

Structural Basis for Ufm1 Processing by UfSP1*

Received for publication, October 23, 2007, and in revised form, February 20, 2008. Published, JBC Papers in Press, March 4, 2008, DOI 10.1074/jbc.M708756200

Byung Hak Ha^{‡§1}, Hee-Chul Ahn[‡], Sung Hwan Kang^{§1}, Keiji Tanaka¹, Chin Ha Chung^{§2},
and Eunice EunKyeong Kim^{‡3}

From the [‡]Life Sciences Division, Korea Institute of Science and Technology, 39-1 Hawolok-dong, Sungbuk-gu, Seoul 136-791, Korea, the [§]School of Biological Sciences, Seoul National University, Seoul 151-742, Korea, and the

¹Department of Molecular Oncology, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113-8613, Japan

Ubiquitin-fold modifier 1 (Ufm1) is a newly identified ubiquitin-like protein. Like ubiquitin and other ubiquitin-like proteins, Ufm1 is synthesized as a precursor that needs to be processed to expose the conserved C-terminal glycine prior to its conjugation to target proteins. Two novel proteases, named UfSP1 and UfSP2, have been shown to be responsible for the release of Ufm1 from Ufm1-conjugated cellular proteins as well as for the processing of its precursor. They show no sequence homology with known proteases. Here, we describe the 1.7 Å resolution crystal structure of mouse UfSP1, consisting of 217 amino acids. The structure reveals that it is a novel cysteine protease having a papain-like fold, with Cys⁵³, Asp¹⁷⁵, and His¹⁷⁷ that form a catalytic triad, and Tyr⁴¹ that participates in the formation of the oxyanion hole. This differs from the canonical catalytic triad of papain-like proteases in that the aspartate and the histidine residues are from the “Asp-Pro-His” box. The Asp-Pro-His configuration seen in UfSP1, together with Atg4B and M48^{USP}, seem to form a new subfamily of the cysteine protease superfamily. The mutagenesis study of the active site residues confirms structural basis for catalysis. The interaction between UfSP1 and Ufm1 appears quite substantial, since the K_D value was estimated to be 1.6 μ M by the isothermal titration calorimetry analysis. Furthermore, the NMR data shows that the loop between β 3 and α 2 in addition to the C-terminal region of Ufm1 plays a role in binding to UfSP1.

In eukaryotes, the posttranslational modification of cellular proteins by ubiquitin (Ub)⁴ is an essential step in regulatory

mechanisms of many cellular processes. These include cell cycle progression, signal transduction, and targeting of misfolded proteins for degradation by the proteasome (1–4). Ub, which is a 76-residue polypeptide with a well defined α/β fold, is highly conserved in all eukaryotes but absent in bacteria or archaea (5). It is synthesized as an inactive precursor and needs to be processed by deubiquitinating enzymes (DUBs) to expose the C-terminal glycine. The C terminus of Ub is then covalently ligated to the ϵ -amino group of lysine residues on target proteins by the sequential action of three classes of enzymes: a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2), and a ubiquitin protein ligase (E3). Ub may be attached to a target protein as a monomer or as a polyubiquitin chain. The ubiquitination pathway is further regulated by DUBs that also remove Ub from Ub-conjugated proteins (6–11).

In addition to Ub, there are a number of distinct ubiquitin-like proteins (UBLs), such as SUMO (small Ub-related modifier), ISG15, NEDD8, and Atg8, that function as protein modifiers. Like Ub, these UBLs also function as critical regulators of many cellular processes, such as transcription, DNA repair, signal transduction, autophagy, and the cell cycle (12, 13). The UBLs share further similarities with Ub. For example, UBLs display tertiary structure similar to that of Ub, although they lack obvious sequence identity. They also require a series of E1, E2, and E3 enzymes in order to be conjugated to a target protein. Like Ub, most UBLs, if not all, are synthesized as precursors that are processed by UBL-specific proteases (ULPs). ULPs also play a role in the release of UBLs from their conjugated proteins. However, unlike Ub, most UBLs are conjugated to their substrates as a monomer. Dysregulation of UBL substrate modification and/or mutations in the UBL conjugation machinery lead to a number of human diseases (11, 14, 15).

Recently, a novel UBL called ubiquitin-fold modifier 1, or Ufm1, has been identified (16). It has the same tertiary structure as Ub despite the fact that they share only 16% sequence identity (17). Similar to other UBLs, Ufm1 is synthesized as a precursor. Interestingly, however, Ufm1 possesses a single glycine at its C terminus followed by a Ser-Cys dipeptide in the precursor form unlike Ub or UBLs that have the conserved diglycine at the C terminus. The matured Ufm1 is specifically activated by a novel E1-like enzyme called Uba5 and then transferred to its cognate E2-like enzyme, called Ufc1. Ufm1 is demonstrated to conjugate several proteins in human HEK293 cells and various mouse tissues, and the identification of these target proteins is under way (16). Since Ufm1, Uba5, and Ufc1 are conserved in metazoa and plants but not in yeast, potential roles of Ufm1

* This work was supported by grants from the Functional Proteomics Center, the 21st Century Frontier Research and Development Program of the Korea Ministry of Science and Technology and the Korea Institute of Science and Technology Institutional Program Grant (to E. E. K.) and by grants from the Korea Research Foundation and the Korea Science and Engineering Foundation (to C. H. C.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2Z84) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ Recipient of the BK21 fellowship from the Korea Ministry of Education.

² To whom correspondence may be addressed. Tel.: 82-2-880-6693; Fax: 82-2-871-9193; E-mail: chchung@snu.ac.kr.

³ To whom correspondence may be addressed. Tel.: 82-2-958-5937; Fax: 82-2-958-5909; E-mail: eunice@kist.re.kr.

⁴ The abbreviations used are: Ub, ubiquitin; UBL, ubiquitin-like protein; E1, ubiquitin-activating enzyme; E2, ubiquitin-conjugating enzyme; E3, ubiquitin ligase; DUB, deubiquitinating enzymes; ULP, UBL-specific protease; ITC, isothermal titration calorimetry; MAD, multiple wavelength anomalous diffraction.

