Human UFSP1 translated from an upstream near-cognate initiation codon functions as an active UFM1-specific protease

Received for publication, December 22, 2021, and in revised form, April 24, 2022  Published, Papers in Press, May 5, 2022, https://doi.org/10.1016/j.jbc.2022.102016

Qian Liang, Yaqi Jin, Shiwen Xu, Junzhi Zhou, Jian Mao, Xiaohe Ma, Miao Wang, and Yu-Sheng Cong*
From the Key Laboratory of Aging and Cancer Biology of Zhejiang Province, School of Basic Medical Sciences, Hangzhou Normal University, Hangzhou, China

Edited by George DeMartino

Ubiquitin-fold modifier 1 (UFM1) is a recently identified ubiquitin-like posttranslational modification with important biological functions. However, the regulatory mechanisms governing UFM1 modification of target proteins (UFMylation) and the cellular processes controlled by UFMylation remain largely unknown. It has been previously shown that a UFM1-specific protease (UFSP2) mediates the maturation of the UFM1 precursor and drives the de-UFMylation reaction. Furthermore, it has long been thought that UFSP1, an ortholog of UFSP2, is inactive in many organisms, including human, because it lacks an apparent protease domain when translated from the canonical start codon (445AUG). Here, we demonstrate using the combination of site-directed mutagenesis, CRISPR/Cas9-mediated genome editing, and mass spectrometry approaches that translation of human UFSP1 initiates from an upstream near-cognate codon, 217CUG, via eukaryotic translation initiation factor eIF2A-mediated translational initiation rather than from the annotated 445AUG, revealing the presence of a catalytic protease domain containing a Cys active site. Moreover, we show that both UFSP1 and UFSP2 mediate maturation of UFM1 and de-UFMylation of target proteins. This study demonstrates that human UFSP1 functions as an active UFM1-specific protease, thus contributing to our understanding of the UFMylation/de-UFMylation process.

The ubiquitin-fold modifier 1 (UFM1) system is a recently identified ubiquitin-like posttranslational modification with essential biological functions (1). Deficiency of this modification leads to embryonic lethality in mice and diseases in humans (2). UFM1 is present in nearly all eukaryotic organisms (except fungi) with a similar tertiary structure to ubiquitin. Similar to ubiquitination, the covalent conjugation of UFM1 (UFMylation) to target proteins involves a three-step enzymatic cascade catalyzed sequentially by UFM1-activating enzyme 5 (UBA5, E1), UFM1-conjugating enzyme 1 (UFC1, E2), and UFM1-specific ligase 1 (UFL1, E3) (2, 3). The UFMylation process is highly conserved in metazoans and plants, implicating its specific roles in multicellular organisms. In human, gene mutations in UFMylation components (UFM1, UFC1, UBA5, DDRGK1, or UFM1-specific protease [UFSP] 2) have been found to be associated with a variety of neurological disorders and skeletal abnormalities (4–7). Accumulating evidence suggests that UFMylation plays a critical role in diverse cellular processes, including erythrocyte differentiation during embryogenesis (8–10), endoplasmic reticulum (ER) homeostasis (11, 12), translational homeostasis (13, 14), DNA damage response, and cancer-related signaling pathways (15–18). So far, only one of the E1, E2, and E3 enzymes each of the UFMylation system have been identified, and a handful of substrates have been reported. The regulatory mechanisms governing UFM1 modification of target proteins and the cellular processes controlled by UFMylation remain largely unknown.

UFMylation is a reversible process because of UFSPs mediated de-UFMylation reaction (19). Although two UFSP genes (UFSP1 and UFSP2) are present in the human genome, it has long been believed that human UFSP1 is inactive or nonfunctional because it lacks a specific protease domain as translated from the annotated 445AUG (2, 17). Therefore, UFSP2 has been regarded as the only active protease that mediates UFM1 precursor (pro-UFM1) maturation and de-UFMylation in human cells. Intriguingly, several studies have shown that KO of UFSP2 resulted in significantly increased protein UFMylation in human cells, indicating that other active UFSPs mediate pro-UFM1 maturation in human cells (13, 20).

Given that only UFSP1 and UFSP2 are present in the human genome, we are naturally concerned about the true identity of UFSP1 in the UFMylation/de-UFMylation process. It has been known for decades that translation can start from codons other than AUG, usually from near-cognate initiation codons, which have a sequence that differs from the AUG codon by one nucleotide (for example, CUG, GUG, and UUG) (21–23). Through sequence alignment, we identified a potential coding region of cysteine protease catalytic domain upstream of the canonical ORF in human UFSP1 gene, which may be derived from near-cognate CUG codons. In this study, we demonstrated that expression of human UFSP1 is initiated through eIF2A-mediated translational initiation, and human UFSP1 is a functional UFSP with distinguishing feature in protein UFMylation/de-UFMylation.

