nature structural & molecular biology

The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints

Claire T Friel[1,2,6], D Alastair Smith[1,3,6], Michele Vendruscolo[4], Joerg Gsponer[5] & Sheena E Radford[1,2]

Many proteins reach their native state through pathways involving the presence of folding intermediates. It is not clear whether this type of folding landscape results from insufficient evolutionary pressure to optimize folding efficiency, or arises from a conflict between functional and folding constraints. Here, using protein-engineering, ultra-rapid mixing and stopped-flow experiments combined with restrained molecular dynamics simulations, we characterize the transition state for the formation of the intermediate populated during the folding of the bacterial immunity protein, Im7, and the subsequent molecular steps leading to the native state. The results provide a comprehensive view of the folding process of this small protein. An analysis of the contributions of native and non-native interactions at different stages of folding reveals how the complexity of the folding landscape arises from concomitant evolutionary pressures for function and folding efficiency.

To fold into their native structures, proteins must undergo a series of extensive conformational changes. Nevertheless, for most small proteins, the experimental manifestations of the folding reaction are simple[1,2]. Theoretical studies suggest that this results from a funnel-like global organization of the landscape of accessible protein conformations[3,4], which is the outcome of an evolutionary selection for sequences that minimize the conflict between different interactions, and leads smoothly toward the native conformation[5]. Indeed, proteins designed computationally or artificially evolved from random libraries, which lack an evolutionary history, fold less cooperatively than their similarly sized naturally occurring counterparts[6–8]. Further selection for landscapes that promote cooperative folding may result from evolutionary pressure against sequences that promote aggregation[9]. Accordingly, sequences that suppress the formation of folding intermediates, thereby disfavoring aggregation, have been identified[10]. Such 'negative design' seems to be essential in the light of findings that non-native structures can have an important role in the formation of amyloid fibrils[11].

Despite evidence of evolutionary selection against long-lived folding intermediates, partially folded states have been identified in the folding of many single-domain proteins[12]. It is unclear whether this reflects insufficient evolutionary selection of a sequence optimal for folding or results from functional constraints on the evolution of the amino acid sequence. The colicin immunity binding proteins of *Escherichia coli*, a family of small four-helix proteins with close sequence similarity (50%)[13], provides an ideal system for investigating this question. One family member, Im7, has been shown to fold via a complex folding landscape involving a highly populated on-pathway intermediate[14–17]. By contrast, for the Im7 homolog Im9, an intermediate becomes detectably populated only under acidic conditions or through targeted substitution of residues to increase hydrophobicity and strengthen non-native interactions during folding[18,19]. Despite the differences in their kinetic folding mechanisms, Im7 and Im9 perform the same function: both bind and inhibit their cognate colicin toxins (E7 and E9 for Im7 and Im9, respectively) with diffusion rate–limited binding and dissociation constants of $\sim 10^{-14}$ M[20]. Thus, the evolutionary pressure for the selection of binding-competent sequences of these proteins is crucial for the survival of the organism[21].

The Im7 folding landscape has been characterized using protein engineering, hydrogen exchange and molecular dynamics simulations[14,17,22,23]. The results revealed that the rate-limiting transition state (TS2) and the preceding intermediate (I) contain three of the four native helices (I, II and IV) (**Fig. 1a**), with the intermediate being stabilized by both native and non-native interactions. Despite this information, it remained unclear why the folding landscape of Im7 involves an intermediate that is conserved within this family of proteins, and which interactions are responsible for its formation. Addressing these questions requires detailed structural insights into the early events in folding that are responsible for the formation of the intermediate state. By combining ultra-rapid mixing with stopped-flow measurements of the folding of Im7 and 16 site-specific variants followed by analysis of the resulting Φ-values using restrained molecular dynamics simulations, we provide an all-atom description of

[1]Astbury Centre for Structural Molecular Biology, [2]Institute of Molecular and Cellular Biology, University of Leeds, Mount Preston Street, Leeds LS2 9JT, UK. [3]School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK. [4]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB1 1EW, UK. [5]Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. [6]Present addresses: Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany (C.T.F.); Avacta Group Plc, York Biocentre, Innovation Way, York YO10 5NY, UK (D.A.S.). Correspondence should be addressed to S.E.R. (s.e.radford@leeds.ac.uk) or J.G. (jag70@cam.ac.uk).

Received 15 May 2008; accepted 21 January 2009; published online 1 March 2009; doi:10.1038/nsmb.1562

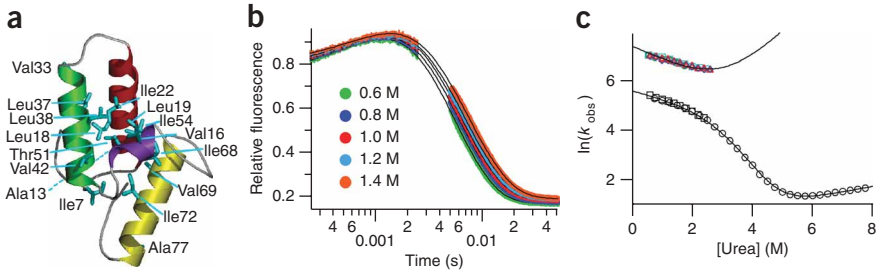

Figure 1 Native structure of Im7 and representative kinetic traces. (**a**) Structure of Im7 (PDB 1AYI) showing residues that were mutated in this study; Ala13 is at the back of the molecule in this orientation. Helices are colored red (helix I, residues 11–27), green (helix II, residues 31–46), magenta (helix III, residues 51–56) and yellow (helix IV, residues 64–79). (**b**) Refolding transients for wild-type Im7 at various concentrations of urea; black lines show the best fit of the data to a double exponential. (**c**) Denaturant dependence of the natural logarithm of the observed rate constants for the folding and unfolding of wild-type Im7. The colored symbols and black squares represent rate constants obtained using the combination of continuous and stopped-flow results. The results of four replicate experiments are shown for the faster rate constants to highlight the reproducibility of the method. Black circles represent rate constants obtained using stopped-flow fluorescence alone.

the entire folding landscape of this protein, including the early transition state for intermediate formation (TS1). In turn, we show how functional constraints have a central role in determining the ruggedness of the folding landscape of this family of proteins.

