nature structural & molecular biology

Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain

Stefano Gianni[1,4], Ylva Ivarsson[1,3,4], Alfonso De Simone[2], Carlo Travaglini-Allocatelli[1], Maurizio Brunori[1] & Michele Vendruscolo[2]

Incorrectly folded states transiently populated during the protein folding process are potentially prone to aggregation and have been implicated in a range of misfolding disorders that include Alzheimer's and Parkinson's diseases. Despite their importance, however, the structures of these states and the mechanism of their formation have largely escaped detailed characterization because of their short-lived nature. Here we present the structures of all the major states involved in the folding process of a PDZ domain, which include an off-pathway misfolded intermediate. By using a combination of kinetic, protein engineering, biophysical and computational techniques, we show that the misfolded intermediate is characterized by an alternative packing of the N-terminal β-hairpin onto an otherwise native-like scaffold. Our results suggest a mechanism of formation of incorrectly folded transient compact states by which misfolded structural elements are assembled together with more extended native-like regions.

The large majority of biochemical processes that take place in living cells rely on the ability of proteins to fold into their functional native states[1]. When folding fails, the identification and elimination of the misfolded species becomes a critical task, as the accumulation of misfolded proteins is associated with a wide range of human disorders such as Alzheimer's and Parkinson's diseases and type II diabetes[2,3]. To distinguish between native and non-native conformations of proteins, the cell uses various sensor molecules, including molecular chaperones, that selectively recognize misfolded species[4–6]. Thus, the quality control system does not degrade functional species, but it is triggered by the presence of folding defects[7,8]. Although it is generally assumed that the exposure of aggregation-prone hydrophobic regions is the hallmark of misfolding[9], only recently is it becoming possible to outline the structural characteristics of folding defects that mediate the recognition of misfolded species by the quality control system[10–12]. Although major advances have been made in determining the structural features of on-pathway folding intermediates by a variety of techniques, including NMR spectroscopy and Φ-value analysis[13–18], the primary aim of these studies has been to define the rules that govern the folding process. Thus, despite their key role in initiating pathological events, we are just beginning to clarify the structural features of misfolded metastable states[19,20].

In this study we consider the problem of characterizing the structure of a misfolded intermediate and of defining the molecular events that determine whether nonproductive misfolding pathways are taken, rather than the correct folding pathway. In order to address these questions, we consider a PDZ domain protein that has been recently shown to populate an off-pathway intermediate[21]. PDZ domains are small (90–100 amino acid) domains that adopt a six-stranded β-sandwich fold flanked by two α-helices[22], with the typical linear arrangement of the secondary structure elements being ββββαββαβ. However, the PDZ domain family also comprises circularly permuted variants, whereby the canonical β1 strand is shifted to the C terminus of the sequence[23,24]. A number of canonical PDZ domains fold via a conserved three-state mechanism, proceeding through two sequential transition states and one unstable on-pathway intermediate[25–30]. By contrast, the folding process of a naturally occurring, circularly permuted PDZ domain—the D1 C-terminal processing protease (D1p) of the green alga *Scenedesmus obliquus*—involves a kinetic competition between folding to the native state and misfolding to a low-energy off-pathway intermediate[31]. This type of misfolding is likely to be relatively rare in small protein domains, but it may be more common with larger proteins that have more complex topologies . We recently engineered a circularly permuted variant of PTP-BL PDZ2, mimicking the sequence connectivity of the naturally occurring circularly permuted PDZs. This variant is characterized by the presence of an intermediate, which is stabilized by the circular permutation[16,21]. The PDZ family thus offers the possibility to investigate the determinants of folding and misfolding in the context of divergent sequence and sequence connectivity, within the same three-dimensional frame.

To map the structural features of both the productive and nonproductive folding pathways of this small protein domain, we carried out a Φ-value analysis, whose results were then incorporated as structural restraints in molecular dynamics simulations of D1pPDZ. Our results indicate that the structural features in the productive folding transition state of D1pPDZ resemble the early transition state for folding

[1]Istituto Pasteur–Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche 'A. Rossi Fanelli', Università di Roma 'La Sapienza', Rome, Italy. [2]Department of Chemistry, University of Cambridge, Cambridge, UK. [3]Present address: Department of Human Genetics, Katholieke Universiteit Leuven, Herestraat, Leuven, Belgium. [4]These authors contributed equally to this work. Correspondence should be addressed to M.V. (mv245@cam.ac.uk).

Received 31 March; accepted 8 October; published online 14 November 2010; doi:10.1038/nsmb.1956

of the canonical PTP-BL PDZ2, which we described previously[28], and thus suggest the existence of a conserved productive folding pathway, despite the low sequence identity and alternative sequence connectivity . Our study, however, also reveals how specific events can cause the failure of this conserved folding mechanism. Our analysis of the nonproductive pathway indicates that the misfolded intermediate of D1pPDZ, despite having an overall native-like topology, is characterized by a non-native and nonfunctional arrangement of the ligand-binding pocket region, caused by the incorrect docking of the N-terminal β-hairpin onto the remainder of the structure. Taken together, our results indicate that the productive folding of D1pPDZ is driven by the formation of interactions in the main β-sheet and its docking with α2, whereas misfolding is caused by the incorrect docking of the N-terminal β-hairpin.

RESULTS

Φ-value analysis

A general and challenging problem in studying the structure of intermediates in the folding process is that these reactions are generally highly cooperative, so that any other state populated during the process is extremely difficult to observe by standard methods of structural biology[1]. A powerful strategy to address this problem, and to obtain structural information about intermediate and transition states, is to carry out a systematic comparison of the folding and unfolding rate constants for a series of mutational variants, from which it is possible to map out the interaction patterns in the intermediate and transition states. In this approach, the so-called Φ-value analysis[32], the strength of the contacts is measured by a parameter—the Φ value—that represents the change in stability of an intermediate (or of a transition state) relative to that of the native state. Thus, Φ = 1 indicates that the site of mutation is fully structured in the intermediate (or in the transition state), whereas Φ = 0 indicates that such a site is essentially unstructured, and intermediate values usually represent a partial degree of formation of the native interactions in the intermediate (or in the transition state).