* For correspondence: Yu-Sheng Cong, yscong@hznu.edu.cn.


**Results**

**Human UFSP1 is an active UFSP**

To explore the possibility that the near-cognate start codon makes human UFSP1 acquired functional protease activity, we constructed two UFSP1 expression plasmids, one containing the canonical coding region (UFSP1-Short) and the other containing the full length of the complementary DNA (cDNA) with the 5′ untranslated region (5′ UTR) (UFSP1-Long). The canonical coding region of human UFSP1 is expected to generate the peptide with the molecular weight (MW) approximately 17 kDa. Intriguingly, we only detected a ~23 kDa specific protein band in UFSP1-L construct transfected cells but no expected size of ~17 kDa protein band was detected in UFSP1-S construct–transfected cells (Fig. 1A). Similarly, we constructed GFP-tag fused UFSP1-S or UFSP1-L vectors and only detected a prominent ~50 kDa fusion protein band (GFP-tag, ~27 kDa) in UFSP1-L-GFP construct–transfected cells (Fig. S1A). Meanwhile, the levels of protein UFMylation were significantly decreased by the expression of UFSP1-L or UFSP1-L-GFP, but not by that of UFSP1-S or UFSP1-S-GFP (Figs. 1A and S1A), and this decreased UFMylation associated with UFSP1-L expression is fully abrogated by the potential active site cysteine (240 TGC) to alanine (GCC) mutation (Fig. 1B). These results suggest that an upstream near-cognate codon, other than the annotated 445 AUG codon, was used for human UFSP1 translation initiation and for the production of an active UFSP.

During pro-UFM1 maturation, the C-terminal Ser-Cys dipeptide of pro-UFM1 (~9.1 kDa) is cleaved by the UFSPs to expose its C-terminal conserved Gly residue (24). Matured UFM1 (~8.9 kDa) is required for conjugation to its target proteins. To further confirm the functionality of human UFSP1 in pro-UFM1 maturation, we transfected a C-terminal hemagglutinin (HA)-tagged pro-UFM1 plasmid (pSG5-UFM1-HA) or control vector into HEK293T cells depleted UFSP1, UFSP2, or both UFSP1 and UFSP2, respectively. We found that the patterns of protein UFMylation were significantly changed with distinctive features in UFSP1, UFSP2, or double KO cells (Figs. 1C and S1B). UFSP2 KO resulted in considerably increased levels of protein UFMylation in cells (Fig. 1C, lines 3 and 7). In addition, the levels of protein UFMylation in UFSP1 KO cells were slightly increased or comparable to those in the parental cells (Fig. 1C, lines 2 and 6). However, UFMylation was completely blocked in UFSP1/UFSP2 double KO cells (Fig. 1C, lines 4 and 8). These results further demonstrate that human UFSP1 is an active UFSP with distinguished specificity in protein UFMylation/de-UFMylation; and UFSP1 and UFSP2 are the only UFSPs in human cells.

**Human UFSP1 is translationally initiated from 217 CUG codon**

In eukaryotic cells, CUG initiation codon is the most common among the various near-cognate codons identified (23, 25). We evaluated the 5′ UTR sequence of human UFSP1 mRNA for translation start site and found a total of seven CUG codons in-frame coding region containing the upstream catalytic domain. According to the MW, we have narrowed down the list to four candidate CUG codons, 190 CUG, 217 CUG, 226 CUG, and 247 CUG (Fig. 2A). To identify the true initiation site, we created a set of mutations on each of these four CUG codons (as CTG in plasmids) individually to determine which is required for human UFSP1 expression. Mutations at 217 CTG fully abrogate UFSP1 expression, whereas mutations at