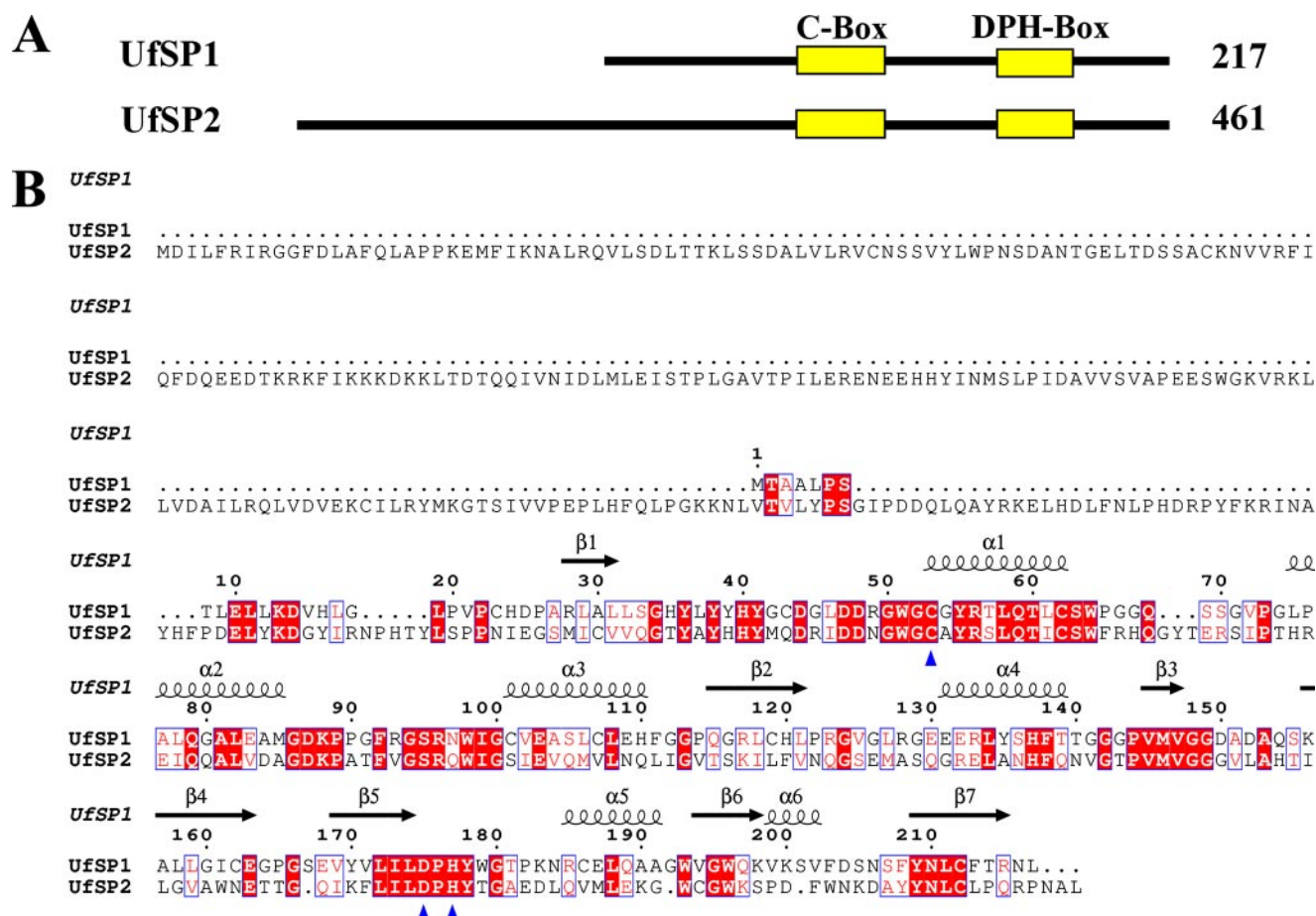


FIGURE 1. **UfSP1 and UfSP2.** A, primary structures of UfSP1 and UfSP2. B, amino acid sequence alignment of mouse UfSP1 (SwissProt accession code Q9CZP0) and UfSP2 (SwissProt accession code Q99K23). Secondary structures of mouse UfSP1 are depicted above the sequences; coils indicate α -helices, and arrows indicate β -strands. The catalytic Cys, Asp, and His are indicated by triangles at the bottom. The residues in the red box are strictly conserved, whereas the residues in the blue box are relatively conserved. The alignment was derived using ClustalW.

modification in various multicellular organisms are of special interest.

Earlier two novel proteases specific for Ufm1, named UfSP1 and UfSP2, were identified (18). Both UfSP1 and UfSP2 cleave only the C-terminal extension of Ufm1 but not Ub or UBLs. In addition, they are both capable of releasing Ufm1 from Ufm1-conjugated cellular proteins. The long UfSP2 is present in most, if not all, multicellular organisms, whereas the short UfSP1 is not found in plants or nematodes. Both UfSP1 and UfSP2 possess highly conserved Cys and His residues, indicating that they are cysteine proteases, and biochemical and mutational analysis confirmed that they are indeed thiol proteases (see Fig. 1). However, they share no sequence homology with other known proteases, including DUBs or ULPs, suggesting a new subfamily of cysteine proteases. Here, we report the crystal structure of mouse UfSP1 at 1.7 Å resolution and discuss the structural basis for Ufm1 processing.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The cDNAs for Ufm1 (SwissProt entry P61961) and UfSP1 (SwissProt entry Q9CZP0) from mice were cloned into pET28a and pET22b (Novagen) to generate N- or C-terminal His-tagged proteins. The resulting vectors were transformed to *Escherichia coli* BL21(DE3) codon plus

RIL (Stratagene) cells, and the His-tagged proteins were purified using nickel affinity resins (GE Healthcare) equilibrated with 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 1 mM Tris(2-carboxyethyl)phosphine hydrochloride. They were further purified by gel filtration on a Superdex 75 26/60 column (GE Healthcare). The purified UfSP1 was concentrated to 40 mg/ml in a buffer containing 20 mM HEPES (pH 7.5), 100 mM NaCl, and 1 mM dithiothreitol using Amicon Ultra-15 (Millipore). Selenomethionine-substituted UfSP1 was generated as described previously (19). The enzymatic activity was measured as described earlier (18).

Crystallization—Initial screening for the crystallization was carried out by using 96-well Intelli plates (Hampton Research) and a Hydra II Plus One (MATRIX Technology) robotics system at 295 K, and they were further optimized using the hanging drop methods. Diffraction quality crystals were obtained by mixing equal volumes of 40 mg/ml mouse UfSP1 in 20 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM dithiothreitol with a reservoir solution containing 0.1 M succinic acid (pH 7.0), and 5% (v/v) polyethylene glycol 3350. Attempts to crystallize UfSP1 complexed with Ufm1 did not yield crystals suitable for high resolution data collection.

X-ray Data Collection and Processing—The crystals of UfSP1 belong to the space group *R*32, with $a = b = 71.70$ Å, $c = 209.23$

Å, and $\alpha = \beta = 90^\circ$ and $\gamma = 120^\circ$. There is one molecule per asymmetric unit. Crystals were equilibrated in a cryoprotectant buffer containing reservoir buffer plus 20% (v/v) ethylene glycol and then flash-cryocooled in a cold nitrogen stream at 100 K. Both the native and the MAD data sets were collected at beamline 4A of Pohang Accelerator Laboratory (Pohang, Korea). The data were processed and scaled by using the HKL2000 program suite (20), and the statistics are summarized in Table 1.

