

# SCIENTIFIC REPORTS



OPEN

## The AAA+ protease ClpXP can easily degrade a $3_1$ and a $5_2$ -knotted protein

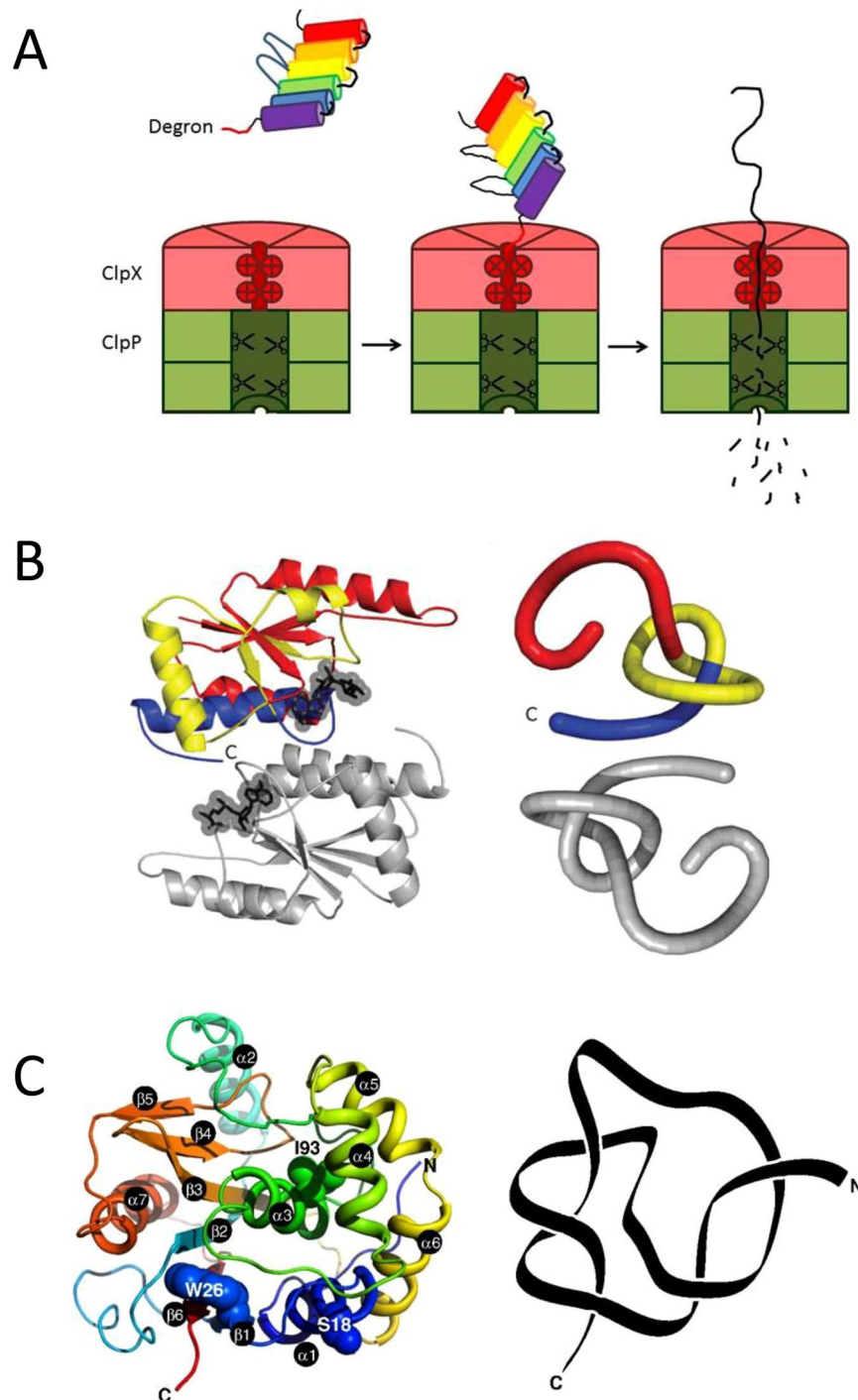
Elin M. Sivertsson<sup>1</sup>, Sophie E. Jackson<sup>2</sup> & Laura S. Itzhaki<sup>1</sup>

Knots in proteins are hypothesized to make them resistant to enzymatic degradation by ATP-dependent proteases and recent studies have shown that whereas ClpXP can easily degrade a protein with a shallow  $3_1$  knot, it cannot degrade  $5_2$ -knotted proteins if degradation is initiated at the C-terminus. Here, we present detailed studies of the degradation of both  $3_1$ - and  $5_2$ -knotted proteins by ClpXP using numerous constructs where proteins are tagged for degradation at both N- and C-termini. Our results confirm and extend earlier work and show that ClpXP can easily degrade a deeply  $3_1$ -knotted protein. In contrast to recently published work on the degradation of  $5_2$ -knotted proteins, our results show that the ClpXP machinery can also easily degrade these proteins. However, the degradation depends critically on the location of the degradation tag and the local stability near the tag. Our results are consistent with mechanisms in which either the knot simply slips along the polypeptide chain and falls off the free terminus, or one in which the tightened knot enters the translocation pore of ClpXP. Results of experiments on knotted protein fusions with a highly stable domain show partial degradation and the formation of degradation intermediates.

Protein degradation is an essential process in the cell, required to clear faulty or obsolete proteins, recycle amino acids and exert spatial and temporal control over cellular processes. ATP-dependent proteases, such as the 26S proteasome<sup>1</sup> in eukaryotes and ClpXP in bacteria<sup>2,3</sup>, carry out protein degradation guided by specific degradation signals. ATP-dependent proteases share the same overall architecture resembling a barrel, with a narrow central pore leading to an inner chamber where proteolytic active sites are located<sup>4,5</sup>. According to the model of translocation-coupled unfolding (Fig. 1A), the ATP-dependent protease recognizes a substrate from its degradation signal or degron, engages it in an unstructured region and uses the energy from ATP to mechanically pull the substrate towards the pore opening<sup>6</sup>. As a folded protein is too large to enter the pore, the pulling results in an unravelling of the protein structure at the same time as the polypeptide chain is translocated into the central chamber for proteolytic cleavage<sup>7</sup>. This model explains the finding that local stability is a more important determinant of degradation resistance than is global thermodynamic stability<sup>8</sup>.

Knotted proteins, in which the peptide backbone forms a topological knot, were first discovered in 1994 when Mansfield identified a single protein structure within the PDB with a very shallow knot, carbonic anhydrase<sup>9</sup>. The fact that carbonic anhydrase was unique in having a knotted chain, and that only a few residues needed to pass through a loop to form the very shallow knot, lead Mansfield to conclude that topological knots in proteins, particularly deep ones, were incompatible with protein folding mechanisms<sup>9</sup>. However, in 2000, Taylor identified many more proteins in the PDB with knots, some with deep knots<sup>10</sup>. Since then, an increasing number of knotted proteins have been found in the PDB using computational methods and mathematical knot theory<sup>11,12</sup>. Knots in proteins can be classified based on the number of times the polypeptide chain crosses itself when the chain is projected onto a two-dimensional plane. So far, knots with 3 (trefoil), 4 (figure-of-eight), 5 (Gordian) and 6 (stevedore) crossings have been discovered<sup>13</sup>. In addition, knots have been found with very different depths, depending upon the number of residues that can be removed from each terminus before the knot disappears<sup>14</sup>. The structure that remains after this procedure is referred to as the knotted core. For an up-to-date list of proteins which have either a knotted or slip-knotted structure there is an excellent database – KnotProt<sup>15</sup>. At the time of writing, there are a total of 1539 protein chains with knots in the database, 1040 of which are knotted, and 499 of which form slipknots<sup>15</sup>.

<sup>1</sup>Department of Pharmacology, Tennis Court Road, Cambridge, CB2 1PD, UK. <sup>2</sup>Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW, UK. Correspondence and requests for materials should be addressed to S.E.J. (email: [sej13@cam.ac.uk](mailto:sej13@cam.ac.uk)) or L.S.I. (email: [lsi10@cam.ac.uk](mailto:lsi10@cam.ac.uk))



**Figure 1.** ATP-dependent protease ClpXP and knotted protein substrates. (A) Cartoon of protein degradation by ClpXP. (B,C) Structure and reduced backbone representations of the knotted protein substrates used in this study. (B) *E. coli*  $\alpha/\beta$ -knot methyl transferase YbeA (dimeric, PDB 1NS5). Image from Mallam and Jackson<sup>29</sup>. (C) Human neuronal ubiquitin C-terminal hydrolase (UCH) UCH-L1 (PDB 2ETL). Image from Andersson *et al.*<sup>29</sup>.

How knotted proteins unfold and fold, and in the latter case how the chain threads through loops to form the different knotted structures observed, has been receiving increasing attention<sup>13,16</sup>. So far, experimental studies have investigated the folding of proteins with  $3_1$ <sup>17–29</sup>  $5_2$ <sup>30–34</sup> and  $6_1$ <sup>35</sup> knotted topologies, and computational studies have tackled  $3_1$ <sup>36–47</sup>  $5_2$ <sup>48,49</sup> and  $6_1$ <sup>50</sup> knotted systems. The first folding studies of knotted proteins were on YibK and YbeA,  $3_1$ -knotted methyltransferases from *Haemophilus influenzae* and *Escherichia coli*, respectively, both of which have a deep trefoil knot<sup>17,19,20</sup>. Both belong to the  $\alpha/\beta$ -knot methyltransferases (MTases) superfamily of proteins, and like all other family members, they dimerise in the native state, with the knotted region forming a large part of the dimer interface, Fig. 1B<sup>51</sup>. Here, we use YbeA as a model of a deeply  $3_1$ -knotted protein.

