Knotting and Unknotting of a Protein in Single Molecule Experiments

Fabian Ziegler[a,1], Nicole C H Lim[b,c,1], Soumit S Mandal[a,1], Benjamin Pelz[a], Wei-Ping Ng[b], Michael Schlierf[e], Sophie E Jackson[b,2] and Matthias Rief[a,d,2]

[a]Physik Department E22, Technische Universität München, 85748 Garching, Germany [b]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom [c]Faculty of Sciences, Universiti Brunei Darussalam, Gadong, BE 1410, Brunei Darussalam [d]Munich Center for Integrated Protein Science, 81377 München, Germany [e]B Cube – Center for Molecular Bioengineering, 01307 Dresden, Germany

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Spontaneous folding of a polypeptide chain into a knotted structure remains one of the most puzzling and fascinating features of protein folding. The folding of knotted proteins is on the timescale of minutes and thus hard to reproduce with atomistic simulations that have been able to reproduce features of ultrafast folding in great detail. Furthermore, it is generally not possible to control the topology of the unfolded state. Single-molecule force spectroscopy is an ideal tool for overcoming this problem: By variation of pulling directions, we have controlled the knotting topology of the unfolded state of the 5_2-knotted protein UCH-L1 and have therefore been able to quantify the influence of knotting on its folding rate. Here, we provide direct evidence that a threading event associated with formation of either a 3_1 or 5_2 knot, or a step closely associated with it, significantly slows down the folding of UCH-L1. The results of the optical tweezers experiments highlight the complex nature of the folding pathway, many additional intermediate structures being detected that cannot be resolved by intrinsic fluorescence. Mechanical stretching of knotted proteins is also of importance for understanding the possible implications of knots in proteins for cellular degradation. Compared to a simple 3_1 knot, we measure a significantly larger size for the 5_2 knot in the unfolded state that can be further tightened with higher forces. Our results highlight the potential difficulties in degrading a 5_2 knot compared with a 3_1 knot.

Knotted proteins | Protein folding | Single molecule | Optical tweezers

Introduction

Over the past few decades, protein folding studies have focused largely on relatively small, monomeric model systems that possess simple topologies. Extensive experimental and computational studies have provided significant mechanistic insight into folding pathways (1-5). These proteins, which fold rapidly, have been shown to possess relatively smooth energy landscapes (6, 7). However, the emergence of classes of topologically complex knotted protein structures in recent years (8-11) challenges some long-standing views in the field as such proteins have to avoid kinetic traps and also overcome significant topological barriers during folding. To date, over 750 knotted proteins have been discovered within the PDB, each containing either a trefoil (3_1), figure-of-eight (4_1), Gordian (5_2) or stevedore (6_1) knot in their structures (10-12). The conservation of these knotted topologies across different families has suggested that the knot may, in some way, be advantageous and important to the stability and/or function of the protein (12). As yet, not much is known about the advantages, if any, of knotted structures over their unknotted counterparts. It has been proposed that the knot may play a part in enhancing a protein's thermodynamic, kinetic, mechanical or cellular stability (13-15).

To date, relatively few experimental studies on the knotting and folding of topologically knotted proteins have been published (14). Most of these investigations have been focused on the trefoil-knotted bacterial methyltransferases YibK and YbeA (16-24). Extensive *in silico* studies have provided insights into how these 3_1-knotted conformations may be formed (13, 14). Some simulations report formation of intermediate 'slipknot' configurations, partial unfolding (backtracking) events or specific, non-native interactions which promote threading, during the folding of 3_1-knotted proteins (25, 26). Other experimental studies have also examined the folding of designed and naturally occurring trefoil-knotted proteins (27-29). In contrast with the trefoil-knotted proteins, much less is known about the folding of proteins with more complex knots.

UCH-L1 is a monomeric cysteine protease that belongs to the ubiquitin C-terminal hydrolase (UCH) family, a subgroup of deubiquitinating enzymes. The overall structure of UCH-L1 consists of six β-strands forming a central β-sheet that is flanked on either side by seven α-helices. The polypeptide chain threads through itself to create a 5_2 knot in the native state (Fig. 1). UCH-L1 is highly expressed in neurons and accounts for 1-2% of the brain proteome (30), and has been linked to neurodegenerative diseases such as Parkinson's Disease (PD) and Alzheimer's Disease (AD) (31, 32).

Previously, we have probed the folding mechanism of UCH-L1 and shown that it reversibly unfolds *in vitro* upon addition of chemical denaturant to a state with no detectable secondary or tertiary structure (33). The same study established that chemical denaturant-induced unfolding is three-state and an intermediate populated under equilibrium conditions (33). Using NMR hydrogen-deuterium exchange (HDX) experiments, the intermediate state was characterised indirectly and it was found that the central β-sheet core of the protein remains structured whilst

Significance

Knots and entanglements are common in biology but little is known about their formation or function. Here, we use state-of-the-art single-molecule force spectroscopy to mechanically unfold a protein, UCH-L1, which has a 5_2 knot in its structure. The exquisite control inherent in this technique enables us to specifically unfold the protein to different knotted and unknotted denatured states, from which we can then watch the protein refold. These experiments establish the effect of knots on folding pathways. Many intermediate structures are detected suggesting both on- and off-pathway intermediates are populated. Our results also highlight the potential difficulties in degrading a 5_2-knotted protein and therefore may have implications for some diseases.