Structure Determination and Refinement—The crystal structure of UfSP1 was determined by the MAD phasing method. All three selenium sites were found and refined, and the initial phases were calculated using the program SOLVE (21) and RESOLVE (22). About 84% of the residues were automatically modeled as a polyalanine chain by RESOLVE and further constructed using the molecular modeling program COOT (23). The refinement was then performed using the CNS and REFMAC (24, 25). The final refinement statistics are summarized in Table 1.

Site-directed Mutagenesis and UfSP1 Activity Assay—Site-directed mutagenesis on the residues that might be involved in the catalysis was carried out using QuikChange site-directed mutagenesis kit (Stratagene) by following the manufacturer's instructions. Mutants of UfSP1 were produced as MBP-fused proteins with single point mutations at positions Tyr⁴¹, Cys⁵³, Trp⁹⁸, Gln¹⁵⁴, Asp¹⁷⁵, and His¹⁷⁷. Ufm1 processing activity was assayed by using GST-Ufm1-HA as a substrate as described previously (18). Briefly, 100 ng of proteins of wild type and mutant forms of MBP-UfSP1 were incubated with 5 μ g of GST-Ufm1-HA for 1 h at 37 °C. The reaction was stopped by the addition of SDS sample buffer and analyzed using SDS-PAGE. The gels were then stained with Coomassie Blue R250.

Ufm1 Binding by NMR Chemical Shift Perturbation—Uniformly labeled ¹⁵N-labeled or ¹³C/¹⁵N-labeled mouse Ufm1 was produced in M9 minimal medium with ¹⁵NH₄Cl or a mixture of ¹⁵NH₄Cl and ¹³C-labeled glucose, respectively. The C53S mutant of mouse UfSP1 was used. They were purified as described above and concentrated to 0.3–0.8 mM. For backbone resonance assignment, CBCA(CO)NH and HNCACB experiments were performed using 0.8 mM ¹³C/¹⁵N-enriched Ufm1 in 20 mM deuterated Tris-HCl, pH 7.5, with 100 mM NaCl, 1 mM dithiothreitol, and 10% (v/v) D₂O. NMR titration experiments were performed by measurement of ¹H-¹⁵N TROSY on 0.3 mM ¹⁵N-Ufm1 with various amounts of unlabeled UfSP1 in 20 mM sodium phosphate, 150 mM NaCl, 1 mM dithiothreitol, and 10% (v/v) D₂O at pH 6.8. The molar ratio of unlabeled UfSP1 to ¹⁵N-labeled UfSP1 was about 0.1, 0.2, 0.3, 1:0.5, and 1:1, respectively. ¹H-¹⁵N resonances of Ufm1 at pH 6.8 with or without UfSP1 were assigned from the analysis of ¹⁵N-edited NOESY-HSQC spectrum of the respective sample. All NMR spectra were collected at 25 °C on a Varian Unity Inova 900 MHz spectrometer equipped with Coldprobe™ at KIST. All NMR data were processed using NMRPipe software (26) and analyzed using NMRView software (27).

Isothermal Titration Calorimetry—Isothermal titration calorimetric experiments were performed using a MicroCal VP isothermal titration calorimetry (ITC) instrument (MicroCal, Northampton, MA) at 22 °C, and the data were analyzed using the program ORIGIN. Protein samples were prepared in 20 mM

HEPES (pH 7.5) and 200 mM NaCl buffer. The purified C53S mutant of UfSP1 protein was concentrated to 1.5 mM, whereas Ufm1 was concentrated to 0.1 mM using an Amicon Ultra-15 concentrator. All protein solutions were degassed for 10 min under vacuum with a ThermoVac accessory prior to performing the experiment at 22 °C. During a typical titration experiment, 250 μ l of Ufm1 protein solution (or the same volume of the buffer as a control) was added to 1.4 ml of UfSP1 protein, and 250 μ l buffer solutions were added to UfSP1 protein as a control under the same condition. Typically, 5 μ l injections, with 10 s duration, were made every 180 s, and the sample cell was stirred with the flat tip of the syringe rotating at 307 rpm.

RESULTS

Both Ufm1-(1–85) and UfSP1-(6–215) from mice were expressed in *E. coli* and purified. The apo form of UfSP1 was crystallized in rhombohedral form, and the crystal structure has been determined using multiple-wavelength anomalous dispersion data collected from the selenomethionine-substituted UfSP1 and refined to a final *R* value of 18.6% (*R*_{free} = 22.7%) at 1.7 Å resolution. The electron density map is clearly defined for all atoms except for residues 149–154. For those six residues, the electron density was weak but enough to trace, so they were included in the final model. The side chain of Trp⁹⁸ can be modeled in two distinct orientations. Table 1 summarizes statistics on the crystallographic data. Binding of Ufm1 to UfSP1 has been analyzed using ITC and NMR chemical shift perturbation.

Overall Crystal Structure of Mouse UfSP1—The overall structure, with dimensions of 45 × 30 × 38 Å, consists of an α/β -fold with a seven-stranded antiparallel β -sheet flanked by α -helices on either side of the sheet and the N terminus forming an antiparallel β -structure. The central antiparallel β -sheet is composed of β 1, β 4, β 5, and β 6; the C-terminal antiparallel β -sheet consists of β 2, β 3, and β 7 (Fig. 2A). The helices α 1, α 2, and α 3 are located at one side of the β -sheet, whereas the other three are on the opposite side (Fig. 2B). The active site cysteine is located at the N terminus of helix α 1, whereas histidine is at the loop between β 5 and α 5. The overall structure mimics the papain fold despite the fact that they share no obvious sequence homology (11%). Since the C-terminal catalytic domain of UfSP2 shares 36% sequence identity with UfSP1, the catalytic domain of UfSP2 is expected to assume the same structure as UfSP1.

Comparison of UfSP1 with other structures in the Protein Data Bank using the DALI algorithm (28) yielded a few structural homologues from the cysteine protease superfamily. The most significant match, based on having the highest Z-score, was Atg4B (Protein Data Bank accession code 2CY7 (29) and Protein Data Bank accession code 2D1I (30)), which showed a Z-score of 12.6. Atg4B, an essential enzyme in autophagy, cleaves nascent Atg8 at its C-terminal arginine residue and also deconjugates Atg8 family proteins from a small adduct, phosphatidylethanolamine. The second highest hit was the recently reported protease domain of murine cytomegalovirus, M48^{USP}, with a Z-score of 6.7 (Protein Data Bank accession code 2J7Q (31)). DUBs, such as a 40-kDa catalytic core domain of HAUSP (Protein Data Bank accession code 1NBF (32)) and UCH-L3

TABLE 1

Data collection and crystallographic refinement statistics

Values in parentheses refer to the highest resolution shell.