In addition to the trefoil-knotted proteins, experimental unfolding and folding studies have also been performed on more complex protein knots. The neuronal ubiquitin C-terminal hydrolase UCH-L1 is monomeric and contains a knot with five crossings in a  $5_2$  conformation (also called a Gordian knot), Fig. 1C<sup>52,53</sup>. For more complex knots such as this one, it is helpful to consider not only the knotted core of the chain as a whole, but also the knotting fingerprint of the protein<sup>54</sup>. The knotting fingerprint is represented by a matrix that shows the location and type of any knotted regions formed by the whole chain and sub-regions of the chain. Almost the entire chain of UCH-L1 forms a  $5_2$  knot. If 4 residues are removed from the C-terminus the  $5_2$  knot is abolished, but a  $3_1$  knot remains until a further 56 residues are removed. In contrast, if residues are progressively removed from the N-terminal end, only 5 residues need to be removed before the chain becomes completely unknotted<sup>54</sup>. Despite its complex topology, UCH-L1 can be reversibly unfolded *in vitro* with chemical denaturants<sup>31,55</sup> and experimental unfolding and refolding studies have elucidated many details of its folding pathway<sup>32,34,55</sup>. Of particular relevance to this study, recently optical tweezers were used to investigate the mechanical unfolding and subsequent refolding of UCH-L1<sup>33</sup>. Numerous intermediate states were observed in both unfolding and refolding experiments illustrating that the energy landscape for folding of UCH-L1 is very complex.

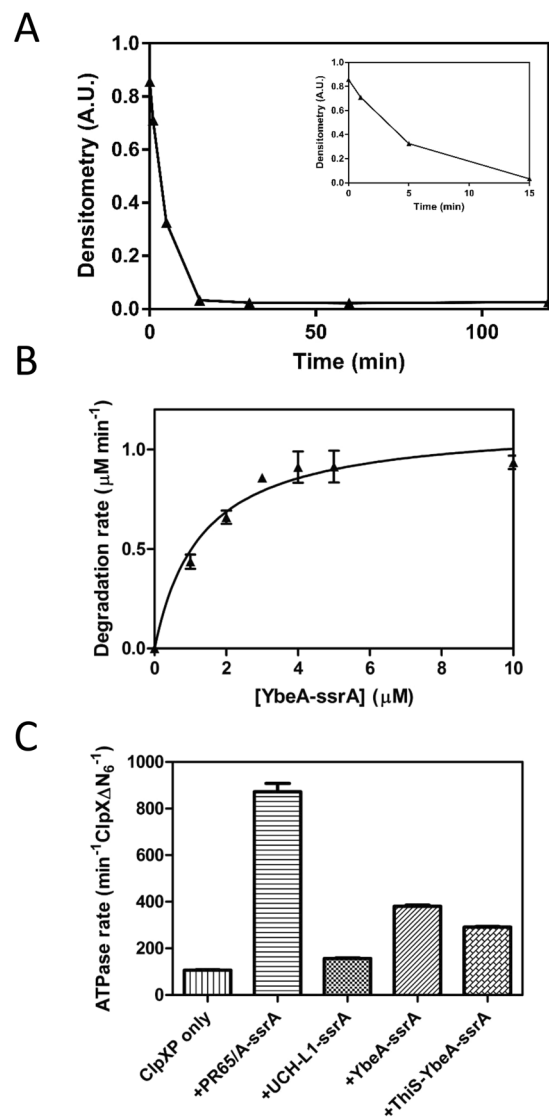
Both experimental and computational approaches have been used to investigate the effect of knotted topologies on the mechanical unfolding properties of the proteins. Simulations of two proteins having similar structures where one is knotted and the other is not showed that a knot can impart additional stability towards mechanical force<sup>36</sup>. In this case, the knot was observed to move during pulling simulations, preferentially ending up at positions of the chain with sharp turns, usually at proline or glycine residues<sup>36,56</sup>. Computational approaches have also been employed to investigate the probability of untying a knot in a protein and its dependence on pulling site, pulling speed and temperature<sup>57</sup>. AFM and more recently optical tweezers experiments have investigated the force-induced unfolding of a number of different knotted proteins. AFM experiments on phytochrome c, a protein with a figure-of-eight<sup>41</sup> knot, first established that mechanical pulling leads to unfolding and then knot tightening demonstrating that very tight knots which are densely packed can form even at biologically relevant forces<sup>58</sup>. In this case, the tightened knot comprised of 17 residues<sup>58</sup>. More recently the Li group mechanically unfolded a slip-knotted protein to a denatured state containing a trefoil knot and showed that the tightened knot comprised of 13 residues<sup>59</sup>. Most recently, a study using optical tweezers to force unfold the  $5_2$ -knotted protein UCH-L1 into  $5_2$ ,  $3_1$  and unknotted denatured states was published<sup>33</sup>. The trefoil-knotted state comprised of 12-13 residues in agreement with the Li study, and the  $5_2$ -knotted denatured state showed complex behaviour in which the knot first tightens to a state involving some 40 residues but upon the application of higher forces further tightens to a state comprising of 23 residues<sup>33</sup>.

The biological function of protein knots is not known, however, it has been hypothesised that the presence of a knot in a polypeptide chain may render a protein resistant to degradation by ATP-dependent proteases<sup>34</sup>. A number of published *in silico* studies of knotted proteins suggest this may be possible. For example, the results of simulations which mimic the pulling and translocation of a trefoil-knotted protein through a narrow pore have shown that the knot either tightens and blocks further translocation or it slides off the free end of the polypeptide chain as the protein is pulled through the pore depending upon the force used<sup>60-62</sup>. Recently, computational studies have simulated the translocation of knotted proteins through the pore in a simple model of a cellular degradation machine<sup>63</sup>. In this case, the simulations showed that the presence of a knot can hinder or even jam translocation<sup>63</sup>. More importantly, a recent experimental study showed that a shallow  $3_1$ -knotted protein MJ0366 could be degraded by ClpXP, and that a number of C-terminal fusions of MJ0366 with GFP resulted in stalling and partial degradation<sup>64</sup>. However, another recent study showed that when the degradation tag was attached to the C-terminus of  $5_2$ -knotted proteins from the ubiquitin C-terminal hydrolase family (UCH) little degradation was observed leading the authors to conclude that this class of knotted protein has unprecedented mechanostability<sup>65</sup>.

Here, we examine the ClpXP-catalysed degradation of two different knotted proteins, the deeply knotted YbeA which contains a  $3_1$  knot and the larger and more complex  $5_2$ -knotted protein UCH-L1. In contrast to previous studies, we initiate degradation from both N- and C-termini, we use destabilised variants of UCH-L1 to establish the role of local stability versus the knotted structure on degradation rates, and also we investigate the degradation of  $3_1$ - and  $5_2$ -knotted proteins fused to a small but highly stable domain that we show can withstand the mechanical unfolding force of ClpXP and therefore degradation. For the  $3_1$ -knotted YbeA we observe results similar in some ways to those recently published on MJ0366: the ClpXP degrades the protein with relative ease but when a super-stable ThiS domain is fused at the opposite termini to the degradation tag, we see partial degradation, and the formation of a degradation intermediate. For the  $5_2$ -knotted UCH-L1, we initiate degradation from both N- and C-termini and find significant differences in degradation rates. Using a number of destabilised variants, we can attribute this to differences in local stability rather than the knotted state of the protein, demonstrating for the first time that ClpXP can easily degrade larger more complex knotted structures as well as smaller  $3_1$ -knotted proteins. Thus, we demonstrate that the origin of the “unprecedented mechanostability” stated in previous work is local stability of the  $\beta$ -sheet structure and is not associated with the  $5_2$  knot. With UCH-L1 fusions, we also see partial degradation similar to the results for the  $3_1$  systems. The fact that even large knotted chains can enter ClpXP points to the remarkable flexibility of its central pore.

## Results

**Degradation of a trefoil-knotted protein.** First, we studied the degradation of the  $3_1$ -knotted methyltransferase YbeA, which contains a deep trefoil knot, with 70 and 34 residues lying to the N- and C-terminus of the knotted core, respectively (Fig. 1B). This is in contrast to the shallow trefoil knot of MJ0366 previously studied<sup>64</sup> where the numbers are 10 and 6 residues for the distance of the knot from the N- and C-terminus, respectively. YbeA was fused with the 11-amino acid *ssrA* degraon<sup>66,67</sup> at the C-terminus to produce YbeA-*ssrA*. To confirm that the knot in YbeA is still present after the addition of the *ssrA*-tag at the C-terminus, analytical gel filtration of YbeA-*ssrA* was performed (Fig. S1A). YbeA eluted at a volume of 16.3 ml, which corresponds to a  $K_{av}$

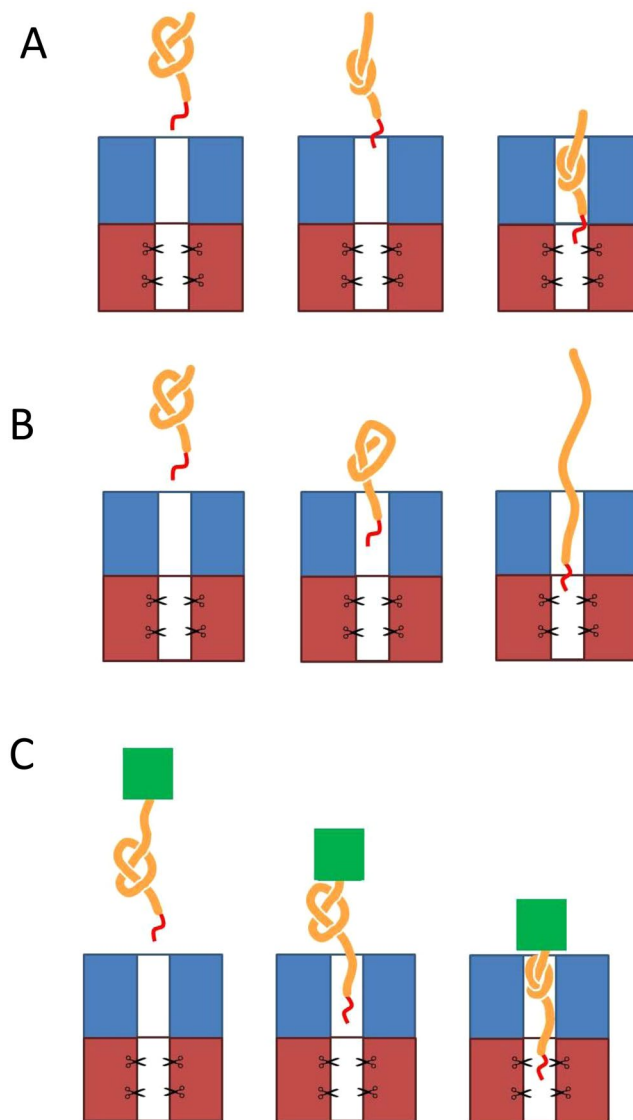


**Figure 2.** Degradation of the trefoil knotted protein YbeA-ssrA by ClpXP. **(A)** Degradation of YbeA-ssrA ( $10\mu\text{M}$ ) by ClpXP ( $0.5\mu\text{M ClpX}\Delta\text{N}_6$ ,  $1\mu\text{M ClpP}_{14}$ ), monitored by SDS-PAGE and densitometry. The reaction contained  $4\text{mM ATP}$  with regeneration system. Inset shows the first 15 minutes of the reaction. **(B)** Michaelis-Menten plot of YbeA-ssrA degradation. Initial rates of degradation were calculated from loss of gel band density of SDS-PAGE gels. Error bars represent SEM of triplicate measurements. The data are fitted to the Michaelis-Menten equation ( $v = V_{\text{max}} [S] / (K_M + [S])$ ), and the rate constant for degradation was calculated as  $k_{\text{deg}} = V_{\text{max}} / [\text{ClpX}_6]$ . **(C)** Steady state ATP hydrolysis rate of ClpXP ( $0.4\mu\text{M ClpX}\Delta\text{N}_6$ ,  $2.2\mu\text{M ClpP}_{14}$ ) at  $30^\circ\text{C}$ , alone or in the presence of ssrA-tagged substrate ( $10\mu\text{M}$ ). Error bars represent the SEM of at least three measurements.

of  $0.52$  and a molecular mass of  $38.9\text{kDa}$ . This value is close to the calculated molecular mass of a dimer,  $36.9\text{kDa}$ . As the dimer interface is largely made up of the knotted region<sup>18</sup>, the results suggest that the knot remains intact in YbeA-ssrA.

