

Visualizing transient protein-folding intermediates by tryptophan-scanning mutagenesis

Alexis Vallée-Bélisle^{1,3} & Stephen W Michnick^{1,2}

To understand how proteins fold, assemble and function, it is necessary to characterize the structure and dynamics of each state they adopt during their lifetime. Experimental characterization of the transient states of proteins remains a major challenge because high-resolution structural techniques, including NMR and X-ray crystallography, cannot be directly applied to study short-lived protein states. To circumvent this limitation, we show that transient states during protein folding can be characterized by measuring the fluorescence of tryptophan residues, introduced at many solvent-exposed positions to determine whether each position is native-like, denatured-like or non-native-like in the intermediate state. We use this approach to characterize a late-folding-intermediate state of the small globular mammalian protein ubiquitin, and we show the presence of productive non-native interactions that suggest a ‘flycatcher’ mechanism of concerted binding and folding.

A protein can adopt many structures, from unfolded, semifolded and misfolded states to a correctly folded state that can display various functionally relevant conformations and oligomeric states^{1–4}. Studies of these intermediate states are crucial to our ability to understand how proteins have evolved to assemble into functional complexes and to develop realistic and predictive *in silico* molecular models^{2,3,5–7}. It is difficult to characterize the structures of transient intermediate states of proteins, particularly folding intermediates, because their transience makes high-resolution techniques such as NMR⁸ or X-ray crystallography⁹ unsuitable, whereas more sensitive methods, such as fluorescence, CD or infrared spectroscopy, lack structural resolution. Having more efficient approaches to characterize protein folding intermediates and other transitory states would therefore help address key questions including: Do proteins fold via similar intermediate states and via similar rate-limiting steps^{10–12}? What is the role of non-native interactions (that is, interactions that are not present in the final folded structure) during protein folding; can they stabilize productive folding intermediate states and accelerate folding¹³, or do they simply add errors or roughness in the folding landscape due to functional constraints^{14–16}?

Over the last 50 years, measuring the fluorescence intensity of tryptophan has been the method of choice for detection and characterization of short-lived intermediate states of proteins^{1,17,18}. For example, folding studies on small proteins typically use the intrinsic fluorescence of naturally contained tryptophan amino acids to monitor protein refolding and use deletion mutations of other amino acids (such as valine to alanine) to identify the native interactions that contribute the most to folding kinetics¹. In the early 1990s, Smith *et al.* pioneered the use of different single-tryptophan mutants, with tryptophan substitutions for phenylalanine or tyrosine, to gain insight into the structures of intermediate states that are populated at equilibrium¹⁹. More recently, we also demonstrated that multiple single-tryptophan mutants can enable determination of whether transient folding intermediates are on- or off-pathway²⁰. But the full potential of using tryptophan as a probe to deduce transient-state structures has yet to be realized²¹.

Here, we describe a tryptophan-scanning strategy to characterize the structure of a transient intermediate (I) state (Fig. 1). Our strategy uses the fluorescence of individual nonperturbing tryptophans to detect the formation of native or non-native structures in transient intermediates that accumulate during protein folding (Fig. 1, Supplementary Fig. 1a). In order to minimize native-structure perturbation and assess the folding of the protein over its entire structure, we first constructed, by site-directed mutagenesis, a library of single-tryptophan mutants, with tryptophans substituted for amino acids that have solvent-exposed side chains. Substitution of large amino acids at such positions has been shown not to substantially affect the folding landscape²². Furthermore, these sites are ideal probe locations for detection of non-native structure formation because they are likely to remain solvent exposed in both the unfolded and folded states. The tryptophan-scanning strategy works by determining whether a tryptophan probe at the depicted site x (or y and z) produces unfolded (U)-like, native (N)-like, or non-native-like fluorescence in each state (Fig. 1a). For example, the fluorescence of tryptophan x should be mostly at N-like levels in the I state because the local environment around x is already in an N-like conformation after the first transition (k_1). In contrast, the fluorescence of tryptophan y should remain mostly at unfolded-like levels in the I state because the local environment only achieves its N-like structure during the second

¹Département de Biochimie, Université de Montréal, Montréal, Québec, Canada. ²Centre Robert-Cedergren en Bio-Informatique et Génomique, Université de Montréal, Montréal, Québec, Canada. ³Present address: Department of Chemistry and Biochemistry, University of California, Santa Barbara, California, USA. Correspondence should be addressed to S.W.M. (stephen.michnick@umontreal.ca).

Received 28 February; accepted 11 May; published online 10 June 2012; doi:10.1038/nsmb.2322

Figure 1 A tryptophan-scanning strategy to characterize the transient intermediate states of proteins. (a) The tryptophan-scanning strategy uses the fluorescence of many nonperturbing, single-tryptophan substitutions to detect the formation of native and nonnative structures at different locations of the surface of a protein as it folds. The fluorescence of a tryptophan in the unfolded, intermediate and native states (F_U , F_I and F_N) can be determined at various solvent-exposed locations (that is, non-perturbing tryptophan substitution sites) through a simple modeling of the folding kinetic traces of distinct single tryptophan mutants (three different tryptophan positions are shown). (b) F_I , the relative fluorescence intensity of a tryptophan at a specific location in the I state, should be similar to F_N if the local environment around this tryptophan already reached its native conformation in I (Trp x; see purple locations). F_I should be similar to F_U if the local environment around the tryptophan did not significantly change following the first transition (Trp y; see red location), and it should be distinct from both F_U and F_N if the local environment formed nonnative interactions in the intermediate (Trp z; see yellow location).

transition (k_2). Finally, the fluorescence of tryptophan z is likely to display distinct fluorescence levels in the I state compared to those in the unfolded and folded states as a result of the non-native interactions made by residue z. By determining the relative fluorescence of many tryptophan mutants, one can therefore obtain structural insights about intermediates (**Fig. 1b**).