---

**Figure 1. Human UFSP1 is an active UFM1-specific protease. A**, de-UFMylation activity of the human UFSP1 in HEK293T cells (left) and HeLa cells (right). B, mutation of 240 TGC in human UFSP1 gene fully abrogates its de-UFMylation function. Human UFSP1-L expression plasmid with 240 TGC to GCC mutation, UFSP1-L, UFSP1-S expression plasmid, and the control vector were introduced into HEK293T cells. C, human UFSP1 and/or UFSP2 KO changed the patterns of protein UFMylation and pro-UFM1 maturation. Human UFSP1 can cleave the pro-UFM1 (~11.8 kDa) to produce the mature form UFM1 (~8.9 kDa). UFSP1 KO (UFSP1ΔC), UFSP2 KO (UFSP2ΔC), and UFSP1/UFSP2 double KO (UFSP1ΔC/UFSP2ΔC) cell lines were generated by CRISPR/Cas9-mediated genome editing and monoclonal screening in HEK293T cells. The C-terminal HA-tagged pro-UFM1 plasmid (pSG5-UFM1-HA) or control vector (pSGS-HA) were transfected into the WT or gene KO cells. After 48 h post-transfection, cell lysates were analyzed by Western blot analysis using the indicated antibodies. HA, hemagglutinin; UFM1, ubiquitin-fold modifier 1; UFSP, UFM1-specific protease.
190CTG, 226CTG, or 247CTG have no significant effect on UFSP1 expression (Fig. 2B). These results indicate that 217CUG codon, but not the other three near-cognate codons, is the translation initiation site of human UFSP1.

In order to confirm the translation initiation site, we sought to obtain direct evidence using mass spectrometry (MS) for peptide sequencing (Fig. S2A). We generated a homozygous FLAG-tag knock-in (KI) cell line (FLAG-KI-6#) using CRISPR/Cas9 system mediated knock-in of the FLAG-tag at the C-terminal end of the UFSP1 locus in HEK293T cells. The oligodeoxyribonucleotides (ssODNs) functioned as the linear donors and were used to introduce FLAG-tag insertion during homology-dependent repair (HDR) after the Cas9/sgRNA–mediated site-specific double-strand break (DSB). D, verification of expression of FLAG-tagged endogenous UFSP1 by mass spectrometry (MS). Cell lysates of FLAG-KI-6# or control cells were subjected to immunoprecipitation with anti-FLAG M2 affinity gel. The bound proteins were separated with SDS-PAGE and stained with Coomassie blue. The 25 kDa specific band (in red box) was sliced and analyzed using MS, and six peptides in the UFSP1-FLAG knock-in sample that match the region between the potential 217CTG start codon and the canonical 445ATG codon (in red font) were identified. UFSP, UFM1-specific protease.
Cas9 system, in which a FLAG-tag coding sequence (CDS) was inserted at the C terminus of the human UFSP1 locus (Figs. 2, C and D and S2B). The MS results of endogenous UFSP1-FLAG reveal a total of eleven peptides covering 75% of the predicted full-length human UFSP1, among them six peptides were covered more than 90% of the predicted N-terminal region (from CUG to the canonical AUG) (Fig. 2E). The MS data confirmed that CUG is the human UFSP1 translation start site, which encodes the amino acid methionine (Met), but not leucine (Leu). CUG normally encodes for a Leu but when used as an alternative start site, CUG can encode for Met (22). The MS data with exogenous C-terminal HA-tagged UFSP1 (UFSP1-HA) showed a similar result (Fig. S2C). Theoretically, all potential UFSP1 isoforms with the same FLAG-tagged or HA-tagged C terminus were immunoprecipitated using anti-FLAG or anti-HA affinity gel. However, our data showed that only one tagged UFSP1 band was specifically detectable with a MW corresponding to the upstream near-cognate codon CUG instead of annotated AUG (Figs. 2, D and E and S2C), suggesting that UFSP1 expressed only the full-length enzyme-active form in human cells.

The 5' UTR sequence is essential for human UFSP1 expression

While human UFSP1 translationally initiated from CUG, we found the efficiency of exogenous expression was extremely low without the 5' UTR sequence in the constructs. Thus, a certain 5' UTR sequence may be required for the efficient translation of human UFSP1. The Kozak consensus sequence (RCCUUGG, where R is a purine) plays an important role in translation initiation (26, 27). In UFSP1 mRNA, we noted that CUG is embedded in an optimal Kozak consensus sequence (ACCGCCUGG), whereas no such motif exists surrounding the AUG and other near-cognate codons (Fig. 2A). We constructed a series of UFSP1 plasmids with full-length or truncated 5' UTR (Fig. 3A). After transfection, we found that the Kozak consensus sequence can only sustain basal UFSP1 expression, and efficient protein expression needs longer 5' UTR sequence (Fig. 3B). We found that approximate 126 nt long sequence upstream of CUG is sufficient for high efficiency of human UFSP1 translation as full-length 5' UTR (Fig. 3C). In particular, the sequence between 97 and 126 nt upstream of CUG seems to be required for efficient UFSP1 expression.

