THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 12, pp. 10248–10257, March 25, 2011

Structure of Ubiquitin-fold Modifier 1-specific Protease UfSP2*[S]

Received for publication, August 5, 2010, and in revised form, December 16, 2010 Published, JBC Papers in Press, January 12, 2011, DOI 10.1074/jbc.M110.172171

Byung Hak Ha[‡§1,2], Young Joo Jeon[§2], Sang Chul Shin[‡], Kanako Tatsumi[¶], Masaaki Komatsu[¶], Keiji Tanaka[¶], Christopher M. Watson[‖], Gillian Wallis[‖], Chin Ha Chung[§3], and Eunice EunKyeong Kim[‡4]

From the [‡]Life Sciences Division, Korea Institute of Science and Technology, Seoul 136-791, Korea, the [§]School of Biological Sciences, Seoul National University, Seoul 151-742, Korea, the [¶]Tokyo Metropolitan Institute of Medical Science, Tokyo 156-8506, Japan, and [‖]The Wellcome Trust Centre for Cell Matrix Research and School of Translational Medicine, University of Manchester, Manchester M13 9PT, United Kingdom

Ubiquitin-fold modifier 1 (Ufm1)-specific protease 2 (UfSP2) is a cysteine protease that is responsible for the release of Ufm1 from Ufm1-conjugated cellular proteins, as well as for the generation of mature Ufm1 from its precursor. The 2.6 Å resolution crystal structure of mouse UfSP2 reveals that it is composed of two domains. The C-terminal catalytic domain is similar to UfSP1 with Cys^{294}, Asp^{418}, His^{420}, Tyr^{282}, and a regulatory loop participating in catalysis. The novel N-terminal domain shows a unique structure and plays a role in the recognition of its cellular substrate C20orf116 and thus in the recruitment of UfSP2 to the endoplasmic reticulum, where C20orf116 predominantly localizes. Mutagenesis studies were carried out to provide the structural basis for understanding the loss of catalytic activity observed in a recently identified UfSP2 mutation that is associated with an autosomal dominant form of hip dysplasia.

Ubiquitin-fold modifier 1 (Ufm1) is a recently identified ubiquitin-like protein (UBL)[5] (1). It shares several common properties with ubiquitin (Ub) and other UBLs. It is synthesized as an inactive precursor protein composed of 85 residues, with two amino acids following a highly conserved glycine in its C terminus, the exposure of which is required for its subsequent conjugation. The NMR structure of Ufm1 shows a similar tertiary structure to Ub and other UBLs despite the fact that it shares very little sequence identity (2). However, Ufm1 displays different surface features from Ub and UBLs, suggesting that it may recognize different partners. It has been demonstrated that Ufm1 is ligated to a number of proteins in HEK293 cells and mouse tissues via a conjugation mechanism similar to that of Ub and UBLs. Mature Ufm1 is activated by a novel E1-like activating enzyme, Uba5, and then transferred to its cognate E2-like conjugating enzyme Ufc1. Recently, a Ufm1-specific E3 ligase, Ufl1, and its cellular substrate, C20orf116, have also been identified (3). Although the biological function of Ufm1 conjugation has yet to be identified, the fact that both Ufm1 and its conjugating system are conserved in both metazoans and plants suggests potential roles in various multicellular organisms.

Like Ub and UBLs, Ufm1 requires specific proteolytic cleavage to remove two C-terminal residues to become its mature form. Two cysteine proteases of different lengths, UfSP1 and UfSP2, have been identified (4). The longer UfSP2 is present in most, if not all, multicellular organisms, whereas the shorter UfSP1 is not found in plants or nematodes. These proteases are also responsible for the removal of Ufm1 from native intracellular conjugates. Neither protease shares sequence homology with any of the five categorized deubiquitinating enzymes identified thus far nor with any previously known proteases. However, the crystal structure of mouse UfSP1 at 1.7 Å resolution revealed a papain-like fold with a unique active site that is composed of Cys and an Asp-Pro-His conserved box instead of the canonical Cys-His-Asp triad, and this Cys and Asp-Pro-His configuration of the catalytic residues seems to form a new subfamily of the cysteine protease superfamily (5).

A mutation within the human *UFSP2* gene has recently been identified in a family with an autosomal dominant form of hip dysplasia, termed Beukes familial hip dysplasia (BFHD; MIM142669) (6), which is characterized by severe premature degenerative osteoarthritis of the hip joint.[6] The *UFSP2* mutation predicts the substitution of the highly conserved Tyr^{290} by His in the encoded protein. Sequence alignments indicated that the human *UFSP2* Tyr^{290} is equivalent to Tyr^{282} in the mouse and also corresponds to the highly conserved Tyr^{41} of mouse *Ufsp1*. The crystal structure of mouse UfSP1 suggested that Tyr^{41} plays a role in oxyanion hole formation. Interestingly, the Y282H substitution in UfSP2 abolished the *in vitro* Ufm1-processing activity of mouse UfSP2, whereas the corresponding Y41H mutation in mouse UfSP1 reduced but did not abolish the activity.[6]

* This work was supported in part by Functional Proteomics Center Grant FPR08B2-280, the 21C Frontier Research and Development Program of the Korea Ministry of Science and Technology, a Korea Institute of Science and Technology Institutional Program grant (to E. E. K.), Korea Research Foundation Grant KRF-2005-084-C00025, and Korea Science and Engineering Foundation Grant M10533010001-05N3301 (to C. H. C.).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S3.

The atomic coordinates and structure factors (code 3OQC) 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 a BK21 fellowship.

[2] Both authors contributed equally to this work.

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

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

[5] The abbreviations used are: UBL, ubiquitin-like protein; Ub, ubiquitin; ER, endoplasmic reticulum; BFHD, Beukes familial hip dysplasia; PDB, Protein Data Bank; TRITC, tetramethylrhodamine isothiocyanate.

[6] C. M. Watson and G. Wallis, manuscript in preparation.

Here, we report the crystal structure of mouse UfSP2 at 2.6 Å resolution, which shows a unique protein fold for the N-terminal domain linked to the catalytic domain that is similar to UfSP1. We also show that the novel N-terminal domain plays a role in the interaction with its cellular substrate C20orf116 and thus in the recruitment of UfSP2 to the endoplasmic reticulum, where C20orf116 almost exclusively resides. A comparison of the crystal structures of UfSP1 and UfSP2 coupled with the results from a series of mutagenesis experiments on both UfSP2 and UfSP1 defines the structural requirements for the substrate recognition and catalysis and explains the loss of activity of the UfSP2 mutation associated with BFHD.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The cDNAs for *Ufm1* (Swiss Protein Database code P61961) and *Ufsp2* (Swiss Protein Database code Q99K23) from mouse were cloned into pET28a (Novagen) to generate N-terminal His-tagged proteins. In the case of *Ufsp2*, because the expressed protein was cleaved at Lys^{94} as confirmed by N-terminal amino acid sequencing, we have replaced it with Arg at this position to avoid cleavage. In addition, we added another mutation of Arg^{128} to Ala to evade cleavage upon standing for crystallization. The resulting vectors were transformed to *Escherichia coli* BL21(DE3) codon plus RIL (Stratagene) cells. The histidine-tagged proteins were purified initially 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 and further by Mono Q and gel filtration on a Superdex 75 26/60 column (GE Healthcare). The purified UfSP2 was concentrated to 10 mg/ml in a buffer containing 20 mM MES (pH 6.5), 100 mM NaCl, and 1 mM DTT. Selenomethionine-substituted UfSP2 was generated as described previously (29).