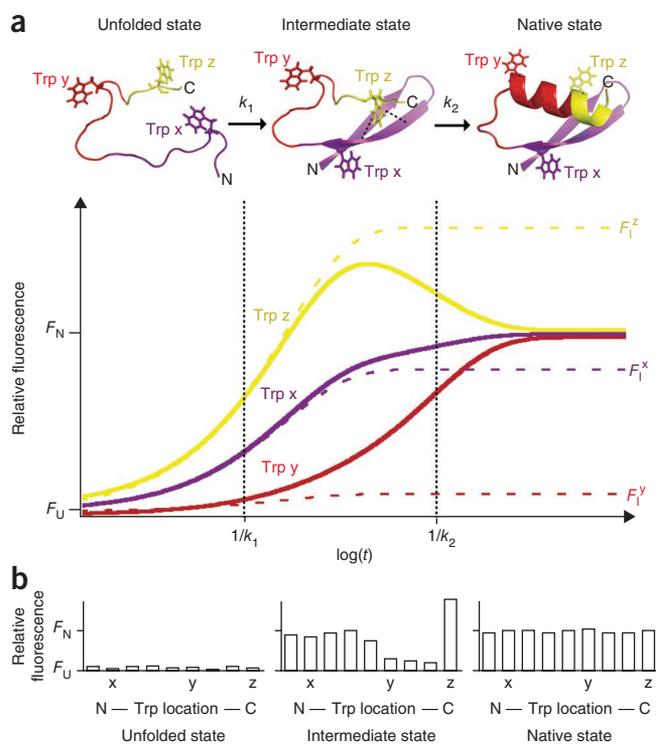
In the present work, we apply the tryptophan-scanning strategy to probe native and non-native structure formation during the folding of mammalian ubiquitin, a well-studied small protein²³ involved in several cellular processes²⁴ (**Fig. 2a**). Ubiquitin is ideally suited for a tryptophan-scanning analysis because it lacks a natural tryptophan and its folding mechanism is disputed, though it has been extensively studied using many approaches (**Supplementary Fig. 1b**)^{20,25–30}. Recently, a late on-pathway intermediate (I_L) has been shown to be highly populated during folding of ubiquitin (~80% of ubiquitin population at 10 ms; **Fig. 2a**)^{20,31}. On the other hand, the accumulation of an early folding intermediate during its refolding remains highly controversial (**Supplementary Fig. 1b**).

RESULTS

Solvent-exposed tryptophans as fluorescent probes

We engineered and expressed 27 different single-tryptophan mutants of ubiquitin, with tryptophan located at 75% of the most solvent-accessible positions (**Fig. 2a**). We first monitored the equilibrium unfolding transitions of all 27 single-tryptophan ubiquitin mutants using tryptophan fluorescence spectroscopy, with guanidine-HCl as a chemical denaturant (**Fig. 2b**). In ~80% (22 of 27) of the mutants, solvent-exposed tryptophan produced higher fluorescence intensities in the native than in the unfolded state (**Supplementary Fig. 2**). All tryptophans were sensitive to only a single unfolding transition centered around 3.2 M guanidine-HCl (**Fig. 2b** and **Supplementary Fig. 2**), suggesting that no I state is significantly populated at equilibrium upon chemical denaturation of ubiquitin^{20,27,30,32,33}. The unfolding free energies of all the single-tryptophan mutants, extrapolated from a two-state fit of the equilibrium curve, were found to be narrowly distributed around the value obtained for wild-type ubiquitin (average 28 ± 3 kJ mol⁻¹, as compared to 28.0 kJ mol⁻¹ for wild type³²; **Fig. 2c**). These results demonstrate that tryptophan substitution at solvent-exposed locations is minimally perturbing compared to, for example, typical mutations used in ϕ -value analysis (**Fig. 2c**)^{1,33}.

We then monitored the folding kinetics of the 27 tryptophan mutants using guanidine-HCl dilution-jump, stopped-flow experiments in order to monitor the formation of the various structural environments around each of our tryptophan probes (**Fig. 3a**). All of the tryptophan probes were sensitive to only two structural

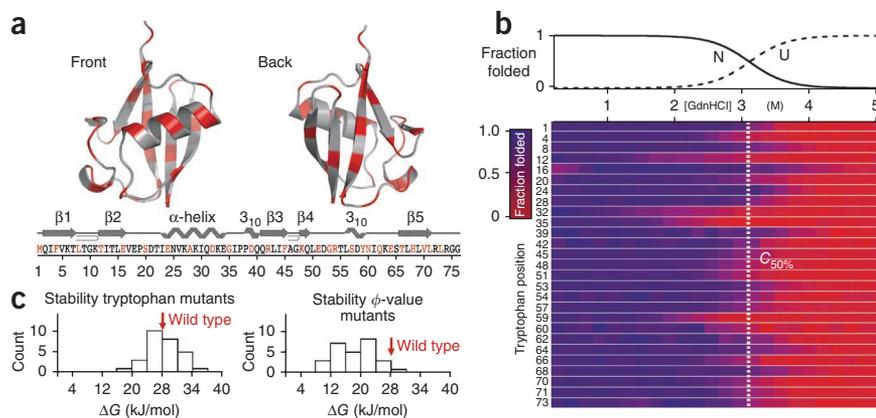


transitions at low denaturant concentration (**Supplementary Fig. 3a**), with rate constants similar to those for the formation of the I_L (k_1 : 200 ± 77 s⁻¹) and the native state (k_2 : 24 ± 13 s⁻¹; **Fig. 3b**, left)²⁰. Of the 27 tryptophan probes, 23 were sensitive to both transitions (exhibiting biexponential refolding curves), whereas T66W and L71W were sensitive only to the second transition (T_2) and D32W and D39W were sensitive only to the first transition (T_1) (**Supplementary Fig. 3**). Only mutant E16W showed significant changes in the folding kinetics (faster k_1 ; **Fig. 3b**), again highlighting the robustness of tryptophan substitutions at solvent-exposed positions, with regard to protein folding^{20,22}.

