties of intermediates can be determined using a number of methods, which are most powerful when used in combination. In particular, \( \phi \)-value analysis, ultra-rapid mixing methods, laser temperature-jump and hydrogen-exchange techniques have all helped to shed light on the nature and importance of early events in folding. However, despite the advances made using these methods, these approaches provide only limited structural detail about transient intermediate states. For example, whilst \( \phi \)-value analysis can be used to provide a measure of the contribution of an individual side-chain to the stability of an intermediate, the identity of the side-chains it interacts with cannot be determined without recourse to detailed analysis using double-mutant cycles, which can be difficult or impossible for marginally stable states. Similarly, whilst hydrogen-exchange methods can be used to identify amide protons that are protected from exchange early during folding, the hydrogen bond acceptor cannot be identified using this approach, and no information is gained about the structure of the core. Finally, detailed structural analysis of intermediates using NMR is often obviated by the rapidity of folding and/or line-broadening effects caused by conformational exchange. Therefore, studying partially folded states that are trapped at equilibrium offers the most promising approach to understanding intermediates on a structural level, provided that the trapped and transient intermediates are highly related in both structural and energetic terms.

In previous studies, we have used both stopped-flow and ultra-rapid mixing methods to analyse the folding of the bacterial immunity protein, Im7, and have shown that the protein folds by a three-state mechanism in which an intermediate is populated transiently during folding (\( k_{\text{f}} = 3000 \text{ s}^{-1} \), \( k_{\text{uf}} = 21.6 \text{ s}^{-1} \) at pH 7.0 in 0.4 M Na\( _2 \)SO\( _4 \) (Figure 1). This analysis demonstrated that the intermediate is hyperfluorescent (it is 1.4 times more fluorescent than the denatured state and 5.8-fold more fluorescent than native Im7), stable (\( \Delta G_{\text{UI}} = -11.9 \text{ kJ mol}^{-1} \)), compact (\( \beta_i = 0.79 \)) and, importantly, on-pathway. Additionally, using \( \phi \)-value analysis and native-state hydrogen exchange, we have shown that the intermediate contains three of the four native helices (I, II and IV), but lacks helix III (Figures 1 and 2). On the basis of the observation that more than ten of the 30 \( \phi \)-values measured decrease between the intermediate and rate-limiting transition states, we have suggested that the on-pathway intermediate of Im7 is stabilised, in part, by non-native interactions; helices I, II and IV docking in a non-native manner so as to sequester hydrophobic residues buried in the native state by helix III. The rate-determining step in folding then involves re-orientation of helices I and IV, to reveal the hydrophobic binding site for helix III. This short, six residue, helix then docks onto the developing structure and effectively locks the protein in the native state (Figure 1).

Despite the wealth of information about the folding intermediate of Im7 obtained from kinetic experiments, key questions remain unresolved, including the mechanism by which the intermediate forms and a detailed structural analysis of the intermediate species itself. Here, we use mutation of residues within helix III of native Im7* (a hexahistidine-tagged version of the protein), that are

\[
D = \frac{k_{\text{f}}}{k_{\text{uf}}} = \frac{21.6}{3000} = 0.0072
\]

\[
E = \frac{k_{\text{f}}}{k_{\text{uf}}} = \frac{3000}{21.6} = 140.4
\]

\[
\Delta G_{\text{UI}} = -RT \ln \left( \frac{k_{\text{f}}}{k_{\text{uf}}} \right) = -298 \times \ln (0.0072) = -11.9 \text{ kJ mol}^{-1}
\]

\[
\beta_i = \frac{\Delta G_{\text{UI}}}{\Delta H_{\text{UI}}} = \frac{-11.9}{10} = 0.79
\]

\[
k_{\text{f}} = 3000 \text{ s}^{-1}
\]

\[
k_{\text{uf}} = 21.6 \text{ s}^{-1}
\]

\[
pH 7.0, 0.4 \text{ M Na}_2\text{SO}_4
\]

\[
\text{U, unfolded state; I, intermediate state; N, native state; TS}_1 \text{ and TS}_2, \text{ early and rate-determining transition states, respectively. How helices I, II and IV dock in I and TS}_2 \text{ is not known and their orientations shown here are arbitrary.}
\]
known or predicted from $\phi$-value analysis to destabilise selectively the native protein relative to the transient intermediate, to trap the kinetic on-pathway intermediate at equilibrium. Four mutants have been generated based on two different design strategies, and the thermodynamic, spectroscopic, and structural properties of the proteins have been studied using a wide range of biophysical methods. The results demonstrate that by removing hydrophobic residues from helix III or replacing this helix with a three or six residue glycine linker, an intermediate is trapped at equilibrium that resembles closely the transient intermediate populated during refolding of the wild-type protein. However, despite the remarkable similarities between the thermodynamic properties of these mutants, differences in their spectroscopic properties suggest that these species possess subtle, but significant, differences in their structural characteristics.