Crystallization—Initial screening for the crystallization was carried out by using 96-well Intelli plates (Hampton Research), and Hydra II Plus One (MATRIX Technology) robotics system at 295 K yielded micro-crystals, and this was further optimized using the hanging drop methods. Diffraction quality crystals were obtained by mixing equal volumes of 10 mg/ml mouse UfSP2 in 20 mM MES (pH 6.5), 100 mM NaCl, and 1 mM DTT with a reservoir solution containing 0.04 M K_2HPO_4, 12% (v/v) PEG3350 in 3 days. The crystals of UfSP2 belong to the space group *C*2, with $a = 184.53$ Å, $b = 56.04$ Å, $c = 143.27$ Å, and $\alpha = \gamma = 90°$ and $\beta = 128.01°$, and it contains two molecules per asymmetric unit, corresponding to a Matthews volume V_m of 2.78 $Å^3\ Da^{-1}$. Attempts to crystallize the UfSP2 complex with Ufm1 did not yield crystals large enough to be suitable for high resolution data collection.

X-ray Data Collection and Processing—The x-ray diffraction data set from the native and selenomethionine crystals were collected at beamline 4A of Pohang Light Source, Pohang, Korea. Crystals were equilibrated in a cryoprotectant buffer containing reservoir buffer plus 30% (v/v) ethylene glycol and then flash-frozen in a cold nitrogen stream at 100 K prior to collection. Data were processed, integrated, and scaled by using HKL2000 program suite (30), and the statistics are summarized in Table 1.

Structure Determination and Refinement—The crystal structure of UfSP2 was determined by the multiple wavelength anomalous diffraction phasing method, because all attempts by molecular replacement using UfSP1 failed. Initially 9 out of 11 possible selenium sites were found, and eventually all selenium sites were refined; the initial phases were calculated using the programs SOLVE (31) and RESOLVE (32). About 54% of the residues were automatically modeled as a polyalanine chain by RESOLVE and further constructed using the molecular modeling program COOT (33). The refinement was then carried out using the CNS and REFMAC (34, 35) to an *R*-value of 23.8% and an R_{free} of 29.8%, and the final model included 6609 protein atoms and 107 water molecules. The final refinement statistics are summarized in Table 1.

Mutagenesis and in Vitro Proteolytic Assay—Site-directed mutagenesis and loop exchanges on the residues that might be involved in the catalysis were carried out using QuickChange site-directed mutagenesis kit (Stratagene) by following the manufacturer's instructions. Mutants of UfSP2 were produced as N-terminally His-tagged proteins with single point mutations at positions Tyr^{282}, Cys^{294}, Asn^{290}, Thr^{422}, and Met^{283}. The chimerical regulatory, upstream, and neighboring loops of UfSP2 were made by substitution with corresponding UfSP1 residues. The regulatory loop of UfSP2 (^{393}GGVLA^{397}) was replaced by ^{149}GDADAQS^{155} of UfSP1 and vice versa. Upstream and neighboring loop of UfSP2 basically having the quadruple mutant (Y282H/M283G/N290R/T422W) was exchanged. The ^{284}QDRI^{287} and ^{423}GAEDL^{427} of UfSP2 changed to ^{43}CDGL^{46} and ^{180}GTPKNR^{185} of UfSP1, respectively. *In vitro* proteolysis assay was performed using GST-Ufm1-HA as a model substrate of Ufm1 precursor as described previously (7). All proteolysis assays were performed by incubating appropriate amounts of UfSP enzymes with 6 μg of GST-Ufm1-HA at 37 °C, and the reaction was stopped by addition of SDS sampling buffer and analyzed using SDS-PAGE. The gels were then stained with Coomassie Brilliant Blue R-250.

Immunocytochemistry—HeLa cells were grown on coverslips and transfected with appropriate vectors. Two days after transfection, they were fixed by incubation for 10 min with 3.7% paraformaldehyde in PBS. Cells were washed three times with PBS containing 0.1% Triton X-100, permeabilized with 0.5% Triton X-100 in PBS for 5 min, and treated with 3% BSA in PBS for 1 h. They were then incubated for 1 h with appropriate antibodies. After washing with PBS containing 0.1% Triton X-100, cells were incubated for 1 h with FITC- or TRITC-conjugated secondary antibody in PBS containing 3% BSA. After washing, cells were observed using a confocal laser scanning microscope (LSM510; Carl Zeiss, Jena, Germany). Images were acquired using an 80× objective and then processed using Photoshop (Adobe Systems, Mountain View, CA).

Immunoprecipitation—For immunoprecipitation, cell lysates were prepared in 50 mM Tris-HCl (pH 7.4) containing 150 mM NaCl, 1 mM EDTA, 1 mM NEM, 1 mM sodium vanadate, 1 mM NaF, 1 mM PMSF, and 1× protease inhibitor mixture (Roche Applied Science). Cell lysates were incubated with appropriate antibodies for 1 h at 4 °C and then with protein A-conjugated agarose for the next 1 h.

TABLE 1

Data collection and crystallographic refinement statistics

Values in parentheses refer to the highest resolution shell. r.m.s.d. means root mean square deviation.

Data sets	Multiple wavelength anomalous diffraction			Native
	Peak	Edge	Remote	
Beam line	BL-6C/BL-4A			
X-ray wavelength	0.97947 Å	0.979613 Å	0.97177 Å	1.0 Å
Energy	12658.3 eV	12656.4 eV	12758.5 eV	12398.9 eV
Resolution range	50 to 2.8 Å	50 to 2.8 Å	50 to 2.8 Å	50 to 2.6 Å
Space group	C2	C2	C2	C2
Unit cell parameters	a = 123.130 Å b = 63.623 Å c = 100.669 Å α = 90° β = 117.7° γ = 90°	a = 123.145 Å b = 63.621 Å c = 100.668 Å α = 90° β = 117.695° γ = 90°	a = 123.152 Å b = 63.617 Å c = 100.677 Å α = 90° β = 117.687° γ = 90°	a = 184.533 Å b = 56.041 Å c = 143.269 Å α = 90.0° β = 128.013° γ = 90.0°
Total/unique reflections	914,333/17,170	920,752/17,139	949,940/17,199	530,451/35,942
Completeness	92.2% (81.2%)	91.7% (80.6%)	91.4% (80.3%)	98.5% (96.6%)
Mean $I/\sigma(I)$	9.3% (1.8%)	8.3% (1.7%)	7.8% (1.7%)	11.0% (2%)
R_{merge} [a]	10.9% (25.4%)	10.6% (25.0%)	10.8% (28.0%)	10.6% (39.2%)
Refinement statistics				
Resolution range				46.1 to 2.60 Å
R/R_{free} [b]				23.8/29.8%
No. of protein atoms				6609
No. of water molecules				107
Average B-factor				42.0 Å^2
r.m.s.d. from ideal geometry: bond length/angle				0.015 Å/1.72°
Ramachandran analysis				
Favored region				84.2%
Additionally allowed				15.4%
Generously allowed				0.1%
Disallowed regions				0.3%

[a] $R_{merge} = \Sigma_h\Sigma_i|I(h,i) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I(h,i)$, where $I(h,i)$ is the intensity of the ith measurement of reflection h, and $\langle I(h)\rangle$ is the mean value of $I(h,i)$ for all i measurements.
[b] R_{free} is calculated from the randomly selected 10% set of reflections not included in the calculation of the R value.