In order to ascertain the role of 97 to 126 nt for efficient UFSP1 expression, we constructed UFSP1-L-Random expression plasmids in which the 30 nt sequence between 97 and 126 nt were substituted by a 30 nt random sequence. After transfection, we found the efficiency of human UFSP1 expression with UFSP1-L-Random plasmids was extremely low as compared to that with WT UFSP1-L or 5' UTR-UFSP1-126 (Fig. 3D). We found that the sequence between 97 and 126 nt contains an E-box motif (CAGCTG), and the expression of human UFSP1 was significantly inhibited by the E-box deletion or mutation (Fig. S3A). These results suggest that the 97 to 126 nt sequence plays an important role in human UFSP1 expression. In addition, we observed that this 97 to 126 nt sequence inserted upstream GFP recombinant construct is able to enhance the GFP expression (Fig. S3B). These data suggest that the specific sequence in 5' UTR is essential for human UFSP1 expression.

**eIF2A mediates CUG initiation of human UFSP1 translation**

The detailed mechanisms of non-AUG initiation are still unclear. It has been reported that eIF2A plays an important role in translational initiation at the CUG start codon, and some chemical compounds differentially regulate protein synthesis initiated at the AUG or CUG start codons (28, 29). For example, acriflavin selectively inhibits CUG initiation, whereas aurintricarboxylic acid (ATA) inhibits AUG initiation but enhances initiation at the CUG codon. We examined the expression levels of human UFSP1 in response to these chemical inhibitors in HEK293T and HeLa cells and found that acriflavin attenuated UFSP1 expression in a dose-dependent manner (Fig. 4, A and B), whereas ATA increased the expression of UFSP1 (Fig. 4, C and D). Whereas the expression of UFSP2 and global UFMylation were barely changed by the treatment with ATAc or were inhibited by treatment with acriflavin only at high dosage (Fig. 4, A–D). Similarly, overexpression of FLAG-eIF2A increases the level of UFSP1 in both HeLa and HEK293T cells, but no changes in UFSP2 expression and in protein UFMylation were observed (Fig. 4, E and F). Moreover, eIF2A knockdown reduced human UFSP1 expression in both in HEK293T (Fig. 4, G and H) and HeLa (Fig. 4, I and J) cells. These results suggest the indispensable role of eIF2A in CUG translation initiation of human UFSP1.

**Human UFSP1 functions in UFMylation/de-UFMylation process**

To assess the function of UFSP1 in the maturation of the pro-UFM1 and de-UFMylated processes in comparison with UFSP2, we analyzed UFMylation/de-UFMylation of the known substrate activating signal cointegrator 1 (ASC-1) and showed that both human UFSP1 and UFSP2 can deconjugate the UF1M1 from UFMylated ASC-1 (Fig. 5A). In addition, we purified human UFSP1 and pro-UFM1-HA from HEK293T cells and showed that human UFSP1 can cleave the pro-UFM1-HA to generate mature UFM1 in in vitro assay (Fig. 5B). We have initially attempted to purify human recombinant UFSP1 from *Escherichia coli*; however, human UFSP1 did not get expressed in *E. coli* (Fig. S4A). Using codon-optimized (CO) UFSP1-L plasmid with CAG initiation codon, we have successfully purified the human recombinant UFSP1 protein from *E. coli* (Fig. S4B) and showed that human recombinant UFSP1 exhibits high enzymatic activity in maturation of pro-UFM1 (Fig. 5C). These results provide direct evidence that human UFSP1, like UFSP2, is an active UFM1-specific protease. In addition, our results showed that human UFSP1 exhibits higher catalytic activity than UFSP2 in de-UFMylation of ASC-1, overexpression of UFSP1 mediated stronger protein de-UFMylation than overexpression of UFSP2 (Fig. 5, A and D).

Both of UFSP1 and UFSP2 are widely expressed in human tissues, and UFSP1 is expressed at a very low level compared with UFSP2 (Fig. 5). Unlike predominantly ER membrane-localized
UFSP2, we found that human UFSP1 is a cytosolic protein (Fig. 5E). Based on the differences of UFSP1 and UFSP2 in catalytic activity, expression levels, intracellular localization, and UFMylation pattern in UFSP1 or UFSP2 KO cells, we believe UFSP2 appears to be mainly involved in the deconjugation of UFMylation, whereas UFSP1 is mainly involved in the maturation of pro-UFM1. Both of UFSP1 and UFSP2 together maintain a dynamic and reversible process of protein UFMylation/de-UFMylation in cells.

