

## STRUCTURE OF UBIQUITIN-FOLD MODIFIER 1 SPECIFIC PROTEASE, UFSP2\*

Byung Hak Ha<sup>1,2,5</sup> Young Joo Jeon<sup>2,5</sup> Sang Chul Shin<sup>1</sup> Kanako Tatsumi<sup>3</sup> Masaaki Komatsu<sup>3</sup> Keiji Tanaka<sup>3</sup> Christopher M. Watson<sup>4</sup> Gillian Wallis<sup>4</sup> Chin Ha Chung<sup>2,6</sup> and Eunice EunKyeong Kim<sup>1,6</sup>

From <sup>1</sup>Life Sciences Division, Korea Institute of Science and Technology, Seoul, 136-791, Korea; <sup>2</sup>School of Biological Sciences, Seoul National University, Seoul, 151-742, Korea; <sup>3</sup>Tokyo Metropolitan Institute of Medical Science, Tokyo, 156-8506, Japan; <sup>4</sup>The Wellcome Trust Centre for Cell Matrix Research and School of Translational Medicine, The University of Manchester, Manchester M13 9PT, United Kingdom

*Running title: Ubiquitin-fold modifier 1 Specific Protease, UfSP2*

Address correspondence to: Eunice E. Kim Tel: +82-2-958-5937; Fax: +82-2-958-5909 E-mail: [eunice@kist.re.kr](mailto:eunice@kist.re.kr) or Chin Ha Chung Tel: +82-2-880-6693; Fax: +82-2-871-9193; E-mail: [chchung@snu.ac.kr](mailto:chchung@snu.ac.kr)

<sup>5</sup>These two authors contributed equally to this work

**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 Cys294, Asp418, His420, Tyr282 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) (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 multi-cellular 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, multi-cellular organisms, while 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, or 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 (6) (BFHD; MIM142669), that is characterized by severe premature degenerative osteoarthritis of the hip joint (Watson *et al.* in preparation). The UfSP2 mutation predicts the substitution of the highly conserved Tyr290 by His in the encoded protein. Sequence alignments indicated that the human UfSP2 Tyr290 is equivalent to Tyr282 in the mouse and also corresponds to the highly conserved Tyr41 of

mouse UfSP1. The crystal structure of mouse UfSP1 suggested that the Tyr41 plays a role in oxyanion hole formation. Interestingly, the Tyr282His substitution in UfSP2 abolished the *in vitro* Ufm1-processing activity of mouse UfSP2, while the corresponding Tyr41His mutation in mouse UfSP1 reduced but not abolished the activity (Watson *et al.* 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 Prot entry: P61961) and UfSP2 (Swiss Prot entry: Q99K23) from mouse were cloned into pET28a (Novagen) to generate N-terminal

His-tagged proteins. In case of UfSP2, since the expressed protein was cleaved at Lys94 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 Arg128 to Ala in order 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) equilibrated with 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 1 mM TCEP, and further by Mono Q and gel filtration on a Superdex75 26/60 column (GE). The purified UfSP2 was concentrated to 10 mg/ml in a buffer containing 20 mM MES (pH 6.5), 100 mM NaCl, and 1mM 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 295K 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<sub>2</sub>HPO<sub>4</sub>, 12% (v/v) PEG3350 in 3 days. The crystals of UfSP2 belong to the space group C2, with a = 184.53 Å, b = 56.04

Å,  $c = 143.27$  Å, and  $\alpha = \gamma = 90^\circ$  and  $\beta = 128.01^\circ$ , and contains two molecules per asymmetric unit, corresponding to a Matthews volume  $V_m$  of  $2.78 \text{ \AA}^3 \text{ Da}^{-1}$ . Attempts to crystallize UfSP2 complex with Ufm1 did not yield crystals large enough suitable for high resolution data collection.

#### **X-ray data collection and processing.**

The X-ray diffraction data set from the native and selenomethione 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 100K 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 MAD phasing method, since 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 program SOLVE (31) and RESOLVE (32). About 54% of the residues were automatically modeled as a polyalanine chain by RESEOLVE 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  $R_{\text{free}}$  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 QuickChang site-directed mutagenesis kit (Stratagene) by following the manufacturer's instruction. Mutants of UfSP2 were produced as N-terminally His-tagged proteins with single point mutations at positions Tyr282, Cys294, Asn290, Thr422, and Met283. The chimerical regulatory, upstream, and neighbouring 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 neighbouring 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  $\mu\text{g}$  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 R250.

**Immunocytochemistry.** HeLa cells were grown on cover-slips 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 80x 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 1X protease inhibitor cocktail (Roche). 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.

