

## Accepted Manuscript

The knotted protein UCH-L1 exhibits partially unfolded forms under native conditions that share common structural features with its kinetic folding intermediates

Shih-Chi Lou, Svava Wetzel, Hongyu Zhang, Elizabeth W. Crone, Yun-Tzai Lee, Sophie E. Jackson, Shang-Te Danny Hsu

PII: S0022-2836(16)30045-6  
DOI: doi: [10.1016/j.jmb.2016.04.002](https://doi.org/10.1016/j.jmb.2016.04.002)  
Reference: YJMBI 65052

To appear in: *Journal of Molecular Biology*

Received date: 21 December 2015  
Revised date: 22 March 2016  
Accepted date: 2 April 2016



Please cite this article as: Lou, S.-C., Wetzel, S., Zhang, H., Crone, E.W., Lee, Y.-T., Jackson, S.E. & Hsu, S.-T.D., 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* (2016), doi: [10.1016/j.jmb.2016.04.002](https://doi.org/10.1016/j.jmb.2016.04.002)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**The knotted protein UCH-L1 exhibits partially unfolded forms  
under native conditions that share common structural features  
with its kinetic folding intermediates**

Shih-Chi Lou,<sup>1,2</sup> Svava Wetzel,<sup>1</sup> Hongyu Zhang,<sup>1</sup> Elizabeth W. Crone,<sup>1</sup> Yun-Tzai  
Lee,<sup>2,3</sup> Sophie E. Jackson<sup>1,\*</sup> and Shang-Te Danny Hsu<sup>1,2,3,\*</sup>

1. Department of Chemistry, University of Cambridge, Lensfield Road,  
Cambridge CB2 1EW, United Kingdom
2. Institute of Biological Chemistry, Academia Sinica, 128, Section 2,  
Academia Road, Taipei 11529, Taiwan
3. Institute of Biochemical Sciences, National Taiwan University, 1, Section 4,  
Roosevelt Road, Taipei 106, Taiwan

\* Corresponding authors: sej13@cam.ac.uk and sthsu@gate.sinica.edu.tw

**ABSTRACT**

Human ubiquitin C-terminal hydrolase, UCH-L1, is an abundant neuronal deubiquitinase that is associated with Parkinson's disease. It contains a complex Gordian knot topology formed by the polypeptide chain alone. Using a combination of fluorescence-based kinetic measurements, we show that UCH-L1 has two distinct kinetic folding intermediates which are transiently populated on parallel pathways between the denatured and native states. NMR hydrogen-deuterium exchange (HDX) experiments indicate the presence of partially unfolded forms (PUFs) of UCH-L1 under native conditions. HDX measurements as a function of urea concentration were used to establish the structure of the PUFs and pulsed-label HDX NMR was used to show that the PUFs and the folding intermediates are likely the same species. In both cases, a similar stable core encompassing most of the central  $\beta$ -sheet is highly structured and  $\alpha$ -helix 3, which is partially formed, packs against it. In contrast to the stable  $\beta$ -sheet core, the peripheral  $\alpha$ -helices display significant local fluctuations leading to rapid exchange. The results also suggest that the main difference between the two kinetic intermediates is structure and packing of  $\alpha$ -helices 3 and 7 and the degree of structure in  $\beta$ -strand 5. Together, the fluorescence and NMR results establish that UCH-L1 neither folds through a continuum of pathways nor by a single discrete pathway. Its folding is complex, the  $\beta$ -sheet core forms early and is present in both intermediate states, and the rate-limiting step which is likely to involve the threading of the chain to form the  $5_2$ -knot occurs late on the folding pathway.

**Keywords:**

Knotted protein; Ubiquitin C-terminal hydrolase; Folding pathway; Partially unfolded form; Folding intermediate; NMR; Hydrogen-deuterium exchange

## INTRODUCTION

Recently, systematic surveys have identified hundreds of topologically knotted proteins in the protein structure databank [1-3]. Relatively little is known about how these proteins attain their topologically knotted conformations and the functions of these topological knots [4-6]. To date, only a handful of knotted proteins have been subject to experimental characterisation with regard to their unfolding/folding equilibria and kinetics: YibK from *Haemophilus influenzae* and YbeA from *Escherichia coli* are the best studied [7-16]. Both proteins are homodimeric and contain a deep trefoil ( $3_1$ ) knot, which is formed by the threading of the C-terminus of the chain through an intramolecular loop. It is remarkable that YibK and YbeA maintain a knotted topology even under chemically denaturing conditions [15] while being random coil-like as revealed by NMR spectroscopy [17, 18] and small angle X-ray scattering [19]. Other examples include MJ0366 from *Methanocaldococcus jannaschii*, the smallest trefoil-knotted protein [20], human ubiquitin C-terminal hydrolase 1 and 3 (UCH-L1 and UCH-L3), which contain a more complex Gordian ( $5_2$ ) knot [21-23]. A tandem repeat of the dimeric hypothetical protein HP0242 from *Helicobacter pylori* has also been designed and found to fold spontaneously from a chemically denatured state into a trefoil knotted structure [24]. In this case, the designed knotted protein exhibits complex folding pathways with off-pathway misfolded intermediate states [25].

A number of computational approaches have been used to obtain insights into the folding pathways of a number of topologically knotted proteins [26-32]. Although various models have been proposed as to how a polypeptide chain can thread itself into a knotted conformation, many of these simulations required *ad*

*hoc* modifications to the potentials in order to successfully fold the proteins into their native knotted states on the timescales used. In light of the emerging interest in understanding the mechanisms by which proteins attain topologically knotted conformations, it is important to experimentally characterise in detail the thermodynamics and kinetics of the folding processes of such proteins. To this end, it is not only important to obtain high-resolution information regarding the structure and dynamics of the native states of knotted proteins but also other states along the folding pathways.

In addition to the fundamental interest in understanding the folding equilibria and kinetics of UCH-L1 as a topologically knotted protein, it is also of biomedical relevance as it has been implicated in the pathogenesis of Parkinson's disease (PD), Alzheimer's disease and other neurodegenerative diseases [33-35]. In addition to PD-associated familial mutations, UCH-L1 has been reported to be post-translationally modified by mono-ubiquitination, farnesylation, glycosylation and other forms of post-translational modifications [36]. In vivo, it is very likely that most of the post-translational modifications occur after the unmodified proteins has first folded into its native knotted state. We have previously investigated the folding equilibria and unfolding kinetics of UCH-L1 and its PD-associated mutants, I93M and S18Y, using intrinsic fluorescence and far-UV circular dichroism (CD) spectroscopy in conjunction with solution-state nuclear magnetic resonance (NMR) spectroscopy [23]. This study identified an intermediate state populated during the unfolding of UCH-L1 under equilibrium conditions. The PD-associated mutants were shown to exhibit lower thermodynamic stabilities and higher unfolding rates compared with wild type. A recent study using methyl NMR spectroscopy and residual dipolar coupling

analysis further revealed that the I93M mutation induces significant structural perturbations within the hydrophobic core of UCH-L1 [37], which was not observed in the crystal structure [38, 39].

Here, we report more detailed kinetic and structural investigations into the folding pathway of human UCH-L1 using a combination of stopped-flow (SF) intrinsic fluorescence measurements and NMR hydrogen/deuterium exchange (HDX) analysis. Single-jump (SJ) and double-jump (DJ) SF fluorescence spectroscopy is used here to establish that UCH-L1 folds into its native conformation along two parallel folding pathways on which two distinct folding intermediates are transiently populated. A series of NMR HDX experiments have been carried out in the presence of various concentrations of urea in order to monitor, at the level of individual residues, the free energy of unfolding as a function of denaturant concentration to provide insight into the structure of intermediate states. In addition, pulsed-label NMR HDX exchange experiments have been employed to probe the structure of UCH-L1 in its non-native states. The results indicate that the UCH-L1 frequently samples partially unfolded forms (PUFs) even under native conditions and that the central  $\beta$ -sheet in the hydrophobic core of UCH-L1 remains structured in the intermediate states.

## EXPERIMENTAL SECTION

### Purification of recombinant UCH-L1

The pGEX-6-*uch-l1* construct for UCH-L1 was a kind gift from Dr Chitta Das (Duke University, USA). Wild-type UCH-L1 was over-expressed in *E. coli* and purified following a published protocol [23]. Briefly, cell pellets from LB culture or  $^{15}\text{N}$ -

labelled minimum media were resuspended thoroughly in 1x PBS, 400 mM KCl (buffer A) with DNase I. The cells were lysed using a sonicator, and the cell debris was removed by centrifugation at 18,000 rpm, 4 °C, for 45 minutes. The supernatant was passed through a 0.22 µm filter and loaded on to three sequentially connected GSTrap™ FF 5 mL columns which were pre-equilibrated in buffer A at a flow rate of 1 mL min<sup>-1</sup> using an AKTA purification system. The columns were then washed with 100 mL buffer A at a flow rate of 1 mL min<sup>-1</sup>, before eluting the recombinant protein with 10 mM reduced glutathione in buffer A at a flow rate of 2 mL min<sup>-1</sup>. Fractions were collected throughout the elution step, and those that contain the fusion protein were pooled for further purification. The fusion protein was incubated with 133 units of PreScission protease overnight in buffer A and 1 mM DTT. GST and any uncleaved fusion protein were removed by a second binding step to the GSTrap™ FF 5 mL columns (which were pre-equilibrated in buffer A at a flow rate of 1 mL min<sup>-1</sup> and washed with buffer A). The flow-through solution was collected and fractionated during both this step and the subsequent washing of the column using buffer A. Fractions with high A<sub>280</sub> absorbance were pooled and concentrated using a Vivaspin 20 (MWCO 5000 Da) concentrator to less than 4 mL. The concentrated protein was further purified using a gel-filtration column (Superdex 75, 26/60, GE Life Science, USA) in 50 mM Tris, 5 mM DTT pH 7.6, 0.5 mM EDTA (buffer B). The purity of fractions containing UCH-L1 eluted from the column was assessed using SDS-PAGE and those that contained pure UCH-L1 were pooled and concentrated using Vivaspin 20 (MWCO 5000 Da) to a final concentration of around 200 µM. The purity of UCH-L1 was over 95 % as judged by SDS-PAGE. The pure protein was aliquoted, flash frozen in liquid nitrogen and stored at -80



°C for further use.

