Characterization of the folding of a 5_2-knotted protein using engineered single-tryptophan variants

H. Zhang & S.E. Jackson

Running title: Folding of a 5_2-knotted protein

Abstract: Proteins containing topological knots have been steadily increasing in number over the past two decades. Despite this, their folding mechanisms are not well understood. UCH-L1 has a 5_2-knotted topology. Here, we constructed a series of mutants that contain a single tryptophan at different locations along the polypeptide chain. A study of the thermodynamic properties of the mutants shows that the structure of UCH-L1 is remarkably tolerant to incorporation of bulky tryptophan side chains.

Comprehensive kinetic studies of the mutants reveal that they fold *via* parallel pathways on which there are two intermediate states very similar to wild-type UCH-L1. The structure of the intermediate states do not change significantly with mutation therefore occupy local minima on the energy landscape that have relatively narrow basins. The kinetic studies also establish that there are considerably more tertiary interactions in the intermediate states than results from previous NMR studies suggested. The two intermediates differ from each other in the extent to which tertiary interactions between the highly stable central β-sheet and flanking α-helices and loop regions are formed. There is strong evidence that these states are aggregation prone.

Both the transition states from I_1 and I_2 to the native state are plastic and change with mutation and denaturant concentration. Previous studies indicate that the threading event that creates the 5_2 knot occurs during these steps, suggesting that there is a broad energy barrier consistent with the chain undergoing some searching of conformational space as the threading takes place.

Introduction

Proteins containing topological knots have become of great interest to the protein folding field as they challenge the conventional view of how proteins fold (1). In classic views of folding, the native conformation of proteins should be not only thermodynamically stable, but also kinetically accessible to accomplish fast and efficient folding on biologically relevant timescales (2). As a consequence, self-threading of the polypeptide backbone to form a knot was thought to be highly unlikely as it would be less kinetically accessible than unknotted protein structures (3).

Since the first topological knots were unambiguously discovered in protein structures in 2000 (4), an increasing number of proteins containing topological knots have been identified (5). Previous investigations on the folding mechanisms of a few knotted proteins have established that they have complex multi-step folding pathways with slow kinetics rather than the simple two-state mechanisms that are commonly seen for small fast-folding single-domain proteins (6-11). Despite experimental and computational work over the last ten years (12-17), the folding mechanisms of knotted proteins remains poorly understood compared with those for small model systems that have been studied in great detail (18-23).

The human ubiquitin C-terminal hydrolase homolog L1 (UCH-L1) is a cysteine protease that is highly expressed in neurons making up some 1-2% of total soluble brain proteins (24). The UCH family are deubiquitinating enzymes that play an important role in maintaining the cellular pool of monomeric ubiquitin (25). The crystal structure of UCH-L1 shows it is composed of a sub-domain that contains five α-helices and another sub-domain comprised of two α-helices and six β-strands, Fig. 1 (26). The β-strands form an anti-parallel β-sheet that forms part of the hydrophobic core of the protein and which is flanked by the α-helices. The structure of UCH-L1 contains a 5_2-knot that is formed by the threading of the polypeptide backbone through a loop formed by the rest of the chain, making it one of the most complicated knotted topologies identified in proteins. However, the knot of UCH-L1 is shallow such that the removal of merely a few residues from the termini removes its 5_2-knotted topology (5).

Earlier experimental investigations of the folding of UCH-L3, a homolog of UCH-L1, have shown that it folds *via* parallel pathways on each of which there is a hyper-fluorescent intermediate state (27). Folding of wild-type UCH-L1 has been investigated using a range of different techniques. In this case, the fluorescence of the single tryptophan, Trp26, was used to probe changes in tertiary structure in both thermodynamic and kinetic studies (28,29). These studies have shown that UCH-L1 unfolds reversibly under equilibrium conditions using urea as a denaturant and at least one stable intermediate state (I) is populated between the native (N) and denatured states (D) (28). More recently, kinetic studies have established that wild-type UCH-L1 folds along parallel pathways, on each of which there is a detectable intermediate state (I_1 and I_2) (29), the same as UCH-L3. However, for wild-type UCH-L1 the change in fluorescence between the denatured and intermediate states is very small, making it difficult to obtain high quality data for the transitions between I_1, I_2 and D.

In order to characterize the folding mechanism of UCH-L1 in further detail, six single tryptophan mutants were constructed: L3W, L70W, F108W, V113W, F160W and F181W. In all these mutants, Trp26 present in wild type was mutated to phenylalanine (W26F). These positions are located throughout the sequence of the protein such that they probe different regions of the structure and, ideally, result in a much larger change in fluorescence for the transition between the denatured and intermediate states than observed for wild type.

Here, detailed thermodynamic and kinetic studies on the six single-Trp mutants of UCH-L1 are presented. In each case, intrinsic fluorescence, far-UV circular dichroism (far-UV CD) and enzymatic activity assays were employed to assess the effect of the Trp substitution on the structure, stability and function of the protein. All the mutants had structure and activities similar to wild type and these were therefore characterized further in detailed kinetic investigations.

Materials and Methods

Materials

Restriction enzymes, PCR reagents and DNA purification kits were purchased from Fermentas, primers from Sigma-Aldrich, pre-cast SDS-PAGE gels were purchased from Invitrogen, protein purification columns were purchased from GE Healthcare. Other materials, including buffer and urea, were of Analytical grade and purchased from Fisher Chemical or Sigma-Aldrich. Millipore-filtered, double-distilled water was used throughout.

Construction of mutants

The sites of mutations were selected based on the solvent accessibility of side chains calculated based on the crystal structure of wild-type UCH-L1 (PDB code: 2ETL) using web-based program GetArea (30), Table S1. Trp26, the single

tryptophan residue in wild type, was replaced by a phenylalanine and residues Leu3, Leu70, Phe108, Val113, Phe160 and Phe181 were individually substituted to tryptophan using site-directed mutagenesis employing standard PCR methods to create six, single-Trp UCH-L1 mutants. For all these mutants, DNA was sequenced by SourceBioscience, Cambridge.

Protein expression and purification

Wild-type and mutant UCH-L1 were purified as described elsewhere (28). The concentration of the purified protein was determined by spectrophotometry using an extinction coefficient of 28086 M^{-1} cm^{-1} at 280 nm. The purity of each protein preparation was assessed by SDS-PAGE and judged to be over 95%. Purified proteins were flash-frozen in 50 mM Tris-HCl, 5 mM DTT, 0.5 mM EDTA, pH 7.6 and stored at -80 °C.

Kinetic and thermodynamic measurements on folding and unfolding

The conformational stabilities of wild type and mutants were measured using urea unfolding experiments conducted under equilibrium conditions where it has been established that the unfolding of wild-type UCH-L1 is fully reversible (28). Unfolding was monitored by tryptophan fluorescence and far-UV CD spectroscopy to probe changes in the tertiary and secondary model, respectively. The data were fit to either a two- or three-state unfolding model dependent upon the probe and mutant, see Supporting Materials for further details. The refolding/unfolding kinetic measurements were carried out as previously described (29). Detailed methods can be found in Supporting Materials.

Enzymatic activity assay

The enzymatic activity of wild-type and mutant UCH-L1 were measured by the hydrolysis of the fluorogenic substrate: ubiquitin-AMC (7-amino-4-methylcoumarin). Stock solutions of enzyme and substrate were manually mixed and diluted such that the final concentrations of ubiquitin-AMC and UCH-L1 were 500 and 25 nM, respectively. The reaction was monitored in a fluorimeter using an excitation wavelength of 380 nm and emission wavelength of 460 nm. Kinetic dat, obtained under steady-state conditions were fitted to a single-exponential equation to obtain a reaction rate constant.

Analytical size-exclusion chromatography

Wild type and mutants were first diluted in 2-3 M urea for 1, 10 and 60 min and then loaded on to a Superdex 200 10/300 analytical size-exclusion column (GE Healthcare) equilibrated with 50 mM Tris and 5 mM DTT, pH 7.6. The absorbance at 280 nm was used to monitor the elution of protein.

Results

The effect of Trp incorporation on the structure and activity of UCH-L1

The mutants (L3W, L70W, F108W, V113W, F160W and F181W) all show comparable catalytic activity to wild-type UCH-L1 for the model substrate ubiquitin-AMC and their far-UV CD spectra are very similar to that of wild type (Fig. S1, S2 & Table S2). Thus, for these mutants, there is no evidence that the mutations have significantly disrupted either the secondary or tertiary structure of the 5_2-knotted native state.

Equilibrium unfolding studies on wild-type and mutant UCH-L1

For each variant, the fluorescence and far-UV CD data, Figures S3 and S4, were fit to either a two- or three-state unfolding model to obtain the thermodynamic parameters associated with the transitions between native, intermediate and denatured states, Table S3. The values for the midpoints of the transitions between native and intermediate states, $[urea]_{I\text{-}N}^{50\%}$, and intermediate and denatured states, $[urea]_{D\text{-}I}^{50\%}$, for each variant are shown in Table 1. These parameters were then used to calculate the degree to which the native and intermediate states are destabilised on incorporation of a bulky Trp side chain, Table 1.

L3W and F160W have comparable stabilities to wild type for both the N to I and I to D transitions, indicating that incorporation of Trp at these positions does not affect protein stability. In contrast, the native states of L70W and V113W are destabilised by some 2.8 and 2.2 kcal mol^{-1} relative to the intermediate states, respectively. In addition, the intermediate states of these mutants are also destabilised relative to the denatured state, by some 1.5 and 2.1 kcal mol^{-1}, respectively. The native states of F108W and F181W are also destabilized by some 2.4 and 1.0 kcal mol^{-1} relative to the intermediate states compared to wild type. In these two cases, the impact of incorporating a Trp at these positions cannot be assessed because the I to D transition can't be probed by fluorescence or far-UV CD.

Φ_I-values provide structural information on the intermediate state

Φ_I-values were calculated using the thermodynamic parameters obtained from the equilibrium unfolding fluorescence data, Tables 1 and S3. The Φ_I value is a measure of the degree of structure at the site of the mutation in the intermediate state relative to the native state (31-36). Although all the Φ_I-values have considerable errors, Table 1, qualitatively the results suggest that local structure in the vicinity of Trp3, Trp70 and Trp113 is partially formed in the intermediate state.

Unfolding and refolding kinetics of wild-type and mutant UCH-L1

Unfolding and refolding kinetics of wild type and the single Trp mutants were probed by single- and double-[urea] jump experiments using a stopped-flow instrument. The intrinsic tryptophan fluorescence was used to probe changes in tertiary structure.

Both single [urea]-jump unfolding and refolding kinetics, and interrupted refolding kinetics have recently been reported for wild-type UCH-L1 (29). The results from these experiments have been used to propose a kinetic scheme for folding, Fig. 1C.

Briefly, single- and double-jump kinetic experiments on wild-type UCH-L1 exhibited two slow unfolding and refolding kinetic phases attributed to the transitions N-I_1 and N-I_2. There were also two fast unfolding and refolding phases which correspond to the transitions between I_1-D and I_2-D (29). We use the kinetic scheme established for wild-type UCH-L1 as a reference point from which to interpret the kinetic data obtained for the mutants presented here.

Results from single [urea]-jump unfolding and refolding experiments

Typical unfolding and refolding kinetic traces from single [urea]-jump experiments are shown in Fig. S6. In the single [urea]-jump unfolding kinetics, the mutants F108W and F181W both show two slow unfolding phases the same as wild-type UCH-L1. However, other mutants show additional faster unfolding phases such that there are three phases in total (Fig. 2). We postulate that these additional phases represent unfolding from the intermediate to denatured states ($I_1 \rightarrow D$ or $I_2 \rightarrow D$) that are not directly observable for wild type because of the small fluorescence change. For L3W, the second slow phase ($N \rightarrow I_2$) is not observed in unfolding experiments, most likely because its unfolding rate

becomes very similar to that of the $N \rightarrow I_1$ transition and therefore these two phases cannot be de-convoluted for this mutant.

Single [urea]-jump refolding kinetics for L3W, L70W, V113W and F160W all showed four phases similar to wild type (29), however, only three refolding phases were observed for F108W and F181W (Fig. 2). This is likely due to the fact that either (i) for the folding of one of these states there is insufficient change in fluorescence for it to be observed, or (ii) two of the refolding phases have similar rate constants. It is unlikely that these two mutants have a different reaction mechanism given the data.

The unfolding and refolding rate constants for all the mutants were measured at different denaturant concentrations and the natural logarithm of the rate constants plotted as a function of denaturant concentration as shown in Fig. 2.

Analysis of the kinetic data of the slow phases corresponding to transitions between native and intermediate states

For wild type and most of the mutants (apart from V113W), the two slow refolding and unfolding phases join smoothly to form characteristic chevron plots, Fig. 2. For wild-type UCH-L1 these phases have been shown to correspond to transitions between the native state and the two intermediate states (29). We assume, and all the data are consistent with, that this is also the case for all the mutants. The unfolding and refolding rate constants for these phases were fitted to a simple two-state kinetic model described by either Equation S8 for wild type or Equation S9 (which takes into account Hammond effects and the curvature observed in the unfolding arm of the chevron plots observed at high urea concentrations) for L3W ($N\text{-}I_1$ only), L70W, F108W, F160W and F181W. This analysis provides values for the unfolding and refolding rate constants in water, $k_U{}^{H_2O}$ and $k_F{}^{H_2O}$, respectively, as well as m_{kU} and m_{kF}, the slopes of the unfolding and refolding arms of the Chevron plots, respectively. When Equation 9 was employed an additional parameter, m_{kU^*} was also obtained which relates to the curvature in the unfolding data observed at high urea concentrations. The kinetic data for V113W could not be fit to either equation, and for L3W, the refolding rate constants corresponding to $I_2 \rightarrow N$ were fit to a linear equation to describe just the refolding arm to obtain values for the refolding rate in water, and the slope as no data for $N \rightarrow I_2$ had been obtained. All the kinetic parameters obtained from the fits are given in Table 2.

The kinetic parameters for the N-I_1 and N-I_2 transitions, Table 2, were then used to calculate the thermodynamic parameters associated with these transitions, also shown in Table 2. This included a value for $[urea]_{50\%}$, the midpoint of each transition that reflects the relative stabilities of the states involved, as well as *m* values, see below for discusssion. For each mutant the $[urea]_{50\%}$ values obtained from the kinetic data are comparable with those obtained from equilibrium unfolding experiments, Table 1. The difference in the stabilities of the intermediate states I_1 and I_2 relative to the native state, between wild type and mutants are calculated and listed in Table S5.

The structures of the intermediate states are not significantly affected by mutation

The *m* value obtained from the kinetic data, m_{kin}, is a measure of the change in solvent accessible surface area between the two states measured (in this case, N-I_1 or N-I_2) (37). Although the values vary a little for the different mutants studied, the m_{kin} values for the N-I_1 transition are all (with the exception of F108W which is still very similar) within error of an average value of 1.9 kcal mol^{-1} M^{-1}. Likewise, the m_{kin} values for the N-I_2 transition are all within error of an average value of 1.7 kcal mol^{-1} M^{-1}. These results suggest that none of the mutations studied significantly affects the structure of the two intermediate states observed in the kinetic experiments.

m_{ku} and m_{kf} values show evidence for plasticity of the transition states between the native and intermediate states

The values of m_{ku} and m_{kf} report on the slope of the refolding and unfolding arms of the chevron plots, respectively, Table 2. These values are a measure of the change in solvent accessible surface area between the ground and transition states (i.e. the native and transition state for m_{ku} and the intermediate and transition state for m_{kf}). There is considerable variability in the values of both parameters for wild type and mutants, which in several cases is greater than the errors, and therefore significant. Since analysis of m_{kin} values suggests that none of the mutations significantly affects the structure of either of the intermediate states therefore the differences in m_{kf} and m_{ku} observed can be attributed to changes in the structure of the transition states. These results are in agreement with the curvature observed in the unfolding arm of the chevron plots for some of the mutants and which can also be attributed to Hammond effects and changes in the transition state, such that the transition state becomes more native like as the difference in energy between the native and transition state decreases (38-40). The Tanford β value that reflects the compactness of the transition state for each mutant was calculated and the values are listed in Table S7.