To simplify the visualization of the refolding traces, we employed a simple color-scale gradient to illustrate the fluorescence values relative to N-like (purple) and U-like (red) levels. (Levels between N- and U-like are pink.) We also represented non-native-like fluorescence as a different color (yellow) when the fluorescence level was higher or lower than both the U- or N-like fluorescence. We found that most tryptophans in the N-terminal region of ubiquitin reach their N-like fluorescence level (purple) in the I_L state (~10 ms), whereas tryptophans located in the 57–66 segment still fluoresce mostly in a U-like manner (red; **Fig. 3a**). Notably, tryptophans in the β -sheet at positions 1 (β_1), 12 (β_2), 42 (β_3), and 73 (β_5) displayed much higher, non-native-like fluorescence intensities in the I_L compared to both the U and N states (yellow). To confirm that these different relative tryptophan fluorescence levels in the I_L state are attributable to the presence of non-native structural environments, rather than to changes in the folding pathway induced by the tryptophan substitution, we also performed additional refolding experiments on selected tryptophan mutants, using an independent fluorescent probe (**Fig. 3c**). Thus, we monitored the refolding kinetics of M1W, A28W, F45W, S57W and T66W by using the structure-sensitive fluorescent dye 8-anilino-1-naphthalenesulfonate (ANS)³⁴, which is known to bind at hydrophobic pockets in transient intermediate states. We confirmed that all traces displayed identical biexponential profiles (**Fig. 3c**)²⁰.

We then performed refolding kinetic experiments over a large range of denaturant concentrations to generate equilibrium folding-unfolding

Figure 2 Tryptophan substitution at solvent-exposed sites on a protein surface minimally perturbs protein stability and serves as a sensitive probe to monitor folding and unfolding of a protein. (a) Ribbon representation of ubiquitin structure ($\beta\beta\alpha\beta\alpha\beta$ topology) showing the 27 solvent-exposed positions (red) where residues are mutated to tryptophan in this study (PDB code: 1UBQ)⁵¹. (b) Equilibrium unfolding curves of each single tryptophan mutant monitored by tryptophan fluorescence (emission above 320 nm; see **Supplementary Fig. 2** for raw data). All tryptophan mutants displayed a single unfolding transition located around 3.2 M guanidine-HCl (GdnHCl)³². (c) Similar extrapolated unfolding free energies (on average: 28.2 ± 3.2 kJ mol⁻¹) and degrees of cooperativity (m values: 9.0 ± 1.2 kJ mol⁻¹ M⁻¹) are observed for each mutant (**Supplementary Table 1**). In comparison, stability perturbing (or deletion) mutants used for the ϕ -value analysis²⁵ destabilized ubiquitin by -9.4 ± 4.5 kJ mol⁻¹ on average (red arrow indicates unfolding free energy of wild type).



curves of the I_L and of the putative early intermediate (**Supplementary Fig. 4**). By plotting the fluorescence values obtained after T_1 (which are proportional to the I_L population) as a function of the denaturant concentration, we were able to directly model the $I_L \rightarrow U$ unfolding equilibrium transition for ten tryptophan mutants (**Supplementary Fig. 4** and **Supplementary Table 1**). Additional global fit of both the equilibrium and kinetic parameters of the tryptophan mutants using a three-state, on-pathway folding model^{14,20} also provided an estimate of the I_L stability for 25 of the 27 tryptophan mutants (**Supplementary Fig. 4**). Overall, we found that the stability of the I_L was little affected

by tryptophan mutations (average $\Delta G_{U-I} = 13 \pm 2$ kJ mol⁻¹), suggesting again that the folding pathway of ubiquitin is relatively insensitive to tryptophan substitutions at native solvent-exposed sites (**Fig. 3b**, right). Finally, we assessed whether a fast structural transition takes place (that is, if an early folding intermediate accumulates) during the dead time of our stopped-flow instrument (**Supplementary Fig. 3b**). To do so, we compared the tryptophan fluorescence intensities extrapolated at the initiation time of the refolding reactions (F_U) to the value F_0 obtained for free tryptophan in solution (**Supplementary Fig. 3b,c**). Notably, all tryptophans showed significantly higher fluorescence intensities following the rapid transfer of the protein from a 5 M to a 0.45 M guanidine-HCl concentration (black bars) relative to the value obtained for free tryptophan (**Supplementary Fig. 3c**). However, this small upward deviation rapidly disappeared as the denaturant concentration increased, perhaps reflecting a nonspecific compaction of the unfolded protein rather than the accumulation of a structurally specific intermediate³⁵.

Fluorescence-based representation of the I_L using ω values

We then used fluorescence intensities for the 27 tryptophan probes to characterize the structure of the I_L (**Fig. 4**). The fluorescence intensity of tryptophan is highly sensitive to the characteristics of the local