## RESULTS

**Overall structure of mUfSP2.** UfSP2 from mouse, containing 461 amino acids, was crystallized, and the structure was determined using multi-wavelength anomalous dispersion data collected from the selenomethionine substituted UfSP2 and refined at 2.6 Å resolution. Crystallized UfSP2 was mutated at three positions: Cys294Ser, Lys94Arg, and Arg128Ala in order to avoid 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 Å × 50 Å × 55 Å, and consists of two domains, with the first domain composed of the 240 residues at the N-terminal, and the second domain comprising the 200 residues at the C-terminal as seen in Figure 1A. The two domains are connected by a linker of about 20 residues, and the C-terminal tail forms additional interactions with the N-terminal at the interface. The N-terminal domain, which is shaped like a rectangular box of 40 Å × 40 Å × 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 diagonally to it, while 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  $\beta$ -sheet facing the  $\alpha$ -helices is highly hydrophobic, while the opposite side of the  $\beta$ -sheet is somewhat polar. The two long helices are amphipathic with the hydrophobic surfaces facing the  $\beta$ -sheet. The linker residues are practically packed against the N-terminal of the  $\beta$ -sheet on the opposite side of the two parallel helices. The catalytic C-terminal domain has seven  $\alpha$ -helices and seven  $\beta$ -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 non-crystallographic 2-fold and show an RMSD (Root Mean Square Deviation) of 1.6 Å (Figure 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) & 2D1I (10)) with a Z-score of 12.7, and murine cytomegalovirus protease, M48<sup>USP</sup> (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 RMSD of 1.9 Å for 210 C $\alpha$  atoms. Some regions show an RMSD 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 UfSP1 (Figure 1B and S2). The helix  $\alpha$ 7, which harbors the catalytic cysteine residue, is longer in UfSP2, and there is a three residue insertion after the helix. The loop between  $\alpha$ 7 and  $\alpha$ 8 interacts with its  $\alpha$ 5 helix which turns 180° in comparison to the loop of UfSP1. A stretch of residues after  $\beta$ 8 reorganizes into a longer helix ( $\alpha$ 11) and is coupled with the changes in residues between  $\beta$ 12 and  $\beta$ 13. The C-terminus of UfSP2 has 3 extra residues, and they make contact with the N-terminal domain, e.g. the backbone amide of Ala460 forms hydrogen bonds with the carboxyl group of Asp231 and the carboxyl terminus of Leu461 forms hydrogen bonds to Arg219.

**Active site of UfSP2.** As seen in Figure 2A, all the atoms near the active site are well-defined in the electron density map. The catalytic Cys294 is located on the N-terminus of an  $\alpha$ -helix utilizing dipole moment, and “Asp418-Pro419-His420” are located at the loop off a  $\beta$ -strand (Figure 2B). 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  $\beta$ -strands of the central  $\beta$ -sheet (16). His398 of UfSP2, which is part of a highly conserved stretch among UfSP2, (Figure S3) is the canonical histidine position, thereby posing a 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 (Figure 2C). Indeed, in the crystal structure, His398 is too far from the catalytic Cys294, i.e. 7.2 Å and 3.9 Å away from the  $S_{\gamma}$  atom of Cys294 and  $N\delta 1$  of His420, respectively. Furthermore, superposition of papain shows that Leu396 is located at the canonical histidine position. The loss of activities by mutations of Cys294, Asp418, and His420 further confirm the identification of the catalytic triad, and Tyr282 is responsible for the formation of the oxyanion hole (Figure 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  $\beta 3$  and  $\beta 4$ , as well as Trp98 may play a role in Ufm1 recognition and/or stabilization (7). These correspond to the loop connecting  $\beta 9$  and  $\beta 10$ , and Trp342 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 (Figure 1B and 3). In order 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-Gly-Gly-Val-Leu-Ala-397 in UfSP2 were replaced by the corresponding loop in UfSP1, namely 149-Gly-Asp-Ala-Asp-Ala-Gln-Ser-155. As shown in Figure 3C, 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 in order to understand the lack of activity observed for the mouse equivalent UfSP2 mutation associated with BFHD (UfSP2 Tyr282His is the mouse equivalent of the human UfSP2 Tyr290His BFHD associated mutation). Of note was the finding that UfSP1 Tyr41His (Y41H), which corresponds to UfSP2 Tyr282His, cleaved GST-Ufm1-HA