Reserved for Publication Footnotes

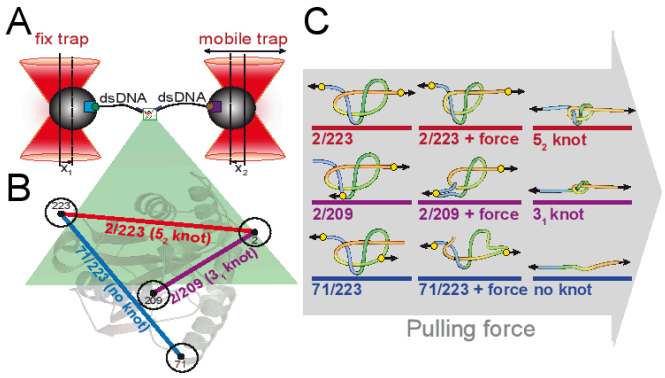

Fig. 1. Design of optical tweezers assay and knotted structure of the variants of UCH-L1 used. (*A*) Schematic experimental setup of the dual-beam optical tweezers assay. Forces are measured through the deflections of the beads (x_1, x_2) out of the trap centers. (*B*) Structure of UCH-L1 with the different attachment points: Pulling at positions 2 and 223 (N and C-termini) leads to an unfolded structure with a 5_2-knotted topology (red), pulling at positions 2 and 209 leads to an unfolded structure with a 3_1-knotted topology (violet) and pulling at positions 71 and 223 leads to an unfolded structure without a knot (blue). (*C*) Schematic representation of the tightening of the mutant protein structures on application of force, colored as in (*B*).

many of the surrounding α-helices have unfolded (33). Recently, a detailed kinetic analysis undertaken by Luo et al. has also established that UCH-L1 exhibits multiphasic kinetics indicative of parallel pathways and the population of at least two, metastable intermediate states (34).

Single-molecule force spectroscopy offers unique control over the conformation of proteins (35-39). In this study, we use high-precision optical tweezers to unfold and refold UCH-L1 in three distinct pulling geometries (Fig. 1 *A* and *B*). By varying the pulling geometry using cysteine engineering (40), we are able to specifically, mechanically unfold the protein chain to a state containing either a 5_2 or 3_1 knot, or completely untie the chain such that the denatured state is the unknot (0). These experiments allow us to i) characterise not only a 5_2 knot in an unfolded polypeptide chain, but also a 3_1 knot created in the same chain, ii) quantify the mechanical stability of 5_2-knotted proteins, iii) directly detect unfolding and refolding intermediates, and most importantly, iv) directly measure the influence of knot formation on the rate of folding of a knotted protein.

Results

Design and ensemble characterization of the UCH-L1 variants

For the optical tweezers experiments, cysteines for oligo-DNA handle attachment were introduced at exposed positions in the native structure. As indicated in Fig. 1*C*, the locations of these residues were chosen in such a way that the conformation of the knotted chain could be manipulated (i.e. converted into a tightened 5_2 or 3_1 knot or untied) when pulling forces are applied. In this study, three double-cysteine variants of UCH-L1 were chosen: Q2C-A223C (2/223), Q2C-Q209C (2/209) and K71C-A223C (71/223) (Fig. 1*C*). For the 2/223 construct, the protein is attached to the DNA handles and pulled from its N- and C-termini, which should lead to the formation of a tightened 5_2 knot after mechanical unfolding. In the 2/209 construct, pulling should lead to an unfolded state with a 3_1 knot, while pulling from the cysteine residues at positions 71 and 223 should result in an unknotted linear polypeptide chain.

Prior to performing the optical tweezers experiments, the double-cysteine variants were characterised using ensemble techniques to assess the impact of the amino-acid substitutions on the structure, stability and unfolding and refolding kinetics of the protein. The secondary structure of the variants was probed by far-UV circular dichroism (*SI Appendix*, Fig. S1), the effect of the