### Single-jump stopped-flow experiments

UCH-L1 was prepared in native buffer (50 mM Tris (pH 7.6), 0.5 mM EDTA and 5 mM DTT) and denatured buffer (the same as native plus 8.5 M urea) with a final protein concentration of 20  $\mu$ M. The native and 8.5 M urea-denatured UCH-L1 were used for unfolding and refolding kinetic measurements using rapid 1:10 dilutions in to various urea concentrations to give final concentrations between 0.5 to 7.5 M depending on whether it was an unfolding (high final concentrations) or refolding (low final concentrations) experiment. Most of the kinetic traces were collected using a SF fluorimeter (Applied Photophysics, UK). For the very slow kinetic phases, manual mixing was done and the reaction monitored by fluorescence measurements performed on a Cary fluorimeter (Varian, USA) or far-UV CD measurements on a ChiralScan CD spectrometer (Applied Photophysics, UK). The reaction rate constants were extracted by non-linear regression of the kinetic traces to a multiple-exponential function. The logarithm of the rate constants were plotted as a function of urea concentration to yield a chevron plot.

For the two slow unfolding and refolding phases, which are attributed to a transition between native and intermediates states, the observed rate constants,  $k_{\text{obs}}$ , for unfolding and folding phases were fit to a two-state kinetic folding model using Equation 1:

$$1 - \frac{k_{\text{U}}}{k_{\text{U}} + k_{\text{R}}} = \frac{k_{\text{U}}}{k_{\text{U}} + k_{\text{R}}} \quad \text{Equation 1}$$

where  $k_f$  and  $k_u$  are the folding and unfolding rate constants, respectively,  $m_f$  and  $m_u$  are constants which are a measure of the degree to which the rate constants change with urea concentration, *i.e.*, the slopes of the folding and unfolding arms of the chevron plots, respectively, and  $k_f^{H_2O}$  and  $k_u^{H_2O}$  the folding rate and unfolding rate constants in the absence of denaturant.  $R$  and  $T$  are the gas constant and sample temperature, respectively. The non-linear regression was achieved using the software package Kaleidagraph (Synergy software) or PRISM (GraphPad, Inc.)

### **Interrupted-refolding experiments using stopped-flow fluorescence**

UCH-L1 (25  $\mu$ M) denatured in 8.5 M urea was rapidly diluted with native buffer (50 mM Tris (pH 7.6), 0.5 mM EDTA and 5 mM DTT) in a 1:5 ratio such that the urea concentration was 1.4 M to initiate refolding. Refolding was allowed to continue for various aging times,  $t_{age}$ , which varied between 0.05 and 500 s. Unfolding was subsequently triggered by a second rapid dilution step into high concentrations of urea using a mixing ratio of 1:10. Final urea concentrations varied between 3.3 and 5.4 M. After the two mixing steps, the unfolding kinetics were monitored at a final protein concentration of 2  $\mu$ M. The DJ-SF fluorescence experiments were performed using a stopped-flow spectrometer (Applied Photophysics, UK). The unfolding traces at different aging times were globally fit to a quadruple exponential function, where the unfolding rate constants were shared across all  $t_{age}$ , whilst the amplitude of each phase was allowed to vary.

### Steady-state NMR HDX experiments

100  $\mu$ M  $^{15}\text{N}$ -labelled UCH-L1 in 50 mM Tris (pH 7.6), 0.5 mM EDTA and 5 mM DTT in  $\text{H}_2\text{O}$  was rapidly buffer-exchanged into urea solutions containing the same buffer but in  $\text{D}_2\text{O}$ . A NAP-5 desalting column (GE Life Science, USA) was used for the rapid buffer exchange. The protein sample was then concentrated using an Amicon concentrator (MWCO 5000 Da; Millipore, USA) at 4  $^{\circ}\text{C}$  at 3,000 rpm to reduce the sample volumes to *ca.* 0.5 mL. These steps took about 50 minutes before the first  $^{15}\text{N}$ - $^1\text{H}$  heteronuclear single-quantum coherence (HSQC) spectrum was acquired. Buffer exchange using the NAP-5 column was necessary to prevent the aggregation of UCH-L1, which was observed when the protein was lyophilised in the presence of urea. All experiments were performed at 25  $^{\circ}\text{C}$  and the samples were stored at 25  $^{\circ}\text{C}$  between measurements. All NMR spectra were acquired on a Bruker AVANCE 500 MHz spectrometer and a Bruker AVNCE 600MHz spectrometer, both equipped with a 5 mm  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  triple-resonance cryogenic probe head. All HSQC spectra were acquired with 1024 ( $t_2$ ) and 128 ( $t_1$ ) complex points and processed by NMRPipe [40] and analyzed using Sparky (Goddard et al., University of California, San Francisco, USA). The HDX rate constants of individual residues were derived by fitting the observed peak intensities as a function of exchange time using a single-exponential decay function using PRISM.

### Pulsed-labelled NMR HDX analysis

To carry out the pulsed-label NMR HDX experiments, native UCH-L1 was quickly exchanged to a  $\text{D}_2\text{O}$ -based urea-containing buffer. Denatured samples were left at 4 or 25  $^{\circ}\text{C}$  for 5, 25, or 40 minutes and then refolded in deuterated buffer

using the NAP-5 column to remove urea followed by NMR data acquisition. In order to ensure that refolding was complete and successful, a one-dimensional proton spectrum of the native state of an untreated UCH-L1 was used as a reference. As it took some time to prepare and transfer the samples into the NMR tubes and to reach temperature equilibrium before NMR data collection, and there was no major differences in the resulting NMR spectra recorded using different parameters, the pulsed-label HDX were performed at 25 °C for 25 minutes to minimize sample variations between different runs.

## RESULTS

Unfolding and refolding experiments using a SF instrument to attain rapid mixing and intrinsic protein fluorescence (UCH-L1 contains only one tryptophan, W26) to monitor conformational changes in the protein structure, were used to characterise the unfolding and refolding kinetics of UCH-L1. Initially, SJ experiments, in which a single, rapid dilution step was employed, were undertaken. In these experiments, native protein was rapidly jumped into highly denaturing conditions to monitor unfolding, or denatured protein was rapidly diluted into native conditions to induce refolding. UCH-L1 exhibits four unfolding phases (the two slower phases 1 and 2, were observed in SJ experiments, and the two additional unfolding phases 3 and 4 were only observed in DJ experiments) and five refolding phases (Figure 1). DJ experiments in GdmCl were used to establish that the slowest phase s (open diamond in Figure 1) may be associated with a slow conformational change that takes place in the denatured state after unfolding, consistent with a process limited by proline isomerisation, considering that UCH-L1 has nine prolines and one of which is in a cis

conformation, P45 (data not shown). This is consistent with previous results obtained for the folding of the structural homologue, UCH-L3 [21], and this phase is not discussed further here. The two unfolding phases and the four refolding phases observed in SJ experiments are numbered as 1-2 or 1-4 from slowest to fastest, respectively and their rate constants were determined over a wide range of urea concentrations (Figure 1).

The observation that there are two slow unfolding phases but four refolding phases, two of which are fast, suggests that at least two intermediates are present on the folding pathway. The fact that the corresponding fast phases are not observed in the SJ unfolding experiments suggests that the rate-limiting step in unfolding is the transition from native (N) to intermediate (I) state. Although the unfolding of intermediate to denatured (D) states cannot be detected in the SJ experiments, DJ interrupted refolding experiments were used to probe these transitions.

### **Double-jump interrupted-refolding stopped-flow experiments**

DJ interrupted-refolding experiments were performed by mixing 8.5 M urea-denatured UCH-L1 with refolding buffer (1.4 M final urea concentration after the first mixing step) to initiate folding. The sample was allowed to refold over a range of timescales, 0.05-100 seconds of aging time,  $t_{age}$ , before it was subsequently mixed with buffer containing high concentrations of urea to trigger unfolding. The final urea concentrations ranged from 3 to 5.5 M. Two additional fast unfolding phases were observed in these DJ experiments (Figure 1). The rate constants for the unfolding for these two new phases,  $k_3$  and  $k_4$ , were measured over a range of denaturant concentrations (Figure 1). In contrast to

the two slow unfolding phases, the rate constants which were approximately 1 and 10 s<sup>-1</sup> respectively, did not vary greatly with denaturant concentration, however, there are relatively large fluctuations in the measured values for these fast phases, which we attribute to the small change in fluorescence associated with these transitions (I  $\leftrightarrow$  D). Thus, the interrupted-refolding measurements identified two hidden kinetic phases that were not detectable in the SJ unfolding experiments, suggesting that these phases are associated with the unfolding of intermediates, which are transiently populated on the folding pathway.

By following the changes of amplitudes associated with the four observed unfolding phases as a function of aging time,  $t_{age}$ , one can establish the relationship between the unfolding and refolding phases and hence delineate the kinetic folding pathway of UCH-L1 [8, 10, 21]. Essentially, monitoring the amplitudes of the unfolding phases with  $t_{age}$  allows the kinetic rate constants for the formation of the species, giving rise to each unfolding phase to be measured and these can then be compared to the results of the SJ refolding experiments. The global-fitting of the resulting transients at 4.9 M urea yields four apparent reaction rates – 0.045, 0.16, 0.46 and 12.53 s<sup>-1</sup> for  $k_1$  to  $k_4$ , respectively – two of which, namely  $k_1$  and  $k_2$ , match very well with SJ-derived reaction rates (Figure 1A). The amplitudes of the four unfolding phases as a function of aging time are shown in Figure 1B. The two fast unfolding phases exhibit very small amplitudes ( $A_3$  and  $A_4$ ), which rise rapidly and then decay with aging time, whereas the amplitude of the two slow unfolding phases ( $A_1$  and  $A_2$ ) increase exponentially with rate constants of 0.014 and 0.1 s<sup>-1</sup>, for  $k_1$  and  $k_2$ , respectively.

