The Mechanism of Folding of Im7 Reveals Competition between Functional and Kinetic Evolutionary Constraints

Claire T. Friel[1,2,3], **D. Alastair Smith**[1,4,5], **Michele Vendruscolo**[6], **Joerg Gsponer**[7,*], and **Sheena E. Radford**[1,2,*]

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

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

[3] Present address: Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, Dresden, Germany

[4]School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK

[5]Present address: Avacta Group Plc, York Biocentre, Innovation Way, York YO10 5NY, UK

[6] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB1 1EW

[7]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

Abstract

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.

Keywords

folding; non-native states; phi values; molecular dynamics; structural ensemble

---

In order 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 rather simple[1,2]. Theoretical studies suggest that this results from a funnel-like global organization of the landscape of accessible protein

conformations3,4 that is the outcome of an evolutionary selection for sequences that minimize the conflict between different interactions, and leads smoothly towards the native conformation5. Indeed, proteins designed computationally or artificially evolved from random libraries, which lack an evolutionary history, fold less cooperatively than their similarly sized naturally occurring counterparts6-8. Further selection for landscapes that promote cooperative folding may result from evolutionary pressure against sequences that promote aggregation9. Accordingly, sequences that suppress the formation of folding intermediates, thereby disfavouring aggregation, have been identified10. Such 'negative design' appears to be essential in the light of findings that non-native structures can play an important role in the formation of amyloid fibrils11.

Despite evidence of evolutionary selection against long-lived folding intermediates, partially folded states have been identified in the folding of many single domain proteins12. 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 intermediate14-17. By contrast, for the Im7 homologue, Im9, an intermediate only becomes detectably populated under acidic conditions, or by targeted substitution of residues to increase hydrophobicity and strengthen non-native interactions during folding18,19. 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 (ref. 20). Thus the evolutionary pressure for the selection of binding-competent sequences of these proteins is critical for the survival of the organism21.

The Im7 folding landscape has been characterized using protein engineering, hydrogen exchange and molecular dynamics (MD) simulations14,17,22,23. 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 and analysis of the resulting Φ-values using restrained MD simulations, we provide the an all-atom description of the entire folding landscape of this protein, including the early transition state for intermediate formation (TS1). In turn, we show how functional constraints play 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, the folding and unfolding kinetics of the wild-type protein and 16 site-specific variants were analyzed (Fig. 1a). At low urea concentrations, where folding is three-state, the refolding kinetics of wild-type Im7 and each variant were analyzed using ultra-rapid, continuous-flow mixing, monitored using the fluorescence of the single tryptophan, Trp75, allowing refolding to be measured between ~200μs and ~2.5ms (Supplementary Fig. S1 online). Stopped-flow

fluorescence measurements were then used to complete the transients (Fig. 1b). The resulting data were fitted globally to a double exponential function (see Methods and Supplementary Methods). At higher urea concentrations, in which the intermediate is no longer populated, the refolding kinetics were measured 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:

Scheme 1

where 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 ($k_{ui} \sim 1600s^{-1}$) 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 hexa-histidine tag[15]. Folding progresses through a highly populated intermediate ($\Delta G_{ui} = -11.7$ kJmol^{-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).

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 Gly and the folding and unfolding kinetics of the variants measured as described above. These substitutions have only a small effect on k_{ui} ($k_{ui} = 1150s^{-1}$ and $1648s^{-1}$ for A13G and A77G, respectively, compared with $1574s^{-1}$ for the wild-type protein) (Fig. 2 and Supplementary Table 1 online), but decrease the stability of I and N (Supplementary Table 2 online), indicating that these residues are not well-ordered in TS1, but form helical structure in I and TS2. Accordingly, 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 substitution V33A increases the helical propensity in the N-terminal region of helix II (Fig. 1a). While accurate Φ-values could not be determined for this substitution since $\Delta\Delta G_{un}$ 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 is not detected in the vicinity of the sites investigated in TS1.

To study the importance of helix III in the folding of Im7, Thr51 and Ile54 were substituted with Ser and Val, respectively (Fig. 1a). Despite residing in the core of native Im7, truncation of these residues does not affect k_{ui} or k_{in}, but increases k_{ni} by ~5 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: ten from helices I, II and IV, plus one (Ile7) that lies 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) reduce k_{ui} by $<500s^{-1}$ (Fig. 2 and Supplementary Table 1). A second group of residues, Val42 (Helix II), Ile68 and Ile72 (Helix IV), reduces k_{ui} by $>500s^{-1}$. The most dramatic changes in k_{ui} are observed for a third group that includes L18A, L19A (Helix I), L37A and L38A (Helix II) for which k_{ui} is reduced by $>1000s^{-1}$ (Fig. 2 and Supplementary Table 1). The Φ_{TS1} values determined for nine of the 11 variants are low (0.1 - 0.4) and, in general, markedly smaller than those for the same residues in I and TS2 (Fig. 3a, 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.