Interrupted-refolding experiments confirm the additional fast unfolding phases and refolding phases correspond to the transitions between intermediate and denatured states.

Additional kinetic experiments, called interrupted-refolding experiments, were undertaken that used two consecutive [urea]-jumps to investigate the unfolding of intermediate states populated during refolding, that were not necessarily directly observed in single-jump experiments (Fig. 2). In this experiment, unfolded protein in high [urea] was rapidly diluted and allowed to refold for some time, the aging time, before another rapid [urea]-jump to high concentrations of urea was used to unfold any species populated during the aging time. At short aging times, most of the mutants (apart from F108W and V113W) have one or two additional fast unfolding phases that were not observed in the single-jump experiments, as is the case for wild type. The additional phases observed correspond to transitions of $I_1 \rightarrow D$ and/or $I_2 \rightarrow D$. At longer aging times, wild type and mutants all showed slow unfolding phases that have similar rate constants to those measured in single-jump unfolding experiments that correspond to $N \rightarrow I_1$ and/or $N \rightarrow I_2$, as expected (Fig. 2).

Using a combination of single and double [urea]-jump experiments it was therefore possible to obtain kinetic information on all of the four phases expected from the kinetic scheme that has been proposed for wild type, Fig. 1C, for most of the mutants. Fig. 2 shows all the kinetic rate constants obtained for the wild type and the mutants, and their dependence on urea concentration.

In addition to the rate constants for the unfolding of I_1 and I_2 to the denatured state, Fig. 2, the interrupted–refolding experiments show how the amplitude of each unfolding phase varies with aging time, Fig. 3. Thus, it shows the temporal evolution of each species along the reaction pathway. By fitting these data to exponential functions, one can, in some cases, obtain the folding rate constants of the different species and these can be compared directly to the results from single-jump refolding experiments, Fig. 2. Thus, in many cases, it is possible to directly link an unfolding phase with its corresponding refolding phase

Interrupted-refolding experiments suggest that transiently populated kinetic intermediates are prone to oligomerization

Of particular interest was the fact that, for some of the mutants, not all of the slow unfolding phases seen in the single [urea]-jump experiments were also observed in the interrupted-refolding experiments For example, for L70W, only aging times less than 10 s could be used as longer times resulted in a dramatic decrease in signal, Fig. 3. This behavior was also observed for other mutants.

For L3W, L70W and V113W, the amplitude corresponding to the formation of the native state from I_1 and I_2 does not behave as expected (e.g. it is not a simple exponential increase with a plateau.). For V113W, the amplitude of the two unfolding phases ($N \rightarrow I_1$ and $N \rightarrow I_2$) increases at short aging times but this is followed by a decrease in signal at longer aging times, Fig. 3. The same phenomenon is observed for L3W and L70W for the single $N \rightarrow I$ transition observed. We postulate that this may be as a result of the fact that, under the conditions used, the intermediate states can aggregate over longer aging times leading to a reduction in the concentration of native protein and therefore a decrease in signal for the $N \rightarrow I_1$ and $N \rightarrow I_2$ unfolding transitions with aging time. Analytical size-exclusion chromatography was used to investigate this phenomenon.

Analytical size-exclusion chromatography of UCH-L1 variants in urea shows their propensity to oligomerize

Size exclusion chromatography was used to assess the size of species of UCH-L1 in solution at different urea concentrations under equilibrium conditions. From equilibrium unfolding experiments, we know the intermediate states are populated between 2-3 M urea. Wild-type and mutant UCH-L1 were therefore incubated in 2-3 M urea for different incubation times and their oligomeric states were analyzed using size-exclusion chromatography, Fig. 4. The results show that wild type and mutants invariably form a species with larger molecular weight, i.e., an oligomer, the concentration of which increases with longer incubation times. It is noted that in general there is a correlation between the propensity of oligomerization and the degree to which the mutation destabilizes the native state. For example, F108W and V113W both destabilize the native state of UCH-L1 and both are highly prone to oligomerization (Fig. 4 and Table S6).

Discussion

Previous studies have established that the 5_2-knot protein UCH-L1 has a complex folding mechanism that involves parallel pathways and multiple intermediate states, suggesting that it has a relatively rough folding energy landscape compared to small proteins with simple folds (29). In this study, we investigated the folding mechanism of UCH-L1 in further detail by comparing the kinetic and thermodynamic behaviors of a series of single tryptophan variants to that of wild type.

The structure and function of UCH-L1 is tolerant to the incorporation of tryptophan in many partially buried sites

The effect of the mutations on the structure of UCH-L1 was assessed using enzymatic activity assays and far-UV CD spectroscopy. In all cases, the structure and activity of the mutants are very similar to that of wild type. This indicates that the partially exposed positions selected for the incorporation of tryptophan are not involved in either substrate binding nor the subsequent conformational change required for activity for this enzyme (26, 41).

Tryptophan fluorescence as a probe of structure in the native and intermediate states

Tryptophan fluorescence is generally used as a probe of tertiary structure that reports on its local environment. For the UCH-L1 variants studied here, we used the tryptophan fluorescence measured in different concentrations of denaturant to report on local structure in the native, intermediate and denatured states. In their native states, wild-type and mutant UCH-L1 have emission maxima (λ_{max}) of 310 and 330-340 nm, respectively, Fig. S1. The emission maximum is related to the hydrophobic/polar environment around the tryptophan side chain. The results are consistent with Trp26 being buried in the native state of wild type, and its fluorescence being quenched by a spatially proximal group. For all the other engineered tryptophan residues the results establish that these are all partially exposed in the native state as expected.

In the intermediate state of wild-type UCH-L1, Trp26 becomes partially exposed and its fluorescence is less quenched, leading to a red-shift in the spectrum and increase in fluorescence, Fig. S1, as reported previously (29). Quenching also provides information on structure in the intermediate states of some of the mutants. The fluorescence of L70W and V113W in moderate concentrations of denaturant are reduced with respect to the denatured state, Fig. S1, suggesting that, at these positions, the tryptophan residues are quenched to some degree in the intermediate state. Therefore, there must be some structure at these sites in

this state.

In high concentrations of denaturant, the emission maxima of wild type and mutants are all between 350-355 nm, Fig. S1, showing that the tryptophan side chains are fully exposed to solvent in the denatured state and there is no indication of significant residual structure at any of the sites studied.

Urea unfolding experiments performed under equilibrium conditions show wild type and mutants follow a similar unfolding pathway and provide further information on structure of the intermediate states

Wild-type UCH-L1 has been shown to populate an intermediate state in unfolding experiments performed under equilibrium conditions and follows a three-state model (29). The fluorescence unfolding data for most of the mutants obtained under the same conditions can be fitted to a three-state model indicating there is a measurable fluorescence change between native and intermediate, and intermediate and denatured states, in these cases, Fig. S1, Tables 1 and S3. The change in fluorescence between the intermediate and denatured states for tryptophan residues at positions 3, 70, 113 and 160 is larger than for Trp 26 in wild type, Table S3. This suggests that the tryptophan side chains in these positions may be more buried in the intermediate states than Trp26 in wild type. Hence, we propose that there is some structure at positions 3, 70, 113 and 160 in the intermediate state.. In contrast, the fluorescence unfolding data for the mutants with tryptophan residues at position 108 and 181 can only be fit to a two-state model as there is little change in fluorescence between intermediate and denatured states. Hence, residues 108 and 181 are likely to be exposed to solvent in the intermediate states.

The *m* values calculated from the unfolding data for the transition between native and intermediate states (or intermediate and denatured states) are a measure of the relative degree of structure (in terms of solvent accessible surface area) in the intermediate and native states (or intermediate and denatured states). The *m* values calculated for the N-I transition using either fluorescence or far-UV CD unfolding data, Table S3, show that all the mutants have comparable changes in solvent exposed surface area between native and intermediate states with respect to wild type. This indicates that the structure of the intermediate states does not change significantly on mutation.

Φ_I values as a measure of structure in the intermediate state

Φ_I values provide useful insight into the structure of an intermediate in the vicinity of a mutated residue (42). L3W, L70W and V113W all have fractional Φ_I

values suggesting that some structure is formed around these residues in the intermediate state.

Structure in the intermediate states from equilibrium unfolding experiments: comparison with earlier NMR studies.

There is strong evidence from the fluorescence spectroscopy and from the thermodynamic stability studies that regions around residues 3, 70 and 113 (located in an N-terminal loop, the C-terminus of α-helix 2 and the N-terminus of α-helix 4, respectively) are, at least, partially structured. The evidence that structure around residue 160 (at the N-terminus of β-strand 3) is formed in the intermediates is much weaker, and mainly comes from the larger change in fluorescence intensity between intermediate and denatured states relative to wild type. In contrast, there is strong evidence that the regions around residues 108 and 181, which are both in loops, are not structured in the intermediate states.

How do these results compare with previously published HDX-NMR (hydrogen-deuterium exchange) investigations on wild-type UCH-L1? (28,29) In general, the NMR results are consistent with an intermediate state in which the central β-sheet is largely formed and that is stabilized by the packing of some regions of the flanking α-helices, however, much of the α-helical structure is lost on unfolding from the native to the intermediate state. In particular, the results demonstrated that there are some tertiary interactions between Leu3, α-helices 3 and 7 and β-strands 2-4 and 6 in the intermediate state (28). Some of these results are consistent with our results from the Trp mutants presented here, particularly with respect to position 3, 108 and 181, and to a lesser extent 160. However, for the other residues there are some discrepancies, including residues 70 and 113. There is no evidence from HDX-NMR measurements that Leu70 or Val113 are involved in any significant structure in the intermediate state, (28, 29), however, here we have provided strong evidence that these two residues are in regions that are partially structured in the intermediate. One possible reason for the apparent discrepancy between the results may be due to the fact that the HDX-NMR experiments report on hydrogen-bonding of the backbone amide group, and therefore secondary structure, whilst the fluorescence spectroscopy and Φ_I values report largely on tertiary structure. Our results suggest that there are considerably more tertiary interactions in the intermediate than suggested from the HDX studies (28, 29). In this case, it is likely that only partial tertiary structure forms, or that the tertiary structure is considerably more dynamic than in the native state. This would give rise to the fractional Φ_I-values observed, but also little protection for the backbone amide groups as seen in the HDX NMR studies.

Analysis of the *m*-values from the thermodynamic stability studies indicates that the structure of the intermediate(s) formed under equilibrium conditions does not change significantly on mutation. Thus, UCH-L1 does not form other structures which are metastable.

The mutants follow the same parallel model for unfolding/folding as wild type with two intermediate states

For all the mutants, the unfolding and refolding kinetic data are consistent with parallel folding routes on which two discrete intermediate states are populated. Therefore, the kinetic data of all the mutants was analyzed as previously undertaken for wild type (29).

The intermediate states of wild-type and mutant UCH-L1 all have similar structures.

The *m*-value obtained from kinetic data, m_{kin}, for both the N-I_1 and N-I_2 transitions for the mutants are similar to those of wild type, indicating that the mutants have comparable changes in solvent exposed surface area between N and I_1/I_2 with respect to wild type. This suggests that the structure of both of the intermediate states populated does not change significantly on mutation and confirms the conclusions from the unfolding studies conducted under equilibrium conditions. Hence, all the mutants have the same general folding mechanism as wild type.

How do the structures of I_1 and I_2 differ?

The kinetic data also provides us with specific information on the nature of I_1 and I_2. In this case, the difference in the change in free energy between N-I_1 or N-I_2 that is obtained from the two-state analysis of the kinetic data corresponding to the two slow unfolding/refolding phases of the mutants, (Table 2) can be compared with the values obtained for wild type (Table 2). The values for $\Delta\Delta G_{kin}^{H_2O}$ for the N-I_1 and N-I_2 transitions calculated are shown in Table S5. Values close to zero indicate that the mutation has the same effect on the intermediate state as the native state as, in these cases, $\Delta G_{\text{N-I}}$ (WT) = $\Delta G_{\text{N-I}}$ (mut). The results suggest that, at position 70, the structure of the intermediate I_2 is similar to the native state whereas in I_1 it is less structured than in the native state, Table S5. In contrast, F160W shows similar changes in free energy to wild type for both the N-I_1 and N-I_2 transitions, indicating that position 160 is likely structured in both intermediate states I_1 and I_2. Interestingly, for L3W, $\Delta\Delta G_{\text{N-I}}$ for I_1 is negative, Table S5, indicating that the mutation stabilizes the intermediate state relative to the native state, again suggesting that there is some structure at this position in I_1 (kinetic data for I_2 was not available). These results are consistent with the information obtained from the thermodynamic studies, but provide additional important structural information on the difference in structure in the two kinetic

intermediates observed. A schematic diagram illustrating the structure of the two kinetic intermediates is shown in Fig. 5B.

A lag phase is observed for F108W confirming the sequential formation of native structure from an intermediate state

With the reaction scheme proposed for wild-type and mutant UCH-L1, a lag phase should be observed for the development of the two refolding phases corresponding to $I_1 \rightarrow N$ and $I_2 \rightarrow N$, corresponding to the formation of the intermediate states I_1 and I_2 from the denatured state. For wild-type and most of the mutants, a lag phase is not observed. Unfortunately, the temporal resolution that can be achieved with the stopped-flow apparatus set up in a double-jump mode is, in these cases, too low compared to the rates of folding from denatured to intermediate states that are generally fast. However, the slowest refolding phase ($I_1 \rightarrow N$) for F108W does show evidence of a lag phase, consistent with the mechanism proposed, Fig. 3. Together, the kinetic data indicate the folding mechanism of UCH-L1 is very robust such that it does not change even on introduction of residues with bulky side chain at numerous sites.

The major folding transition state between the native and intermediate states shows considerable plasticity

For the transitions between the intermediate and native states, many of the mutants show m_{ku} and m_{kf} values that deviate significantly from the values obtained for wild type and from an average value. This suggests that there is significant plasticity in the transition states for both the I_1-N and I_2-N transitions, which is also reflected by the $m_{k_u 2}$ value, Table 2. At high concentrations of urea, curvature is observed in the unfolding arms of the chevron plots for all the mutants indicative of the movement of the transition state along the reaction coordinate as a function of denaturant concentration (38, 40). In these cases, it reflects the degree to which the transition state becomes more native like as the energy between the native state and the transition states leading to the two intermediate states decreases with increasing urea concentration (38). This is also shown by the fact that there is a linear relationship between $\Delta\Delta G_{\text{N-TS1}}$, the difference in the change in free energy between the native state and the transition state to I_1 and the β-Tanford value, Fig. S7, Table S6. For the transition state leading to I_2, this is not observed likely because the range of β-Tanford values is too small in this case to observe a correlation. Therefore, in contrast to the intermediate states that do not differ greatly in structure, the transition states must be represented by quite broad energy barriers such that the position of the highest energy point varies with both denaturant concentration and mutation (40).