The analysis of the amplitudes of the four unfolding phases with aging time also allows us to associate specific unfolding phases with their counterpart

refolding phases. For example, the amplitude associated with the slowest unfolding phase ( $A_1$ ) evolves with aging time with a rate constant of  $0.014 \text{ s}^{-1}$ . This corresponds to the rate constant of the slowest, non-proline isomerisation, refolding phase at 1.4 M urea in the SJ refolding experiments (Figure 1). Thus, these two phases represent the forward and backward transition between the same two states, in this case,  $I'$  and  $N$ . These two phases are therefore both coloured blue in Figure 1. Likewise, the amplitude associated with the second slowest unfolding phase ( $A_2$ ) evolves with aging time with a rate constant of  $0.1 \text{ s}^{-1}$  which corresponds with the second slowest, non-proline isomerisation refolding phase,  $k_2$ , seen in the SJ experiments (Figure 1). Thus, these two phases represent the forward and back reaction between  $I''$  and  $N$ .

For the faster phases, it is not possible to link the unfolding and refolding arms directly in this manner as the rates for the evolution of the two fast unfolding phases are too fast to measure with the SF instrument in DJ mode where the deadtime is  $> 50 \text{ ms}$ . However, as it is highly likely that the fastest unfolding and refolding phases are associated with the same transition, and likewise for the next fastest unfolding and folding phases, they are coloured accordingly (Figure 1). Thus, the two fastest unfolding and refolding phases correspond to the forward and back reactions of the transitions  $I'-D$  and  $I''-D$ , whilst the two slower unfolding and refolding phases correspond to the transitions between  $I'$  and  $I''$  and  $N$ .

The kinetic data for the  $I \leftrightarrow N$  reactions can be fit to a simple, two-state kinetic model to obtain the unfolding and refolding rate constants in water,  $k_u^{H_2O}$  and  $k_f^{H_2O}$ , for the two transitions, as well as the  $m_{kf}$  and  $m_{ku}$  values, which determine how the rate constants for the phases vary with denaturant

concentration. For both the  $I' \leftrightarrow N$  and  $I'' \leftrightarrow N$  transitions, these kinetic parameters can be used to calculate the thermodynamic parameters associated with each reaction (Table 1), and these can be compared with those obtained from the equilibrium unfolding experiments [41]. Based on the data, we propose that UCH-L1 exhibits two parallel folding pathways with two transiently populated folding intermediates,  $I'$  and  $I''$ , whose unfolding rate constants can be measured in the DJ interrupted-refolding experiments (Figure 1C).

### Urea-dependent NMR HDX experiments

To gain further insights into the unfolding of UCH-L1 by urea, a series of equilibrium NMR HDX experiments were performed in the absence and presence of urea. Concentrations ranged from 0.5 to 2.25 M urea, the latter approaching the value of urea where the transition between native and intermediate states occurs based on the intrinsic fluorescence measurements ( $[D]_{5\%}^{I-N} \sim 3$  M) [23]. The HDX rate constants ( $k_{ex}$ ) of about 50 residues were followed as a function of urea concentration (Figure 2). At low urea concentrations, the observed HDX rate constants (measured in units of  $\text{min}^{-1}$ ) span over two orders of magnitude ( $-2.5 > \log k_{ex} > -4$ ) and converge at 2 M urea where  $\log k_{ex} = -2.25 \pm 0.25$ . The large dispersion of the observed HDX rate constants at low urea concentrations is the result of having different local fluctuations and sub-domain unfolding events in addition to global unfolding which give rise to exchange competent states [42-45]. As the denaturant concentration increases,  $k_{op}$  which corresponds to the unfolding rate constant for the native state unfolding to  $I'$  or  $I''$  increases and, at the same time,  $k_{cl}$ , the folding rate decreases. Consequently, amide protons that exchange in the EX2 regime in 0 M urea, switch to an EX1 mechanism in which



the dominating term in the exchange reaction is  $k_{op}$ . This gives rise to the convergence of the rate constants at approximately 2 M urea, as observed (Figure 2).

Englander and co-workers have shown that by following the residue-specific free energy of unfolding as a function of denaturant concentration, one can identify clusters of residues, or foldons, that unfold/fold cooperatively and become transiently unfolded in PUFs [44]. A prerequisite of such analysis however, is that the residues of interest should undergo HDX in rapid equilibrium between folded and unfolded or partially unfolded states, *i.e.*, in the EX2 regime where the rate of HDX is proportional to the concentration of the catalyst; that is,  $[OH^-]$  [46]. The HDX rate constants of UCH-L1 were determined at pH 7.6 and pH 8.6 and compared: these results show that 21 residues can be broadly considered to be in the EX2 regime ( $PF_{pH8.6} - PF_{pH7.6} < 0.3$  and denoted hereafter as EX2 residues). In these cases, the amide groups exchange about ten times faster at pH 8.6 than at pH 7.6. In contrast, 22 residues undergo HDX at rates that are independent of the pH, *i.e.*, in the EX1 regime (defined as  $PF_{pH8.6} - PF_{pH7.6} > 0.75$  and denoted hereafter as EX1 residues), and eight residues exchange between the EX1 and EX2 regimes (Figure 3). The reason why residues in the central  $\beta$ -sheet region are in EX1 not EX2, which might be expected for the most stable part of the structure, is that for UCH-L1 the folding rates constants ( $k_{cl}$ ) are small (the protein is slow to fold as with many larger proteins with complex structures) whilst the NMR experiments are performed under conditions at higher pH where intrinsic exchange rates are fast.

Structural mapping of the observed HDX rate constants shows that most of the EX1 residues reside in the central  $\beta$ -sheet ( $\beta 2$ - $\beta 6$ ) and the C-terminal half

of  $\alpha$ -helix 3 ( $\alpha 3$ ), which is part of the hydrophobic core (Figure 3). L17 in  $\alpha 1$ , which is in contact with  $\alpha 3$ , is also an EX1 residue. Except for a few isolated residues in  $\alpha 1$  (L20),  $\alpha 2$  (E60),  $\alpha 3$  (G94),  $\alpha 5$  (R129) and  $\alpha 7$  (H97 and V200), and  $\beta 1$  (V22, W26 and R27),  $\beta 2$  (A48) and  $\beta 3$  (V168), most of the EX2 residues are generally flanking residues of secondary structure elements, which are less restricted by the hydrogen-bonding network (Table S1). The majority of the slow-exchanging residues are either EX1 or in between EX1 and EX2, rendering it difficult to translate the HDX rates into free energy of unfolding for individual backbone amide-mediated hydrogen bonds.

#### **[Urea]-dependent free energies of unfolding for individual residues**

For the EX2 residues, the HDX rate constants were converted into free energies of unfolding,  $\Delta G_{\text{HDX}}$ , as a function of urea concentration and linear regression analysis was used to extract their corresponding  $m$ -values. This analysis assumes that the free energy of unfolding depends linearly on denaturant concentration,  $\Delta G_{\text{HDX}} = \Delta G_{\text{HDX}}^{\text{H}_2\text{O}} - m \cdot [\text{urea}]$  (Figure 4A). The intercepts at 0 M urea correspond to the free energies of unfolding from the native state to an exchange-competent PUF,  $\Delta G_{\text{HDX}}^{\text{H}_2\text{O}}$ , to which the residue undergoing exchange belongs [43, 44]. Cluster analysis has been used previously to identify residues that exhibit similar  $\Delta G_{\text{HDX}}^{\text{H}_2\text{O}}$  and  $m$ -values, that are close in proximity in the protein structure, in order to identify cooperatively unfolding units, foldons [43, 44]. Such an analysis was performed here for UCH-L1, however, only 17 out of the 21 EX2 residues were analysed as the signal-to-noise ratio for many fast-exchanging residues, e.g., V22, Y80, H171 and D176, was too low in the presence

of urea to extract reliable HDX rate constants. The free energy of unfolding of these EX2 residues ranges between 5 and 10 kcal mol<sup>-1</sup> with *m*-values between 0 and -1.5 kcal mol<sup>-1</sup> M<sup>-1</sup> (Figure 4B). Although several clusters can be defined based on their HDX behaviour, the residues are not close in the three-dimensional structure as one would expect for cooperative unfolding/folding units.

### Use of NMR HDX to investigate the unfolding kinetics of UCH-L1

The pH-dependent NMR HDX analysis reveals two groups of EX1 residues with HDX rate constants differing by almost one order of magnitude (Figure 3). The pH-independent HDX rate constants correspond to the opening rate,  $k_{op}$ , of the amide group undergoing exchange, which is the kinetically controlled rate-limiting step for HDX under these conditions. This corresponds to the global unfolding of the protein from the native to an exchange-competent state and not due to a local unfolding event [44, 47]. For the first group of EX1 residues, I96, A98, H185 and D196, their opening rates in the absence of urea  $k_e^{HO} \rightleftharpoons k_o^{HC}$  are in good agreement with the slowest unfolding rate constant  $k_u^{H_2O}$  observed in the SJ unfolding experiment (Figure 1). These results indicate that the NMR HDX-derived opening rates indeed report on the N→I' unfolding transition. Rather unexpected, however, is the additional group of EX1 residues which exhibit opening rate constants that are almost ten times slower than this value. These EX1 residues also exhibit higher *m*-values, *i.e.*, the linear dependency of the observed exchange rates with respect to urea concentration in the semi-logarithmic plot (Figure 1).

### Pulsed-label NMR HDX analysis

In order to identify the structure of the intermediate states of UCH-L1 that are transiently populated during unfolding and refolding, pulsed-label NMR HDX experiments were carried out [48]. In these experiments, UCH-L1 was rapidly buffer exchanged into deuterated buffer (either in the presence or absence of urea) to allow HDX under native and denaturing conditions. In the latter case, after 25 minutes at 25 °C, a second buffer exchange step was used to remove the urea and return UCH-L1 to native conditions while keeping the protein in a deuterated buffer. Intriguingly, the pulsed-label  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum of UCH-L1 acquired after HDX in the presence of urea (3.3 M) looks virtually identical to the one that has undergone the same pulsed-labelling steps but in the absence of urea, *i.e.*, pulse-labelled under native conditions (Figure S1). The results indicate that a large number of residues that are susceptible and undergo rapid HDX in the PUFs are transiently populated under native conditions and that these PUFs are similar in structure and exchange properties to the intermediate states populated in 3.3 M urea. Thus, both PUFs and intermediate states have a highly stable core structure in which a number of amide groups are protected from exchange.