### Results

#### Trapping the kinetic folding intermediate of Im7* at equilibrium

Previous $\phi$-value analysis has indicated that the kinetic intermediate of Im7* is a compact ($\beta = 0.79$) species that contains the three long helices (I, II, and IV) found in native Im7, but lacks the short helix III (Figure 1).12 Although helix III is only six residues long, it plays a key role in the late stages of Im7 folding by docking residues Thr51, Leu53 and Ile54 onto the open face of helices I, II and IV after the rate-limiting transition state has been traversed (the average values for $\phi_o$ and $\phi_{TS}$ for residues mutated in helix III are $0.05 \pm 0.03$ and $0.06 \pm 0.05$, respectively).12 Mutation of residues in helix III thus selectively stabilises the native state relative to the intermediate and rate-limiting transition states. Interestingly, helix III plays an important role in determining the binding energy of the immunity protein for its cognate and non-cognate colicin toxins.28 As a consequence of its importance in both structure and function, the sequence of helix III is highly conserved in the four E. coli colicin-binding immunity proteins (Im2, Im7, Im8 and Im9).29

To characterise the on-pathway kinetic intermediate of Im7* (I$^{\text{on}}$) in more detail, a series of mutants were designed with the aim of preventing helix III from docking onto the developing native structure, thus trapping the protein in the intermediate state. Two strategies were employed. In the first strategy, the side-chains of Leu53 and Ile54 in helix III (Figure 2(a)), that are critical for stabilising the interaction between helix III and its binding surface on helices I, II and IV, were truncated. Our previous work has shown that mutating Leu53 to alanine destabilises the native state of wild-type Im7* by 13.6 kJ mol$^{-1}$, whilst the stability of the intermediate is not affected ($\Delta G_{\text{UI}} = -11.9$ kJ mol$^{-1}$ and $\Delta G_{\text{UF}} = -11.3$ kJ mol$^{-1}$) for Im7* L53A at 10°C in 0.4 M sodium sulphate, whilst $\Delta G_{\text{UI}} = -11.9$ kJ mol$^{-1}$ and $\Delta G_{\text{UF}} = -24.9$ kJ mol$^{-1}$ for wild-type Im7* under the same conditions).12 Similarly, replacing Ile54 with alanine destabilises the native state by more than 20 kJ mol$^{-1}$, whilst the intermediate is destabilised by only 0.6 kJ mol$^{-1}$.12 The mutation I54A thus destabilises the native state to such an extent that it is no longer the most stable state in the absence of denaturant (denaturant $\Delta G_{\text{UI}} = -11.3$ kJ mol$^{-1}$, $\Delta G_{\text{UF}} = -4.3$ kJ mol$^{-1}$ at 10°C, pH 7.0, 0.4 M sodium sulphate).12 As a consequence, in these conditions only around 6% of I54A molecules are in the native state, whilst more than 90% of molecules populate the partially folded intermediate at equilibrium (I$^{\text{on}}$). Mutation of Ile54 to alanine therefore provides one route by which to populate the kinetic intermediate at equilibrium by substitution of only a single side-chain. To further increase the population of the intermediate at equilibrium, the double mutant L53A54A was created.
On the basis of the thermodynamic parameters estimated from kinetic analysis of each single mutant (and assuming that the effect of each mutation is additive), it is predicted that this double mutation should destabilise native I* by nearly 30 kJ mol$^{-1}$, but destabilise the intermediate by less than 1 kJ mol$^{-1}$, resulting in a population of Ieqm of approximately 99% at equilibrium ($\Delta G_{UI} = -11.3$ kJ mol$^{-1}$, $\Delta G_{Uf} = +3.9$ kJ mol$^{-1}$).

The second strategy for creating Ieqm involved replacing the six residues (Thr51–Tyr56) that comprise helix III in native I*, with a polyglycine sequence. Although $\phi$-value analysis suggests helix III is not docked onto the rest of the structure in I*,$^{12}$ and hydrogen-exchange studies show that residues 53–56 do not form stable hydrogen bonded structure at this stage of folding,$^{27}$ these analyses cannot rule out involvement of residues that were not mutated or analysed in these studies in tailoring the properties of the intermediate ensemble. Two lengths of glycine linker were chosen. In the first mutant, the sequence encoding helix III was replaced with a Gly$_3$ linker (H3G3; Figure 2(b) and (c)) to mimic the length of this sequence in a helical conformation. In the second mutant, the six residue sequence of helix III was replaced with a Gly$_6$ linker (H3G6; Figure 2(b) and (c)), introducing an unfolded chain equivalent in length to unfolded helix III.

**Structural properties of the Ieqm mutants**

All of the I* mutants described above were expressed to high levels and purified to homogeneity (see Materials and Methods). The mutants were shown to be monomeric over the concentration range 1–500 $\mu$M in buffer A (50 mM sodium phosphate buffer (pH 7.0), 0.4 M sodium sulphate, 1 mM EDTA) at 10°C, as judged by analytical ultracentrifugation (Table 1). The solubility of these mutants allowed us to carefully characterise their structural properties using a number of techniques. The near-UV circular dichroism (CD) spectra of native I* and the four variants are shown in Figure 3(a). The near-UV CD spectrum of wild-type I* has two negative peaks, one between 270 nm and 280 nm, and the other at ~295 nm. These peaks are diagnostic of fixed packing, presumably of the two buried tyrosine residues (Tyr10 and Tyr55) and single tryptophan residue (Trp75), respectively, in wild-type I*.28 The near-UV CD spectra of the four Ieqm mutants have substantial negative ellipticity around 270–280 nm, demonstrating that each of these species contains fixed, tertiary packing involving tyrosine residues. Interestingly, however, the intensity of these bands in the spectra of the variants differs significantly from that of the wild-type protein, consistent with the view that the packing of aromatic side-chains is not yet native in the variant proteins. In accord with this view, the intensity of the near-UV CD absorbance band at ~295 nm for the four proteins also differs significantly from that of wild-type I*.