The intermediate states are prone to aggregate

In contrast to wild-type UCH-L1, for three mutants, L3W, L70W and V113W, the amplitudes of the two slower phases (formation of the native state from I_1 and I_2) in the double-jump experiments fails to accumulate and plateau at any significant level at long aging times, Fig. 3. These results suggest that, over long aging times, the intermediates do not proceed to fold to the native state. We propose that this is a result of the intermediates aggregating to form misfolded species under the conditions used in the double-jump experiment. Consistent with this hypothesis, the results of size-exclusion chromatography at moderate urea concentrations show that the intermediate states of all the UCH-L1 variants studied here have a strong propensity to oligomerize, Fig. 4. Qualitatively, the propensity of oligomerization is associated with the degree to which the mutations destabilize the native state of UCH-L1 as the effect is most pronounced for L70W and V113W. Based on these data, it is highly likely that the on-pathway intermediate states of UCH-L1 are able to oligomerize to stable species that cannot refold further to the native state on the timescales used in the experiment, Fig. 5. It is postulated that the formation of the oligomeric species requires some prolonged incubation time and high protein concentrations as employed in the double-jump experiments. For V113W, these results and scheme are consistent with the positive slopes seen in the refolding arm of the chevron plot for all three observable phases, Fig. 3, which suggest that even in single [urea]-jump measurements this variant can misfold to a state from which it must partially unfold in order to achieve productive folding to the native state.

Threading of the polypeptide chain to form the 5_2-knotted structure

The 5_2-knot in UCH-L1 is located between residues 5-219, thus much of the 223-residue polypeptide chain is involved in knot formation. The knot is relatively shallow compared to some other deeply knotted protein structures, only five to six residues need to pass through a loop to make the knot (43). Within the 5_2 knot there is a 3_1 knot that resides between residues 6-163 (43).

Although we cannot state categorically whether the intermediate states observed are knotted or not, it is possible to use the structural information we have to speculate on the knotted state of the intermediates. Consider the structure at the N-terminus of the chain: Leu3, located near the crossing of the chains critical for knot formation, we have shown forms some tertiary interactions in I_1 but not I_2. However, the results presented here and elsewhere (29) show that the loop through which the N-terminus passes to form the knot, that is comprised of α-helices 4, 5 and 6, is completely unstructured in both I_1 and I_2. Therefore, we propose that is unlikely that a knot has formed in either I_1 or I_2. This is consistent with the fact that the folding of I_1 or I_2 to the native state is slow compared to the formation of I_1 or I_2 from the denatured ensemble.

As all the knotted proteins so far discovered have twist knots and require only a single threading event to knot or unknot (44), it is likely that the 5_2-knot in UCH-L1 forms by a single threading (or other) event. If this is the case, then the formation of 3_1 then a 5_2 knot is unfavorable as it requires two threading events. Thus, we propose that the 5_2 knot forms by either threading of the N-terminus of the chain through a loop within a structure that has a twist in it, or that the knot may form by a loop flipping over the N-terminal region of the chain. This loop flipping mechanism has recently been observed in computational studies (45). One might therefore expect the kinetics of L3W, which contains a bulky Trp in the part of the chain that might undergo threading, to show slower kinetics compared to wild type. In fact, the rate constants for folding of L3W are some 3-4 times slower than wild type, however, it should be noted that other variants also show such a decrease in rate. Of more significance, perhaps, is the fact that L3W shows considerable differences in its m_{kf} and m_{ku} values compared to wild type, and much more so than any other variant. These results indicate that the position of the transition state has moved, and become less native-like, possibly due to steric hindrance associated with the bulky Trp side chain.

Conclusion

Folding of topologically knotted proteins is often challenging to study because of their complex multi-phasic kinetic behavior. In this paper, we report a detailed thermodynamic and kinetic characterization of a series of single-Trp mutants of the 5_2-knotted protein UCH-L1. The thermodynamic studies, using urea to unfold the protein under equilibrium conditions, reveal that there is partial structure around residues 3, 70 and 113 in the intermediate states, less structure at residue 160 and no structure at positions 108 or 181. These results suggest that there are considerably more tertiary interactions in the intermediate states, than results from HDX NMR studies have previously indicated (28, 29). This establishes the importance of using multiple probes of structure in characterizing folding intermediates of larger proteins with complex structures. The thermodynamic studies also establish that the structure of the intermediates formed do not change significantly with mutation, thus, the local minima on the energy landscape corresponding to the intermediates has a relatively narrow basin.

The results of our numerous kinetic studies confirm the fact that the intermediate states of the mutants populated during folding share significant structural similarity as those observed for wild type. Comparison of the kinetic data obtained for the mutants with that of wild-type UCH-L1, shows that there is structure around residues 160 in both I_1 and I_2, structure around residue 3 in I_1, and around residue 70 in I_2 but not I_1. Thus, the two kinetic intermediates

appear to differ in the extent to which tertiary interactions between the highly stable central β-sheet core and flanking α-helices and loop regions are formed, Fig. 5B. Additionally, there is strong evidence from the kinetic interrupted-refolding data that these intermediate states and are all aggregation-prone, particularly under the conditions used in the interrupted-unfolding experiments. For V113W, which is highly aggregation prone, our results suggest that it can form misfolded, likely oligomeric species, even under the conditions used in single-jump experiments. By carefully studying the kinetics of a particularly destabilized mutant, F108W, it was also possible to observe the predicted lag phase in the formation of the native state from I_1, thus confirming the sequential folding mechanism proposed. Of particular interest, is the fact that both the transition states from I_1 and I_2 to the native state are quite plastic and change with mutation and denaturant concentration. Given these and previous studie have established that it is likely that the threading event that creates the 5_2 knot occurs during the folding of the intermediates to the native state, our results suggest that there is a very broad energy barrier that is consistent with the chain undergoing some searching of conformational space as the threading takes place.

Author Contributions:
HZ and SEJ designed the experimental approach, HZ performed the research and undertook the data analysis, HZ and SEJ wrote the paper.

References

1. T. O. Yeates, T. S. Norcross, and N. P. King. 2007. Knotted and topologically complex proteins as models for studying folding and stability. Curr. Opin. Chem. Biol, vol. 11, no. 6, pp. 595–603.
2. V. Daggett and A. Fersht. 2003. The present view of the mechanism of protein folding. Nat. Rev. Mol. Cell Biol. vol. 4, no. 6, pp. 497–502.
3. M. L. Mansfield. 1994. Are there knots in proteins. Nature, vol. 1, no. 9, pp. 638–653.
4. W. R. Taylor. 2000. A deeply knotted protein structure and how it might fold. Nature, vol. 406, no. 6798, pp. 916–9.
5. P. Virnau, A. Mallam, and S. Jackson. 2011. Structures and folding pathways of topologically knotted proteins. J. Phys. Condens. Matter, vol. 23, no. 3, p. 033101.
6. A. L. Mallam and S. E. Jackson. 2005. Folding studies on a knotted protein. J. Mol. Biol. vol. 346, no. 5, pp. 1409–21.

7. A. L. Mallam and S. E. Jackson. 2006. Probing nature's knots: the folding pathway of a knotted homodimeric protein. J. Mol. Biol. vol. 359, no. 5, pp. 1420–36.
8. A. L. Mallam and S. E. Jackson. 2007. A Comparison of the Folding of Two Knotted Proteins : YbeA and YibK. J. Mol. Biol. vol. 366, no. 2, pp. 650–665.
9. A. L. Mallam and S. E. Jackson. 2007. The dimerization of an alpha/beta-knotted protein is essential for structure and function. Structure, vol. 15, no. 1, pp. 111–22.
10. A. L. Mallam, E. R. Morris, and S. E. Jackson. 2008. Exploring knotting mechanisms in protein folding. Proc. Natl. Acad. Sci. vol. 105, no. 48.
11. A. L. Mallam, S. C. Onuoha, J. G. Grossmann, and S. E. Jackson. 2008. Knotted fusion proteins reveal unexpected possibilities in protein folding. Mol. Cell, vol. 30, no. 5, pp. 642–8.
12. P. Virnau, L. a Mirny, and M. Kardar. 2006. Intricate knots in proteins: Function and evolution. PLoS Comput. Biol. vol. 2, no. 9, p. e122.
13. P. Szymczak. 2013. Tight knots in proteins: can they block the mitochondrial pores? Biochem. Soc. Trans. vol. 41, no. 2, pp. 620–4.
14. S. Wallin, K. B. Zeldovich, and E. I. Shakhnovich. 2007. The Folding Mechanics of a Knotted Protein. J. Mol. Biol. vol. 368, no. 3, pp. 884–893.
15. M. A. Soler and P. F. N. Faísca. 2013. Effects of Knots on Protein Folding Properties. PLoS One, vol. 8, no. 9, p. e74755.
16. J. I. Sulkowska, E. J. Rawdon, K. C. Millett, J. N. Onuchic, and A. Stasiak. 2012. Conservation of complex knotting and slipknotting patterns in proteins. Proc. Natl. Acad. Sci. vol. 109, no. 26, pp. E1715–E1723.
17. J. I. Sułkowska, P. Sułkowski, and J. Onuchic. 2009. Dodging the crisis of folding proteins with knots. Proc. Natl. Acad. Sci. U. S. A., vol. 106, no. 9, pp. 3119–3124.
18. S. E. Jackson and A. R. Fersht. 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry, vol. 30, no. 43, pp. 10428–35.
19. B. A. Krantz and T. R. Sosnick. 2000. Distinguishing between two-state and three-state models for ubiquitin folding. Biochemistry, vol. 39, no. 38, pp. 11696–11701.
20. N. A. Van Nuland, W. Meijberg, J. Warner, V. Forge, R. M. Scheek, G. T. Robillard, and C. M. Dobson. 1998. Slow cooperative folding of a small globular protein HPr. Biochemistry, vol. 37, no. 2, pp. 622–37.
21. S. W. Englander, T. R. Sosnick, L. C. Mayne, M. Shtilerman, P. X. Qi, and Y. Bai, 1998. Fast and Slow Folding in Cytochrome c. Acc. Chem. Res. vol. 31, pp. 744–767.
22. G. Wildegger and T. Kiefhaber. 1997. Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. J. Mol. Biol. vol. 270, no. 2, pp. 294–304.

23. S. E. Jackson. 1998. How do small single-domain proteins fold? Fold. Des. vol. 3, no. 4, pp. R81–91.
24. K. D. Wilkinson, K. M. Lee, S. Deshpande, P. Duerksen-Hughes, J. M. Boss, and J. Pohl. 1989. The neuron-specific protein PGP 9.5 is a ubiquitin carboxyl-terminal hydrolase. Science, vol. 246, no. 4930, pp. 670–3.
25. C. N. Larsen, B. A. Krantz, and K. D. Wilkinson. 1998. Substrate Specificity of Deubiquitinating Enzymes : Ubiquitin C-Terminal. Biochemistry. vol. 2960, no. 97, pp. 3358–3368.
26. C. Das, Q. Q. Hoang, C. A. Kreinbring, S. J. Luchansky, R. K. Meray, S. S. Ray, P. T. Lansbury, D. Ringe, and G. A. Petsko. 2006. Structural basis for conformational plasticity of the Parkinson ' s disease-associated ubiquitin hydrolase UCH-L1. Proc. Natl. Acad. Sci. vol. 103, no. 12. pp 4675-4680.
27. F. I. Andersson, D. G. Pina, A. L. Mallam, G. Blaser, and S. E. Jackson. 2009. Untangling the folding mechanism of the 5(2)-knotted protein UCH-L3. FEBS J. vol. 276, no. 9, pp. 2625–35.
28. F. I. Andersson, E. F. Werrell, L. Mcmorran, W. J. K. Crone, C. Das, S. D. Hsu, and S. E. Jackson. 2011. The Effect of Parkinson ' s-Disease-Associated Mutations on the Deubiquitinating Enzyme UCH-L1. J. Mol. Biol. vol. 407, no. 2, pp. 261–272.
29. S.-C. Lou, S. Wetzel, H. Zhang, E. W. Crone, Y.-T. Lee, S. E. Jackson, and S.-T. D. Hsu. 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. vol. 428, pp. 2507-20.

30. R. Fraczkiewicz and W. Braun. 1997. Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. J. Comput. Chem. vol. 19, no. 3, pp. 319-333
31. L. Serrano, A. Matouschek, and A. R. Fersht. 1992. The folding of an enzyme. VI. The folding pathway of barnase: comparison with theoretical models. J. Mol. Biol. vol. 224, no. 3, pp. 847–59.
32. A. R. Fersht and L. Serrano. 1991. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding. J. Mol. Biol., vol. 224, pp. 771–782.
33. G. Bulaj and D. P. Goldenberg. 2001. Phi-values for BPTI folding intermediates and implications for transition state analysis. Nat. Struct. Biol. vol. 8, no. 4, pp. 326–30.
34. T. M. Raschke, J. Kho, and S. Marqusee. 1999. Confirmation of the hierarchical folding of RNase H: a protein engineering study. Nat. Struct. Biol., vol. 6, pp. 825–831.
35. A. P. Capaldi, C. Kleanthous, and S. E. Radford. 2002. Im7 folding mechanism: misfolding on a path to the native state. Nat. Struct. Biol., vol. 9, no. 3, pp. 209–16.

36. H. Feng, N.-D. Vu, Z. Zhou, and Y. Bai. 2004. Structural examination of phi-value analysis in protein folding. Biochemistry, vol. 43, pp. 14325–14331.
37. C. Tanford. 1968. Protein denaturation. Adv. Protein Chem., vol. 23, pp. 121–282.
38. I. E. Sánchez and T. Kiefhaber. 2003. Hammond behavior versus ground state effects in protein folding: Evidence for narrow free energy barriers and residual structure in unfolded states. J. Mol. Biol. vol. 327, no. 4, pp. 867–884.
39. H. Taskent, J. H. Cho, and D. P. Raleigh. 2008. Temperature-Dependent Hammond Behavior in a Protein-Folding Reaction: Analysis of Transition-State Movement and Ground-State Effects. J. Mol. Biol., vol. 378, no. 3, pp. 699–706.
40. A. Matouschek, D. E. Otzen, L. S. Itzhaki, S. E. Jackson, and A. R. Fersht. 1995. Movement of the position of the transition state in protein folding. Biochemistry, vol. 34, no. 41, pp. 13656–62.
41. D. A. Boudreaux, T. K. Maiti, C. W. Davies, and C. Das. 2010. Ubiquitin vinyl methyl ester binding orients the misaligned active site of the ubiquitin hydrolase UCHL1 into productive conformation. vol. 107, no. 20.
42. A. Matouschek, L. Serrano, and A. R. Fersht. 1992. The folding of an enzyme. J. Mol. Biol. vol. 224, no. 3, pp. 819–835.
43. M. Jamroz, W. Niemyska, E.J. Rawdon, A. Stasiak, K.C. Millett, P. Sułkowski and J.I. Sulkowska. 2015. KnotProt: a database of proteins with knots and slipknots. Nucleic Acids Res. vol. 43, D306-14.
44. S.E. Jackson, A. Suma and C. Micheletti. 2016. How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. Curr. Opin. Struct. Biol. In press.
45. M. Chwastyk and M. Cieplak. 2015. Multiple folding pathways of proteins with shallow knots and co-translational folding. J Chem Phys. Vol. 143(4) pp045101.

.