**Discussion**

In this study, we demonstrated that translation of human UFSP1 is initiated from the upstream near-cognate 217CUG codon in 5' UTR through eIF2A-mediated translational initiation, rather than the annotated 445AUG codon, producing an active UFMylation protease. The approximate 126 nt long sequence upstream of 217CUG plays an important role in the efficiency of UFSP1 expression. Like UFSP2, human UFSP1 functions in both maturation of the pro-UFM1 and de-UFMylation reaction. Unlike UFSP2, human UFSP1 expresses at low level in cytosol and may have different substrate specificity.

Most CUG-initiated proteins (such as PTEN, FGF2, and BiP) have the canonical AUG-initiated form besides one or two CUG-initiated forms (29–31), whereas human UFSP1 has only the 217CUG-initiated form (~23 kDa), but not the 445AUG-initiated form (~17 kDa). Even during exogenous expression, we can only detect a very small amount of 445AUG-initiated form by overexposure, which is likely caused by the ribosomal leaky scanning mechanism during protein translational initiation (32). Thus, human UFSP1 is translated by using near-cognate codon 217CUG, rather than annotated 445AUG codon.
Although a Kozak consensus sequence exists surrounding the $^{217}$CUG, efficient protein expression (UFSP1 or GFP) still needs longer 5' UTR sequence (Figs. 3B and S3B). The sequence in 97 to 126 nt upstream of the $^{217}$CUG played a pivotal role in regulating UFSP1 expression efficiency. Previous studies showed a correlation between the efficiency of mRNA translation and secondary structure stability of local mRNA sequence near the initiation codon (33, 34). Therefore, it is possible that 5' UTR sequence of human UFSP1 forms secondary structures (such as hairpin loops, bulges, and internal loops) necessary for the efficiency of translation initiation of UFSP1.

eIF2A plays an important role in translational initiation at non-AUG start codon. Like other CUG-initiated proteins, the translation initiation of human UFSP1 is mediated by eIF2A. Given that phosphorylation of eIF2a leads to a general inhibition of translation, and eIF2A is increased during multiple stress conditions (31), suggesting non-AUG translation is a common stress response mechanism. Although the detailed molecular mechanism remains unclear, UFMylation is closely related to ER homeostasis (8–12). Therefore, we propose that eIF2A mediated $^{217}$CUG initiation of human UFSP1 expression (as well as UFSP1 in other primates) may provide a specific regulatory mechanism in ER homeostasis regulation.

Furthermore, sequence alignment showed that the Cys protease domain of UFSP1 is conserved among organisms from fruit flies to humans (Fig. S6A). However, the CUG translation initiation mechanism of UFSP1 only exists in primates but not in other mammals (Fig. S6B and Table S2), suggesting that specific regulatory mechanism is restricted to primates. Unlike the human UFSP1, mouse UFSP1 has a canonical AUG start site upstream of Cys protease domain coding region and expressed as an active UFM1-specific protease with ~23 kDa (19). Although two recent reports suggest
Figure 5. Human UFSP1 functions in UFMylation/de-UFMylation process. A, human UFSP1 can release UFMylation substrate ASC-1. HEK293T cells were transfected with the indicated plasmids, followed by UFMylation assay and immunoblotting with the indicated antibodies. Red asterisks indicate nonspecific band. B, human UFSP1 cleaves pro-UFM1 in vitro. C-terminal HA-tagged UFM1 was purified using anti-HA affinity gel. UFSP1-L-His or control vector were transfected in HEK293T cells. Cells lysates were subjected to immunoprecipitation with Ni-NTA Agarose. Amounts of UFSP1-His or His-tag
that mouse UFSP1 specifically expresses at the neuromuscular junction and functions as regulators of the acetylcholine receptor clustering, implicating that UFSP1 plays some role in the neuromuscular junction formation during development (35, 36), and the physiological roles of UFSP1 remain largely unknown.

In conclusion, we identified human UFSP1 as an active UFM1-specific protease participating in UFMylation/de-UFMylation process. These findings provide additional insights in our understanding of UFMylation/de-UFMylation process in cells.