Thus, whilst the near-UV CD spectra of wild-type I* and L53A I54A contain significant intensity at ~295 nm, I54A, H3G3 and H3G6 show no chirality of Trp75. Together, these data suggest that whilst the four variants contain fixed tertiary structure, the arrangement of aromatic side-chains and/or the dynamics of the ensemble differs in each variant and from the wild-type protein.
The far-UV CD spectra of wild-type Im7* and the four mutant proteins are shown in Figure 3(b). All of the spectra have the signature of an α-helical protein, with strong positive peaks at 222 nm and 208 nm, and a large positive peak at <195 nm. Importantly, all of the mutant proteins show a small, but significant, decrease in ellipticity relative to wild-type Im7*. Estimation of the number of residues in a helical conformation using the mean residue ellipticity relative to the known helical content of native Im7*.

Binding of the hydrophobic dye 1-anilino-8-naphthalene sulphonic acid (ANS) was used to probe structural differences between native, wild-type Im7* and the folded states of the Ieqm mutants. ANS binds to exposed hydrophobic surface area of native Im7* and the folded states of the Ieqm mutants. The addition of ANS to the Im7* mutants results in a blue shift in the ANS fluorescence emission maximum and an increase in fluorescence intensity.

Table 1. Spectroscopic properties of wild-type Im7* and the mutants I54A, L53AI54A, H3G3 and H3G6 determined in buffer A at 10°C

<table>
<thead>
<tr>
<th></th>
<th>Im7*</th>
<th>I54A</th>
<th>L53AI54A</th>
<th>H3G3</th>
<th>H3G6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass (ESI-MS) (Da)</td>
<td>10,846</td>
<td>10,805</td>
<td>10,763</td>
<td>10,250</td>
<td>10,420</td>
</tr>
<tr>
<td>Expected mass (Da)</td>
<td>10,848</td>
<td>10,805</td>
<td>10,763</td>
<td>10,250</td>
<td>10,421</td>
</tr>
<tr>
<td>MW (AUC) (kDa)</td>
<td>8.93 ± 0.39</td>
<td>11.10 ± 0.58</td>
<td>10.34 ± 0.68</td>
<td>11.65 ± 0.81</td>
<td>11.70 ± 0.70</td>
</tr>
<tr>
<td>( \varepsilon_{\text{max}} ) (M⁻¹ cm⁻¹)</td>
<td>8554</td>
<td>8051</td>
<td>8598</td>
<td>6252</td>
<td></td>
</tr>
<tr>
<td>( \lambda_{\text{max}} ) (Trp) (nm)</td>
<td>310 ± 0.3</td>
<td>333 ± 0.3</td>
<td>336 ± 0.3</td>
<td>340 ± 0.3</td>
<td>340 ± 0.6</td>
</tr>
<tr>
<td>Intensity (Trp)</td>
<td>0.29 ± 0.01</td>
<td>1.44 ± 0.01</td>
<td>1.48 ± 0.01</td>
<td>1.02 ± 0.01</td>
<td>0.96 ± 0.01</td>
</tr>
<tr>
<td>Intensity (ANS)</td>
<td>471 ± 2</td>
<td>473 ± 0.6</td>
<td>471 ± 0.6</td>
<td>479 ± 1.2</td>
<td>477 ± 0.7</td>
</tr>
<tr>
<td>Intensity (ANS)</td>
<td>1.0</td>
<td>3.8 ± 0.1</td>
<td>7.5 ± 0.2</td>
<td>3.8 ± 0.1</td>
<td>6.4 ± 0.2</td>
</tr>
<tr>
<td>% Helix</td>
<td>50</td>
<td>40</td>
<td>45</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>( K_{\text{diss}} ) (M⁻¹)</td>
<td>1.5 ± 0.2</td>
<td>2.8 ± 0.1</td>
<td>2.8 ± 0.1</td>
<td>3.8 ± 0.2</td>
<td>4.0 ± 0.2</td>
</tr>
<tr>
<td>( \Delta G_{\text{diss}} ) (kJ mol⁻¹)</td>
<td>-23.6 ± 0.6</td>
<td>-11.4 ± 1.2</td>
<td>-10.1 ± 0.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( M_{\text{helix}} ) (Fl)</td>
<td>4.9 ± 0.14</td>
<td>3.4 ± 0.29</td>
<td>3.4 ± 0.09</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \Delta G_{\text{CD}} ) (kJ mol⁻¹)</td>
<td>-24.0 ± 1.5</td>
<td>-10.4 ± 1.9</td>
<td>-11.1 ± 1.3</td>
<td>-10.0 ± 1.2</td>
<td>-10.2 ± 1.3</td>
</tr>
<tr>
<td>( M_{\text{CD}} )</td>
<td>4.6 ± 0.31</td>
<td>3.4 ± 0.51</td>
<td>3.5 ± 0.35</td>
<td>3.1 ± 0.31</td>
<td>3.1 ± 0.34</td>
</tr>
</tbody>
</table>

a Mass quoted was determined for samples at 500 μM. The error quoted is 1 standard deviation.

b Measured in the folded state.

c The error quoted is the standard error for three replicate experiments.
d Measured in the folded state and normalized to the signal of the denatured protein in 8 M urea.
e Data are normalised to the fluorescence signal of native Im7* in the presence of ANS.
f Determined from the mean residue ellipticity relative to the known helical content of native Im7*.
g The error quoted is the regression error.
h Equilibrium stability (kJ mol⁻¹) and M-value (kJ mol⁻¹ M⁻¹).
near-UV CD, these data suggest that, whilst each of the Im7* variants is folded to a specific structure, the precise packing of residues in the hydrophobic core of each protein and/or the conformational dynamics of each species differs in detail.