at about 3-fold lower rate than wild-type UfSP1 (Figure 4C). Since Y41H retained the enzymatic activity (although significantly reduced), the residues that are not conserved within 6Å of the oxyanion hole Tyr of UfSP2 were examined. These residues included Met283, Asn290, and Thr422 of UfSP2, which correspond to Gly42, Arg49, and Trp179 of UfSP1, respectively (Figure 4A). To the inactive Tyr282His UfSP2 mutant, additional mutations were introduced 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 Figure 4C, incorporation of Arg at amino acid 290 resulted in a slight recovery of activity. Introduction of an additional mutation at position 422, i.e. Tyr282His/Asn290Arg/Thr422Trp led to a further recovery in enzymatic activity. A similar effect was observed in the case of Tyr282His/Met293Gly/Asn290Arg triple mutant. However, incorporation of mutation at all four sites, i.e. Tyr282His/Met283Gly/Asn290Arg/Thr422Trp did not seem to yield additional enhancement. We then tested the possible effect of residues further away from Tyr282, in two adjacent loops that differ from one another as seen in Figure 4B. These residues correspond to Gln284-Ile287 and Gly423-Leu427 in UfSP2, and Cys43-Leu46 and Gly180-Arg185 stretches in UfSP1 (Figure 1B), 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) (Figure 4C). On the other hand, the replacement of ‘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-terminal and C-terminal domains. As shown in Figure 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: while the newly identified Uf11 serves as a Ufm1 E3 ligase for C20orf116, 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  $\beta$ -COP, a marker protein for Golgi bodies (Figure 6). Coexpression 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 (FL) and its N-terminal domain, but not its C-terminal domain, co-precipitated with C20orf116 (Figure 7A), 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 HisMax-tagged UfSP2 and its N- and C-terminal domains. Pull-down analysis by using  $\text{Ni}^{2+}$ -nitrilotriacetic acid (NTA) 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 (Figure 7B). 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 *C. elegans*, the N-terminal domain is even longer, with about 100 extra residues (Figure 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  $\beta$ -sheet with 2  $\alpha$ -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  $\beta$ -sheet and two parallel  $\alpha$ -helices of similar length. Yet, the topology of the domain is quite different, and the groove between the two long helices ( $\alpha 1$  and  $\alpha 3$ ) 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  $\alpha 4$  facing the groove interface, as well as the beginning part of the linker, are conserved amongst all known forms of UfSP2 (Figure 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 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 Cys294Ala and His420Ala mutants, and some residual activity for the Asp418Ala mutant (Figure 2C). Mutation of His398, which is located near the canonical histidine position, to an alanine did not affect the enzyme activity as expected. These results confirm Cys294, Asp418, and His420 as the *bona fide* catalytic residues for UfSP2 as was previously suggested based on the structure of UfSP1 (7). Cys294 and His420 are directly involved in the reaction, serving as a nucleophilic attacking group and a general acid-base catalytic element, respectively, while Asp418 is essential in stabilizing the transition state through electrostatic interactions. The residual activity seen for the Asp418Ala mutant may be due to the aid of His398 residues (Figure 2). In fact, another crystal form of UfSP2 having one molecule per asymmetric unit shows an additional water molecule between His420 and His398 that could make hydrogen bond linkage between the two (data not shown). It is worth mentioning that when the corresponding aspartate in UfSP1 (Asp175) was mutated to an alanine, it showed a complete loss of activity. Additionally, the position of His398 is occupied by Lys156 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 Trp98, and the regulatory loop containing residues 149-Gly-Asp-Ala-Asp-Ala-Gln-Ser-155 in UfSP1, play a role in stabilizing Ufm1 binding, i.e. Ufm1 recognition based on the binding study using NMR and modeling. In UfSP2, these correspond to Trp342 and 393-Gly-Gly-Val-Leu-Ala-397 (Figure 3 and 4). Residues around Trp342, which is part of a highly conserved stretch, practically have the same conformation as is found in UfSP1. However, the regulatory loop connecting  $\beta$ 9 and  $\beta$ 10 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 Figure 3. In UfSP1, the loop is held in place *via* interactions with water molecules as well as the residues connecting  $\alpha$ 6 helix and  $\beta$ 7 strand, while the corresponding loop in UfSP2 does not make significant interactions, except the hydrophobic interaction between Trp342 and Val395. Among the UfSP2s these regulatory loops are highly conserved, and the Val395Ala mutant shows reduced activity (data not shown).

In order 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 hours (Figure 3C). 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 Phe119 of LC3, while overall structures are almost identical. In the case of USP2 and HAUSP, in addition to domain-wide conformational changes, reordering of the catalytic site upon Ub binding was reported (22-23), while 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 Tyr290 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 Tyr282 in mouse UfSP2 was mutated to histidine, UfSP2 almost completely lost its catalytic activity, while the analogous Tyr41His mutation in UfSP1 retained decreased activity (Figure 4C). In the crystal structure of UfSP1, the hydroxyl oxygen of Tyr41 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 Cys53 in UfSP1. The water molecule is 3.2 Å away from the hydroxyl oxygen of Tyr282, between two UfSP2s in the asymmetric unit. When Tyr41 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 6Å from

Tyr282 that are not conserved, such as Met283, Asn290, and Thr422 were mutated to test for their effects on the oxyanion hole. Mutations of Asn290 did not show much effect when replaced by alanine, lysine, or arginine. However, when it was mutated together with Thr422 and/or Met283, the activity of the enzyme was partially restored. When the upstream loop spanning residues 282 to 289 of UfSP2, namely 282-**Tyr-Met-Gln-Asp-Arg-Ile-Asp-Asp**-289, was replaced by that of UfSP1, namely 41-**Tyr-Gly-Cys-Asp-Gly-Leu-Asp-Asp**-48, the enzymatic activity was restored to nearly the same level of wild type UfSP2. This is the relatively conserved loop connecting the  $\beta$ -strand with the oxyanion hole, and the helix with the catalytic cysteine. In UfSP1, Asp47 and Arg49 make two by two hydrogen bonds that are stacked by the side chain of Trp179, shielding the side chain of Tyr41. In UfSP2, Met283 occupies the side chain position of Trp179 of UfSP1, and Trp179 is replaced by Thr422. Arg49 of UfSP1 is replaced by Asn290, so Tyr282 is not as well-shielded. In UfSP2, Gln284, which is involved in the upstream loop, forms hydrogen bonds with the carbonyl oxygen and N $\epsilon$ 2 of His281. Ultimately, these configurations make the loop connected to Tyr282 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 Cys294, Asp418, His420, Tyr282 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. Ile93Met mutation of ubiquitin C-

terminal hydrolase-L1 (UCH-L1), which is a highly abundant neuronal enzyme that is associated with familial Parkinson's 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. I., Natsume, T., Kang, S. H., Chung, C. H., Kasahara, M., Kominami, E., Yamamoto, M., Tanaka, K., and Komatsu, M. (2010) *J Biol Chem* **285**, 5417-27
4. Kang, S. H., Kim, G. R., Seong, M., Baek, S. H., Seol, J. H., Bang, O. S., Ova, 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) *Macromolecular Crystallography, Pt A* **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. 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., 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

#### FOOTNOTES

\*This work was supported by grants from the Functional Proteomics Center (FPR08B2-280), the 21C Frontier Research & Development Program of the Korea Ministry of Science and Technology and the Korea Institute of Science and Technology Institutional Program Grant to E.E. Kim and by grants from the Korea Research Foundation (KRF-2005-084-C00025) and the Korea Science and Engineering Foundation (M10533010001-05N3301) to C.H. Chung. B.H. Ha was the recipient of BK21 fellowship. 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 have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>) The accession code is 3OQC. Supplementary data include three figures and are available on the JBC website.