Intriguingly, 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 $\Phi_{TS1} = 0.56\pm0.13$, although k_{ui} is reduced by only ~500s^{-1}. By contrast, for L19A k_{ui} is reduced by ~1400s^{-1}, to a value of only 184s^{-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 very differently to mutation, with the result that $\Delta\Delta G_{ui}$ and $\Delta\Delta G_{un}$ are not linearly correlated over all residues, contrary to proteins that fold by progressive consolidation of native contacts[24]. Indeed, for I72V $\Delta\Delta G_{ui}$ exceeds $\Delta\Delta G_{un}$, whilst for L19A $\Delta\Delta G_{ui} << \Delta\Delta G_{un}$ (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. Interestingly, 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} 440s^{-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, 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, which are presumably variable between members of the ensemble, involving residues both local to and distant from these sites. Interestingly, 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, 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, ensembles of structures representing TS1 and TS2 were calculated 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 MD simulations for

generating representative structural ensembles of transition states of proteins has been demonstrated previously[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 MD simulations are shown in Fig. 4a,b. These ensembles are fully consistent with the experimental Φ-values. This result is expected for TS1 and TS2, since the Φ-values were used as a source of structural information, but is notable for the intermediate state, since in this case equilibrium hydrogen-exchange protection factors – but not Φ-values – were used as restraints (Supplementary Fig. S2)[22]. To assess the quality of the ensembles generated, Φ-values were back-calculated 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]. Importantly, as some experimental Φ_{TS1} values, especially those of L18A, L19A and L37A, depend critically 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, 0.73 between experimental and back-calculated Φ-values for TS1, I, TS2 (with respect to $\Delta\Delta G_{un}$), and 0.76 for TS1 (with respect to $\Delta\Delta G_{ui}$), highlight the quality of all ensembles (Supplementary Fig. S2b). An additional validation of the structures results from the correct prediction of the experimentally determined β_T values (Supplementary Methods online).

To provide further controls, ensembles for TS1 and TS2 were determined following the same protocol but using either reshuffled Φ-values or a restricted set of eight Φ-values (Supplementary Fig. S3). The new ensembles were then used 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. S3a-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. S3f-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. S4a), 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 20Å 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 experimentally determined 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 be discerned. The results establish that the U 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. S4a). 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). While 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 sub-optimal 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 helix II and helix III, Tyr55 in particular, still form more non-native than native contacts within this ensemble (Figs. 5 and 6b).

To determine more precisely the nature of the reorganizational events leading to and from the intermediate state, the TS1, I and TS2 ensembles were analyzed 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 since Phe41 forms a crucial part of the native hydrophobic core and shows clear evidence for the formation of non-native contacts during folding using both experiment[14] and simulation (Fig. 6b). Tyr55 is partially solvent exposed in native Im7 and is predicted to form non-native contacts throughout the folding process. In addition, the interactions of Trp75 (helix IV) were monitored, since both experiment and simulation suggest that this residue is more buried in I than in any other state (Supplementary Fig. S4b)[22,30].

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 Fig. 7a. These profiles reveal that Trp75 makes a large number of non-native interactions with residues in regions 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, the distribution of distances between Phe41 (helix II) and either I54 (helix III) or Trp75 (helix IV) was determined for I, TS2 and N (Supplementary Fig. S4c online). Whilst Phe41 is

close to Ile54 in TS2 and N, this is not the case for the 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 MD 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 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. While 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 re-organization 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 residues identified here to form non-native contacts during the early stages of folding of Im7 lie in regions that play a vital role in the function of immunity proteins: the recognition and inactivation of colicin toxins (Fig. 6a-d and Supplementary Fig. S5)[20]. An initial docking of the conserved residues Tyr55 and Tyr56 in helix III onto the colicin surface anchors cognate and non-cognate complexes. This is followed by exploration of the second docking site, involving primarily residues in helix II, the binding free energy of which discriminates between cognate and non-cognate 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) providing the capability for colicin inhibition required for survival of the organism, whilst 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 proteins36-38, 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.