In this work, we used Φ-value analysis to study the properties of the main states populated during the folding process of D1pPDZ. Forty positions in the 95-residue sequence of this protein were probed by conservative site-specific mutagenesis; mutants at such positions were constructed, purified and characterized (**Supplementary Table 1** and **Supplementary Fig. 1**). The folding time-course of wild-type D1pPDZ is biphasic (**Fig. 1a**), both in the presence and absence of stabilizing salt[31]. We have previously demonstrated that such a biphasic refolding is caused by the kinetic partitioning between two folding pathways, one leading to the more stable native state and the other to the less stable off-pathway intermediate (**Fig. 1a**). In the present study, we carried out urea-induced kinetic (un)folding experiments in the presence of a stabilizing salt (0.4 M sodium sulfate). The majority of the D1pPDZ variants—30 mutants—had biphasic kinetic (un)folding traces, similar to those observed for the wild type; the time courses were fitted to a double-exponential function, allowing unambiguous determination of all four microscopic rate constants. Reliable Φ values were calculated for 24 variants ($\Delta\Delta G_{N-D}$ > 0.5 kcal mol^{-1}, ref. 33), thereby providing information on interactions formed in the intermediate (Φ_I, **Supplementary Table 1**) and in the two distinct transition states (Φ_{TSN} and Φ_{TSI}, respectively, **Supplementary Table 1**).

In contrast to the biphasic time course in the folding of wild-type D1pPDZ, ten mutants

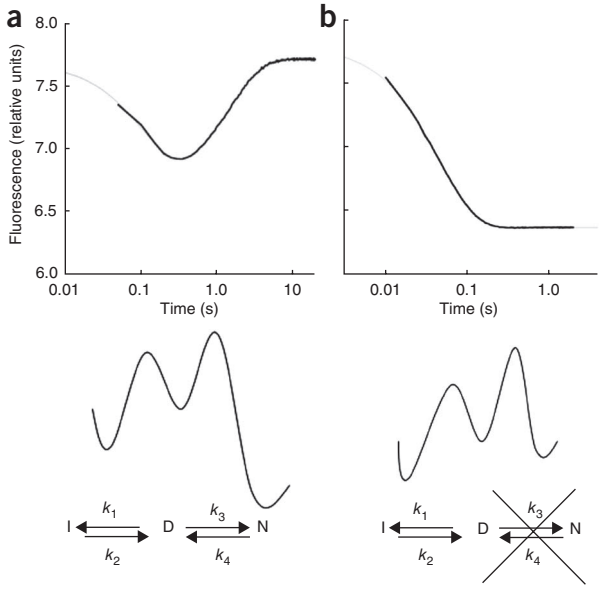

Figure 1 Folding kinetics of D1pPDZ. (**a**) Representative refolding trace of D1pPDZ. Gray line is the best fit to a double exponential. (**b**) Trace of a mutant (L245A) that populates almost exclusively the misfolded intermediate. The gray line is the best fit to a single exponential. Although refolding of D1pPDZ is biphasic, the mutant shows only the fast phase, reflecting the transition from the denatured state to the misfolded intermediate.

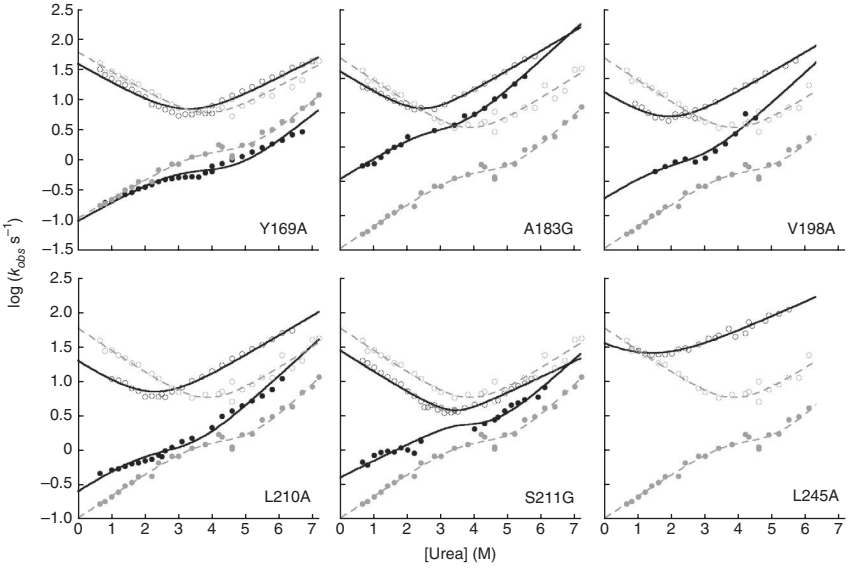

Figure 2 Comparison between the chevron plots of D1pPDZ and of six representative mutants (Y169A, A183G, V198A, L210A, S211G and L245A). Empty circles refer to the kinetic data of intermediate (un)folding, and filled circles to native (un)folding. The folding of D1pPDZ is indicated by gray symbols and lines. Mutant proteins are indicated by black symbols and lines. Data for all variants (except L245A) are fitted to a three-state equation with an off-pathway intermediate. Because of the complexity of the kinetic model, the *m* values for the microscopic rate constants were assumed to be the same as those determined for wild-type D1pPDZ. The kinetic data of L245A (un)folding were fitted to a two-state equation.

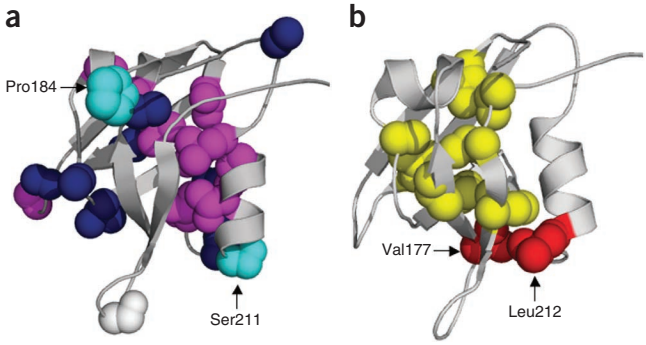

Figure 3 Structural distribution of the Φ_I values in the misfolding reaction of D1pPDZ. (**a**) Mapping of the experimentally measured Φ_I values on the native structure of D1pPDZ. The Φ_I values are divided into four categories and color coded accordingly: white, $0 < \Phi_I < 0.33$; magenta, $0.33 < \Phi_I < 0.66$; deep blue, $0.66 < \Phi_I < 1$; and cyan, $\Phi_I > 1$. (**b**) Mutations at ten positions (shown in yellow or red spheres) destabilize the native state but do not substantially destabilize the intermediate, thereby shifting the equilibrium toward the intermediate (**Supplementary Table 1**). The V177A and L212A mutations (at the positions marked by red spheres) stabilize the intermediate while destabilizing the native state, indicating that the two residues have a negative Φ_I value.