RESULTS

The folding kinetics of Im7 and its variants

To provide an accurate molecular description of the folding mechanism of Im7, including the early stages during which its on-pathway intermediate is formed, we analyzed the folding and unfolding kinetics of the wild-type protein and 16 site-specific variants (**Fig. 1a**). At low urea concentrations, where folding is three state, we analyzed the refolding kinetics of wild-type Im7 and each variant using ultra-rapid, continuous-flow mixing, monitored using the fluorescence of the single tryptophan, Trp75, allowing refolding to be measured between ~200 μs and ~2.5 ms (**Supplementary Fig. 1** online). We then used the stopped-flow fluorescence measurements to complete the transients (**Fig. 1b**). The resulting data were fitted globally to a double-exponential function (see Methods and **Supplementary Methods** online). At higher urea concentrations, in which the intermediate is no longer populated, we measured the refolding kinetics using stopped-flow fluorescence alone. The data were combined with measurements of the rates of unfolding to complete the chevron plot (**Fig. 1c**). Together with the initial and end-point fluorescence signals measured using stopped-flow fluorescence, all data were fitted globally to the analytical solution of the model (Equation (1)).

$$U \underset{k_{iu}}{\overset{k_{ui}}{\rightleftharpoons}} I \underset{k_{ni}}{\overset{k_{in}}{\rightleftharpoons}} N \qquad \text{Equation (1)}$$

U, I and N represent the denatured, intermediate and native states, respectively, and k_{xy} is the microscopic rate constant for the conversion of x to y. The data for the wild-type protein and each variant are described well by this model.

This fitting procedure results in the accurate determination of all four microscopic rate constants (k_{ui}, k_{iu}, k_{in} and k_{ni}) and their respective denaturant dependencies (m_{ui}, m_{iu}, m_{in} and m_{ni}). No assumptions are required about the rate of formation of the intermediate, allowing a more accurate determination of the rate constants and the resulting Φ-values for the intermediate and TS2 than was possible hitherto[14], as well as revealing the first insights into the effect of mutations on TS1. The data obtained for wild-type Im7 fitted in this manner reveals that the protein folds rapidly to the intermediate (k_{ui} ~1,600 s^{-1}) through a transition state (TS1) with a β_T value of 0.24 ($\beta_T^{TS1} = m_{ui}/ m_{ui} + m_{iu} + m_{in} + m_{ni}$), consistent with previous experiments on Im7 lacking a hexahistidine (His_6) tag[15]. Folding progresses through a highly populated intermediate ($\Delta G_{ui} = -11.7$ kJ mol^{-1}) with a β_T value of 0.74 and a subsequent rate-determining transition state (TS2) with a β_T value of 0.9 (**Supplementary Tables 1** and **2** online).

Structure of TS1

To obtain information about the extent of secondary-structure formation in TS1, Ala13 and Ala77, which are solvent exposed in the native state and located in helices I and IV, respectively (**Fig. 1a**), were truncated to glycine, and the folding and unfolding kinetics of the variants were measured as described above. These substitutions have only a small effect on k_{ui} (k_{ui} = 1,150 s^{-1} and 1,648 s^{-1} for the A13G and A77G mutants, respectively, compared with 1,574 s^{-1} for the wild-type protein) (**Fig. 2** and **Supplementary Table 1**) but decrease the stability of I and N (**Supplementary Table 2**), indicating that these residues are not well ordered in TS1 but form a helical structure in I and TS2. Accordingly, the Φ-values calculated for TS1, I and TS2 are 0.29 ± 0.07, 1.33 ± 0.20 and 1.39 ± 0.09, respectively, for A13G and –0.02 ± 0.06, 0.82 ± 0.22 and 0.81 ± 0.10, respectively, for A77G. The

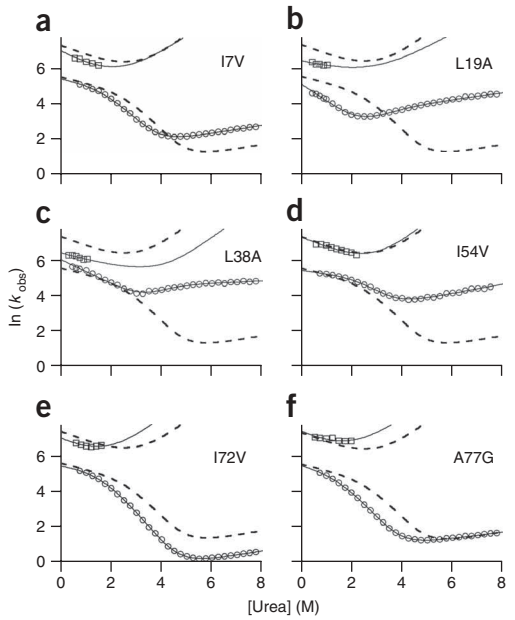

Figure 2 Dependence of the natural logarithm of the observed rate constants for folding/unfolding on the concentration of urea for selected Im7 variants: I7V (**a**), L19A (**b**), L38A (**c**), I54V (**d**), I72V (**e**) and A77G (**f**). Solid lines show the best fit of the data to a three-state model with an on-pathway intermediate. The dashed line shows the best fit to the data for wild-type Im7 for comparison.

