The EMBO Journal (2004) 23, 1977–1986 | © 2004 European Molecular Biology Organization | All Rights Reserved 0261-4189/04
www.embojournal.org

THE EMBO JOURNAL

# A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier

**Masaaki Komatsu[1], Tomoki Chiba[1], Kanako Tatsumi[1], Shun-ichiro Iemura[2], Isei Tanida[3], Noriko Okazaki[4], Takashi Ueno[3], Eiki Kominami[3], Tohru Natsume[2] and Keiji Tanaka[1,*]**

[1]Department of Molecular Oncology, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo, Japan, [2]National Institutes of Advanced Industrial Science and Technology, Biological Information Research Center (JBIRC), Kohtoh-ku, Tokyo, Japan, [3]Department of Biochemistry, Juntendo University School of Medicine, Bunkyo-ku, Tokyo, Japan and [4]Kazusa DNA Research Institute, Kazusa-Kamatari, Kisarazu, Chiba, Japan

**Several studies have addressed the importance of various ubiquitin-like (UBL) post-translational modifiers. These UBLs are covalently linked to most, if not all, target protein(s) through an enzymatic cascade analogous to ubiquitylation, consisting of E1 (activating), E2 (conjugating), and E3 (ligating) enzymes. In this report, we describe the identification of a novel ubiquitin-fold modifier 1 (Ufm1) with a molecular mass of 9.1 kDa, displaying apparently similar tertiary structure, although lacking obvious sequence identity, to ubiquitin. Ufm1 is first cleaved at the C-terminus to expose its conserved Gly residue. This Gly residue is essential for its subsequent conjugating reactions. The C-terminally processed Ufm1 is activated by a novel E1-like enzyme, Uba5, by forming a high-energy thioester bond. Activated Ufm1 is then transferred to its cognate E2-like enzyme, Ufc1, in a similar thioester linkage. Ufm1 forms several complexes in HEK293 cells and mouse tissues, revealing that it conjugates to the target proteins. Ufm1, Uba5, and Ufc1 are all conserved in metazoa and plants but not in yeast, suggesting its potential roles in various multicellular organisms.**
*The EMBO Journal* (2004) **23**, 1977–1986. doi:10.1038/sj.emboj.7600205; Published online 8 April 2004
*Subject Categories*: proteins
*Keywords*: Uba5; ubiquitin; ubiquitin fold; ubiquitin-like protein; Ufm1

## Introduction

Protein modification plays a pivotal role in the regulation and expansion of genetic information. In the past two decades, a new type of post-translational protein-modifying system has been identified whose uniqueness is that protein(s) is used as a ligand, that is, modification of protein, by protein, and for protein. A typical system is the ubiquitylation, a modification system in which a single or multiple ubiquitin molecules are attached to a protein, which serves as a signaling player that controls a variety of cellular functions (Hershko and Ciechanover, 1998; Pickart, 2001). Protein ubiquitylation is catalyzed by an elaborate system highly regulated in the cells, which is catalyzed by a sequential reaction of multiple enzymes consisting of activating (E1), conjugating (E2), and ligating (E3) enzymes. E1, which initiates the reaction, forms a high-energy thiolester bond with ubiquitin via adenylation in an ATP-dependent manner. The E1-activated ubiquitin is then transferred to E2 in a thioester linkage. In some cases, E2 can directly transfer the ubiquitin to substrate proteins in an isopeptide linkage; however, E2s mostly requires the participation of E3 to achieve substrate-specific ubiquitylation reaction in the cells. E3s are defined as enzymes required for recognition of specific substrates for ubiquitylation, other than E1 and E2 (Varshavsky, 1997; Bonifacino and Weissman, 1998; Glickman and Ciechanover, 2002).

A set of novel molecules called ubiquitin-like proteins (UBLs) that have structural similarities to ubiquitin has been recently identified (Jentsch and Pyrowolakis, 2000). They are divided into two subclasses: type-1 UBLs, which ligate to target proteins in a manner similar, but not identical, to the ubiquitylation pathway, such as SUMO, NEDD8, and UCRP/ISG15, and type-2 UBLs (also called UDPs, ubiquitin-domain proteins), which contain ubiquitin-like structure embedded in a variety of different classes of large proteins with apparently distinct functions, such as Rad23, Elongin B, Scythe, Parkin, and HOIL-1 (Tanaka *et al*, 1998; Jentsch and Pyrowolakis, 2000; Yeh *et al*, 2000; Schwartz and Hochstrasser, 2003).

In this report, we describe a unique human UBL-type modifier named ubiquitin-fold modifier 1 (Ufm1) that is synthesized in a precursor form consisting of 85 amino-acid residues. We also identified the human activating (Uba5) and conjugating (Ufc1) enzymes for Ufm1. Prior to activation by Uba5, the extra two amino acids at the C-terminal region of the human proUfm1 protein are removed to expose Gly whose residue is necessary for conjugation to target molecule(s). Lastly, we show that the mature Ufm1 is conjugated to yet unidentified endogenous proteins, forming ~28, 38, 47, and 70 kDa complexes in human HEK293 cells and various mouse tissues.

## Results

### *Identification of a novel protein-activating enzyme, Uba5*

Our initial plan was to identify the molecule(s) that interacts with human Atg8p homolog GATE16, a type-1 UBL modifier required for autophagy (Klionsky and Emr, 2000; Ohsumi, 2001), using a yeast two-hybrid screening. Please note that the nomenclature of the autophagy-related genes was recently unified as *ATG* (Klionsky *et al*, 2003). Among several

*Corresponding author. Department of Molecular Oncology, The Tokyo Metropolitan Institute of Medical Science, 3-18-22 Honkomagome, Bunkyo-ku, Tokyo 113-8613, Japan. Tel.: +81 3 3823 2237; Fax: +81 3 3823 2237; E-mail: tanakak@rinshoken.or.jp

**Received: 1 December 2003; accepted: 15 March 2004; published online: 8 April 2004**

©2004 European Molecular Biology Organization

positive clones, we identified fragments of FLJ23251 (Figure 1A), which encodes a 404-amino-acid protein highly conserved in various multicellular organisms, such as *Homo sapiens, Caenorhabditis elegans, Drosophila melanogaster,* and *Arabidopsis thaliana,* but absent in yeasts (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) (Figure 1B). The sequence of FLJ23251 in the region containing residues 72–229 is highly homologous to the corresponding regions in Uba1 (i.e., E1 for ubiquitin) and other E1-like proteins for UBLs including the ATP-binding motif (GXGXXG) (Figure 1A and B). We named this protein Uba5, because it is a member of the E1-like enzyme family. Uba5 also has a metal-binding motif conserved in other E1-like enzymes such as Uba2, Uba3, Uba4, and Atg7. Most of E1-like enzymes have an active site Cys residue within the conserved 10–20 amino-acid residues downstream from the metal-binding