showed single-exponential refolding and unfolding time-courses under all investigated conditions (**Fig. 1b** and **Supplementary Table 1**). These variants were characterized by a fluorescence change associated with the transition from the denatured state to the intermediate, during the refolding reaction, and vice versa during unfolding (**Fig. 1b**). In analogy with what was previously observed for the Engrailed homeodomain[34], we conclude that the mutations destabilize the native state much more than the intermediate, modifying the folding process by increasing the lifetime of the latter. The experimental traces of these ten mutants conformed to single-exponential time courses, and their chevron plots were fitted to a two-state equation (**Fig. 2**).

Because unusual Φ values, such as $\Phi > 1$ or $\Phi < 0$, are often indicative of non-native contact formation[35–37], they can in principle be used to detect misfolded regions. The folding kinetics of wild-type D1pPDZ clearly indicates that the intermediate represents a kinetic trap that needs to unfold before the folding process can proceed to the native state. We did therefore expect the Φ-value analysis to produce a set of unusual Φ values indicating the existence of large regions of the structure of the intermediate characterized by the presence of non-native interactions. Notably, among the 24 reliable Φ_I values obtained, only two (for A184G, $\Phi_I = 1.2 \pm 0.2$, and for S211G, $\Phi_I = 1.9 \pm 0.6$) indicated non-native contact formation. The great majority of Φ_I values range from 0.5 to 1, suggesting that the overall fold of the intermediate is largely similar to those found in partially or fully formed native-like interactions (**Fig. 3a** and **Supplementary Table 1**).

Additional insight about the structure of the misfolded intermediate is provided by the set of ten mutants that destabilize the native state and increase the lifetime of the intermediate (**Supplementary Table 1** and **Figs. 1** and **3b**); most of these positions are part of the hydrophobic core of native D1pPDZ. We also found, however, that two of the mutations (V177A and L212A, **Fig. 3b**) stabilize the intermediate and destabilize the native one, implying a negative Φ value. In the native protein, these two positions are engaged in long-range interactions (**Fig. 3b**) and are also close to position 211, where the S211G mutation is associated with an unusual Φ_I value ($\Phi_I = 1.9 \pm 0.6$).

Taken together, these observations suggest that the intermediate is largely stabilized by native-like interactions and misfolding is limited to the specific region involving the packing of the N-terminal β-hairpin and α2. In this sense, the cause of misfolding is the improper docking of the N-terminal β-hairpin, which acquires a swapped orientation in the intermediate.

A structural model of the misfolded intermediate

The structure of the intermediate was defined by using molecular modeling and molecular dynamics simulations with Φ-value restraints[38–41]. In this method, the experimental Φ values are assumed to reflect the fraction of native contacts that the corresponding protein residues form in the intermediate (see Online Methods). We used the Φ_I values (**Supplementary Table 1**) to calculate an ensemble of conformations representing the structure of the intermediate; its topology was assumed to comprise an inverted N-terminal β-hairpin (strands β1 and β2) packed onto the rest of the structure, which was considered to be largely native-like. Whereas the native state and the intermediate share essentially identical secondary structure content, the single tryptophan residue at position 178 (Trp178) adopts a different orientation in the native state (**Fig. 4a**) than in the intermediate (**Fig. 4b**), being partly solvent exposed in the native state and almost completely buried in the intermediate. We further observe that incorrect docking of the N-terminal β-hairpin onto α2 compromises the architecture of the ligand-binding pocket, which is located at the interface between α2 and β1 in circularly permuted PDZ domains.

Validation of the structural model of the misfolded intermediate

To validate the structure of the intermediate that we calculated by incorporating Φ values as restraints in molecular dynamics simulations, we exploited the opportunities offered by the fact that ten of the mutants presented in this work destabilize the native state

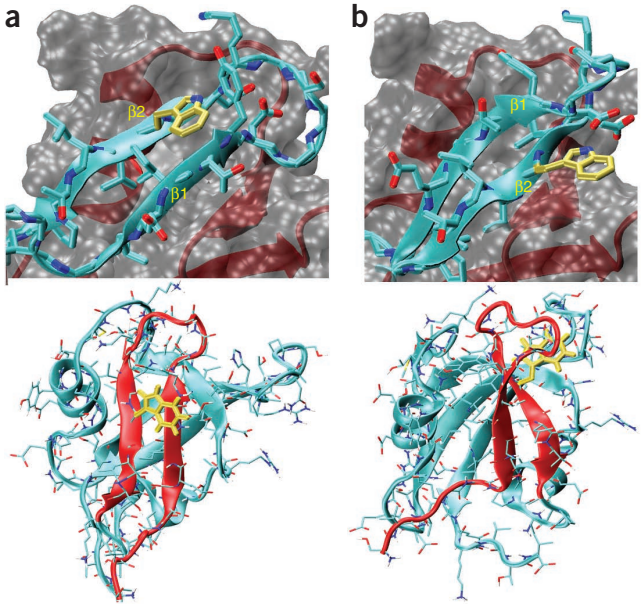

Figure 4 Comparison of the structures of the intermediate and the native states. (**a**) Intermediate state. (**b**) Native state. The upper panels illustrate the packing between the N-terminal β1-β2 hairpin and the rest of the protein. The β1-β2 hairpin is drawn as cyan ribbons, with side chains explicitly shown; the remaining part of the protein is drawn as red ribbons and gray surfaces. The structures of the intermediate and native states are shown in the lower panels. The β1-β2 hairpin is drawn as red ribbons and the rest of the protein as cyan ribbons. The single tryptophan residue in position 178 is partly solvent exposed in the native state yet is completely buried in the intermediate state.