We next carried out the same pulsed-label NMR HDX experiments by incubating UCH-L1 in the presence of 2.0, 3.0, 3.3, 3.7, 4.0 and 4.5 M urea for 25 minutes at 25 °C to probe the structure of UCH-L1 over a range of non-native conditions (Figure 5). In the presence of 2.0 M urea, where UCH-L1 remains largely in the native state, the residual crosspeak distributions after pulsed-labelling remain the same as those without urea, suggesting that the intermediate state(s) are not yet populated. In the presence of 3.0 M urea, where

the intermediate state(s) of UCH-L1 begin to become populated based on the fluorescence measurements, there is a loss of about 40% of the initial crosspeak intensities after pulsed-labelling, consistent with the fact that intermediate state(s) are becoming populated. On increasing the concentration of urea to 3.3 and further to 3.7 M urea, where the denatured state of UCH-L1 starts to become populated, a large decrease of the residual crosspeak intensities is observed (a loss of more than 80%). When UCH-L1 is subject to HDX in the presence of 4.0 and 4.5 M urea, where the population of the denatured state becomes greater than that of the intermediate state(s), the amide proton resonances could no longer be detected, indicating that the amide protons in the protein have fully exchanged with the bulk deuterons within 25 minutes.

According to the kinetic analysis, the time constants for unfolding of UCH-L1 from N to I' or I'' are on the scale of hundreds of seconds at 3.0 M urea or below and tens of seconds at 3.3 M urea or higher (Figure 1). Therefore, the 25-minute incubation time is sufficient for the native protein to unfold into intermediate state(s) at 3.0 M urea in which some of the amide protons fully exchange with the bulk solvent. At higher concentrations of urea, the denatured state becomes populated and exchange can occur for the set of amide protons that remain protected in the intermediate state(s).

## DISCUSSION

The folding pathways of small, monomeric proteins have been studied in detail and much is known about the mechanisms and factors that influence folding in these cases [49, 50]. For small proteins, typically 100 residues or less in length with relatively simple topologies, there is considerable evidence that such

proteins have comparatively smooth energy landscapes for folding, [51, 52] and they do not populate highly stable intermediate states or get stuck in kinetic traps. In contrast, much less is known about the folding pathways of larger proteins with more complex topologies, including proteins with knots in their structures.

To date, the folding pathways of only a few knotted proteins have been characterised in any detail experimentally. These include YibK and YbeA [8, 10], MJ0366, the smallest protein that forms a knot [20] and the engineered knotted protein, tandem HP0242 [25] all of which contain a  $3_1$  trefoil knot. Furthermore, the folding pathway of UCH-L3, which contains a Gordian knot and is a homolog of UCH-L1, has also been reported [21]. Its folding pathway involves the formation of two hyper-fluorescent folding intermediates that are transiently populated on two parallel folding pathways connecting the native and denatured states [21].

In this study, we characterise in some detail the kinetic pathway and energy landscape for folding of the topologically knotted protein UCH-L1. Our previous studies on wild-type and mutant UCH-L1 have established that the protein reversibly unfolds in urea and populates an intermediate under equilibrium conditions, we also demonstrated that two unfolding phases could be observed kinetically and that the native states of mutants of UCH-L1 associated with PD were destabilized leading to an increase in the amount of intermediate state present under native conditions [23]. However, a comprehensive analysis of the folding pathway of UCH-L1 and a full characterization of the intermediate state(s) were not undertaken in this study.

Here, using results from single- and double-[denaturant] jump kinetic experiments and NMR HDX methods under both equilibrium and non-equilibrium conditions, we present both a full analysis of the kinetic pathway of the folding of wild-type UCH-L1 and also residue-specific information relating to the intermediate states observed. The results of the single- and double-[denaturant] jump experiments, which used SF methodology and intrinsic fluorescence to probe transitions between different states, supports a kinetic model of folding shown in Figure 1C. In this case, UCH-L1 folds along two parallel pathways on each of which there is a metastable intermediate state, similar to the kinetic scheme proposed for its homologue UCH-L3. Evidence for this model comes from the following results:

1. In the SJ unfolding kinetics, two phases are observed which are attributed to transitions between  $N \rightarrow I'$  and  $N \rightarrow I''$ . In this case, these are rate-determining steps in unfolding and the transitions from  $I'$  and  $I''$  to D are not observed.
2. Excluding the phase attributed to proline isomerization, there are four refolding phases in the SJ refolding experiments. The two slower phases corresponding to the folding of  $I'$  and  $I''$  to the native state, whilst the faster phases are attributed to folding from the denatured state to  $I'$  and  $I''$ . In this case, all four phases are observed in SJ refolding experiments as the transitions from I to N are rate limiting and the timescales for the formation of the intermediate states can be measured by SF kinetics.
3. The DJ interrupted-refolding experiments demonstrate that there are two additional fast phases corresponding to the unfolding of  $I'$  and  $I''$  as we would expect from the kinetic scheme proposed.

4. An analysis of the amplitude data from the DJ interrupted-refolding experiments indicates that the slowest kinetic phases observed in unfolding and refolding experiments (corresponding to the I'-N transition) are linked, as the rate constant for the formation of the slowest unfolding species ( $A_1$  in Figure 1B) is the same as the slowest refolding rate constant observed in the SJ refolding experiments. The same is true for the I''-N transition and the second slowest unfolding and refolding phases.
5. The parallel nature of the folding pathway is established by the fact that no lag is observed for the formation of I' or I'' or N (within the timescales employed)
6. Phases corresponding to transitions between I' (or I'') and D and N can be linked from the interrupted refolding experiments, as the rate of depletion of I' or I'' at longer aging times must have the same rate constant as the formation of N from I' or I'' (Figure 1).

It should be noted that we cannot rule out more complex folding pathways, for example, one in which heterogeneity in the denatured state means that I' and I'' form from specific subpopulations within the denatured state. Nor can we rule out that there may be additional early intermediate states formed before I' and I'' within the dead time of the SF measurements. The kinetic scheme shown in Figure 1C is the simplest scheme that fits the data.

Having established a kinetic scheme for the folding of UCH-L1, it is possible to use the kinetic data to obtain information on the intermediate states populated during folding. This was done by fitting the unfolding and refolding rate constants for the transition between I' or I'' and N to a two-state kinetic model



(Table 1). The stabilities of I' and I'' with respect to N, calculated using the refolding and unfolding rate constants extrapolated to water, are 5.04 and 4.15 kcal mol<sup>-1</sup> respectively. In addition, an *m*-value for the transition between I' or I'' and N can be calculated from the slopes of the refolding and unfolding arms for each phase taking into account of a factor of RT (Table 1). The summation yielded kinetic *m*-values of 3.05 and 2.59 kcal mol<sup>-1</sup> M<sup>-1</sup>, respectively, which are consistent with the value obtained from the equilibrium unfolding experiment (about 3 kcal mol<sup>-1</sup> M<sup>-1</sup>) [23]. Thus, the intermediates form quickly from the denatured state, with rate constants of approximately 1 and 10 s<sup>-1</sup> (D→I' and D→I'', respectively). Under native conditions the intermediates are lower in energy than the denatured state but they are only metastable with respect to the native state, and fold further on timescales of 0.07 and 0.43 s<sup>-1</sup> for I' and I'', respectively, to the native 5<sub>2</sub> knotted structure.

Whereas the equilibrium and kinetic fluorescence measurements provide strong evidence that intermediate states are populated both under equilibrium and non-equilibrium conditions, these experiments provide little, and only global, information on the structure of these states. In order to get higher-resolution data on PUFs and intermediate states, a series of NMR HDX measurements were undertaken in which the exchange rates of amide protons in different regions of the structure were measured under native conditions, in the presence of increasing amounts of denaturant, and under non-equilibrium conditions using pulsed-labelling techniques.

Under native conditions, *i.e.*, in the absence of denaturant (or in the presence of low concentrations of denaturant), amide protons in proteins can exchange with solvent D<sub>2</sub>O by a variety of different mechanisms including: (i) exchange

only on global unfolding, (ii) exchange after sub-domain unfolding to form a partially unfolded form, and (iii) exchange after a local fluctuation that exposes the backbone amide group. Exchange data collected under native conditions can therefore be used to inform on the structure of PUFs that are higher in energy than, but accessible from, the native state. Here, we also measured exchange rates under mildly to strongly denaturing conditions using pulsed-labelling techniques thereby probing the structure of intermediate states directly.

By comparing the data obtained from these experiments, in which the intermediate state is populated and exchange occurs from it directly, with the NMR HDX data acquired under native conditions, where exchange occurs from a high energy PUF, we have been able to show that the PUFs and intermediate states have similar properties and NMR spectra (Figure S1). Thus, it is highly likely that the PUFs are very similar in structure to the intermediate states populated during folding, and a structural characterisation of the PUFs can be used to inform on the structure of the intermediate states observed under non-equilibrium conditions.

To obtain the maximum amount of structural information on the PUFs from the exchange data, the exchange regime of the amide protons under study was established. Amide protons in proteins can exchange with an EX1 or EX2 mechanism, and the two mechanisms can be distinguished by their dependence of the exchange rate on pH (Figure 3). In the EX1 regime, the exchange rate is limited by the unfolding of the protein to an exchange-competent form. This may be a PUF or the fully denatured state. In this case, the exchange rates provide information on unfolding rates, which can be compared with values measured directly in the SF experiments. Four EX1 residues, I96, A98, H185 and D196

have exchange rate constants (equivalent to an opening or unfolding rate constant) that in water are within error of the value for  $k_u^{H_2O}$  for the slowest unfolding phase extrapolated from the SF measurements for unfolding in high concentrations of denaturant (Figure 1A). These results suggest that I96, A98, H185 and D196 are not protected in a PUF which has similar properties to I'. Therefore, these residues are in regions of the protein that are not structured in this state. As the unfolding rate for the formation of I'' from the native state is faster than for I', the fact that these four residues exchange with the lower rate suggests that in the PUF corresponding to I'' these residues are in regions of structure and cannot therefore undergo exchange.