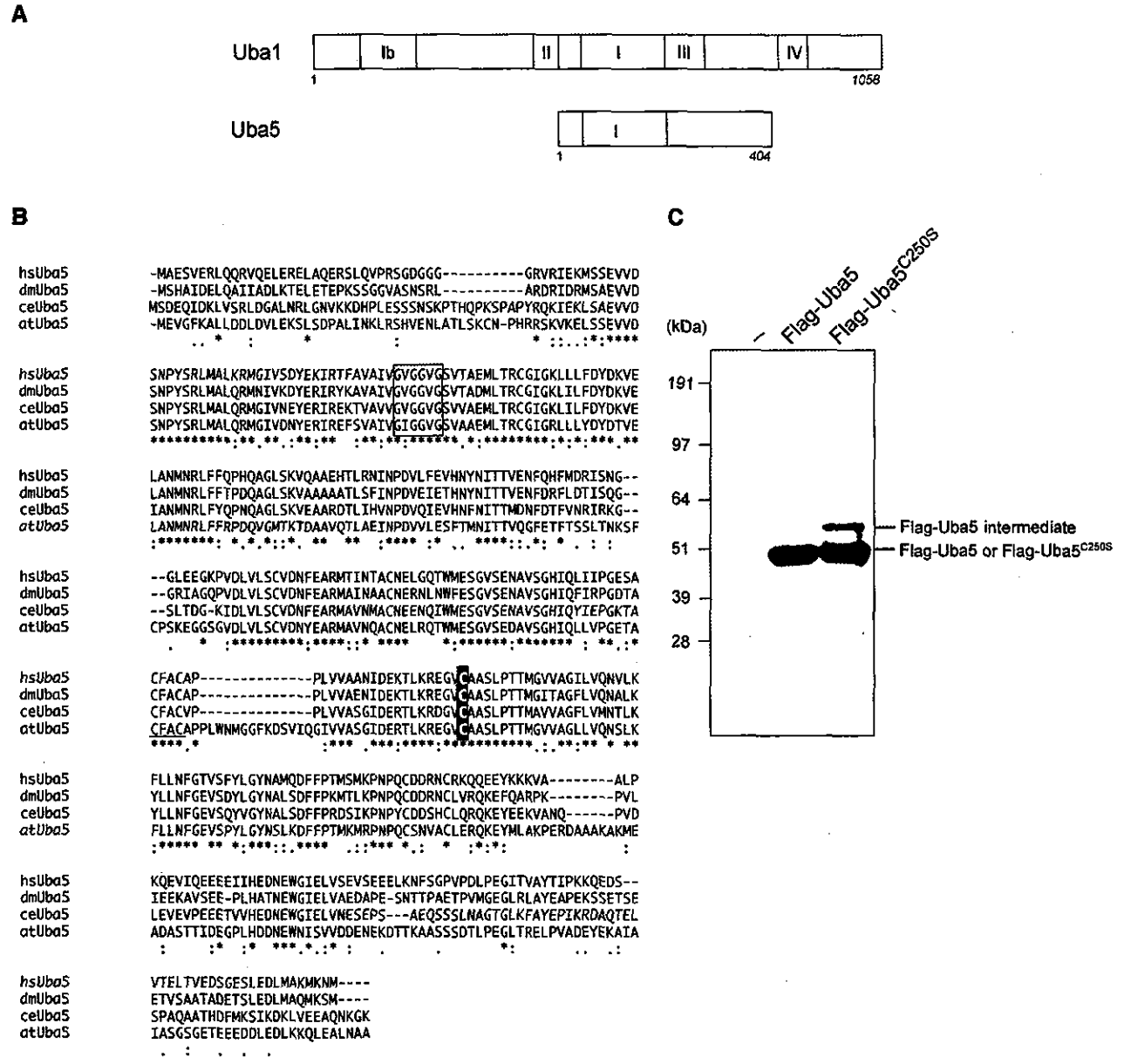



**Figure 1** Uba5, a novel E1-like enzyme. (**A**) Schematic representation of Uba1 and Uba5 in *H. sapiens*. Uba1 is divided into several domains, including I, Ib, II, III, and IV boxes, which are conserved in other E1-like enzymes, and other regions without obvious similarity, described previously (Komatsu *et al*, 2001). Note that Uba5 is of a relatively small size and includes the box I and two other parts. The box I region of Uba1 (amino acids 459–611) has 48.4% similarity and 22.3% identity to amino acids 72–229 of Uba5, which includes the conserved ATP-binding motif (GXGXXG). The sequence of Uba5 is available from GenBank™ under the accession number AK026904. hs, *H. sapiens*; ce, *C. elegans*; dm, *D. melanogaster*; at, *A. thaliana*. (**B**) Sequence alignment of hsUba5 and its homologs of other species (dm, NM_132494; ce, NM_058847; at, NM_100414). The amino-acid sequence of hsUba5 is compared by the ClustalW program. *Asterisks*, identical amino acids; *single* and *double dots*, weakly and strongly similar amino acids, respectively, determined by the criteria of ClustalW program. *Open box* indicates an ATP-binding motif. The putative active site Cys residue is boxed in black. The metal-binding motif is underlined. (**C**) Identification of the intermediate linked to Uba5 in HEK293 cells. Both Uba5 and Uba5$^{C250S}$, in which the predicted active site Cys positioned at 250 was changed to Ser by site-directed mutagenesis, were tagged with Flag peptide at N-terminus, resulting in Flag-Uba5 and Flag-Uba5$^{C250S}$, respectively. Each Flag-Uba5 and Flag-Uba5$^{C250S}$ was expressed in HEK293 cells. The cell lysates were subjected to SDS-PAGE and analyzed by immunoblotting with anti-Flag antibody.


©2004 European Molecular Biology Organization

motif. In the case of Uba5, the $Cys^{250}$ seems to be the most possible active site Cys residue (Figure 1B). If an active site Cys residue within an E1 and E1-like enzymes is changed to Ser, an O-ester bond instead of a thioester bond is formed with its respective modifier protein and the intermediates become stable even under reducing conditions. Therefore, we mutated $Cys^{250}$ within Uba5 to Ser and expressed it as a Flag-fused $Uba5^{C250S}$ (Flag-$Uba5^{C250S}$) or Flag-Uba5 as control in HEK293 cells. As shown in Figure 1C, both Flag-Uba5 and Flag-$Uba5^{C250S}$ were expressed as ~50 kDa proteins in HEK293 cells. When Flag-$Uba5^{C250S}$ was expressed, an additional band with a higher molecular mass of ~60 kDa was clearly observed, indicating that Flag-$Uba5^{C250S}$ forms an intermediate complex with an endogenous protein. These results suggest that Uba5 is indeed a novel protein-activating enzyme for a presumptive modifier (see below).

### *Identification of a novel ubiquitin-fold molecule, Ufm1*

Because Uba5 was identified as GATE-16-binding protein, we initially assumed that Uba5 is another GATE-16-activating enzyme, in addition to Atg7. To test this possibility, we examined whether $Uba5^{C250S}$ (the presumptive active site Cys at position 250 was replaced by Ser) forms an intermediate complex with GATE-16 or not. Unexpectedly, we could not identify a stable complex between $Uba5^{C250S}$ and GATE-16 (data not shown). Therefore, we attempted to identify a protein(s) that physically associates with Uba5 in the cells. To do this, Flag-Uba5 was expressed in HEK293 cells, then immunoprecipitated by anti-Flag antibody. The immunoprecipitates were eluted with a Flag peptide, then digested with Lys-C endopeptides (*Achromobacter* protease I) and the cleaved fragments were directly analyzed using a highly sensitive 'direct nano-flow LC–MS/MS' system as described in Materials and methods. Following database search, a total of 28 peptides were assigned to MS/MS spectra obtained from four nano-LC–MS/MS analyses for the Flag-Uba5-associated complexes. These peptide data identified three proteins as Uba5-associated components: GATE-16, and hypothetical proteins BM-002 and CGI-126 (excluding the bait protein Uba5 and the background proteins, such as HSP70 and keratins).

One of these identified proteins, BM-002, is an 85-amino-acid protein with a predicted molecular mass of ~9.1 kDa. This protein is conserved in multicellular organisms, but not in yeasts, like Uba5 (Figure 2A). The human BM-002 has high identity over the species in the central region but has elongated sequences at both N- and C-terminal regions in some species. Although the protein shows no clear overall sequence identity to ubiquitin or other modifiers (Figure 2B), the tertiary structure of BM-002 displays a striking resemblance to human ubiquitin (Figure 2C). The human structure of BM-002 was constructed by a computer-assisted modeling, based on the structure of its *C. elegans* homolog that has been analyzed previously, as a protein possessing 'ubiquitin-like fold' with secondary structure elements ordered β–β–α–β–β–α (α-helix and β-sheet) along the sequence (Cort *et al*, 2002). Thus, we named human BM-002 as Ufm1.