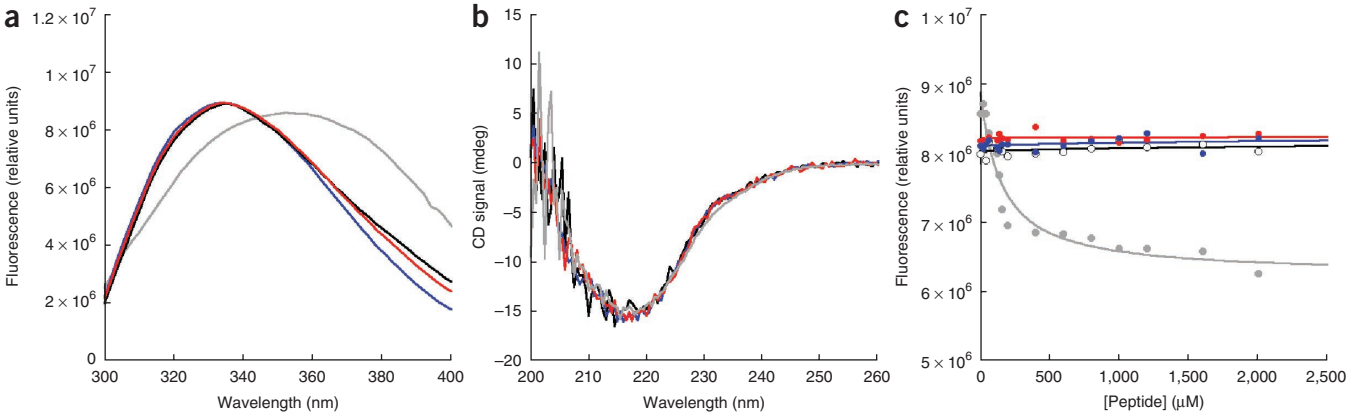

Figure 5 Comparison of the spectroscopic and functional properties of D1pPDZ (gray) and three mutants that populate the intermediate state, V163A (blue), A193G (red) and L231A (black). (**a**) Fluorescence emission spectra of D1pPDZ. (**b**) Near-UV circular dichroism spectra. (**c**) Equilibrium binding titration to the peptide EAPSVNA monitored by the change in intrinsic fluorescence at 350 nm. The experiments were done in 50 mM sodium phosphate, pH 7.2, at a constant protein concentration of 5 μM. Whereas D1pPDZ yields a simple hyperbolic transition, no detectable change in fluorescence could be observed with the mutants.

much more than the intermediate, thus shifting the equilibrium toward the latter. Among these ten mutational variants, we selected V163A, A193G and L231A, which are located at different positions along the polypeptide chain, for the analysis.

Consistent with the predictions made on the basis of the structure of the intermediate that we presented above, the fluorescence emission spectra of the three mutational variants are blue shifted relative to the case of the native state (**Fig. 5a**), indicating that Trp178 lies in a more hydrophobic environment in the intermediate, which thus confirms the orientation of the Trp178 side chain reported in **Figure 4b**. In addition, the three mutational variants and the wild-type protein have nearly identical far-UV circular dichroism spectra (**Fig. 5b**), indicating that the secondary structure content in the intermediate is similar to that in the native state, which is consistent with the structure of the intermediate that we calculated. Furthermore, a synthetic peptide (EAPSVNA) mimicking the natural binding target, the C-terminal extension of the D1 polypeptide of photosystem II[31], was used in ligand-binding equilibrium experiments. As previously shown for other PDZ domains[31], binding of the ligand produces a conformational change often sufficient to alter the intrinsic protein fluorescence . Although the wild-type protein binds the peptide with a hyperbolic behavior with an apparent K_D of about 200 μM, none of the three mutants that we analyzed showed a detectable change in fluorescence, even in the presence of up to 2 mM peptide (**Fig. 5c**). This result indicates clearly that no binding takes place in the intermediate, as expected on the basis of the structure of the intermediate that we calculated, in which the architecture of the binding pocket is destroyed by the misfolding of the N-terminal β-hairpin.

Because all three of the mutational variants that we analyzed (V163A, A193G and L231A) have similar spectroscopic and functional properties, these experimental data support the structure of the intermediate that we presented, and they indicate that this state, although retaining a native-like secondary structure, has a misfolded ligand-binding pocket, which arises from a swapped orientation of the N-terminal β-hairpin.

Structural models of the TSN and TSI transition states

The analysis of the folding kinetics of the series of mutational variants that we considered in this work also enable us not just to define the structure of the misfolded intermediate but also to understand the molecular events that lead to its formation and that are distinct from those leading to the correct folding to the native structure.

As the key events in the productive folding or in the nonproductive misfolding pathways are the formation of the respective transition states, we used the Φ-value analysis to characterize the structure of such states; we denote as TSN the transition state between the unfolded and the native state and as TSI the transition state between the unfolded and the misfolded intermediate. As in the case of the intermediate, we calculated the TSN and TSI structures by using molecular dynamics simulations with Φ-value restraints (see Online Methods). We used Φ_{TSI} and Φ_{TSN} values (**Supplementary Table 1**)

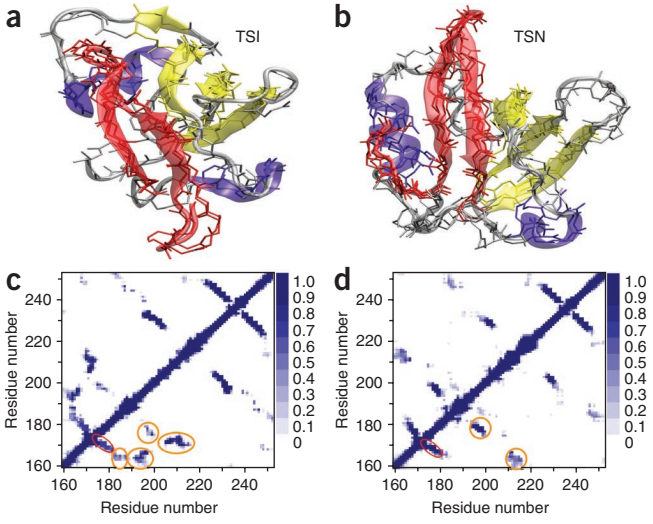

Figure 6 Comparison of the TSN and TSI structures of D1pPDZ. (**a**) Conformational ensemble of TSI. (**b**) Conformational ensemble of TSN. The color codes are gray (loops), purple (helices), yellow (β-strands) and red (β1 and β2). (**c**) Contact map of TSI. (**d**) Contact map of TSN. The contact maps are averaged over the structural ensembles; for each member of the ensemble, a contact is assigned if at least an inter-residue distance (heavy atoms only) is below 6.5 Å. Probabilities are normalized to range from 0 (for not contacting residues) to 1 (for tightly contacting residues). Red circles indicate contacts between strands β1 and β2. Orange circles indicate contacts between the β1-β2 hairpin and the rest of the protein.