Interestingly, NMR HDX experiments also reveal another set of EX1 residues that display slower opening rates (Figure 3). These residues are located in the central  $\beta$ -sheet and include E174 and V217. These results indicate that these residues only become exchange competent in the denatured state, suggesting that there is structure around E174 and V217 in all the PUFs. The  $m$ -values, which were derived from the urea-dependent NMR HDX analysis (Figure 4), are significantly larger than the  $m_u$  value corresponding to the unfolding from N to I' (Figure 1A) consistent with a model in which these very slow-exchanging EX1 residues exchange only after global unfolding from the native to denatured state (N→D), which would not only be much slower than the N→I' or I'' transitions but have a higher  $m$ -value due to the larger change in solvent accessible surface area between N and D reflecting the cooperative unfolding of the entire structure under these conditions.

Other amide protons in UCH-L1 exchange by an EX2 mechanism, and in these cases, the free energy of unfolding  $\Delta G_{HDX}$  can be calculated from the NMR

HDX exchange rates (Figure 4). This provides a measure of the stability of any high-energy PUFs on the energy landscape for folding/unfolding relative to the native state. The variation of  $\Delta G_{\text{HDX}}$  with denaturant concentration was also measured for conditions in which the protein remains mainly in the native state, *i.e.*, up to 2.2 M urea (Figure 4). This type of information has been used previously with several model systems to identify multiple unfolded forms of a given protein and, in some cases, an atomic view of sequential unfolding events can be constructed under the assumption that clusters of residues with similar properties represent foldons, which are cooperatively unfolding regions of structure.

For wild-type UCH-L1, only 21 residues are in the EX2 regime and these residues are sparsely distributed in the loop regions and flanking regions in the secondary structure elements (Figure 2 and Table S1). Even for these residues, no well-defined clusters with similar  $m$ -values and  $\Delta G_{\text{HDX}}^{\text{H}_2\text{O}}$  could be identified; instead, a rather continuous distribution of values for  $m$ -values and  $\Delta G_{\text{HDX}}^{\text{H}_2\text{O}}$  is observed (Figure 4). Such a continuum of unfolding stabilities has been reported for other proteins such as T4 lysozyme [45]. Despite us being unable to identify clear foldons in the structure of UCH-L1, structural information on the PUFs was obtained. For example, R129 in  $\alpha 5$  is an EX2 residue with a sizeable  $m$ -value and a free energy of unfolding  $\Delta G_{\text{HDX}} > 8.5 \text{ kcal mol}^{-1}$  (Figure 4). Therefore, it is likely that at least some regions of  $\alpha 5$  remain structured (or in some way protected from solvent exchange) in the intermediate states and that this residue exchanges only from the denatured state.

The NMR HDX-derived free energies of unfolding  $\Delta G_{HDX}^{H,O}$  for A48, N101, and Q103 are all close to 10 kcal mol<sup>-1</sup> (Figure 4B) suggesting that they are in structured regions in the two intermediate states and exchange only on complete unfolding of the protein to the denatured state. It is interesting to note that N101 and Q103 are in the loop preceding  $\alpha 3$  in the native state,  $\alpha 3$  containing structure in the intermediate states (Figure 3). Consistent with these results, A48, N101, Q103 are in close proximity to the very slow-exchanging EX1 residues in the central  $\beta$ -sheet, which likely undergo exchange only after global unfolding as discussed above. These data provide further evidence to suggest that the central  $\beta$ -sheet, and some regions close to it, are the main structural elements in the PUFs, and therefore also the intermediate states.

Even for those amide protons that cannot be unambiguously assigned as exchanging by an EX1 or EX2 mechanism, information relevant to the folding pathway can be obtained. For example, many of the residues in  $\alpha 2$  and  $\alpha 6$  (with the exception of E60) rapidly exchanged beyond detection within the dead time of 2-3 minutes in the NMR HDX experiments. This suggests that these helices undergo rapid fluctuations and unfolding events even under highly native conditions. Thus, as these helices are some of the first regions of structure to unfold, this means that these elements are unlikely to be structured in either of the intermediate states and are likely the last regions of structure to fold, probably after the final transition state barrier.

The NMR HDX results allow us to build up a picture of the likely structure of PUFs transiently populated under native conditions, which we have shown are similar to the intermediate states populated during folding. Many of the amide protons in UCH-L1 have relatively fast exchange rates, *i.e.*, low protection factors,

and therefore exchange by local fluctuations in the structure. Many of these residues are in  $\alpha$ -helices, in particular in  $\alpha 2$ ,  $\alpha 4$  and  $\alpha 6$ . However, for some of the other  $\alpha$ -helices, namely  $\alpha 1$ ,  $\alpha 3$ ,  $\alpha 5$ , and  $\alpha 7$ , the protection factors of some of the residues in the helices are of the order of 5 to 7.5, indicating that these residues do not exchange by local fluctuation but likely only after they are exposed in the formation of either I' or I''. In either case, this enables us to conclude that the folding intermediates of UCH-L1 have lost considerable  $\alpha$ -helical structure (Figure 3), consistent with previous studies using far-UV CD to characterise unfolding of UCH-L1 under equilibrium conditions [23].

In contrast to the  $\alpha$ -helices, there is considerable evidence that the central  $\beta$ -sheet of UCH-L1 retains significant structure in both intermediate states. This is shown by the exchange rates of residues in the  $\beta$ -sheet that exchange with an EX1 mechanism with rates corresponding to the global unfolding of the protein to the denatured state, as well as the  $\Delta G_{HDX}^{H_2O}$  values for residues that are in EX2 exchange and which are  $> 10 \text{ kcal mol}^{-1}$  which are also located close to or in the  $\beta$ -sheet.

It is also of interest to look at the HDX results of residues in loop regions. Many of these exchange very rapidly, within the dead time of the experiment, however, some show protection towards exchange suggesting that they are associated with regions of structure that are, at least, metastable. In most cases, these residues lie just before or after an element of secondary structure that is partially or fully formed in the intermediate state, with the exception of Y80 and Q84, which are located in the middle of loop 5 (Figure 3 and Table S1).

Using the NMR results, it is also possible to speculate on whether the two intermediate states detected are knotted or not. For UCH-L1, a  $3_1$  knot exists in the structure between residues 6-163, whilst a  $5_2$  knot exists between residues 5-219. Our results provide strong evidence that the C-terminal strand  $\beta 6$  (including residue L219) is structured in both intermediate states. In contrast, our results show that residues 5 and 6, that are in the N-terminal tail, are not structured in either intermediate state. This lack of structure at residues 5 or 6, which in the native state gives rise to a topological crossing of the polypeptide chain that is key for formation of both the  $3_1$  or  $5_2$  knot, means it is unlikely that either intermediate states are knotted. This is consistent with the threading, event occurring during the folding of I' or I'' to the native state. We have demonstrated here that, these steps are slow and, rate limiting in the folding of UCH-L1 and, therefore consistent with a threading event taking place. Threading has also been shown to be rate-limiting for other knotted systems [12, 24].

In summary, we have established the kinetic folding pathway of UCH-L1, which involves the formation of multiple folding intermediates along at least two parallel folding pathways, and have obtained structural information on the two intermediate states indirectly using NMR HDX methods. The results show us that the most structured region within both of the intermediate states is the central  $\beta$ -sheet. It is likely that this  $\beta$ -sheet is stabilised in part by a number of partially formed  $\alpha$ -helices and maybe even a loop packing against it. In particular, it is likely that  $\alpha 3$ , that has a number of slowly exchanging amides with high protection factors, but even transient interactions with regions of  $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 5$  and  $\alpha 7$  may also play some role.

The results of the analysis of the NMR data are consistent with both intermediate states containing a similar structured  $\beta$ -sheet core. This is consistent with our kinetic data, which establishes that the two intermediate states have similar stabilities with respect to the native state. However, the NMR data also reveals differences in structure between I' and I''. In particular, I96 and A98 in  $\alpha$ 3, H185 in  $\beta$ 5 and D196 in  $\alpha$ 7 exchange only in I' not I'', suggesting that there is additional structure around these sites in I' compared with I'', and that the packing of partially-formed  $\alpha$ -helices against the central  $\beta$ -sheet may differ in the two intermediates.

Taken together, the fluorescence-based kinetic results and the HDX determined by NMR in native and non-native conditions presented here, allows us to make some more general statements on the nature of the folding pathway of UCH-L1. In contrast to the well-defined and highly hierarchical pathways proposed for cytochrome c based on the identification of clear unfolding units (foldons) [44], our HDX results indicate that this is not the case for UCH-L1. However, given that two discrete intermediate states can be resolved in kinetic experiments, the folding of UCH-L1 is unlikely to occur by a continuum of pathways as suggested by the energy landscape studies on small proteins with simple topologies. Instead, our results are consistent with a mechanism in which there is clearly more than one pathway to form the native structure, however, there are not an infinite number of pathways. The lack of foldons identified for UCH-L1 may be as a result of the parallel pathways observed. In addition, it is quite likely that there are more than two intermediate states on the folding energy landscape, but that only two, which have the right kinetic and



spectroscopic properties are observed in our fluorescence experiments. This would also make the identification of foldons challenging. Alternatively, foldons may comprise non-local elements of structure making them more difficult to identify. In either case, we can conclude that the folding of UCH-L1 neither folds through a continuum of pathways, or a single discrete pathway. We have shown that the folding of UCH-L1 is complex, that the  $\beta$ -sheet core forms early and is present in all intermediate states, and that the rate-limiting step which involves the threading of the chain to form the  $5_2$  knot likely occurs during the late stages of folding. We postulate that the folding of UCH-L1, which is large (223 residues) in length and knotted, is typical of larger proteins with complex topologies in general.

## **ASSOCIATED CONTENT**

### **Supporting Information**

Additional experimental data and analyses (Figure S1 and Table S1).

## **AUTHOR INFORMATION**

### **Corresponding Authors**

sej13@cam.ac.uk and sthsu@gate.sinica.edu.tw

## **NOTES**

The authors declare no competing financial interest.