Ubiquitin is synthesized in a precursor form that must be processed by de-ubiquitylating enzymes (DUBs) to generate a Gly–Gly sequence at the C-terminus. Similarly, Ufm1 has a single Gly residue conserved across species at the C-terminal region, although the length and sequences of amino acids extending from this Gly residue vary among species. To test whether the C-terminus of Ufm1 is post-translationally cleaved, we constructed an expression vector for Ufm1 tagged at both the N- and C-ends, that is, a Flag epitope at the N-terminus and an HA epitope at the C-terminus (Flag-Ufm1-HA) (Figure 2D). After transfection of Flag-Ufm1-HA into HEK293 cells, the cell lysate was subjected to SDS–PAGE, and Flag-Ufm1-HA was detected by immunoblotting. A 10-kDa protein corresponding to Ufm1 was recognized with anti-Flag antibody, while no appreciable protein was observed with anti-HA antibody (Figure 2E, lanes 2 and 7). The mobility on SDS–PAGE was similar to that of Flag-Ufm1ΔC2 (equivalent to mature $Ufm1^{1-83}$ protein) lacking the C-terminal $Ser^{84}$ and $Cys^{85}$ of proUfm1 (Figure 2E, lane 4). These results suggested that the C-terminus of Ufm1 is post-translationally cleaved in the cells, producing mature Ufm1 with the C-terminal $Gly^{83}$ residue. It is known that the replacement of C-terminal Gly residue of Ub and other UBLs with an Ala residue inhibits the C-terminal processing (Kabeya *et al*, 2000; Tanida *et al*, 2003). To examine whether $Gly^{83}$ of Ufm1 is essential for the cleavage, $Gly^{83}$ of Flag-Ufm1-HA was mutated to Ala, and expressed in HEK293 cells (Figure 2D, Flag-$Ufm1^{G83A}$-HA). The mobility of most Flag-$Ufm1^{G83A}$-HA on SDS–PAGE was apparently slower than that of Flag-Ufm1-HA (Figure 2E, lane 3). This mutant was recognized by immunoblotting with anti-HA antibody as well as anti-Flag antibody, suggesting that mutation $Gly^{83}$ to Ala confers resistance to its C-terminal cleavage.

### *Uba5 is an Ufm1-activating enzyme*

We next investigated whether Uba5 forms an intermediate complex with Ufm1. We expressed Flag-Uba5 or Flag-$Uba5^{C250S}$ with Myc-tagged Ufm1 (Myc-Ufm1) in HEK293 cells. Myc-tagged Ufm1ΔC3 lacking the C-terminal $Gly^{83}$ of mature Ufm1 (Myc-Ufm1ΔC3; i.e., deletion form of three residues from precursor $Ufm1^{1-85}$ protein) was used as control. Each cell lysate was prepared and analyzed by immunoblotting with anti-Flag antibody. Flag-$Uba5^{C250S}$ formed an intermediate with an endogenous protein as shown in Figure 1 (Figure 3A, lane 7). When Flag-$Uba5^{C250S}$ was coexpressed with Myc-Ufm1, the intermediate shifted to higher molecular weight (Figure 3A, lane 8). The higher band was not detected when Myc-Ufm1ΔC3 was coexpressed (Figure 3A, lane 9). To verify that the intermediate is indeed the Uba5–Ufm1 complex, Flag-$Uba5^{C250S}$ was immunoprecipitated and blotted with anti-Flag and anti-Myc antibody. Consistent with the above data, a higher sized intermediate was observed when Flag-$Uba5^{C250S}$ was coexpressed with Myc-Ufm1 (Figure 3B, top panel, lane 5), but not alone or with Myc-Ufm1ΔC3 (Figure 3B, top panel, lanes 4 and 6). The intermediate was also recognized by anti-Myc antibody (Figure 3B, lower panel, lane 5), indicating the existence of the Flag-$Uba5^{C250S}$–Myc-Ufm1 complex. Note that the small-sized intermediate is presumably a complex with an endogenous Ufm1, as mentioned. These results indicate that Uba5 forms an intermediate with Ufm1 and the $Gly^{83}$ residue of Ufm1 is essential for the formation of the intermediate with Uba5 *in vivo*.

We subsequently tested whether Uba5 can activate Ufm1 *in vitro*. The thioester formation assay was performed using recombinant proteins expressed in *Escherichia coli*. Recombinant GST-tagged Uba5 and mature Ufm1