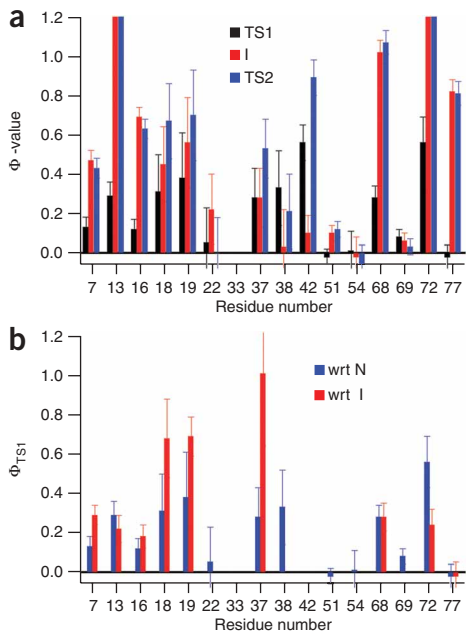

Figure 3 Calculated Φ-values. (**a**) Φ-values with respect to $\Delta\Delta G_{un}$ for TS1, I and TS2 (black, red and blue, respectively). (**b**) Φ-values for TS1 with respect to (wrt) $\Delta\Delta G_{ui}$ (red) or $\Delta\Delta G_{un}$ (blue). Error bars indicate 1 s.d.

substitution V33A increases the helical propensity in the N-terminal region of helix II (**Fig. 1a**). Although accurate Φ-values could not be determined for this substitution because $\Delta\Delta G_{un}$ is small (**Supplementary Table 1**), this substitution also has little effect on k_{ui}. By contrast with I and TS2, which contain native-like helices I, II and IV[14,17], helical structure was not detected in the vicinity of the sites investigated in TS1.

To study the importance of helix III in the folding of Im7, we substituted Thr51 and Ile54 with serine and valine, respectively (**Fig. 1a**). Despite their position in the core of native Im7, truncation of these residues did not affect k_{ui} or k_{in} but increased k_{ni} by approximately 5-fold and 15-fold for T51S and I54A, respectively (**Fig. 2** and **Supplementary Table 1**). These substitutions result in Φ-values for TS1, I and TS2 of –0.02 ± 0.04, 0.10 ± 0.11 and 0.12 ± 0.15, respectively, for T51S and 0.01 ± 0.10, –0.02 ± 0.13 and –0.06 ± 0.12, respectively, for I54V (**Fig. 3a** and **Supplementary Table 1**), indicating that these residues make few stabilizing contacts until the native state is formed.

Key residues in folding to the on-pathway intermediate

Further to the substitutions described above, 11 buried or partially buried hydrophobic residues were truncated: 10 from helices I, II and IV, plus 1 (Ile7) that lies in the N-terminal region of the protein and forms stabilizing interactions in the hydrophobic core of the native structure (**Fig. 1a**). Four of these substitutions (I7V, V16A, I22V and V69A) reduced k_{ui} by <500 s^{-1} (**Fig. 2** and **Supplementary Table 1**). A second group of substitutions, V42A (helix II), I68V and I72V (helix IV), reduces k_{ui} by >500 s^{-1}. The most dramatic changes in k_{ui} were observed for a third group that includes L18A, L19A (helix I), L37A and L38A (helix II) for which k_{ui} is reduced by $>1{,}000$ s^{-1} (**Fig. 2** and **Supplementary Table 1**). The Φ_{TS1}-values determined for 9 of the 11 variants were low (0.1–0.4) and, in general, markedly smaller than those for the same residues in I and TS2 (**Fig. 3a** and **Supplementary Table 1**). Overall, therefore, side chain packing is less well ordered in TS1 compared with I and TS2, consistent with the low β_T value of this state.

Notably, substitutions that result in the most dramatic changes in k_{ui} do not give rise to the largest values of Φ_{TS1}. For example, for I72V, $\Phi_{TS1} = 0.56 \pm 0.13$, although k_{ui} is reduced by only $\sim$500 s^{-1}. By contrast, for L19A k_{ui} is reduced by $\sim$1,400 s^{-1}, to a value of only 184 s^{-1} (**Fig. 2**), yet the resultant Φ_{TS1} value is only 0.38 ± 0.23 (**Supplementary Table 1**). Consideration of k_{ui} and Φ_{TS1} thus provides contrasting views of the relative importance of different residues in stabilizing TS1. These results question which ground state should best be used as the reference for determination of Φ_{TS1}. The calculation of Φ-values relative to $\Delta\Delta G_{un}$ allows direct comparison of Φ_{TS1}, Φ_I and Φ_{TS2} (**Fig. 3a**). However, the Im7 intermediate has previously been shown to be stabilized by both native and non-native contacts[14]. Therefore, for some variants, I and N respond differently to mutation, with the result that $\Delta\Delta G_{ui}$ and $\Delta\Delta G_{un}$ are not linearly correlated over all residues, in contrast to proteins that fold by progressive consolidation of native contacts[24]. Indeed, for I72V, $\Delta\Delta G_{ui}$ exceeds $\Delta\Delta G_{un}$, whereas for L19A, $\Delta\Delta G_{ui} \ll \Delta\Delta G_{un}$ (**Supplementary Table 2**). When Φ_{TS1} values are calculated using $\Delta\Delta G_{ui}$ as the normalization factor (**Supplementary Table 2**), a different picture emerges (**Fig. 3b**). L18A, L19A and L37A now have Φ_{TS1} values of 0.7–1.0, highlighting the importance of these residues in stabilizing TS1. Notably, each of these variants gives rise to a chevron plot with pronounced curvature in the unfolding branch, a feature that becomes apparent when $k_{ui} < k_{in}$. This is also seen for L38A (k_{ui} 440 s^{-1}) (**Fig. 2**), suggesting that this residue is also important in stabilizing TS1.