## FIGURE LEGENDS

**Figure 1. Overall structure of UfSP2.** (A) Ribbon diagram of the overall structure of UfSP2 with red  $\alpha$ -helices and green  $\beta$ -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 while the residues in the blue box are relatively conserved. Secondary structures are depicted above the sequences, coils indicate  $\alpha$ -helices and arrows indicate  $\beta$ -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.

**Figure 2. The Active site of UfSP2.** (A) The electron density around the active site.  $2F_o - F_c$  map is contoured at  $1.0\sigma$ . (B) The catalytic triad is formed by Cys 294, Asp418, and His 420 with Tyr282 participating in the formation of oxyanion hole. (C) *In vitro* processing activities of Cys294Ser, Asp418Ala, His420A, and His398Ala mutants were assayed by incubating  $1.5 \mu\text{g}$  of the proteins for 2 h as described under “Experimental Procedures.” S denotes substrate (GST-Ufm1-HA) while W/T is for wild-type UfSP2.

**Figure 3. The regulatory loop.** The R-loop of (A) UfSP1 and (B) UfSP2 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  $1 \mu\text{g}$  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.

**Figure 4. The oxyanion hole mutation in BFHD.** (A) The comparison of the active sites of UfSP2 (pink) and UfSP1 (grey). Residues which are not conserved are highlighted in brown (Asn280, Met283 and Thr422 of UfSP2). Ufm1 is modeled based earlier binding results<sup>7</sup> and shown in green. (B) Comparison of UfSP2 (pink) and UfSP1 (grey) 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  $0.2 \mu\text{g}$  of UfSP1 proteins and  $1.5 \mu\text{g}$  of UfSP2 proteins for the indicated periods.

**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  $\mu\text{m}$ .

**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- $\beta$ -COP and anti-calreticulin antibodies. Bars, 10  $\mu\text{m}$ .

**Figure 7. Interaction of UfSP2 and its N-terminal domain with C20orf116.** (A) Myc-tagged UfSP2 and its N- and C-terminal domains were expressed in HeLa cells with Flag-C20orf116. Cell lysates were subjected to immunoprecipitation with anti-Flag antibody followed by immunoblot 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  $\text{Ni}^{2+}$ -nitrilotriacetic acid (NTA)-conjugated resin followed by immunoblot with anti-Myc antibody.