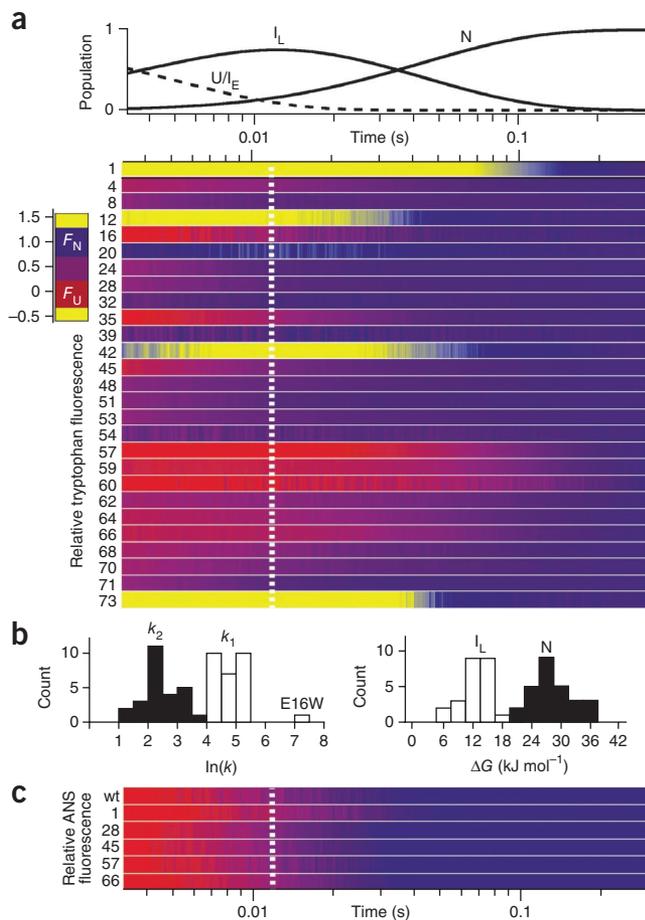
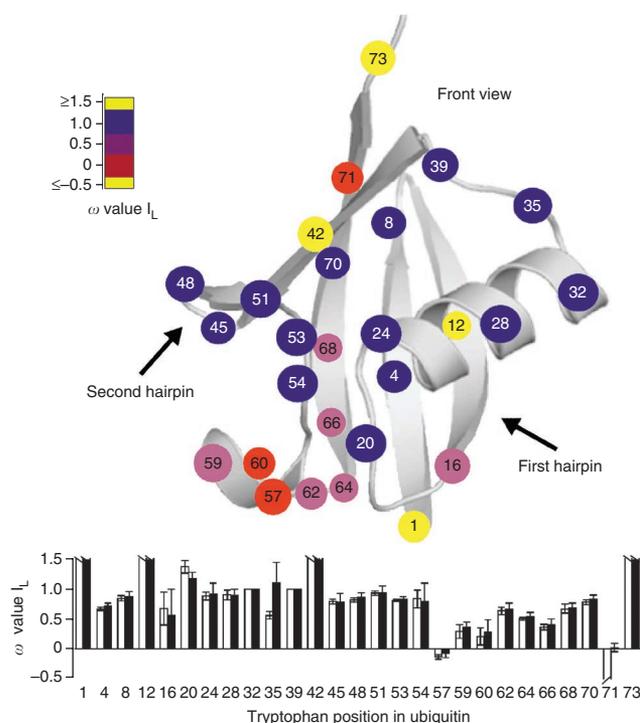


Figure 3 Folding kinetics of ubiquitin probed using 27 solvent-exposed tryptophan mutants. (a) Folding kinetics of each tryptophan mutant were monitored by tryptophan fluorescence spectroscopy ($\lambda_{\text{excitation}} = 281$ nm; $\lambda_{\text{emission}} \leq 320$ nm). Results are expressed relative to the fluorescence of the unfolded (U) and native (N) states ($F_U = 0$; $F_N = 1$; see **Supplementary Figs. 3** and **4** for analysis of kinetic curves). These folding traces are represented in a color code that uses red and purple to represent the fluorescence intensities of the U and N states, respectively; a pink gradient to represent fluorescence intensities between the U- and N-like states; and yellow to represent non-native fluorescence intensities, which are neither U or N like. (b) Left, 23 of the 27 tryptophans were sensitive to two refolding transitions at low denaturant concentrations (biexponential fit), with average rates (k_1 : 202 ± 77 s⁻¹ and k_2 : 24 ± 13 s⁻¹) similar to those for the formation of I_L and N, respectively (**Supplementary Table 1**)²⁰. Right, stability of the I_L for all tryptophan mutants as determined by a global fitting analysis of the folding traces (see **Supplementary Fig. 4**) (average $\Delta G_{U-I} = 13 \pm 2$ kJ mol⁻¹). (c) Folding kinetics of wild-type (wt) ubiquitin and five representative tryptophan mutants that display distinct relative tryptophan fluorescence in the I_L state show identical biphasic kinetics profiles (k_1 : 136 ± 15 s⁻¹; k_2 : 20 ± 2 s⁻¹)²⁰ when monitored using the fluorescence of the external dye ANS ($\lambda_{\text{excitation}} = 350$ nm; $\lambda_{\text{emission}} \leq 395$ nm)³⁴, which is known to bind hydrophobic pockets in transient intermediate states.

Figure 4 Fluorescence-based representation of the I_L using ω values. ω value = $\Delta F_{I-U}(i)/\Delta F_{N-U}(i)$, where $\Delta F_{I-U}(i)$ represents the difference in fluorescence between the I and U states for tryptophan i , and $\Delta F_{N-U}(i)$ represents the difference in fluorescence between the N and U state for tryptophan i . Thus ω value provides a measure of whether a tryptophan probe at position i in the I state has a fluorescence that is N or U like ($\omega = \sim 1$ or $\omega = \sim 0$, respectively) or whether it is distinct from these two states ($\omega \leq -0.25$ or $\omega \geq 1.25$). White bars, ω value determined using $\Delta F_{I-U}(i)$ and $\Delta F_{N-U}(i)$ obtained from the folding trace at 0.45 M guanidine-HCl (**Supplementary Fig. 3**). Black bars, ω value determined using $\Delta F_{I-U}(i)$ and $\Delta F_{N-U}(i)$ obtained from the simultaneous global fit of all equilibrium and kinetic parameters (**Supplementary Fig. 4**). Residue 71 has a low ω value (red) despite adopting its native-state fluorescence quite rapidly as ubiquitin refolds (see **Fig. 3a**). This is because tryptophan insertion at this location also increases the rate of rearrangement of the intermediate (see **Fig. 5a**, right). s.d. were obtained from the best fit of the data.