## **ACKNOWLEDGEMENTS**

This work was supported by the National Science Council (99-2911-I-007-034 and 104-2113-M-001-016), National Tsing Hua University and Academia Sinica, Taiwan. S.-T.D.H. was supported by a Career Development Award (CDA-00025/2010-C) from the International Human Frontier Science Program. The NMR spectra were obtained at the NMR facility of the Department of Chemistry, University of Cambridge and at the Core Facility for Protein Structural Analysis, supported by the National Core Facility Program for Biotechnology, Taiwan.

## REFERENCES

- [1] Virnau P, Mirny LA, Kardar M. Intricate knots in proteins: Function and evolution. *PLoS Comp. Biol.* 2006, 2, 1074-9.
- [2] Lai YL, Yen SC, Yu SH, Hwang JK. pKNOT: the protein KNOT web server. *Nucleic Acids Res.* 2007, 35, W420-W4.
- [3] Jamroz M, Niemyska W, Rawdon EJ, Stasiak A, Millett KC, Sulkowski P, et al. KnotProt: a database of proteins with knots and slipknots. *Nucleic Acids Res.* 2015, 43, D306-14.
- [4] Mallam AL. How does a knotted protein fold? *FEBS J.* 2009, 276, 365-75.
- [5] Virnau P, Mallam A, Jackson S. Structures and folding pathways of topologically knotted proteins. *J. Phys. Condens. Matter.* 2011, 23, 033101.
- [6] Lim NC, Jackson SE. Molecular knots in biology and chemistry. *J. Phys. Condens. Matter.* 2015, 27, 354101.

- [7] Mallam AL, Jackson SE. Folding studies on a knotted protein. *J. Mol. Biol.* 2005, 346, 1409-21.
- [8] Mallam AL, Jackson SE. Probing Nature's knots: The folding pathway of a knotted homodimeric protein. *J. Mol. Biol.* 2006, 359, 1420-36.
- [9] Mallam AL, Jackson SE. The dimerization of an alpha/beta-knotted protein is essential for structure and function. *Structure.* 2007, 15, 111-22.
- [10] Mallam AL, Jackson SE. A comparison of the folding of two knotted proteins: YbeA and YibK. *J. Mol. Biol.* 2007, 366, 650-65.
- [11] Mallam AL, Jackson SE. Use of protein engineering techniques to elucidate protein folding pathways. *Prog. Mol. Biol. Transl. Sci.* 2008, 84, 57-113.
- [12] Mallam AL, Jackson SE. Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat. Chem. Biol.* 2012, 8, 147-53.
- [13] Mallam AL, Morris ER, Jackson SE. Exploring knotting mechanisms in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 18740-5.
- [14] Mallam AL, Onuoha SC, Grossmann JG, Jackson SE. Knotted fusion proteins reveal unexpected possibilities in protein folding. *Mol. Cell.* 2008, 30, 642-8.
- [15] Mallam AL, Rogers JM, Jackson SE. Experimental detection of knotted conformations in denatured proteins. *Proc. Natl. Acad. Sci. U. S. A.* 2010, 107, 8189-94.
- [16] Lim NC, Jackson SE. Mechanistic insights into the folding of knotted proteins in vitro and in vivo. *J Mol Biol.* 2015, 427, 248-58.

- [17] Hsieh SJ, Mallam AL, Jackson SE, Hsu STD. Backbone NMR assignments of a topologically knotted protein in urea-denatured state. *Biomol NMR Assign.* 2014, 8, 439-42.
- [18] Hsieh SJ, Mallam AL, Jackson SE, Hsu STD. Backbone NMR assignments of a topologically knotted protein in urea-denatured state. *Biomol NMR Assign.* 2014, 8, 283-5.
- [19] Shih PM, Wang I, Lee YT, Hsieh SJ, Chen SY, Wang LW, et al. Random-Coil Behavior of Chemically Denatured Topologically Knotted Proteins Revealed by Small-Angle X-ray Scattering. *J. Phys. Chem. B.* 2015, 119, 5437-43.
- [20] Wang I, Chen SY, Hsu STD. Unraveling the folding mechanism of the smallest knotted protein, MJ0366. *J. Phys. Chem. B.* 2015, 119, 4359-70.
- [21] Andersson FI, Pina DG, Mallam AL, Blaser G, Jackson SE. Untangling the folding mechanism of the 5(2)-knotted protein UCH-L3. *FEBS J.* 2009, 276, 2625-35.
- [22] Andersson FI, Jackson SE, Hsu ST. Backbone assignments of the 26 kDa neuron-specific ubiquitin carboxyl-terminal hydrolase L1 (UCH-L1). *Biomol NMR Assign.* 2010, 4, 41-3.
- [23] Andersson FI, Werrell EF, McMorran L, Crone WJ, Das C, Hsu STD, Jackson SE. The effect of Parkinson's-disease-associated mutations on the deubiquitinating enzyme UCH-L1. *J. Mol. Biol.* 2011, 407, 261-72.
- [24] King NP, Jacobitz AW, Sawaya MR, Goldschmidt L, Yeates TO. Structure and folding of a designed knotted protein. *Proc. Natl. Acad. Sci. U. S. A.* 2010.

- [25] Wang LW, Liu YN, Lyu PC, Jackson SE, Hsu ST. Comparative analysis of the folding dynamics and kinetics of an engineered knotted protein and its variants derived from HP0242 of *Helicobacter pylori*. *J. Phys. Condens. Matter*. 2015, 27, 354106.
- [26] Li W, Terakawa T, Wang W, Takada S. Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc. Natl. Acad. Sci. U.S.A.*. 2012, 109, 17789-94.
- [27] Sulkowska JI, Noel JK, Ramirez-Sarmiento CA, Rawdon EJ, Millett KC, Onuchic JN. Knotting pathways in proteins. *Biochem. Soc. Trans.* 2013, 41, 523-7.
- [28] Sulkowska JI, Rawdon EJ, Millett KC, Onuchic JN, Stasiak A. Conservation of complex knotting and slipknotting patterns in proteins. *Proc. Natl. Acad. Sci. U. S. A.* 2012, 109, E1715-23.
- [29] Sulkowska JI, Noel JK, Onuchic JN. Energy landscape of knotted protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 2012, 109, 17783-8.
- [30] Noel JK, Sulkowska JI, Onuchic JN. Slipknotting upon native-like loop formation in a trefoil knot protein. *Proc. Natl. Acad. Sci. U. S. A.* 2010, 107, 15403-8.
- [31] Wallin S, Zeldovich KB, Shakhnovich EI. The folding mechanics of a knotted protein. *J. Mol. Biol.* 2007, 368, 884-93.
- [32] Bolinger D, Sulkowska JI, Hsu HP, Mirny LA, Kardar M, Onuchic JN, et al. A Stevedore's protein knot. *PLoS Comp. Biol.* 2010, 6, e1000731.

- [33] Setsuie R, Wada K. The functions of UCH-L1 and its relation to neurodegenerative diseases. *Neurochem. Int.* 2007, 51, 105-11.
- [34] Leroy E, Boyer R, Auburger G, Leube B, Ulm G, Mezey E, et al. The ubiquitin pathway in Parkinson's disease. *Nature.* 1998, 395, 451-2.
- [35] Hsu S-TD. Folding dynamics and structural basis of the enzyme mechanism of ubiquitin C-terminal hydrolyases. In: Svendsen A, editor. *Understanding enzymes, Function, Design, Engineering and Analysis*: Pan Stanford Publishing, 2015.
- [36] Lee YTC, Hsu STD. Familial Mutations and Post-translational Modifications of UCH-L1 in Parkinson's Disease and Neurodegenerative Disorders. *Curr. Protein. Pept. Sci.* 2016, in press.
- [37] Kumar SM, Lyu PC, Hsu S-TD. Structural perturbation of the Parkinson's disease-associated I93M mutation in human UCH-L1 revealed by solution state NMR spectroscopy. *Chinese Journal of Magnetic Resonance.* 2015, 32, 329-41.
- [38] Das C, Hoang QQ, Kreinbring CA, Luchansky SJ, Meray RK, Ray SS, et al. Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1. *Proc. Natl. Acad. Sci. U. S. A.* 2006, 103, 4675-80.
- [39] Boudreaux DA, Maiti TK, Davies CW, Das C. Ubiquitin vinyl methyl ester binding orients the misaligned active site of the ubiquitin hydrolase UCHL1 into productive conformation. *Proc. Natl. Acad. Sci. U. S. A.* 2010, 107, 9117-22.

- [40] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*. 1995, 6, 277-93.
- [41] Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*. 1991, 30, 10428-35.
- [42] Feng H, Zhou Z, Bai Y. A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* 2005, 102, 5026-31.
- [43] Bollen YJ, Kamphuis MB, van Mierlo CP. The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proc. Natl. Acad. Sci. U. S. A.* 2006, 103, 4095-100.
- [44] Bai Y, Sosnick TR, Mayne L, Englander SW. Protein folding intermediates: native-state hydrogen exchange. *Science*. 1995, 269, 192-7.
- [45] Llinas M, Gillespie B, Dahlquist FW, Marqusee S. The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat. Struct. Biol.* 1999, 6, 1072-8.
- [46] Krishna MMG, Hoang L, Lin Y, Englander SW. Hydrogen exchange methods to study protein folding. *Methods*. 2004, 34, 51.
- [47] Wildes D, Marqusee S. Hydrogen-exchange strategies applied to energetics of intermediate processes in protein folding. *Methods Enzymol.* 2004, 380, 328-49.

- [48] Huang JR, Hsu STD, Christodoulou J, Jackson SE. The extremely slow-exchanging core and acid-denatured state of green fluorescent protein. *HFSP J.* 2008, 2, 378-87.
- [49] Kuhlman B, Baker D. Exploring folding free energy landscapes using computational protein design. *Curr. Opin. Struct. Biol.* 2004, 14, 89-95.
- [50] Schaeffer RD, Fersht A, Daggett V. Combining experiment and simulation in protein folding: closing the gap for small model systems. *Curr. Opin. Struct. Biol.* 2008, 18, 4.
- [51] Oliveberg M, Wolynes PG. The experimental survey of protein-folding energy landscapes. *Q. Rev. Biophys.* 2005, 38, 245-88.
- [52] Onuchic JN, Wolynes PG. Theory of protein folding. *Curr. Opin. Struct. Biol.* 2004, 14, 70-5.