Materials and Methods

Data collection and analysis

Im7 variants were created, expressed and purified as described14,39. Kinetic measurements were performed at 10°C in 50mM sodium phosphate buffer, pH 7.0 containing 0.4M sodium sulfate using a custom built continuous flow instrument40 and an Applied Photophysics SX18.MV stopped flow instrument39. Final protein concentrations were ~20μM for the continuous-flow and ~5-20μM 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). In order to constrain the endpoint of the fit to the refolding transient obtained by continuous-flow mixing, the fluorescence signal at equilibrium at each concentration of denaturant was measured using premixed samples and the endpoint constrained to this value (see Supplementary Methods online). 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 (see Supplementary Methods online). Φ-values for TS1, I and TS2 were then calculated using the microscopic rate constants determined (see Supplementary Methods online). Errors were propagated mathematically from the errors determined on the fit parameters.

Restrained MD simulations

The CHARMM22 (ref. 41) force-field was used to carry out MD simulations with Φ-value restraints25 using an all-atom protein representation, the TIP3P water model and periodic boundary conditions41. All calculations used an atom-based truncation scheme with a list cut-off of 14Å, a non-bond cut-off of 12Å, and the Lennard-Jones smoothing function initiated at 10Å. Electrostatic and Lennard-Jones interactions were force switched. Molecular dynamics simulations used a 2fs integration time step and SHAKE of covalent bonds involving hydrogen atoms. For more detailed information see the Supplementary Methods online.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Roder H, Maki K, Cheng H. Early events in protein folding explored by rapid mixing methods. Chem. Rev. 2006; 106:1836–1861. [PubMed: 16683757]

2. Schuler B, Eaton WA. Protein folding studied by single-molecule FRET. Curr. Opin. Struct. Biol. 2008; 18:16–26. [PubMed: 18221865]

3. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat. Struct. Biol. 1997; 4:10–19. [PubMed: 8989315]

4. Onuchic JN, Wolynes PG. Theory of protein folding. Curr. Opin. Struct. Biol. 2004; 14:70–75. [PubMed: 15102452]

5. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. Proc. Natl. Acad. Sci. USA. 1987; 84:7524–7528. [PubMed: 3478708]

6. Hill RB, et al. De novo design of helical bundles as models for understanding protein folding and function. Acc. Chem. Res. 2000; 33:745–754. [PubMed: 11087311]

7. Sauer RT. Protein folding from a combinatorial perspective. Fold. Des. 1996; 1:R27–30. [PubMed: 9079366]

8. Watters AL, et al. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. Cell. 2007; 128:613–624. [PubMed: 17289578]

9. Monsellier E, Chiti F. Prevention of amyloid-like aggregation as a driving force of protein evolution. EMBO Rep. 2007; 8:737–742. [PubMed: 17668004]

10. Mitraki A, et al. Global suppression of protein folding defects and inclusion body formation. Science. 1991; 253:54–58. [PubMed: 1648264]

11. Jahn TR, Radford SE. Folding versus aggregation: polypeptide conformations on competing pathways. Arch. Biochem. Biophys. 2008; 469:100–117. [PubMed: 17588526]

12. Brockwell DJ, Radford SE. Intermediates: ubiquitous species on folding energy landscapes? Curr. Opin. Struct. Biol. 2007; 17:30–37. [PubMed: 17239580]

13. Dennis CA, 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. 1998; 333:183–191. [PubMed: 9639578]

14. Capaldi AP, Kleanthous C, Radford SE. Im7 folding mechanism: misfolding on a path to the native state. Nat. Struct. Biol. 2002; 9:209–216. [PubMed: 11875516]

15. Capaldi AP, et al. Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. Nat. Struct. Biol. 2001; 8:68–72. [PubMed: 11135674]

16. Ferguson N, et al. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. J. Mol. Biol. 1999; 286:1597–1608. [PubMed: 10064717]

17. Gorski SA, et al. Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7. J. Mol. Biol. 2004; 337:183–193. [PubMed: 15001361]

18. Cranz-Mileva S, Friel CT, Radford SE. Helix stability and hydrophobicity in the folding mechanism of the bacterial immunity protein Im9. Protein Eng. Des. Sel. 2005; 18:41–50. [PubMed: 15790579]

19. Friel CT, Beddard GS, Radford SE. 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. 2004; 342:261–273. [PubMed: 15313622]

20. Li W, et al. Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. J. Mol. Biol. 2004; 337:743–759. [PubMed: 15019791]

21. Goh CS, Cohen FE. Co-evolutionary analysis reveals insights into protein-protein interactions. J Mol Biol. 2002; 324:177–192. [PubMed: 12421567]

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. 2006; 103:99–104. [PubMed: 16371468]

23. Whittaker SB, et al. NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. J. Mol. Biol. 2007; 366:1001–1015. [PubMed: 17188712]