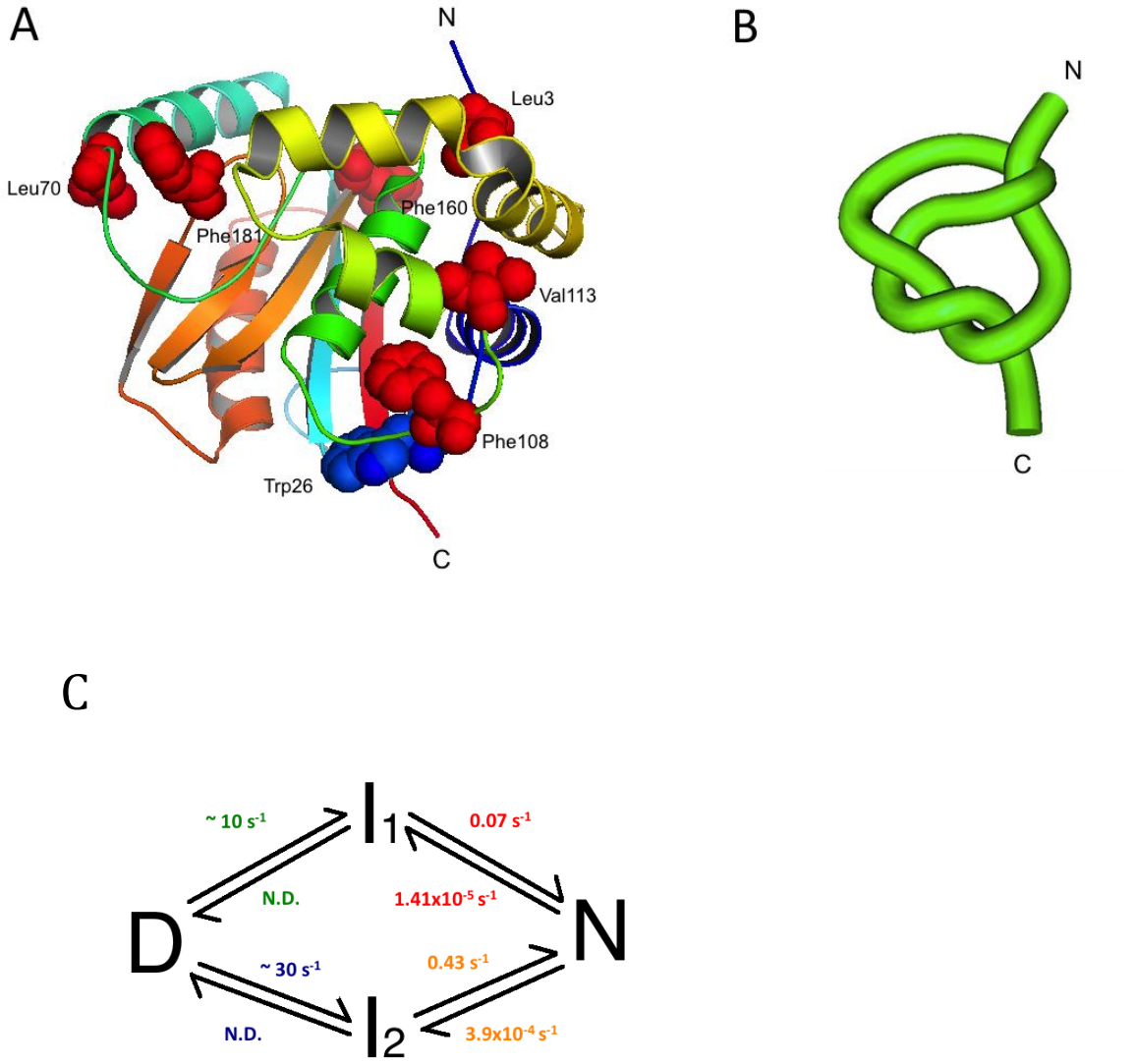

Figure 1: Structure of human wild-type UCH-L1. (A) The main chain is shown as a ribbon, colored from N-terminus (dark blue) to C-terminus (red). The side chains of the residues mutated in this study are labeled and shown in red in a space-filling fashion. Trp26 in wild-type UCH-L1 is displayed in blue. (B) Schematic representation of the 5_2-knotted topology of UCH-L1. (C) Folding mechanism proposed for wild-type UCH-L1 (29). The forward and reverse first-order rate constants for each transition are shown above and below the arrows for each transition. The values are color coded such that the rate constants for the N-I_1 and N-I_2 transitions are shown in red and orange, respectively, and the rate constants for the D-I_1 and D-I_2 transitions are shown in green and blue, respectively. This color scheme is used throughout.

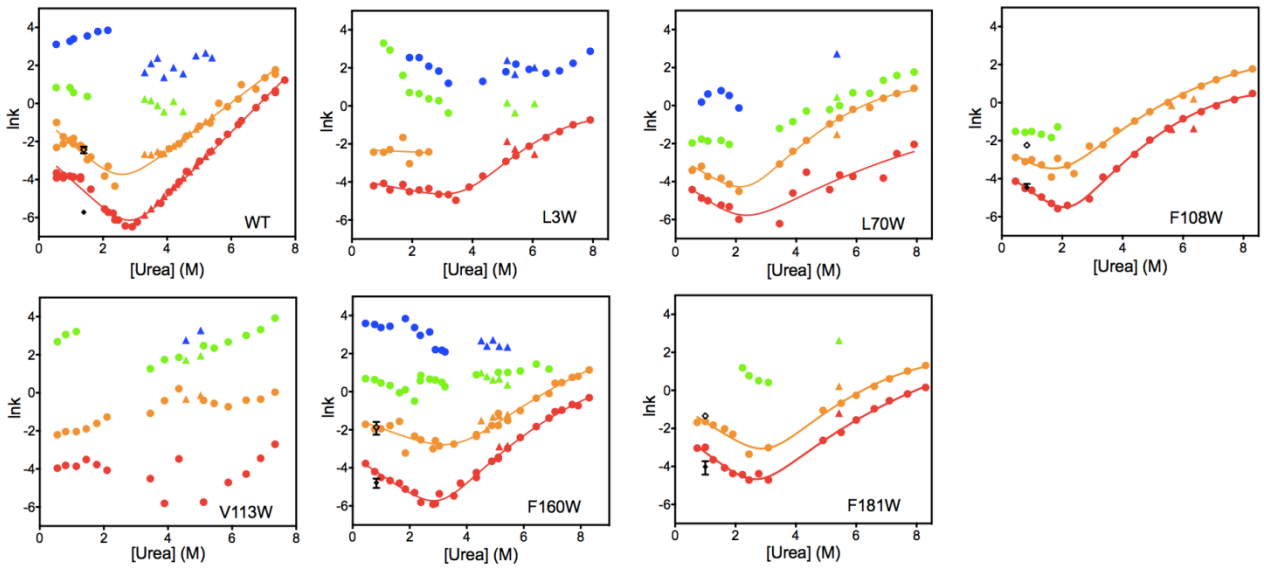

Figure 2: Chevron plots of the single- and double-jump kinetic data for wild-type and mutant UCH-L1. Each kinetic phase has been assigned to a different structural transition and is shown in distinct colors: N-I_1 (red), N-I_2 (orange), I_1-D (green) and I_2-D (blue). Unfolding and refolding rate constants from single-jump experiments are shown as filled circles. Unfolding rate constants measured in interrupted-refolding experiments are shown as filled triangles. The rate constants associated with the evolution of each unfolding phase measured in the interrupted-refolding experiments are shown as open or filled diamonds with their associated fitting errors displayed. Wild-type data were taken from an earlier investigation (28).

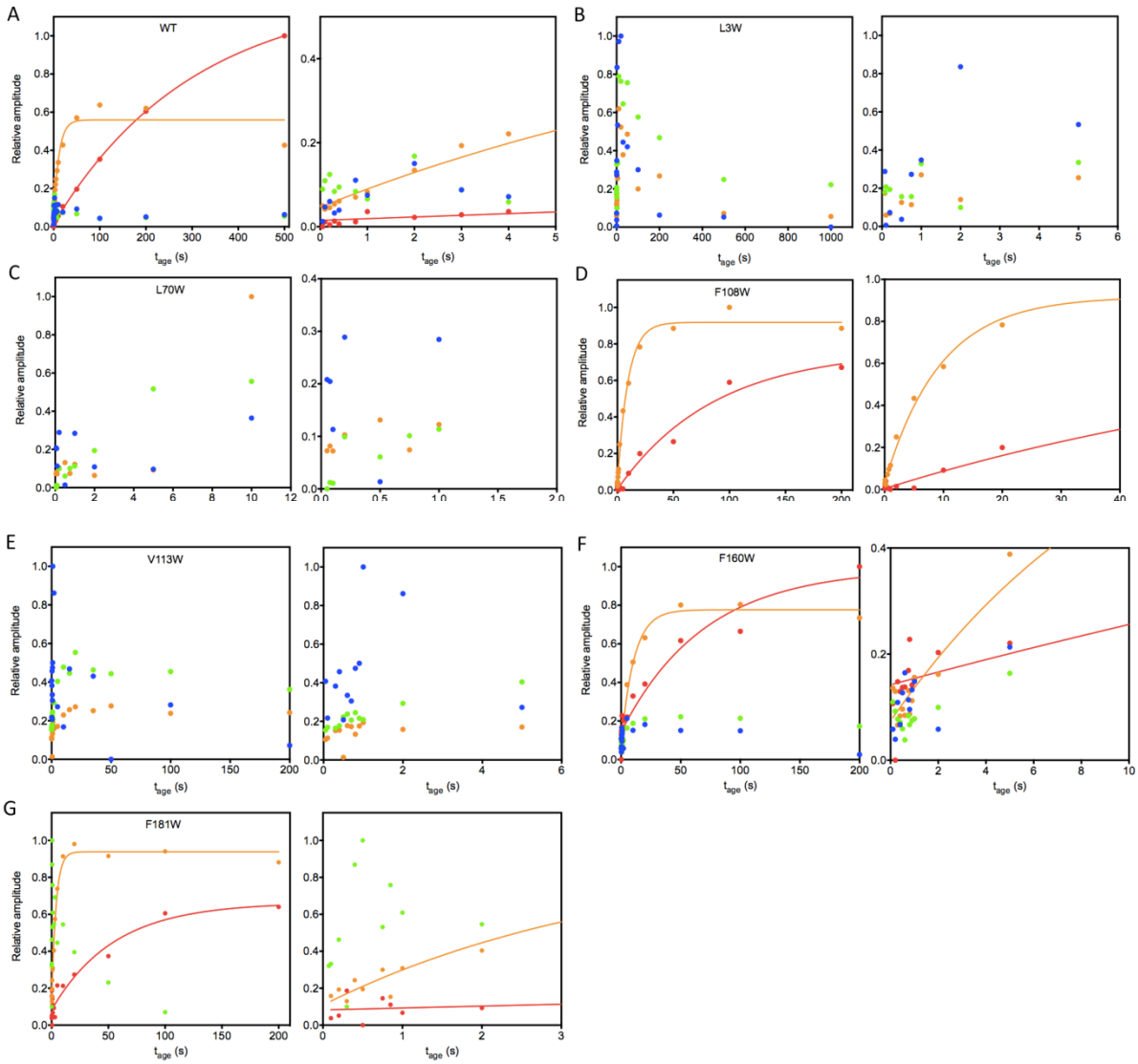

Figure 3: Amplitudes of the UCH-L1 unfolding phases from interrupted-refolding experiments. Data for wild type were taken from earlier investigation and are shown in (A) (29). Data for the variants are shown in (B)-(G), respectively. Each unfolding phase is colored according to the transition with which it is associated: N-I_1 (red), N-I_2 (orange), I_1-D (green) and I_2-D (blue). The continuous lines represent the best fit of the amplitude change against aging time to a single-exponential equation, Equation S7. The data were normalized by setting the highest amplitude as 1 and lowest amplitude as 0. For wild type and each UCH-L1 variant, the panel on the right hand side of the main figure shows the amplitude change over short aging times

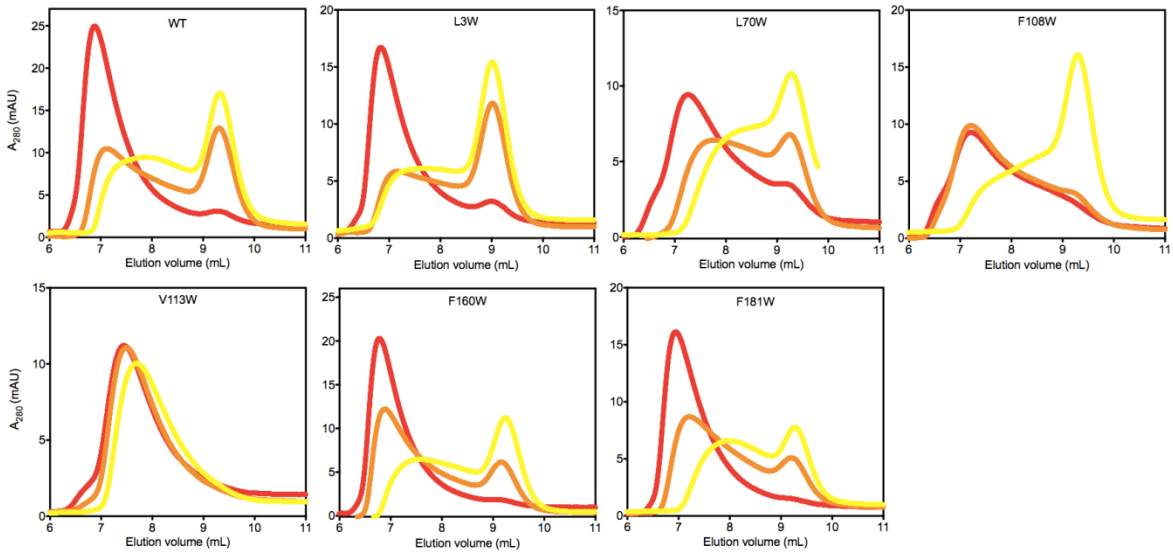

Figure 4: Analytical size-exclusion chromatography of wild-type and mutant UCH-L1 at 2-3 M urea. The chromatograms are colored by equilibration time: 1 min (yellow), 10 min (orange) and 60 min (red).