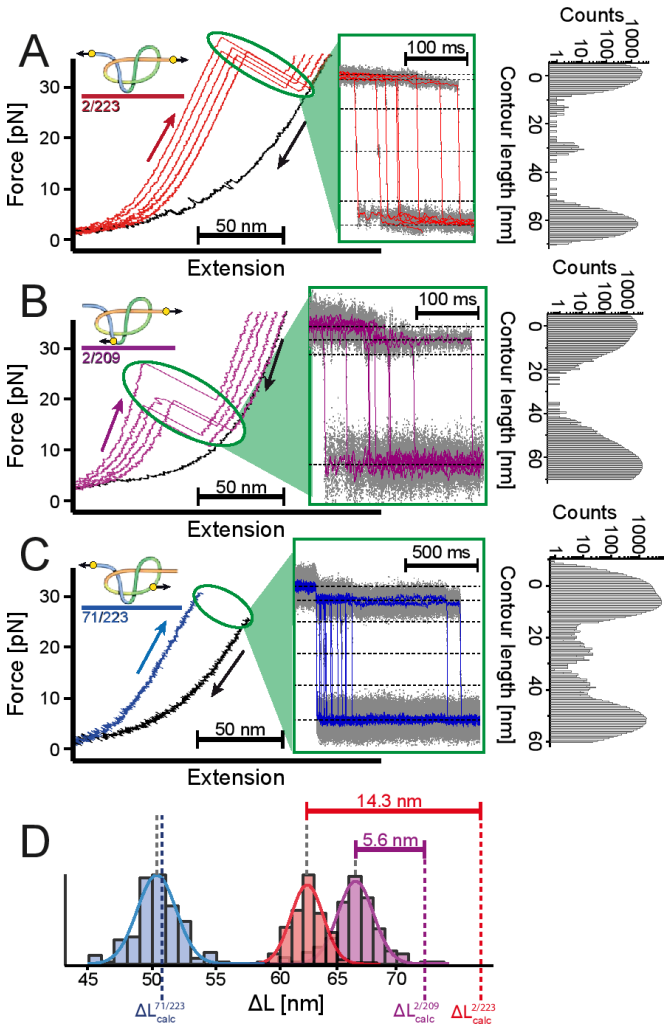

Fig. 2. Sample force-extension traces of experiments with the three UCH-L1 variants. UCH-L1 2/223 (A), UCH-L1 2/209 (B) and UCH-L1 71/223 (C). Force-extension curves are shown on the left. The middle panel shows a zoom into the unfolding phase of force extension curves transformed into a contour length vs. time representation (see §2.7 of the *SI Appendix*). Unfolding intermediates are marked by dotted lines. Right panels show position histogram of the unfolding phase of those traces. Note that that for UCH-L1 71/223 (C), waiting times of seconds to minutes at forces of ~35 pN were required. Therefore, the inset shows an overlay of unfolding events in constant distance mode, exhibiting the characteristic unfolding pattern (see also *SI Appendix*, Fig. S6). (D) Contour length increases associated with the unfolding of all three UCH-L1 variants: 2/223 (red), 2/209 (violet) and 71/223 (blue). The calculated values (see text) are indicated as colored dotted lines. Difference between calculated and observed distances leads to an estimation for the size of the tightened 5_2 (red) and 3_1 (violet) knot.

amino-acid substitutions on the thermodynamic stability of the native and intermediate states was assessed using urea-induced unfolding under equilibrium conditions (*SI Appendix*, Fig. S2), and the unfolding and refolding kinetics were investigated using single [denaturant]-jump stopped flow experiments (*SI Appendix*, Fig. S3). In the latter two cases, intrinsic fluorescence was used to probe the unfolding/folding of the protein. Collectively, these experiments established that the amino-acid substitutions do not significantly affect the structure of the native state, the stability of the intermediate state, nor the folding pathway. In all cases, the amino-acid substitutions destabilised the native state of UCH-L1 to a similar degree, but the protein still formed a highly stable native state relative to the intermediate state. More importantly, the folding rate constants in water calculated for the

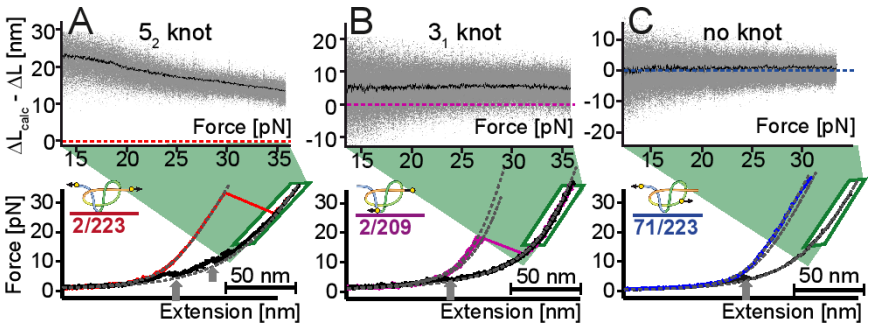

Fig. 3. The effect of force on the size of the 5_2 and 3_1 knots in the unfolded state of UCH-L1. Plotting the difference between calculated and measured contour length (ΔL_{calc}- ΔL) against force shows a significantly different behaviour of (*A*) the construct with a 5_2 knot in the unfolded state compared to (*B*) the construct with the smaller and more compact 3_1 knot and (*C*) the construct without a knot in the unfolded state. Top panel: Black curves are an average of 15 to 20 different measurements in the same experiment and gray dots are overlay of all the original traces. Lower panel: Sample unfolding (colored) and relaxation curves (black) for all three constructs. The gray arrows indicate the region where fast equilibrium transitions between pre-formed structures occur. These show different behaviour than the knot tightening at higher forces.

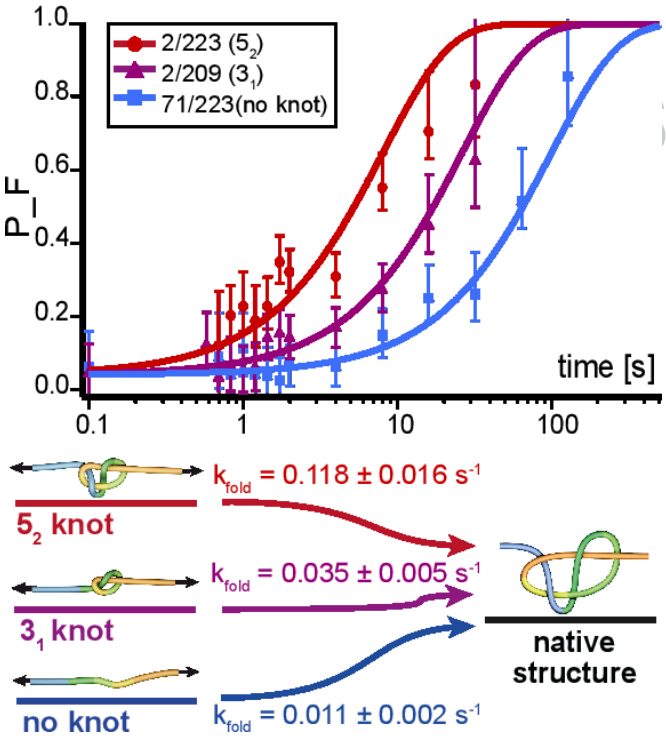

Fig. 4. Probability P_F folding to the native state plotted against waiting time at zero force for all three constructs (colored as in Fig. 2). Refolding rate constants of each construct were obtained from a single-exponential fit of the optical tweezers data.

three variants used in the optical tweezers experiments do not differ significantly; see *SI Appendix*, §3.3 and Table S2 for further details.