RESULTS

Overall Structure of mUfSP2—UfSP2 from mouse, containing 461 amino acids, was crystallized, and the structure was determined using multiple wavelength anomalous dispersion data collected from the selenomethionine-substituted UfSP2 and refined at 2.6 Å resolution. Crystallized UfSP2 was mutated at three positions: C294S, K94R, and R128A to void cleavage during expression and crystallization, and all atoms were well defined in the electron density map except for the residues 53–55, 62–64, 81–102, and 117–133. Table 1 summarizes statistics on the crystallographic data. The overall structure has dimensions of 80 × 50 × 55 Å and consists of two domains, with the first domain composed of the 240 residues at the N terminus and the second domain consisting of the 200 residues at the C terminus as seen in Fig. 1*A*. The two domains are connected by a linker of about 20 residues, and the C-terminal tail forms additional interactions with the N terminus at the interface. The N-terminal domain, which is shaped like a rectangular box of 40 × 40 × 20 Å, has a six-stranded β-sheet with five helices. The strands are in the order of β2-β3-β1-β4-β5-β6, and the two long helices, α1 and α3, are on the somewhat concave face of the sheet, running diagonal to it, whereas the third helix, α4, packs at one end of the β-sheet (β5-β6) in the same direction as the sheet. The inner surface of the β-sheet facing the α-helices is highly hydrophobic, whereas the opposite side of the β-sheet is somewhat polar. The two long helices are amphipathic with the hydrophobic surfaces facing the β-sheet. The linker residues are practically packed against the N terminus of the β-sheet on the opposite side of the two parallel helices. The catalytic C-terminal domain has seven α-helices and seven β-strands and resembles the papain-like structure that was previously reported in mouse UfSP1 (7), and the catalytic triad is positioned on the surface cleft of the opposite side of the N-terminal domain. In this crystal form, there are two UfSP2 molecules per asymmetric unit and the two are related by a noncrystallographic 2-fold and show a root mean square deviation of 1.6 Å (supplemental Fig. S1).

As expected, comparison of UfSP2 with other structures in the Protein Data Bank using the DALI algorithm (8) yielded UfSP1 as the most significant match. The Z-score was 27.2. The next highest matches were Atg4B (PDB code 2CY7 (9) and 2D1I (10)) with a Z-score of 12.7 and murine cytomegalovirus protease M48^{USP} (PDB code 2J7Q) (11) with a score of 7.6. Atg4B is an essential enzyme in autophagy that cleaves nascent Atg8 at its C-terminal arginine residue and deconjugates Atg8 family proteins from a small adduct, phosphatidylethanolamine (12). Other deubiquitinating enzymes such as USP14 (PDB code 2AYN) (13) and OTU1(PDB code 3C0R) (14) had much lower Z-scores, 4.6 and 3.6, respectively. When the N-terminal domain of UfSP2 alone was tested, there were no significant hits on structural similarity search using either DALI or TM-align (15). The highest similarities were found in a putative lipoprotein B and major histocompatibility complex (MHC) class I molecules but with Z-scores less than 5.

The catalytic domain of UfSP2 shows almost identical overall structure with UfSP1, as expected from the sequence identity of 36%. However, there are significant differences as indicated by the root mean square deviation of 1.9 Å for 210 Cα atoms. Some regions show a root mean square deviation greater than 3 Å, but they are mostly on the surface and more than 15 Å away from the active site. One surprising point is that UfSP2 has more prominent secondary structures than

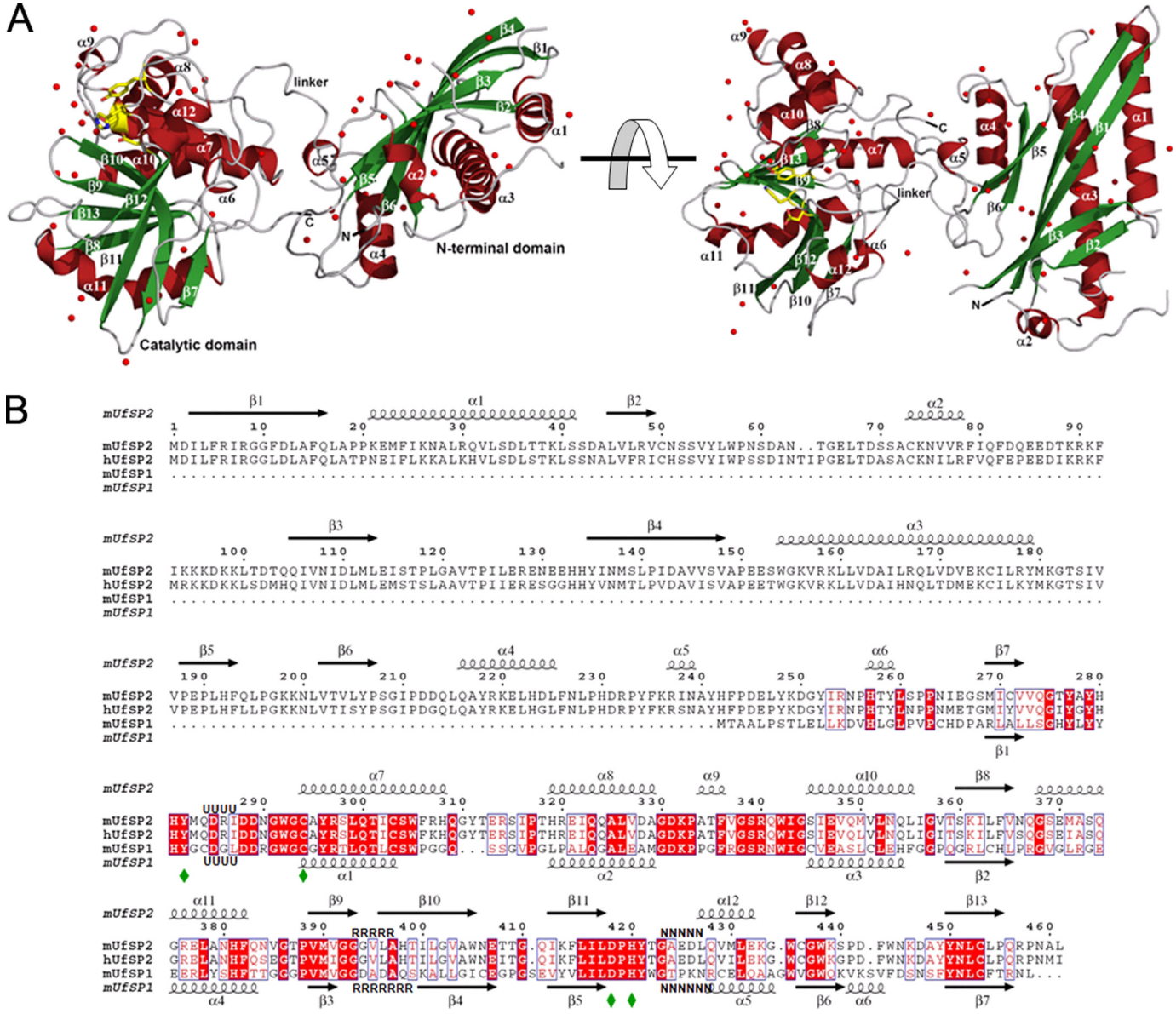

FIGURE 1. **Overall structure of UfSP2.** *A*, ribbon diagram of the overall structure of UfSP2 with *red* α-helices and *green* β-strands. The side chains of the catalytic residues are shown in *yellow*. *B*, amino acid sequence alignment of mouse UfSP1, UfSP2, and human UfSP2. The residues in the *red box* are strictly conserved, whereas the residues in the *blue box* are relatively conserved. Secondary structures are depicted *above* the sequences; *coils* indicate α-helices, and *arrows* indicate β-strands. The catalytic residues are indicated by *green diamonds* at the *bottom*, and the regulatory, upstream, and neighboring loops are indicated by *R*, *U*, and *N*, respectively.