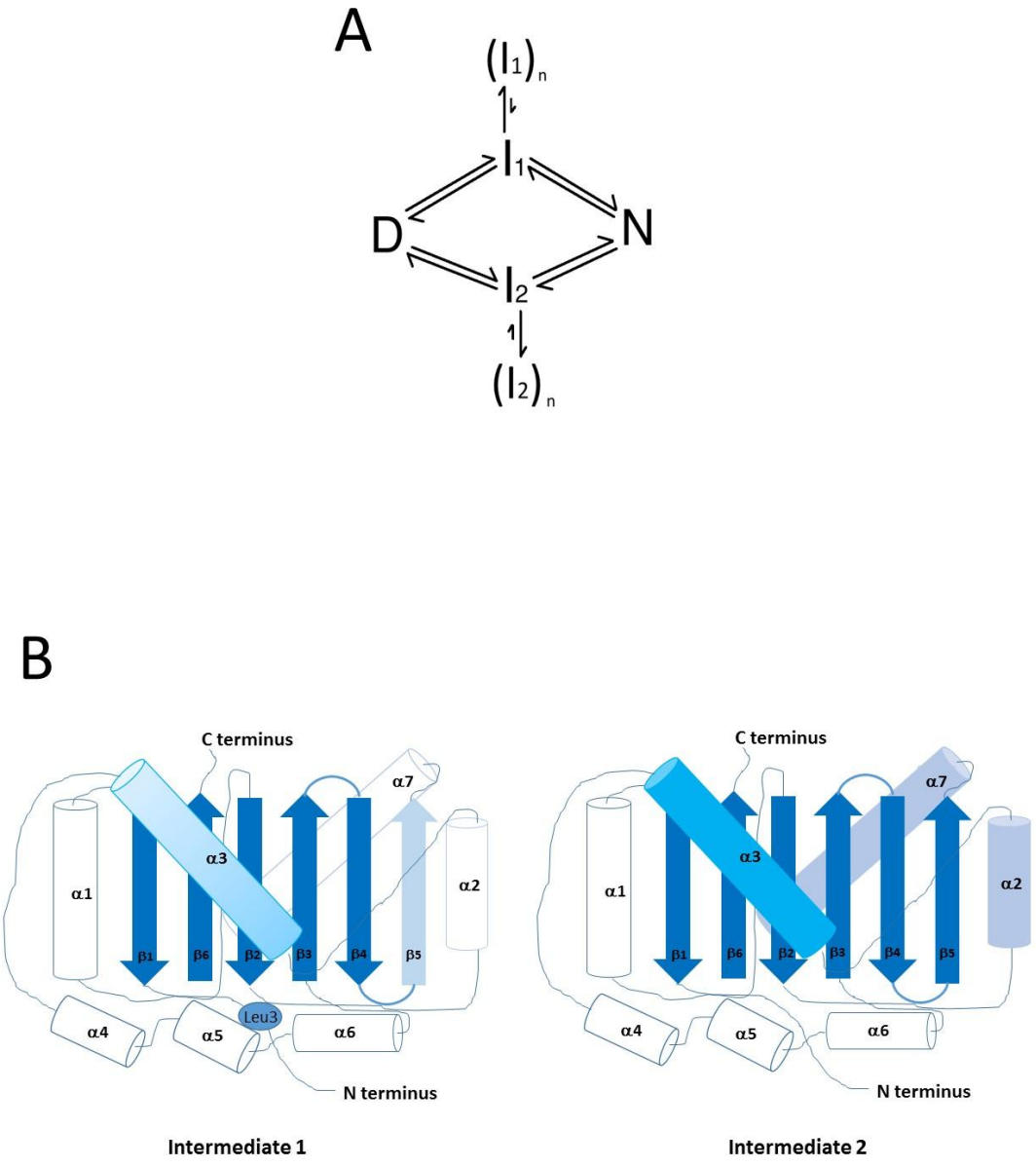

Figure 5: Proposed folding pathway of UCH-L1 and structures of the intermediate states. **(A)** Kinetic scheme consistent with the kinetic data. The off-pathway oligomers $(I_1)_n$ and $(I_2)_n$ are formed from the aggregation-prone intermediate states I_1 and I_2. (B) Schematic diagram of structures of the intermediate states I_1 and I_2 based on kinetic data presented here and previous HDX NMR data (29). The differing degrees of structure in the intermediates is represented by the color: dark blue (highly structured), mid blue (some secondary and tertiary structure), light blue (some tertiary interactions but no stable secondary structure). In I_1, the gradation in color along α-helix 3 that the N-terminal region is more structured than the C-terminus of the helix.

	[a][urea]$_{I\text{-}N}^{50\%}$ (M) Fl	[a][urea]$_{D\text{-}I}^{50\%}$ (M) FL	[b][urea]$_{I\text{-}N}^{50\%}$ (M) far-UV CD	[c]$\Delta\Delta G_{I\text{-}N}^{H_2O}$ (kcal mol^{-1}) Fl	[c]$\Delta\Delta G_{D\text{-}I}^{H_2O}$ (kcal mol^{-1}) Fl	[d]$\Delta\Delta G_{D\text{-}N}^{H_2O}$ (kcal mol^{-1}) Fl	[e]Φ_I Fl
WT	2.5 ± 0.1	4.1 ± 0.2	2.60 ± 0.02				
L3W	2.4 ± 0.1	3.7 ± 0.1	2.60 ± 0.02	0.3 ± 0.4	0.6 ± 0.3	0.9 ± 0.5	0.7 ± 0.6
W26F	2.5 ± 0.1	ND	2.60 ± 0.02				
L70W	1.5 ± 0.1	3.1 ± 0.6	2.10 ± 0.02	2.8 ± 0.6	1.5 ± 1.0	4.3 ± 1.1	0.3 ± 0.2
F108W	1.9 ± 0.1	ND	2.20 ± 0.02	2.4 ± 1.1	ND	ND	ND
V113W	1.7 ± 0.1	2.7 ± 0.2	1.80 ± 0.02	2.2 ± 0.5	2.1 ± 0.5	4.3 ± 0.7	0.5 ± 0.1
F160W	2.7 ± 0.1	4.0 ± 0.5	2.90 ± 0.02	-0.8 ± 0.4	0.2 ± 0.8	-0.6 ± 0.9	-0.3 ± 1.2
F181W	2.5 ± 0.1	ND	2.50 ±0.02	1.1 ± 1.0	ND	ND	ND

TABLE 1: Φ_I values and the difference in unfolding free energies between wild-type and variant UCH-L1.

a. The concentration of urea corresponding to the midpoint of unfolding between the native and intermediate states as calculated from fluorescence data. See Table S3 for further thermodynamic parameters.
b. The concentration of urea corresponding to the midpoint of unfolding between the intermediate and denatured states as calculated from fluorescence data. See Table S3 for further thermodynamic parameters.
c. The concentration of urea corresponding to the midpoint of unfolding between the native and intermediate states as calculated from the far-UV CD data. See Table S3 for further thermodynamic parameters.
d. The difference in the change in unfolding free energies for both N-I and I-D transitions based on fluorescence equilibrium unfolding measurements were calculated using Equation S5.
e. $\Delta\Delta G_{D\text{-}N}^{H_2O} = \Delta\Delta G_{I\text{-}N}^{H_2O} + \Delta\Delta G_{D\text{-}I}^{H_2O}$
f. The Φ_I value was obtained using Equation S6.

Note that all the mutants also contain a second mutation W26F. This mutant was also made and stability measurements undertaken, Fig. S5. In this case, the mutation had little effect on the stability of either the native or intermediate states, such that wild-type UCH-L1 is a suitable reference state.

Fl = fluorescence

ND = not determined

	$k_f^{H_2O}$ (s^{-1})	m_{kf} (M^{-1})	$k_u^{H_2O}$ (s^{-1})	m_{ku} (M^{-1})	[c] m_{kin} (kcal $mol^{-1}M^{-1}$)	[d] $[urea]_{I\text{-}N}$ (M)
[a]WT (N-I_1)	0.09 ± 0.01	1.5 ± 0.1	0.000009 ± 0.000003	1.7 ± 0.1	1.9 ± 0.1	2.9 ± 0.2
[a]WT (N-I_2)	0.6 ± 0.1	1.5 ± 0.2	0.0005 ± 0.0002	1.3 ± 0.1	1.7 ± 0.2	2.5 ± 0.4
[b]L3W (N-I_1)	0.020 ± 0.004	0.3 ± 0.1	0.0000004 ± 0.0000008	3.3 ± 0.7	2.1 ± 0.5	3.0 ± 0.7
[b]L3W (N-I_2)	0.2 ± 0.1	0.1 ± 0.3	ND	ND	ND	ND
[b]L70W (N-I_1)	0.02 ± 0.02	1.2 ± 0.9	0.0002 ± 0.0006	1.1 ± 1.4	1.4 ± 1.4	2.0 ± 2.1
[b]L70W (N-I_2)	0.07 ± 0.01	0.9 ± 0.2	0.00004 ± 0.00004	2.5 ± 0.4	2.0 ± 0.4	2.2 ± 0.4
[b]F108W (N-I_1)	0.028 ± 0.004	1.1 ± 0.1	0.000014 ± 0.000006	2.6 ± 0.2	2.2 ± 0.2	2.1 ± 0.2
[b]F108W (N-I_2)	0.09 ± 0.03	0.9 ± 0.4	0.0007 ± 0.0007	1.8 ± 0.4	1.6 ± 0.5	1.8 ± 0.6
[b]F160W (N-I_1)	0.03 ± 0.01	1.0 ±0.1	0.000004 ± 0.000004	2.3 ± 0.3	2.0 ± 0.2	2.7 ± 0.4
[b]F160W (N-I_2)	0.3± 0.1	0.5 ± 0.2	0.0001 ± 0.0004	1.7 ± 0.9	1.3 ± 0.7	3.6 ± 2.0
[b]F181W (N-I_1)	0.13 ± 0.03	1.2 ±0.2	0.00007 ± 0.00006	1.8 ±0.3	1.8 ± 0.3	2.5 ± 0.8
[b]F181W (N-I_2)	0.5 ± 0.2	1.0 ± 0.2	0.00008 ± 0.00016	2.2 ± 0.7	1.9 ±0.5	2.7 ± 0.8

The sum of parameters	[f] m_{kin}^{sum} (kcal $mol^{-1}M^{-1}$)	[d] $[urea]_{I\text{-}N}^{50\%}$ (M)	[g] ΔG_{I} (kca
WT (N-I)	3.6 ± 0.2	2.7 ± 0.2	9.7
L70W (N-I)	3.4 ± 1.4	2.1 ± 0.4	7.2
F108W (N-I)	3.8 ± 0.5	2.0 ± 0.2	7.4
F160W (N-I)	3.3 ±0.7	3.1 ± 0.2	10.
F181W (N-I)	3.7 ± 0.6	2.6 ± 0.2	9.6

TABLE 2: Kinetic and thermodynamic parameters of the unfolding and

refolding of wild-type and mutant UCH-L1.

a. Wild-type data were taken from an earlier investigation (29) and analyzed using a simple two-state model, Equation S8.
b. Data of all mutants were analyzed using Equation S9.
c. Calculated using Equation S11.
d. Calculated using Equation S2.
e. Calculated using Equation S10.
f. $m_{kin}^{sum} = m_{kin}(\text{N-I}_1) + m_{kin}(\text{N-I}_2)$
g. $\Delta G_{kin}^{H_2O}{}_{\text{sum}} = \Delta G_{kin}^{H_2O}(\text{N-I}_1) + \Delta G_{kin}^{H_2O}(\text{N-I}_2)$

ND = not determined

Supporting Material

Materials and Methods

Equilibrium unfolding experiments

The conformational stabilities of wild type and mutants were measured using urea unfolding experiments conducted under equilibrium conditions where it has been established that the unfolding of wild-type UCH-L1 is fully reversible. Unfolding was monitored by tryptophan fluorescence and far-UV CD spectroscopy to probe changes in the tertiary and secondary model, respectively

An 8 M urea stock, the exact concentration of which was determined using a refractometer, was diluted volumetrically with 50 mM Tris, 5 mM DTT, 0.5 mM EDTA, pH 7.6 (for fluorescence measurements) or with 50 mM phosphate, 5 mM DTT, pH 7.6 (for far-UV CD measurements). This was used to prepare approximately 60-70 aliquots of 400 μL spanning a concentration range between 0-7 M urea. 50 μL protein solution in buffer was added to each aliquot such that the final protein concentration was 3 or 10 μM for fluorescence or far-UV CD measurements, respectively. Samples were thermostatted at 25 °C for more than 5 h to allow them to reach equilibrium. Fluorescence measurements were performed in a 3 mm pathlength cuvette using a Cary Eclipse fluorimeter. The protein was excited at 280 nm and the emission spectrum between 300-400 nm were recorded at each urea concentration. In addition, the wavelength at which there is the largest difference between the native and denatured signal was selected individually for each mutant and the fluorescence signal intensity of a 5 nm range around this particular wavelength was integrated. These data were then normalized, such that the minimum signal was set at zero and the maximal signal as one, to produce the relative fluorescence as a function of urea concentration. Far-UV CD measurements were performed in a 1 mm pathlength cuvette using an Applied Photophysics Chirascan spectrometer. The signal between 222 and 225 nm was integrated and plotted against urea concentration.

The fluorescence spectra of wild-type and mutant UCH-L1 at different urea concentrations are shown in Fig. S3. For each variant, the fluorescence at the wavelength where there was the maximal difference in signal between the native and denatured states was normalized and plotted against urea concentration, Fig. S4A. For wild type, L3W, L70W, V113W and F160W, two transitions were observed and the data for these variants were globally fit to a three-state unfolding model to obtain values for the thermodynamic parameters associated with the unfolding transitions from native to intermediate state, and intermediate to denatured state. The thermodynamic parameters obtained - free

energy change between native and intermediate states, $\Delta G_{\text{I-N}}{}^{\text{H}_2\text{O}}$, the intermediate and denatured states, $\Delta G_{\text{D-I}}{}^{\text{H}_2\text{O}}$, the urea concentrations corresponding to the midpoints of the transitions between native and intermediate and intermediate and denatured states, $[\text{urea}]_{\text{I-N}}{}^{50\%}$ and $[\text{urea}]_{\text{D-I}}{}^{50\%}$, respectively, and $m_{\text{I-N}}$ and $m_{\text{D-I}}$ constants related to the change in solvent-accessible surface area between the native and intermediate, and intermediate and denatured states, for wild-type and variants are shown in Table S3.

For W26F, F108W and F181W, fluorescence only probes the N to I transition, Fig. S3A, so in these cases, the relative fluorescence data were globally fit to a two-state model, Fig S4A, to obtain thermodynamic parameters for the transition between native and intermediate state only, Table S3. In addition, for all the variants the far-UV CD probes only the N to I transition, and the mean-residue ellipticity as a function of urea concentration, Fig. S4B, were also fit to a two-state model to obtain the thermodynamic parameters for the N to I transition, Table S1. For most mutants, the midpoints of unfolding determined by far-UV CD and fluorescence for the N-I transition were similar but some were not within the fitting errors. In order to assess whether this was significant, we carried out three independent fluorescence and far-UV CD measurements for one mutant, F160W, such that we obtained average values and standard deviations for the different thermodynamic parameters, Table S4. In this case, the midpoints of unfolding using the different structural probes are all within the standard deviations measured. .

Single [urea]-jump unfolding and refolding kinetics

Single [urea]-jump unfolding and refolding kinetic measurements were performed using a stopped-flow instrument (SX-20, Applied Photophysics) at 25 °C. Unfolding and refolding solutions were prepared by volumetric dilution of a 10 M urea stock solution into 50 mM Tris, 5 mM DTT, pH 7.6. Native or denatured protein was rapidly mixed in a 1:10 ratio with unfolding or refolding solutions such that the final protein concentration was 2 μM. For slow reactions, where data were acquired for over 1000 s, the experiments were performed by manual mixing using a Cary Eclipse fluorimeter thermostatted at 25 °C.

Interrupted-refolding experiments (double-jump)

72 μM of wild-type or mutant UCH-L1 was denatured in 5-6 M urea. The denatured protein was then diluted rapidly in a 1:5 ratio in 50 mM Tris, 5 mM DTT, pH 7.6 in the stopped-flow SX-20 instrument. This initiated refolding between 0.83-1 M urea for various aging times (t_{age} = 50 ms-1000 s). Unfolding was then initiated by further mixing in a 1:5 ratio into high concentrations of urea such that the final urea concentration varied between 4-7 M and the final

protein concentration was 2 μM.