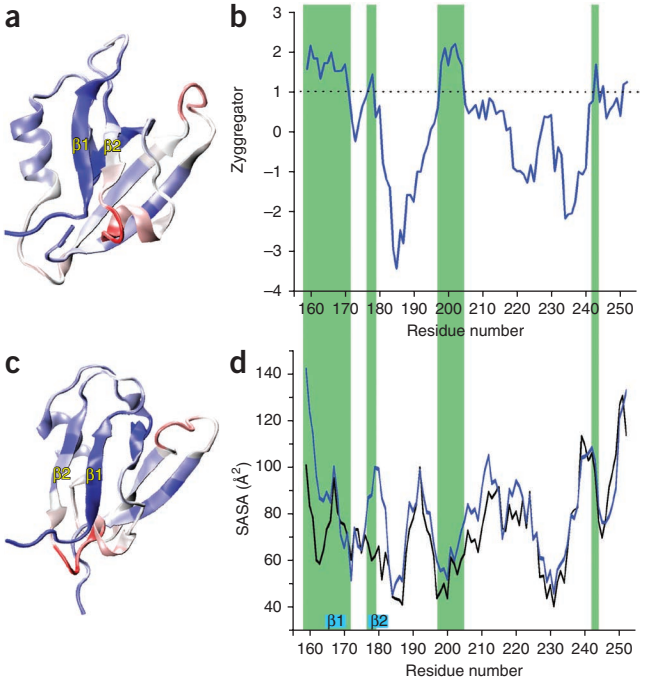

Figure 7 Comparison of the solvent exposure of the aggregation-prone regions of D1pPDZ in the native and in the misfolded intermediate case. (**a**) Native structure, which has been color coded according to the sequence-dependent aggregation propensity[45], which is shown in **b**; blue indicates highly aggregation-prone regions, and red indicates aggregation-resistant regions. (**b**) Sequence-dependent aggregation propensity of D1pPDZ; green shaded areas indicate regions that are highly aggregation prone. (**c**) Misfolded intermediate structure, color coded in the same manner as for the native state. (**d**) Comparison of the solvent-exposed surface area (SASA) between the native state (black line) and the misfolded intermediate (blue line); the aggregation-prone regions in the N-terminal part of the protein, which corresponds to the β1-β2 hairpin, are substantially more solvent exposed in the misfolded intermediate than in the native state.

to calculate structural ensembles representing TSI and TSN, starting from the structures of the intermediate and the native state, respectively (**Fig. 6**). In the calculated TSI structure, the alternative packing of the N-terminal β-hairpin (strands β1 and β2) on the native-like protein scaffold leads to the formation of stabilizing backbone and hydrophobic interactions similar in magnitude to the native ones. This arrangement is due to the symmetry of the backbone trace of β-hairpins that allows swapping the strands β1 and β2 without affecting the hydrogen-bonding network of the protein backbone (**Fig. 6**). However, such an alternative misfolded docking of the β-hairpin occurs while preserving the pattern of side chain packing, thereby keeping an optimized exposure of hydrophobic residues also in TSI (**Fig. 6**). These features indicate that the TSI structure is itself optimized, conserving favorable interactions similar to those evolutionarily selected for the native state. As expected from the starting conformation and the experimental data, the two ensembles have a high structural similarity, except in the region where the β-hairpin docks on the protein scaffold. Analysis of the contact maps provided an additional assessment of the structural features of the ensembles (**Fig. 6**). Here the occurrence of inter-residue contacts is reported for the two structured ensembles of TSI (**Fig. 6c**) and TSN (**Fig. 6d**). The results indicate that in TSI, stable, native-like interactions are formed not only between the regions corresponding to the native secondary elements (α1, α2, β3, β4 and β5) but also at the β1-β2 hairpin interface (red circles). However, although these interactions are highly conserved in both TSI and TSN, the β1-β2 hairpin in these two structures is interfaced with the rest of the protein in a completely different orientation (orange circles).

A comparison of the structural features of TSN with those of TS1 of PDZ2 from PTP-BL, a previously characterized canonical PDZ domain, suggests the presence of common features in the productive folding of PDZ domains at large. In particular, both proteins have high Φ values in the regions of α2, β3 and β5 (β4 and β6 in the canonical PDZ domain; see **Supplementary Table 1** and compare with Table 1 from ref. 28), which represent the early folding nucleus for both proteins. A similar behavior was also observed for the early transition state of PDZ3 from PSD95, whose folding was however characterized by the presence of an additional nucleus at the level of the interaction between β1 and β2 (β2 and β3 in canonical PDZ domains[25]).

DISCUSSION

In this work, we have characterized all the major states on the folding pathway of the small protein domain D1pPDZ, which is a natural permutant of the canonical PDZ domain, by using Φ-value analysis combined with restrained molecular dynamics simulations. The folding of this protein involves the presence of an off-pathway intermediate and two transition states, the first (TSI) leading from the denatured state to the intermediate and the second (TSN) from the denatured to the native state. In this protein, the direct pathway to the native state is in competition with the one populating the intermediate, which is misfolded and should thus unfold to enable productive folding.

Our results suggest that the misfolding intermediate observed for D1pPDZ is largely stabilized by non-native interactions concentrated in the ligand-binding region, leading to a misfolded structure characterized by an incorrect docking of the N-terminal β-hairpin with α2. The high-resolution structural information that we have presented of this misfolded state shows that the intermediate has mostly native-like features, such as formation of the main β-sheet and partial formation of the α-helices, but is less compact than the native state and characterized by a crucial alternative docking of the N-terminal β-hairpin (**Fig. 4**). In the context of the free energy landscape theory[42], folding takes place over a funneled landscape, and folding pathways are robust against mutations. However, the distinct demands of folding and function appear to trigger conflicting requirements of the amino acid sequence of D1pPDZ, and the formation of non-native interactions between residues making up the ligand-binding pocket in the native state are the main cause for the formation of the misfolded intermediate. We therefore suggest that functional demands involved in shaping the ligand-binding pocket may compromise an otherwise robust folding process. This view is in agreement with recent theoretical studies showing that highly frustrated regions often correspond to those binding other macromolecules or small ligands[43,44].