The experimental results suggest that the docking of the side chains of Leu18, Leu19 and Leu37 (and possibly also Leu38) is the first key event in the folding of Im7. The TS1 ensemble is stabilized by numerous weak hydrophobic interactions that are presumably variable between members of the ensemble, involving residues both local to and distant from these sites. Notably, for side chains that form the native helix I (Val16, Leu18 and Leu19), $\Phi_{TS1} < \Phi_I < \Phi_{TS2}$ (**Fig. 3a**), as might be expected given the increasing β_T value (0.2, 0.7 and 0.9 for TS1, I and TS2, respectively (**Supplementary Table 2**). However, for residues that ultimately form the native helix II (Leu37, Leu38 and Val42), such a pattern is less clear. The data reinforce the view that the folding of Im7 does not progress by a straightforward consolidation of native contacts[14], even in the earliest stages in which the on-pathway intermediate is formed from TS1.

The folding of Im7 in atomistic detail

To elucidate which residue-residue interactions are involved in different stages of folding, we calculated the ensembles of structures representing TS1 and TS2 using the newly derived Φ_{TS1} and Φ_{TS2} values described above (calculated relative to ΔG_{un}) as restraints[25] (see Methods). Equilibrium hydrogen-exchange protection factors have previously been used to model the intermediate ensemble[22]. The validity of restrained molecular dynamics simulations for generating representative structural ensembles of transition states of proteins has been demonstrated[22,26,27] and shown to be consistent with experimentally measured quantities that were not used in the simulations[22] or used to design mutants with prescribed folding properties[27].

Ensembles representative of TS1, I and TS2 determined by restrained molecular dynamics simulations are shown in **Figure 4a,b**. These ensembles are fully consistent with the experimental Φ-values. This result is expected for TS1 and TS2, because the Φ-values were used as a source of structural information, but it is notable for the intermediate

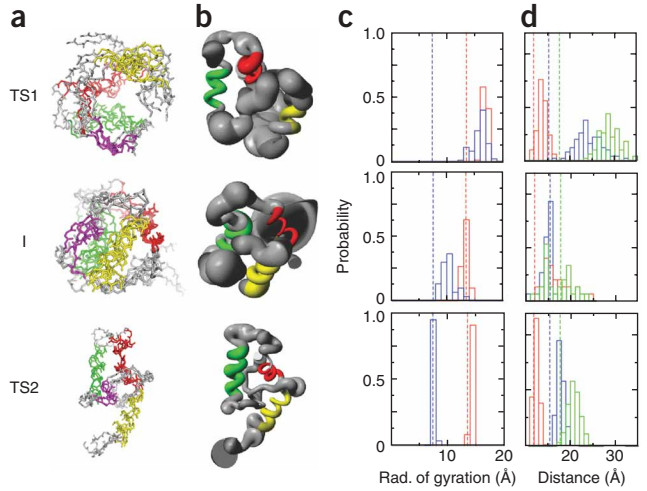

Figure 4 Comparison of structural properties of TS1, I and TS2. (**a**) Representative members of the three ensembles. The segments forming the four helices in N are shown in red (I), green (II), magenta (III) and yellow (IV). (**b**) Diagram showing the heterogeneity of different ensembles; the tube diameter is scaled by the average displacement to the mean structure. (**c**) Distribution of the radius (rad.) of gyration of the entire polypeptide chain (red) and the core residues, 15, 18, 19, 22, 34, 38, 41, 44, 53, 54, 55, 68 and 72 (blue). The dashed lines indicate the radius of gyration of each set of residues calculated for N. (**d**) Distribution of the distances between the centers of mass of different helices: helices I–II, red; helices II–IV, green; helices I–IV, blue. The dashed lines indicate the distances calculated for N.

state, because in this case equilibrium hydrogen-exchange protection factors—but not Φ-values—were used as restraints[22] (**Supplementary Fig. 2** online). To assess the quality of the ensembles generated, we back-calculated the Φ-values using FoldX[28]. In contrast to the native contact approximation used to restrain Φ-values during the structure calculations (see Methods), the free energy–based back-calculation of Φ-values using FoldX is indifferent to whether contacts are native or non-native[26]. Notably, as some experimental Φ_{TS1} values, especially those of L18A, L19A and L37A, depend crucially on the reference state used for their determination (see above), a correct prediction of these Φ_{TS1} values relative to $\Delta\Delta G_{ui}$ and $\Delta\Delta G_{un}$ computed over the ensembles generated provides a stringent test for the quality of the TS1 and I ensembles. Correlations of 0.79, 0.74 and 0.73 between experimental and back-calculated Φ-values for TS1, I and TS2 (with respect to $\Delta\Delta G_{un}$), respectively, and 0.76 for TS1 (with respect to $\Delta\Delta G_{ui}$), highlight the quality of all ensembles (**Supplementary Fig. 2b**). An additional validation of the structures results from the correct prediction of the experimentally determined β_T values (**Supplementary Methods**).

To provide further controls, we determined ensembles for TS1 and TS2 following the same protocol but using either reshuffled Φ-values or a restricted set of eight Φ-values (**Supplementary Fig. 3** online). We then used the new ensembles to back-calculate Φ-values using FoldX. The use of reshuffled Φ-values generated putative structures of TS1 and TS2 that differ markedly from those derived from the experimental Φ-values (compare **Fig. 4** and **Supplementary Fig. 3a–e**). The control carried out using a reduced set of Φ-values resulted in structures for which FoldX predicts the Φ_{TS1} and Φ_{TS2} less well than the ensembles determined using the full set of Φ-vales (**Supplementary Fig. 3f–j**). These control calculations demonstrate the necessity of using an extended set of Φ-values to produce ensembles accurate enough to enable an analysis of the interactions made at different stages of folding in all-atom detail.