YbeA-ssrA was degraded rapidly by ClpXP (Figs 2A and S2A). The initial degradation rate at different concentrations of YbeA-ssrA was measured and the data were fitted to the Michaelis-Menten equation to extract the kinetic parameters (Fig. 2B,  $k_{\text{deg}} = 2.3 \pm 0.1 \text{ min}^{-1} \text{ ClpX}_6^{-1}$ ,  $K_M = 1.3 \pm 0.3 \mu\text{M}$ ). These numbers are comparable to those reported for the degradation of a series of variants of titin by ClpXP (ranging from  $0.25 < k_{\text{deg}} < 5 \text{ min}^{-1} \text{ ClpX}_6^{-1}$  and  $1.1 < K_M < 2.9 \mu\text{M}$ )<sup>68</sup>, and that recently reported for the  $3_1$ -knotted MJ0366 ( $k_{\text{cat}} = 3 \text{ min}^{-1}$ )<sup>64</sup>. The rate of ATP-hydrolysis by ClpXP was moderately stimulated during the degradation process (Fig. 2C).

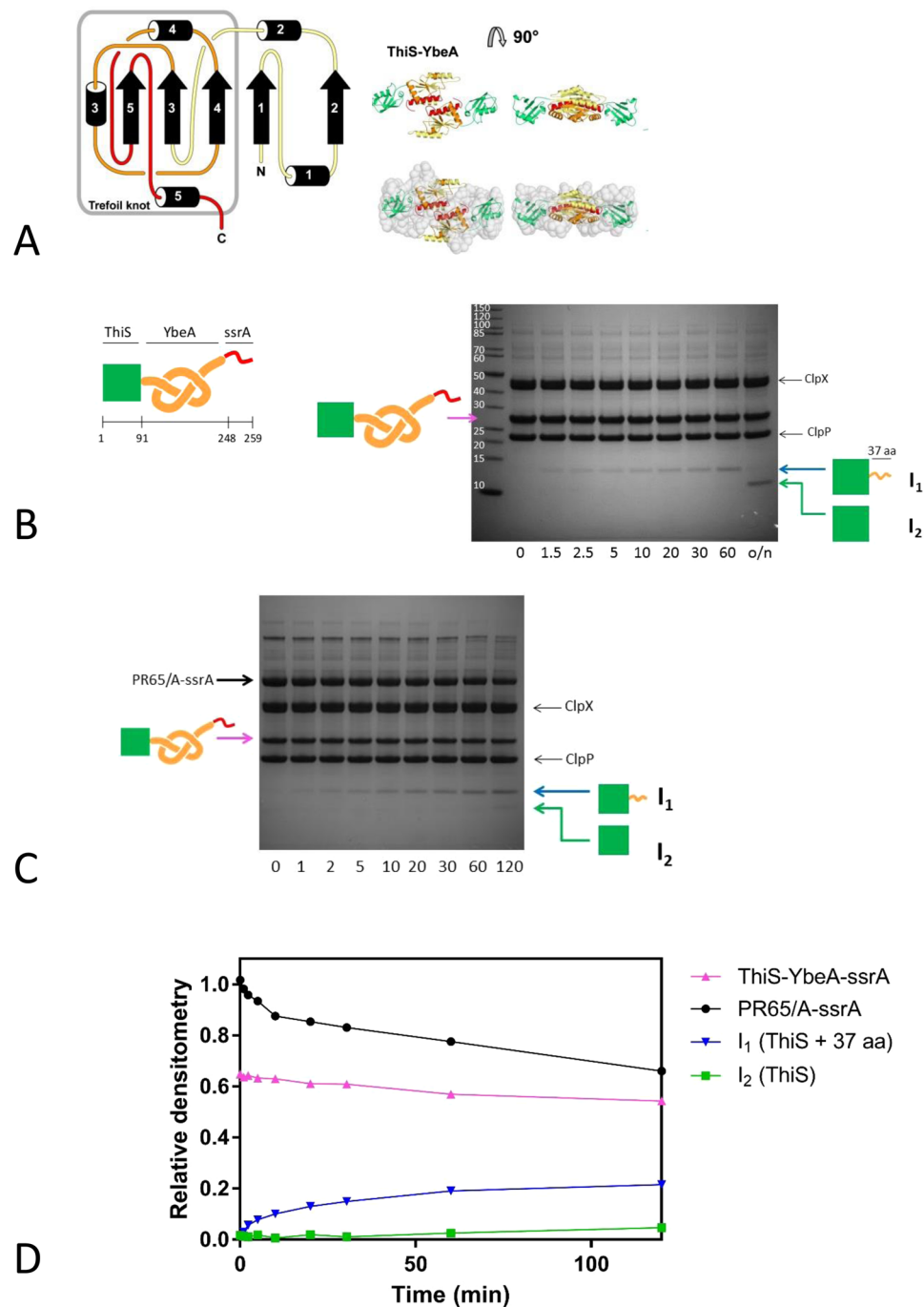
*Degradation of a trefoil-knotted fusion protein.* Is YbeA-ssrA degraded by ClpXP because the knot is small enough to fit through the translocation pore, Fig. 3A? Alternatively, does the knot slide off the free N-terminal end of the protein as its C-terminus is pulled through the ClpXP pore, Fig. 3B? To further investigate these questions, we made a ssrA-tagged fusion protein of YbeA and the small, very stable protein ThiS from the thermophile *Archaeoglobus fulgidus*<sup>69</sup> (Fig. 4A). It has been shown that despite an additional ThiS domain at either



**Figure 3.** Models of ClpXP-catalysed degradation for trefoil knotted proteins. Trefoil knotted protein (orange), ssrA-tag (red), ATP-dependent protease (blue/red), highly stable ThiS ‘plug’ domain (green). (A) Knot tightening and translocation into the degradation machinery. (B) Knot slips off the end of the polypeptide chain as the protein is unfolded and translocated into ClpXP. (C) Knot slipping along the chain before tightening and being blocked by the addition of a highly stable ThiS domain.

the N- or C-terminus of YbeA, the fusion proteins can refold into their native knotted structures both after urea denaturation<sup>69</sup> and after *in vitro* translation<sup>70</sup>. Here, we used a construct in which the ThiS domain was fused to the N-terminus of YbeA with a ssrA tag at its C-terminus (ThiS-YbeA-ssrA). During mechanical pulling of the protein at the C-terminus by the ClpXP machine, ThiS is expected to function as a ‘molecular plug’ blocking the N-terminus of YbeA and preventing the knot from sliding off the end of the chain (Fig. 3C), similar to the use of GFP fusions in the recent study on the shallow 3<sub>1</sub>-knotted MJ0366<sup>64</sup>. The fusion protein, ThiS-YbeA-ssrA ran as a dimer in size-exclusion chromatography (Fig. S1B), suggesting that it had assumed its native knotted state.

ThiS-YbeA-ssrA displayed complex degradation behaviour. There was loss of signal corresponding to the full-length protein and a gradual appearance of a smaller species in the gel when the reaction time points were analysed by SDS-PAGE (Fig. 4B). This small species was processed into an even shorter form after overnight incubation, Fig. 4B. Mass spectrometric analysis of the respective gel bands showed that the larger species (referred to as degradation intermediate I<sub>1</sub>) is a truncation of the fusion protein ending in position 128, Fig. S10A,B. As ThiS is 91-residues long and there is a short GlySer linker, this species corresponds to ThiS with an additional 37 residues (35 from YbeA and two from the GlySer linker) at the C-terminus. These 37 residues are consistent with the 30–37 residues known to span the distance from the ClpX opening to the proteolysis site in ClpP<sup>71,72</sup>. The smaller species (referred to as degradation intermediate I<sub>2</sub>) is a further truncation ending between positions 95 and 105, Fig. S10C,E, which is therefore likely to be ThiS on its own without the pore-spanning tail. We attribute this further processing to a contaminating protease as the I<sub>1</sub> intermediate does not have a ssrA tag. The ATPase



**Figure 4.** Degradation of the knotted fusion protein ThiS-YbeA. **(A)** Modelled structure of ThiS-YbeA, from Mallam *et al.*<sup>21</sup>. Left, topology of YbeA. The knotting loop is shown in orange, with the knotted chain in red. Right, modelled structure of ThiS-YbeA. YbeA is coloured as in the topology diagram, ThiS is in green. **(B)** Degradation of ThiS-YbeA-ssrA. Left, schematic representation of ThiS-YbeA-ssrA: ThiS (green square), YbeA (orange knot) and ssrA-tag (red line), with residue numbering. Right, degradation of ThiS-YbeA-ssrA (10  $\mu$ M) by ClpXP monitored by SDS-PAGE. **(C)** Degradation competition experiment with equimolar amounts of PR65/A-ssrA and ThiS-YbeA-ssrA (5  $\mu$ M each), monitored by SDS-PAGE and **(D)** subsequent densitometry. The reactions contained 0.5  $\mu$ M ClpX $\Delta$ N<sub>6</sub>, 1  $\mu$ M ClpP<sub>14</sub> and 4 mM ATP with regeneration system. Times in minutes indicated below the gel. Arrows indicate position of protein species as confirmed by gel band mass spectrometry analysis: PR65/A-ssrA (black circles), full-length ThiS-YbeA-ssrA (pink triangles), degradation intermediate I<sub>1</sub> (ThiS + 37 residues, blue inverted triangles), degradation intermediate I<sub>2</sub> (ThiS, green squares). In theory, the total amount of ThiS-YbeA-ssrA, I<sub>1</sub> and I<sub>2</sub> should equal 1, but it is greater than 1. We think that this discrepancy lies in the fact that the band for the full-length ThiS-YbeA-ssrA overlaps with the band in the gel from the ATP regeneration system, thus leading to some inaccuracies in the exact quantities of the species present. As MS data clearly show I<sub>1</sub> and I<sub>2</sub> come from the parent ThiS-YbeA-ssrA construct, we are not too concerned about this small difference.

rate during degradation of ThiS-YbeA-ssrA was stimulated moderately ( $291 \pm 3 \text{ min}^{-1} \text{ ClpX}_6^{-1}$ ), though lower than compared to YbeA-ssrA ( $380 \pm 5 \text{ min}^{-1} \text{ ClpX}_6^{-1}$ ) (Fig. 2C).