structural environment, such as the proximity to charges and the polarity (for example, whether a tryptophan is surrounded by water or by hydrophobic residues)^{21,36}. If the fluorescence of the tryptophan in the I state (F_I) is similar to its fluorescence in the U state (F_U), we can assume that few or no structural changes take place in the vicinity of the tryptophan location as it goes from the U state to the I state. In contrast, if most of the fluorescence change between F_U and the fluorescence in the N state (F_N) takes place during the U→I transition (and not during the I→N transition), then we can assume that the local structure in the vicinity of the tryptophan in the I state is mostly N-like. On the other hand, if F_I is very distinct from both F_U and F_N , then it is most likely that the structural environment around the tryptophan in the I state is non-native (that is, distinct from the U and N structural environments). Given that the tryptophan side chains in our strategy are highly exposed to solvent in both the U and N states, such non-native fluorescence could suggest that the tryptophan may be buried while in the I state, or may come into close proximity to a charged amino acid that is distant while in either the U or N states. In order to translate fluorescence signal into such structural information, we defined the ratio $\Delta F_{I-U}(i)/\Delta F_{N-U}(i)$, which we call the ω value and which provides a measure of whether a tryptophan probe at position i in the I state has an N- or U-like fluorescence ($\omega = \sim 1$ or $\omega = \sim 0$, respectively) or whether it is distinct from these two states ($\omega \leq -0.25$ or $\omega \geq 1.25$; ω is also comparable to the relative fluorescence introduced in **Fig. 1**). We then mapped the ω values onto the native structure of ubiquitin (**Fig. 4**).

All tryptophans with U-like fluorescence in the I_L ($\omega = \sim 0$; residues 57, 60 and 71) are located at the C-terminal extremity, which contains an $\alpha_{3/10}$ -helix and the central C-terminal β -strand. Five of the six tryptophans that display medium ω values in the I_L ($\omega = \sim 0.5$; residues 59, 62, 64, 66, 68) are also located in the same segment (see **Supplementary Notes**). In contrast, 13 of the 14 tryptophan positions that display N-like fluorescence in the I_L are located in the N-terminal segment (1–54). Also, the four tryptophan probes that revealed non-native environments in the I_L ($\omega \geq 1.25$) either are located in the C-terminal region (residue 73), interact with this region (residues 1 and 42) or are in a strand in the same β -sheet (residue 12). These results suggest that the small, single-domain ubiquitin protein folds through much more energetic frustration than previously reported^{27,29,33}. The C-terminal segment, containing the $\alpha_{3/10}$ -helix (residues 57–60) and the central β -strand (residues 64–71), adopts its native conformation at a rate, k_2 , that is approximately 20-fold slower than k_1 , the rate of formation of the N-terminal two-hairpin-helix motif. These results are also in agreement with an earlier NMR H/D pulse-exchange study of wild-type ubiquitin folding, which suggested that, upon refolding, amide hydrogen bonds of residues 59, 61 and 69 are protected at slower rates than for the rest of the amide protons³⁷.



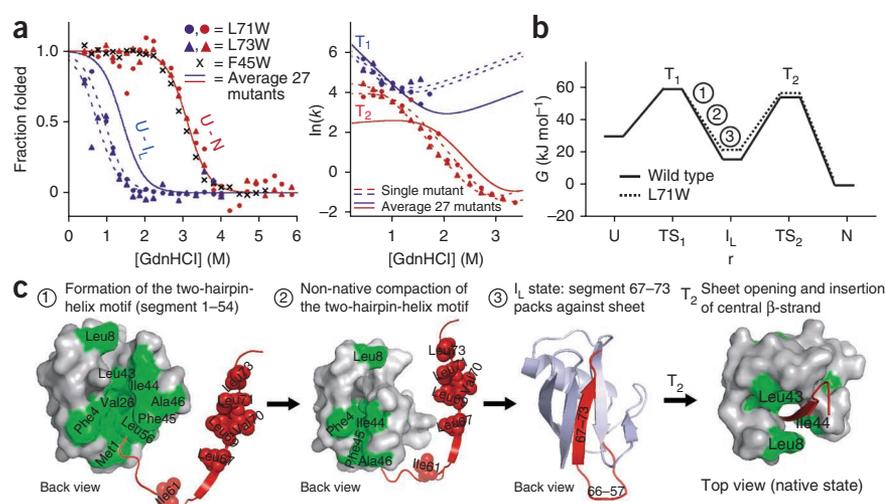
Though the majority of the tryptophan substitutions did not significantly affect the stability of the N state and I state of ubiquitin, seven of them caused sufficient perturbation (>4 kJ mol⁻¹) to allow us to cross-validate the ω values with conventional ϕ -value measurements¹ (that is, the level of energetic involvement of an amino acid in a particular state). We first performed ϕ -value measurements of the apparent two-state transition state observed above 2 M guanidine-HCl, using the seven tryptophan mutations that displayed sufficient structure perturbation of the N state (**Supplementary Table 2**). The ϕ values obtained for these tryptophan mutants are in good agreement with those obtained in previously published work, which used conventional deletion mutations³³. These ϕ values suggest that the N-terminal part of ubiquitin is already formed in the transition state at high guanidine-HCl concentration, but that the C-terminal part remains mostly unstructured (**Supplementary Table 2**)³³. We then explored the extent to which ω values enable us to quantify the degree of nativeness of different regions of ubiquitin in the I_L by comparing the ω values we obtained for the I_L to nine conventional ϕ values ($\Delta\Delta G_{U-I}/\Delta\Delta G_{U-N}$) that we could determine with sufficient precision for this I_L (**Supplementary Fig. 5b** and **Supplementary Table 3**). We found that the ω values for I_L were in excellent agreement with the ϕ values determined for the same tryptophan substitution, suggesting that tryptophan fluorescence may be used as an accurate indicator of the degree of nativeness (or non-nativeness) of a protein's structure (**Supplementary Fig. 5b**).