The formation of the misfolded intermediate may further compromise the reliability of the folding process of D1pPDZ, because such a non-native conformation could be less protected against aggregation than the native state. This expectation is supported by the analysis of the sequence-based aggregation propensity[45] of D1pPDZ, which suggests that the aggregation-prone regions in the N-terminal part of the protein are substantially more solvent exposed in the misfolded intermediate than in the native state (**Fig. 7**). By contrast, the other aggregation-prone regions in the sequence of D1pPDZ (green shaded areas in **Fig. 7b**,**d**) have essentially the same degree of solvent exposure in the native state as in the misfolded intermediate.

The results presented in this work provide a structural perspective to assess previous work on the folding of the PDZ domain family.

Folding studies on sequence-divergent variants of the PDZ family have revealed that a number of canonical variants share a common three-state mechanism comprising an on-pathway folding intermediate[26,30]. Furthermore, Φ-value analysis of the two three-state folders PDZ2 of PTP-BL and PDZ3 of PSD-95 have demonstrated that although the early folding events are clearly distinct, the folding pathways converge at the late transition state[25]. In the present study we have observed that the structure of the productive folding transition state of the naturally occurring circularly permuted D1pPDZ clearly resembles TS1 of PTP-BL PDZ2, despite the alternative connectivity of secondary structure elements and low sequence identity. This observation illustrates the importance of the native state structure in guiding the folding pathway.

Studies of topological variants, such as engineered circularly permuted proteins, have demonstrated that loop entropy perturbations may redirect the folding to alternative nuclei when such quasi-species are available[46–49]. Furthermore, theoretical studies have highlighted the significance of a symmetric distribution of secondary structure elements along the sequence in promoting the presence of parallel folding pathways[50]. Notably, the D1pPDZ has a symmetric linear organization of the α-helical and β-sheet elements, and it is plausible that this configuration may promote alternative folding pathways. This conclusion is supported by our recent work on a circularly permuted variant of PTP-BL PDZ2 (cpPDZ2) that shares the sequence connectivity of D1pPDZ. Indeed, similar to what is observed for D1pPDZ, in the case of cpPDZ2, an intermediate is involved in the folding process[16,21]. It thus appears as if the circularly permuted PDZ domains are more prone to folding through stable intermediates, and thus show a reduced folding cooperativity. On the basis of our results, we may speculate that the symmetric connectivity of circularly permuted PDZ domains facilitates the formation of alternative folding nuclei, which in the case of D1pPDZ leads to a misfolded intermediate. Misfolding would thus result from a combination of frustrated interactions in the ligand-binding region[43,44] and a symmetric sequence connectivity that allows for alternative nucleation[50].

We have presented the structure of all the major states populated during the folding process of the circularly permuted PDZ domain of D1p, and characterized in particular the structure of its misfolded intermediate. These results were achieved by conducting an extensive Φ-value analysis to map the interactions formed in the different transient species and by using the resulting values as structural restraints in molecular dynamics simulations. Our findings reveal that although the productive folding of D1pPDZ is similar to that observed in another canonical PDZ domain (PDZ2 from PTP-BL), the off-pathway intermediate, whose topology largely resembles that of the native state, is less compact and shows a misfolded docking of the N-terminal β-hairpin onto an otherwise native-like scaffold. These results suggest that misfolded intermediates may often be quite compact, rather than substantially unstructured and highly dynamic, and may contain misassembled structural elements along with more extended correctly folded regions .

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

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

ACKNOWLEDGMENTS

We acknowledge financial support from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (2007B57EAB_004, 20074TJ3ZB_005, RBRN07BMCT_007) (S.G., Y.I., C.T.-A., M.B.), the Wenner Green Foundations (Y.I.), the Istituto Pasteur-Fondazione Cenci Bolognetti (Y.I.), European Molecular Biology Organization (Y.I.), Engineering and Physical Sciences Research Council (A.D.S.), Biotechnology and Biological Sciences Research Council (M.V.) and the Royal Society (M.V.).

AUTHOR CONTRIBUTIONS

S.G., Y.I., C.T.-A. and M.B. conceived and designed the experimental work; A.D.S. and M.V. conceived and designed the computational work; Y.I. expressed and purified the protein samples; S.G. and Y.I. conducted the experiments; A.D.S. did the simulations; S.G., Y.I., A.D.S., C.T.-A., M.B. and M.V. analyzed the data and wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Fersht, A.R. *Structure and Mechanism in Protein Science* (Freeman, New York, 1999).
2. Selkoe, D.J. Folding proteins in fatal ways. *Nature* **426**, 900–904 (2003).
3. Chiti, F. & Dobson, C.M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).
4. Frydman, J. Folding of newly translated proteins *in vivo*: the role of molecular chaperones. *Annu. Rev. Biochem.* **70**, 603–647 (2001).
5. Hartl, F.U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
6. Bukau, B., Weissman, J. & Horwich, A. Molecular chaperones and protein quality control. *Cell* **125**, 443–451 (2006).
7. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **4**, 181–191 (2003).
8. Goldberg, A.L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–899 (2003).
9. Chiti, F. & Dobson, C.M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5**, 15–22 (2009).
10. Elad, N. *et al.* Topologies of a substrate protein bound to the chaperonin GroEL. *Mol. Cell* **26**, 415–426 (2007).
11. Hartl, F.U. & Hayer-Hartl, M. Converging concepts of protein folding *in vitro* and *in vivo*. *Nat. Struct. Mol. Biol.* **16**, 574–581 (2009).
12. Horwich, A.L. & Fenton, W.A. Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q. Rev. Biophys.* **42**, 83–116 (2009).
13. Matouschek, A., Kellis, J.T. Jr., Serrano, L., Bycroft, M. & Fersht, A.R. Transient folding intermediates characterized by protein engineering. *Nature* **346**, 440–445 (1990).
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. Friel, C.T., Smith, D.A., Vendruscolo, M., Gsponer, J. & Radford, S.E. The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat. Struct. Mol. Biol.* **16**, 318–324 (2009).
16. Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M. & Gianni, S. Engineered symmetric connectivity of secondary structure elements highlights malleability of protein folding pathways. *J. Am. Chem. Soc.* **131**, 11727–11733 (2009).
17. Korzhnev, D.M., Religa, T.L., Banachewicz, W., Fersht, A.R. & Kay, L.E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).
18. Religa, T.L., Markson, J.S., Mayor, U., Freund, S.M. & Fersht, A.R. Solution structure of a protein denatured state and folding intermediate. *Nature* **437**, 1053–1056 (2005).
19. Wu, Y., Vadrevu, R., Kathuria, S., Yang, X.Y. & Matthews, C.R. A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the alpha subunit of tryptophan synthase, a TIM barrel protein. *J. Mol. Biol.* **366**, 1624–1638 (2007).
20. Nabuurs, S.M., Westphal, A.H. & van Mierlo, C.P.M. Noncooperative formation of the off-pathway molten globule during folding of the alpha-beta parallel protein apoflavodoxin. *J. Am. Chem. Soc.* **131**, 2739–2746 (2009).
21. Ivarsson, Y., Travaglini-Allocatelli, C., Morea, V., Brunori, M. & Gianni, S. The folding pathway of an engineered circularly permuted PDZ domain. *Protein Eng. Des. Sel.* **21**, 155–160 (2008).
22. Morais Cabral, J.H. *et al.* Crystal structure of a PDZ domain. *Nature* **382**, 649–652 (1996).
23. Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M. & Clausen, T. Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* **416**, 455–459 (2002).
24. Liao, D.I., Qian, J., Chisholm, D.A., Jordan, D.B. & Diner, B.A. Crystal structures of the photosystem II D1 C-terminal processing protease. *Nat. Struct. Biol.* **7**, 749–753 (2000).