Molecular description of TS1

Analysis of the ensemble of structures representing TS1 showed that this species is almost devoid of ordered secondary structure, a characteristic common to all the members of this ensemble (**Fig. 4a,b**). The large majority of residues remain solvent exposed in TS1 (**Supplementary Fig. 4a** online), consistent with its expanded nature (β_T), large radius of gyration (**Fig. 4c**) and lack of a stable hydrophobic core (**Fig. 5**). This conclusion is supported by the large radius of gyration of residues that comprise the native hydrophobic core of TS1 (**Fig. 4c**). Moreover, the helix-forming regions of the protein sequence are more than 20 Å apart in TS1, except for the nascent helices I and II, which contact each other via long-range side chain interactions between residues 16–20 and 37–42 (**Figs. 4d** and **5**). The presence of these side chain contacts in TS1 is consistent with the high Φ-values determined experimentally for residues 18, 19 and 37 (**Fig. 3b**). Although these residues form some native-like contacts in this early transition state, many interactions are non-native (**Fig. 5**).

The U-to-I transition through TS1

Knowledge of the structure of TS1 allows the molecular rearrangements associated with the transition from TS1 to the intermediate to

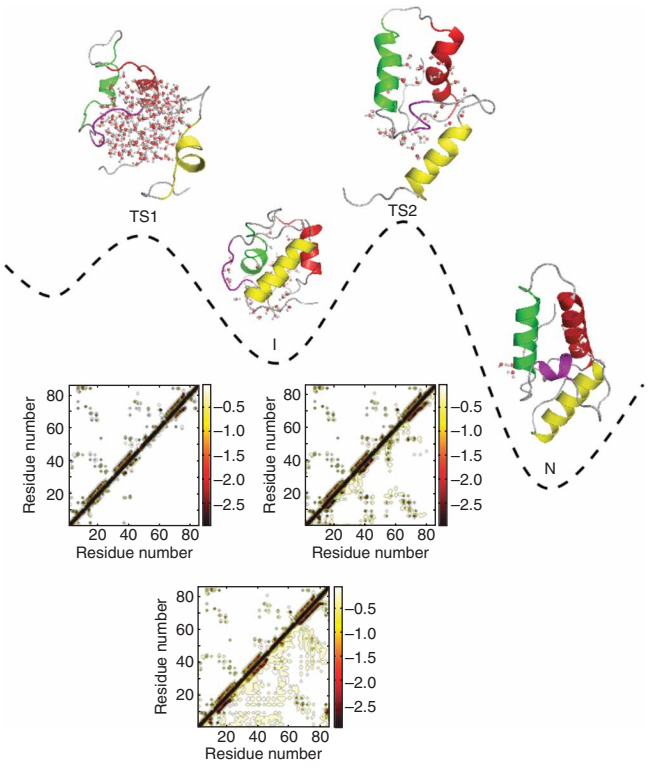

Figure 5 Schematic illustration of the folding landscape of Im7. Ribbon diagram representations of selected cluster centers of TS1, I, TS2 and N are shown. The helix-forming segments are colored red (helix I), green (helix II), magenta (helix III) and yellow (helix IV). Water molecules within a 12-Å sphere around the centers of mass of TS1, I, TS2 and N are shown in ball-and-stick representation. Below, the average interaction-energy maps of TS1, I and TS2 (below the diagonal) are compared with that of N (above the diagonal). The scale on the right is in kcal mol^{-1}.

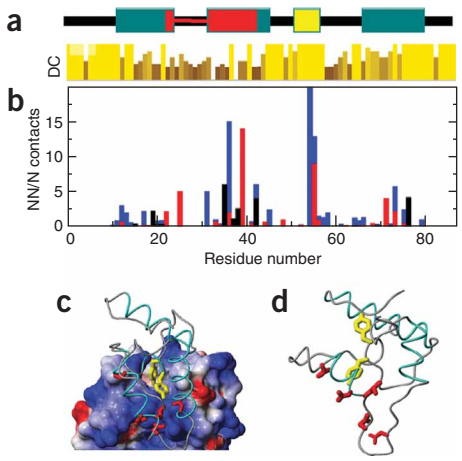

Figure 6 Non-nativeness during folding versus functionality. (**a**) Degree of conservation (DC), reflecting the conservation of physicochemical properties in each column of the alignment of Im2, Im7, Im8 and Im9 (**Supplementary Fig. 5**). The positions of the native helices are shown above; segments that provide binding affinity and specificity are highlighted in yellow and red, respectively. (**b**) Fraction of non-native (NN) versus native (N) atomic contacts per residue in TS1 (black), I (red) and TS2 (blue). Only contacts present in more than 25% of the structures in each ensemble are considered. (**c**,**d**) Comparison of the relative positions of Tyr55 and Tyr56 (in yellow) and of the specificity-determining residues Asn26, Asp31, Leu34, Asp35 and Val42 (in red) in the native complex of Im7 with colicin E7 (**c**) and in a representative member of the I state ensemble (**d**).