Evidence for productive non-native interactions

Three tryptophan mutations also generated sufficient destabilization in the I_L ($\Delta\Delta G_{U-I} > 4$ kJ mol⁻¹) to allow characterization of the extent to which these mutated amino acids participate in the folding mechanism¹. For example, G35W destabilized the I_L to about the same extent as for the native state (7.5 versus 8.7 kJ mol⁻¹, respectively), suggesting that the C-terminal helix cap formed by Gly35 is already present in the I_L (**Supplementary Table 1**). In contrast, the two other tryptophan substitutions that most destabilized the I_L ($\Delta\Delta G_{U-I} = 5.6$ and 4.1 kJ mol⁻¹, respectively), L73W and L71W, both located at or proximal to positions displaying a non-native environment in the I_L (see Leu73 in **Fig. 4**),

Figure 5 Ubiquitin folds via a late intermediate (I_L) through the formation of non-native interactions. (a) Stability curves (left) and folding kinetics (right) of tryptophan mutants L71W (circles) and L73W (triangles). A tryptophan was added at location 45 in the mutant L71W in order to detect both T_1 and T_2 (Supplementary Fig. 4).

(b) Folding free-energy profile of mutant L71W compared to wild-type ubiquitin (average of all 27 tryptophan mutants). TS_1 and TS_2 represent the transition states for T_1 and T_2 . (c) Proposed structural model for the formation (1, 2) and rearrangement (T_2) of the late intermediate state (I_L ; 3) detected in the folding mechanism of ubiquitin. Green color represents hydrophobic residues in the 1–56 segment; red represents those in the 57–76 segment.



have no effect on the native-state stability ($\Delta\Delta G_{U-N} = 1$ and 0 kJ mol⁻¹, respectively; Fig. 5a, left). This suggests that Leu73 and Leu71 stabilize the I_L via non-native interactions. Consistent with this observation, disruption of these non-native interactions significantly accelerated the insertion of the C-terminal β -strand into the β -sheet (increased rate for T_2 ; see Fig. 5a, right), suggesting that these non-native interactions create frustration in the refolding pathway of ubiquitin. However, and perhaps more interestingly, disruption of these non-native interactions also increases the folding rate of ubiquitin under conditions where the I_L no longer accumulates (that is, above 2 M guanidine-HCl; see T_2 in Fig. 5a, right, and Fig. 5b). These results therefore reinforce the notion that non-native interactions may accelerate folding by stabilizing productive folding intermediates, but they may also create frustration by trapping the intermediate if these interactions are too strong^{38–41}.

A closer look at the putative structure of a partially formed two-hairpin-helix motif provides a compelling structural representation of the energetic frustration associated with the correct insertion of the central β -strand in the β -sheet (Fig. 5c). First, the initial formation of the two-hairpin-helix motif (residues 1–56)³³ may lead to a compact native-like intermediate state that buries most of the hydrophobic residues (green) of the hydrophobic surfaces of both the hairpin and the α -helix (Fig. 5c, inset 1). The reopening of such a compact structure could therefore occur at the beginning of the slow insertion of the C-terminal segment. Interestingly, another group⁴² reported a structurally similar late folding intermediate in a folding simulation of ubiquitin in which the C-terminal segment of the protein packs against the relatively hydrophobic, partially formed β -sheet before being correctly inserted into the core of the protein in a noncooperative manner. Indeed, the surface of the putative nascent β -sheet formed by the two-hairpin-helix motif reveals a hydrophobic surface on which the hydrophobic-amphipathic tail at residues 67–73 (LHLVLRRL) could rapidly collapse, as suggested by the evidence of non-native interactions found for residue L71 and L73 (Fig. 5c, inset 2). The highly hydrophilic segment at residues 57–66 (SDYNIQKEST) would be likely to remain in a U-like conformation, as suggested by the low ω values found for these residues (Fig. 5c, inset 3). The correct insertion of the central β -strand between the two hairpins would then require an energetically costly reopening of the ubiquitin core (see T_2 , Fig. 5c). Consistent with this hypothesis, the guanidine-HCl dependence of the rate constant for the rearrangement of the I_L suggests that the I_L must re-expose ~10% of its buried side chains to the solvent in order to reach the native state²⁰. The main finding of this study, however, remains that the proposed binding of residues Leu71 and Leu73

to the nascent β -sheet resembles the ‘flycatcher’ mechanism of concerted binding and folding observed for unstructured proteins^{43,44}. In this mechanism, non-native interactions first ‘catch’ an unstructured peptide and then accelerate its structural transition into a native-like complex by bringing the two interacting elements closer to each other. These results thus suggest a close relationship between intramolecular folding and intermolecular binding-folding mechanisms.

DISCUSSION

The role of non-native interactions and transient intermediate states in protein folding has been the subject of much debate because few examples have been detected or characterized during the folding of small proteins^{1,13–16,35,40,45,46}. In the present study, however, we demonstrated that a tryptophan-scanning strategy allowed us to detect hidden non-native interactions and to characterize the structure of a folding intermediate in the folding pathway of ubiquitin, a well-studied protein that has long been considered to fold via an apparent two-state mechanism^{27,29,33}. Our results also highlight the dual role of non-native interactions: creating frustration in the folding pathway of proteins but also accelerating protein folding by stabilizing productive intermediates without overstabilizing or trapping them^{13,38}. Interestingly, a previous study found that more than 50% of proteins reported to display an apparent two-state folding mechanism also display deviations in folding or unfolding rate constants¹², suggesting the presence of a late intermediate state, similar to the one observed in ubiquitin and other proteins^{14,47,48}. It will be interesting to see whether our tryptophan-scanning strategy will also enable the direct detection of non-native interactions and late transient intermediates for other proteins exhibiting two-state folding. In such cases, late frustration and the ‘flycatcher’ mechanism of concerted binding and folding might well represent a common mechanism of protein folding^{43,44}.