	MAD			Native
	Peak	Edge	Remote	
Data sets				
Beam line	BL-4A	BL-4A	BL-4A	
X-ray wavelength (Å)	0.979512	0.979628	0.971834	1.0
Energy (eV)	12657.8	12656.3	12757.8	12398.9
Resolution range (Å)	50–1.9	50–1.9	50–1.9	50–1.64
Space group	R32	R32	R32	R32
Unit cell parameters (Å)	$a = 71.617, b = 71.617, c = 208.919, \alpha = 90, \beta = 90, \gamma = 120$	$a = 71.629, b = 71.629, c = 208.999, \alpha = 90, \beta = 90, \gamma = 120$	$a = 71.630, b = 71.630, c = 209.024, \alpha = 90, \beta = 90, \gamma = 120$	$a = 71.704, b = 71.704, c = 209.228, \alpha = 90, \beta = 90, \gamma = 120$
Total/unique reflections	816,996/16,707	821,201/16,727	850,449/16,740	602,580/25,883
Completeness (%)	99.7 (99.2)	99.5 (98.7)	99.2 (97.7)	98.7 (89.2)
Mean $I/\sigma(I)$ (%)	29.8 (5.5)	27.5 (4.6)	21.8 (3.3)	31.4 (5.6)
R_{merge}^a (%)	7.2 (23.4)	6.7 (26.2)	7.4 (31.8)	8.6 (24.3)
Refinement statistics				
Resolution range (Å)				40–1.7
R/R_{free}^b (%)				18.6/22.7
No. of protein atoms				1599
No. of water molecules				111
Average B -factor (Å ²)				23.2
Root mean square deviation from ideal geometry:				0.01/1.37
bond length (Å)/angle(degrees)				
Ramachandran analysis (%)				
Favored region				91.3
Additionally allowed				8.7
Generously allowed				0
Disallowed regions				0

^a $R_{\text{merge}} = \sum_i \sum_h |I(h,i) - \langle I(h) \rangle| / \sum_i \sum_h I(h,i)$, where $I(h,i)$ is the intensity of the i th measurement of reflection h , and $\langle I(h) \rangle$ is the mean value of $I(h,i)$ for all i measurements.

^b R_{free} was calculated from the randomly selected 10% set of reflections not included in the calculation of the R value.

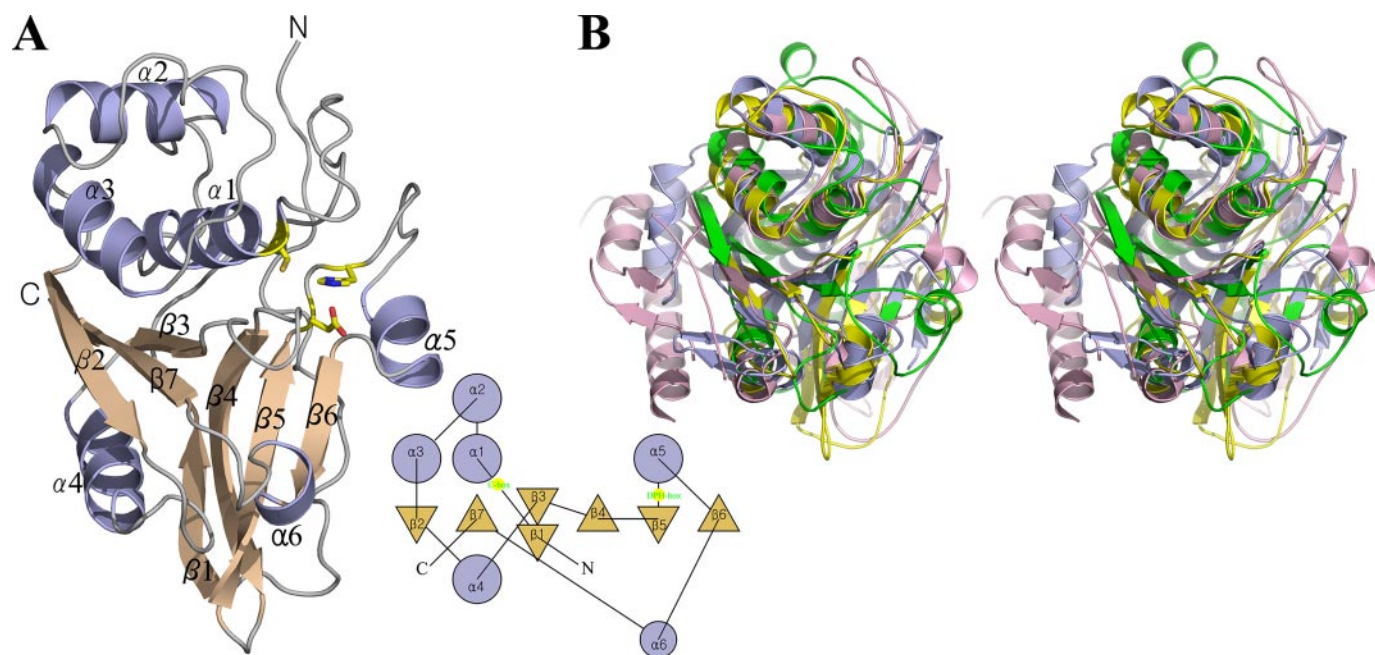


FIGURE 2. **Structure of UfSP1.** A, ribbon diagram of the overall structure of UfSP1. The secondary structural elements are labeled. The side chains of the catalytic residues, Cys⁵³, Asp¹⁷⁵, and His¹⁷⁷, are included. Topology of the protein structure is shown with α -helices indicated by blue circles and β -strands shown by yellow triangles. B, comparison with other cysteine proteases in stereo. The UfSP1 in yellow, human Atg4B (Protein Data Bank code 2CY7) in pink, and cytomegalovirus M48^{USP} (Protein Data Bank code 2J7Q) in blue and papain (Protein Data Bank code 1KHP) in green were structurally aligned and superposed.

(Protein Data Bank accession codes 1UCH and 1XD3 (33, 34)) showed only low structural similarities with UfSP1, as indicated by the Z-scores of 3.9 and 1.9, respectively. Fig. 2B shows the superposition of Atg4B, M48^{USP}, and papain on UfSP1.