be discerned. The results establish that the TS1-to-I transition is a dramatic step in the folding of Im7, which is characterized by hydrophobic collapse and the expulsion of water from the core (**Fig. 5** and **Supplementary Fig. 4a**). During this transition, Im7 adopts a radius of gyration (computed over all residues) that is close to that of the native state (**Fig. 4c**), and native-like secondary structure forms in the regions of the sequence-defining helices I, II and IV (**Figs. 4a**,**b** and **5**). Whereas the sequences spanning the native helices I and II are already in close contact in TS1, crossing of the first transition-state barrier results in the additional docking of helix IV and the formation of the three-helical intermediate. The non-native proximity of helices II and IV in members of the intermediate ensemble (**Fig. 4d**) and a radius of gyration of the core residues that is larger than that of the native state (**Fig. 4c**) are indicative of suboptimal packing of side chains in the intermediate. In addition to non-native contacts already formed in TS1 between residues in the native helices I and II, the engagement of helix II with residues of helix IV provides additional non-native contacts that stabilize the native-like topology of the intermediate state (**Figs. 5** and **6a**,**b**). The fact that helix III does not rapidly dock onto the three-helical structure, allowing folding to proceed directly to the native state without delaying in a stable intermediate, suggests that the non-native interactions prove an impediment to rapid folding.

The I-to-N transition through TS2

Subtle rearrangements of the core take place in the folding step from I to TS2, which results in a native-like positioning of helices I, II and IV and a native-like radius of gyration for core hydrophobic residues (**Fig. 4a–d**). The rate-limiting step in folding occurs at TS2 and involves the formation of the binding site for residues that dock onto the already-formed three-helix bundle in order for helix III to form (this sequence has no propensity to exist as a helix in the absence of tertiary interactions[29]). Despite the overall native-like topology of TS2, many residues in helices II and III, Tyr55 in particular, still form more non-native than native contacts within this ensemble (**Figs. 5** and **6**).

To determine more precisely the nature of the reorganizational events leading to and from the intermediate state, we analyzed the TS1, I and TS2 ensembles in more detail, focusing on Phe41 (helix II) and Tyr55 (helix III) as representatives of residues that form non-native interactions and may interfere with the docking and formation of helix III during folding. These residues were chosen because Phe41 forms a crucial part of the native hydrophobic core and shows clear evidence for the formation of non-native contacts during folding both experimentally[14] and by simulation (**Fig. 6**). Tyr55 is partially solvent exposed in native Im7 and is predicted to form non-native contacts throughout the folding process. In addition, we monitored the interactions of Trp75 (helix IV), because both experimental and simulation data suggest that this residue is more buried in I than in any other state[22,30] (**Supplementary Fig. 4b**).

The number of side chain–side chain interactions between Phe41, Tyr55 and Trp75 and all other residues in TS1, I, TS2 and N are shown in **Figure 7a**. These profiles reveal that Trp75 makes many non-native interactions with residues in the regions of residues 37–45 (helix II) and 51–56 (helix III) in the intermediate. Moreover, inspection of representative structures from each ensemble (**Fig. 7b**) suggests that the non-native interactions formed between Trp75 and side chains of residues in helix II hinder residues in helix III (represented here by Tyr55) from adopting their native position in which these residues dock against buried side chains of residues in helices II and IV. To investigate this mechanism further, we determined the distribution of distances between Phe41 (helix II) and either Ile54 (helix III) or Trp75 (helix IV) for I, TS2 and N (**Supplementary Fig. 4c** online). Whereas Phe41 is close to Ile54 in TS2 and N, this is not the case for the

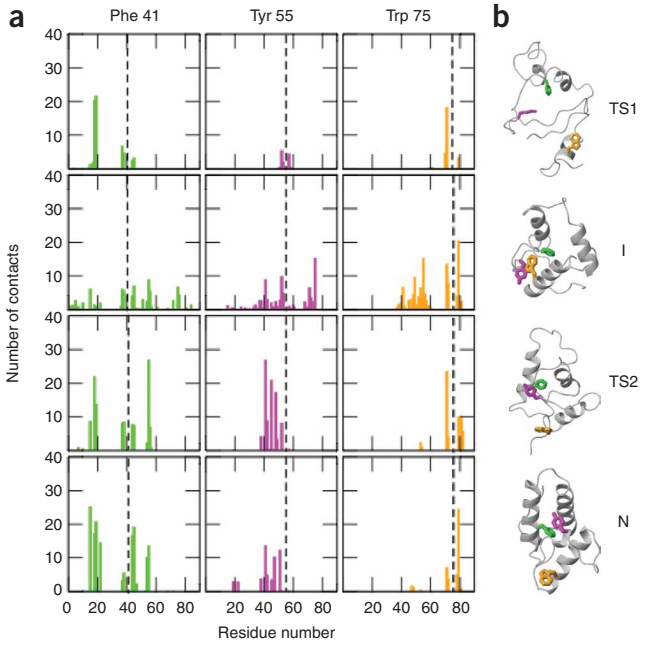

Figure 7 Interaction patterns of selected residues in TS1, I, TS2 and N. (**a**) Number of atomic contacts (native and non-native) formed between residues Phe41, Tyr55, Trp75 and the other residues of Im7 in TS1, I, TS2 and N. (**b**) Structures showing the positions of residues Phe41 (green), Tyr55 (magenta) and Trp75 (orange) in each ensemble.

intermediate. In fact, in many conformations of the intermediate ensemble, Trp75 is closer to Phe41 than is Ile54. These results confirm that residues in the C-terminal region of helix II form substantial non-native interactions with Trp75 in the intermediate, thereby inhibiting helix III from finding its native interaction partners and temporarily trapping Im7 in the intermediate state.