**Table 1.** Data collection and crystallographic refinement statistics.

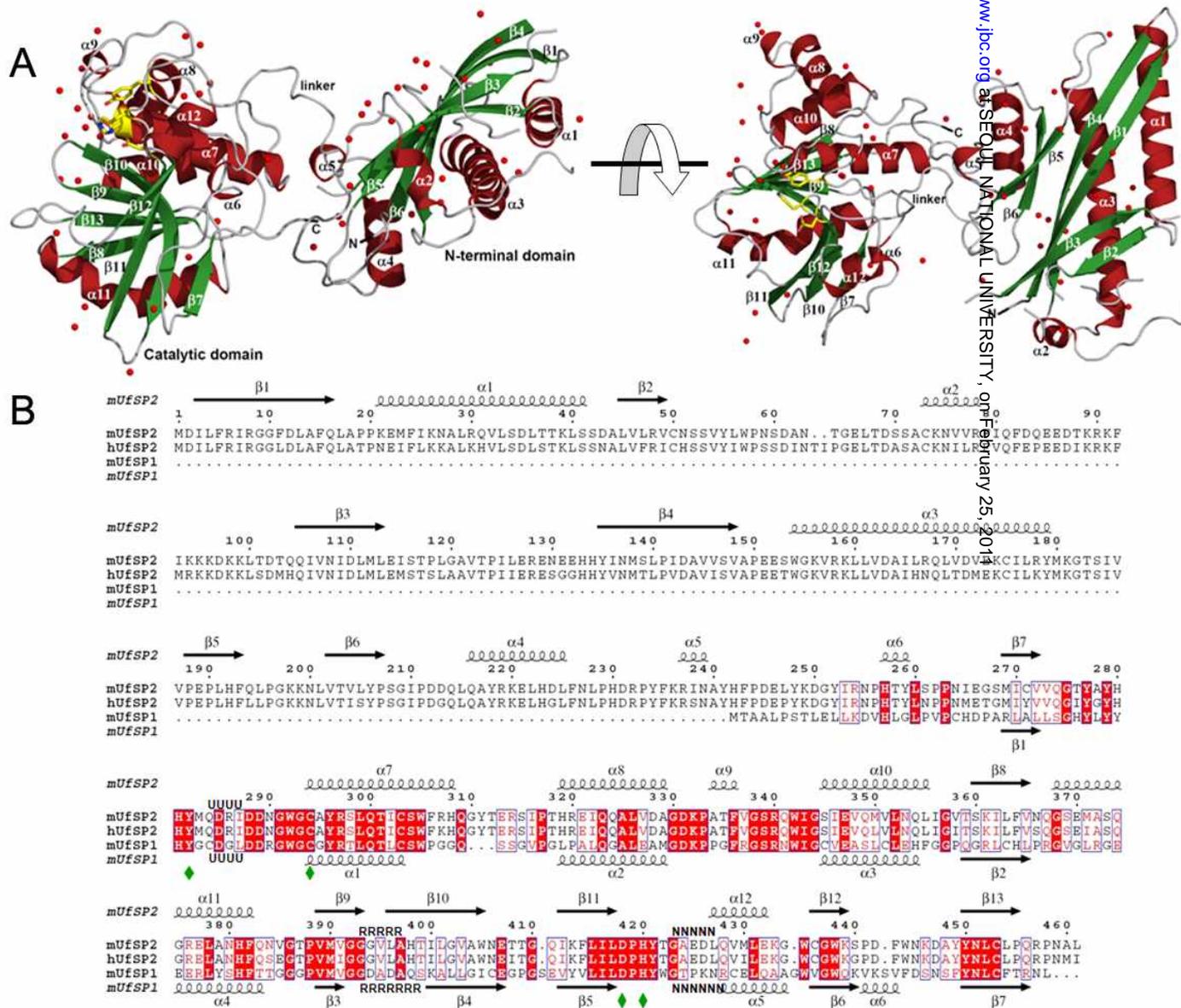
Data sets	MAD			Native
	Peak	Edge	Remote	
Beam Line	BL-6C/BL-4A			
X-ray wavelength (Å)	0.97947	0.979613	0.97177	1.0
Energy (eV)	12658.3	12656.4	12758.5	12398.9
Resolution range (Å)	50-2.8	50-2.8	50-2.8	50 – 2.6
Space group	C2	C2	C2	C2
Unit-cell parameters (Å)	$a = 123.130$	$a = 123.145$	$a = 123.152$	$a = 184.533$
	$b = 63.623$	$b = 63.621$	$b = 63.617$	$b = 56.041$
	$c = 100.669$	$c = 100.668$	$c = 100.677$	$c = 143.269$
	$\alpha = 90$	$\alpha = 90$	$\alpha = 90$	$\alpha = 90.0$
	$\beta = 117.7$	$\beta = 117.695$	$\beta = 117.687$	$\beta = 128.013$
	$\gamma = 90$	$\gamma = 90$	$\gamma = 90$	$\gamma = 90.0$
Total/unique reflections	914333/17170	920752/17139	949940/17199	530451/35942
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.)
$R_{\text{merge}}^{\dagger}$ (%)	10.9 (25.4)	10.6 (25.0)	10.8 (28.0)	10.6 (39.2)
<b>Refinement statistics</b>				
Resolution range (Å)				46.1 – 2.60
$R/R_{\text{free}}^{\ddagger}$ (%)				23.8/29.8
No. of protein atoms				6609
No. of water molecules				107
Average B-factor (Å <sup>2</sup> )				42.0
RMSD from ideal geometry				
: bond length(Å)/angle(deg.)				0.015/1.72
Ramachandran analysis (%)				
Favored region				84.2
Additionally allowed				15.4
Generously allowed				0.1
Disallowed regions				0.3

Values in parentheses refer to the highest resolution shell

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

$R_{\text{free}}^{\ddagger}$  is calculated from the randomly selected 10% set of reflections not included in the calculation of the R-value.

Figure 1. Overall structure of UfSP2. (Ha *et al*)