25. Calosci, N. *et al.* Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc. Natl. Acad. Sci. USA* **105**, 19241–19246 (2008).
26. Chi, C.N. *et al.* A conserved folding mechanism for PDZ domains. *FEBS Lett.* **581**, 1109–1113 (2007).
27. Gianni, S. *et al.* Kinetic folding mechanism of PDZ2 from PTP-BL. *Protein Eng. Des. Sel.* **18**, 389–395 (2005).
28. Gianni, S. *et al.* A PDZ domain recapitulates a unifying mechanism for protein folding. *Proc. Natl. Acad. Sci. USA* **104**, 128–133 (2007).
29. Ivarsson, Y. *et al.* An on-pathway intermediate in the folding of a PDZ domain. *J. Biol. Chem.* **282**, 8568–8572 (2007).
30. Jemth, P. & Gianni, S. PDZ domains: folding and binding. *Biochemistry* **46**, 8701–8708 (2007).
31. Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M. & Gianni, S. Folding and misfolding in a naturally occurring circularly permuted PDZ domain. *J. Biol. Chem.* **283**, 8954–8960 (2008).
32. Fersht, A.R., Matouschek, A. & Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782 (1992).
33. Fersht, A.R. & Sato, S. Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. USA* **101**, 7976–7981 (2004).
34. Mayor, U. *et al.* The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**, 863–867 (2003).
35. Li, L., Mirny, L.A. & Shakhnovich, E.I. Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. *Nat. Struct. Biol.* **7**, 336–342 (2000).
36. Ozkan, S.B., Bahar, I. & Dill, K.A. Transition states and the meaning of Phi-values in protein folding kinetics. *Nat. Struct. Biol.* **8**, 765–769 (2001).
37. Ventura, S. *et al.* Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat. Struct. Biol.* **9**, 485–493 (2002).
38. Paci, E., Vendruscolo, M., Dobson, C.M. & Karplus, M. Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol.* **324**, 151–163 (2002).
39. Vendruscolo, M., Paci, E., Dobson, C.M. & Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **409**, 641–645 (2001).
40. 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).
41. Salvatella, X., Dobson, C.M., Fersht, A.R. & Vendruscolo, M. Determination of the folding transition states of barnase by using Phil-value-restrained simulations validated by double mutant PhiIJ-values. *Proc. Natl. Acad. Sci. USA* **102**, 12389–12394 (2005).
42. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).
43. Ferreiro, D.U., Hegler, J.A., Komives, E.A. & Wolynes, P.G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. USA* **104**, 19819–19824 (2007).
44. Sutto, L., Lätzer, J., Hegler, J.A., Ferreiro, D.U. & Wolynes, P.G. Consequences of localized frustration for the folding mechanism of the IM7 protein. *Proc. Natl. Acad. Sci. USA* **104**, 19825–19830 (2007).
45. Tartaglia, G.G. *et al.* Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* **380**, 425–436 (2008).
46. Hubner, I.A., Lindberg, M., Haglund, E., Oliveberg, M. & Shakhnovich, E.I. Common motifs and topological effects in the protein folding transition state. *J. Mol. Biol.* **359**, 1075–1085 (2006).
47. Lindberg, M., Tangrot, J. & Oliveberg, M. Complete change of the protein folding transition state upon circular permutation. *Nat. Struct. Biol.* **9**, 818–822 (2002).
48. Lindberg, M.O. & Oliveberg, M. Malleability of protein folding pathways: a simple reason for complex behaviour. *Curr. Opin. Struct. Biol.* **17**, 21–29 (2007).
49. Lindberg, M.O. *et al.* Folding of circular permutants with decreased contact order: general trend balanced by protein stability. *J. Mol. Biol.* **314**, 891–900 (2001).
50. Klimov, D.K. & Thirumalai, D. Symmetric connectivity of secondary structure elements enhances the diversity of folding pathways. *J. Mol. Biol.* **353**, 1171–1186 (2005).

ONLINE METHODS

Mutagenesis, expression and purification. Mutants were made using a QuikChange site-directed mutagenesis kit (Stratagene) and the synthetic gene encoding pseudo–wild type D1pPDZ as template. This construct encodes for residues 159–253 of D1p protease and was expressed and purified as described in reference 31. The mutations were confirmed by DNA sequencing and the purity of the proteins was checked by SDS-PAGE. All reagents were of analytical grade.

Binding assays. A peptide mimicking the natural substrate of D1pPDZ (EAPSVNA) was purchased from JPT. Equilibrium binding experiments were carried out by exciting the tryptophan at 280 nm and measuring the fluorescence emission of the protein sample, at a 5 μM protein concentration, in the presence of different peptide concentrations, ranging from 1 μM to 2 mM.