As the loss of ThiS-YbeA-ssrA is hard to observe, we repeated the degradation experiment in the presence of a competitive substrate, PR65/A-ssrA, which is a topologically simple helical repeat protein that we know is easily degraded by ClpXP (Fig. S4). In this way, we ascertained that ThiS-YbeA-ssrA did engage with the ClpXP machinery (Figs 4C and S3A) as the degradation of PR65/A-ssrA was slower in the presence of ThiS-YbeA-ssrA. The degradation intermediates appeared with time as expected and, in this experiment, it was clearer that the larger intermediate  $I_1$  was gradually being processed into the smaller intermediate  $I_2$  as expected (Fig. S3A). Densitometric analysis of the gel (Fig. 4D) showed that some of ThiS-YbeA-ssrA as well as some PR65/A-ssrA was degraded, but neither protein was degraded to completion. We repeated the experiment using three times the concentration of enzyme in the degradation assay. Densitometric analysis of SDS-PAGE gels (Fig. S3B,C) confirmed that the ratio of the  $I_1$  (or  $I_2$ ) band at maximum intensity and the ClpP band was the same in this experiment as when using the lower enzyme concentration, indicating that the increase in the amount of enzyme has resulted in a proportional increase in the amount of intermediate. Moreover, a larger fraction of full-length ThiS-YbeA-ssrA was degraded before the reaction came to a halt. Degradation-tagged ThiS on its own (ThiS-ssrA) was not degraded by ClpXP (Fig. S4).

**Degradation of a more complex 5<sub>2</sub>-knotted protein by ClpXP.** As our results, and those of the recent study by San Martin *et al.*<sup>64</sup> showed that ClpXP is able to degrade a 3<sub>1</sub>-knotted protein with relative ease, we moved on to study the degradation of a protein with a larger, more complex knot. UCH-L1 (ubiquitin C-terminal hydrolase L1) has a 5<sub>2</sub> knot that is shallow from the N-terminal end (only 5 residues need to be deleted to result in an unknotted chain) but deeper from the C-terminus (removal of some 4 residues results in formation of a deep trefoil knot which requires further deletion of more than 56 residues for the chain to become completely unknotted) (Fig. 1C). Recently, the Hsu group has shown that members of the UCH family including UCH-L1 withstand degradation by ClpXP when the ssrA degon is placed at the C-terminus<sup>65</sup>. Here, we carried out similar experiments and UCH-L1 was fused with the 11-amino acid ssrA degon at the C-terminus, but we additionally made a second variant in which we attached the ssrA degon at the N-terminus of UCH-L1 using a covalent chemical linkage strategy<sup>8</sup>.

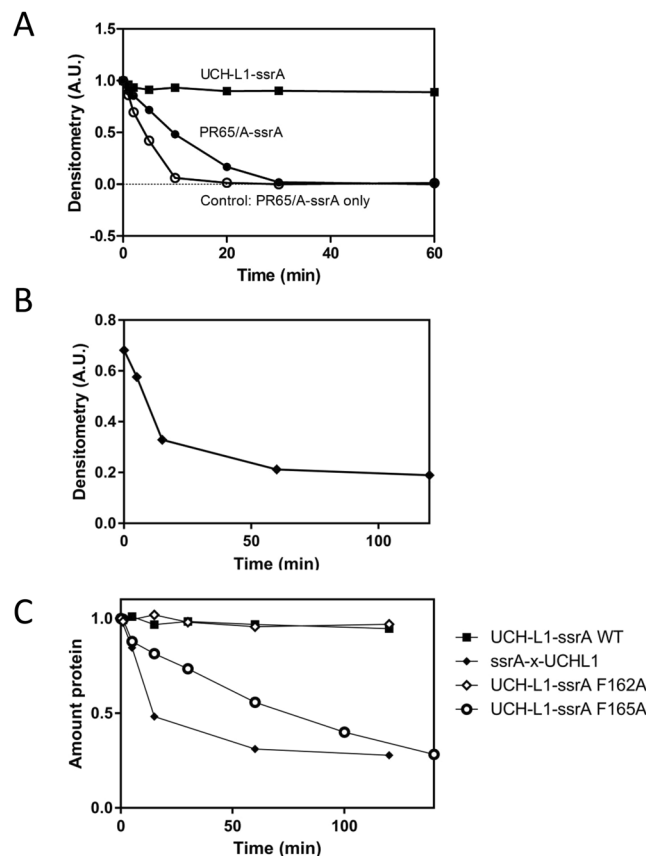
*Degradation of 5<sub>2</sub>-knotted UCH-L1.* First, we studied the degradation of UCH-L1-ssrA where the degon is located at the C-terminus of the knotted protein. We obtained very similar results to the Hsu group and did not detect any significant degradation of UCH-L1-ssrA by ClpXP over 60 mins, Fig. 5A. In a competition experiment with equimolar concentrations of UCH-L1-ssrA and the topologically simple helical repeat protein PR65/A-ssrA (Fig. 5A), the presence of UCH-L1-ssrA slowed down the degradation of PR65/A-ssrA about two-fold. This result indicates that ClpXP was able to bind to UCH-L1-ssrA but not degrade it.

ClpXP displays a basal rate of ATP hydrolysis even when not engaged in protein degradation and in the presence of ssrA-tagged substrate, the ATPase rate is stimulated<sup>73</sup>. The rate of ATP hydrolysis of ClpXP was not stimulated to any greater degree by the addition of UCH-L1-ssrA compared to the stimulation upon addition of PR65/A-ssrA (Fig. 2C), suggesting that although UCH-L1-ssrA engages with ClpXP it cannot be unfolded and translocated by it.

The 5<sub>2</sub> knot in UCH-L1 knot is located between residues 5-219, whereas a 3<sub>1</sub> knot exists between residues 5-163. Thus, deletion of a much smaller length of chain is required to fully unknot the protein from the N-terminal end (5 residues) than from the C-terminal end (more than 60 residues). Therefore, we hypothesized that the terminus at which degradation is initiated may make a difference for knot unfolding. To examine this question, a mutant of UCH-L1 with a single cysteine located close to the N-terminus (named Q2C) was employed. Circular dichroism showed that the native structure of the Q2C mutant is unperturbed (Fig. S5A) and that its thermal stability is identical to that of wild-type UCH-L1-ssrA (Fig. S5B).

A synthetic ssrA peptide was chemically crosslinked to the cysteine of the mutant protein<sup>8</sup>. The yield from the crosslinking reaction was approximately 50% (Fig. S6). The crosslinked species (ssrA-x-UCH-L1) was purified by nickel affinity using a hexahistidine tag in the peptide sequence prior to the degon sequence and used in degradation experiments with ClpXP. In contrast to the results obtained for UCH-L1-ssrA, there was visible degradation of ssrA-x-UCH-L1 in the presence of ClpXP and ATP (Figs 5B and S2C). The half-life was estimated to be 10 minutes for 10  $\mu\text{M}$  substrate protein. Low amounts of non-crosslinked protein without the ssrA-tag, visible as a weak band of slightly lower molecular weight than ssrA-x-UCH-L1, were present but not degraded by ClpXP, Fig. S2C.

*UCH-L1 mutants.* C-terminally tagged UCH-L1-ssrA was not noticeably degraded by ClpXP, Fig. 5A. N-terminally tagged ssrA-x-UCH-L1, on the other hand, was degraded by ClpXP, Fig. 5B. The fact that the C-terminal ssrA-tag is attached directly to  $\beta$ -strand 6, which is located at the centre of the core  $\beta$ -sheet structure, may explain the resistance of UCH-L1-ssrA to ClpXP-induced degradation. UCH-L1 unfolding by chemical denaturants proceeds via a stable intermediate, where the central part of this  $\beta$ -sheet remains highly structured<sup>31</sup>. It has previously been shown that the local structure adjacent to the degradation signal is more important than the overall thermodynamic stability of a protein in determining its resistance to degradation by ATP-dependent proteases<sup>71</sup>. It is possible that the stability of the secondary structure of the region where the tag is attached, rather than the overall knotted topology of the protein as a whole, may be a more important factor in the resistance of UCH-L1-ssrA to degradation by ClpXP. To test this possibility, we designed UCH-L1 mutants where the  $\beta$ -sheet is destabilised by the replacement of a large hydrophobic residue with an alanine. Mutants UCH-L1-ssrA F162A and UCH-L1-ssrA F165A were selected based on their theoretical destabilisation of 5.4 kcal mol<sup>-1</sup> and 5.9 kcal mol<sup>-1</sup>, respectively (calculated using FoldX<sup>74,75</sup>). The thermodynamic stability of these different substrates was compared in a thermal melt assay, monitored by far-UV circular dichroism (Fig. S5C). Although



**Figure 5.** Degradation of N- and C-terminally ssrA tagged variants of the  $5_2$ -knotted protein UCH-L1 by ClpXP. (A) ClpXP degradation competition experiment with equimolar amounts of UCH-L1-ssrA (closed squares) and PR65/A-ssrA (closed circles) ( $5\ \mu\text{M}$  each). Control reaction with PR65/A-ssrA ( $5\ \mu\text{M}$ ) without competitor shown as open circles. (B) Degradation of UCH-L1 with a crosslinked N-terminal ssrA-tag (ssrA-x-UCH-L1,  $10\ \mu\text{M}$ , closed diamonds). All reactions contained  $0.5\ \mu\text{M}$  ClpX $\Delta\text{N}_6$ ,  $1\ \mu\text{M}$  ClpP $_{14}$  and  $4\ \text{mM}$  ATP with regeneration system and were monitored by SDS-PAGE and densitometry. (C) Degradation of UCH-L1-ssrA variants by ClpXP ( $0.5\ \mu\text{M}$  ClpX $\Delta\text{N}_6$ ,  $1\ \mu\text{M}$  ClpP $_{14}$ ) monitored by SDS-PAGE and densitometry: UCH-L1-ssrA wild type ( $10\ \mu\text{M}$ , solid squares); F165A ( $9\ \mu\text{M}$ , open circles), F162A ( $10\ \mu\text{M}$ , open diamonds), ssrA-x-UCH-L1 ( $10\ \mu\text{M}$ , closed diamonds). The reactions contained  $4\ \text{mM}$  ATP with a regeneration system.