**Figure 2. The Active site of UfSP2.** (Ha *et al*)

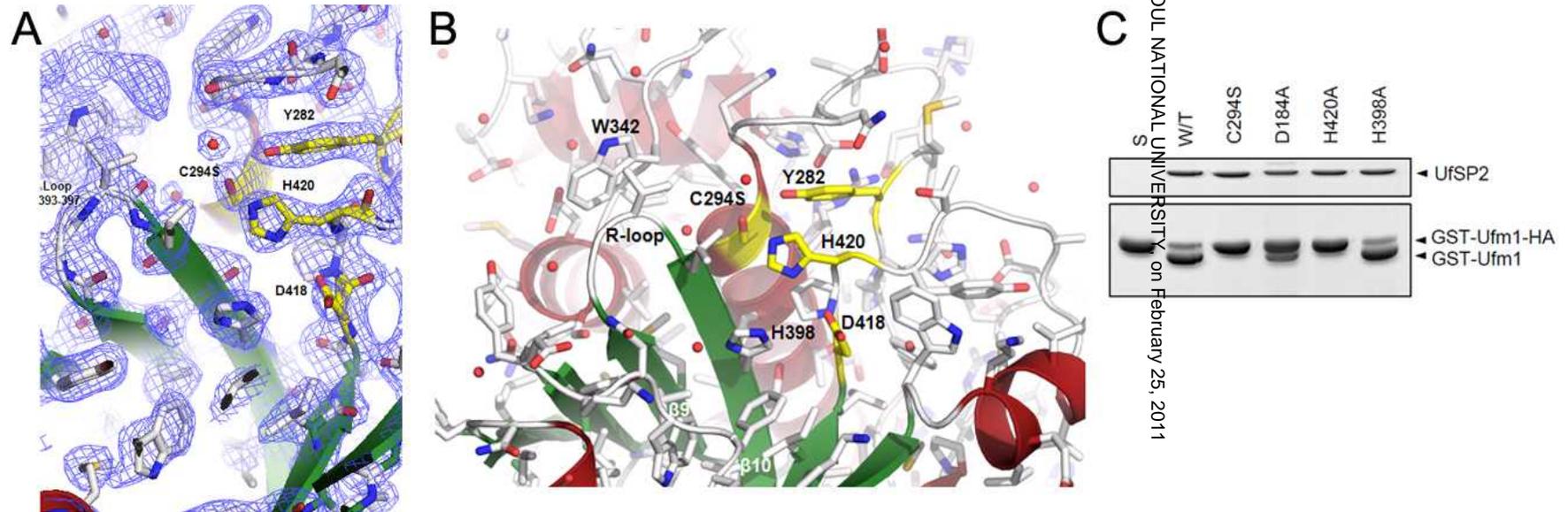


Figure 3. The regulatory loop. (Ha *et al*)