UfSP1 (Fig. 1*B* and supplemental Fig. S2). The helix α7, which harbors the catalytic cysteine residue, is longer in UfSP2, and there is a three-residue insertion after the helix. The loop between α7 and α8 interacts with its α5 helix, which turns 180° in comparison with the loop of UfSP1. A stretch of residues after β8 reorganizes into a longer helix (α11) and is coupled with the changes in residues between β12 and β13. The C terminus of UfSP2 has three extra residues, and they make contact with the N-terminal domain, *e.g.* the backbone amide of Ala^{460} forms hydrogen bonds with the carboxyl group of Asp^{231}, and the C terminus of Leu^{461} forms hydrogen bonds to Arg^{219}.

Active Site of UfSP2—As seen in Fig. 2*A*, all the atoms near the active site are well defined in the electron density map. The catalytic Cys^{294} is located on the N terminus of an α-helix utilizing dipole moment, and "Asp^{418}–Pro^{419}–His^{420}" are located at the loop off a β-strand (Fig. 2*B*). This is the same as what is found in UfSP1 (5, 7). This differs from the canonical triad of the cysteine proteases where the Asp and His are located at two separate β-strands of the central β-sheet (16). His^{398} of UfSP2, which is part of a highly conserved stretch among UfSP2 (supplemental Fig. S3), is the canonical histidine position, thereby posing doubt on the identification of catalytic residues. However, when this residue was replaced by alanine, the *in vitro* activity of the mutant was the same as that of wild type UfSP2 (Fig. 2*C*). Indeed, in the crystal structure, His^{398} is too far from the catalytic Cys^{294}, *i.e.* 7.2 and 3.9 Å away from the Sγ atom of Cys^{294} and Nδ1 of His^{420}, respectively. Furthermore, superposition of papain shows that Leu^{396} is located at the canonical histidine position. The loss of activities by mutations of Cys^{294},

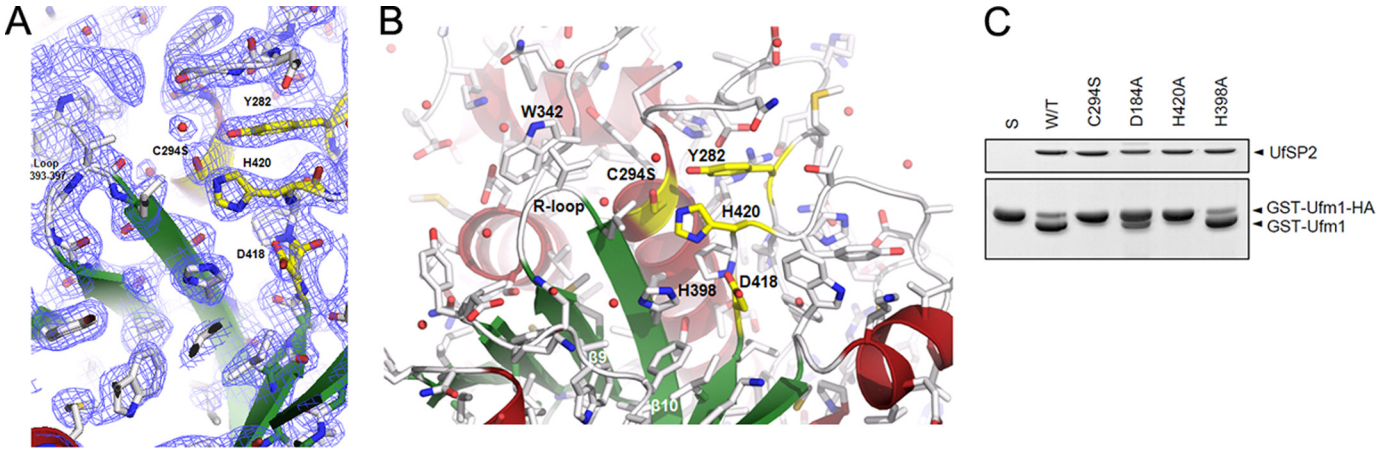

FIGURE 2. **Active site of UfSP2.** *A*, electron density around the active site. $2F_o - F_c$ map is contoured at 1.0σ. *B*, catalytic triad is formed by Cys^{294}, Asp^{418}, and His^{420} with Tyr^{282} participating in the formation of oxyanion hole. *C*, *in vitro* processing activities of C294S, D418A, H420A, and H398A mutants were assayed by incubating 1.5 μg of the proteins for 2 h as described under "Experimental Procedures." *S* denotes substrate (GST-Ufm1-HA), and *W/T* is for wild type UfSP2.

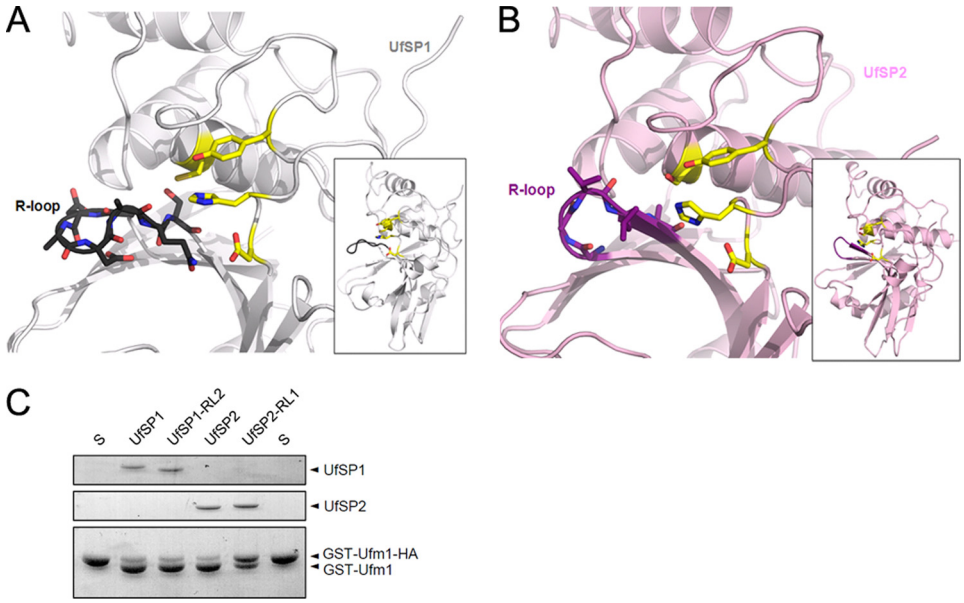

FIGURE 3. **Regulatory loop.** The R-loop of UfSP1 (*A*) and UfSP2 (*B*) are highlighted in *black* and *purple*, respectively, with the catalytic residues shown in *yellow*. *C*, *in vitro* processing activities of the chimeric UfSP1 and UfSP2 were assayed by incubating 1 μg of the proteins for 2 h. *S*, GST-Ufm1-HA; UfSP1-RL2, UfSP1 with the R-loop replaced by the R-loop of UfSP2; UfSP2-RL1, UfSP2 with the R-loop replaced by that of UfSP1.

Asp^{418}, and His^{420} further confirm the identification of the catalytic triad, and Tyr^{282} is responsible for the formation of the oxyanion hole (Fig. 2). The *in vitro* activity was assessed by using GST-Ufm1-HA as a model substrate for Ufm1 precursor as was used in previous studies (4).

Based on the crystal structure of UfSP1 and the NMR peak shifts in the UfSP1-Ufm1 complex, we predicted that the loop connecting β3 and β4 as well as Trp^{98} may play a role in Ufm1 recognition and/or stabilization (7). These correspond to the loop connecting β9 and β10 and Trp^{342} in UfSP2. This loop is referred to as the "R-loop" (regulatory loop) hereafter. On the outset, the R-loop in UfSP2 is slightly shorter than that of UfSP1 (Figs. 1*B* and 3). To test whether this loop indeed participates in Ufm1 recognition, it was mutated. The R-loop of UfSP2 was swapped with that of UfSP1, *i.e.* the residues ^{393}GGVLA^{397} in UfSP2 were replaced by the corresponding loop in UfSP1, namely ^{149}GDADAQS^{155}. As shown in Fig. 3*C*, both the chimera UfSP1 with UfSP2 R-loop (UfSP1-RL2) and wild type UfSP1 digested the substrate completely within 2 h, whereas the chimera UfSP2 with UfSP1 R-loop (UfSP2-RL1) showed limited activity. These results suggest that although this loop is not strictly conserved, it plays a role in the recognition of Ufm1 precursor.