Data analysis

Equilibrium unfolding data: fitting to a three-state or two-state unfolding model

Equilibrium unfolding data were analyzed by plotting the fluorescence against denaturant concentration. For wild-type, V113W and F160W, the data were fitted to a three-state model using Equation S1 using the non-linear, least-squared fitting program GraphPad Prism, version 5 (GraphPad Software):

$$F_{obs} = F_N + (F_I - F_N)\frac{K_{I-N}}{1+K_{I-N}+K_{I-N}K_{D-I}} + (F_D - F_N)\frac{K_{I-N}K_{D-I}}{1+K_{I-N}+K_{I-N}K_{D-I}} \quad \text{Equation S1}$$

where F is the fluorescence signal for each state, N, I and D. The equilibrium constants, K_{I-N} and K_{D-I}, for the N-I and I-D transition, respectively can be related to $\Delta G_{I-N}^{H_2O}$ and $\Delta G_{D-I}^{H_2O}$, which are the difference in free energies between the I and N states and between the D and I states in water, which depend on denaturant concentration as shown by:

$$\Delta G_{I-N} = -RT\ln(K_{I-N}) = \Delta G_{I-N}^{H_2O} - m_{I-N}[urea]$$

$$\Delta G_{D-I} = -RT\ln(K_{D-I}) = \Delta G_{D-I}^{H_2O} - m_{D-I}[urea]$$

where the m_{I-N} and m_{D-I} values are constants and related to the change in solvent accessible surface area on unfolding for each transition. R is the gas constant and is 1.987×10^{-3} kcal K^{-1} mol^{-1}. T is the temperature 298 K.

When ΔG_{I-N} or ΔG_{D-I} are zero, this defines the point at which the concentration of N and I, or I and D, are the same, which corresponds to the midpoints of the transitions. Therefore:

$$\Delta G_{I-N}^{H_2O} = m_{I-N}[urea]_{50\%}^{I-N}$$

$$\Delta G_{D-I}^{H_2O} = m_{D-I}[urea]_{50\%}^{D-I}$$

Equation S2

where $[urea]_{50\%}$ is the denaturant concentration for the midpoint of each

unfolding transition. Therefore, the equilibrium constants in Equation S1 can be expressed in terms of *m*-values and $[urea]_{50\%}$:

$$K_{I\text{-}N} = \exp\left(\frac{m_{I\text{-}N}([urea] - [urea]_{50\%}^{I\text{-}N})}{RT}\right)$$

$$K_{D\text{-}I} = \exp\left(\frac{m_{D\text{-}I}([urea] - [urea]_{50\%}^{D\text{-}I})}{RT}\right)$$

Therefore, the change of fluorescence can be fitted as a function denaturant concentration.

For ease of comparison, the fluorescence data for each mutant were normalized to give a relative fluorescence:

$$S_r = \frac{S - S_N}{S_D - S_N}$$

Equation S3

where S_r is the relative fluorescence, S is the fluorescence signal at each urea concentration and S_D and S_N are the fluorescence of denatured and native states, respectively.

The far-UV CD equilibrium unfolding data for all the mutants, as well as the fluorescence unfolding data for W26F, F108W and F181W, where the only observable fluorescence change is between N and I, were analyzed using a two-state model with a sloping baseline for the denatured state:

$$F_{obs} = F_N + ((F_D + A_D[urea]) - F_N)\frac{1}{1+K_{D-N}}$$

Equation S4

where the equilibrium constant between N and D, $K_{D\text{-}N}$, can be expressed as:

$$K_{D\text{-}N} = \exp\left(\frac{m_{D\text{-}N}([urea] - [urea]_{50\%}^{D\text{-}N})}{RT}\right)$$

Calculation of the effects of mutation on the thermodynamic stabilities of native and intermediate state.

In order to compare the thermodynamic stabilities of wild type and mutants, to determine the degree to which the mutation has destabilized the native relative to the intermediate state, and the intermediate relative to the denatured state, the differences between the unfolding free energies of wild type and mutants for both N-I and I-D transitions are defined as:

$$\Delta\Delta G_{N-I}^{H_2O} = \Delta G_{N-I}^{H_2O}(WT) - \Delta G_{N-I}^{H_2O}(mut)$$

$$\Delta\Delta G_{I-D}^{H_2O} = \Delta G_{I-D}^{H_2O}(WT) - \Delta G_{I-D}^{H_2O}(mut)$$

Equation S5

From these energies, a Φ value for the intermediate state, Φ_I, can be calculated:

$$\Phi_I = \frac{\Delta\Delta G_{D-I}}{\Delta\Delta G_{D-N}} \quad \text{or} \quad \Phi_I = 1 - \frac{\Delta\Delta G_{I-N}}{\Delta\Delta G_{D-N}}$$ Equation S6

Kinetics

Kinetic traces were fitted using SX-20 Pro-Data viewer software (Applied Photophysics) or KaleidaGraph, Version 5 (Synergy Software). The kinetic traces of the wild type and mutants were fit individually to a first-order reaction with the required number of exponentials:

$$F_t = \sum_{i=1}^{N} A_i \exp(-k_i t) + C$$

Equation S7

where F_t is the fluorescence at time t, C is the fluorescence at t = 0, A_i is the amplitude of the corresponding kinetic phase and k_i is the rate constant.

Thus, one can obtain a number of unfolding and refolding rate constants as a function of denaturant concentration. When the natural logarithm of the unfolding and refolding rate constants is plotted against denaturant concentration, unfolding phases can sometimes be directly associated with a refolding phase, as they produce the well-known chevron plot. In these cases, the unfolding and refolding kinetic data for a single unfolding/refolding transition (rate constants versus denaturant concentration) were fitted to Equation S8 which describes a two-state kinetic model (wild type kinetic data only).

$$\ln k_{obs} = \ln(k_f^{H_2O} \exp(-m_{k_f}[urea]) + k_u^{H_2O} \exp(m_{k_u}[urea]))$$ Equation S8

where k_{obs} is the observed rate constant; $k_f^{H_2O}$ and $k_u^{H_2O}$ are the refolding and unfolding rate constants in water, which were obtained by extrapolation to y-axis, respectively; m_{k_f} and m_{k_u} are the slopes of the refolding or unfolding arms of the chevron plots for one particular unfolding/refolding transition (phase)

For the mutants, the kinetic data for each chevron plot were fit to Equation S9. This equation describes a two-state kinetic model with an additional term representing the curvature observed at high urea concentrations:

$$\ln k_{obs} = \ln(k_f^{H_2O} \exp(- m_{k_f}[urea]) + k_u^{H_2O} \exp(m_{k_u}[urea] - m_{k_u 2}[urea]^2)) \quad \text{Equation S9}$$

where $m_{k_u 2}[urea]^2$ is an additional term defining the curvature at high urea concentrations.

The free energy of unfolding in water, $\Delta G_{kin}{}^{H_2O}$, was calculated from the kinetic data using:

$$\Delta G_{kin}^{H_2O} = - RT \ln(k_u^{H_2O} / k_f^{H_2O}) \quad \text{Equation S10}$$

where $k_u{}^{H_2O}$ and $k_f{}^{H_2O}$ are the unfolding and refolding rate constants in water, respectively.

A m_{kin} value was calculated using the kinetic data using Equation S11:

$$m_{kin} = RT(m_{ku} + m_{kf}) \quad \text{Equation S11}$$

The Tanford β value, β_T, is a measure of the average degree of exposure in the transition state relative to that of the denatured state and the native state. It is a useful index of the compactness of the transition state and was calculated using Equation S12:

$$b_T = \frac{m_{kf}}{m_{kf} + m_{ku}} \quad \text{Equation S12}$$

Supporting Material: Figures and Tables

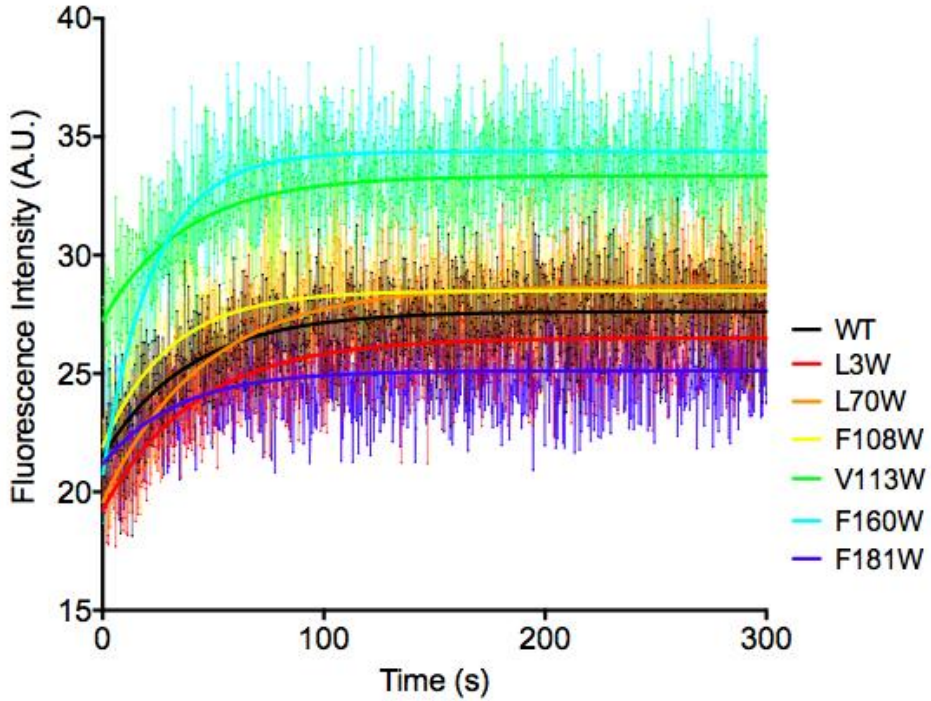

Figure S1: Enzymatic activity assay of wild-type and variant UCH-L1. The figure shows the steady-state kinetics for the hydrolysis of the model susbtrate Ubq-AMC. Data are presented in the following colors: wild type (black), L3W (red), L70W (orange), F108W (yellow), V113W (green), F160W (cyan) and F181W (blue). The solid lines show the best of the data to a single exponential fit.

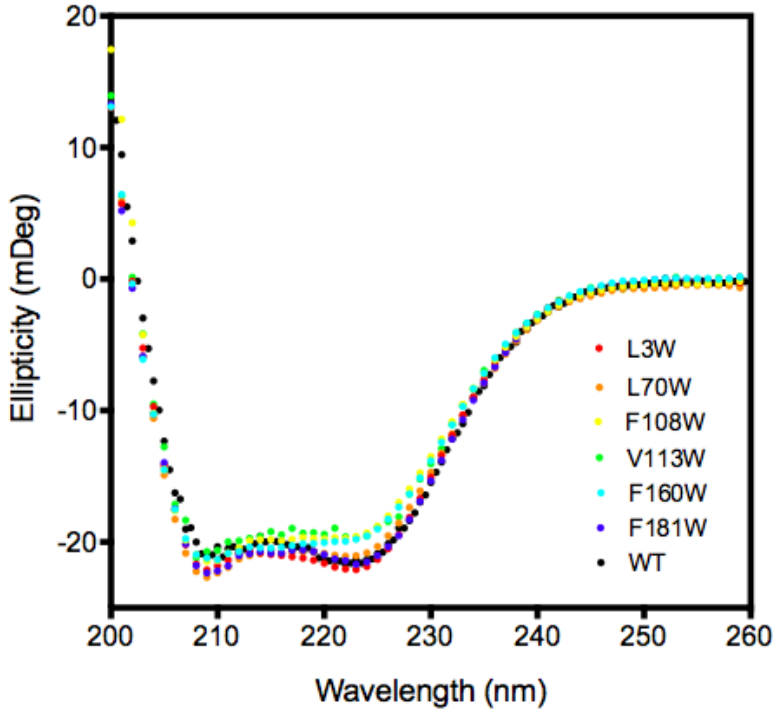

Figure S2: Far-UV CD spectra of wild-type and mutant UCH-L1. The variants are presented in the following colors: wild type (black), L3W (red), L70W (orange), F108W (yellow), V113W (green), F160W (cyan) and F181W (blue). In all cases, the protein concentration was 10 μM.

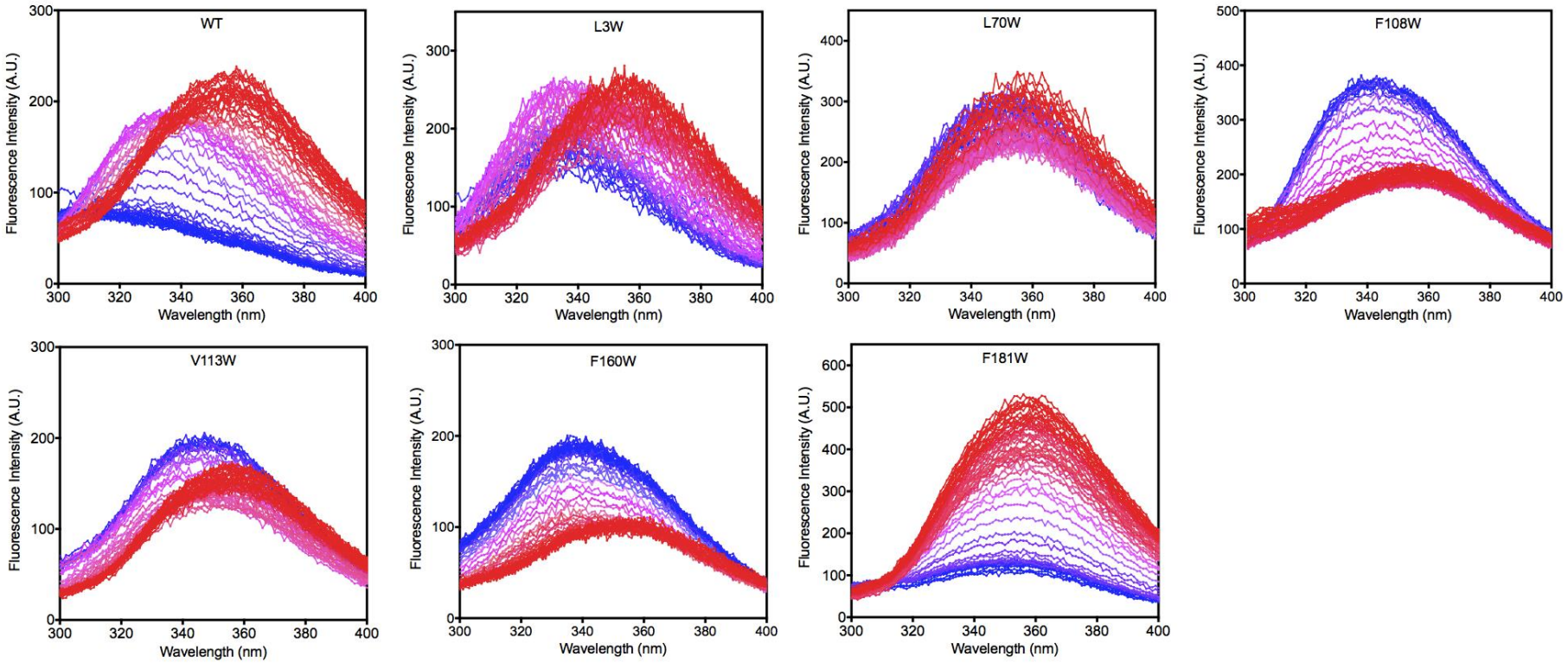

Figure S3: Fluorescence spectra of wild-type and mutant UCH-L1 at different urea concentrations. As the protein unfolds, the λ_{max} red shifts from lower to higher wavelengths. The spectra are colored blue at lower urea concentrations (where the protein is native) to red at higher urea concentrations (where the protein is fully unfolded). For wild type, the λ_{max} is 310, 335 and 355 nm for the native, intermediate and denatured states, respectively. All of the mutants have emission maxima around 330-340 nm under native conditions, and there is a red-shift on unfolding to 355-360 nm.