©2004 European Molecular Biology Organization

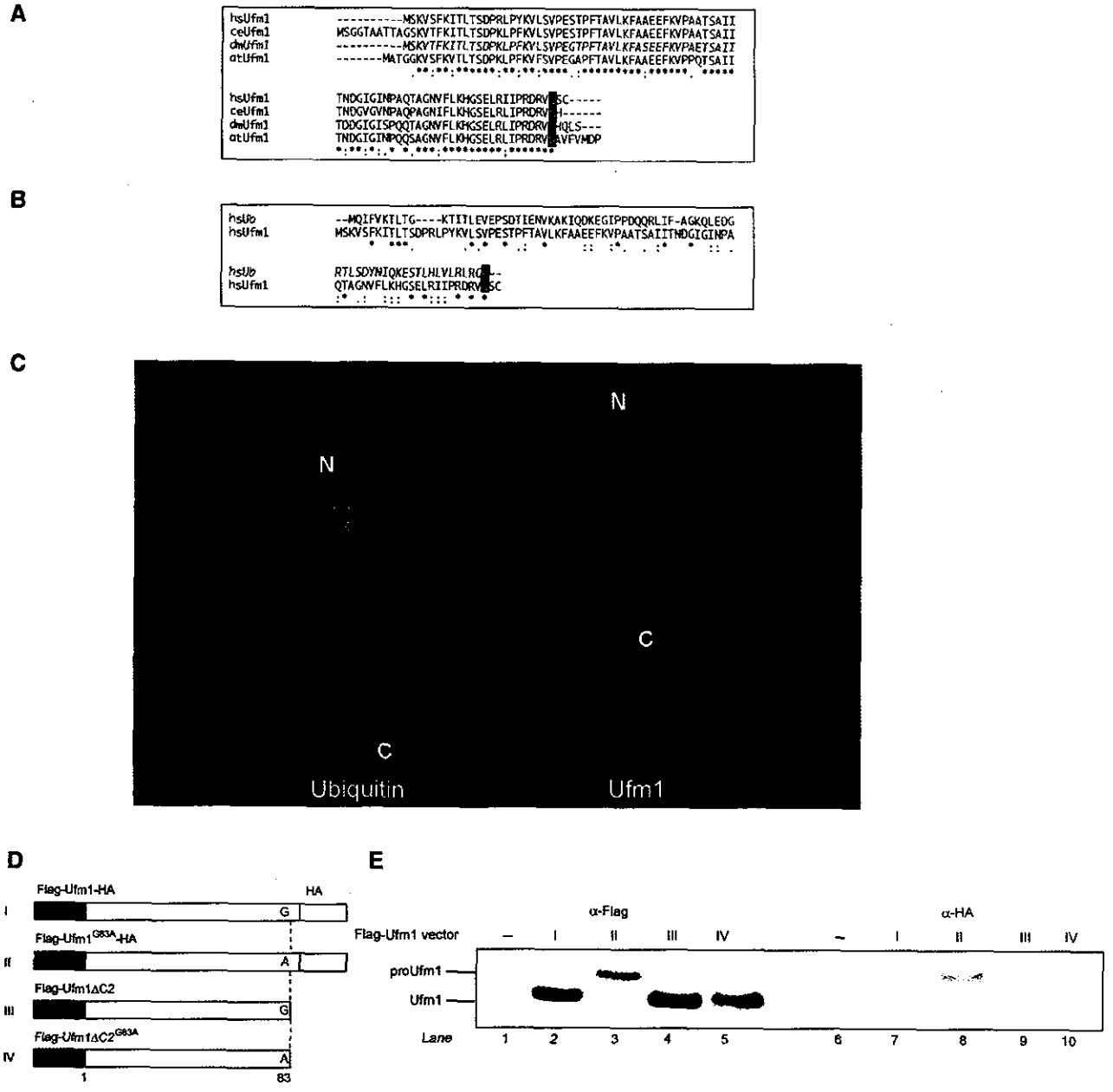



**Figure 2** Ufm1, a novel ubiquitin-fold molecule. (**A**) Sequence alignment of hsUfm1 and its homologs. The sequence of hsUfm1 is available from GenBank™ under the accession number BC005193 (dm, a coding region of dmUfm1 was found from *D. melanogaster* genomic sequence; ce, NM_066304; at, NM_106420). The homology analysis was performed as described in Figure 1B. The C-terminal conserved Gly residue is boxed in black. (**B**) Sequence alignment of hsUbiquitin with hsUfm1. The homology analysis was performed as described in Figure 1B. The C-terminal conserved Gly residue is boxed in black. (**C**) Structural ribbon of hsUbiquitin and predicted structural ribbon of hsUfm1. α-Helices and β-strands are shown in green and yellow, respectively. The homology model of hsUfm1 was created from the *C. elegans* Ufm1 structure (Cort *et al*, 2002) by using MOE program (2003.02; Chemical Computing Group Inc., Montreal, Quebec, Canada). (**D**) Schematic representation of mammalian expression plasmids for Ufm1 and the derivative mutants. Flag epitope tags at the N-terminus, HA epitope tags at the C-terminus, and putative cleavage site $Gly^{83}$ residue (vertical dotted lines) are indicated. To construct $Ufm1^{G83A}$, a single point mutation was introduced into Ufm1, which led to an amino-acid substitution from Gly to Ala at position 83. To construct Ufm1ΔC2, the two C-terminal residues were deleted by PCR. $Ufm1\Delta C2^{G83A}$ was also produced by site-directed mutagenesis of Ufm1ΔC2. The ΔC2 mutants were tagged with the Flag epitopes at N-terminus. (**E**) ProUfm1 processing. HEK293 cells were transfected with Flag-Ufm1-HA, Flag-$Ufm1^{G83A}$-HA, Flag-Ufm1ΔC2, or Flag-$Ufm1\Delta C2^{G83A}$. The cell lysates were subjected to SDS–PAGE and analyzed by immunoblots with anti-Flag and anti-HA antibodies. ProUfm1 and mature Ufm1 are indicated on the left. The numbers at the top from I to IV are similar to those in (D).

(Ufm1ΔC2) with exposed C-terminal $Gly^{83}$ residue were purified, mixed and incubated in the presence of ATP and then subjected to SDS–PAGE at either reducing or nonreducing conditions. GST-Ufm1ΔC3 was used as control. An ~100 kDa band corresponding to the GST-Ufm1ΔC2–GST-Uba5 intermediate complex was clearly observed when the mixture was applied at nonreducing conditions (Figure 3C, lane 3). This intermediate was not observed when ATP or GST-Uba5 was excluded from the mixture (Figure 3C, lanes 1 and 2), or when the mixture was loaded in the presence of


©2004 European Molecular Biology Organization

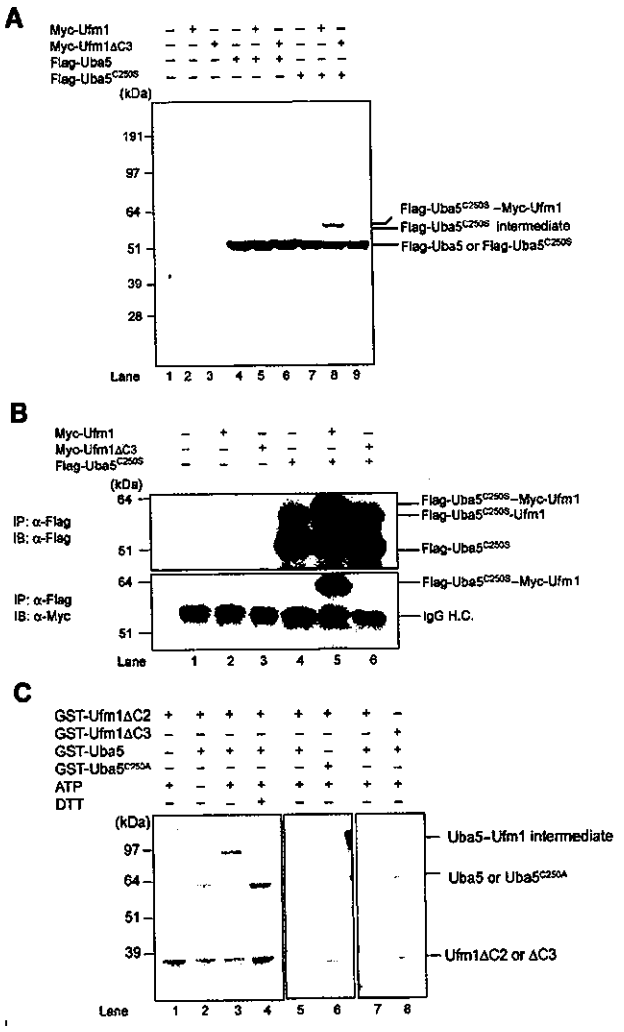



**Figure 3** Demonstration that Uba5 is an Ufm1-activating enzyme. (**A**) Immunoblotting analysis. Each Myc-tagged Ufm1 (Myc-Ufm1) and Myc-Ufm1ΔC3 was expressed alone (lanes 2 and 3, respectively), and coexpressed with Flag-Uba5 (lanes 5 and 6, respectively) or Flag-Uba5$^{C250S}$ (lanes 8 and 9, respectively). Each Flag-Uba5 and Flag-Uba5$^{C250S}$ was also expressed alone (lanes 4 and 7, respectively). The cell lysates were subjected to SDS–PAGE and analyzed by immunoblotting with anti-Flag antibody. The bands corresponding to Flag-Uba5, Flag-Uba5$^{C250S}$, and Flag-Uba5$^{C250S}$ intermediates are indicated on the right. (**B**) Immunoblotting analysis after immunoprecipitation. Each Myc-Ufm1 and Myc-Ufm1ΔC3 was expressed alone (lanes 2 and 3, respectively), and coexpressed with Flag-Uba5$^{C250S}$ (lanes 5 and 6, respectively). Flag-Uba5$^{C250S}$ was also expressed alone (lane 4). The cell lysates were immunoprecipitated with anti-Flag antibody. The resulting immunoprecipitates were subjected to SDS–PAGE and analyzed by immunoblotting with anti-Flag and anti-Myc antibodies. The bands corresponding to Flag-Uba5$^{C250S}$, Flag-Uba5$^{C250S}$–endogenous Ufm1, and Flag-Uba5$^{C250S}$–Myc-Ufm1 intermediates are indicated. (**C**) *In vitro* activating assay of Ufm1 by Uba5. Purified recombinant GST-Ufm1ΔC2 (2 μg) (lanes 1–7) was incubated for 30 min at 25°C with some of the following: 2 μg of purified recombinant GST-Uba5 (lanes 2–5, 7, and 8), GST-Uba5$^{C250A}$ (lane 6), and 5 mM ATP (lanes 1 and 3–8). Lane 8 was conducted similar to lane 7, except that GST-Ufm1ΔC3 was used instead of GST-Ufm1ΔC2. Reactions were then incubated with SDS loading buffer lacking reducing agent (lanes 1–3 and 5–8) or containing 100 mM DTT (lane 4). The presence or absence of various components is indicated above the lanes. The bands corresponding to free GST-Uba5, GST-Uba5$^{C250A}$, GST-Ufm1ΔC2 (mature Ufm1), GST-Ufm1ΔC3, and GST-Uba5–GST-Ufm1ΔC2 thioester product are indicated on the right.

a reducing agent dithiothreitol (DTT) (Figure 3C, lane 4). Furthermore, GST-tagged Uba5$^{C250A}$ mutant, a presumptive active site Cys mutant, could not form the intermediate even at nonreducing conditions (Figure 3C, lane 6). GST-tagged Ufm1ΔC3 was also incapable of forming the intermediate in this reaction (Figure 3C, lane 8). Taken together, we concluded that Uba5 is an Ufm1-activating enzyme and has the active site in Cys$^{250}$.