The active site was further dissected to understand the lack of activity observed for the mouse equivalent *Ufsp2* mutation associated with BFHD (*Ufsp2* Y282H is the mouse equivalent of the human *UFSP2* Y290H BFHD-associated mutation). Of note was the finding that UfSP1 Y41H, which corresponds to UfSP2 Y282H, cleaved GST-Ufm1-HA at about a 3-fold lower rate than wild type UfSP1 (Fig. 4*C*). Because Y41H retained the enzymatic activity (although significantly reduced), the residues that are not conserved within 6 Å of the oxyanion hole Tyr of UfSP2 were examined. These residues included Met^{283}, Asn^{290}, and Thr^{422} of UfSP2, which correspond to Gly^{42}, Arg^{49}, and Trp^{179} of UfSP1, respectively (Fig. 4*A*). To the inactive Y282H UfSP2 mutant, additional mutations were introduced

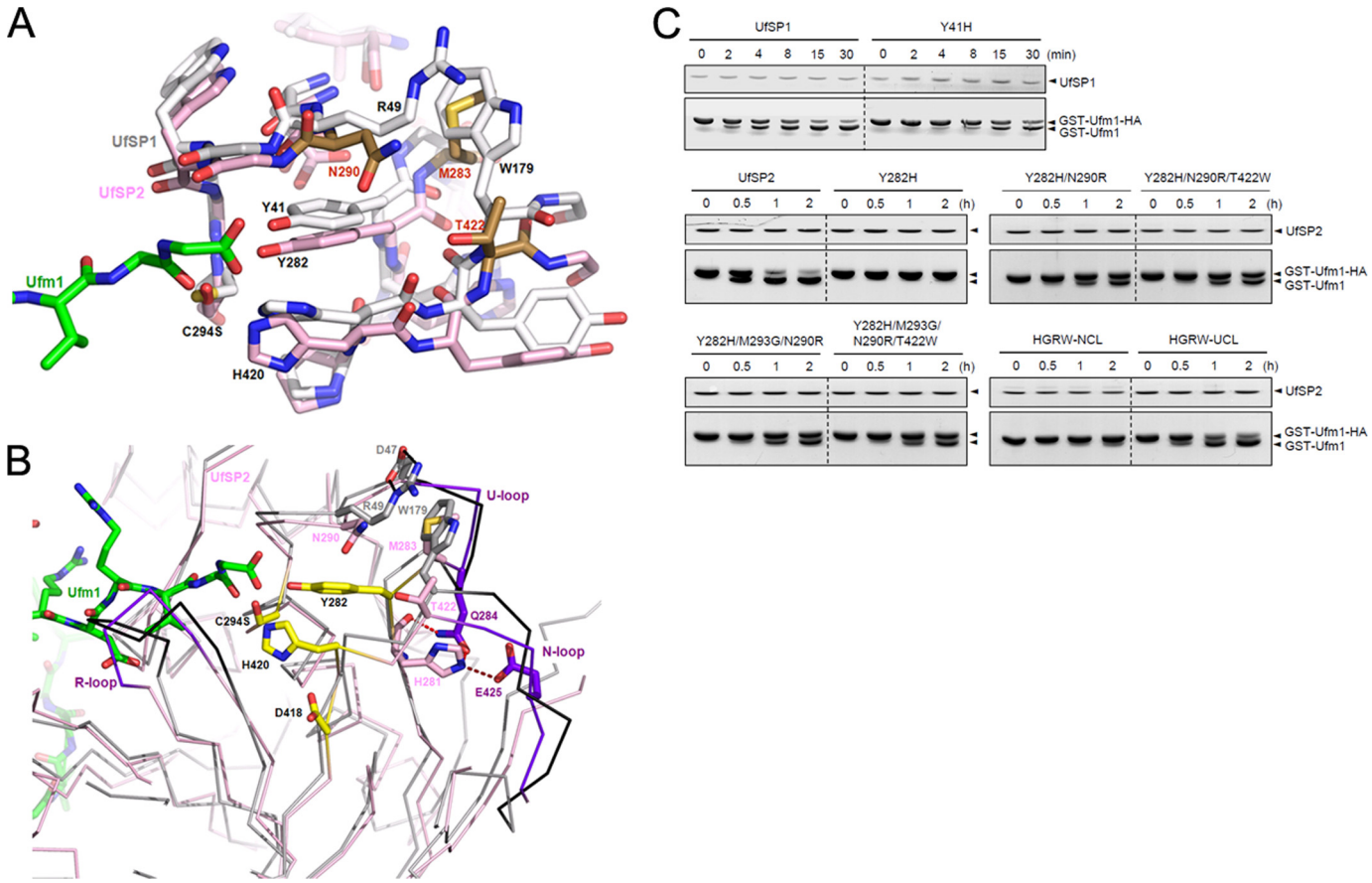

FIGURE 4. **Oxyanion hole mutation in BFHD.** *A,* comparison of the active sites of UfSP2 (*pink*) and UfSP1 (*gray*). Residues that are not conserved are highlighted in *brown* (Asn^{280}, Met^{283}, and Thr^{422} of UfSP2). Ufm1 is modeled based on earlier binding results (7) and is shown in *green*. *B,* comparison of UfSP2 (*pink*) and UfSP1 (*gray*) near the oxyanion hole. The regulatory, upstream, and neighboring loops of UfSP2 and UfSP1 are highlighted in *purple* and *black*, respectively. The side chain atoms of the catalytic residues and the mutants of UfSP2 are shown. *C, in vitro* processing activities of oxyanion hole mutants of UfSP2 (Y282H) and UfSP1 (Y41H) and various other mutants that rescue the activity for BFHD mutation were assayed by incubation of 0.2 μg of UfSP1 proteins and 1.5 μg of UfSP2 proteins for the indicated periods.

and tested for their *in vitro* activities against GST-Ufm1-HA to see whether the enzymatic activity gets restored, as was the case in UfSP1. As shown in Fig. 4*C*, incorporation of Arg at amino acid 290 resulted in a slight recovery of activity. Introduction of an additional mutation at position 422, *i.e.* Y282H/N290R/T422W led to a further recovery in enzymatic activity. A similar effect was observed in the case of Y282H/M293G/N290R triple mutant. However, incorporation of mutation at all four sites, *i.e.* Y282H/M283G/N290R/T422W, did not seem to yield additional enhancement. We then tested the possible effect of residues further away from Tyr^{282}, in two adjacent loops that differ from one another as seen in Fig. 4*B*. These residues correspond to Gln^{284}–Ile^{287} and Gly^{423}–Leu^{427} in UfSP2, and Cys^{43}–Leu^{46} and Gly^{180}–Arg^{185} stretches in UfSP1 (Fig. 1*B*), referred to as the "U-loop" (upstream loop) and "N-loop" (neighboring loop), respectively. When the U-loop of UfSP2 was replaced with the corresponding loops of UfSP1 in addition to the quadruple mutant above, it showed a dramatic restoration of enzymatic activity (*i.e.* the chimera of UfSP2 with its U-loop replaced by that of UfSP1 showed activity that was nearly the same as that of wild type UfSP2) (Fig. 4*C*). On the other hand, the replacement of the N-loop showed relatively little effect on activity.