**Equilibrium denaturation of the Im7^* proteins**

In order to determine the stability of the mutant proteins, equilibrium denaturation curves were measured using both intrinsic tryptophan fluorescence and circular dichroism (Figure 5(a) and (b)). The equilibrium denaturation curve of wild-type Im7* determined by intrinsic tryptophan fluorescence is shown in Figure 5(a). The data show that native Im7* is significantly less fluorescent than its denatured ensemble, probably due to quenching from the close interaction of His47 with the single tryptophan residue (Trp75) in the native protein.\(^{28}\) As a consequence, denaturation of wild-type Im7* leads to a substantial increase in fluorescence intensity (Figure 5(a)). Fitting these data to a two-state transition yields a stability of \(\Delta G_{\text{UF}} = 23.6(\pm 0.6) \text{ kJ mol}^{-1}\) and an \(M_{\text{UF}}\)-value of 4.9 (\pm 0.14) \text{ kJ mol}^{-1} \text{ M}^{-1}\ (Table 1). These values accord well with previous data acquired under identical conditions using equilibrium methods\(^{37}\) or stopped-flow fluorescence.\(^{12,37}\) Further, following this transition by far-UV CD gives a \(\Delta G_{\text{UF}} = -24.0(\pm 1.5) \text{ kJ mol}^{-1}\) and \(M_{\text{UF}}\) of 4.6 (\pm 0.31) \text{ kJ mol}^{-1} \text{ M}^{-1}\, confirming the validity of the two-state fit (Table 1). The equilibrium unfolding curves of I54A and L53AI54A are shown in Figure 5(a). Interestingly, both of these proteins show a decrease in fluorescence intensity upon denaturation (compared with an increase observed during denaturation of wild-type Im7*) indicating that, as with I54A\(^{12}\) these proteins are hyperfluorescent. The thermodynamic parameters for I54A estimated by fitting these data to a two-state transition (\(\Delta G_{\text{UF}} = -11.4(\pm 1.2) \text{ kJ mol}^{-1}\), \(M_{\text{UF}} = 3.4(\pm 0.29) \text{ kJ mol}^{-1} \text{ M}^{-1}\); Table 1) agree closely with those determined using CD (\(\Delta G_{\text{UF}} = -10.4(\pm 1.9) \text{ kJ mol}^{-1}\), \(M_{\text{UF}} = 3.4(\pm 0.51) \text{ kJ mol}^{-1} \text{ M}^{-1}\); Table 1), suggesting that this protein has adopted a cooperatively stabilised structure that unfolds with a two-state transition. The unfolding behaviour of the double mutant L53AI54A is very similar to that of I54A (Table 1). The fact that mutation of Leu53 to Ala in wild-type Im7* leads to a

---

**Figure 4.** The 1D \(^1\)H NMR spectra (500 MHz) of native wild-type Im7* and the helix III variants I54A, L53AI54A, H3G3 and H3G6, showing the aromatic and amide region of the spectra (11.0 to 6.0 ppm) and methyl region (3.0 to 0.75 ppm). The spectra were acquired in buffer A containing 10% (v/v) \(^2\)H\(_2\)O at 10 °C (pH\(_{\text{meas}}\) 7.0).

**Figure 5.** Urea-induced denaturation of wild-type Im7* and the four helix III mutants (coloured as in Figure 2), measured using (a) intrinsic tryptophan fluorescence and (b) the change in CD signal at 225 nm. Fluorescence data are normalised to the fluorescence signal of each protein in 8 M urea. The different intensities of the proteins in 0 M urea reflect their different fluorescence emission spectra (see Figure 6). The CD data are scaled to the CD signal at 225 nm of wild-type Im7* in the absence of urea. The continuous lines show the best fit of the data to a model describing a two-state transition, as described in Materials and Methods. Where there is little change in fluorescence between the native and denatured states (H3G3 and H3G6 in (a)), the data were not fitted directly. For these proteins the continuous line is calculated from parameters determined from the denaturation transitions of these proteins monitored using far-UV CD. The resulting thermodynamic parameters are shown in Table 1.
destabilisation of more than 13 kJ mol\(^{-1}\), but has a negligible effect on the stability of I54A, demonstrates further that both I54A and L53AI54A fold into a conformation in which the sequence encompassing the native helix III is unstructured. Furthermore, the remarkable similarity between the values for \(\Delta G_{\text{UF}}\) and \(M_{\text{UF}}\) for both I54A and L53AI54A with those determined for the on-pathway intermediate populated during Im7\(^*\) folding calculated from stopped-flow fluorescence (\(\Delta G_{\text{UF}} = -11.8(\pm 1.7)\) kJ mol\(^{-1}\), \(M_{\text{UF}} = 3.87(\pm 0.27)\) kJ mol\(^{-1}\) M\(^{-1}\)) (measured here, see Materials and Methods),\(^{12}\) demonstrates that the mutant proteins and the kinetic intermediate are highly related species.