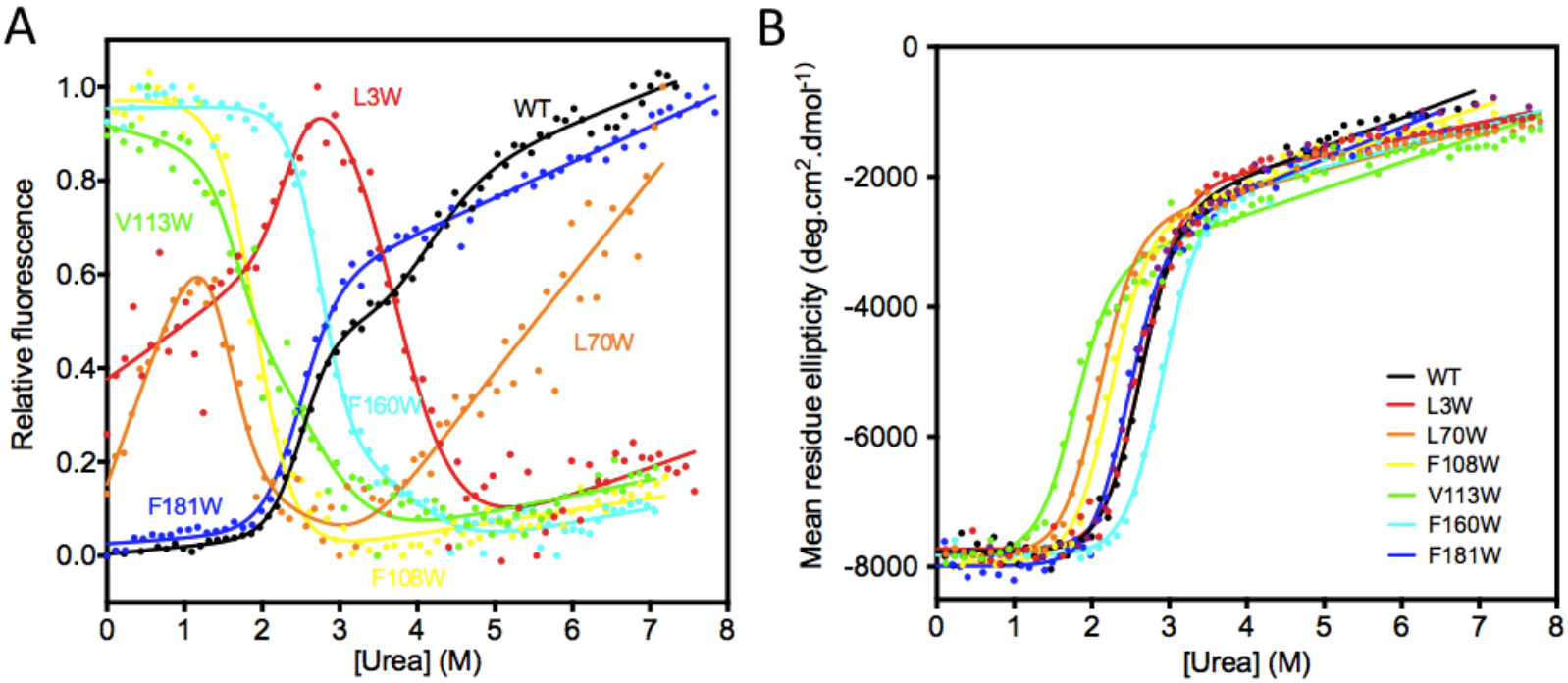

Figure S4: Equilibrium urea denaturation of wild-type and mutant UCH-L1 studied by (A) fluorescence and (B) far-UV CD.
For each mutant, the fluorescence signal over a 4 nm wavelength range was integrated and is shown. The wavelength range varied from mutant to mutant and corresponds to the wavelength range where there is a maximal difference in the signal between the native and denatured states. The wavelength ranges were 361-365 nm (wild type), 318-322 nm (L3W), 361-365 nm (L70W), 33-337 nm (F108W), 327-331 nm (V113W), 332-336 nm (F180W), 354-358 nm (F181W). For ease of comparison, these fluorescence signals were then normalized using Equation S3. The solid lines represent the best fit of the data to a three- or two-state model. For wild type, L3W, L70W, V113W and F160W, a fluorescence change is observed for the transitions between native and intermediate, and intermediate and denatured states; data were fit to a three-state model, Equation S1. The fluorescence data for the denaturation of F108W and F180W and

the far-UV CD data for wild type and all the mutants were all analyzed using a two-state unfolding model, Equation S4. This is because the change in signal corresponding to the I-D transition is too small for the data to be fit to a three-state model. The thermodynamic parameters obtained from the best fits are given in Table 1.

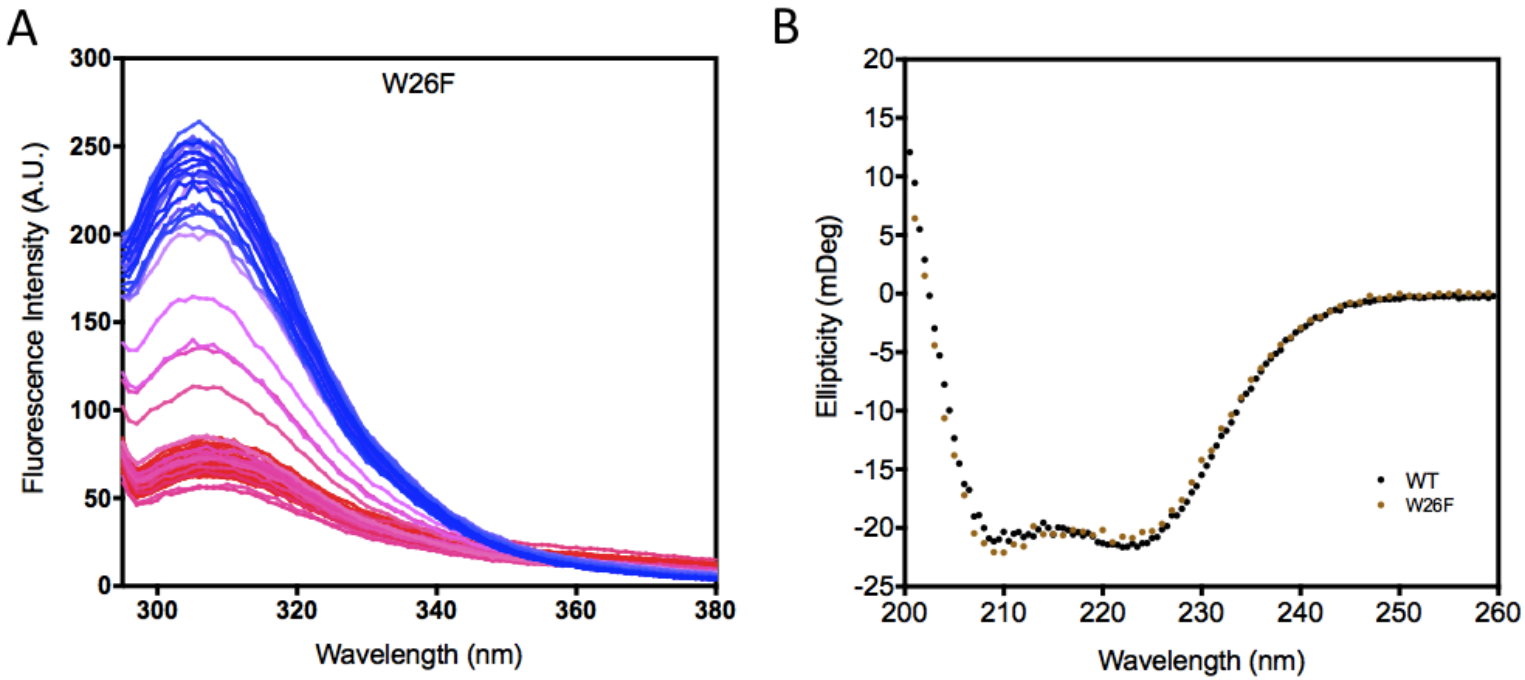

Figure S5 (and overleaf): Structure and stability of W26F, a Trp-less mutant of UCH-L. (A) Fluorescence spectra of W26F at low (blue) to high (red) urea concentrations. The fluorescence spectra of W26F differ from wild type as in the latter the emission from Trp dominates the fluorescence spectra, whilst for W26F, which contains no Trp, tyrosine fluorescence dominates. (B) Far-UV CD spectra of 10 μM wild type (black-filled circles) and W26F (brown-filled circles) under native conditions.

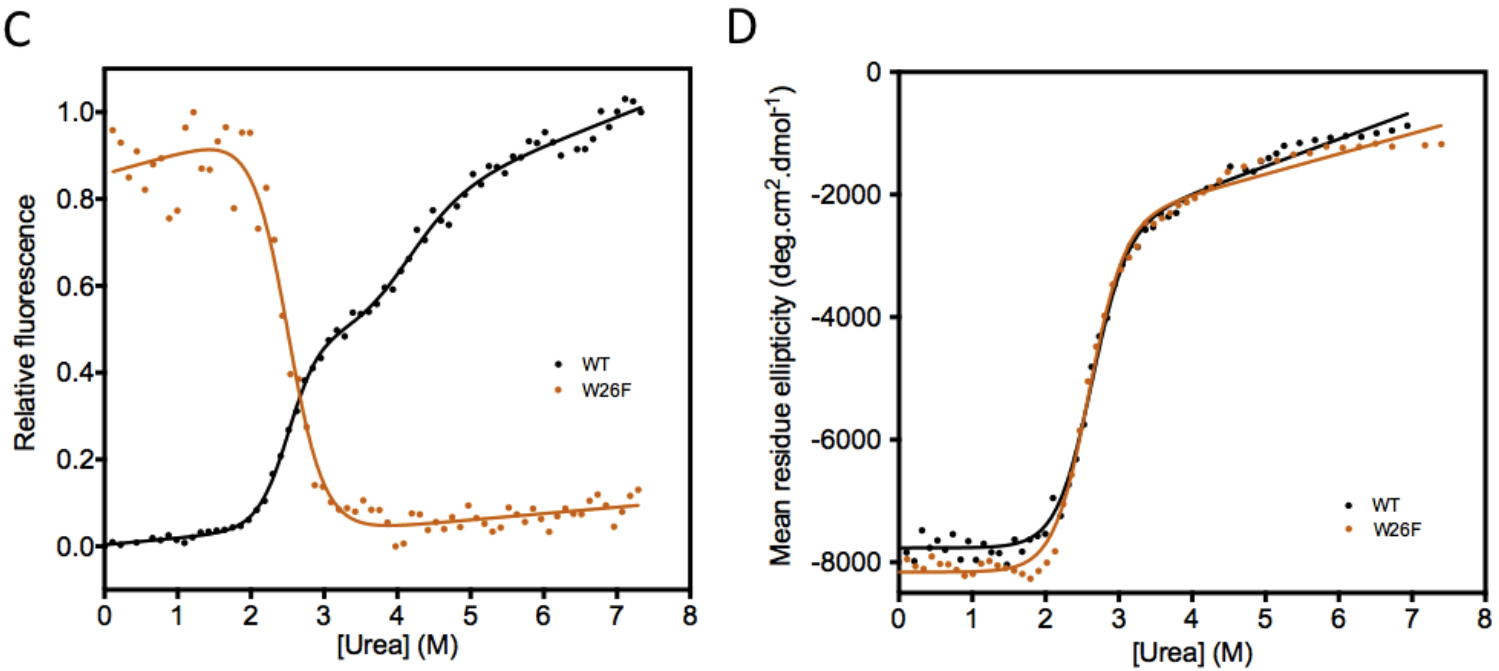

Figure S5 (continued): Structure and stability of W26F, a Trp-less mutant of UCH-L1. (C) Equilibrium unfolding experiments showing the normalized change in relative fluorescence between 361-365 nm for wild type (black-filled circles) and 302-306 nm for W26F (brown-filled circles). The solid lines represent the best fit of the wild-type data to a three-state model, Equation S1 or for W26F a two-state model, Equation S4 . (D) Equilibrium unfolding experiments showing the change in ellipticity for wild type (black-filled circles) and W26F (brown -filled circles). The solid lines represent the best fit of data to a two-state model, Equation S4. The thermodynamic parameters calculated from the best fits are shown in Table 1.

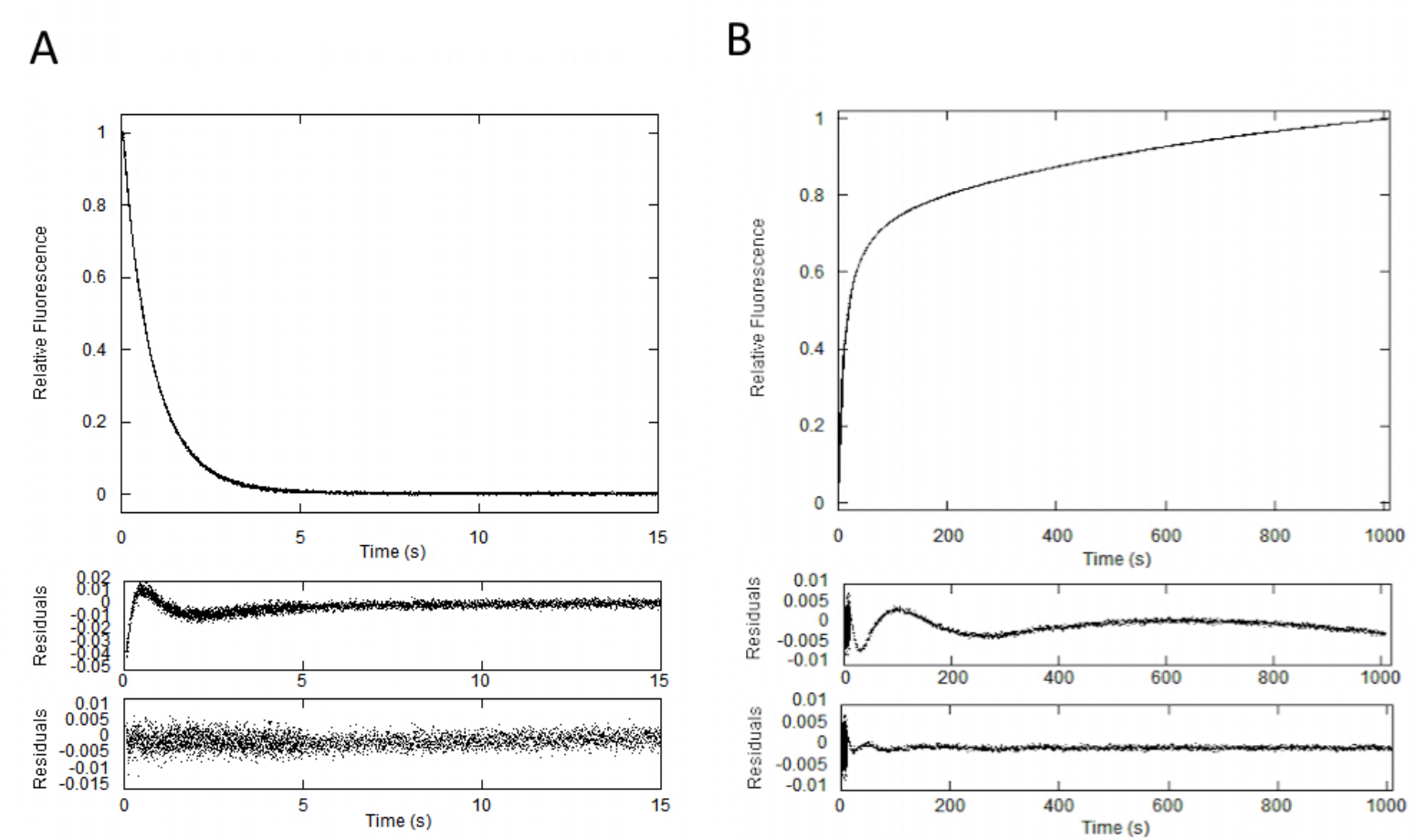

Figure S6: Typical kinetic traces for the unfolding and refolding of UCH-L1. (A) Unfolding trace (B) Refolding trace. The residuals show the best fit of the data to either (A) a single (upper) or double (lower) exponential function, or (B) a triple (upper) or quadruple (lower) exponential function..