Subcellular Localization of UfSP2—In an attempt to determine the role of the unique N-terminal domain of UfSP2, we first examined the subcellular localization of UfSP2 and its N- and C-terminal domains. As shown in Fig. 5, all are localized in both the nucleus and the cytoplasm. Significantly, a portion of the N-terminal domain in the cytoplasm appeared as speckles, raising the possibility that the N-terminal domain plays a role in the localization of UfSP2 in subcellular organelles, such as the ER and the Golgi apparatus (see below).

Recently C20orf116, a protein of unknown function, has been identified as a target for Ufm1 modification; whereas the newly identified Ufl1 serves as a Ufm1 E3 ligase for C20orf116, and UfSP2 deconjugates Ufm1 from its cellular substrate (3). It has also been found that C20orf116 predominantly localizes in the ER. In accordance with this finding, C20orf116 almost completely co-localized with calreticulin, a marker protein for the ER in HeLa cells but minimally with β-COP, a marker protein for Golgi bodies (Fig. 6). Co-expression with C20orf116 revealed that UfSP2 and its N-terminal domain, but not the C-terminal domain, strongly co-localize in the ER. These results suggest that C20orf116 possibly interacts with the N-terminal domain of UfSP2 and recruits it to the ER. To test this possibility, FLAG-C20orf116 was expressed in HeLa cells with and without Myc-tagged UfSP2 and its N- and C-terminal domains. Immunoprecipitation analysis by using an anti-FLAG antibody revealed that full-length UfSP2 and its N-terminal

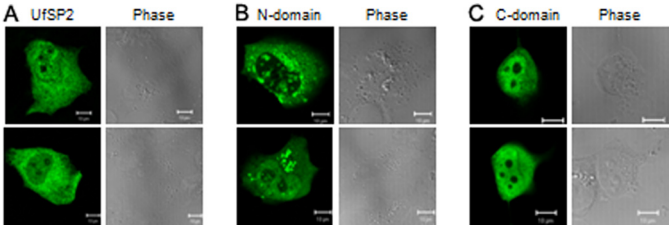

FIGURE 5. **Subcellular localization of UfSP2 and its N- and C-terminal domains.** Myc-tagged UfSP2 and its N- and C-terminal domains were expressed in HeLa cells and stained with anti-Myc antibody as described under "Experimental Procedures." *Bars,* 10 μm.

domain, but not its C-terminal domain, co-precipitated with C20orf116 (Fig. 7*A*), indicating that the N-terminal domain of UfSP2 interacts with C20orf116. To confirm this finding, Myc-C20orf116 was expressed in HeLa cells with and without His-Max-tagged UfSP2 and its N- and C-terminal domains. Pull-down analysis using Ni^{2+}-nitrilotriacetic acid-conjugated agarose showed that the amount of C20orf116 that co-precipitated with UfSP2 or its N-terminal domain was much higher than that with the C-terminal domain of UfSP2 (Fig. 7*B*). Collectively, these results suggest that the N-terminal domain of UfSP2 plays a key role in the recognition of C20orf116 and thus in the recruitment of UfSP2 to the ER, where C20orf116 is predominantly localized.

DISCUSSION

Most of the deubiquitinating enzymes that cleave ubiquitin or ubiquitin-like proteins from their precursors or protein conjugates contain not only the catalytic domains necessary for proteolytic activity but also additional N- or C-terminal extensions (17). Some of these extensions include ubiquitin binding domains, ubiquitin-like domains, and others that may participate in protein-protein interactions, but quite often they are not well characterized despite the fact that they may play an important role in modulating substrate specificity, cellular localization, or other physiological functions (18–20). In the case of UfSPs, unlike UfSP1, which only has a catalytic domain, UfSP2 has an N-terminal extension of about 250 residues that has no homologous proteins other than UfSPs when searched using BLAST. It is worthwhile mentioning that in some species such as *Drosophila*, rice, and *Caenorhabditis elegans,* the N-terminal domain is even longer, with about 100 extra residues (supplemental Fig. S3).

N-terminal Domain of UfSP2 Plays a Role in the Recognition of a Cellular Substrate—The crystal structure of mouse UfSP2 reveals a unique fold that is not found in any presently known cellular protein. The DALI search gave putative lipoproteins as the closest match; however, they are quite different in that these two proteins have a three stranded β-sheet with two α-helices of different lengths stacked next to each other. Again, the heavy chains of MHC class I are somewhat similar, as these have a more extensive β-sheet and two parallel α-helices of similar length. Yet the topology of the domain is quite different, and the groove between the two long helices (α1 and α3) is too narrow to fit anything like a peptide. It is worth mentioning that the sequence identity of the N-terminal domain is much lower than that of the catalytic domain, yet the residues of α4

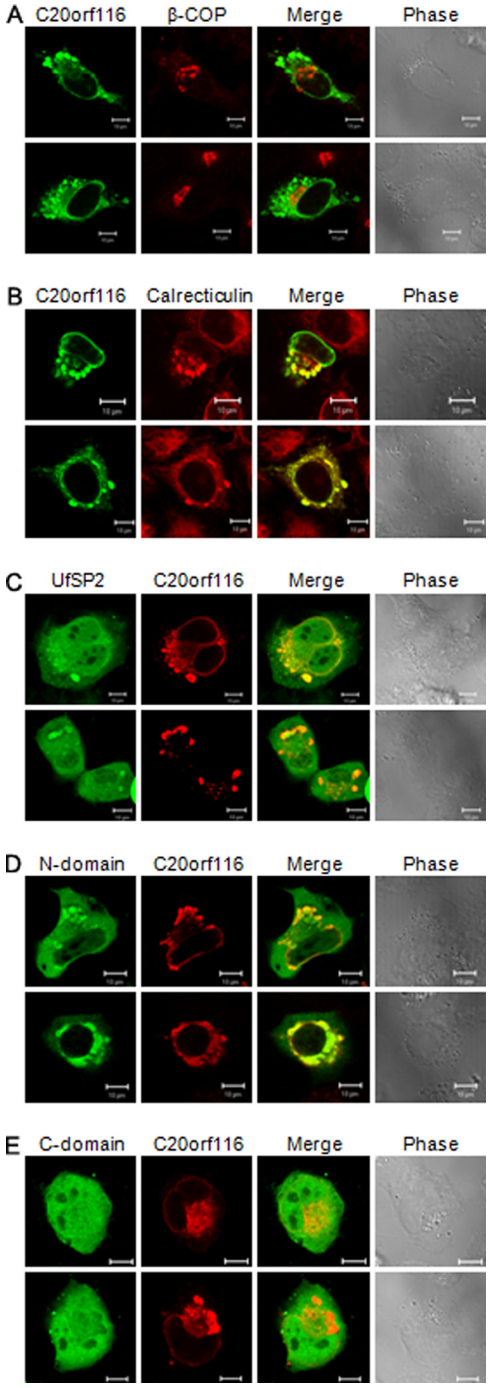

FIGURE 6. **Co-localization of UfSP2 and its N-terminal domain with C20orf116 in ER.** Myc-tagged UfSP2 and its N- and C-terminal domains were expressed in HeLa cells with FLAG-C20orf116. C20orf116 was stained with anti-FLAG antibody. Cells were also stained by anti-β-COP and anti-calreticulin antibodies. *Bars,* 10 μm.

facing the groove interface, as well as the beginning part of the linker, are conserved among all known forms of UfSP2 (supplemental Fig. S3).