24. Fersht AR. Nucleation mechanisms in protein folding. Curr. Opin. Struct. Biol. 1997; 7:3–9. [PubMed: 9032066]

25. Vendruscolo M, et al. Three key residues form a critical contact network in a protein folding transition state. Nature. 2001; 409:641–645. [PubMed: 11214326]

26. Lindorff-Larsen K, et al. Calculation of mutational free energy changes in transition states for protein folding. Biophys. J. 2003; 85:1207–1214. [PubMed: 12885664]

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. 2005; 102:12389–12394. [PubMed: 16116084]

28. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J. Mol. Biol. 2002; 320:369–387. [PubMed: 12079393]

29. Munoz V, Serrano L. Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. Biopolymers. 1997; 41:495–509. [PubMed: 9095674]

30. Rodriguez-Mendieta IR, 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. 2005; 44:3306–3315. [PubMed: 15736941]

31. Dobson CM. Protein folding and misfolding. Nature. 2003; 426:884–890. [PubMed: 14685248]

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. 2005; 350:363–378. [PubMed: 15935381]

33. Sutto L, et al. Consequences of localized frustration for the folding mechanism of the Im7 protein. Proc. Natl. Acad. Sci. USA. 2007; 104:19825–19830. [PubMed: 18077415]

34. Keeble AH, Kleanthous C. The kinetic basis for dual recognition in colicin endonuclease-immunity protein complexes. J. Mol. Biol. 2005; 352:656–671. [PubMed: 16109424]

35. Kuhlmann UC, et al. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. J. Mol. Biol. 2000; 301:1163–1178. [PubMed: 10966813]

36. Di Nardo AA, et al. Dramatic acceleration of protein folding by stabilization of a non-native backbone conformation. Proc. Natl. Acad. Sci. USA. 2004; 101:7954–7959. [PubMed: 15148398]

37. Gosavi S, et al. Extracting function from a beta-trefoil folding motif. Proc. Natl. Acad. Sci. USA. 2008; 105:10384–10389. [PubMed: 18650393]

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. 2007; 104:15717–15722. [PubMed: 17898173]

39. Friel CT, Capaldi AP, Radford SE. Structural analysis of the rate-limiting transition states in the folding of lm7 and lm9: Similarities and differences in the folding of homologous proteins. J. Mol. Biol. 2003; 326:293–305. [PubMed: 12547210]

40. Masca SI, et al. Detailed evaluation of the performance of microfluidic T mixers using fluorescence and ultraviolet resonance Raman spectroscopy. Rev. Sci. Instrum. 2006; 77

41. Brooks BR, et al. CHARMM- A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 1983; 4:187–217.

$$U \xrightleftharpoons[k_{iu}]{k_{ui}} I \xrightleftharpoons[k_{ni}]{k_{in}} N$$
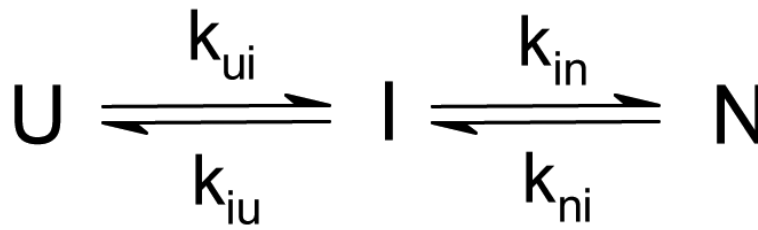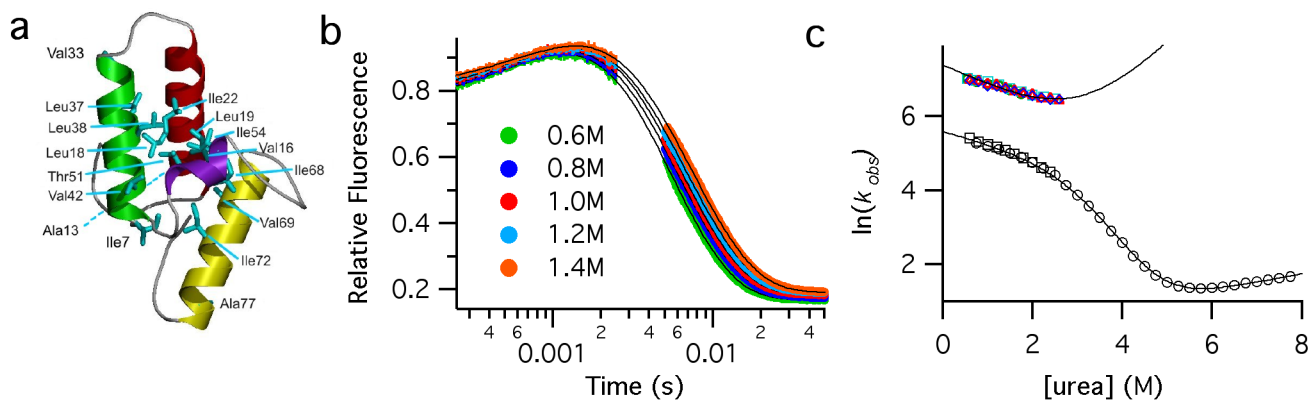
Scheme 1.