In a broader context, the tryptophan-scanning strategy presented here is an ideal complement to classic perturbation-based strategies to study transition states in protein folding, such as ϕ -value¹ or ψ -value analysis^{49,50}. For example, the tryptophan-scanning strategy allows generation of fluorescent probes that enable efficient detection of all structural transitions²⁰. In contrast to ϕ -value analysis, which necessitates mutations that destabilize the N state, the tryptophan-scanning strategy also enables assessment of the role of surface-exposed residues, which often don’t contribute much to native-state stability. We envisage adaptation of a general fluorophore-scanning approach (using, for example, structure sensitive fluorescent nucleotides) to study the transient states in the folding of complex DNA or RNA molecules or even

to characterize any structural transitions of macromolecules formed as a result of chemical or physical perturbation. Finally, as our understanding of the structural determinants of fluorescence of tryptophan (or other fluorophores)^{21,36} progresses, the tryptophan-scanning strategy will generate more specific structural constraints that, once integrated with other experimental data and computer simulations, will allow us to test models of transient states with unprecedented detail⁸.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

The authors acknowledge A. Bonham for help with Mathematica; T. Sosnick, H. Roder, K. Plaxco, F.-X. Campbell-Valois, S. Chteinberg, J.W. Keillor, H. Bhaskarah, C. Lawrence and H. Watkins for helpful discussions; M. Fyfe for sequencing; and J.W. Keillor for providing access to the stopped-flow apparatus. This work was supported by the National Science and Engineering Research Council of Canada (Grant 194582-SWM). A.V.-B. acknowledges the financial support of the Fonds Québécois de Recherche Nature et Technologies.

AUTHOR CONTRIBUTIONS

A.V.-B. performed experiments and mathematical modeling. A.V.-B. and S.W.M. designed experiments, analyzed results, and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/nsmb.2322>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Fersht, A.R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W.H. Freeman, New York, 1999).
2. Vendruscolo, M. & Dobson, C.M. Towards complete descriptions of the free-energy landscapes of proteins. *Philos. Transact. A Math. Phys. Eng. Sci.* **363**, 433–452 (2005).
3. Karplus, M., Gao, Y.Q., Ma, J., van der Vaart, A. & Yang, W. Protein structural transitions and their functional role. *Philos. Transact. A Math. Phys. Eng. Sci.* **363**, 331–356 (2005).
4. Jahn, T.R. & Radford, S.E. Folding versus aggregation: polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **469**, 100–117 (2008).
5. Schaeffer, R.D., Fersht, A. & Daggett, V. Combining experiment and simulation in protein folding: closing the gap for small model systems. *Curr. Opin. Struct. Biol.* **18**, 4–9 (2008).
6. Bowman, G.R., Voelz, V.A. & Pande, V.S. Taming the complexity of protein folding. *Curr. Opin. Struct. Biol.* **21**, 4–11 (2011).
7. Fleishman, S.J. & Baker, D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **149**, 262–273 (2012).
8. Korzhnev, D.M. *et al.* Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586–590 (2004).
9. Schotte, F. *et al.* Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science* **300**, 1944–1947 (2003).
10. Brockwell, D.J. & Radford, S.E. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* **17**, 30–37 (2007).
11. Plaxco, K.W., Simons, K.T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994 (1998).
12. Sanchez, I.E. & Kiefhaber, T. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376 (2003).
13. Zarrine-Afsar, A. *et al.* Theoretical and experimental demonstration of the importance of specific non-native interactions in protein folding. *Proc. Natl. Acad. Sci. USA* **105**, 9999–10004 (2008).
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. Krishna, M.M. & Englander, S.W. A unified mechanism for protein folding: predetermined pathways with optional errors. *Protein Sci.* **16**, 449–464 (2007).
16. 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).