Despite their overall structural similarity, Atg4B and M48^{USP} exhibit a number of significant local structural differences from UfSP1 (Fig. 2B). Atg4B has an additional domain called the “short fingers domain” or “auxiliary domain,” which is com-

posed of three α -helices and two β -stands forming an α/β structure whose function is unknown (29, 30). On the other hand, M48^{USP} has an “extended β -hairpin” structure that participates in extensive hydrophobic interaction with the Ub core that is bound. This interaction is quite distinct from what has been seen in the DUB·Ub or ULP·UBL complexes, and this Ub binding mode has been suggested to contribute to Ub specificity of M48^{USP} (31). HAUSP has three domains, described as

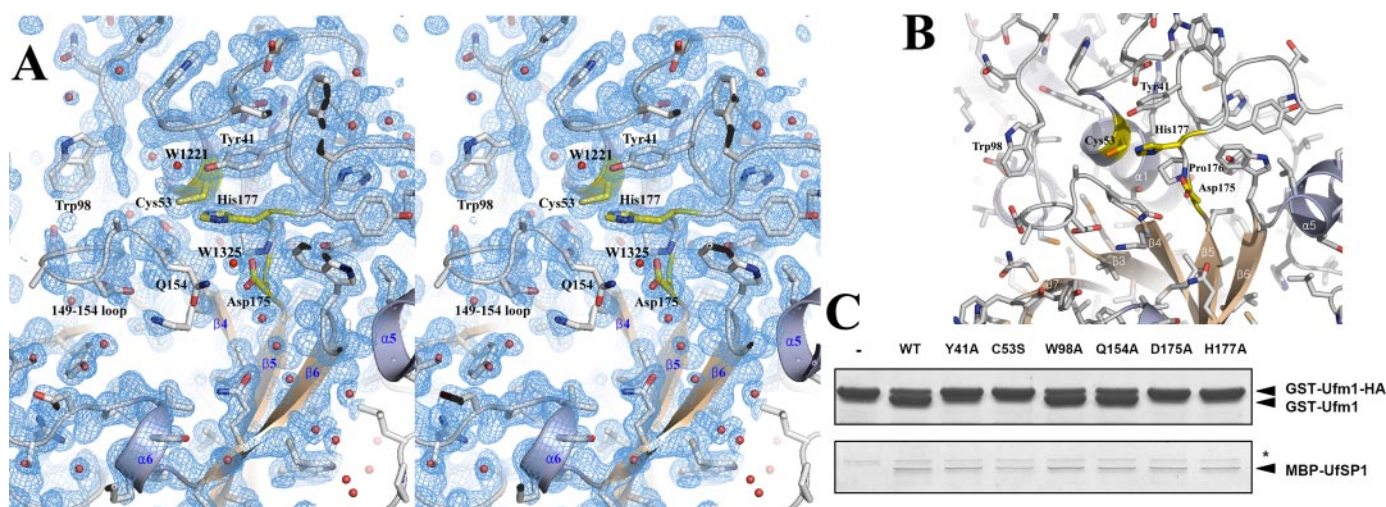


FIGURE 3. Active site of UfSP1. A, the electron density is a $2F_o - F_c$ map contoured at 1.0σ in stereo. The electron density map is clear for all atoms, except for residues 149–154, and the side chain of Trp⁹⁸ can be modeled in two positions. B, the catalytic triad is formed by Cys⁵³, Asp¹⁷⁵, and His¹⁷⁷, and Tyr⁴¹ participates in the formation of the oxyanion hole crucial for catalysis. Although the catalytic triad residues are located in a similar geometry as those from papain and some DUBs, the aspartate and histidine are juxtaposed with a proline in between. This “Asp-Pro-His” is a signature of new subfamily of cysteine protease superfamily. Similar arrangements were observed in M48^{USP} and Atg4B. C, *in vitro* processing activities of C53S, D175A, H177A, Y41A, Q154A, and W98A mutants. Proteins were produced as MBP-UfSP1, and GST-Ufm1-HA was used as substrate. WT, wild type.

fingers, palm, and thumb, and its active site is located between the palm and thumb. In this case, however, binding of Ub aldehyde induces a drastic conformational change in the active site that realigns the catalytic triad residues for catalysis (32). The root mean square deviations for the core structures are 2.8 Å for 138 C α atoms in papain, 2.9 Å for 169 C α atoms in Atg4B, 3.9 Å for 137 C α atoms in M48^{USP}, and 4.0 Å for 121 C α atoms in HAUSP.

Active Site of UfSP1—Cys⁵³ and His¹⁷⁷ of UfSP1 were identified as catalytic residues earlier (18). They are part of highly conserved stretches (*i.e.* **Gly-Trp-Cys**⁵³-Ala/**Gly-Tyr-Arg-Ser/Thr**, and **Ile-Leu-Asp**¹⁷⁵-**Pro-His**¹⁷⁷-**Tyr-Thr/Trp-Gly**, in which strictly conserved residues are shown in boldface type or else sequences are shown in the order of UfSP1/UfSP2. As shown in Fig. 3A, Cys⁵³ is located at the N-terminal end of the α 1 helix, and His¹⁷⁷ is situated at the loop between the strand β 5 and helix α 5. Indeed, the structure shows that the Ne2 atom in the imidazole ring of His¹⁷⁷ is only 3.4 Å away from the Sy atom of the catalytic Cys⁵³ and is further hydrogen-bonded to a water molecule (W1325) through N δ 1. The Asp¹⁷⁵ is 4.5 Å away from the N δ 1 of His¹⁷⁷ in the current structure, but with a simple χ 1 and χ 2 rotation of His¹⁷⁷ it can easily make a hydrogen bond to the side chain oxygen of Asp¹⁷⁵. Thus, Cys⁵³, His¹⁷⁷, and Asp¹⁷⁵ are reasonably well posed for catalysis. The proline residue between Asp¹⁷⁵ and His¹⁷⁷ plays a role in ensuring the proper positioning of the two. In addition to the catalytic triad, cysteine proteases have another conserved residue, typically Asn or Gln, which participates in the formation of the oxyanion hole that is critical for catalysis. In UfSP1, the hydroxyl oxygen of Tyr⁴¹ makes a hydrogen bond to a water molecule (W1221), which in turn makes a hydrogen bond to the amide backbone of Cys⁵³. This tyrosine is conserved throughout UfSP1 and UfSP2. Therefore, the residues involved in catalysis are Cys⁵³, His¹⁷⁷, Asp¹⁷⁵, and Tyr⁴¹, and the catalytic mechanism of UfSP1 appears to be parallel to that of the papain family of cysteine proteases.

This was somewhat surprising, since the catalytic triad of canonical cysteine proteases is such that cysteine is located at the N terminus of an α -helix, and histidine and aspartate are at the end of two adjacent β -strands that form the central β -sheet. In papain, the residues corresponding to the triad are Cys²⁵, Asn¹⁷⁵, and His¹⁵⁹, whereas Gln¹⁹ participates in forming the oxyanion hole. In UfSP1, the position of canonical histidine is occupied by Gln¹⁵⁴, and this is histidine in all known UfSP2 and alanine in *Drosophila* UfSP1. This suggests that it is unlikely that Gln¹⁵⁴ participates in catalysis. In the current structure, the electron density in this region (residues 149–154) is not as well defined but enough to trace the chain, and the side chain of Gln¹⁵⁴ is facing away from the active site cysteine and histidine.