Mechanical unfolding of UCH-L1 with different pulling directions shows different unfolding patterns

Fig. 2*A* shows a force-extension curve for unfolding of UCH-L1 in the N- and C-terminal direction (2/223). We observe unfolding forces of about 37 pN at a pulling velocity of 200 nm s^{-1} and four very short-lived unfolding intermediates. The first two can be shown to be due to rupturing of the N-terminal α−helix, followed by the N-terminal β-strand (see Fig. 2A and *SI Appendix* S9A). The folded distance between residues 2 and 223 calculated from the crystal structure is 3.7 nm (PDB code 2ETL) (41). Together with the length of 0.365 nm for a single peptide bond in an unfolded polypeptide chain (42), this allows us to estimate the contour length gain one would expect for full unfolding of UCH-L1 from residues 2 to 223:

$$\Delta L_p = (223 - 2) \cdot 0.365\ nm - 3.7\ nm = 77.0\ nm$$

Experimentally, a contour length gain of 62.4 nm is observed. The "missing contour length", i.e. the difference between calculated and measured values of 14.6 nm corresponds to a chain approximately 40 residues long. This can be explained by either i) 40 residues forming a tightened knotted structure after unfolding, or ii) the unfolding of a 40-residue segment of protein structure at forces below the detection threshold of our setup (< 0.5 pN). The second possibility can be excluded as such a flexible segment is not observed for either of the two other constructs 71/223 and 2/209. Since 71/223 contains the full C-terminal region and 2/209 the full N-terminal region, if present in the structure, a flexible segment would be detectable in at least one of these other two constructs.

Fig. 2*B* shows the unfolding of native UCH-L1 in the 2/209 direction. The major unfolding force peak is in the range of 18 pN at a pulling velocity of 200 nm s^{-1}. Unfolding also proceeds through several intermediates (see Fig. 2B and *SI Appendix* S9C). In contrast to the results for the 2/223 construct, the major unfolding peak is preceded by a transition where the protein rapidly samples both native and intermediate states (see zoom in *SI Appendix* Fig. S9C). This transition shows a contour length increase of 6.3 nm that is likely caused by unfolding of the first N-terminal α-helix (see *SI Appendix*, Fig. S9D and §3.5). Note that the direction of force application to this α-helix is different than in 2/223 explaining the different forces and kinetics for α-helix detachment in the two variants (*SI Appendix*, Fig. S9). The total contour-length increase for full unfolding is 66.7 nm, shorter by a missing contour length of 5.7 nm than the contour length gain for an extended, unknotted polypeptide chain as calculated using

$$\Delta L_p = (209 - 2) \cdot 0.365\ nm - 3.2\ nm = 72.4\ nm$$

This result establishes that the 3_1 knot, comprising of approximately 16 residues, remains in the mechanically unfolded state of 2/209 as predicted.

The 71/223 construct, designed for complete unknotting of UCH-L1 on unfolding, requires a significantly higher unfolding force. At the pulling velocities used (200 nm s^{-1}) unfolding of the native structure in most cases did not occur at forces below 35 pN (Fig. 2*C*). The structure then could only be unfolded by waiting for seconds to minutes at forces between 35 and 40 pN. Since

our optical tweezers set-up does not allow application of higher forces, it was not possible to obtain a distribution of unfolding forces as for the other two variants. However, the rare unfolding events that occur while pulling the 71/223 construct with constant velocity (*SI Appendix* Fig. S5C, right panel) show exactly the same fingerprint as additional unfolding experiments performed at a constant distance: all unfolding events exhibit a characteristic first unfolding step to an intermediate associated with a contour length increase of 6.2 nm (middle and right panel of Fig. 2*C*), sometimes followed by up to three, very short-lived, additional unfolding intermediates consistent with a kinetic partitioning mechanism for unfolding (43, 44).

A contour length increase of 50.4 nm was obtained for the 71/223 construct, consistent with the calculated contour length gain for an extended, unknotted polypeptide chain:

$$\Delta L_p = (223 - 71) \cdot 0.365\ nm - 4.1\ nm = 51.4\ nm.$$

Distinct from the results obtained for the 2/223 and 2/209 constructs, in 71/223 we find no evidence for a remaining knotted structure in the unfolded chain.

Tightening of the 5_2 knot

As shown above, the presence of a knot in the taut unfolded chains of the 2/223 and 2/209 constructs manifests itself indirectly in the difference between contour length observed and that calculated for an unknotted chain of the same length. Since a 5_2 knot contains five crossings of the chain one might expect to observe more contributions from steric clashes of side chains during the compaction of the knot upon tightening it at high loads compared to a 3_1 knot. If tightening of the knot proceeds through discrete transitions, we expect to measure a growing contour length of the unfolded chain owing to the stepwise compaction of the knot towards high forces. In other words, the missing contour length should shrink with force. Indeed, plotting the missing contour length of unfolded polypeptide chain vs the applied force (Fig. 3, top panel) reveals a clear difference between the 2/223 construct and both the other constructs: The measured contour lengths of 71/223 and 2/209 do not depend on force and can thus be well described by the typical elasticity of an entropic polymer chain (see *SI Appendix*, §2.5 for a detailed description). Note that the occurrence of the 3_1 knot in 2/209 leads to a missing contour length of 5.7 nm (Fig. 3*B*, top panel), however, this value does not change with force. In contrast, the missing contour length decreases with force in the unfolded 2/223 chain indicating further tightening of the 5_2 knot at high forces. Between 20 and 36 pN, the missing contour length reduces by $\sim$ 6 nm (Fig. 3*A*, top panel). A close inspection of the data (*SI Appendix*, Fig. S8B) reveals a series of discrete transitions, likely intermediate states with different degrees of compactness populated during the tightening of the 5_2 knot. Since the tightening transitions occur close to equilibrium, integrating the force vs extension traces in this transition regime allows estimation of the free energy required for compaction of the knot between 20 and 36 pN to approximately 23.1 kT (see *SI Appendix* § 2.8).