### *Identification of a novel protein-conjugating enzyme, Ufc1*

The LC–MS/MS analysis revealed CGI-126 protein as another Uba5 interacting protein. CGI-126 is a protein of 167-amino-acid residues with a predicted molecular mass of 19.4 kDa. This protein is also conserved in multicellular organisms, like Uba5 and Ufm1 (Figure 4A). The C-terminal half of human CGI-126 has a high identity across species as shown in Figure 4A. CGI-126 has a highly conserved region, for example, residues 113–126, with limited similarity to the region of Ubc's that encodes an active site Cys residue capable of forming a thioester bond (Figure 4A). We assumed that this protein may be an E2-like conjugating enzyme for Ufm1 and thus named it Ufm1-conjugating enzyme 1 (Ufc1). If Ufc1 is an authentic E2 enzyme for Ufm1, it is expected to form an intermediate complex with Ufm1 via a thioester linkage. To test this possibility in the same way as Uba5, we mutated the predicted active site Cys residue within Ufc1 (Figure 4A, Cys$^{116}$) to Ser. We expressed Flag-Ufc1 or Flag-Ufc1$^{C116S}$ (a presumptive active site Cys at position 116 was replaced by Ser) in combination with Myc-Ufm1 or Myc-Ufm1ΔC3 in HEK293 cells. Flag-Ufc1$^{C116S}$ formed a stable intermediate band when coexpressed with Myc-Ufm1 (Figure 4B, lane 8), but not alone or with Myc-Ufm1ΔC3 (Figure 4B, lanes 7 and 9). To ascertain that this is the Flag-Ufc1$^{C116S}$–Myc-Ufm1 intermediate, Flag-Ufc1$^{C116S}$ was immunoprecipitated and blotted with anti-Myc antibody (Figure 4C). Indeed, Myc-Ufm1, but not Myc-Ufm1ΔC3, formed a complex with Flag-Ufc1$^{C116S}$ (Figure 4C, lanes 5 and 6, top and bottom panels). Note that Flag-Ufc1$^{C116S}$ intermediate with a faster electrophoretic mobility than the Flag-Ufc1$^{C116S}$–Myc-Ufm1 complex is presumably the intermediate with the endogenous Ufm1 (Figure 4C, lanes 4–6, upper panel). These results indicate that Ufc1 forms an intermediate with Ufm1 *in vivo*.

To confirm that Ufc1 is indeed an E2-like enzyme that conjugates with Ufm1 *via* a thioester linkage, we conducted an *in vitro* Ufm1 conjugation assay. Recombinant GST-Uba5, GST-Ufc1, and GST-Ufm1ΔC2 were mixed and incubated in the presence of ATP. GST-Ufc1$^{C116A}$ mutant and GST-Ufm1ΔC3 were used as negative controls. Under nonreducing conditions, an ~70 kDa band corresponding to GST-Ufm1ΔC2–GST-Ufc1 intermediate was observed (Figure 4D, lane 4). This product was not formed at reducing conditions, or when any of the components was omitted from the reaction (Figure 4D, lanes 1–3 and 5). GST-tagged Ufc1$^{C116A}$ mutant could not form the intermediate, suggesting that Cys116 is indeed the active site (Figure 4D, lane 7). GST-Ufm1ΔC3 was again unable to form the intermediate complex in this reaction (Figure 4D, lane 9). Taken together, we concluded that Ufc1 functions as an Ufm1-conjugating enzyme and has the active site in Cys$^{116}$.