Surprisingly, and in contrast to I54A and L53AI54A, the fluorescence intensity of Trp75 in the folded states of H3G3 and H3G6 is similar to that of their denatured states (Figure 5(a)). The relatively small change in the fluorescence intensity of Trp75 upon denaturation of these mutants obviates quantitative analysis of their equilibrium denaturation properties using fluorescence. However, these proteins are highly helical (Figure 3(b)), allowing analysis of their equilibrium unfolding properties using far-UV CD (Figure 5(b)). Measuring the unfolding of H3G3 and H3G6 in this manner reveals that \(\Delta G_{\text{UF}}\) for H3G3 and H3G6 is \(-10.0(\pm 1.2)\) kJ mol\(^{-1}\) and \(-10.2(\pm 1.3)\) kJ mol\(^{-1}\), respectively, while \(M_{\text{UF}}\) is \(3.1(\pm 0.31)\) kJ mol\(^{-1}\) M\(^{-1}\) and \(3.1(\pm 0.34)\) kJ mol\(^{-1}\) M\(^{-1}\), respectively. These values are similar to the corresponding parameters for I54A and L53AI54A (Table 1), suggesting that the glycine-containing proteins adopt a structure that is similar to that of the two site-directed mutants. The similarity in the stability and \(M_{\text{UF}}\)-value of the four Im7\(^*\) variants with the kinetic intermediate populated during Im7\(^*\) folding provides strong evidence to support the hypothesis that these species are closely related, and confirms the validity of the two-state assumption used previously to describe the transition between the unfolded and intermediate states during the kinetic folding of wild-type Im7.\(^{11}\)

The equilibrium \(M_{\text{UF}}\)-value reflects the change in exposed hydrophobic surface area between the unfolded and folded states.\(^{28}\) Measurement of the \(M_{\text{UF}}\)-value, therefore, can be used to test the compactness of the mutants compared with wild-type Im7\(^*\) and I\(^{\text{kin}}\). The native state of wild-type Im7\(^*\) has a kinetic \(M_{\text{UF}} = 4.9(\pm 0.29)\) kJ mol\(^{-1}\) M\(^{-1}\) (measured here, see Materials and Methods)\(^{12}\) compared with \(4.9(\pm 0.14)\) kJ mol\(^{-1}\) M\(^{-1}\) determined here and elsewhere using equilibrium experiments (Table 1).\(^{27}\) By contrast, the kinetic intermediate has \(M_{\text{UF}} = 3.87(\pm 0.27)\) kJ mol\(^{-1}\) M\(^{-1}\), demonstrating that this species is about 20% expanded relative to native wild-type Im7\(^*\) (\(\beta = 0.79\)). Consistent with this, measurement of the hydrodynamic radii of wild-type Im7\(^*\) and the four variants in buffer A at 10°C (see Materials and Methods) indicated that the mutant proteins are expanded by \(~30\%) relative to wild-type Im7\(^*\) (\(R_\text{H}\) values \(~17(\pm 1)\) Å and \(~22(\pm 1)\) Å for wild-type Im7\(^*\) and the four variants, respectively (S. Whittaker, R. Boezi, G. R. Moore, G.R.S. & S.E.R., unpublished results)). Equilibrium measurements made here using fluorescence and CD for all four helix III variants result in \(M_{\text{UF}}\) values closely similar to that of I\(^{\text{kin}}\) and significantly smaller than that of wild-type Im7\(^*\) (Table 1). These data confirm that the folded structures of all four variants are similarly compact and, in common with I\(^{\text{kin}}\), are about 20–30% expanded relative to the native state of wild-type Im7\(^*\).

### Origins of the hyperfluorescence of I\(^{\text{kin}}\)

The equilibrium denaturation experiments shown in Figure 5 demonstrate that, whilst the folded states of I54A and L53AI54A are hyperfluorescent, the fluorescence intensities of H3G3 and H3G6 in their folded and unfolded states are similar. To investigate this further, fluorescence emission spectra of the folded states of wild-type Im7\(^*\) and the four I\(^{\text{kin}}\) mutants were measured and compared (Figure 6(a)). The fluorescence emission of native wild-type Im7\(^*\) is highly quenched, presumably as a consequence of a ring-stacking interaction between Trp75 and His47 (D. Gill, G.R.S. & S.E.R., unpublished results),\(^{28}\) leading to depopulation of the excited state of tryptophan through non-radiative pathways. The single, broad emission peak has a \(\lambda_{\text{max}}\) of approximately 310 nm, presumably reflecting the fact that the major contribution to the fluorescence emission of native wild-type Im7\(^*\) is the three tyrosine residues. The fluorescence emission spectra of the folded states of I54A and L53AI54A differ markedly from that of wild-type Im7\(^*\). Both spectra are now dominated by tryptophan emission, with a \(\lambda_{\text{max}}\) of approximately 334 nm and intensities that are 1.5-fold greater than their unfolded states (the latter in 8 M urea) (Table 1). These data demonstrate that Trp75 is buried from solvent in these mutants, the large increase in fluorescence presumably occurring because Trp75 is no longer quenched by His47. In accord with this view, Stern–Volmer analysis indicates that Trp75 is relatively inaccessible to iodide ions in these mutants (compared with N-acetyltryptophanamide), although the Trp is significantly more accessible to iodide ions than in the native state of wild-type Im7\(^*\) (Figure 6(b) and Table 1). By contrast with these results, and consistent with the data shown in Figure 5, the fluorescence intensity of Trp75 in the mutants H3G3 and H3G6 is similar to that of their denatured states. Nonetheless, the emission \(\lambda_{\text{max}}\) is blue-shifted relative to denatured Im7\(^*\) and red-shifted relative to I54A and L53AI54A. These data are consistent with at least partial burial of Trp75 from solvent in the glycine-containing mutants. Accordingly, the Stern–Volmer constant (\(K_{\text{SV}}\)) for H3G3 and H3G6 is increased relative to that of I54A and L53AI54A (Figure 6(b) and Table 1). These data suggest that, despite having similar overall structural properties,
the local environment of Trp75 differs significantly in the glycine-containing and site-directed mutant proteins.