UCH-L1-ssrA F162A was somewhat stabilised by the mutation, UCH-L1-ssrA F165A was very destabilised. A UCH-L1 functional assay, which measures the hydrolysis of the fluorescent substrate Ub-AMC (Fig. S7), showed that both mutants retained enzymatic activity, suggesting that their native structures, and therefore knotted state, are unaltered.

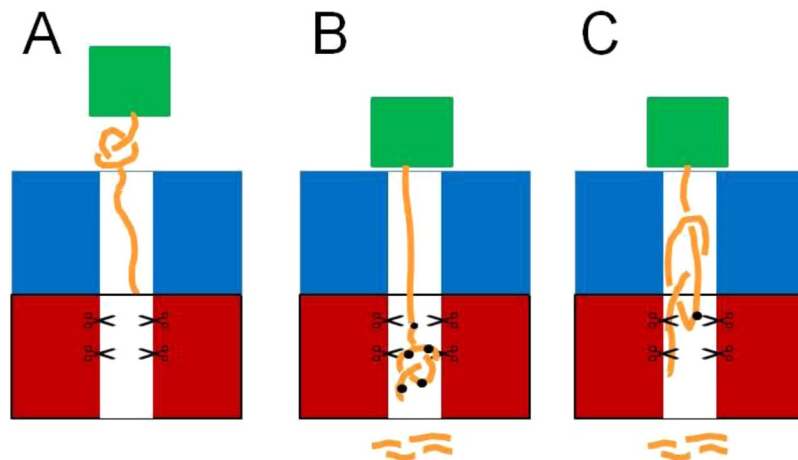
The somewhat stabilised UCH-L1-ssrA F162A was not degraded by ClpXP (Fig. 5C). However, the significantly destabilised UCH-L1-ssrA F165A was degraded by ClpXP (Fig. 5C), suggesting that the reduced thermodynamic stability of the  $\beta$ -sheet is sufficient to allow degradation.

Lastly, we examined an N-terminal fusion of UCH-L1 and the extremely stable ThiS domain. The fusion protein ThiS-UCH-L1-ssrA also displayed the appearance of a degradation intermediate (referred to as  $I_{\text{TU}}$ ) over time (Fig. S8). Mass spectrometry detected three fragments in the gel band, all C-terminal truncations starting at residue 1 and ending at residues 140, 142 and 143, respectively (Fig. S11). In the ThiS-UCH-L1-ssrA construct, the ThiS protein ends at position 109. That is, the detected intermediates are ThiS with a tail of 32–35 residues. The tightened knot in UCH-L1 is estimated from AFM studies to comprise roughly 40 residues up to  $40\ \text{pN}$  of force and is further tightened to 23–24 residues at higher forces<sup>33</sup>. Thus, the observed tail of ThiS-UCH-L1-ssrA is not long enough to accommodate a  $5_2$  knot, even if fully tightened. In fact, the tail is just long enough to span the distance from the ClpX opening to the ClpP proteolytic sites provided.

## Discussion and Conclusions

A significant number of knotted proteins have now been identified with differing knot complexities and sizes<sup>13,15</sup>. Mechanical unfolding studies have established that applying force to the N- and C-termini of these proteins results in unfolding and the subsequent formation of a tight knot in the denatured state<sup>33,58,59</sup>. The size of the tightened knot varies depending upon the knot type and for  $3_1$ -knotted proteins has been shown to comprise approximately of 12–14 residues and simulations suggest that tightened trefoil knots in proteins have a radius of gyration of around  $7\ \text{\AA}$ <sup>33,59,76,77</sup>. In contrast,  $4_1$ -knotted proteins have larger knots comprising some 15–16 residues<sup>58</sup> and the even larger  $5_2$ -knotted protein UCH-L1 shows complex behaviour and initially adopts a tightened





**Figure 6.** Models of the ClpXP-catalysed degradation of knotted fusion proteins. Trefoil knotted protein (orange), ATP-dependent protease (blue/red), highly stable ThiS ‘plug’ domain (green). (A) Knot tightening and abutting the stable ThiS domain outside of the ClpXP machinery. In this model, once the YbeA domain has been unfolded, the knot slips along the chain as translocation takes place until it reaches the ThiS domain at which point it can go no further. The knot does not enter the translocation pore. (B) The tightened knot enters the translocation pore and when it reaches the proteolytic active sites it is hydrolysed. The chain continues to translocate and be degraded until the stable ThiS domain abuts the surface of ClpXP. (C) The tightened knot slips along the polypeptide chain until it can slip no further as it abuts the ThiS domain. At this point, the knot can enter the translocation pore and part of the knotted region reaches the proteolytic sites and is hydrolysed.

knot comprising around 40 residues but tightens further to a smaller knot comprising some 23 residues at high forces<sup>33</sup>. Recently, San Martin and co-workers showed that, under some conditions the ClpXP machinery could easily degrade the shallow  $3_1$ -knotted protein MJ0366, whereas under other conditions the knot impaired degradation of specific multi-domain protein constructs<sup>64</sup>. In addition, another recently published paper reported that the ClpXP machinery only degrades a class of  $5_2$ -knotted proteins with unprecedented slow kinetics which the authors attribute to the remarkable mechanostability of the protein due to the large complex  $5_2$ -knot. However, in that study degradation was only initiated from the C-terminus<sup>65</sup>.

In this study, we have addressed the question of whether the ClpXP machine can degrade  $3_1$ - (trefoil) knotted proteins which contain very deep knots in their chains and also  $5_2$  knotted proteins if degradation is initiated from either N- or C-termini, and if the local stability near the degradation tag is reduced. Rapid degradation of the C-terminally *ssrA*-tagged YbeA was observed (Fig. 2) similar to the results on a C-terminally *ssrA*-tagged MJ0366, demonstrating that trefoil-knotted proteins, if they are not fused to any other domains, can be easily degraded by ClpXP and that it does not matter whether the  $3_1$  knot is shallow or, as we have shown here, deep. In the case of the rapid degradation of YbeA-*ssrA* by ClpXP we can imagine two different scenarios (we ignore the possibility that partial degradation occurs followed by release, unknotting and then rebinding and further degradation as the protein loses its degenon as soon as degradation begins). Scenario (i) Once the ClpXP machinery has unfolded the YbeA domain, the knot is displaced along the chain as translocation takes place until it falls off the free end (Fig. 3B) and (ii) the knot is sufficiently small to be translocated through the ClpXP pore and is thus degraded when it reaches the proteolytic domain (Fig. 3A).

In order to investigate whether scenario (i) or (ii) takes place, we engineered a variant of YbeA-*ssrA* in which a small but highly stable domain, ThiS, was fused to the N-terminus of YbeA. We tested ThiS on its own with a C-terminal degenon, ThiS-*ssrA*, and found that it is not degraded by ClpXP to any degree (Fig. S4). We assume that this is due to the very high kinetic and thermodynamic stability of ThiS (Fig. S5D) and<sup>69,70</sup>. ThiS-YbeA-*ssrA* showed complex behaviour in the degradation assays (Fig. 4) and the results shed light on whether scenario (i) or (ii) is occurring.

Scenario (i): In this case, ClpXP engages with the fusion protein through the *ssrA* degenon at the C-terminus and unfolds the YbeA domain retaining the  $3_1$  knot. It subsequently translocates and degrades the chain with the knot slipping along the chain until it abuts the very stable ThiS domain (Fig. 6A). At this point, the knot should become tightened and the degradation machinery stall until the chain dissociates from it. If this happens, a degradation intermediate corresponding to the ThiS domain plus approximately 50 residues (the length of the tightened knot plus the length of chain that spans ClpXP from the pore opening to the proteolytic sites). We do observe a degradation intermediate but it is significantly shorter than this, suggesting that this scenario is incorrect. We can also rule out similar scenarios where the knot tightens before it has reached the ThiS domain (Fig. S9) that would result in degradation intermediates of even longer length.

In scenario (ii), after the ClpXP machinery engages with the C-terminal *ssrA* tag and unfolds the YbeA domain, the knot (which we assume will be tightened at some point in the chain) can translocate into the pore and to the proteolytic sites (Fig. 6B). Translocation and degradation will continue until the stable folded ThiS domain abuts the ClpXP and then will stop (Fig. 6B). In this case, one would expect to observe a degradation intermediate that is the length of ThiS plus approximately 37 residues, the amount of chain required to span the non-proteolytic

domain of ClpXP. This is what we observe (Figs 4B and S3) and this result together with recent results on the shallow  $3_1$ -knotted MJ0366 suggests that the pore of ClpXP is large and flexible enough to accommodate a tightened  $3_1$  knot. This is perhaps not surprising given that it has been shown in computational studies that the diameter of a tightened  $3_1$ -knot is approximately  $14 \text{ \AA}$ <sup>76,77</sup> similar to the diameter of the translocation pore of ClpXP<sup>78</sup>. In addition, it has been shown that ClpXP can translocate disulphide-linked chains, meaning that at least two, or even three, polypeptide chains can fit in the ClpXP pore at the same time<sup>73</sup>. Indeed, the 26S proteasome is able to degrade disulphide-linked substrates with at least three, possibly up to five, chains passing through the translocation channel at once, albeit with lower degradation rate and efficiency<sup>79</sup>.