Folding Intermediates

Relaxation traces in force spectroscopy experiments allow observation of rapidly forming intermediate states. Even though overall folding to the native state was slow (see next paragraph), all three constructs exhibit rapid fluctuations between extended and collapsed structures when relaxing the force after unfolding (black curves in Fig. 3). These close-to-equilibrium transitions (see arrows in Fig. 3) appear in a force range between 3 and 12 pN indicating the rapid and early population of intermediate states. 2/223 shows two regions at 12 and 6 pN respectively, while 2/209 shows only one such region at around 4 pN, and 71/223 one region at 5 pN (see *SI Appendix*, Fig. S8 for more details). Those rapidly forming intermediates involve large regions of the protein which, judged by the large change in contour lengths, are of almost the same compactness as the native state.

We can estimate the $^{\triangle}$G of the rapidly forming equilibrium structures by integrating the area enclosed by the relaxation force curve and the WLC curve of the fully unfolded state. We find a $^{\triangle}$G of 8 kT for the low-force intermediates and 11 kT for the high-force intermediates in the 2/223 construct and a $^{\triangle}$G of 6-7 kT and 8 kT for the (2/209) and (71/223) constructs, respectively (see *SI Appendix*, §2.8). It is important to note that those intermediates are heterogeneous in both contour length as well as kinetics indicating that there is an ensemble of intermediate states: analysis of a sample trace for the low-force region of the force-relaxation curves of the 2/223 construct (*SI Appendix*, Fig. S10) using a Hidden-Markov-Model (45) shows the heterogeneity of this ensemble. We find that even direct transitions between the intermediate states are possible without visiting the unfolded state reminiscent of the folding network model proposed in (46). Apparently, this ensemble also contains longer-lived intermediate states that are stable for many seconds even under load (see long dwell time in the trace of *SI Appendix*, Fig. S8D). A similar heterogeneous population can be observed in traces where we relax the chain to zero load and, after a certain waiting time, rapidly pull and unfold the chain (sample traces in *SI Appendix*, Fig. S7). Longer-lived intermediate states then lead to pronounced force peaks at non-native contour lengths while more dynamic intermediates lead to low force peaks (see *SI Appendix*, Fig. S7D).

Refolding kinetics

In the next set of experiments, the effect of the three different knotted unfolded states on the refolding kinetics of the protein was investigated. To this end, we first unfolded the various constructs and subsequently relaxed the tension to zero force and allowed the protein a certain time to refold. The folding state of the molecule was then probed in another force-ramp experiment (for details of the experimental protocol see SI Methods). A summary of the time-dependent refolding probability for all three constructs is shown in Fig. 4. The refolding kinetics strongly differ with pulling direction. Using a simplified two-state model of folding, the following 'global' folding rate constants were calculated from the data: $k_{2/223} = 0.118 \pm 0.016\ s^{-1}$, $k_{2/209} = 0.035 \pm 0.005\ s^{-1}$, and $k_{71/223} = 0.011 \pm 0.002\ s^{-1}$. Thus, a pre-formed 5_2 knot accelerates folding to the native state by one order of magnitude compared to refolding from a fully unfolded and unknotted chain. A pre-formed trefoil knot also leads to an increase in folding rate compared to the unknotted denatured state, however, the effect is smaller than for the 5_2 knot. It is important to note that, in all constructs, a broad range of different meta-stable folding intermediates is observed. These intermediates account for the deviations of the data in Fig. 4 from a single-exponential fit. Typical sequences of unfolding/refolding cycles with varying waiting times are shown in *SI Appendix*, Fig. S4.

Discussion

Size of knots in the stretched polypeptide chain

While the contour length increase upon unfolding the 71/223 construct is in good agreement with the calculated length of the fully unfolded polypeptide chain, the contour length increase for the 2/223 and 2/209 constructs are shorter than expected for unknotted, unfolded chains containing the same number of residues. Pulling in the 2/209 direction provides a size estimate for the 3_1 knot of 5.7 nm, which corresponds to approximately 16 amino-acid residues. This is in good agreement with AFM-measurements on AFV3-109 (47) and phytochrome C (48) as well as values from simulations of tight knots in polypeptide chains under force (49).

In contrast to the results for the 2/209 and 71/223 constructs, the missing length measured for the residual 5_2 knot in the 2/223

construct has a surprisingly large size of 14.6 nm which corresponds to roughly 40 residues, significantly larger than simple estimates on knotted ropes predict (which for a 5_2 knot vary from 6.4 to 9.2 nm) (50). This suggests that the dense network of crossing strands in the 5_2 knot makes it more difficult to tighten the knot leaving its size larger than expected under certain loads.