**Discussion**

**Summary of the properties of the Ieqm mutants**

Whilst the kinetic intermediate observed during the folding of Ieqm is populated only transiently on the sub-millisecond timescale, a number of its properties have been inferred from detailed kinetic studies.\(^{11,12,27,37}\) Ultra-rapid mixing experiments monitored using fluorescence, together with protein engineering and stopped-flow experiments, showed that the Ieqm folding intermediate is on-pathway and contains three of the four α-helices that comprise the native structure. This species is hyperfluorescent, contains specifically packed and co-operatively stabilised tertiary structure and is stabilised by both native and non-native interactions.\(^{12}\) However, obtaining more detailed structural information about this species is limited by the transient nature of the kinetic intermediate and the rapidity of Ieqm folding, even at 10°C.

Here, we have exploited our knowledge of the Ieqm folding mechanism determined using kinetic experiments to engineer the Ieqm sequence by specific redesign such that the kinetic intermediate becomes populated at equilibrium. Two design strategies were employed with the aim of destabilising the native state selectively by preventing the binding of helix III, such that the on-pathway intermediate becomes the ground-state of the folding reaction. The success of both strategies provides strong evidence for the validity of the three state on-pathway model for Ieqm folding. Moreover, the close structural and thermodynamic similarities of Ieqm and Ikin allows more detailed insights into the structural properties of the intermediate than was possible hitherto. Thus, the data have shown that the partially folded intermediate of Ieqm has a stable, co-operative and specifically packed core, indicating that this species is not a randomly collapsed state. Moreover, even at a concentration of 500 μM, the Ieqm species are monomeric, ruling out transient intermolecular association as the cause of the burst-phase kinetics in the folding of the wild-type protein, consistent with previous data that showed no dependence of the rates or amplitudes of folding of the wild-type protein over a wide range of concentrations (50 nM to 50 μM; C. T. Friel & S.E.R., unpublished results). In addition, we show that the Ieqm folding intermediate has a high α-helical content, the far-UV CD spectra of the Ieqm states suggesting that the mutants have five to ten residues less helical structure than wild-type Ieqm, in remarkable agreement with prediction of the structure of the intermediate from ψ-value analysis and hydrogen-exchange measurements.\(^{12,27}\) We propose, therefore, that akin to the kinetic intermediate in Ieqm folding, the variants of Ieqm created here are also three-helical species. Finally, the four Ieqm species denature with two-state transitions characterised by a stability of 11 kJ mol\(^{-1}\) and \(\Delta G_{\text{UI}}\)-values of 3.1 to 3.5 kJ mol\(^{-1}\) M\(^{-1}\); values that are remarkably similar to the stability (\(\Delta G_{\text{UI}} = -11.8\) kJ mol\(^{-1}\)) and compactness (\(M_{\text{UI}} = 3.87\) kJ mol\(^{-1}\) M\(^{-1}\)) of the on-pathway kinetic intermediate. Overall, therefore, the data demonstrate that Ieqm and Ikin are structurally and thermodynamically closely related species.
The on-pathway intermediate is a relatively malleable state

Whilst the data presented above show that the folded states of all four mutants are structurally similar, subtle but significant differences exist in the properties of the different variants. Thus, whilst all four variants are highly helical, the proteins differ markedly in their spectroscopic properties. For example, the four variants have different near-UV CD spectra, different abilities to bind the hydrophobic dye ANS and different chemical shift dispersion. Despite the fact the mutations made are distant from the single tryptophan residue, Trp75, the variants also differ in their fluorescence properties, I54A and I53A being hyperfluorescent, whilst H3G3 and H3G6 are not. Interestingly, previous studies of Im7* using $\phi$-value analysis has shown that the single-point mutations T51S, L53A and I54V, which are buried in the core of native Im7* and form part of helix III, have no effect on the fluorescence properties of Im7*. Together with the data presented above, these results suggest that the side-chains of Asp52, Tyr55 and/or Tyr56, all of which are solvent-exposed in native Im7*, may be responsible for the dramatic effect of replacing the sequence of helix III with glycine residues in H3G3 and H3G6 on the fluorescence intensity of Trp75. This effect could be either direct by contributing non-native interactions in the core of the partially folded state (these residues are $>12.5 \text{ Å}$ apart in native Im7*), or indirect by altering the conformational properties of this ensemble, perhaps enabled by the insertion of a highly flexible glycine sequence. Interestingly, molecular dynamics simulations suggest that Tyr55 and Tyr56 form significant non-native interactions with Trp75 in the rate-limiting transition state formed during the folding of wild-type Im7*, and Tyr55 also makes a strong non-native interaction with Glu46 in this state. These data raise the intriguing possibility that the residues could contribute to the structural properties of Im7*. Further experiments using site-directed mutagenesis will now be needed to determine the role of these natively solvent-exposed residues in Im7* folding and in tailoring the fluorescence properties of Im7* and Im7**.

In summary, the data presented above demonstrate that mutation of only a single residue is sufficient to trap the kinetic intermediate formed during Im7* folding such that it becomes the most highly populated state at equilibrium. These results will now permit more detailed structural analysis of the on-pathway kinetic intermediate using multi-dimensional NMR methods, and offer a route towards determining the mechanism of intermediate formation using $\phi$-value analysis of the folding and unfolding properties of the Im7** species. The difference in spectroscopic properties of the Im7** species formed by the four variants described here highlight the sensitivity of using tryptophan fluorescence as a probe of the structure of the immunity proteins, and demonstrate that, whilst the intermediate populated during Im7* folding is remarkably well-defined, it is nonetheless more malleable than the native state. This raises the intriguing possibility that the intermediate populated during Im7 folding represents a pri-mordial three-helical species that evolved to display new functional or structural attributes by the insertion or substitution of a small number of amino acid residues. The ability to trap the kinetic intermediate species of Im7* at equilibrium will now permit an unprecedented experimental view of the early stages in immunity protein folding and provide a platform with which to test the hypothesis that new structures and/or functional domains can be evolved from the three-helical intermediate state.