DISCUSSION

The folding landscape of Im7 is unusually rugged

Effective folding of proteins to their native states in the cellular environment is essential for their function. Furthermore, the avoidance of long-lived, partially folded states helps prevent potentially harmful misfolding and aggregation[10,31]. In this context, the folding landscape of Im7 is unusual, as this small, single-domain protein folds with an unexpectedly complex energy landscape. Here, by combining detailed and complete kinetic analysis of the folding of Im7 with molecular dynamics simulations, we provide detailed molecular insights into the entire folding landscape from the earliest (least-compact) transition state examined to date ($\beta_T = 0.2$), through the three-helix intermediate ($\beta_T = 0.7$), to the highly native-like, rate-limiting transition state ($\beta_T = 0.9$). The results reveal that the transition state for intermediate formation is expanded, containing long-range stabilizing contacts between residues in regions corresponding to the native helices I and II that are supported by further, weak interactions with residues in helix IV. These interactions are not yet sufficient to establish a stable native-like topology. Substantial further collapse and mispacking of hydrophobic residues (in particular aromatic side chains) occurs as the intermediate state forms. Although native and non-native interactions stabilize TS1, further non-native interactions are formed in the transition from TS1 to I. These interactions occlude the binding site required for the formation of helix III, but establish the formation of a native-like topology in which the fully formed helices I, II and IV remain misaligned. The reorganization of the packing of helices I, II and IV to establish the helix III binding site determines the rate-limiting step in the overall folding reaction for Im7, and presumably for the rest of the family of immunity proteins. Rather than forming an increasing number of native contacts during folding, as is commonly found for small proteins[32], the sequence of Im7 is not optimized for efficient folding. Consistent with this finding, recent simulations of a coarse-grained representation of Im7 also indicate that frustrated interactions give rise to a rugged folding-energy landscape[33].

Functional constraints hinder folding efficiency

Many of the residues identified here as forming non-native contacts during the early stages of folding of Im7 lie in regions that have a vital role in the function of immunity proteins: the recognition and inactivation of colicin toxins[20] (**Fig. 6a–d** and **Supplementary Fig. 5** online). An initial docking of the conserved residues Tyr55 and Tyr56 in helix III onto the colicin surface anchors cognate and noncognate complexes. This is followed by exploration of the second docking site, involving primarily residues in helix II, whose binding free energy discriminates between cognate and noncognate pairs[34,35]. This so-called dual-recognition mechanism offers a selective advantage to the organism: maintenance of the sequence of helix III (>80% conserved over its six residues across four DNase-type immunity proteins) provides the capability for colicin inhibition required for survival of the organism, whereas changing motifs of charged and hydrophobic residues in helix II (<30% conserved over its 14 residues) allow for the evolution of specificity in partner recognition. The characteristics of the variable residues of helix II tailor the competition between native and non-native interactions, determining the degree to which an intermediate is populated during folding across the immunity protein family[19]. These functional constraints therefore not only result in the presence of an intermediate in folding, but also determine its structural and energetic features and rationalize why this species is evolutionarily conserved. The need to maintain and evolve function has thus influenced the selection of immunity-protein sequences resulting in a rugged landscape to the detriment of folding efficiency. Such a scenario has been proposed for the folding of other small proteins[36–38], suggesting that the evolutionary pressures for function and for folding can be conflicting and providing a rationale for the formation of folding intermediates in many single-domain proteins.

METHODS

Data collection and analysis. Im7 variants were created, expressed and purified as described[14,39]. Kinetic measurements were performed at 10 °C in 50 mM sodium phosphate buffer, pH 7.0, containing 0.4 M sodium sulfate using a custom-built continuous-flow instrument[40] and an Applied Photophysics SX18.MV stopped-flow instrument[39]. Final protein concentrations were ~20 μM for the continuous-flow and ~5–20 μM for the stopped-flow measurements. Data from the two instruments were fitted globally to a double-exponential function sharing both rate constants in IgorPro (Wavemetrics). To constrain the endpoint of the fit to the refolding transient obtained by continuous-flow mixing, the fluorescence signal was measured at equilibrium at each concentration of denaturant using premixed samples and the endpoint constrained to this value (**Supplementary Methods**). For denaturant concentrations in which folding is two state and for all unfolding experiments, a single observed rate constant was determined using stopped-flow measurements alone.

The two sets of observed rate constants determined for wild-type Im7 and its variants, and also the end-point and initial signals from the refolding traces measured using stopped-flow alone, were fitted, using the global fitting package in IgorPro (Wavemetrics), to the two roots of the analytical solution for an on-pathway, three-state model (**Supplementary Methods**). Φ-values for TS1, I and TS2 were then calculated using the microscopic rate constants determined (**Supplementary Methods**). Errors were propagated mathematically from the errors determined on the fit parameters.

Restrained molecular dynamics simulations. The CHARMM22 (ref. 41) force-field was used to carry out molecular dynamics simulations with Φ-value restraints[25] using an all-atom protein representation, the TIP3P water model and periodic boundary conditions[41]. All calculations used an atom-based truncation scheme with a list cut-off of 14 Å, a non-bond cut-off of 12 Å and the Lennard-Jones smoothing function initiated at 10 Å. Electrostatic and Lennard-Jones interactions were force switched. Molecular dynamics simulations used a 2-fs integration-time step and SHAKE of covalent bonds involving hydrogen atoms. For more detailed information see **Supplementary Methods**.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank C. Kleanthous and members of the Radford group for helpful discussions, S. Masca and I. Rodriguez-Mendieta for much help with the design and construction of the ultra-rapid mixing device and C. Gell for help with data analysis. C.T.F. was supported by the UK Biotechnology and Biological Sciences Research Council BBSRC (24/B17145), M.V. by the European Molecular Biology Organization, the Leverhulme Trust and the Royal Society, and J.G. by the UK Medical Research Council.