Tightening the 5_2 knot

Consistent with the putative "bulkiness" of the 5_2 knot compared to the simpler 3_1 knot, additional transitions are observed that relate to compaction of this knot at high forces (Fig. 3). It is important to note that knot compaction occurs at much higher forces (>20 pN) than those at which the early refolding intermediates fold/unfold (<12 pN, see arrows in Fig. 3), enabling us to clearly distinguish between the processes. Despite the small contour length change of roughly 6 nm upon compaction of the 5_2 knot between 20 and 36 pN pulling force, the associated equilibrium free energy changes are large (approximately 23.1 kT or 13.7 kcal mol^{-1}), a value far too high to be explained by folding/unfolding of a small element of protein structure such as an α-helix or a small β-sheet (51). The increasing contour length for the 2/223 construct towards high forces shows the knot is not yet compact and offers an explanation for the apparently overly large size of the 5_2 knot at low forces. This compaction is only observed for the 5_2 knot indicating that the simpler 3_1 knot has no significant free energy barriers opposing compaction and thus easily assumes a tight, compact structure at forces of 5 pN and above. It is possible to speculate on the potential biological consequences of this large difference in knot size between the 5_2 and the 3_1 knot. Our results suggest that 3_1-knotted proteins may be degradable by either the bacterial or eukaryotic degradation machinery as the tightened knot is approximately the same size as the channel in which a polypeptide chain has to be translocated through in order to reach the central proteolytic site (52-54). In contrast, degradation may be challenging for a 5_2-knotted protein, where a large 5_2 knot might block the entry pore of the degradation machine thus preventing or slowing degradation and also potentially thereby hindering the degradation of other cellular proteins. This is interesting as UCH-L1 is an abundant protein in human neurons and has been associated with a number of neurodegenerative diseases (31, 32). Additionally, an isoform of UCH-L1 is bound to the proteasome and had better protection against accidental degradation (55).

Influence of knotting on the folding kinetics of UCH-L1

The single-molecule mechanical assay used here provides a unique direct method to measure the effect of knot formation on folding rates. For UCH-L1, we observe significant effects of both 3_1 and 5_2 knots in the unfolded states on the folding rates. Folding from an unknotted denatured state is one order of magnitude slower than from knotted unfolded states (Fig. 4). These results establish that knotting, or a step associated with knotting, is rate limiting for the folding of this family of knotted proteins. This is consistent with recent *in vitro* translation and folding studies on trefoil-knotted methyltransferases (16). Note that the conclusion of knotting as the rate limiting step only holds if the back-reactions (unknotting and unfolding) are assumed to be slow compared to the overall timescale of folding. The overall timescale for folding observed here using single-molecule force spectroscopy (between 10 and 100 s) agrees well with previous measurements of folding rates in solution from a chemically denatured state (34). Our results can also be compared with values from computational studies where it is facile to create an unknotted variant of a knotted protein with the same secondary structural elements packed against each other in the same manner but with a different connectivity, essentially "rewiring" the polypeptide chain. Coarse-grained lattice-based approaches have been used to estimate the effect knots on the folding and unfolding rates of a model system (56). A 5_2 knot had an effect on both the folding rate ($k_f^{knot} = 0.61\ k_f^{unknot}$) and unfolding rate ($k_u^{knot} = 0.59\ k_u^{unknot}$). Remarkably, the folding rate was only decreased by a factor of two, compared with the factor of ten observed in this experimental study of a naturally occurring 5_2 knotted protein.

Folding Pathways

Both the mechanical unfolding and refolding studies conducted here establish that there are a large number of intermediate states populated during force unfolding measurements and during refolding at low force. These vary in contour length, therefore structure, and mechanical stability. Ideally, the data from the optical tweezers experiments could be used to reconstruct the energy landscape for folding/unfolding of UCH-L1 as was achieved for calmodulin (57). Unfortunately, the number and heterogeneity of intermediates observed, together with the fact that timescales of unfolding/folding events do not allow us to monitor equilibrium transitions that include folding to the native state, means it is currently not possible to do this for UCH-L1. However, important information on the folding pathway can still be extracted (see also discussion in the SI section §3.7).

The results of the optical tweezers experiments presented here are consistent with our unfolding and refolding kinetic studies on UCH-L1 (34). The overall timeframes for formation of the native state are similar and intermediate states are observed during unfolding and refolding in both cases. The results of the optical tweezers experiments focused on the refolding of mechanically unfolded UCH-L1 show that there are long-lived intermediate states, stable for quite some time even under load. It is highly likely that these are the same as the meta-stable states observed in both equilibrium and kinetic measurements on UCH-L1 folding using chemical denaturants and intrinsic fluorescence as a probe of structure. Thus, despite the fact that the two studies used different methods to unfold the protein, the results suggest that both report on the same key features of the folding energy landscape for this knotted protein. However, the single-molecule optical tweezers results provide considerably more detail on this landscape. Our results support the view that the folding of knotted proteins is a complex process with a large number of intermediate structures that may include non-native contacts (58, 59) and consists very likely of on-pathway states as well as off-pathway, kinetically trapped, states.

Conclusions

The exquisite control inherent in single-molecule force spectroscopy experiments has enabled us to control the knotted topology of an unfolded state of a protein. This has generated the unique ability to study the folding of a knotted protein from three different knotted states thereby establishing the effect of different knot types on folding rates and pathways. Here, we provide direct evidence that a threading event associated with formation of either a 3_1 or 5_2 knot, or a step closely associated with it, significantly slows down folding of UCH-L1. The results of the optical tweezers experiments highlight the complex nature of the folding of a knotted protein, and detect many additional intermediate structures that cannot be resolved by intrinsic fluorescence. Given the number of intermediates observed, it is likely that some of these are off pathway and we can speculate that these species may have a significant number of non-native contacts (58, 59). Mechanical stretching of knotted proteins is also of importance for understanding the possible implications of knots in proteins for cellular degradation. Our results highlight the potential difficulties in degrading a 5_2 knot compared with a 3_1 knot and therefore have possibly important implications for knotted proteins in proteostasis and associated disease states (14, 15, 32).