In order to validate our structural analysis, we employed site-directed mutagenesis to alter the key residues in UfSP1 predicted to participate in catalysis. Alanine mutants were made for Tyr⁴¹, Gln¹⁵⁴, Asp¹⁷⁵, His¹⁷⁷, and Trp⁹⁸ in addition to a serine mutant of Cys⁵³. *In vitro* enzymatic activity was assessed by using GST-Ufm1-HA as a substrate. As shown in Fig. 3C, C53S, H177A, D175A, and Y41A mutants show no activity, whereas Q154A exhibits activity, therefore confirming the conclusions from the structural data. Since the residues at the active site are highly conserved, UfSP2 is expected to have the same active site and is expected to operate via the same mechanism. It is striking to note that the active sites of Atg4B (29, 30) and M48^{USP} (31) have similar arrangements of the active site residues; namely, the catalytic histidine and the aspartate residues are from “Asp-Pro-His,” which is located at the tip of a β -strand. The oxyanion position is filled by a glutamine (Gln¹⁰) in the case of Atg4B and a tyrosine (Tyr⁵⁴) in M48^{USP}.

Interaction of UfSP1 with Ufm1—In order to examine Ufm1 binding to UfSP1, we first analyzed the solution structure of mouse Ufm1 by NMR. As reported previously (17), most of the backbone resonances of Ufm1 could be assigned except for the two residues at the N terminus, residues 9–18, which correspond to the flexible loop, and the two residues at the C termi-

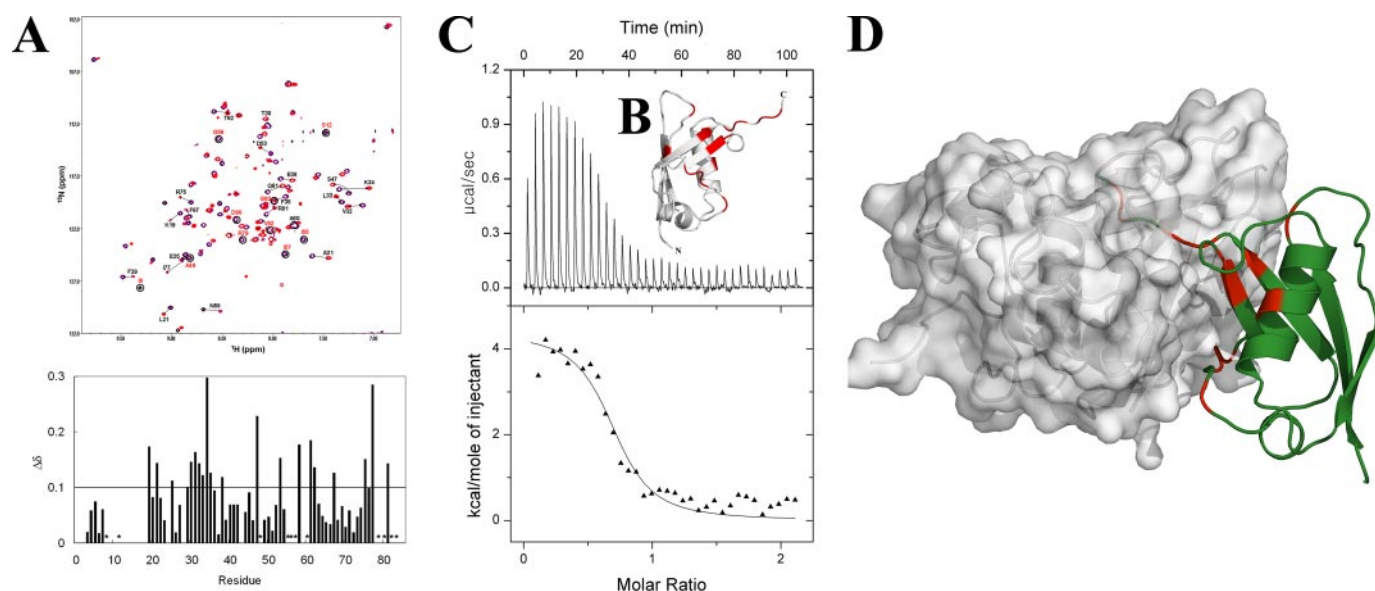


FIGURE 4. Ufm1 binding to UfSP1. *A*, chemical shift perturbations of Ufm1 upon UfSP1 binding using HSQC. The *top* shows the resonances of free Ufm1 in *black* and Ufm1 complexed with UfSP1 in 1:0.5 in *purple* and 1:1 in *red*. The *bottom* shows the weighted average of chemical shift perturbation ($\Delta\delta$) calculated on the individual residues. *B*, Ufm1 with significant chemical shift perturbation ($\Delta\delta > 0.2$) is marked in *red*. *C*, calorimetric titration of UfSP1 with Ufm1 using ITC. The *top* shows the heat effects associated with the injection of Ufm1 into the calorimetric cell containing UfSP1, whereas the *bottom* shows the corresponding binding isotherm and the best fitted curve. The binding is endothermic, and the analysis of the data yields a binding affinity of $K_D = 1.6 \mu\text{M}$. *D*, manual docking of Ufm1 onto UfSP1.

nus. Second, the interaction between UfSP1 and Ufm1 was characterized by a chemical shift change in the two-dimensional ^1H - ^{15}N TROSY experiment using ^{15}N -labeled Ufm1 with unlabeled UfSP1. Fig. 4*A* shows the overlay of the HSQC spectra of Ufm1 in the free form and complexed to the C53S mutant of UfSP1. Upon the addition of UfSP1, most of the backbone resonances of Ufm1 changed, and this was completed at a 1:1 molar ratio of Ufm1 and UfSP1. In fact, when the molar ratio of Ufm1 and UfSP1 is 1:0.5, most of the residues showed two resonances (shown in *purple* in Fig. 4*A*), which represents both the unbound and the bound states. The *bottom* of Fig. 4*A* shows the weighted average of chemical shift perturbation ($\Delta\delta = (\delta\text{H}^2 + \delta\text{N}^2/5)/[\text{suf}(\text{r}(\text{a}(\text{x}),1,2)]$) calculated on the individual residues. We also tried an NMR titration experiment using ^{15}N -labeled UfSP1 with the addition of Ufm1. Although ^1H - ^{15}N TROSY experiments gave well dispersed amide proton-nitrogen correlations (data not shown), we were not able to assign the resonances of UfSP1. Nevertheless, we were able to confirm the binding between UfSP1 and Ufm1 using ^{15}N labeled UfSP1 as a probe for the chemical shift perturbation experiment (data not shown).