ClpXP was shown to degrade the trefoil-knotted protein YbeA with relative ease, perhaps unsurprising given that it has been estimated that the ClpXP opening is able to accommodate up to three disulphide-bonded chains that would require an opening of at least  $20\text{--}25 \text{ \AA}$ . With an estimated diameter of approximately  $14 \text{ \AA}$ , a trefoil knotted chain can translocate into the ClpXP pore and be degraded, as we and others<sup>64</sup> have observed. Recently, other groups have shown that the ClpXP machinery degrades  $5_2$ -knotted proteins from the UCH family with unprecedentedly slow kinetics and have attributed this to extremely high mechanostability due to the knotted structure<sup>65</sup>. Remarkably though, and in contrast to the earlier study, our results clearly establish that ClpXP can also degrade the more complex  $5_2$ -knotted protein UCH-L1 if the local stability of the protein near the degron is not too high (Fig. 5C) or if degradation is initiated from the N-terminus (Fig. 5B). We note that our experiments were performed in a buffer that did not contain monovalent cation whilst those of Hsu and coworkers used  $100 \text{ mM KCl}$ <sup>65</sup>; however, we have strong evidence that this is unlikely to be the origin of the differences observed. First, the results of our degradation assays on the same constructs (UCH-L1-ssrA) are the same as those from the Hsu group, second, when we destabilise the C-terminal region of the  $\beta$ -sheet by mutation we see efficient degradation even when degradation is initiated from the C-terminus.

Our results conclusively show that the  $5_2$  knot in UCH-L1 does not confer mechanostability on the protein as degradation is efficient if initiated from the N-terminus or from the C-terminus if the local stability is decreased by mutation. This leads to the question of how the ClpXP machinery degrades this  $5_2$ -knotted protein. It is possible that, in the absence of a stable structured domain blocking the knot from falling off the end of the polypeptide chain, the  $5_2$  knot simply slips along the chain until it drops off the end, Fig. 3B. These results suggest that, after unfolding of the UCH-L1 structure, the  $5_2$  knot is sufficiently dynamic to move along the chain as translocation takes place or that the  $5_2$  knot, although large, can enter the translocation pore.

Finally, we also studied an N-terminal fusion of the  $5_2$ -knotted UCH-L1 with the very stable ThiS domain and obtained similar results to those obtained for the ThiS-YbeA-ssrA. In this case, incomplete degradation was observed in addition to a degradation intermediate (referred to as  $I_{TU}$ ), Fig. S8. Mass spectrometric analysis of this intermediate revealed that it corresponded to the ThiS domain with a C-terminal tail of some  $32\text{--}35$  residues, Fig. S11. From these results, we can conclude that the tightened knot must be able to penetrate the ClpXP translocation pore as it has been estimated from optical tweezer studies on the mechanical unfolding of UCH-L1 that the  $5_2$  knot comprises roughly 40 residues up to  $40 \text{ pN}$  of force and even if it is further tightened by higher forces only reduces to some  $23\text{--}24$  residues<sup>33</sup>. Thus, the observed tail of ThiS-UCH-L1-ssrA is not long enough to accommodate a  $5_2$  knot, even if fully tightened. In fact, the tail is just long enough to span the distance from the ClpX opening to the ClpP proteolytic sites provided it is in an extended unfolded conformation. Therefore, the  $5_2$ -knot must either be able to pass through the translocation pore and be hydrolysed by the ClpP machinery, Fig. 6B, or it becomes stuck within the ClpXP machine but at least parts of the knotted regions are sufficiently close to the proteolytic sites to be hydrolysed, Fig. 6C.

The degradation of the knotted fusion proteins results in the formation of intermediates which have also been seen in the degradation of other multi-domain proteins<sup>80–82</sup> and also fusions of the shallow trefoil-knotted MJ0366<sup>64</sup>. Here, we have demonstrated that the incomplete degradation of deeply-knotted trefoil fusions and also fusions of  $5_2$ -knotted proteins can result in the production of stable partially degraded intermediate states. It has been proposed that these processes can give rise to potential new biological activities<sup>64,80–82</sup>. It is interesting to speculate whether this might be the case for the knotted proteins studied here. Inspection of the Pfam database shows that there exist three architectures (3055 sequences) that have a deep trefoil-knotted MTase domain fused to another domain at either N- or C-termini (PF02590). Remarkably, there are 46 different architectures (2769 sequences) where the more complex  $5_2$ -knot found in the UCH family of deubiquitinating enzymes is fused to another structured domain (PF01088). These results raise the possibility that partial degradation of a knotted fusion protein may play a role in the regulation of biological activity of the additional domain.

To the best of our knowledge, this is the first experimental study of the degradation of a trefoil-knotted protein with a deep knot in its structure and also the first study of the degradation of a  $5_2$ -knotted protein by an ATP-dependent protease initiating degradation from both termini and also investigating the role of local stability versus knotted structures in determining degradation rates. Overall, our data suggest that the ClpXP machinery is easily able to degrade a deeply  $3_1$ -knotted protein and a  $5_2$ -knotted protein. It is possible that, in these cases, once the knotted protein has unfolded the knot simply slips along the polypeptide chain and falls off the free terminus, although our results on knotted fusion proteins have also shown that in the case of both  $3_1$ - and  $5_2$ -knotted proteins it is possible for the tightened knot to enter the translocation pore.

## Experimental Procedures

**Plasmid and protein preparation.** A truncated form of ClpX lacking the N-terminal domain (referred to as ClpX $\Delta$ N) was used in all experiments. Residues  $1\text{--}62$  of full-length *Escherichia coli* ClpX in plasmid pRSETA were deleted using PCR mutagenesis to construct ClpX $\Delta$ N. His-tagged ClpX $\Delta$ N was expressed in *E. coli* BL-21 (DE3) cells. Cultures were grown in  $2\times$ TY media until  $OD_{600} = 0.7$ , when the temperature was reduced to  $20^\circ\text{C}$  before induction with  $100 \mu\text{M}$  isopropyl- $\beta$ -D-thiogalactoside (IPTG) overnight. The protein was purified using Ni-NTA affinity (Qiagen) followed by size-exclusion chromatography on a Superdex 75 column (GE Healthcare)

in ClpX storage buffer (25 mM Tris pH 7.6, 200 mM KCl, 2 mM EDTA, 1 mM DTT). His-tagged ClpP from *E. coli* was expressed from the pET21a plasmid in *E. coli* C41 (DE3) cells. Cultures were grown until  $OD_{600} = 0.8$  and induced with 100  $\mu$ M IPTG overnight at 26 °C. Purification was carried out essentially as described previously<sup>83</sup>, using Ni-NTA affinity and ion exchange chromatography on a monoQ column (GE Healthcare), before dialysis into ClpP storage buffer (50 mM Tris-HCl pH 7.6, 1 mM DTT, 0.5 mM EDTA, 100 mM KCl, 10% glycerol).

The *ssrA* sequence (ANDENYALAA) was added to the C-terminus of the protein substrates (UCH-L1 in pRSET; YbeA, ThiS and ThiSYbeA in pET17b) using PCR mutagenesis. All tagged substrate proteins were expressed in *E. coli* C41 (DE3) cells.

The protocol for purification of His-tagged UCH-L1-*ssrA* was based on methods reported previously<sup>30,31,52,84</sup>. Bacterial cultures were grown until  $OD_{600} = 0.8$  before induction (400  $\mu$ M IPTG, 2.5 h, 37 °C). Collected cell pellets were resuspended in UCH-L1 lysis buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM EDTA, 1 mM DTT) with protease inhibitor. The cells were lysed using an Emulsiflex cell cracker, and soluble protein was purified by Ni-NTA affinity before size-exclusion chromatography on a Superdex 75 column in UCH-L1 storage buffer (50 mM Tris-HCl pH 7.6, 0.5 mM EDTA, 5 mM DTT).

YbeA-*ssrA*, YbeA-ThiS-*ssrA* and ThiS-*ssrA* were purified essentially as described in<sup>20,21</sup>. Cultures were grown until  $OD_{600} = 0.8$  and induced with 400  $\mu$ M IPTG for 5 h at 37 °C. Collected cell pellets were resuspended in YbeA lysis buffer (20 mM Tris-HCl pH 7.5, 200 mM KCl, 10% glycerol, 1 mM DTT) before cell cracking. For YbeA-*ssrA*, the soluble fraction of the cell lysate was loaded on a Q Sepharose FF column (GE Healthcare). As expected from the theoretical pI of YbeA-*ssrA* (pI = 6.84), the protein appeared in the flow-through. The flow-through was diluted in 50 mM Tris-HCl pH 8.7, 1 mM DTT and loaded on to a monoQ column. Unexpectedly, the protein did not bind the column but appeared in the flow-through. However, several contaminants were removed by this purification step. Finally, the flow-through was concentrated using spin concentrators (Vivaspin, GE Healthcare) before size-exclusion chromatography on a HiLoad 26/60 Superdex 75 column in YbeA storage buffer (50 mM Tris-HCl pH 7.5, 200 mM KCl, 10% glycerol, 1 mM DTT). For ThiS-YbeA-*ssrA* (theoretical pI = 5.70), cell lysate was diluted in 50 mM Tris-HCl pH 8.4, 50 mM KCl, 1 mM DTT, and loaded on to a Q Sepharose FF column and eluted with a gradient to 1 M KCl. Protein-containing fractions were diluted in 50 mM Tris-HCl pH 8.4, 1 mM DTT and purified further on a monoQ column with a shallow gradient to 1 M KCl. Finally, the protein was desalted on a HiLoad 26/60 Superdex 75 column in YbeA storage buffer.

For ThiS-*ssrA*, cell lysate was diluted in 50 mM Tris-HCl pH 7.5, loaded on a Q Sepharose FF column and eluted with a gradient to 1 M NaCl. Protein-containing fractions were pooled and run on a HiLoad 26/60 Superdex 75 column in ThiS storage buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl).

Protein purity was assayed by SDS-PAGE, and protein masses were confirmed by mass spectrometry (PNAC Facility, Cambridge). Protein concentrations were determined spectrophotometrically at 280 nm, using extinction coefficients calculated with the ProtParam tool (<http://web.expasy.org/protparam/>).