©2004 European Molecular Biology Organization

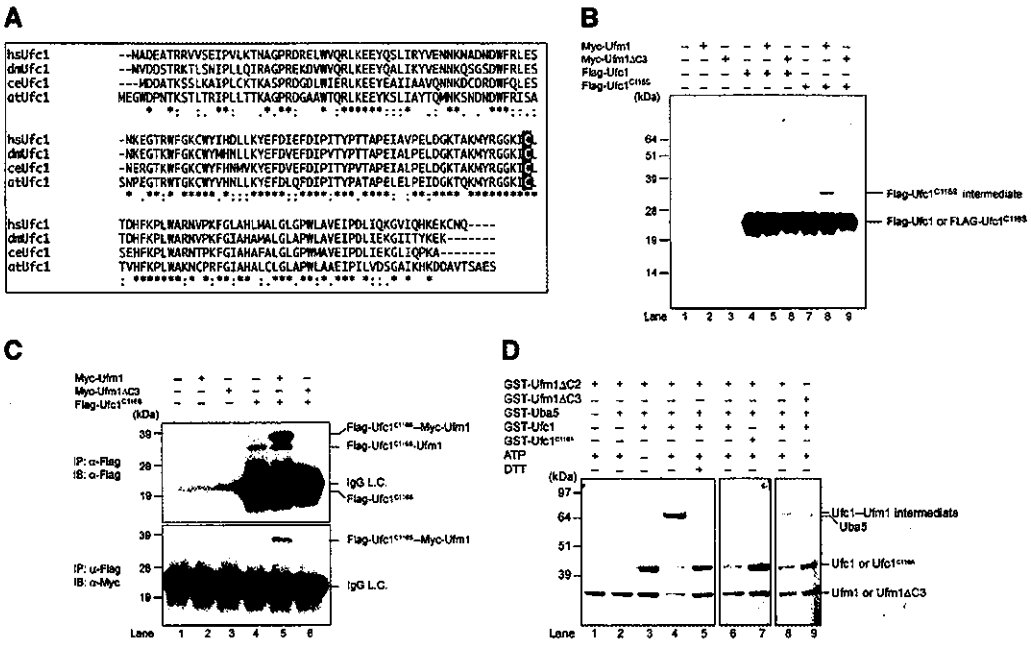



**Figure 4** Ufc1, a novel E2-like enzyme. (**A**) Sequence alignment of hsUfc1 and its homologs. The sequence of Ufc1 is available from GenBank™ under the accession number BC005187 (dm, NM_137230; ce, NM_066654; at, BT001180). The homology analysis was performed as described in Figure 1B. The putative active site Cys residue is boxed in black. (**B**) Immunoblotting analysis. Each Myc-tagged Ufm1 (Myc-Ufm1) and Myc-Ufm1ΔC3 was expressed alone (lanes 2 and 3, respectively), and coexpressed with Flag-Ufc1 (lanes 5 and 6, respectively) or Flag-Ufc1$^{C116S}$ (lanes 8 and 9, respectively). Each Flag-Ufc1 and Flag-Ufc1$^{C116S}$ was also expressed alone (lanes 4 and 7, respectively). The cell lysates were subjected to SDS–PAGE and analyzed by immunoblotting with anti-Flag antibody. The bands corresponding to Flag-Ufc1, Flag-Ufc1$^{C116S}$, and Flag-Ufc1$^{C116S}$ intermediates are indicated on the right. (**C**) Immunoblotting analysis after immunoprecipitation. Each Myc-Ufm1 and Myc-Ufm1ΔC3 was expressed alone (lanes 2 and 3, respectively), and coexpressed with Flag-Ufc1$^{C116S}$ (lanes 5 and 6, respectively). Flag-Ufc1$^{C116S}$ was also expressed alone (lane 4). The cell lysates were immunoprecipitated with anti-Flag antibody. The resulting immunoprecipitates were subjected to SDS–PAGE and analyzed by immunoblots with anti-Flag and anti-Myc antibodies. The bands corresponding to Flag-Ufc1$^{C116S}$, Flag-Ufc1$^{C116S}$–endogenous Ufm1, and Flag-Ufc1$^{C116S}$–Myc-Ufm1 intermediates are indicated. (**D**) *In vitro* thioester bond formation assay of Ufm1 by Ufc1. Purified recombinant GST-Ufm1ΔC2 (2 μg) (lanes 1–8) was incubated for 30 min at 25°C with the following: purified recombinant GST-Uba5 (0.2 μg) (lanes 2–9), GST-Ufc1 (2 μg) (lanes 3–6, 8, and 9), GST-Ufc1$^{C116S}$ (2 μg) (lane 7), and 5 mM ATP (lanes 1, 2, and 4–9). Lane 9 was conducted similar to lane 8, except that GST-Ufm1ΔC3 was used instead of GST-Ufm1ΔC2. Reactions were then incubated with SDS loading buffer lacking reducing agent (lanes 1–4 and 6–9) or containing 100 mM DTT (lane 5). The presence or absence of various components is indicated above the lanes. The bands corresponding to free GST-Ufm1ΔC2 (mature Ufm1), GST-Ufm1ΔC3, GST-Uba5, GST-Ufc1, GST-Ufc1$^{C116S}$, and GST-Ufc1-GST-Ufm1ΔC2 thioester product are indicated on the right.

### *Conjugation of Ufm1 to cellular protein(s)*

We next examined whether Ufm1 conjugates to the target protein(s) in cells. To this end, we expressed Flag- and 6xHis-tagged Ufm1 constructs in HEK293 cells, and purified them under denaturing conditions by $Ni^{2+}$ beads. The resulting precipitates were then analyzed by immunoblotting with anti-Flag antibody. When FlagHis-Ufm1-HA (proUfm1) or FlagHis-Ufm1ΔC2 (mature form) was expressed, several proteins with sizes of about 28, 38, and 47 kDa were detected, in addition to the 10 kDa corresponding to free FlagHis-Ufm1ΔC2 (Figure 5A, lanes 2 and 4). These bands were not detected by FlagHis-Ufm1$^{G83A}$-HA and FlagHis-Ufm1ΔC3, suggesting that both C-terminal cleavage and C-terminal Gly residue are required for the conjugation reaction (Figure 5A, lanes 3 and 6). Moreover, these protein bands were resistant to reducing agents, such as DTT and β-mercaptoethanol. These results indicate that Ufm1 is covalently attached to some target proteins, probably through an isopeptide bond between the C-terminal Gly$^{83}$ of Ufm1 and a Lys residue in the cellular proteins. It is of note that FlagHis-Ufm1$^{G83A}$ mutant with exposed C-terminal Ala instead of Gly can conjugate to target proteins (Figure 5A, lane 5), consistent with the previous report on ubiquitin and SUMO (Hodgins *et al*, 1992; Kamitani *et al*, 1997). Since C-terminal Gly to Ala mutation confers resistance to the Ufm1 processing, the conjugates with FlagHis-Ufm1$^{G83A}$ mutant may be more stable than those with FlagHis-Ufm1ΔC2 (Figure 5A, compare lanes 4 and 5). These results suggest that the Ufm1 conjugation is also a reversible reaction.

We further investigated the expression of Ufm1 and its conjugated proteins in mouse tissues using anti-Ufm1 serum. Ufm1 was widely expressed in all tissues examined, such as brain, heart, lung, liver, and kidney (Figure 5B, left panel). In addition, several bands with striking similarity to proteins detected in HEK293 cells were observed. These bands were not detected by preimmune or preabsorbed antisera (Figure 5B, right panel), suggesting that they are likely the Ufm1 conjugates. Although the intensity of each band varied among tissues and HEK293 cells, 28 and 38 kDa proteins were commonly detected. The 70-kDa band observed in all tissues was also detected faintly in HEK293 cells (Figure 5A, lane 5). The 47-kDa band observed in HEK293 cells was not clear. These protein bands were resistant to reducing agents, such as DTT and β-mercaptoethanol, indicating that Ufm1 covalently attaches to cellular proteins like other Ubl proteins. The targets of Ufm1 appeared to be common in a variety

 ©2004 European Molecular Biology Organization

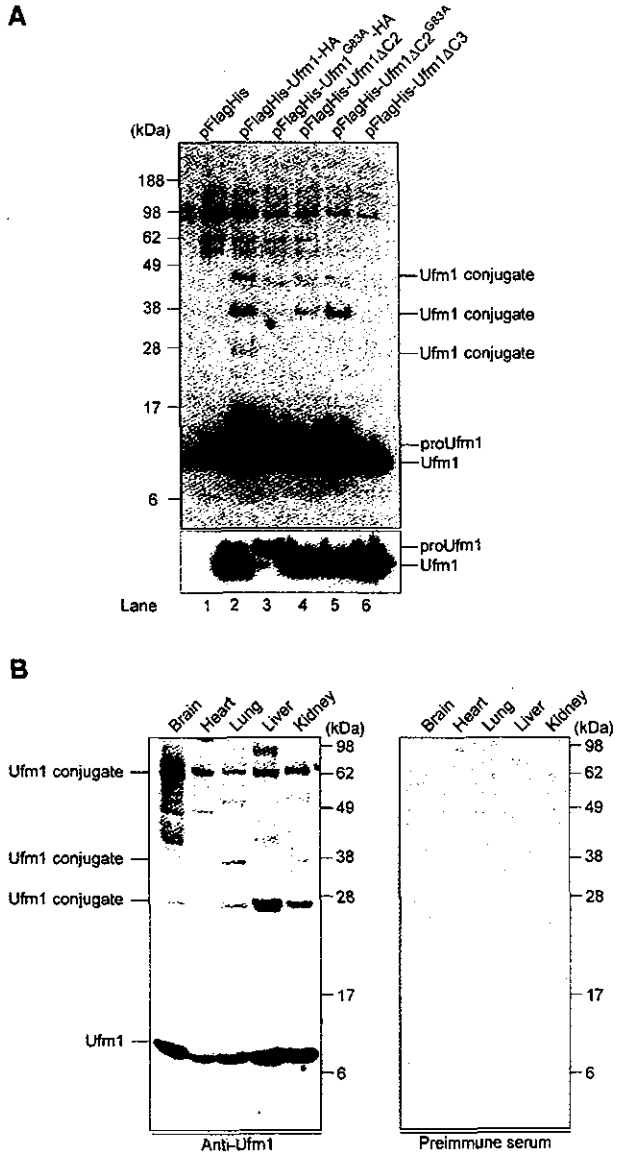



**Figure 5** Formation of a covalent protein conjugate(s) with Ufm1 in HEK293 cells and mouse tissues. (**A**) Ufm1 conjugates in human HEK293 cells. HEK293 cells were transfected with FlagHis-Ufm1-HA, FlagHis-Ufm1$^{G83A}$-HA, FlagHis-Ufm1ΔC2, FlagHis-Ufm1ΔC2$^{G83A}$, or FlagHis-Ufm1ΔC3 expression plasmids. These cells were lysed under denaturing conditions, and the lysates were precipitated with $Ni^{2+}$ beads. The precipitates were subjected to SDS–PAGE and analyzed by immunoblotting with anti-Flag antibody. The bottom panel shows the short exposure of the upper panel. The bands corresponding to mature Ufm1, proUfm, and Ufm1 conjugates are indicated on the right. (**B**) Ufm1 conjugates in various mouse tissues. Homogenates from mouse tissues as indicated were prepared and subjected to SDS–PAGE and analyzed by immunoblotting with anti-Ufm1 serum (left panel) or preimmune serum (right panel). The bands corresponding to Ufm1 and conjugates between Ufm1 and target proteins are indicated on the left.

of tissues. These results suggest the universal roles of Ufm1 in the regulation of cellular function in multicellular organisms.