Many deubiquitinating enzymes contain additional domains other than catalytic domains, and these extra domains have been suggested to function in the regulation of subcellular localization, substrate specificity, or physiological function (17, 18). The 2.6 Å resolution crystal structure of mouse UfSP2 reveals that it is composed of two domains that are connected

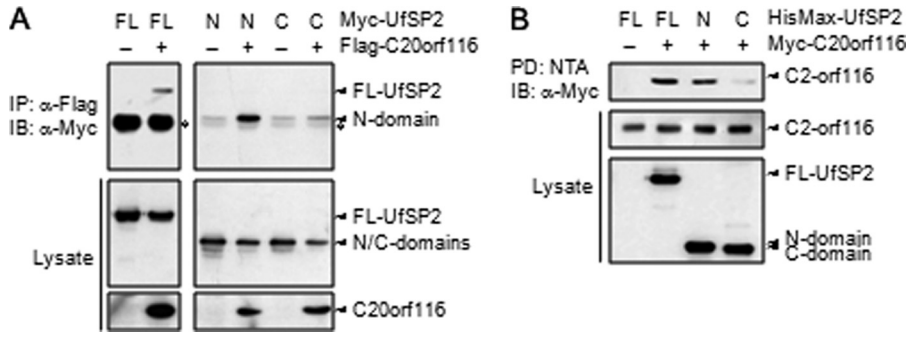

FIGURE 7. **Interaction of UfSP2 and its N-terminal domain with C20orf116.** *A,* Myc-tagged UfSP2 and its N (*N*)- and C (*C*)-terminal domains were expressed in HeLa cells with FLAG-C20orf116. Cell lysates were subjected to immunoprecipitation (*IP*) with anti-FLAG antibody followed by immunoblot (*IB*) with anti-Myc antibody. The lysates were also directly probed with the same antibodies. The *asterisks* indicate IgG heavy chain. *B,* HisMax-tagged UfSP2 and its N- and C-terminal domains were expressed in HeLa cells with Myc-C20orf116. Cell lysates were subjected to pull down with Ni^{2+}-nitrilotriacetic acid-conjugated resin followed by immunoblot with anti-Myc antibody. *FL*, full length.

by a 20-residue-long linker. C20orf116 is a cellular substrate of UfSP2 as well as a target protein for Ufm1 modification by the Ufm1 E3 ligase Ufl1 (3). Ectopically expressed C20orf116 interacts with UfSP2 and its N-terminal domain but much less with its C-terminal domain *in vivo*. These findings strongly suggest that the N-terminal extension of UfSP2 plays a critical role in the recognition of its cellular substrate C20orf116, and thus in the recruitment of UfSP2 to the ER, where C20orf116 predominantly localizes.

UfSP2 Shares the Same Catalytic Machinery as UfSP1—The catalytic domain of UfSP2 has a papain-like fold. Mutagenesis of the active site residues shows complete loss of activity for C294A and H420A mutants and some residual activity for the D418A mutant (Fig. 2*C*). Mutation of His^{398}, which is located near the canonical histidine position, to an alanine did not affect the enzyme activity as expected. These results confirm Cys^{294}, Asp^{418}, and His^{420} as the *bona fide* catalytic residues for UfSP2 as was previously suggested based on the structure of UfSP1 (7). Cys^{294} and His^{420} are directly involved in the reaction, serving as a nucleophilic attacking group and a general acid-base catalytic element, respectively, whereas Asp^{418} is essential in stabilizing the transition state through electrostatic interactions. The residual activity seen for the D418A mutant may be due to the aid of His^{398} residues (Fig. 2). In fact, another crystal form of UfSP2 having one molecule per asymmetric unit shows an additional water molecule between His^{420} and His^{398} that could make hydrogen bond linkage between the two (data not shown). It is worth mentioning that when the corresponding aspartate in UfSP1 (Asp^{175}) was mutated to an alanine, it showed a complete loss of activity. Additionally, the position of His^{398} is occupied by Lys^{156}, and the water molecule is not present in UfSP1.

UfSP2 Differs from UfSP1 in Ufm1 Recognition—In addition to the catalytic triad, we suggested that Trp^{98}, and the regulatory loop containing residues ^{149}GDADAQS^{155} in UfSP1, plays a role in stabilizing Ufm1 binding, *i.e.* Ufm1 recognition based on the binding study using NMR and modeling. In UfSP2, these correspond to Trp^{342} and ^{393}GGVLA^{397} (Figs. 3 and 4). Residues around Trp^{342}, which is part of a highly conserved stretch, practically have the same conformation as is found in UfSP1. However, the regulatory loop connecting β9 and β10 in UfSP2 is two residues shorter, and it takes up a somewhat different conformation. In the crystal structures there are differences between the two as shown in Fig. 3. In UfSP1, the loop is held in place via interactions with water molecules as well as the residues connecting α6 helix and β7 strand, whereas the corresponding loop in UfSP2 does not make significant interactions, except the hydrophobic interaction between Trp^{342} and Val^{395}. Among the UfSP2s these regulatory loops are highly conserved, and the V395A mutant shows reduced activity (data not shown).

To test how important the loop is in Ufm1 recognition, we swapped the regulatory loops between the two, *i.e.* the loop (residues 393–397) in UfSP2 was replaced by that of UfSP1 (residues 149–155), and vice versa. Although we expected some changes in enzymatic activities, the results were dramatic. In the case of UfSP1, when the loop is replaced with the shorter loop, there was no significant difference in activity; however, the UfSP2 chimera with a longer loop showed decreased enzymatic activity within 2 h (Fig. 3*C*). Therefore, the length, as well as the composition of this loop, seems to be important in Ufm1 processing. It is also worthwhile mentioning that UfSP1 is more active than UfSP2 in processing Ufm1 (4).

It is interesting to note that in the recently determined crystal structure of Atg4B complexed with LC3, a mammalian ortholog of yeast Atg8, the "regulatory loop," showed a large conformational change upon LC3 binding (21). In this case, the loop masking the entrance to the active site of free Atg4B is lifted by Phe^{119} of LC3, whereas overall structures are almost identical. In the case of UfSP2 and HAUSP, in addition to domain-wide conformational changes, reordering of the catalytic site upon Ub binding was reported (22, 23), whereas SENPs appear to require relatively minor local structural rearrangement at the catalytic site in response to the binding of SUMO (24).

Structural Basis for Inactivity Seen in UfSP2 Mutant of BFHD—During cysteine protease-mediated catalysis, the catalytic cysteine performs a nucleophilic attack on the carbonyl carbon of the scissile peptide bond. The histidine in the catalytic triad facilitates a proton transfer, and the aspartate stabilizes the transition state. In addition to the catalytic triad, another important component of the active site is the oxyanion hole, which is typically provided by glutamine/glutamate or asparagine (25, 26). In the case of both UfSP1 and UfSP2, the

oxyanion hole is formed by the backbone amide of catalytic cysteine and tyrosine. The reaction is completed by the attack of a water molecule, which results in the release of free Ufm1 molecules.

The identification of Tyr^{290} to histidine mutation associated with BFHD led us to consider the role of this residue and others in its vicinity during catalysis. First, when the equivalent Tyr^{282} in mouse UfSP2 was mutated to histidine, UfSP2 almost completely lost its catalytic activity, whereas the analogous Y41H mutation in UfSP1 retained decreased activity (Fig. 4*C*). In the crystal structure of UfSP1, the hydroxyl oxygen of Tyr^{41} makes a tight hydrogen bond to a water molecule (W1221) with a length of 2.8 Å, which in turn makes a hydrogen bond to the amide backbone of Cys^{53} in UfSP1. The water molecule is 3.2 Å away from the hydroxyl oxygen of Tyr^{282}, between two UfSP2s in the asymmetric unit. When Tyr^{41} is mutated to histidine, it is most likely that the water molecule is still within reasonable distance of the functional groups of histidine to stabilize the tetrahedral intermediate of Ufm1 as a part of the oxyanion hole in UfSP1, whereas it could be not in the case of UfSP2.