**Enzymatic assays.** Degradation assays were performed in degradation buffer (25 mM HEPES-KOH pH 7.6, 5 mM MgCl<sub>2</sub>, 0.032% Igepal CA-630, 10% glycerol). A reaction containing 0.5  $\mu$ M ClpX $\Delta$ N<sub>6</sub>, 1  $\mu$ M ClpP<sub>14</sub> and an ATP regeneration system (4 mM ATP, 16 mM creatine phosphate, 0.32 mg/ml creatine kinase) was pre-incubated at 25 °C for 2 min. To start the reaction, pre-warmed substrate in degradation buffer was added. The degradation process was monitored by SDS-PAGE and subsequent densitometry analysis (ImageJ). The amount of ClpP at each time point was used as an internal standard to which the amount of remaining substrate protein was normalised. The data were plotted in GraphPad Prism and fitted to a single exponential. The data were normalized using the equation:

$$y_{norm} = \frac{y_t - y_{\infty}}{y_0 - y_{\infty}}$$

where  $y_t$  is the signal at time  $t$ ,  $y_{\infty}$  is the signal at infinite times and  $y_0$  is the initial signal at  $t = 0$ .

The initial rates of the reactions were calculated from the slope of the initial linear phase, taking into account the starting concentration of substrate and the total concentration of enzyme.

Although degradation assays were frequently repeated to ensure reproducibility, often the time points used were different and, therefore, calculations of errors difficult. However, the degradation assays and rates used in the calculation of Michaelis-Menten kinetics were repeated in triplicate under the same conditions. In this case, the largest errors were no more than 20% and frequently smaller than this, see Fig. 2B. For experiments on UCHL1-*ssrA*, repeats employed the same time points and here we have calculated the mean and the standard deviation, Fig. S12.

**Stability assays.** Thermal stability assays were monitored by far-UV circular dichroism on a Chirascan spectrometer (Applied Photophysics). Protein samples (10  $\mu$ M) in a 1 mm pathlength cuvette were heated from 20–90 °C in 2 °C steps with 2 min equilibration at each temperature. Ellipticity scans were performed between 220–240 nm, stepsize 1 nm.

**Mass spectrometry.** Identification of gel band proteins by peptide mass fingerprinting using trypsin and/or LysC digestion and MALDI mass spectrometry was performed by the Protein & Nucleic Acid Chemistry Facility (PNAC), Cambridge.

**Size-exclusion chromatography.** Analytical size exclusion chromatography (SEC) was performed on an S200 10/300 GL column (GE Healthcare). A 500  $\mu$ L loop was used for sample loading, and the sample was eluted at a flow rate of 0.5–1 mL min<sup>-1</sup>. The gel phase distribution parameter,  $K_{av}$ , was calculated as

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

where  $V_e$  is the elution volume measured from the center of the eluted peak,  $V_0$  is the void volume of the column determined from the elution volume of Blue Dextran 2000 and  $V_t$  is the total volume of the column. A calibration curve was prepared using the proteins in the Low Molecular Weight Gel Filtration Kit (GE Healthcare). The  $K_{av}$ -value of each protein standard was plotted against the logarithm of its molecular weight. The linear regression of this plot was used to determine the molecular weight of the YbeA-ssrA and ThiS-YbeA-ssrA.

**Chemical crosslinking of ssrA-peptide.** A synthesised ssrA peptide was attached to a mutant of UCHL-1 where all natural cysteines had been removed and a single cysteine had been inserted in a position close to the N-terminus (referred to as UCH-L1 Q2C). Chemical crosslinking of the ssrA peptide to the cysteine was performed in a two-step reaction using the heterobifunctional crosslinker sulfo-SMCC (Pierce, Thermo Scientific). This crosslinker has an amine-reactive succinimidyl ester at one end and a sulfhydryl reactive maleimide group at the other, separated by a cyclohexane spacer arm. The ssrA-peptide ( $\text{NH}_2$ -GGWDHHHHHHAANDENYALAA-COOH, Pepceuticals) was dissolved in 100 mM sodium phosphate, pH 6.5, 30 mM NaCl to a final concentration of 6.4 mM. The Sulfo-SMCC crosslinker was dissolved in ultrapure water to a final concentration of 30 mM. UCH-L1 Q2C (45  $\mu\text{M}$ ) in UCH-L1 storage buffer was reduced with 5 mM TCEP for 1 h at room temperature. The protein was then buffer exchanged by two passes through a 7 kDa cut-off spin desalting column (Zeba spin, Thermo Scientific) into PBS, 2 mM EDTA. Reaction 1 (conjugation of crosslinker to peptide) was carried out by mixing crosslinker in 10 times molar excess with peptide and incubating for 40 min at room temperature with agitation. Reaction 1 was quenched by adding glycine (100 mM final concentration). Excess unreacted crosslinker was removed from the peptide by chromatography on a column with a 1.8 kDa exclusion limit (Pierce, Thermo Scientific). Fractions containing peptide were identified by absorbance at 280 nm. Reaction 2 (conjugation of activated peptide-crosslinker to the protein cysteine) was carried out by mixing reduced protein with activated, purified peptide in 20 times molar excess before overnight incubation at 4 °C with rotation. Reaction 2 was quenched by addition of 5 mM DTT. Excess activated peptide was removed by desalting on a PD-10 column (GE Healthcare) into PBS, 1 mM DTT. Fractions containing the final ssrA-tagged protein were identified by measuring the absorbance at 280 nm and at 240 nm (indicative of the crosslinker), and by SDS-PAGE. The ssrA-tagged protein was separated from unreacted protein using Ni-NTA affinity chromatography.

## Data Availability

All materials and data are available from SEJ and LSI (sej13@cam.ac.uk and lsi10@cam.ac.uk).

## References

- Ciechanover, A. Intracellular protein degradation: from a vague idea through the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Bioorganic & medicinal chemistry* **21**, 3400–3410 (2013).
- Bittner, L. M., Arends, J. & Narberhaus, F. Mini review: ATP-dependent proteases in bacteria. *Biopolymers* **105**, 505–517 (2016).
- Olivares, A. O., Baker, T. A. & Sauer, R. T. Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nature reviews. Microbiology* **14**, 33–44 (2016).
- Baker, T. A. & Sauer, R. T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochimica et biophysica acta* **1823**, 15–28 (2012).
- Wehmer, M. & Sakata, E. Recent advances in the structural biology of the 26S proteasome. *The international journal of biochemistry & cell biology* **79**, 437–442 (2016).
- Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A. & Sauer, R. T. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**, 744–756 (2009).
- Cordova, J. C. *et al.* Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* **158**, 647–658 (2014).
- Kenniston, J. A., Burton, R. E., Siddiqui, S. M., Baker, T. A. & Sauer, R. T. Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. *Journal of structural biology* **146**, 130–140 (2004).
- Mansfield, M. L. Are there knots in proteins? *Nature structural biology* **1**, 213–214 (1994).
- Taylor, W. R. A deeply knotted protein structure and how it might fold. *Nature* **406**, 916–919 (2000).
- Kolesov, G., Virnau, P., Kardar, M. & Mirny, L. A. Protein knot server: detection of knots in protein structures. *Nucleic acids research* **35**, W425–428 (2007).
- Lai, Y. L., Chen, C. C. & Hwang, J. K. pKNOT v.2: the protein KNOT web server. *Nucleic acids research* **40**, W228–231 (2012).
- Jackson, S. E., Suma, A. & Micheletti, C. How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. *Current opinion in structural biology* **42**, 6–14 (2016).
- Millett, K. C., Rawdon, E. J., Stasiak, A. & Sulkowska, J. I. Identifying knots in proteins. *Biochemical Society transactions* **41**, 533–537 (2013).
- Jamroz, M. *et al.* KnotProt: a database of proteins with knots and slipknots. *Nucleic acids research* **43**, D306–314 (2015).
- Faisca, P. F. Knotted proteins: A tangled tale of Structural Biology. *Computational and structural biotechnology journal* **13**, 459–468 (2015).
- Mallam, A. L. & Jackson, S. E. A comparison of the folding of two knotted proteins: YbeA and YibK. *Journal of molecular biology* **366**, 650–665 (2007).
- Mallam, A. L. & Jackson, S. E. The dimerization of an alpha/beta-knotted protein is essential for structure and function. *Structure* **15**, 111–122 (2007).
- Mallam, A. L. & Jackson, S. E. Probing nature's knots: the folding pathway of a knotted homodimeric protein. *Journal of molecular biology* **359**, 1420–1436 (2006).
- Mallam, A. L. & Jackson, S. E. Folding studies on a knotted protein. *Journal of molecular biology* **346**, 1409–1421 (2005).
- Mallam, A. L., Morris, E. R. & Jackson, S. E. Exploring knotting mechanisms in protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 18740–18745 (2008).