Materials and Methods

Expression, purification and mutagenesis

Im7*, wild-type Im7 containing an N-terminal histidine tag, was used as the background for all studies. Residue numbers stated are those for the wild-type protein. All mutants, apart from H3G6, were created using the Stratagene QuickChange site-directed mutagenesis kit following the manufacturer’s instructions, except that the incubation time with DpnI was increased from one hour to overnight. H3G6 was created using mega-primer PCR. In all cases, sequencing established that the genes contained only the desired changes. Proteins were over-expressed and purified as described, and shown to be $>$95% pure (using SDS-PAGE) and within 2 Da of the expected mass (using electrospray ionisation mass spectrometry (ESI-MS); Table 1). The yield of pure protein was approximately 20 mg/l of culture medium. All proteins were dissolved in buffer A (50 mM sodium phosphate, 0.4 M sodium sulphate, 1 mM EDTA (pH 7.0)) and were analysed at 10 °C, unless stated otherwise.

Determination of molar extinction coefficients

The extinction coefficient of each folded protein was determined as described. Briefly, the concentration of a sample of each protein denatured in 20 mM sodium phosphate and 6 M guanidinium chloride was determined by measuring its absorbance at 280 nm. Each sample was then refolded by tenfold dilution into 20 mM sodium phosphate buffer and the extinction coefficient of the folded state calculated from measurement of the absorbance of this sample.

Analytical ultracentrifugation

Sedimentation velocity experiments were carried out using a Beckman Optima XL-A analytical ultracentrifuge (Beckman, Palo Alto, CA). To determine the molecular mass distribution of each protein, double-sector cells were filled with $420 \mu l$ of protein (500 μM, 100 μM, 10 μM and 1 μM) and $420 \mu l$ of buffer A in their respective channels. Samples were centrifuged at between 50,000 and 60,000 rpm at 10 to 13 °C using an An-60 Ti rotor and solute distributions were scanned using both
absorbance and Rayleigh interference optics, depending on protein concentration. A series of scans were recorded at fixed time intervals over a period of 110 minutes to eight hours. The results were analysed by g(τ) analysis using the program DCDT + version 1.13 by John Philo.

**Fluorescence spectra**

Fluorescence emission spectra of each protein were measured using a Photon Technology International fluorimeter (Ford, West Sussex, UK). For spectra of native and denatured Im7* species, proteins were dissolved in buffer A or buffer A containing 8 M urea, respectively, at a known concentration of protein (approximately 20 μM). Excitation and emission slits were set to 2 nm and 3 nm, respectively. Each spectrum was recorded from 270 nm to 450 nm in 1 nm increments, integrating for one second at each wavelength, exciting fluorescence at 280 nm. Spectra of all the denatured states were assumed to have the same maximum intensity at 350 nm. Spectra of each protein in the absence of urea were then normalised to the intensity of their respective denatured states. In this manner, a direct comparison of the fluorescence intensity of the variants with that of Im7* can be made.

**Circular dichroism**

Spectra were acquired on a Jasco J715 CD spectropolarimeter (Great Dunmow, Essex, UK). For measurements in the far-UV (195 nm to 260 nm) the CD signal was recorded on a 1 mm path-length cell using a protein concentration of 0.6 mg/ml. For measurement in the near-UV (250 nm to 350 nm) spectra were recorded at 0.6 mg/ml, using a 1 cm path-length cell. All spectra were recorded using a 1 nm bandwidth, 1 nm resolution, 20 nm/minute scan speed and a response time of eight seconds, averaging eight scans. For the near-UV, a 1 nm bandwidth, 0.5 nm resolution, 10 nm/minute scan speed and response time of four seconds were used, averaging 32 scans and each sample was corrected for concentration using the experimentally determined extinction coefficients (Table 1). Helical content was estimated from the measured mean residue molar ellipticity (MRE) using equation (1). The scaling factor 50 is included as 50% of the residues in wild-type Im7* are in a helical conformation, as defined in the PDB entry 1AYI.28

\[
\% \text{helical content} = \frac{MRE_{\text{variant}}}{MRE_{\text{Im7}*}} \times 50
\]

**Equilibrium denaturation curves**

Measurements were made using a Photon Technology International fluorimeter, with a 1 cm path-length cuvette, and a Jasco J715 CD spectropolarimeter, with a 1 mm path-length cuvette. Solutions contained 0–8 M urea in 0.2 M increments in buffer A. The final protein concentration was 0.2 mg/ml and all samples were incubated overnight at 10 °C prior to their measurement. Fluorescence was excited at 280 nm and emission at 350 nm was measured and averaged for one minute. The CD signal at 220 nm was measured for one minute and averaged using a response time of eight seconds and a bandwidth of 1 nm. These data were fit to a model describing a two-state transition:

\[
\text{Signal} = \frac{(a(\text{urea}) + b) \exp(\Delta G_{\text{UF}} - M_M(\text{urea}))/RT + (c(\text{urea}) + d)}{(1 + \exp(\Delta G_{\text{UF}} - M_M(\text{urea}))/RT)}
\]

where \(\Delta G_{\text{UF}}\) (kJ mol\(^{-1}\)) is the equilibrium stability, and \(M_M\) (kJ mol\(^{-1}\) M\(^{-1}\)) is the equilibrium m-value, \(a\) and \(c\) represent the denaturant-dependence of the folded and unfolded fluorescence intensities, respectively, and \(b\) and \(d\) are the fluorescence intensities of the folded and unfolded states, respectively, in the absence of denaturant. Errors quoted are the regression error.