Methods

The engineering, expression and purification of the double-cysteine variants

of UCH-L1 were performed as described in the *SI Materials and Methods*. Protein characterization using far-UV CD, thermodynamic and kinetic folding experiments was carried out as described in the *SI Materials and Methods*. For the single-molecule mechanical measurements, a dumbbell configuration was generated by attaching the biotin/digoxigenin functionalized end of the DNA handles to micrometer-sized streptavidin/anti-digoxigenin silica beads (Fig. 1A). The beads were trapped in the foci of a custom-built dual beam optical-tweezer setup and subjected to stretch-and-relax cycles at a constant velocity or at a constant force bias with fixed trap positions. All measurements were conducted in PBS, pH 7.4. A complete description of the methods used is given in *SI Materials and Methods*.

Acknowledgements.

N.C.H.L. is supported by a UBD Chancellor's Scholarship from the Brunei Government. W.N. was supported by the Ministry of Education (MoE) Singapore. S.S.M acknowledges funding from the Alexander von Humboldt Foundation. This work was supported by an SFB 863 A2 grant of Deutsche Forschungsgemeinschaft to M.R.

1. Kuhlman B & Baker D (2004) Exploring folding free energy landscapes using computational protein design. *Curr Opin Struct Biol* 14(1):89-95.
2. Nickson AA, Wensley BG, & Clarke J (2013) Take home lessons from studies of related proteins. *Curr Opin Struct Biol* 23(1):66-74.
3. Reddy G, Liu Z, & Thirumalai D (2012) Denaturant-dependent folding of GFP. *Proc Natl Acad Sci U S A* 109(44):17832-17838.
4. Schaeffer RD, Fersht A, & Daggett V (2008) Combining experiment and simulation in protein folding: closing the gap for small model systems. *Curr Opin Struct Biol* 18(1):4-9.
5. Wolynes PG, Onuchic JN, & Thirumalai D (1995) Navigating the folding routes. *Science* 267(5204):1619-1620.
6. Whitford PC & Onuchic JN (2015) What protein folding teaches us about biological function and molecular machines. *Curr Opin Struct Biol* 30:57-62.
7. Wolynes PG (2015) Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie* 119:218-30.
8. Taylor WR (2007) Protein knots and fold complexity: Some new twists. *Comput Biol Chem* 31(3):151-162.
9. Mallam AL (2009) How does a knotted protein fold? *FEBS Journal* 276(2):365-375.
10. Boelinger D, *et al.* (2010) A Stevedore's Protein Knot. *PLoS Comput Biol* 6(4).
11. Virnau P, Mallam A, & Jackson S (2011) Structures and folding pathways of topologically knotted proteins. *J Phys Condens Matter.* 23(3):033101 (033117 pp.)-033101 (033117 pp.).
12. Sulkowska JI, Rawdon EJ, Millett KC, Onuchic JN, & Stasiak A (2012) Conservation of complex knotting and slipknotting patterns in proteins. *Proc Natl Acad Sci U S A* 109(26):E1715-E1723.
13. Faisca PF (2015) Knotted proteins: A tangled tale of Structural Biology. *Comput Struct Biotechnol J* 13:459-468.
14. Lim NC & Jackson SE (2015) Molecular knots in biology and chemistry. *J Phys Condens Matter* 27(35):354101.
15. Virnau P, Mirny LA, & Kardar M (2006) Intricate knots in proteins: Function and evolution. *PLoS Comput Biol* 2(9):e122.
16. Lim NCH & Jackson SE (2015) Mechanistic insights into the folding of knotted proteins *in vitro* and *in vivo*. *J Mol Biol* 427(2):248-258.
17. Mallam A, Rogers JM, & Jackson S (2010) Experimental detection of knotted conformations in denatured proteins. *Proc Natl Acad Sci U S A* 107(18):8189-8194.
18. Mallam AL & Jackson SE (2005) Folding studies on a knotted protein. *J Mol Biol* 346(5):1409-1421.
19. Mallam AL & Jackson SE (2006) Probing Nature's knots: The folding pathway of a knotted homodimeric protein. *J Mol Biol* 359(5):1420-1436.
20. Mallam AL & Jackson SE (2007) The dimerization of an alpha/beta-knotted protein is essential for structure and function. *Structure* 15(1):111-122.
21. Mallam AL & Jackson SE (2007) A comparison of the folding of two knotted proteins: YbeA and YibK. *J Mol Biol* 366:650-665.
22. Mallam AL & Jackson SE (2012) Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat Chem Biol* 8(2):147-153.
23. Mallam AL, Morris ER, & Jackson SE (2008) Exploring knotting mechanisms in protein folding. *Proc Natl Acad Sci U S A* 105(48):18740-18745.
24. Mallam AL, Onuoha SC, Grossmann JG, & Jackson SE (2008) Knotted fusion proteins reveal unexpected possibilities in protein folding. *Mol Cell* 30(5):642-648.
25. Noel JK, Sulkowska JI, & Onuchic JN (2010) Slipknotting upon native-like loop formation in a trefoil knot protein. *Proc Natl Acad Sci U S A* 107(35):15403-15408.
26. Sulkowska JI, Sulkowski P, & Onuchic J (2009) Dodging the crisis of folding proteins with knots. *Proc Natl Acad Sci U S A* 106(9):3119-3124.