17. Weber, G. Fluorescence-polarization spectrum and electronic-energy transfer in proteins. *Biochem. J.* **75**, 345–352 (1960).
18. Royer, C.A. Probing protein folding and conformational transitions with fluorescence. *Chem. Rev.* **106**, 1769–1784 (2006).
19. Smith, C.J. *et al.* Detection and characterization of intermediates in the folding of large proteins by the use of genetically inserted tryptophan probes. *Biochemistry* **30**, 1028–1036 (1991).
20. Vallée-Bélisle, A. & Michnick, S.W. Multiple tryptophan probes reveal that ubiquitin folds via a late misfolded intermediate. *J. Mol. Biol.* **374**, 791–805 (2007).
21. Beechem, J.M. Picosecond fluorescence decay curves collected on millisecond time scale: direct measurement of hydrodynamic radii, local/global mobility, and intramolecular distances during protein-folding reactions. *Methods Enzymol.* **278**, 24–49 (1997).
22. Gu, H. *et al.* Robustness of protein folding kinetics to surface hydrophobic substitutions. *Protein Sci.* **8**, 2734–2741 (1999).
23. Jackson, S.E. Ubiquitin: a small protein folding paradigm. *Org. Biomol. Chem.* **4**, 1845–1853 (2006).
24. Hicke, L., Schubert, H.L. & Hill, C.P. Ubiquitin-binding domains. *Nat. Rev. Mol. Cell Biol.* **6**, 610–621 (2005).
25. Went, H.M., Benitez-Cardoza, C.G. & Jackson, S.E. Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Lett.* **567**, 333–338 (2004).
26. Khorasanizadeh, S., Peters, I.D. & Roder, H. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat. Struct. Biol.* **3**, 193–205 (1996).
27. Krantz, B.A. & Sosnick, T.R. Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* **39**, 11696–11701 (2000).
28. Roder, H., Maki, K., Cheng, H. & Shastry, M.C. Rapid mixing methods for exploring the kinetics of protein folding. *Methods* **34**, 15–27 (2004).
29. Sosnick, T.R., Dohager, R.S. & Krantz, B.A. Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *Proc. Natl. Acad. Sci. USA* **101**, 17377–17382 (2004).
30. Vallée-Bélisle, A., Turcotte, J.F. & Michnick, S.W. raf RBD and ubiquitin proteins share similar folds, folding rates and mechanisms despite having unrelated amino acid sequences. *Biochemistry* **43**, 8447–8458 (2004).
31. Rea, A.M., Simpson, E.R., Crespo, M.D. & Searle, M.S. Helix mutations stabilize a late productive intermediate on the folding pathway of ubiquitin. *Biochemistry* **47**, 8225–8236 (2008).
32. Khorasanizadeh, S., Peters, I.D., Butt, T.R. & Roder, H. Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **32**, 7054–7063 (1993).
33. Went, H.M. & Jackson, S.E. Ubiquitin folds through a highly polarized transition state. *Protein Eng. Des. Sel.* **18**, 229–237 (2005).
34. Semisotnov, G.V. *et al.* Study of the “molten globule” intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119–128 (1991).
35. Jennings, P.A. Speeding along the protein folding highway, are we reading the signs correctly? *Nat. Struct. Biol.* **5**, 846–848 (1998).
36. Callis, P.R. & Liu, T. Quantitative prediction of fluorescence quantum yields for tryptophan in proteins. *J. Phys. Chem. B* **108**, 4248–4259 (2004).
37. Briggs, M.S. & Roder, H. Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc. Natl. Acad. Sci. USA* **89**, 2017–2021 (1992).
38. Clementi, C. & Plotkin, S.S. The effects of non-native interactions on protein folding rates: theory and simulation. *Protein Sci.* **13**, 1750–1766 (2004).
39. Wolynes, P.G., Onuchic, J.N. & Thirumalai, D. Navigating the folding routes. *Science* **267**, 1619–1620 (1995).
40. Onuchic, J.N. & Wolynes, P.G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).
41. Oliveberg, M. & Wolynes, P.G. The experimental survey of protein-folding energy landscapes. *Q. Rev. Biophys.* **38**, 245–288 (2005).
42. Sorenson, J.M. & Head-Gordon, T. Toward minimalist models of larger proteins: a ubiquitin-like protein. *Proteins* **46**, 368–379 (2002).
43. Sugase, K., Dyson, H.J. & Wright, P.E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025 (2007).
44. Shoemaker, B.A., Portman, J.J. & Wolynes, P.G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA* **97**, 8868–8873 (2000).
45. Baker, D. A surprising simplicity to protein folding. *Nature* **405**, 39–42 (2000).
46. Watters, A.L. *et al.* The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* **128**, 613–624 (2007).
47. 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).
48. Wensley, B.G. *et al.* Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **463**, 685–688 (2010).
49. Krantz, B.A., Dohager, R.S. & Sosnick, T.R. Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J. Mol. Biol.* **337**, 463–475 (2004).
50. Sosnick, T.R., Krantz, B.A., Dohager, R.S. & Baxa, M. Characterizing the protein folding transition state using psi analysis. *Chem. Rev.* **106**, 1862–1876 (2006).
51. Vijay-Kumar, S., Bugg, C.E. & Cook, W.J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531–544 (1987).

© 2012 Nature America, Inc. All rights reserved.



ONLINE METHODS

Protein constructs and expression. Mammalian ubiquitin was fused to a His₆ tag at its N terminus, (MHHHHHHG). Tryptophan mutations and expression of the different ubiquitin mutants were carried out as described²⁰.

Data collection. All experiments were performed in 50 mM sodium phosphate buffer (pH 7.0) at 30 ± 0.1 °C to allow for proper kinetic resolution between T₁ and T₂. All folding data are the average of three guanidine-HCl-jump experiments acquired using an Applied Photophysics SX18.MV stopped-flow fluorimeter (tryptophan fluorescence: $\lambda_{\text{ex}} = 281 \pm 2.5$ nm with a 320-nm-cutoff filter for λ_{em} ; ANS fluorescence (200 μM): $\lambda_{\text{ex}} = 350 \pm 2.5$ nm with a 395-nm cutoff for λ_{em}). Denatured protein (~150 μM) in 5.0 M guanidine-HCl (ultrapure grade) was mixed 1:10 with different concentrations of guanidine-HCl. For the ANS-monitored refolding experiment, the unfolded protein was mixed 1:10 into a buffered solution of 220 μM ANS. Dead-time calibration of the stopped-flow instrument was performed as described²⁰. Data analysis was performed using the nonlinear regression analysis program in Kaleidagraph (version 3.6 Synergy Software, pCS Inc.). Refolding kinetic traces at low concentrations of denaturant were fit with two exponential terms between 3.2 ms and 1,000 ms. The refolding traces obtained for each tryptophan mutant were set relative to the fluorescence of the unfolded state obtained between 2 M and 6 M guanidine-HCl, which was

set to zero in the absence of guanidine-HCl. The fluorescence levels F_1 and F_2 , extrapolated from T₁ and T₂, respectively, were fit to a standard two-state equilibrium curve to obtain an estimate of m_I and ΔG_{UI} and m_N and ΔG_{UN} , respectively (see **Supplementary Fig. 4**). Global fitting analysis of equilibrium and kinetics data was also performed for all ubiquitin variants (see, for example, **Fig. 5a**). In brief, the observed rate constant k_1 , for formation of I_L (fitted to a chevron curve⁵²), the observed rate constant k_2 , for formation of the N state (fitted to a three-state, on-pathway mechanism⁵³), and both the equilibrium curves for I_L and N states (both fitted to a two-state transition) were simultaneously fitted, thus providing estimates for m_I and ΔG_{UI} and m_N and ΔG_{UN} . See **Supplementary Figure 4** and previously described methods²⁰ for more details on the global fitting procedure. *N*-Acetyltryptophanamide (Sigma-Aldrich) was used as a model for free tryptophan to study its fluorescence dependence on the concentration of guanidine-HCl (**Supplementary Fig. 3c**).

52. Maxwell, K.L. *et al.* Protein folding: defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **14**, 602–616 (2005).

53. Bofill, R., Simpson, E.R., Platt, G.W., Crespo, M.D. & Searle, M.S. Extending the folding nucleus of ubiquitin with an independently folding beta-hairpin finger: hurdles to rapid folding arising from the stabilisation of local interactions. *J. Mol. Biol.* **349**, 205–221 (2005).