## FIGURES LEGENDS

**Figure 1. Folding kinetics of UCH-L1 by stopped-flow fluorescence measurements.** (A) UCH-L1 exhibits four unfolding phases, namely phases 1 (blue circle) and 2 (open orange circle), 3 (red inverted triangle) and 4 (purple open triangle), and five refolding phases, namely phases 1 to 4 that are paired with the four unfolding phases plus the slowest phase s (open diamond; tentatively assigned to proline *trans-cis* isomerisation) in addition to the four refolding phases that are paired with the unfolding phases in the same symbols. The DJ experiments-derived amplitude build-up rates ( $(DJ)_1$ ) shown in (B) are



indicated in larger symbols; likewise, the global-fitted reaction rates ( $DJ_2$ ) are also indicated in larger symbols. The black dashed line corresponds to the linear dependency of the slowest unfolding phase of UCH-L1 from the native to intermediate state. Extrapolation of the unfolding arm to zero denaturant gives the global unfolding rate of UCH-L1 in water,  $k_{1u}^{H_2O}$ , and is in good agreement with the opening rates,  $k_{op}$ , of residues 96 (filled black circle) and 185 (filled green triangle) derived from NMR HDX experiment under native conditions. The [urea]-dependent HDX rates of 174 (filled blue diamond) and 217 (filled red square), which are also in the EX1 regime but which have much slower opening rates, are also shown for comparison. (B) Amplitude analysis of interrupted-refolding analysis of UCH-L1. Refolding of 8.5 M urea-denatured UCH-L1 was initiated in the presence of 1.0 M urea for different aging times ( $t_{age}$ ) ranging from 50 ms to 100 s, before a second mixing step to unfold UCH-L1 in the presence of 4.9 M urea. The kinetic traces were globally fit to a quadruple exponential function in which the four unfolding rates (0.045, 0.16, 0.46 and 12.53 s<sup>-1</sup> for  $k_1$  to  $k_4$ , respectively) are shared between different kinetic traces. The four amplitudes,  $A_1$ - $A_4$ , corresponding to the four kinetic unfolding phases are plotted against  $t_{age}$  with the same colours and symbols as in (A).  $A_1$  and  $A_2$  are fit to a single-exponential function and their rate constants are indicated.  $A_3$  and  $A_4$  are fit to a double-exponential function with the slower decaying rate fixed to the build-up rate of 0.014 s<sup>-1</sup> as for  $A_1$ . This constrain is necessary to obtain converging fitting results as the initial amplitudes of  $A_3$  and  $A_4$  are too small and scattered (inset) to yield reliable fitting results. (C) Proposed kinetic folding pathway of UCH-L1. The reaction rates between the two intermediates, *i.e.*, I' and I'', and denatured state, D, were estimated by manual extrapolations of the

observed rates to zero molar urea. The colours of the arrows correspond to the same colour scheme in (A).

**Figure 2. [Urea]-dependent NMR HDX analysis.** (A) Plot of HDX rate constants,  $k_{ex}$ , in the presence of urea *versus* those that are measured under native conditions, *i.e.*, 0 M urea. The HDX constants are coloured from blue to red on increasing urea concentration as indicated in the lower right corner. Dashed line defines the condition in which the HDX rate constants are not affected by the presence of urea, *i.e.*,  $k_{ex}$  (0 M urea) =  $k_{ex}$  (x M urea). (B) Arrow plot of the HDX rate constants of specific residues as a function of urea concentration. The data points are coloured in the same way as in (A) and connected by the arrows.

**Figure 3. NMR HDX analysis of UCH-L1.** (A) Correlation plot of the protection factors obtained at pH 7.6 (horizontal axis) and at pH 8.6 (vertical axis). Residues that correspond to EX1, EX2 and in between (EX1-EX2) are shown in blue circles, red triangles and purple squares, respectively. The expected values for EX1 and EX2 regimes are shown in blue and red lines, respectively. (B) Topological mapping of the HDX regimes. The colouring scheme is the same as in (A).

**Figure 4. [Urea]-dependent NMR HDX analysis of UCH-L1.** (A) Free energy of unfolding  $\Delta G_{HDX}$  as a function of urea concentration. The free energies of unfolding are calculated for the EX2 residues and the corresponding values are subject to linear regression to extract the  $m$ -values and  $\Delta G_{HDX}^{H_2O}$  according to  $\Delta G_{HDX}[Urea] = \Delta G_{HDX}^{H_2O} - m \cdot [Urea]$ . (B) Cluster analysis of  $m$ -values as a function of  $\Delta G_{HDX}^{H_2O}$ . Residues that display similar  $m$ -values and  $\Delta G_{HDX}^{H_2O}$  are clustered and coloured the same. The cluster analysis was made by visual inspection.

**Figure 5. Pulsed-label NMR HDX analysis of UCH-L1.** The crosspeak intensities of individual residues (open green circles) are normalised with respect to the native state and the overall distributions are represented using the box plot. The fractional populations of native (N), intermediate (I) and denatured (D) states are calculated using the previously reported  $m$ -values and  $[D]_{50\%}$  obtained from the three-state fitting the data from an equilibrium unfolding experiment where UCH-L1 is denatured by urea using intrinsic fluorescence as the structural probe.