Fig. 1.
Native structure of Im7 and representative kinetic traces. (a) Structure of Im7 (1AYI) showing residues that were mutated in this study; Ala13 is at the back of the molecule in this orientation. Helices are coloured as 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/unfolding of wild-type Im7. The coloured 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.
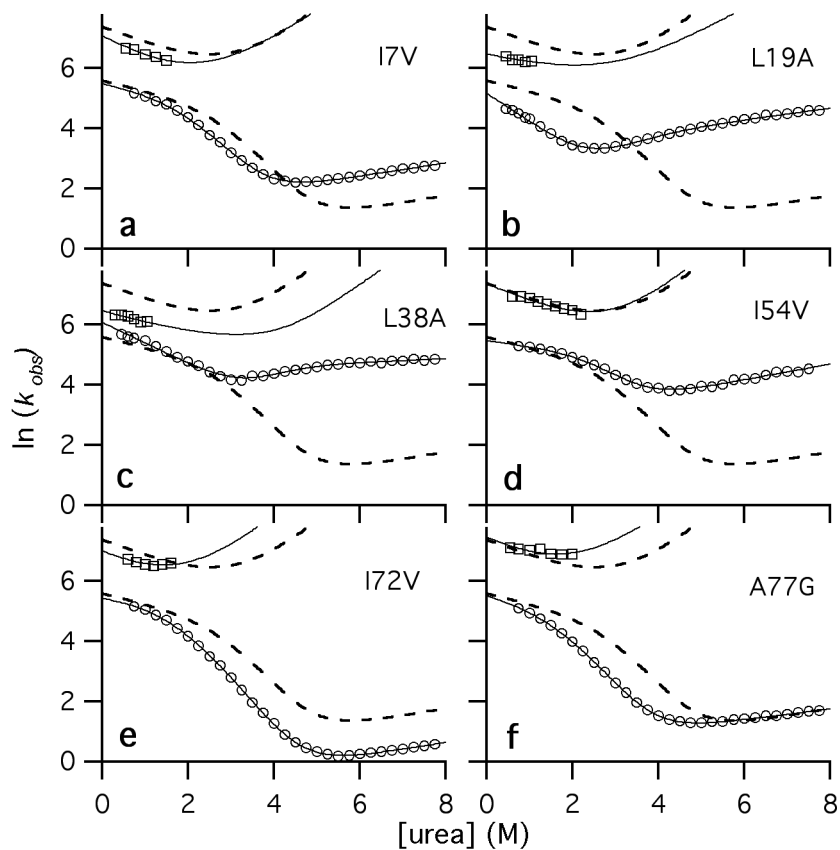
Fig. 2.
Dependence of the natural logarithm of the observed rate constants on the concentration of urea for selected Im7 variants. Solid lines show the best fit of the data to a three-state model with an on-pathway intermediate. The dashed line shows the best fit to the data for wild-type Im7 for comparison. (a) I7V, (b) L19A, (c) L38A, (d) I54V, (e) I72V, (f) A77G.