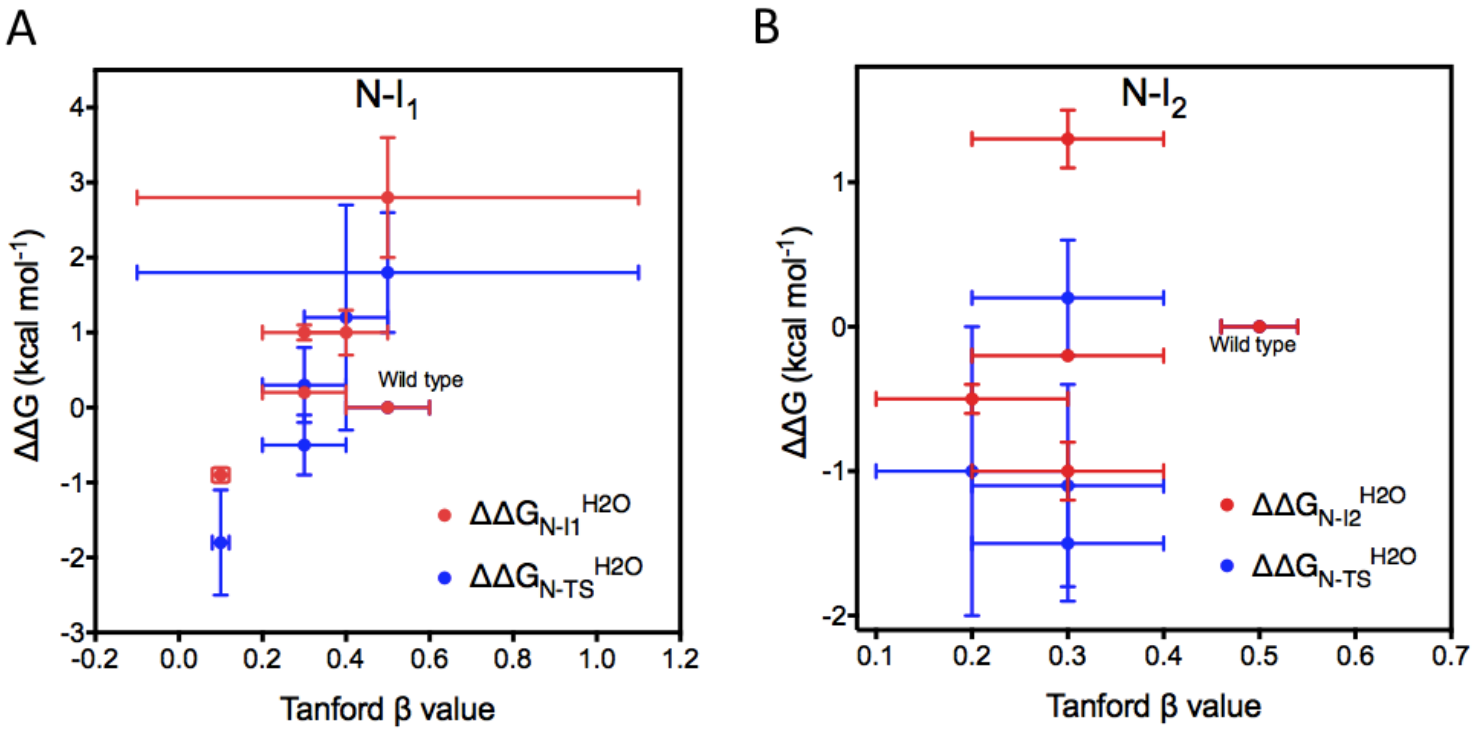

Figure S7: Correlation between the Tanford β value and the difference in change in free energy between either the native state and I_1 or I_2 (red- filled circles) or the native state and the transition state between N and I_1 or I_2 (blue-filled circles) .. Tanford β values for each mutant are taken from Table S6 which were calculated based on the data presented in Table 3 and Equation S12. (A) Shows the correlation for the transition between native state and I_1. (B) Shows the correlation between native state and I_2.

	Solvent accessible surface area of residue in a Gly-X-Gly tripeptide reference (Å^2)	Solvent accessible surface area from crystal structure (Å^2)	Ratio (%)	Location in the protein structure
Trp26	217	14.9	7	N-terminus of β-strand 1
Leu3	137	9.2	7	N-terminal coil region
Leu70	137	49.3	36	C-terminus of α-helix 2
Phe108	175	15.3	9	Loop
Val113	117	38.8	33	N-terminus of α-helix 4
Phe160	175	57.4	33	N-terminus of β-strand 3
Phe181	175	55.3	32	Loop

TABLE S1: The solvent-accessible surface area of residues mutated in this study. The ratios listed are the percentage between the solvent accessible surface area of the side chain in the protein structure and the theoretical solvent accessible area of the residue in a Gly-X-Gly sequence. The side chain of a residue is considered partially exposed if the ratio falls between 20-50% (30). Leu3 and Phe108 were also chosen for mutagenesis because: i) There is evidence that Leu3 is involved in some structure in the intermediate state from previous HDX-NMR studies (28), and ii) Phe108 is located in a loop, mutation to Trp is likely have minimal effect on the structure.

	Wild type	L3W	L70W	F108W	V113W	F160W	F181W
Rate (s^{-1})	0.025 ± 0.002	0.024 ± 0.001	0.022 ± 0.001	0.033 ± 0.002	0.027 ± 0.002	0.044 ± 0.002	0.030 ± 0.003

TABLE S2: Catalytic rates of the hydrolysis of Ubq-AMC by wild-type and mutant UCH-L1. Data were obtained from fitting the steady-state kinetics of the hydrolysis of the model substrate Ubq-AMC to a single-exponential equation. The errors shown are fitting errors from the program.

		WT	L3W	L70W	V113W	F160W	W26F	F108W	F181W
[a] N-I (Fluorescence)	$m_{I\text{-}N}$ (kcal mol^{-1}M^{-1})	2.8 ± 0.4 (shared)					2.4 ± 0.2 (shared)		
	$[\text{urea}]_{I\text{-}N}^{50\%}$ (M)	2.5 ± 0.1	2.4 ± 0.1	1.5 ± 0.1	1.7 ± 0.1	2.7 ± 0.1	2.5 ± 0.1	1.9 ± 0.1	2.5 ± 0.1
	[c] $\Delta G_{\text{I-N}}^{\text{H}_2\text{O}}$ (kcal mol^{-1})	7.0 ± 1.0	6.7 ± 1.0	4.2 ± 0.7	4.8 ± 0.7	7.8 ± 1.2	6.0 ± 0.5	4.6 ± 0.4	6.0 ± 0.6
[a] I-D (Fluorescence)	$m_{\text{D-I}}$ (kcal mol^{-1}M^{-1})	1.5 ± 0.2 (shared)					ND		
	$[\text{urea}]_{\text{D-I}}^{50\%}$ (M)	4.1 ± 0.2	3.7 ± 0.1	3.1 ± 0.6	2.7 ± 0.2	4.0 ± 0.5	ND	ND	ND
	[c] $\Delta G_{\text{D-I}}^{\text{H}_2\text{O}}$ (kcal mol^{-1})	6.2 ± 0.9	5.6 ± 0.8	4.7 ± 1.1	4.1 ± 0.6	6.0 ± 1.1	ND	ND	ND
[b] N-I (Far-UV CD)	$m_{\text{I-N}}$ (kcal mol^{-1}M^{-1})	2.4 ± 0.1 (shared)							
	$[\text{urea}]_{\text{I-N}}^{50\%}$ (M)	2.60 ± 0.02	2.60 ± 0.02	2.10 ± 0.02	1.80 ± 0.02	2.90 ± 0.02	2.60 ± 0.02	2.20 ± 0.02	2.50 ± 0.02
	[c] $\Delta G_{\text{I-N}}^{\text{H}_2\text{O}}$ (kcal mol^{-1})	6.2 ± 0.3	6.2 ± 0.3	5.0 ± 0.2	4.3 ± 0.2	7.0 ± 0.3	6.2 ± 0.3	5.3 ± 0.2	6.0 ± 0.3
Change in fluorescence (N-I)		86 ± 8	86 ± 13	3 ± 15	32 ± 17	339 ± 24	ND	ND	ND
Change in fluorescence (I-D)		9 ± 22	164 ± 21	91 ± 16	41 ± 19	165 ± 32	ND	ND	ND

TABLE S3: Thermodynamic parameters for the unfolding of wild-type and mutant UCH-L1 by urea under equilibrium conditions. The native, intermediate and denatured states are represented by N, I and D, respectively. Fitting errors are given..

a. Fluorescence data were fit to either a three- or two-state unfolding model using Equation S1 or S4, respectively.
b. Far-UV CD data were fit to a two-state unfolding model using Equation S4.
c. ΔG for each transition was calculated using Equation S2.

ND = not determined

F160W					
N-I (Fluorescence)	$m_{\text{I-N}}$ (kcal mol^{-1}M^{-1})	2.2 ± 0.1			Average
	$[urea]_{\text{I-N}}^{50\%}$ (M)	2.80 ± 0.03	2.80 ± 0.03	3.00 ± 0.02	2.9 ± 0.1
	$\Delta G_{\text{I-N}}^{H_2O}$ (kcal mol^{-1})	6.2 ± 0.3	6.2 ± 0.3	6.6 ± 0.3	6.3 ± 0.5
I-D (Fluorescence)	$m_{\text{D-I}}$ (kcal mol^{-1}M^{-1})	1.8 ± 0.5			
	$[urea]_{\text{D-I}}^{50\%}$ (M)	4.3 ± 0.3	4.2 ± 0.2	5.0 ± 2.2	4.5 ± 2.0
	$\Delta G_{\text{D-I}}^{H_2O}$ (kcal mol^{-1})	7.7 ± 2.2	7.6 ± 2.1	9.0 ± 4.7	8.1 ± 5.0
N-I	$m_{\text{I-N}}$ (kcal mol^{-1}M^{-1})	2.5 ± 0.1			

(Far-UV CD)	$[urea]_{I\text{-}N}{}^{50\%}$ (M)	2.90 ± 0.02	2.80 ± 0.02	2.80 ± 0.02	2.80 ± 0.03
	$\Delta G_{I\text{-}N}{}^{H_2O}$ (kcal mol^{-1})	7.3 ± 0.3	7.0 ± 0.3	7.0 ± 0.3	7.1 ±0.5

TABLE S4: Thermodynamic parameters of three independent measurements of the urea-induced unfolding of the mutant F160W under equilibrium conditions. All the errors shown are fitting errors except the average where the error shown is the standard deviation from the mean. F160W was chosen as we had lots of purified protein available to us in this case for repeated measurements.

Mutant	[a]$\Delta\Delta G_{kin}^{H_2O}$ (N-I_1) (kcal mol^{-1})	[a]$\Delta\Delta G_{kin}^{H_2O}$ (N-I_2) (kcal mol^{-1})	[b]$\Delta\Delta G_{D\text{-}N}^{H_2O}$ (kcal mol^{-1})
L3W	-1.0 ± 0.7	ND	0.9 ± 0.5
L70W	2.7 ± 0.9	-0.2 ± 0.6	4.3 ± 1.1
F160W	0.2 ± 0.5	-0.5 ± 1.0	-0.7 ± 0.9

a) The difference in free energies for N-I_1 or N-I_2 transition based on kinetic measurements were calculated using Equation S5. A value close to zero indicates that there is no difference in the change in free energy between native and intermediate state (I_1 or I_2) on mutation. Therefore, it is unlikely that the intermediate state is not structured in the vicinity of the mutation.

b) $\Delta\Delta G_{D\text{-}N}^{H_2O} = \Delta\Delta G_{I\text{-}N}^{H_2O} + \Delta\Delta G_{D\text{-}I}^{H_2O}$ based on equilibrium unfolding data.
ND = not determined.

Table S5: Difference in the change free energy between the native and intermediate states for wild type and mutants based on kinetic measurements.

	[a]$\Delta\Delta G_{I\text{-}N}^{H_2O}$ (kcal mol^{-1})	[a]$\Delta\Delta G_{D\text{-}I}^{H_2O}$ (kcal mol^{-1})	[b]$\Delta\Delta G_{D\text{-}N}^{H_2O}$ (kcal mol^{-1})	[c]Oligomer formation propensity of mutants compared to wild type
L3W	0.3 ± 0.4	0.6 ± 0.3	0.9 ± 0.5	similar
L70W	2.8 ± 0.6	1.5 ± 1.0	4.3 ± 1.1	+
F108W	2.4 ± 1.1	ND	ND	++
V113W	2.2 ± 0.5	2.1 ± 0.5	4.3 ± 0.7	+++
F160W	-0.8 ± 0.4	0.2 ± 0.8	-0.6 ± 0.9	similar
F181W	1.1 ± 1.0	ND	ND	similar

TABLE S6: Conformational stability and oligomerization propensity of UCH-L1 mutants compared to wild type.

(a) Calculated using Equation S5.

(b) $\Delta\Delta G_{D\text{-}N}^{H_2O} = \Delta\Delta G_{I\text{-}N}^{H_2O} + \Delta\Delta G_{D\text{-}I}^{H_2O}$

(c) An increased propensity to aggregate relative to wild-type UCH-L1 is shown by +, ++, +++ (the highest aggregation propensity observed).

All errors are compound errors using the fitting errors of the parameters used, shown in Table 1.

	Variant	[a]Tanford β value	[b]$\Delta\Delta G_{I\text{-}N}^{H2O}$ (kcal mol^{-1})	[b]$\Delta\Delta G_{N\text{-}TS}^{H2O}$ (kcal mol^{-1})
N-I_1	L3W	0.10 ± 0.02	-0.9 ± 0.1	-1.8 ± 0.7
	L70W	0.5 ± 0.6	2.8 ± 0.8	1.8 ± 0.8
	F108W	0.3 ± 0.1	1.0 ± 0.1	0.3 ± 0.5
	F160W	0.3 ± 0.1	0.2 ± 0.02	-0.5 ± 0.4
	F181W	0.4 ± 0.1	1.0 ± 0.3	1.2 ± 1.5
N-I_2	L3W	ND	ND	ND
	L70W	0.3 ± 0.1	-0.2 ± 0.03	-1.5 ± 0.4
	F108W	0.3 ± 0.1	1.3 ± 0.2	0.2 ± 0.4
	F160W	0.2 ± 0.1	-0.5 ± 0.1	-1.0 ± 1.0
	F181W	0.3 ± 0.1	-1.0 ± 0.2	-1.1 ± 0.7

Table S7: Tanford β value and the difference in the free energy change between different states on the folding pathway.

a) Tanford β value is obtained using Equation S12 and the values for wild type for N-I_1 and N-I_2 are 0.5 ± 0.1 and 0.50 ± 0.04, respectively.

b) The difference in the change in free energy between the native and intermediate (I_1 and I_2) states N-I and between the native state and the transition state for the transition between native and intermediate states, N-TS, based on the kinetic data were calculated using a modified Equation S5.

TS = transition state

ND = not determined