### *Subcellular localization of Ufm1 in HeLa cells*

We finally examined the subcellular distribution of Ufm1 in HeLa cells. Immunocytochemical analysis using anti-Ufm1

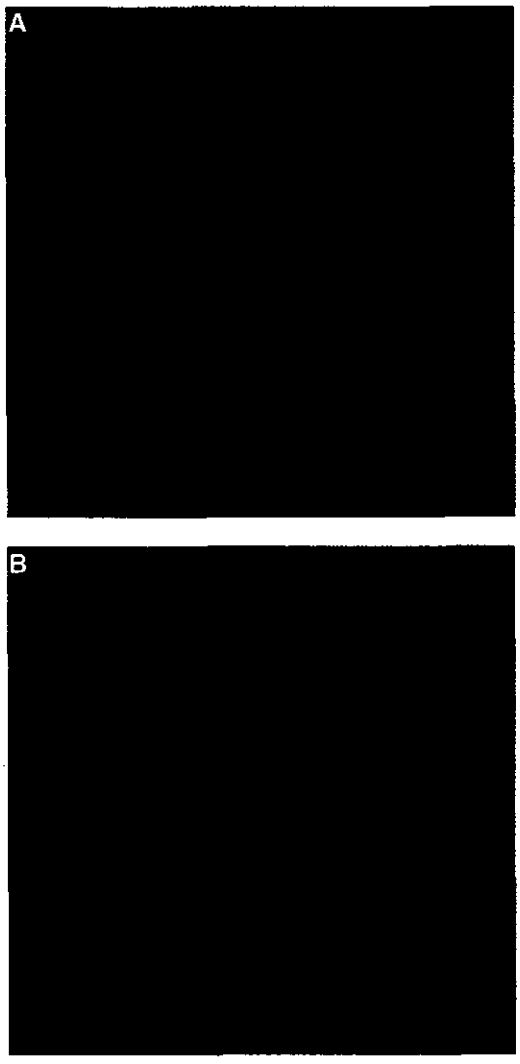



**Figure 6** Intracellular distribution of Ufm1 in HeLa cells. (**A**) HeLa cells were seeded on coverslips 24 h before fixation for immunostaining. Ufm1 was detected with anti-Ufm1 serum and visualized with Alexa 488 nm anti-rabbit antibody. (**B**) Immunocytochemical analysis was conducted as for (A), except that preimmune serum was used. Cells were observed using a fluorescence microscope. Magnification, × 400.

serum revealed that Ufm1 was predominantly localized in the nucleus and diffusely in the cytoplasm (Figure 6A). These staining patterns were not observed when anti-Ufm1 serum had been preadsorbed with excess amounts of recombinant Ufm1 protein or preimmune serum was used instead of anti-Ufm1 serum (Figure 6B). Moreover, Ufm1 localization in the cytoplasm and nucleus was similar to the localization of exogenously expressed GFP-tagged Ufm1 in HeLa cells (data not shown). In the nucleus, strong immunoreactivity to anti-Ufm1 serum was observed as a dot-like structure. Although such dots-like structures were detected by preimmune serum, those intensities were weak. Thus, some of these dot-like structures may represent conjugates of Ufm1.

## Discussion

In the present study, we reported that Ufm1 acts as a new post-translational UBL modifier, based on the following

©2004 European Molecular Biology Organization

criteria: (1) It is a small protein of 9.1 kDa with a ubiquitin-fold structure. (2) Ufm1 is synthesized in a precursor form, and the extra amino-acid residues at the C-terminal side need to be processed to expose the Gly residue. (3) The C-terminal processing and exposure of glycine residue are essential to the formation of Ufm conjugates in the cells. (4) Ufm1 has specific E1-like (Uba5) and E2-like (Ufc1) enzymes for activation and conjugation, respectively. Intriguingly, many UBL modifiers are evolutionarily conserved from yeast to human, except interferon-inducible UBL modifiers, such as UCRP/ISG15, Fat10, and Fau1/MNSFβ (Nakamura *et al*, 1995; D'Cunha *et al*, 1996; Liu *et al*, 1999). Ufm1, Uba5, and Ufc1 found in the present study are conserved in various multicellular organisms (Figures 1B, 2A, and 4A), but not in both budding and fission yeasts, suggesting that they all have been generated by coevolution.

We identified Uba5 as an E1 enzyme for Ufm1. This enzyme is relatively small compared to Uba1, that is, an E1 for ubiquitin (Figure 1A). In the *in vitro* assay, the recombinant Uba5 protein formed a thioester linkage with recombinant Ufm1 (Figure 3C) and transferred the activated Ufm1 to recombinant Ufc1 (an E2 enzyme) (Figure 4D), indicating that Uba5 can activate Ufm1 as a single molecule. This is in marked contrast to other E1s such as Uba2 and Uba3, which retain obvious similarities to the C-terminal half of Uba1 but require the formation of heterodimer complexes with respective partner molecules, AOS1 and APP-BP1, respectively, with similarities to the N-terminal half of Uba1 (Johnson *et al*, 1997; Liakopoulos *et al*, 1998; Osaka *et al*, 1998). Another E1-like enzyme, Uba4 that activates Urm1, is of similar size to Uba5 (Furukawa *et al*, 2000), but it remains unknown whether Uba4 acts as a single molecule or needs a partner subunit. The homology of Uba5 to Uba1 is less than those of Uba2 and Uba3, except their ThiF domain conserved in E1s, and thus it is likely that Uba5 may uniquely activate Ufm1, differing from other E1s such as Uba1, Uba2/AOS1, and Uba3/APP-BP1. Thus, although the structure of APP-BP1/Uba3 heterodimer is determined and the mechanism by which E1s activate their cognate UBLs was proposed (Walden *et al*, 2003a, b), the weak homology of Uba5 with other E1s hampered the computer-assisted structural analysis. To clarify this issue, structural analysis of Uba5 is required. This issue is currently under investigation in our laboratories. So far, most E1-like enzymes activate single species of UBL protein, although Atg7 is exception, which can activate both Atg8 and Atg12 (Mizushima *et al*, 1998; Tanida *et al*, 1999; Ichimura *et al*, 2000). A total of 10 E1-like enzymes can be identified in the human genome by computer analysis. Considering the limited number of E1-like proteins, it is possible that some E1-like proteins can activate a distinct set of UBL proteins. Whether or not Uba5 is capable of activating proteins other than Ufm1 remains to be clarified.

There are more than a dozen of E2 family genes in human genomes. In the budding yeast, 13 different E2s, namely Ubc1–Ubc13, have been documented and functionally characterized. Functionally, most of them catalyze the conjugation of ubiquitin, except that Ubc9 and Ubc12 are for SUMO and NEDD8/Rub1, respectively (Johnson and Blobel, 1997; Lammer *et al*, 1998; Osaka *et al*, 1998). In addition, in the autophagic pathway, Atg3 and Atg10 are both E2 enzymes for Atg8 and Atg12, respectively, but they do not have obvious sequence similarities to known Ubc's, except for a short region encompassing an active Cys residue (Shintani *et al*, 1999; Ichimura *et al*, 2000). Similarly, Ufc1 is a unique E2-like enzyme with no obvious sequence homology with other E2s, except approximately 10 amino-acid residues encompassing the active site Cys residue.