Next, the residues within 6 Å from Tyr^{282} that are not conserved, such as Met^{283}, Asn^{290}, and Thr^{422}, were mutated to test for their effects on the oxyanion hole. Mutations of Asn^{290} did not show much effect when replaced by alanine, lysine, or arginine. However, when it was mutated together with Thr^{422} and/or Met^{283}, the activity of the enzyme was partially restored. When the upstream loop spanning residues 282–289 of UfSP2, namely ^{282}**Y**MQ**D**RI**DD**^{289}, was replaced by that of UfSP1, namely ^{41}**Y**GC**D**GL**DD**^{48}, the enzymatic activity was restored to nearly the same level of wild type UfSP2, in which strictly conserved residues are shown in boldface type. This is the relatively conserved loop connecting the β-strand with the oxyanion hole and the helix with the catalytic cysteine. In UfSP1, Asp^{47} and Arg^{49} make two by two hydrogen bonds that are stacked by the side chain of Trp^{179}, shielding the side chain of Tyr^{41}. In UfSP2, Met^{283} occupies the side chain position of Trp^{179} of UfSP1, and Trp^{179} is replaced by Thr^{422}. Arg^{49} of UfSP1 is replaced by Asn^{290}, so Tyr^{282} is not as well shielded. In UfSP2, Gln^{284}, which is involved in the upstream loop, forms hydrogen bonds with the carbonyl oxygen and Nϵ2 of His^{281}. Ultimately, these configurations make the loop connected to Tyr^{282} less flexible and might restrict the movement of the side chain of this residue to adopt the available conformation for an oxyanion hole.

Conclusion—The crystal structure of UfSP2 reveals a two-domain structure with the N-terminal domain having a novel fold connected by a 20-residue-long loop to the catalytic C-terminal domain, which is similar to UfSP1 with Cys^{294}, Asp^{418}, His^{420}, Tyr^{282}, and a regulatory loop, that participates in catalysis. The novel N-terminal domain plays a role in the recognition of its cellular substrate C20orf116 and in the recruitment of UfSP2 to the endoplasmic reticulum, where the substrate predominantly localizes. Of some 80–90 deubiquitinating enzymes, a number of them are linked to physiological disorders, some by mutation, through altered expression levels, and/or as part of regulatory complexes (27), *e.g.* I93M mutation of ubiquitin C-terminal hydrolase-L1 (UCH-L1), which is a highly abundant neuronal enzyme that is associated with familial Parkinson disease (28). A roughly 50% reduction in catalytic activity resulted in that case. Although the exact mechanism of how UfSP2 inactivity relates to BFHD has yet to be identified, our mutagenesis results provide a structural basis for understanding the loss of catalytic activity that is observed in the UfSP2 mutant that is associated with BFHD.

Acknowledgments—We thank Drs. K. Joo and S.H. Kang for discussions and the staff at the 4A beamline of Pohang Light Source for help during data collection.

REFERENCES

1. 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
2. Sasakawa, H., Sakata, E., Yamaguchi, Y., Komatsu, M., Tatsumi, K., Kominami, E., Tanaka, K., and Kato, K. (2006) *Biochem. Biophys. Res. Commun.* **343,** 21–26
3. Tatsumi, K., Sou, Y. S., Tada, N., Nakamura, E., Iemura, S., Natsume, T., Kang, S. H., Chung, C. H., Kasahara, M., Kominami, E., Yamamoto, M., Tanaka, K., and Komatsu, M. (2010) *J. Biol. Chem.* **285,** 5417–5427
4. Kang, S. H., Kim, G. R., Seong, 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
5. Ha, B. H., and Kim, E. E. (2008) *BMB Rep.* **41,** 435–443
6. Roby, P., Eyre, S., Worthington, J., Ramesar, R., Cilliers, H., Beighton, P., Grant, M., and Wallis, G. (1999) *Am. J. Hum. Genet.* **64,** 904–908
7. Ha, B. H., Ahn, H. C., Kang, S. H., Tanaka, K., Chung, C. H., and Kim, E. E. (2008) *J. Biol. Chem.* **283,** 14893–14900
8. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
9. Sugawara, K., Suzuki, N. N., Fujioka, Y., Mizushima, N., Ohsumi, Y., and Inagaki, F. (2005) *J. Biol. Chem.* **280,** 40058–40065
10. Kumanomidou, T., Mizushima, T., Komatsu, M., Suzuki, A., Tanida, I., Sou, Y. S., Ueno, T., Kominami, E., Tanaka, K., and Yamane, T. (2006) *J. Mol. Biol.* **355,** 612–618
11. Schlieker, C., Weihofen, W. A., Frijns, E., Kattenhorn, L. M., Gaudet, R., and Ploegh, H. L. (2007) *Mol. Cell* **25,** 677–687
12. Ohsumi, Y. (2001) *Nat. Rev. Mol. Cell Biol.* **2,** 211–216
13. Hu, M., Li, P., Song, L., Jeffrey, P. D., Chenova, T. A., Wilkinson, K. D., Cohen, R. E., and Shi, Y. (2005) *EMBO J.* **24,** 3747–3756
14. Messick, T. E., Russell, N. S., Iwata, A. J., Sarachan, K. L., Shiekhattar, R., Shanks, J. R., Reyes-Turcu, F. E., Wilkinson, K. D., and Marmorstein, R. (2008) *J. Biol. Chem.* **283,** 11038–11049
15. Zhang, Y., and Skolnick, J. (2005) *Nucleic Acids Res.* **33,** 2302–2309
16. Kraut, J. (1977) *Annu. Rev. Biochem.* **46,** 331–358
17. Reyes-Turcu, F. E., Ventii, K. H., and Wilkinson, K. D. (2009) *Annu. Rev. Biochem.* **78,** 363–397
18. 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
19. Drag, M., and Salvesen, G. S. (2008) *IUBMB Life* **60,** 734–742
20. Reyes-Turcu, F. E., and Wilkinson, K. D. (2009) *Chem. Rev.* **109,** 1495–1508
21. Satoo, K., Noda, N. N., Kumeta, H., Fujioka, Y., Mizushima, N., Ohsumi, Y., and Inagaki, F. (2009) *EMBO J.* **28,** 1341–1350
22. Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J. W., Gu, W., Cohen, R. E., and Shi, Y. (2002) *Cell* **111,** 1041–1054
23. Renatus, M., Parrado, S. G., D'Arcy, A., Eidhoff, U., Gerhartz, B., Hassiepen, U., Pierrat, B., Riedl, R., Vinzenz, D., Worpenberg, S., and Kroemer, M. (2006) *Structure* **14,** 1293–1302
24. Reverter, D., and Lima, C. D. (2006) *Nat. Struct. Mol. Biol.* **13,** 1060–1068
25. Hedstrom, L. (2002) *Chem. Rev.* **102,** 4501–4524
26. Otto, H. H., and Schirmeister, T. (1997) *Chem. Rev.* **97,** 133–172
27. Singhal, S., Taylor, M. C., and Baker, R. T. (2008) *BMC Biochem.* **9,** Suppl. 1, S3

28. Leroy, E., Boyer, R., Auburger, G., Leube, B., Ulm, G., Mezey, E., Harta, G., Brownstein, M. J., Jonnalagada, S., Chernova, T., Dehejia, A., Lavedan, C., Gasser, T., Steinbach, P. J., Wilkinson, K. D., and Polymeropoulos, M. H. (1998) *Nature* **395,** 451–452
29. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) *EMBO J.* **9,** 1665–1672
30. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
31. Terwilliger, T. C., and Berendzen, J. (1996) *Acta Crystallogr. D Biol. Crystallogr.* **52,** 749–757
32. Terwilliger, T. C. (2000) *Acta Crystallogr. D Biol. Crystallogr.* **56,** 965–972
33. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
34. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921
35. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763