27. King NP, Jacobitz AW, Sawaya MR, Goldschmidt L, & Yeates TO (2010) Structure and folding of a designed knotted protein. *Proc Natl Acad Sci U S A* 107(48):20732-20737.
28. Wang I, Chen SY, & Hsu ST (2015) Unraveling the folding mechanism of the smallest knotted protein, MJ0366. *J Phys Chem B* 119(12):4359-4370.
29. Wang LW, Liu YN, Lyu PC, Jackson SE, & Hsu ST (2015) Comparative analysis of the folding dynamics and kinetics of an engineered knotted protein and its variants derived from HP0242 of Helicobacter pylori. *J Phys Condens Matter* 27(35):354106.
30. Wilkinson KD, *et al.* (1989) The neuron-specific protein PGP 9.5 is a ubiquitin carboxyl-terminal hydrolase. *Science* 246(4930):670-673.
31. Gong B & Leznik E (2007) The role of ubiquitin C-terminal hydrolase L1 in neurodegenerative disorders. *Drug News Perspect* 20(6):365-370.
32. Setsuie R & Wada K (2007) The functions of UCH-L1 and its relation to neurodegenerative diseases. *Neurochem Int* 51(2-4):105-111.
33. Andersson FI, *et al.* (2011) The Effect of Parkinson's-Disease-Associated Mutations on the Deubiquitinating Enzyme UCH-L1. *J Mol Biol* 407(2):261-272.
34. Luo SC, *et al.* (2016) The knotted protein UCH-L1 exhibits partially unfolded forms under native conditions that share common structural features with its kinetic folding intermediates. *J Mol Biol* S0022-2836(16)30045-6.
35. Cecconi C, Shank EA, Bustamante C, & Marqusee S (2005) Direct observation of the three-state folding of a single protein molecule. *Science* 309(5743):2057-2060.
36. Junker JP, Ziegler F, & Rief M (2009) Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science* 323(5914):633-637.
37. Rief M, Pascual J, Saraste M, & Gaub HE (1999) Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J Mol Biol* 286(2):553-561.
38. Schlierf M, Li H, & Fernandez JM (2004) The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc Natl Acad Sci U S A* 101(19):7299-7304.
39. Yu H, Dee DR, & Woodside MT (2013) Single-molecule approaches to prion protein misfolding. *Prion* 7(2):140-146.
40. Dietz H & Rief M (2006) Protein structure by mechanical triangulation. *Proc Natl Acad Sci U S A* 103(5):1244-1247.
41. Das C, *et al.* (2006) Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1. *Proc Natl Acad Sci U S A* 103(12):4675-4680.
42. Dietz H & Rief M (2004) Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc Natl Acad Sci U S A* 101(46):16192-16197.
43. Guo Z & Thirumalai D (1995) Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. *Biopolymers* 36(1):83-102.
44. Thirumalai D, Klimov KD, & Woodson AS (1997) Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theor Chem Acc* 96(1):14-22.
45. Stigler J & Rief M (2012) Hidden Markov analysis of trajectories in single-molecule experiments and the effects of missed events. *Chemphyschem* 13(4):1079-1086.
46. Bowman GR & Pande VS (2010) Protein folded states are kinetic hubs. *Proc Natl Acad Sci U S A* 107(24):10890-10895.
47. He C, Lamour G, Xiao A, Gsponer J, & Li H (2014) Mechanically tightening a protein slipknot into a trefoil knot. *J Am Chem Soc* 136(34):11946-11955.
48. Bornschlogl T, *et al.* (2009) Tightening the knot in phytochrome by single-molecule atomic force microscopy. *Biophys J* 96(4):1508-1514.
49. Dzubiella J (2009) Sequence-specific size, structure, and stability of tight protein knots. *Biophys J* 96(3):831-839.
50. Pieranski P, Przybyl S, & Stasiak A (2001) Tight open knots. *Eur Phys J E* 6(2):123-128.
51. Kubelka J, Hofrichter J, & Eaton WA (2004) The protein folding 'speed limit'. *Curr Opin Struct Biol* 14(1):76-88.
52. Lowe J, *et al.* (1995) Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. *Science* 268(5210):533-539.
53. Wenzel T & Baumeister W (1995) Conformational constraints in protein degradation by the 20S proteasome. *Nat Struct Biol* 2(3):199-204.
54. Wang J, Hartling JA, & Flanagan JM (1997) The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* 91(4):447-456.
55. Nishio K, *et al.* (2009) Crystal structure of the de-ubiquitinating enzyme UCH37 (human UCH-L5) catalytic domain. *Biochem Biophys Res Commun* 390(3):855-860.
56. Soler MA, Nunes A, & Faisca PF (2014) Effects of knot type in the folding of topologically complex lattice proteins. *J Chem Phys* 141(2):025101.
57. Stigler J, Ziegler F, Gieseke A, Gebhardt JC, & Rief M (2011) The complex folding network of single calmodulin molecules. *Science* 334(6055):512-516.
58. Skrbic T, Micheletti C, & Faccioli P (2012) The role of non-native interactions in the folding of knotted proteins. *PLoS Comput Biol* 8(6):e1002504.
59. Wallin S, Zeldovich KB, & Shakhnovich EI (2007) The folding mechanics of a knotted protein. *J Mol Biol* 368(3):884-893.