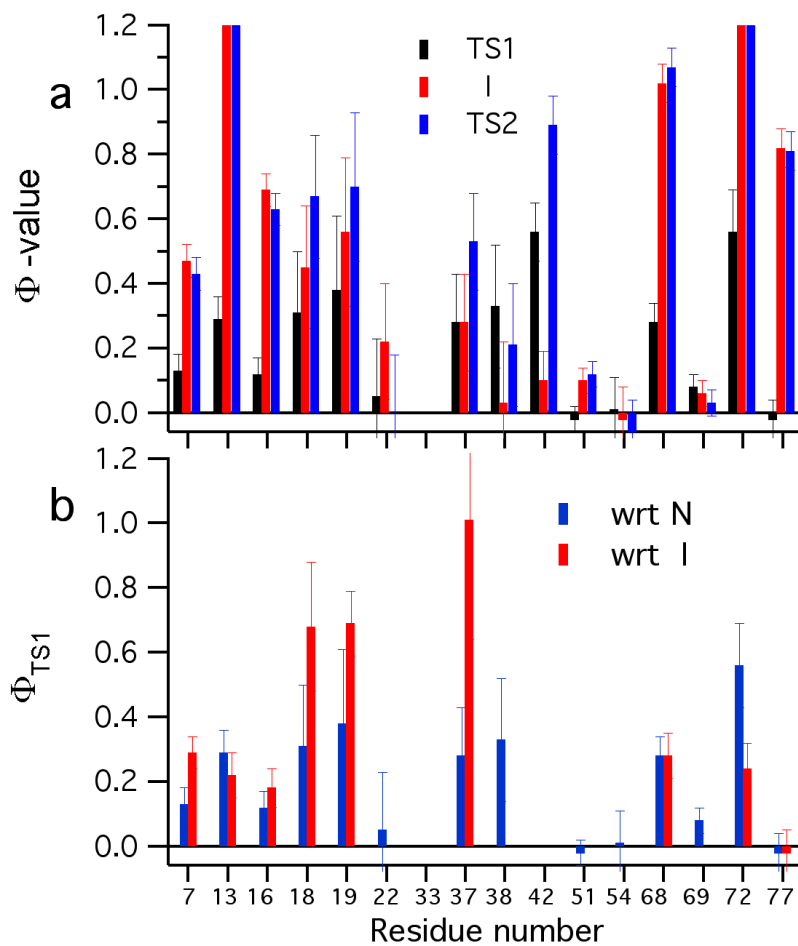
Fig. 3.
Calculated Φ-values. Bar plots of (a) Φ-values normalized to $\Delta\Delta G_{un}$ for TS1, I and TS2 (black, red and blue, respectively). (b) Φ-values for TS1 normalized to $\Delta\Delta G_{ui}$ (red) or $\Delta\Delta G_{un}$ (blue). Error bars depict 1 s.d..
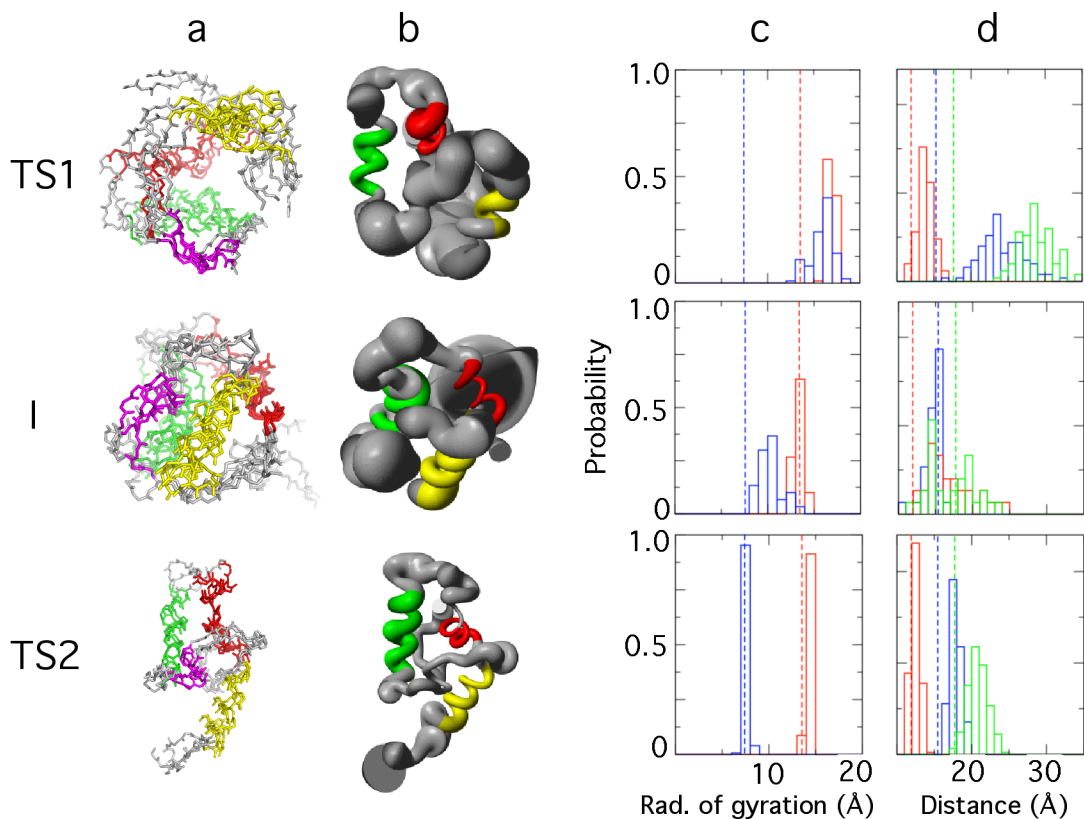
Fig. 4.
Comparison of structural properties of TS1, I and TS2. (a) Representative members of the three ensembles. The segments forming the four helices in N are shown in red (I), green (II) magenta (III) and yellow (IV). (b) Diagram displaying the heterogeneity of different ensembles: the tube diameter is scaled by the average displacement to the mean structure. (c) Distribution of