**Stern–Volmer analysis**

Samples of each protein were prepared in buffer A containing 0–0.5 M KI in 0.1 M increments. The concentration of protein was approximately 20 μM and the total ionic strength was kept constant by including an appropriate concentration of KCl. Tryptophan fluorescence emission was measured as described above, except excitation was at 295 nm to excite the tryptophan fluorophore specifically. The data were fit to:

\[
\frac{F_0}{F} = 1 + K_{SV}[\Gamma]
\]

where \(F_0\) is the fluorescence emission in the absence of quencher, \(F\) is the fluorescence emission in the presence of quencher, \(K_{SV}\) is the Stern–Volmer constant and [\(\Gamma\)] is the concentration of iodide ions. Quenching of 20 μM N-acetyl-tryptophanamide was included as a control.

**ANS binding**

Fluorescence emission spectra of 1-anilino-8-naphthalene sulphonate (ANS; 250 μM final concentration) in buffer A in the presence and in the absence of 1 mM protein were measured between 400 nm and 600 nm, with an excitation wavelength of 389 nm. Excitation and emission slit-widths were 5 nm. The spectrum of ANS in buffer A alone was subtracted from spectra of ANS containing protein, and all spectra were normalised to that of ANS in the presence of 1 mM native wild-type Im7*.

**Kinetic analysis**

The refolding and unfolding kinetics of wild-type Im7* were measured in buffer A using an Applied Photophysics SX18.MV stopped-flow fluorimeter with the temperature held at 10 °C, as described.22 Data were fitted to an on-pathway three-state model, as described.12 Errors on the kinetic parameters were calculated using a combination of Monte Carlo simulation and weighted non-linear, least-squares analysis as described.27 These errors were propagated mathematically.

**1D 1H NMR**

15N-decoupled 1D 1H NMR spectra were recorded at 10 °C on a Varian Inova 500 MHz spectrometer using the shaka1dpresat pulse sequence,28 averaging 256 transients. Samples contained approximately 0.5 mM protein dissolved in buffer A containing 10% 2H2O at pH 7.0. Spectra were processed and plotted using the NMRPipe suite of programs,44 and required zero-order phase
correction. Pulsed-field gradient experiments were performed at 10 °C on a Varian Inova 600 MHz spectrometer in order to determine the radius of hydration, using the pulse sequence water_sLED_fm_v2_500. Samples contained approximately 0.5 mM protein dissolved in buffer A containing 10% (v/v) D2O and approximately 0.05% (v/v) 1,4-dioxane at pH 7.0. Each spectrum comprised 256 scans. Typically, 13 gradient strengths were used, ranging from 1.85–46.25 G cm⁻¹. The water signal was suppressed with a watergate sequence. The FIDs were processed using scripts kindly provided by Dr I. Guijarro (Institut Pasteur, France) in VNMR (Varian Inc., USA). The integrated areas for the 1,4-dioxane peak were normalised and fit according to:

\[
\text{Signal} = S \exp^{-\frac{m_1}{2}}\]

where \( S \) is the signal intensity, \( m \) is the rate of signal decay and \( g \) is the gradient strength. The integrated area for all protein peaks in the aliphatic and methyl region were normalised and fitted to equation (5), which contains a second term to allow for a small molecule contaminant present in all of the samples:

\[
\text{Signal} = S_1\exp^{-\frac{m_1}{2}} + S_2\exp^{-\frac{m_2}{2}}
\]

where \( S_1 \) and \( S_2 \) are the signal intensities for the small molecule and the protein, respectively, and \( m_1 \) and \( m_2 \) are the decay rates for the small molecule and the protein, respectively.

By comparing the decay rate of 1,4-dioxane, with a known \( K_{H} \), the \( K_{H} \) for the protein can be calculated according to:

\[
K_{H}^{\text{protein}} = \frac{m_1^{1,4-\text{dioxane}}}{m_1^{\text{H}}} \cdot \frac{r^{1,4-\text{dioxane}}}{r^{\text{H}}}
\]

where \( K_{H}^{1,4-\text{dioxane}} \) is assumed to be 2.12 Å. 46

Acknowledgements

We thank Claire Friel and Stanislaw Gorski for many helpful discussions, and Eva Sanchez-Cobos, Stuart Knowling and Victoria Morton for commenting on drafts of the manuscript. We thank Keith Ainley for preparing growth media, Alison Ashcroft for ESI-MS analysis, and Andy Baron and Thomas Jahn for performing the analytical ultracentrifugation. We acknowledge Arnout Kalverda, Sara Whittaker, Ruth Boetzel and Geoff Moore for help in both acquiring and interpreting the NMR spectra. G.R.S. and A.P.C. were supported by the BBSRC. S.E.R. is a BBSRC Professio- nal Fellow. G.R.S., A.P.C. and S.E.R. are members of the Astbury Centre for Structural Molecular Biology, which is part of the North of England Structural Biology Centre and was funded by the BBSRC.

References

20. Jemth, P., Gianni, S., Day, R., Li, B., Johnson, C. M.,


*Edited by A. R. Fersht*

(Received 27 February 2004; received in revised form 20 May 2004; accepted 26 May 2004)