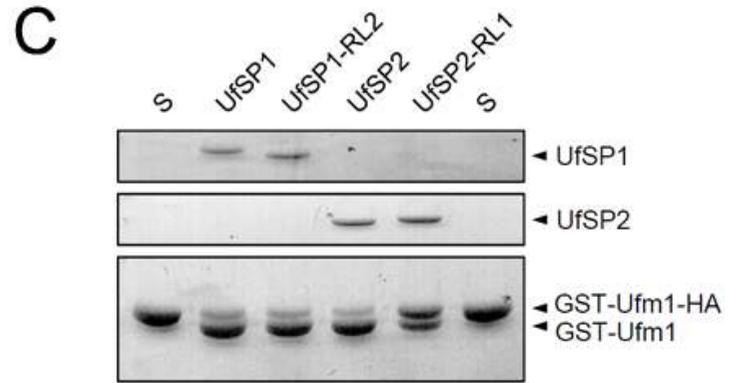
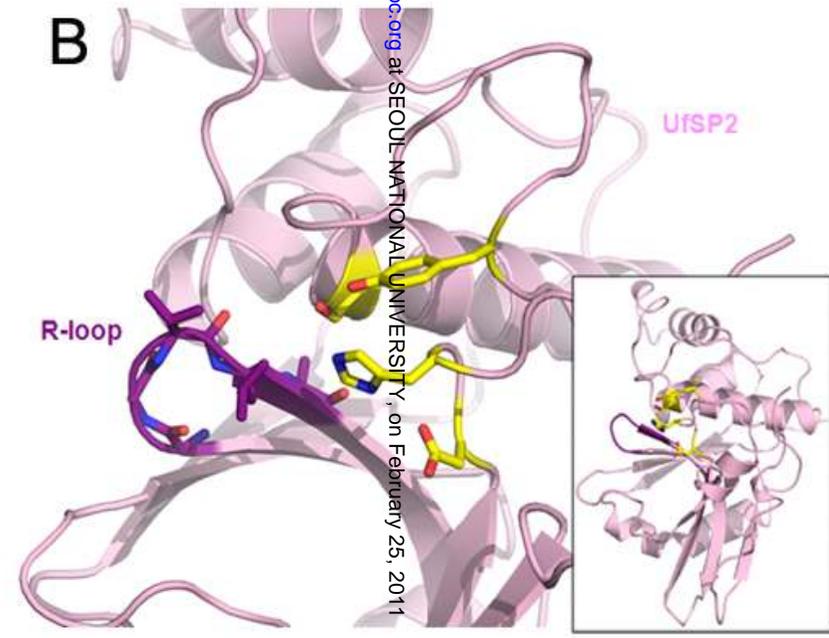
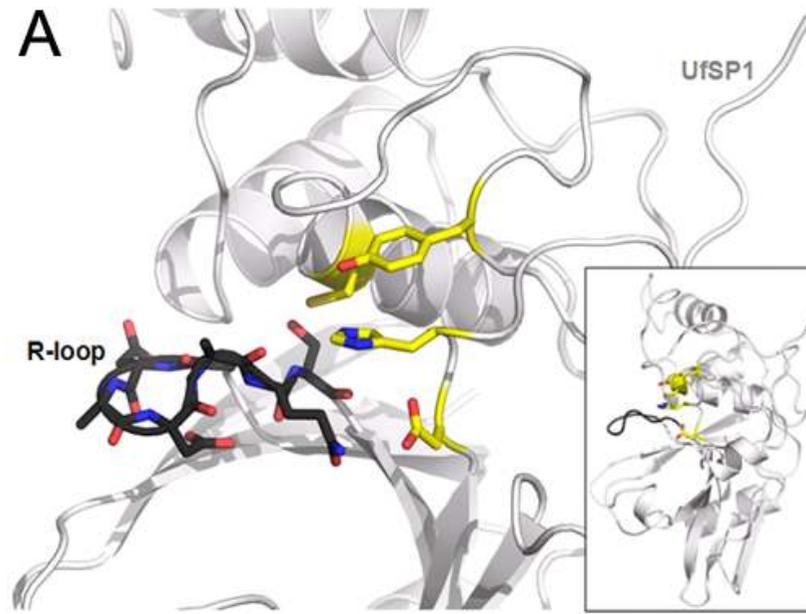
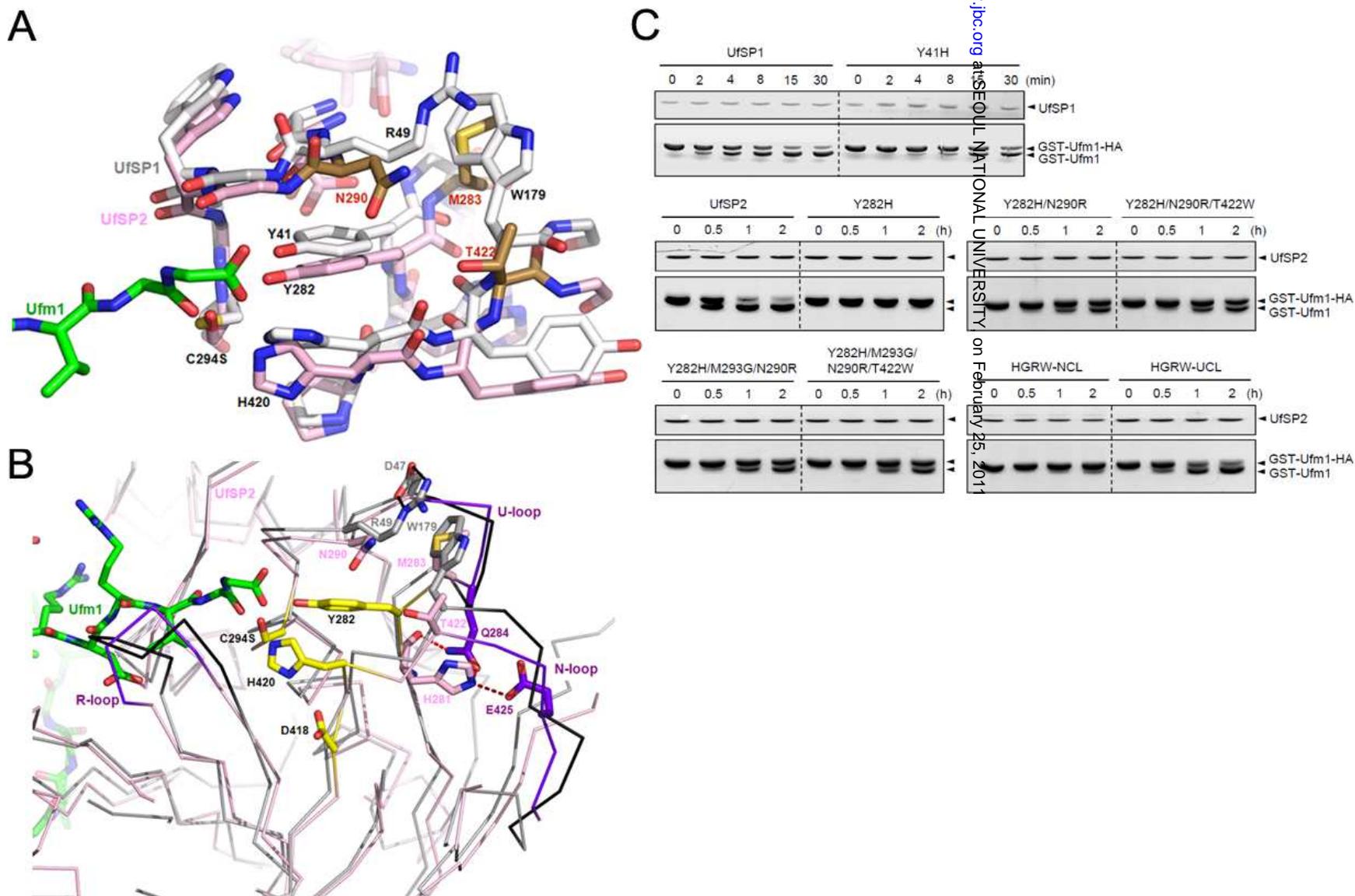