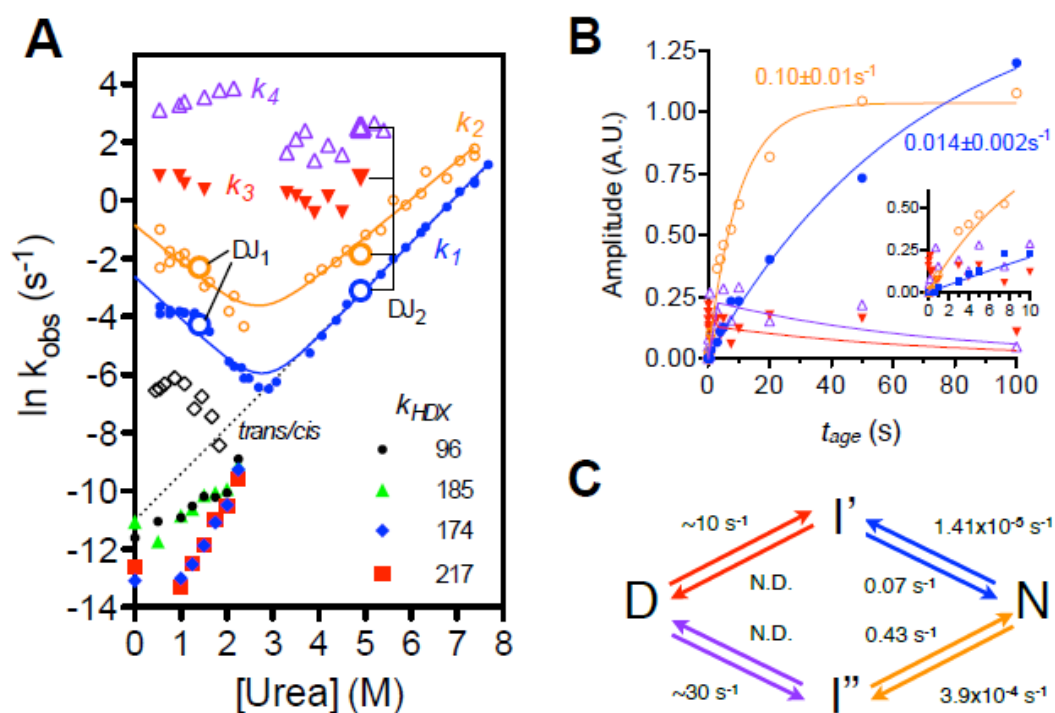


Figure 1

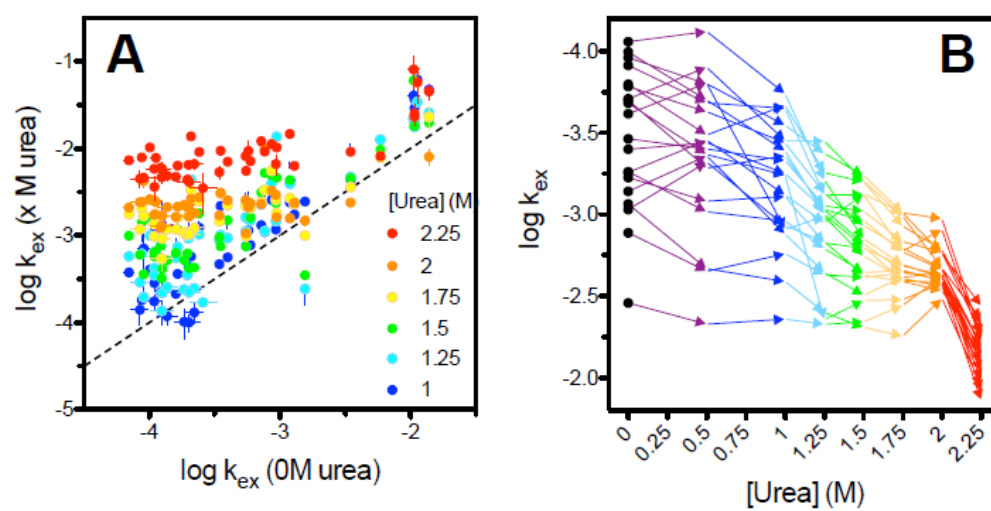


Figure 2

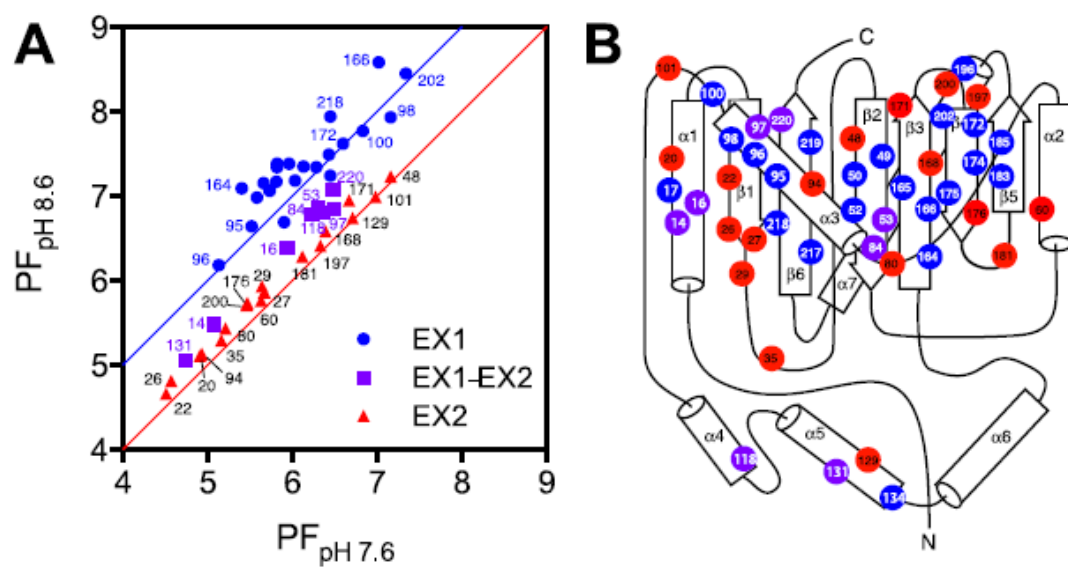


Figure 3

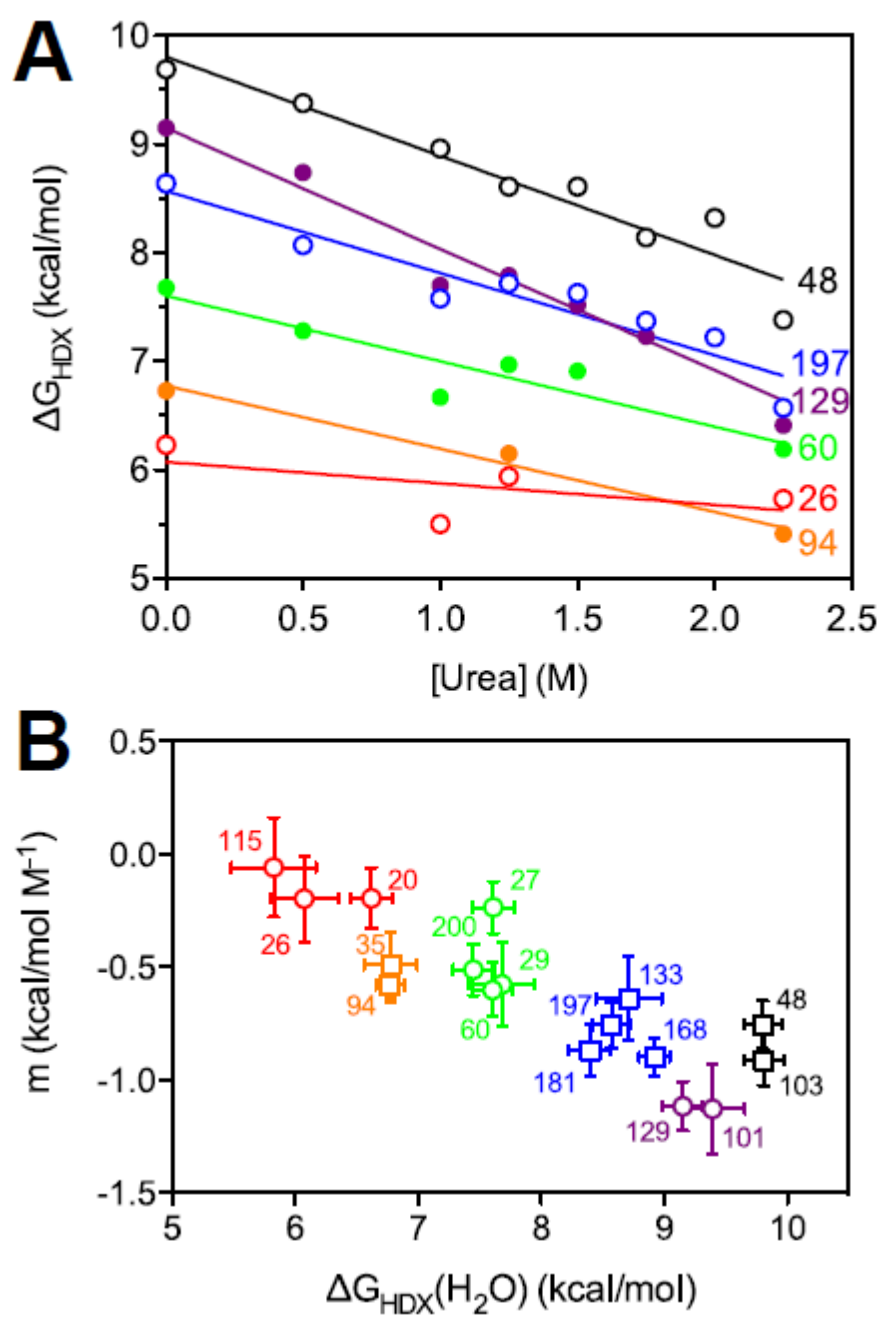


Figure 4

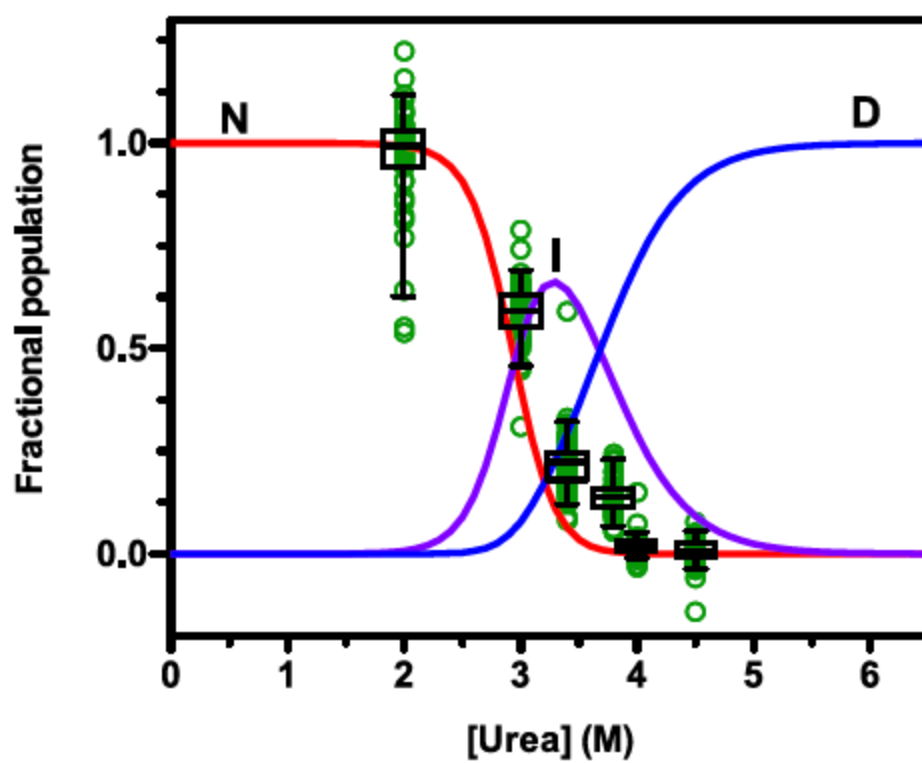
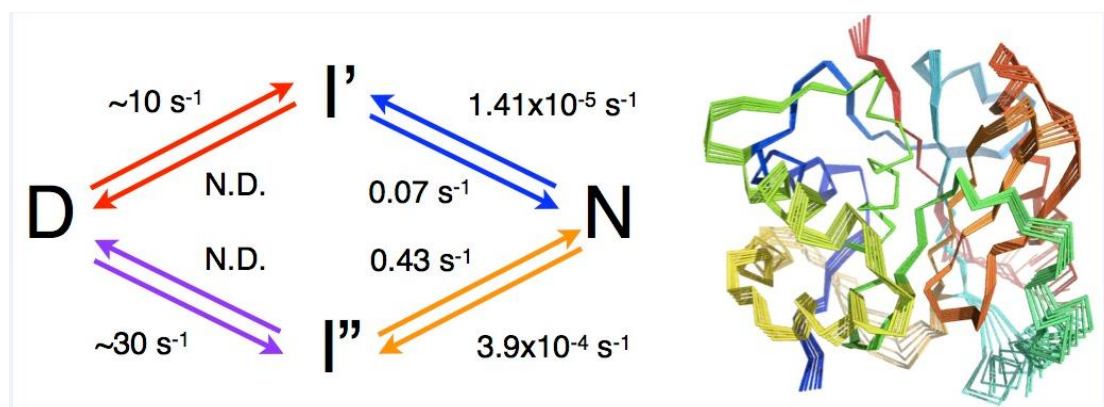


Figure 5



**Table 1. Kinetic parameters from the analysis of unfolding and refolding rate constants for the phases corresponding to transitions between native and intermediate states**

	Phase 1	Phase 2
$k_f^{H_2O}$ (s <sup>-1</sup> )	0.07±0.01	0.43±0.23
$m_f$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	1.43±0.18	1.31±0.13
$k_u^{H_2O}$ (s <sup>-1</sup> )	(1.41±0.37)*10 <sup>-5</sup>	(3.9±3.1)*10 <sup>-4</sup>
$m_u$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	-1.62±0.06	-1.28±0.51
[D] <sub>50%</sub> (M)	2.79±0.18	2.70±0.66
$\Delta G$ (kcal mol <sup>-1</sup> )	5.04±0.18	4.15±0.57



Graphical abstract

## Highlights

UCH-L1 is an abundant neuronal 5<sub>2</sub>-knotted protein implicated in Parkinson's disease

UCH-L1 exhibits parallel folding pathways with two distinct folding intermediates

Highly populated partially unfolded forms (PUFs) was observed under native conditions

The PUFs share similar structural features with the folding intermediates

Perturbation of the PUFs of UCH-L1 could lead to increased aggregation propensity