the radius 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 centres 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.
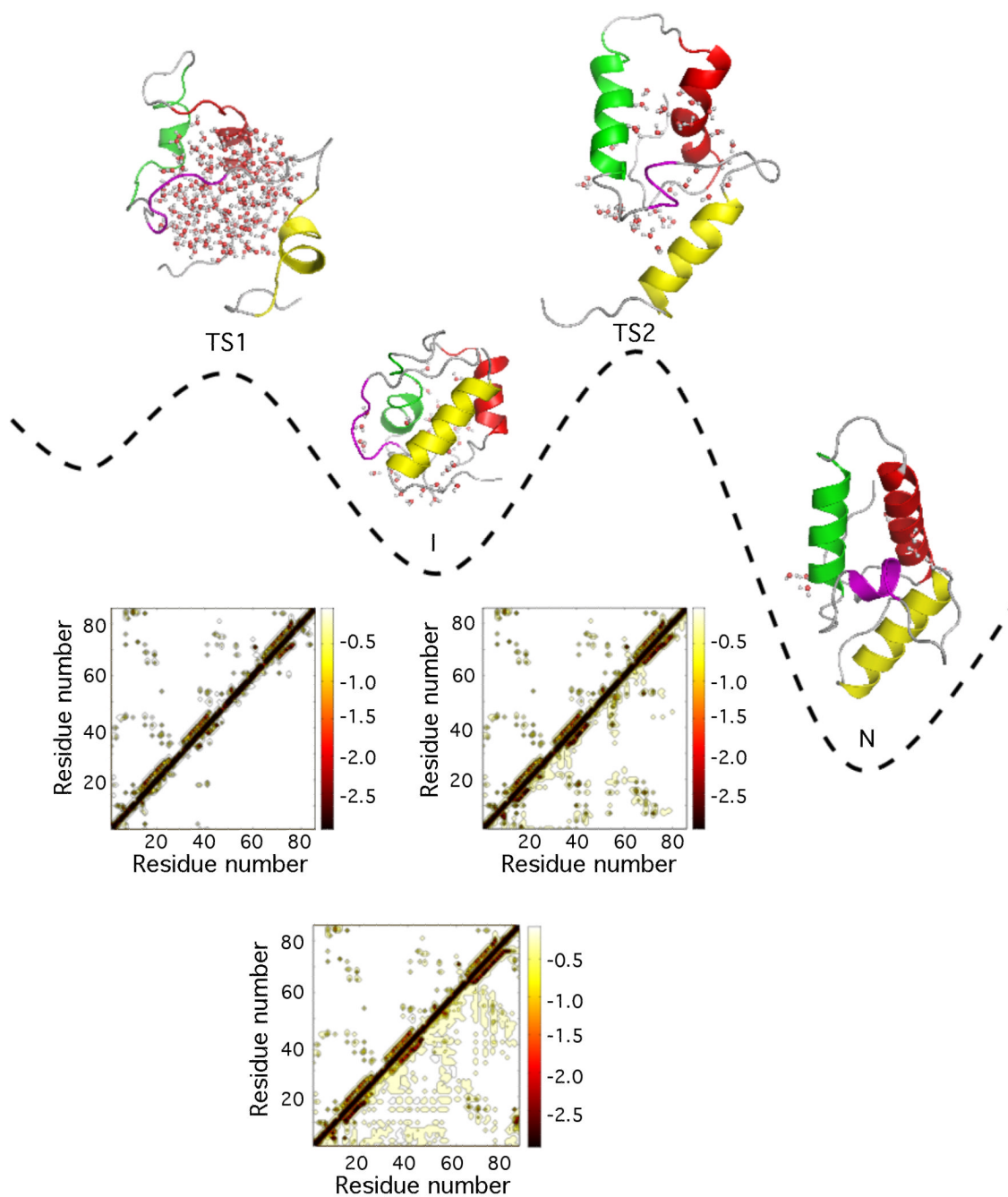
Fig. 5.
Schematic illustration of the folding landscape of Im7. Ribbon diagram representations of selected cluster centers of TS1, I, TS2 and N are shown. The helix forming segments are coloured red (helix I), green (helix II), purple (helix III) and yellow (helix IV). Water molecules within a 12Å sphere around the centre of masses of TS1, I, TS2 and N are shown in ball and stick representation. Lower plots: the average interaction energy maps of TS1, I and TS2 (below the diagonal) are compared with that of N (above the diagonal). The scale is in kcal mol^{-1}.
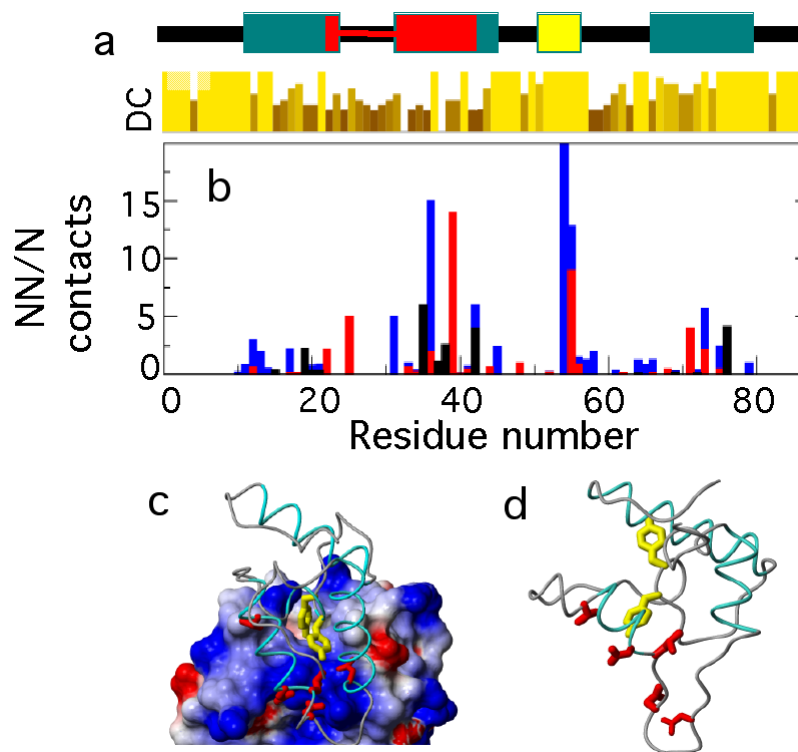