Residues whose resonances completely disappeared include Ile⁸, Ser¹², Ala⁴⁸, Ile⁵⁵, Gly⁵⁶, Ile⁵⁷, and the residues from Arg⁷⁹ to Gly⁸³ excluding Arg⁸¹, and the residues with $\Delta\delta$ of >0.1 are Lys¹⁹, Leu²¹, Phe²⁹-Lys³⁴, Glu³⁸, Ser⁴⁷, Asp⁵³, Asn⁵⁸, Gln⁶¹, Thr⁶², Phe⁶⁷, Arg⁷⁵, Ile⁷⁷, and Arg⁸¹. When these residues are mapped on the Ufm1, they are mostly located on one side of the Ufm1 surface, forming a somewhat contiguous region, as seen in Fig. 4*B*. These chemical shift perturbations may be the result of direct interactions between Ufm1 and UfSP1 or indirect structural changes due to complex formation. But the results strongly suggest that the C-terminal part of $\beta 3$ and the loop between the $\beta 3$ and $\alpha 2$ of Ufm1 play an important role in UfSP1 binding. Also, the intermediate or slow exchange time scale

observed suggests that the interactions between Ufm1 and UfSP1 are relatively strong.

ITC was carried out using C53S mutant of UfSP1 and Ufm1, and the results of a calorimetric titration of Ufm1 into the UfSP1 are shown in Fig. 4*C*. The *top panel* shows the heat effects associated with the injection of Ufm1 into the calorimetric cell containing UfSP1, whereas the *bottom panel* shows the binding isotherm corresponding to the data in the *top panel* and the best fitted curve. Almost identical results were obtained when UfSP1 was added to Ufm1 (data not shown). The binding is endothermic, and analysis of the data yields a binding affinity of $K_D = 1.6 \mu\text{M}$. The dissociation constants (K_D) for SENP1, using the C604A mutant, were reported to be 0.79 and 0.49 μM toward the full-length SUMO1 and SUMO2, respectively (35).

DISCUSSION

Another Cysteine Protease, UfSP1—Among the five different families of proteases described for DUBs so far, cysteine protease is the most common (36). A large number of them have conserved Cys and His boxes, and the Cys box contains a catalytic cysteine residue, which is thought to undergo deprotonation before the nucleophilic attack on the carbonyl carbon atoms of the incoming substrate. This deprotonation is assisted by histidine, which, in turn, is stabilized by either aspartate or asparagine. Although ULP family proteases have different topology, their catalytic triad residues are located in the same geometry as those of papain.

The x-ray structure of UfSP1 shows that it also has a papain-like fold, with the catalytic residues being Cys⁵³, His¹⁷⁷, Asp¹⁷⁵, and Tyr⁴¹, as confirmed by mutagenesis analysis. The arrangement of the catalytic residues is similar to that observed in the canonical catalytic triad of cysteine proteases. However, unlike the canonical triad, the aspartate and histidine residues in the active site of UfSP1 are from the highly

conserved "Asp-Pro-His" box. This unique arrangement is also found in the recently reported structure for Atg4B, with the catalytic residues being Cys⁷⁴, Asp²⁷⁸, His²⁸⁰, and Tyr⁵⁴ (29, 30), and M48^{USP}, with the catalytic residues being Cys²³, Asp¹⁵⁶, and His¹⁵⁸ with Gln¹⁰ (31). Although there is a histidine (His¹⁴¹) at the position equivalent to the canonical histidine in the case of M48^{USP}, it is not conserved. Therefore, in the two enzymes, the catalytic aspartate and histidine come from a strictly conserved "Asp-Pro-His" box as well. Collectively, these results indicate that UfSP1 together with Atg4B and M48^{USP} form a new subfamily of the cysteine protease papain superfamily.

Ufm1 Binding to UfSP1—Both NMR and ITC data suggest that there is relatively strong interaction between UfSP1 and Ufm1, and this interaction involves not only the C-terminal of Ufm1 but also the residues of β 3 and the loop between β 3 and α 2, including both hydrophobic and polar residues. The endothermic nature of the interaction shown by ITC data suggests that the entropy change has to not only compensate for the unfavorable enthalpy change but also provide necessary binding energy to achieve the observed binding affinity. Main factors contributing to the entropy change would be lost in conformational degree of freedom as well as from the solvation entropy coming from the release of water molecules upon binding. Upon the addition of UfSP1, most of the backbone resonances of Ufm1 changed. These changes may well account for the necessary entropy change.

It is difficult to predict structural changes in the complex structure of UfSP1 and Ufm1, since in some complexes rather large changes are observed in DUBs or ULPs upon Ub or UBL binding (19, 32). For example, in the case of USP14, in the free form, the active site is blocked by two surface loops, but upon Ub binding, a significant conformational change occurs that translocates the two surface loops, thereby allowing the C terminus of Ub access to the active site (19). In the case of UfSP1, since it has a relatively open configuration at the active site, and it lacks an analogous crossover loop seen in UCH enzymes, it is most likely that UfSP1 could allow large protein conjugates to come within a distance suitable for recognition of the isopeptide bond by the active site. Assuming this, we tried manual docking of the Ufm1 onto the active site without any modification. Since the sequence of mouse Ufm1 is identical to that of human Ufm1, we used the earlier reported NMR structure (Protein Data Bank code 1WXS (18)), and the results from the peak shift as well as the surface electrostatic potentials of both Ufm1 and UfSP1 were used as a general guide. Fig. 4D shows the result.

In the resulting model, the surfaces of the two proteins are reasonably complementary to each other. For example, the hydrophobic patch around residues Ala⁴⁸ and Ile⁴⁹ of Ufm1 complements the Val¹⁰², Leu¹⁰⁶, and Leu¹⁰⁸ of UfSP1, and Ile⁷⁷ of Ufm1 faces Phe²⁰⁸ and His¹²⁰ of UfSP1. These residues of UfSP1 are not strictly conserved, but they are substituted by amino acids with hydrophobic side chains in UfSP2 (see Fig. 1B). Also, Arg⁷⁹ of Ufm1 comes near Asp⁸⁷ or Glu¹⁰³ of UfSP1. In this case, both Asp⁸⁷ (Gly-Asp⁸⁷-Lys-Pro) and Glu¹⁰³ of UfSP1 are conserved, suggesting a possible formation of a salt bridge. This model puts the C terminus of Ufm1 near the shal-

low and narrow "cleft" formed by the loop between α 2 and α 3 and the loop between β 3 and β 4 of UfSP1. The residues between α 2 and α 3 are generally conserved, whereas the residues on the loop between β 3 and β 4 are not (see Fig. 1B). Especially, C α of Trp⁹⁸ (in Gly-Ser-Arg-Gln/Asn-Trp⁹⁸-Ile-Gly) is 6.4 Å away from the C α of Cys⁵³ (see Fig. 3, A and B), and it is tempting to suggest that it may play a role in stabilizing the substrate binding, although the mutagenesis study on Trp⁹⁸ showed no significant effect on the proteolysis (see Fig. 3C). In the current crystal structure, there is not enough room at the cleft to accommodate the C terminus of Ufm1. However, noting that residues 149–154 are somewhat disordered in the current model, one can easily imagine rearrangement of these residues.