Experimental procedures

Plasmids, siRNAs, antibodies, and chemicals

The 5′ UTR and CDS of human UFSP1 gene was amplified by PCR from HeLa cDNA. The fragment of UFSP1 canonical CDS from annotated 445 ATG codon to TGA stop codon is referred to UFSP1-Short. The fragment that contains the canonical CDS and 5′ UTR sequence is referred to UFSP1-Long. A series of truncated 5′ UTR sequence of human UFSP1 with UFSP1 or GFP (without ATG) CDS were constructed and named with the length of 5′ UTR. All of the amplified fragments were subcloned into the indicated vectors. Human UFSP1-L expression plasmids with specific point mutation were constructed using Q5 Site-Directed Mutagenesis Kit (New England Biolabs). UFSP1-L- Random expression plasmids in which the 30 nt sequence between 97 and 126 nt were substituted by a 30 nt random sequence. Prokaryotic expression plasmids of human UFSP1-L were subcloned in the WT or CO UFSP1-L cDNA with 217CTG or 217ATG initiation codon into pET-28a (+) vector using recombinant DNA technology, which destroyed the ATG initiation codon of the pET-28a (+) vector and seamlessly integrated 6 × His-tag at the C terminus. Human UFSP2 cDNA was subcloned into pSG5-HA vector. Human UFSP1-L expression plasmids were transfected into HEK293T or HeLa cells via Lipofectamine 2000 (Invitrogen) transfection. For transduction, target cells (HEK293T cells or HeLa cells) were seeded in 12-well plates and allowed to adhere overnight. The concentrated supernatant and fresh medium were added to the target cells with 8 μg/ml polybrene. Cells were incubated with the virus-containing medium overnight, followed by replacement with fresh medium. After 48 h, puromycin (1 μg/ml for HEK293T cells or HeLa cells) was added to select stable cell lines.

CRISPR/Cas9 mediated KI of FLAG-tag at the C-terminal end of the UFSP1 locus

CRISPR/Cas9–mediated homology-dependent repair was performed as previously described (37). A DNA sequence encoding the FLAG tag (DYKDDDDK) was introduced at the C-terminal end of the UFSP1 locus by CRISPR/Cas9 mediated homology-dependent repair. The single guide RNAs (sgRNAs) targeting human UFSP1, UFSP2, and UFSP1 genes as well as nontargeting control were synthesized and subcloned into the BsmB1 site of lentiCRISPR-sgRNA (Addgene).

The siRNAs targeting human eIF2A and ASC-1 cDNAs were subcloned into p3 × FLAG-HA vector. Human UFSP1 cDNAs were subcloned into pET-28a (+) vector and seamlessly integrated 6 × His-tag at the C terminus. Human UFSP1-L cDNA with 217CTG or 217ATG initiation codon were constructed and transfected into HEK293T cells or HeLa cells via Lipofectamine 2000 (Invitrogen) transfection. For transduction, target cells (HEK293T cells or HeLa cells) were seeded in 12-well plates and allowed to adhere overnight. The concentrated supernatant and fresh medium were added to the target cells with 8 μg/ml polybrene. Cells were incubated with the virus-containing medium overnight, followed by replacement with fresh medium. After 48 h, puromycin (1 μg/ml for HEK293T cells or HeLa cells) was added to select stable cell lines.
homology-dependent repair–mediated insertion of FLAG-tag.

The ssODN is 94 nt, with two 35 nt flanking homology arms and a 24 nt FLAG-tag insertion (5′-GCTTTAGCTCCCAGCAGCAGCCACCTTGGACATTACAAAGGCAGACG

ATGACAAGTGAGGAGCTAGTTAGAATCTCTTCCT

CGGGTC-3′).

HEK293T cells were seeded at 5 × 10^5 per well in a 6-well plate. After 24 h, cells were cotransfected with Cas9 vector expressing USP1-sgRNA-5# and ssODN donor template. Forty-eight hours after transfection, individual cells are sorted into 96-well plates and grown out as single-cell clones. The single-cell clones were lysed and genotyped by FLAG-tag specific PCR screen. For FLAG-tag–positive candidates, the second round of PCR was performed, and PCR products were directly cloned into a blunt-end vector. The homozygous KO clones were selected by Sanger sequencing.

**Generation of CRISPR KO cell lines**

For USP1 or USP2 single gene KO, HEK293T cells with stable expression of Cas9-sgRNA were generated by lentivirus infection. A population of USP1-sgRNA and USP2-sgRNA stable cells was obtained by selecting with 1 μg/ml puromycin. After selecting for 1 to 2 weeks, clones were isolated by limiting dilution and screened for KO by Western blotting analysis. USP1 or USP2 single-gene deficiency monoclonal cell lines were named USP1^KO and USP2^KO, respectively. For USP1 and USP2 double gene KO, USP2^KO cell line was infected with USP1-sgRNA lentivirus. Without puromycin selection, cells were single-cell seeded into 96-wells at 48 h after infection. Clones derived from single cells were screened by Western blotting analysis. Finally, the double gene KO cell line was named USP1^KO/USP2^KO.