22. Mallam, A. L., Rogers, J. M. & Jackson, S. E. Experimental detection of knotted conformations in denatured proteins. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8189–8194 (2010).
23. Wang, L. W., Liu, Y. N., Lyu, P. C., Jackson, S. E. & Hsu, S. T. Comparative analysis of the folding dynamics and kinetics of an engineered knotted protein and its variants derived from HP0242 of *Helicobacter pylori*. *Journal of physics. Condensed matter: an Institute of Physics journal* **27**, 354106 (2015).
24. Andrews, B. T., Capraro, D. T., Sulkowska, J. I., Onuchic, J. N. & Jennings, P. A. Hysteresis as a Marker for Complex, Overlapping Landscapes in Proteins. *The journal of physical chemistry letters* **4**, 180–188 (2013).
25. Capraro, D. T. & Jennings, P. A. Untangling the Influence of a Protein Knot on Folding. *Biophysical journal* **110**, 1044–1051 (2016).
26. Burbán, D. J., Haglund, E., Capraro, D. T. & Jennings, P. A. Heterogeneous side chain conformation highlights a network of interactions implicated in hysteresis of the knotted protein, minimal tied trefoil. *Journal of physics. Condensed matter: an Institute of Physics journal* **27**, 354108 (2015).
27. Wang, L., Chen, S. Y. & Hsu, S. T. Unraveling the folding mechanism of the smallest knotted protein, MJ0366. *The journal of physical chemistry. B* **119**, 4359–4370 (2015).
28. King, N. P., Jacobitz, A. W., Sawaya, M. R., Goldschmidt, L. & Yeates, T. O. Structure and folding of a designed knotted protein. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20732–20737 (2010).
29. Mallam, A. L. & Jackson, S. E. Knot formation in newlytranslated proteins is spontaneous and accelerated by chaperonins. *Nature chemical biology* **8**, (147–153 (2011)).
30. Andersson, F. I., Jackson, S. E. & Hsu, S. T. Backbone assignments of the 26 kDa neuron-specific ubiquitin carboxyl-terminal hydrolase L1 (UCH-L1). *Biomolecular NMR assignments* **4**, 41–43 (2010).
31. Andersson, F. I. *et al.* The effect of Parkinson's-disease-associated mutations on the deubiquitinating enzyme UCH-L1. *Journal of molecular biology* **407**, 261–272 (2011).
32. Zhang, H. & Jackson, S. E. Characterization of the Folding of a 52-Knotted Protein Using Engineered Single-Tryptophan Variants. *Biophysical journal* **111**, 2587–2599 (2016).
33. Ziegler, F. *et al.* Knotting and unknotting of a protein in single molecule experiments. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 7533–7538 (2016).
34. Lou, S. C. *et al.* The Knotted Protein UCH-L1 Exhibits Partially Unfolded Forms under Native Conditions that Share Common Structural Features with Its Kinetic Folding Intermediates. *Journal of molecular biology* **428**, 2507–2520 (2016).
35. Wang, L., Chen, S. Y. & Hsu, S. T. Folding analysis of the most complex Stevedore's protein knot. *Scientific reports* **6**, 31514 (2016).
36. Sulkowska, J. I., Sulkowski, P., Szymczak, P. & Cieplak, M. Stabilizing effect of knots on proteins. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19714–19719 (2008).
37. Skrbic, T., Micheletti, C. & Faccioli, P. The role of non-native interactions in the folding of knotted proteins. *PLoS computational biology* **8**, e1002504 (2012).
38. Sulkowska, J. I., Sulkowski, P. & Onuchic, J. Dodging the crisis of folding proteins with knots. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 3119–3124 (2009).
39. Wallin, S., Zeldovich, K. B. & Shakhnovich, E. I. The folding mechanics of a knotted protein. *Journal of molecular biology* **368**, 884–893 (2007).
40. Prentiss, M. C., Wales, D. J. & Wolynes, P. G. The energy landscape, folding pathways and the kinetics of a knotted protein. *PLoS computational biology* **6**, e1000835 (2010).
41. Chwastyk, M. & Cieplak, M. Cotranslational folding of deeply knotted proteins. *Journal of physics. Condensed matter: an Institute of Physics journal* **27**, 354105 (2015).
42. Covino, R., Skrbic, T., Beccara, S. A., Faccioli, P. & Micheletti, C. The role of non-native interactions in the folding of knotted proteins: insights from molecular dynamics simulations. *Biomolecules* **4**, 1–19 (2013).
43. Noel, J. K., Sulkowska, J. I. & Onuchic, J. N. Slipknotting upon native-like loop formation in a trefoil knot protein. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15403–15408 (2010).
44. Sulkowska, J. I. *et al.* Knotting pathways in proteins. *Biochemical Society transactions* **41**, 523–527 (2013).
45. Najafi, S. & Potestio, R. Folding of small knotted proteins: Insights from a mean field coarse-grained model. *The Journal of chemical physics* **143**, 243121 (2015).
46. Chwastyk, M. & Cieplak, M. Multiple folding pathways of proteins with shallow knots and co-translational folding. *The Journal of chemical physics* **143**, 045101 (2015).
47. Li, W., Terakawa, T., Wang, W. & Takada, S. Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and zouf-knot. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17789–17794 (2012).
48. Soler, M. A., Nunes, A. & Faisca, P. F. Effects of knot type in the folding of topologically complex lattice proteins. *The Journal of chemical physics* **141**, 025101 (2014).
49. Faisca, P. F., Travasso, R. D., Charters, T., Nunes, A. & Cieplak, M. The folding of knotted proteins: insights from lattice simulations. *Physical biology* **7**, 16009 (2010).
50. Bolinger, D. *et al.* A Stevedore's protein knot. *PLoS computational biology* **6**, e1000731 (2010).
51. Lim, K. *et al.* Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot. *Proteins* **51**, 56–67 (2003).
52. Das, C. *et al.* Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4675–4680 (2006).
53. Virnau, P., Mirny, L. A. & Kardar, M. Intricate knots in proteins: Function and evolution. *PLoS computational biology* **2**, e122 (2006).
54. Rawdon, E. J., Millett, K. C., Sulkowska, J. I. & Stasiak, A. Knot localization in proteins. *Biochemical Society transactions* **41**, 538–541 (2013).
55. Andersson, F. I., Pina, D. G., Mallam, A. L., Blaser, G. & Jackson, S. E. Untangling the folding mechanism of the 5(2)-knotted protein UCH-L3. *The FEBS journal* **276**, 2625–2635 (2009).
56. Sulkowska, J. I., Sulkowski, P., Szymczak, P., Cieplak, M. Tightening of knots in proteins. *Physical Review Letters* **100** (2008).
57. Sulkowska, J. I., Sulkowski, P., Szymczak, P. & Cieplak, M. Untying Knots in Proteins. *Journal of the American Chemical Society* **132**, 13954–13956 (2010).
58. Bornschlogl, T. *et al.* Tightening the knot in phytochrome by single-molecule atomic force microscopy. *Biophysical journal* **96**, 1508–1514 (2009).
59. He, C., Lamour, G., Xiao, A., Gsponer, J. & Li, H. Mechanically tightening a protein slipknot into a trefoil knot. *Journal of the American Chemical Society* **136**, 11946–11955 (2014).
60. Szymczak, P. Tight knots in proteins: can they block the mitochondrial pores? *Biochemical Society transactions* **41**, 620–624 (2013).
61. Szymczak, P. Periodic forces trigger knot untying during translocation of knotted proteins. *Scientific reports* **6**, 21702 (2016).
62. Huang, L. & Makarov, D. E. Translocation of a knotted polypeptide through a pore. *The Journal of chemical physics* **129**, 121107 (2008).
63. Wojciechowski, M., Gomez-Sicilia, A., Carrion-Vazquez, M. & Cieplak, M. Unfolding knots by proteasome-like systems: simulations of the behaviour of folded and neurotoxic proteins. *Molecular bioSystems* **12**, 2700–2712 (2016).
64. San Martín, A. *et al.* Knots can impair protein degradation by ATP-dependent proteases. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 9864–9869 (2017).

65. Sriramoju, M. K., Chen, Y., Lee, Y. C. & Hsu, S. D. Topologically knotted deubiquitinases exhibit unprecedented mechanostability to withstand the proteolysis by an AAA+ protease. *Sci Rep* **8**, 7076 (2018).
66. Keiler, K. C. & Sauer, R. T. Sequence determinants of C-terminal substrate recognition by the Tsp protease. *The Journal of biological chemistry* **271**, 2589–2593 (1996).
67. Karzai, A. W., Roche, E. D. & Sauer, R. T. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nature structural biology* **7**, 449–455 (2000).
68. Kenniston, J. A., Baker, T. A., Fernandez, J. M. & Sauer, R. T. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**, 511–520 (2003).
69. Mallam, A. L., Onuoha, S. C., Grossmann, J. G. & Jackson, S. E. Knotted fusion proteins reveal unexpected possibilities in protein folding. *Molecular cell* **30**, 642–648 (2008).
70. Lim, N. C. & Jackson, S. E. Mechanistic insights into the folding of knotted proteins *in vitro* and *in vivo*. *Journal of molecular biology* **427**, 248–258 (2015).
71. Lee, C., Schwartz, M. P., Prakash, S., Iwakura, M. & Matouschek, A. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Molecular cell* **7**, 627–637 (2001).
72. Kenniston, J. A., Baker, T. A. & Sauer, R. T. Partitioning between unfolding and release of native domains during ClpXP degradation determines substrate selectivity and partial processing. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1390–1395 (2005).
73. Burton, R. E., Siddiqui, S. M., Kim, Y. I., Baker, T. A. & Sauer, R. T. Effects of protein stability and structure on substrate processing by the ClpXP unfolding and degradation machine. *The EMBO journal* **20**, 3092–3100 (2001).
74. Kiel, C. *et al.* Recognizing and defining true Ras binding domains II: *in silico* prediction based on homology modelling and energy calculations. *Journal of molecular biology* **348**, 759–775 (2005).
75. Rousseau, F., Schymkowitz, J. W., Sanchez del Pino, M. & Itzhaki, L. S. Stability and folding of the cell cycle regulatory protein, p13(suc1). *Journal of molecular biology* **284**, 503–519 (1998).
76. Dzubilla, J. Tightening and Untying the Knot in Human Carbonic Anhydrase III. *The journal of physical chemistry letters* **4**, 1829–1833 (2013).
77. Dzubilla, J. Sequence-specific size, structure, and stability of tight protein knots. *Biophysical journal* **96**, 831–839 (2009).
78. Sauer, R. T. & Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annual review of biochemistry* **80**, 587–612 (2011).
79. Lee, C., Prakash, S. & Matouschek, A. Concurrent translocation of multiple polypeptide chains through the proteasomal degradation channel. *The Journal of biological chemistry* **277**, 34760–34765 (2002).
80. Wilson, M. D. *et al.* Proteasome-mediated processing of Def1, a critical step in the cellular response to transcription stress. *Cell* **154**, 983–995 (2013).
81. Piwko, W. & Jentsch, S. Proteasome-mediated protein processing by bidirectional degradation initiated from an internal site. *Nature structural & molecular biology* **13**, 691–697 (2006).
82. Tian, L., Holmgren, R. A. & Matouschek, A. A conserved processing mechanism regulates the activity of transcription factors Cubitus interruptus and NF-kappaB. *Nature structural & molecular biology* **12**, 1045–1053 (2005).
83. Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T. & Baker, T. A. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Molecular cell* **5**, 639–648 (2000).
84. Larsen, C. N., Price, J. S. & Wilkinson, K. D. Substrate binding and catalysis by ubiquitin C-terminal hydrolases: identification of two active site residues. *Biochemistry* **35**, 6735–6744 (1996).

## Acknowledgements

E.M.S. was supported by a PhD studentship from the Medical Research Council of the UK (MRC), and grants from Fund for Women Graduates (FfWG), the Lundgren Research Fund and the Cambridge Philosophical Society. L.S.I. was supported by a Senior Fellowship from the Medical Research Foundation. We would also like to thank Dr. Nicole Lim and Dr. Anna Mallam for the original expression vectors containing the different knotted fusion proteins.

## Author Contributions

E.M.S., L.S.I. and S.E.J. conceived the experiments. E.M.S. conducted all the experimental work and undertook data analysis. All authors wrote and reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-38173-3>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019