Figure 4. The oxyanion hole mutation in BFHD. (Ha *et al*)



**Figure 5. Subcellular localization of UfSP2 and its N- and C-terminal domains. (Ha *et al*)**

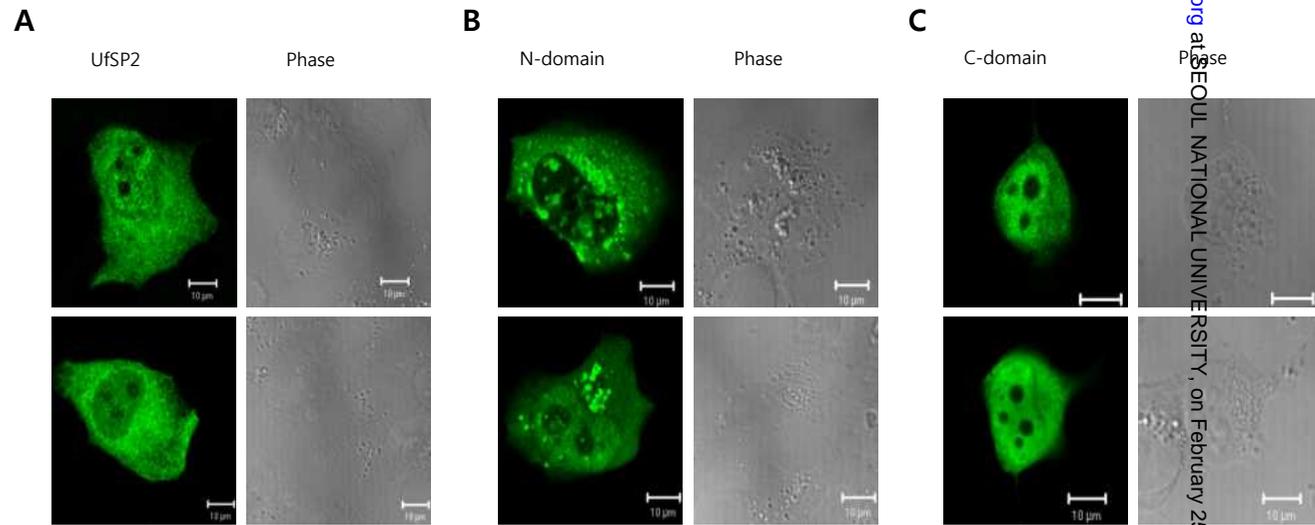
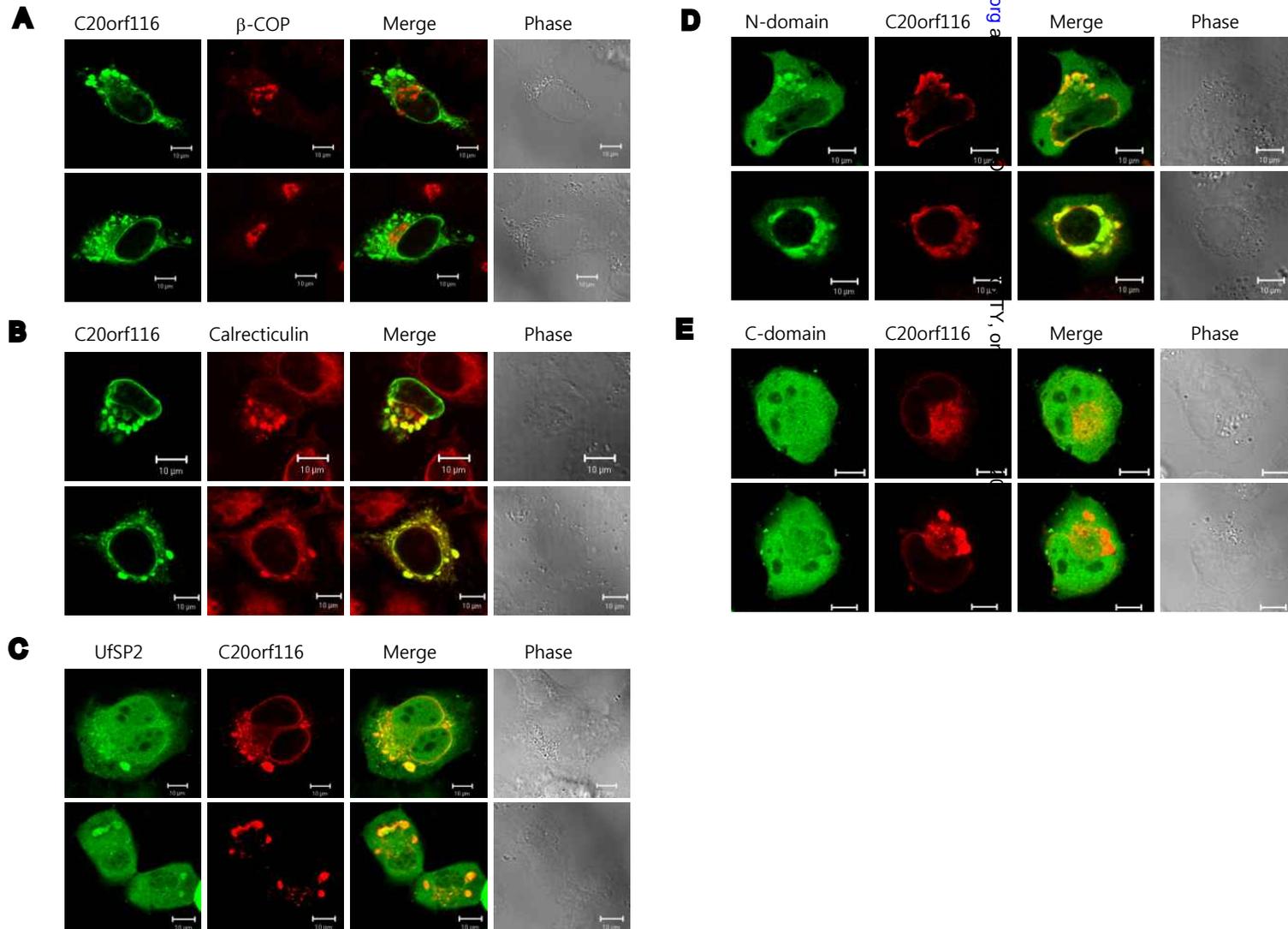


Figure 6. Co-localization of UfSP2 and its N-terminal domain with C20orf116 in ER. (Ha *et al*)



**Figure 7. Interaction of UfSP2 and its N-terminal domain with C20orf116.** (Ha *et al*)