Fig. 6.
Non-nativeness during folding *versus* functionality. (a) Degree of conservation (DC); reflecting the conservation of physico-chemical properties in each column of the alignment of Im2, Im7, Im8 and Im9 (see also Supplementary Fig. S5 online). The positions of the native helices are shown above; 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).
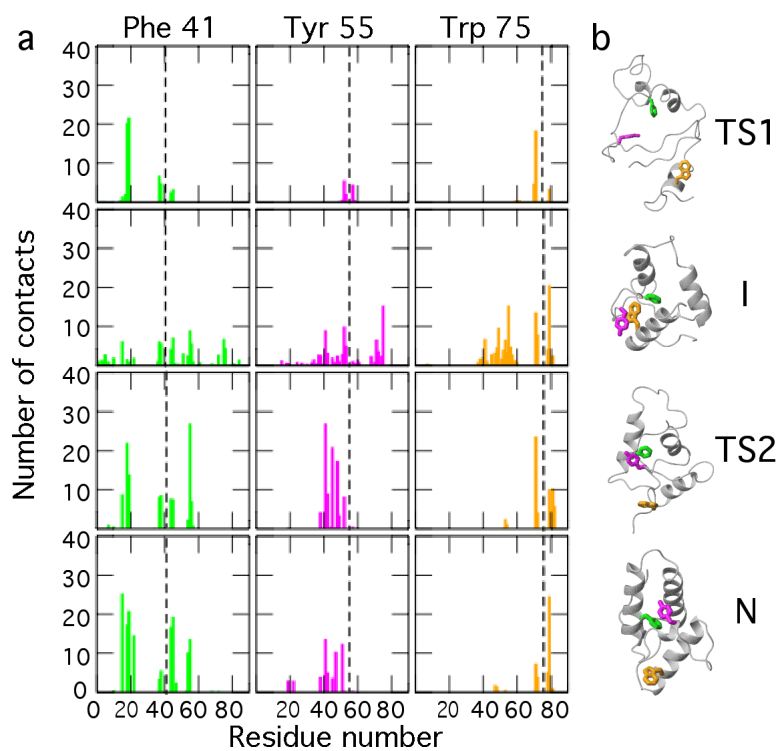
Fig. 7.
Interaction patterns of selected residues in TS1, I, TS2 and N. (a) Number of atomic contacts (native and non-native) formed between residues Phe41, Tyr55, Trp75 and the other residues of Im7 in TS1, I, TS2, and N. (b) Structures illustrating the position of residues Phe41 (green), Tyr55 (magenta), and Trp75 (orange) in each ensemble.