**Western blotting**

For Western blotting assay, cells were lysed in radioimmunoprecipitation assay buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 1% nonidet P-40 [NP-40], 0.5% sodium deoxycholate, 1 mM EDTA, and 0.1% SDS) containing 1 μg/ml of protease and phosphatase inhibitor cocktail (Roche). Protein concentrations were determined using Pierce BCA Protein Assay Kit (Thermo Scientific). Protein lysates were resolved by SDS-PAGE, transferred to polyvinylidene fluoride membranes (Millipore), blocked in 5% milk, and probed with primary antibody overnight at 4 °C, and then incubated with horseradish peroxidase–conjugated secondary antibodies. Western blots were visualized with Immobilon Western horseradish peroxidase substrate (Millipore).

**In cell UFMylation and de-UFMylation assay**

ASC-1 was identified as an UFMylation substrate (17). FLAG-tagged ASC-1 was constructed and used for the in vivo UFMylation and de-UFMylation assay. HEK293T cells were transfected with the appropriate constructs. After 48 h, cells were harvested and lysed by boiling in buffer (150 mM Tris–HCl pH 8.0, 5% SDS, and 30% glycerol) for 10 min. Cell lysates were diluted 20-fold with buffer A (50 mM Tris–HCl pH 8.0, 150 mM NaCl, 0.5% NP-40, and 2 mM N-ethylmaleimide) containing 1 μg/ml protease inhibitor cocktail, as described (17). After incubation with anti-FLAG M2 affinity gel (Sigma–Aldrich) overnight at 4 °C, the immunoprecipitates were resolved by SDS-PAGE followed by Western blotting analysis.

**Prokaryotic expression and purification of human USP1**

Prokaryotic expression plasmids of human USP1-L (pET-28a-USP1-217CTG-His, pET-28a-USP1-217ATG-His, pET-28a-USP1-217CTG-His-CO, and pET-28a-USP1-217ATG-His-CO) were transformed into E. coli BL21(DE3) cells for IPTG-induced expression. The positive clones were inoculated into LB medium (containing 50 μg/ml kanamycin) and cultured overnight. Then the bacteria solution was used to inoculate (1% v/v) 5 ml of fresh LB medium in 15 ml tubes and incubated at 37 °C to an A_{600} of 0.6. The expression of human USP1 was induced by the addition of IPTG (0.2 mM). After induction, the bacteria were collected, lysed, and centrifuged. The soluble (supernatant) and insoluble (pellet) fractions were subjected to SDS-PAGE, followed by Coomassie blue staining. Only pET-28a-USP1-217ATG-His-CO plasmid transformed strain expressed human USP1 protein after IPTG induction. And, we used this strain for protein purification of USP1.

The IPTG concentration, temperature, and time of incubation with IPTG were optimized for human USP1 expression in E. coli BL21(DE3) cells. For USP1 protein purification, 200 ml bacteria solution was taken in 300 ml conical flask and incubated at 37 °C to an A_{600} of 0.6. After precocooning, IPTG was added with a final concentration of 0.2 mM and incubation for 16 to 20 h at 16 °C. USP1 protein was purified with Nitritotriacetic acid (NTA) Agarose (Qiagen) as described by the manufacturer. Purified proteins were concentrated and desalted using 10 kDa molecular mass cut-off centrifugal filter units (Amicon Ultra 15 ml). The proteins were then quantified, diluted, and stored at –80 °C in aliquots.

**In vitro enzymatic activity assay for USP1**

The human pro-UFM1 with a C-terminal HA-tag (pSG5-USP1-HA) and USP1-L with a C-terminal 6 × His-tag (USP1-L-His) plasmids were constructed and transfected to HEK293T cells. For the purification of pro-UFM1-HA, cell lysates were incubated with anti-HA affinity gel (Sigma–Aldrich) overnight at 4 °C. The bead-bound pro-UFM1-HA protein was washed and eluted by addition of HA peptide (100 μg/ml). For the purification of USP1-L-His, cell lysates were subjected to Ni-NTA Agarose pulldown overnight at 4 °C. The Ni-NTA resin-bound USP1-L-His proteins were resolved by SDS-PAGE followed by Western blotting analysis.
and 0.1 mM DTT. The reaction products were subjected to SDS-PAGE and evaluated by Western blotting analysis with antibody recognizing UFSP1, His, and UFM1.

**Cell fractionation**

The same amount of HEK293T WT and gene KO cells were collected in cold PBS and pelleted by centrifuging at 1000g. Cytoplasmic, membrane/organellar, and nuclear/cytoskeletal fractions were separated by Cell Fractionation Kit (Cell Signaling Technology) according to the manufacturer’s protocol. Equal volumes of the collected fractions were analyzed by SDS-PAGE and Western blotting analysis.