Conclusion—In this study, we reported the crystal structure of mouse UfSP1 at 1.7 Å resolution. Despite no obvious sequence homology with known proteases, the structure of UfSP1 shows a papain-like fold, with catalytic residues from Cys and Asp-Pro-His boxes. This is similar to what is seen for Atg4B and M48^{USP}. Thus, UfSP1 together with Atg4B and M48^{USP} seem to form a new subfamily of the cysteine protease superfamily. Since the catalytic domains of UfSP1 and UfSP2 share 36% sequence identity (45.6% similarity) with the mentioned residues conserved throughout, UfSP2 appears to also belong to this new family. The ITC data suggest that the K_D between Ufm1 and UfSP1 is 1.6 μ M. The NMR data also suggest that about 10 residues between β 3 and α 2, in addition to the 10 or so residues at the C terminus of Ufm1, may be important in binding of Ufm1 to UfSP1.

Acknowledgments—We thank the staffs at the 4A beamline of PAL, and the 900-MHz NMR facility at Korea Institute of Science and Technology (KIST) was supported by a KIST institutional grant.

REFERENCES

- Glickman, M. H., and Ciechanover, A. (2002) *Physiol. Rev.* **82**, 373–428
- Hershko, A., Ciechanover, A., and Varshavsky, A. (2000) *Nat. Med.* **6**, 1073–1081
- Pickart, C. M. (2004) *Cell* **116**, 181–190
- Kerscher, O., Felberbaum, R., and Hochstrasser, M. (2006) *Annu. Rev. Cell Biol.* **22**, 159–180
- Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987) *J. Biol. Chem.* **262**, 396–6399
- Galan, J. M., and Haguenaer-Tsapis, R. (1997) *EMBO J.* **16**, 5847–5854
- Terrell, J., Shih, S., Dunn, R., and Hicke, L. (1998) *Mol. Cell* **1**, 193–202
- Hofmann, R. M., and Pickart, C. M. (1999) *Cell* **96**, 645–653
- Deng, L., Wang, C., Spencer, E., Yang, L., Braun, A., You, J., Slaughter, C., Pickart, C., and Chen, Z. J. (2000) *Cell* **103**, 351–361
- Spence, J., Gali, R. R., Dittmar, G., Sherman, F., Karin, M., and Finley, D. (2000) *Cell* **102**, 67–76
- Hicke, L. (2001) *Nat. Rev. Mol. Cell Biol.* **2**, 195–201
- Jentsch, S., and Pyrowolakis, G. (2000) *Trends Cell Biol.* **10**, 335–342
- Welchman, R. L., Gordon, C., and Mayer, R. J. (2005) *Nat. Rev. Mol. Cell Biol.* **6**, 599–609
- Schwartz, A. L., and Ciechanover, A. (1999) *Annu. Rev. Med.* **50**, 57–74
- Chung, K. K., Dawson, V. L., and Dawson, T. M. (2001) *Trends Neurosci.* **24**, (suppl.) 7–14
- Komatsu, M., Chiba, T., Tatsumi, K., Iemura, S., Tanida, I., Okazaki, N., Ueno, T., Kominami, E., Natsume, T., and Tanaka, K. (2004) *EMBO J.* **23**, 1977–1986
- Sasakawa, H., Sakata, E., Yamaguchi, Y., Komatsu, M., Tatsumi, K., Komi-

- nam, E., Tanaka, K., and Kato, K. (2006) *Biochem. Biophys. Res. Commun.* **343**, 21–26
18. Kang, S. H., Kim, G. R., Song, M., Baek, S. H., Seol, J. H., Bang, O. S., Ovaa, H., Tatsumi, K., Komatsu, M., Tanaka, K., and Chung, C. H. (2007) *J. Biol. Chem.* **282**, 5256–5262
19. Hu, M., Li, D., Song, L., Jeffrey, P. D., Chernova, T. A., Wilkinson, K. D., Cohen, R. E., and Shi, Y. (2005) *EMBO J.* **24**, 3747–3756
20. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
21. Terwilliger, T. C., and Berendzen, J. (1996) *Acta Crystallogr. Sect. D* **52**, 749–757
22. Terwilliger, T. C. (2000) *Acta Crystallogr. Sect. D* **56**, 965–972
23. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D* **60**, 2126–2132
24. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., and Pannu, N. S. (1998) *Acta Crystallogr. Sect. D* **54**, 905–921
25. Collaborative Computational Project 4 (1994) *Acta Crystallogr. Sect. D* **50**, 760–763
26. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6**, 277–293
27. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR* **4**, 603–614
28. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233**, 123–138
29. Sugawara, K., Suzuki, N. N., Fujioka, Y., Mizushima, N., Ohsumi, Y., and Inagaki, F. (2005) *J. Biol. Chem.* **280**, 40058–40065
30. Kumanomidou, T., Mizushima, T., Komatsu, M., Suzuki, A., Tanida, I., Sou, U. S., Ueno, T., Kominami, E., Tanaka, K., and Yamane, T. (2006) *J. Mol. Biol.* **355**, 612–618
31. Schlieker, C., Weihofen, W. A., Frijns, E., Kattenhorn, L. M., Gaudet, R., and Ploegh, H. L. (2007) *Mol. Cell.* **25**, 677–687
32. Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J. W., Gu, W., Cohen, E., and Shi, Y. (2002) *Cell* **111**, 1041–1054
33. Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D., and Hill, C. P. (1997) *EMBO J.* **16**, 3787–3796
34. Misaghi, S., Galardy, P. J., Meester, W. J. N., Ovaa, H., Ploegh, H. L., and Gaudet, R. (2005) *J. Biol. Chem.* **280**, 1512–1520
35. Shen, L., Tatham, M. H., Dong, C., Zagorska, A., Naismith, J. H., and Hay, R. T. (2006) *Nat. Struct. Mol. Biol.* **13**, 1069–1077
36. Nijman, S. M., Luna-Vargas, M. P., Velds, A., Brummelkamp, T. R., Dirac, A. M., Sixma, T. K., and Bernards, R. (2005) *Cell* **123**, 773–786