AUTHOR CONTRIBUTIONS

C.T.F. performed all experiments and their analysis and interpretation, and manuscript preparation; D.A.S. performed instrument development and analysis; M.V. performed molecular dynamics data interpretation and manuscript preparation; J.G. performed all molecular dynamics simulations and their analysis, interpretation and manuscript preparation; S.E.R. conceived of the project and prepared the manuscript.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Roder, H., Maki, K. & Cheng, H. Early events in protein folding explored by rapid mixing methods. *Chem. Rev.* **106**, 1836–1861 (2006).
2. Schuler, B. & Eaton, W.A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008).
3. Dill, K.A. & Chan, H.S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
4. Onuchic, J.N. & Wolynes, P.G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).
5. Bryngelson, J.D. & Wolynes, P.G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA* **84**, 7524–7528 (1987).
6. Hill, R.B. *et al. De novo* design of helical bundles as models for understanding protein folding and function. *Acc. Chem. Res.* **33**, 745–754 (2000).
7. Sauer, R.T. Protein folding from a combinatorial perspective. *Fold. Des.* **1**, R27–R30 (1996).
8. Watters, A.L. *et al.* The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* **128**, 613–624 (2007).
9. Monsellier, E. & Chiti, F. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.* **8**, 737–742 (2007).
10. Mitraki, A. *et al.* Global suppression of protein folding defects and inclusion body formation. *Science* **253**, 54–58 (1991).
11. Jahn, T.R. & Radford, S.E. Folding versus aggregation: polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **469**, 100–117 (2008).
12. Brockwell, D.J. & Radford, S.E. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* **17**, 30–37 (2007).
13. Dennis, C.A. *et al.* A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity-protein specificity. *Biochem. J.* **333**, 183–191 (1998).
14. Capaldi, A.P., Kleanthous, C. & Radford, S.E. Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.* **9**, 209–216 (2002).
15. Capaldi, A.P. *et al.* Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat. Struct. Biol.* **8**, 68–72 (2001).
16. Ferguson, N. *et al.* Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* **286**, 1597–1608 (1999).
17. Gorski, S.A. *et al.* Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7. *J. Mol. Biol.* **337**, 183–193 (2004).
18. Cranz-Mileva, S., Friel, C.T. & Radford, S.E. Helix stability and hydrophobicity in the folding mechanism of the bacterial immunity protein Im9. *Protein Eng. Des. Sel.* **18**, 41–50 (2005).
19. Friel, C.T., Beddard, G.S. & Radford, S.E. Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J. Mol. Biol.* **342**, 261–273 (2004).
20. Li, W. *et al.* Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. *J. Mol. Biol.* **337**, 743–759 (2004).
21. Goh, C.S. & Cohen, F.E. Co-evolutionary analysis reveals insights into protein-protein interactions. *J. Mol. Biol.* **324**, 177–192 (2002).
22. Gsponer, J. *et al.* Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl. Acad. Sci. USA* **103**, 99–104 (2006).
23. Whittaker, S.B. *et al.* NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J. Mol. Biol.* **366**, 1001–1015 (2007).
24. Fersht, A.R. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3–9 (1997).
25. Vendruscolo, M. *et al.* Three key residues form a critical contact network in a protein folding transition state. *Nature* **409**, 641–645 (2001).
26. Lindorff-Larsen, K. *et al.* Calculation of mutational free energy changes in transition states for protein folding. *Biophys. J.* **85**, 1207–1214 (2003).
27. Salvatella, X. *et al.* Determination of the folding transition states of barnase by using PhiI-value-restrained simulations validated by double mutant PhiIJ-values. *Proc. Natl. Acad. Sci. USA* **102**, 12389–12394 (2005).
28. Guerois, R., Nielsen, J.E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387 (2002).
29. Munoz, V. & Serrano, L. Development of the multiple sequence approximation within the AGADIR model of α-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* **41**, 495–509 (1997).
30. Rodriguez-Mendieta, I.R. *et al.* Ultraviolet resonance Raman studies reveal the environment of tryptophan and tyrosine residues in the native and partially folded states of the E. colicin-binding immunity protein Im7. *Biochemistry* **44**, 3306–3315 (2005).
31. Dobson, C.M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
32. Jemth, P. *et al.* The structure of the major transition state for folding of an FF domain from experiment and simulation. *J. Mol. Biol.* **350**, 363–378 (2005).
33. Sutto, L. *et al.* Consequences of localized frustration for the folding mechanism of the Im7 protein. *Proc. Natl. Acad. Sci. USA* **104**, 19825–19830 (2007).
34. Keeble, A.H. & Kleanthous, C. The kinetic basis for dual recognition in colicin endonuclease-immunity protein complexes. *J. Mol. Biol.* **352**, 656–671 (2005).
35. Kuhlmann, U.C. *et al.* Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J. Mol. Biol.* **301**, 1163–1178 (2000).
36. Di Nardo, A.A. *et al.* Dramatic acceleration of protein folding by stabilization of a non-native backbone conformation. *Proc. Natl. Acad. Sci. USA* **101**, 7954–7959 (2004).
37. Gosavi, S. *et al.* Extracting function from a β-trefoil folding motif. *Proc. Natl. Acad. Sci. USA* **105**, 10384–10389 (2008).
38. Neudecker, P. *et al.* Phi-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc. Natl. Acad. Sci. USA* **104**, 15717–15722 (2007).
39. Friel, C.T., Capaldi, A.P. & Radford, S.E. Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: Similarities and differences in the folding of homologous proteins. *J. Mol. Biol.* **326**, 293–305 (2003).
40. Masca, S.I. *et al.* Detailed evaluation of the performance of microfluidic T mixers using fluorescence and ultraviolet resonance Raman spectroscopy. *Rev. Sci. Instrum.* **77**, 055105 (2006).
41. Brooks, B.R. *et al.* CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).