**LC-MS/MS**

For affinity enrichment of exogenously expressed UFSP1, HEK293T cells were transfected with the C-terminal HA-tagged UFSP1-L construct (UFSP1-L-HA). The cells were lysed with lysis buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, and 0.5% NP-40). UFSP1-HA protein was immunoprecipitated by incubation with anti-HA affinity gel overnight at 4 °C. For affinity enrichment of endogenously expressed UFSP1, C-terminal FLAG-tagged HEK293T cells (FLAG-KI-6#) were lysed and immunoprecipitated by incubation with anti-FLAG M2 affinity gel. All of the aforementioned eluates were resolved by SDS-PAGE, followed by Coomassie blue staining. Desired bands were excised from gels and dissected into small squared blocks. The proteins in excised gel pieces were subjected to in-gel tryptic digestion procedure that involves destaining, reduction, alkylation, digestion, and finally, extraction of peptides for LC-MS/MS analysis. For LC-MS/MS analysis, the peptides were loaded onto a homemade reversed-phase analytical column and separated using a linearly programmed gradient mobile phase consisting of solvent A (0.1% formic acid and 2% acetonitrile in water) and solvent B (0.1% formic acid and 80% acetonitrile in water). The gradient was comprised of an increase from 8% to 35% solvent B over 60 min, 35% to 80% in 5 min, and then holding at 80% for the last 3 min, all at a constant flow rate of 300 nl/min on an EASY-nLC 1000 UPLC system. The peptides were subjected to MS on a Q Exactive HF Mass Spectrometer (Thermo Fisher Scientific). For data processing, the resulting MS/MS data were processed using Maxquant (Max-Planck-Institute of Biochemistry). Tandem mass spectra were searched against Uniprot-Human. The MS analysis was performed by Micrometer Biotech Company, and the lists of human UFSP1 peptide sequences identified are shown in Supplementary Data.

**Statistics and reproducibility**

The ChemiDoc MP Imaging system (Bio-Rad) was used to collect the Western blot images, and Image Lab 6.0.1 (Bio-Rad) and ImageJ 1.52a (National Institutes of Health) were used for image processing and densitometry quantifications. GraphPad Prism 5 (GraphPad Software) and Excel 2007 software (Microsoft) was used for all of the statistical analysis. Data are mean ± SD. Differences in means between two groups were analyzed using two-sided unpaired t test (*p < 0.05). Each experiment was repeated independently with similar results.

**Data availability**

All data that support the findings of this study are available from the corresponding authors upon reasonable request. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD033083 (38).

**Supporting information**—This article contains supporting information.

**Acknowledgments**—We thank members of the Cong lab for helpful discussions. This work was supported by the National Natural Science Foundation of China (32170776, 31730020, 31900520) and the Hangzhou Science and Technology Bureau [20182014B01].

**Author contributions**—Q.L. and Y.-S.C. conceived and designed the experiments; Q.L. and Y.J. performed all of the experiments with help from S.X., J.Z., J.M., X.M and M.W.; Q.L., Y.J. and Y.-S.C. analyzed data; Q.L. and Y.-S.C. wrote the manuscript.

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: 5’ UTR, 5’ untranslated region; ASC-1, activating signal cointegrator 1; ATA, aurintricarboxylic acid; cDNA, complementary DNA; CDS, coding region; ASC-1, activating signal cointegrator 1; ATA, aurintricarboxylic acid; cDNA, complementary DNA; CDS, coding sequence; CO, codon-optimized; ER, endoplasmic reticulum; HA, hemagglutinin; KI, knockin; MS, mass spectrometry; MW, molecular weight; NP-40, nonidet P-40; NTA, nitrilotriacetic acid; sgRNA, single guide RNA; ssODN, oligodeoxyribonucleotide; UBA5, UFM1-activating enzyme 5; UFC1, UFM1-conjugating enzyme 1; UFL1, UFM1-specific ligase 1; UFMB, UFM1-specific protease.
Editors’ Pick: Human UFSP1 is an active UFM1-specific protease

Qian Liang earned his PhD from Northeast Normal University and conducted his postdoc training in Tsinghua University and University of Colorado Boulder. Dr Liang is currently a faculty member in Key Laboratory of Aging and Cancer Biology of Zhejiang Province, Hangzhou Normal University School of Basic Medical Sciences. His research is focusing on the functions and regulatory mechanisms of UFMylation, a recently identified ubiquitin-like posttranslational modification with important biological functions.