Kinetic experiments. All kinetic experiments were conducted on a Pi-star stopped-flow apparatus (Applied Photophysics). (Un)folding was followed by stopped-flow experiments in which the protein sample was excited at 280 nm and the change in fluorescence was followed using a 360-nm cutoff filter. Using this cutoff filter, the amplitudes of the observed phases are of opposite sign[31], allowing unequivocal distinction between the misfolding and the productive folding events. Dilution of denatured or native protein by a factor of 11 in appropriate buffer initiated refolding and unfolding . Final protein concentrations were typically in the range of 2–5 μM . As described in reference 31, (un)folding of D1pPDZ was always independent of protein concentration. All kinetic experiments were done in 50 mM sodium phosphate buffer (pH 7.2) containing 0.4 M sodium sulfate and at 25 °C. Biphasic refolding and unfolding traces were observed for 23 mutants. The combined chevron plots, obtained by analyzing the urea dependence of the two observed rate constants, were adequately fitted to the analytical solution of the quadratic equation describing folding with an off-pathway intermediate, as previously described[51]. Owing to the complexity of the kinetic model, the *m* values for the microscopic rate constants of all mutants were assumed to be the same as those determined for wild-type D1pPDZ[31]. Ten mutants showed single-exponential (un)folding kinetics, and the kinetic data for these mutants were fitted to a two-state folding mechanism[52]. The $\Delta\Delta G$ and Φ-values for each mutant were calculated from kinetic parameters using the following equations:

$$\Delta\Delta G_{D-N}^{wt-mut} = RT \cdot \ln\left(\frac{k_{D-N}^{wt} \cdot k_{N-D}^{mut}}{k_{N-D}^{wt} \cdot k_{D-N}^{mut}}\right) \quad (1)$$

$$\Delta\Delta G_{D-TSN}^{wt-mut} = RT \cdot \ln\left(\frac{k_{D-N}^{wt}}{k_{D-N}^{mut}}\right) \quad (2)$$

$$\Phi_{TSN} = \frac{\Delta\Delta G_{D-TSN}^{wt-mut}}{\Delta\Delta G_{D-N}^{wt-mut}} \quad (3)$$

$$\Delta\Delta G_{D-I}^{wt-mut} = RT \cdot \ln\left(\frac{k_{D-I}^{wt} \cdot k_{I-D}^{mut}}{k_{I-D}^{wt} \cdot k_{D-I}^{mut}}\right) \quad (4)$$

$$\Phi_{I} = \frac{\Delta\Delta G_{D-I}^{wt-mut}}{\Delta\Delta G_{D-N}^{wt-mut}} \quad (5)$$

$$\Delta\Delta G_{D-TSI}^{wt-mut} = RT \cdot \ln\left(\frac{k_{D-I}^{wt}}{k_{D-I}^{mut}}\right) \quad (6)$$

$$\Phi_{TSI} = \frac{\Delta\Delta G_{D-TSI}^{wt-mut}}{\Delta\Delta G_{D-N}^{wt-mut}} \quad (7)$$

where *k*D–I and *k*I–D are the forward and reverse microscopic rate constants for the reaction leading to the intermediate and *k*D–N and *k*N–D are the forward and reverse microscopic rate constants for productive folding to the native state.

Molecular dynamics simulations. Molecular dynamics simulations of TSN were started from the atomic-resolution (2.0-Å) X-ray structure of PDZ (PDB code 1FC7, fragment 159–253). For calculating the ensemble TSI, the starting model was generated by docking the swapped N-terminal β-hairpin (strands β1 and β2), optimizing the residues packing in the cleft, and equilibrating by a 50-ns molecular dynamics simulation in explicit solvent (using the CHARMM22 force field, NPT ensemble, 300 K and 1 atm). To compute the transition state ensembles, experimental Φ values have been treated as restraints in molecular dynamics simulations. In this method, Φ values are defined as[38,39]

$$\Phi_r^{calc}(t) = \frac{N_r(t)}{N_r^{nat}} \quad (8)$$

where the Φ value of residue r is modeled as the ratio between the number of native contacts, $N_r(t)$ (heavy side chain atom contacts within a cutoff distance of 6.5 Å), at time t during the simulation, and the number of native contacts, N_r^{nat}, of residue *r*. These are estimated from an unrestrained 50 ns of molecular dynamics simulation (TIP3P explicit solvent, NPT ensemble, 300 K and 1 atm) by selecting contacts presenting an occurrence in the simulation of 0.8 or more[38,40,41]. Because our definition of Φ_r^{calc} is not compatible with Φ values larger than 1, experimental Φ values larger than 1 have been excluded from the dataset used in the simulations. The restrained simulations are carried out by using a total energy[38,39], E_{tot}, in which a restraining energy term, E_{res}, is added to the conventional force field, E_{ff}, so that the total energy is $E_{tot} = E_{ff} + \alpha E_{res}$, where α is a weighting factor. The additional term is enforced by minimizing the discrepancy between calculated and experimental data[38,39]

$$E_{res} = \sum_r \left(\Phi_r^{exp} - \Phi_r^{calc}\right)^2 \quad (9)$$

Restrained molecular dynamics simulations have been conducted with the CHARMM program by using the CHARMM22 force field and an implicit model of the solvent[38,40,41]. All calculations used an atom-based truncation scheme with a list cutoff of 14 Å, a non-bonded cutoff of 12 Å, and the Lennard-Jones smoothing function initiated at 10 Å[38,40,41]. Electrostatic and Lennard-Jones interactions were force-switched. Covalent bonds were constrained with SHAKE[38,40,41]. The structure calculations consisted of two parts, a preparation part followed by cycles of simulated annealing. An initial equilibration simulation at 298 K was run, during which the calculated results were allowed to converge on the experimental results. This was achieved by gently raising the restraint force constant. The initial velocities were randomly assigned from a Maxwell-Boltzmann distribution at 298 K. Subsequently, a series of 100 cycles of simulated annealing between 298 and 498 K were carried out to effectively sample the conformational space. Each cycle was carried out for 150 ps by using an integration step of 1 fs. The first 20 cycles were not considered for the analysis.

51. Jemth, P. *et al.* Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc. Natl. Acad. Sci. USA* **101**, 6450–6455 (2004).
52. Jackson, S.E. & Fersht, A.R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428–10435 (1991).

doi:10.1038/nsmb.1956