In assessing the biological roles of the Ufm1-modifying system, characterization of the target molecule(s) is of particular importance. Regarding this issue, we identified several putative proteins that are conjugated with Ufm1 in human HEK293 cells and various mouse tissues. It is noteworthy that the sizes of these bands (28, 38, 47 kDa) increase by 10 kDa, which is consistent with the size of Ufm1. Considering that several Ubl modifiers can attach to target proteins as a polymer, it is possible that these bands correspond to multi- or poly-Ufm1 conjugates. In fact, Ufm1 has six Lys residues. Whether Ufm1 is conjugated to several distinct proteins or multiple Lys residues in a single target or polymerized in a single Lys residue awaits future study. Unfortunately, we could not identify the protein, and detailed analysis of the cellular function of Ufm1 conjugation awaits future study. It was recently reported that Uba5 is induced by certain reagents that induce stress in the endoplasmic reticulum (ER), a so-called 'unfolded protein response' (Harding *et al*, 2003). However, we could not observe the induction of Uba5, Ufc1, and Ufm1 by treatment with various compounds known to induce ER stress in mammalian cells (data not shown). In addition, exposure to other stresses including high temperature or heavy metals also did not induce the appearance of obvious new conjugation band(s) of Ufm1, by immunoblot analysis. Further studies on the biological roles of the Ufm1 conjugation pathway are under investigation in our laboratories.

## Materials and methods

### *DNA construction*

The cDNA encoding human Uba5 was obtained by PCR from human liver cDNA with the Uba5-s5′ primer (5′-CGGAGGGATCCC CATGGCGGAGTCTGTGGAG-3′) and the Uba5-r3′ primer (5′-CAGTCCTCGAGCTACATATTCTTCATTTT-3′). It was then subcloned into pcDNA3 vector (Invitrogen, San Diego, CA). A point mutation for Cys at position 250 to Ser or Ala was generated by PCR-based site-directed mutagenesis. The Flag tag was introduced at the N-terminus of Uba5 or Uba5$^{C250S}$. Similarly, cDNA encoding human Ufm1 was amplified by PCR from human liver cDNA with the Ufm1-s5′ primer (5′-TTCCGGGATCCCCATGTCGAAGGTTTCCTTT-3′) and the Ufm1-r3′ primer (5′-AGTAGCTCGAGTTAACAACTTCCAA CACGAT-3′), and subcloned into pcDNA3 vector. The Flag, FlagHis, or Myc tags were introduced at the N-terminus of Ufm1. The HA tag was introduced at the C-terminus of Ufm1. The C-terminal deletion mutants of Ufm1 named Ufm1ΔC2 and Ufm1ΔC3, encoding amino acids 1–83 and 1–82, respectively, were generated by PCR. A point mutation for Gly at position 83 to Ala of Ufm1 and Ufm1ΔC2 (Ufm1$^{G83A}$ and Ufm1ΔC2$^{G83A}$, respectively) was generated by PCR-based site-directed mutagenesis. The cDNA encoding human Ufc1 was obtained by PCR from human liver cDNA with the Ufc1-s5′ primer (5′-GCCCTGGATCCAGATGGCGGATGAAGCCACG-3′) and the Ufc1-r3′ primer (5′-TTCTCGAGTCATTGGTTGCATTTCTCTT-3′). It was then subcloned into pcDNA3 vector. A point mutation for Cys at position 116 to Ser or Ala was generated by PCR-based site-directed mutagenesis. The Flag tag was introduced at the N-terminus of Ufc1 and Ufc1$^{C116S}$. To express GST-fused Ufm1ΔC2, Ufm1ΔC3, Uba5, Uba5$^{C250A}$, Ufc1, and Ufc1$^{C116A}$ in *E. coli*, these cDNAs were subcloned into pGEX-6p vector (Amersham Biosciences). All mutations mentioned above were confirmed by DNA sequencing.


©2004 European Molecular Biology Organization

### *Cell culture and transfection*

Media and reagents for cell culture were purchased from Life Technologies (Grand Island, NY). HEK293 cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS), 5 U/ml penicillin, and 50 μg/ml streptomycin. HEK293 cells at subconfluence were transfected with the indicated plasmids using Fugene 6 reagent (Roche Molecular Biochemicals, Mannheim, Germany). Cells were analyzed at 20–24 h after transfection.

### *Immunological analysis*

For immunoblot analysis, cells were lysed with ice-cold TNE buffer (10 mM Tris–HCl, pH 7.5, 1% Nonidet P-40, 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and protease inhibitors) and the lysates were separated by SDS–PAGE (12% gel or 4–12% gradient gel) and transferred to a polyvinylidene difluoride (PVDF) membrane. Mouse monoclonal anti-Flag antibody (M2; Sigma Chemical Co., St Louis, MO), anti-HA antibody (F7; Santa Cruz Biotechnology, Santa Cruz, CA), and rabbit polyclonal anti-Myc antibody (N14; Santa Cruz) were used for immunodetection. Development was performed by the Western lighting detection methods.

For immunoprecipitation analysis, cells were lysed by 200 μl of TNE, and the lysate was then centrifuged at 10 000 *g* for 10 min at 4°C to remove debris. In the next step, 800 μl of TNE and 30 μl of M2-agarose (Sigma) were added to the lysate, and the mixture was mixed under constant rotation for 12 h at 4°C. The immunoprecipitates were washed five times with ice-cold TNE. The complex was boiled for 10 min in SDS sample buffer in the presence of β-mercaptoethanol to elute proteins and centrifuged at 10 000 *g* for 10 min at 4°C. The supernatant was subjected to SDS–PAGE, transferred to PVDF membrane, and analyzed by immunoblots with anti-Flag (M2) or anti-Myc (N14) antibody.

For purification of 6xHis-tagged proteins under denaturing conditions, cells were lysed by 1 ml of denaturing lysis buffer (8 M urea, 0.1 M $NaH_2PO_4$, and 0.01 M Tris–HCl, pH 8.0) in the presence of 20 mM *N*-ethylmaleimide as an inhibitor of isopeptidases, and the lysate was sonicated briefly and then centrifuged at 10 000 *g* for 10 min at room temperature to remove debris. Then, 30 μl of Ni-NTA Superflow (QIAGEN) was added to the lysate, and the mixture was shaken under constant rotation for 30 min at room temperature. The precipitates were washed five times with denaturing wash buffer (8 M urea, 0.1 M $NaH_2PO_4$, and 0.01 M Tris–HCl, pH 5.9). To elute proteins, elution buffer (8 M urea, 0.1 M $NaH_2PO_4$, and 0.01 M Tris–HCl, pH 4.5) was added to the complex, and the mixture was centrifuged at 10 000 *g* for 10 min at room temperature. The resulting supernatant was subjected to SDS–PAGE, transferred to PVDF membrane, and analyzed by immunoblots with anti-Flag (M2).

Freshly isolated tissues from mice were homogenized in lysis buffer (50 mM Tris–HCl, pH 7.5, 1% SDS, 5 mM EDTA, and 10 mM β-mercaptoethanol) using potter-Elvehjem homogenizer. The homogenate was centrifuged at 10 000 *g* for 10 min to remove debris. The resulting supernatant was subjected to SDS–PAGE, transferred to PVDF membrane, and analyzed by immunoblotting with anti-Ufm1 or preimmune serum. The anti-Ufm1 polyclonal antibody was raised in rabbits using the recombinant protein produced in *E. coli* as an antigen.

### *In vitro thioester formation assay*

Recombinant GST-Ufm1ΔC2, GST-Ufm1ΔC3, GST-Uba5, GST-Uba5$^{C250A}$, GST-Ufc1, and GST-Ufc1$^{C116A}$ (tagged N-terminally with GST) were produced in *E. coli* and recombinant proteins were purified by chromatography on glutathione sepharose 4B (Amersham Biosciences). After elution of proteins from the beads, the preparations were dialyzed against 50 mM BisTris (pH 6.5), 100 mM NaCl, 10 mM $MgCl_2$, and 0.1 mM DTT (reaction buffer). Most thioester formation reactions contained reaction buffer with 4 μg GST-Ufm1ΔC2 or GST-Ufm1ΔC3 and some of the following: 5 mM ATP, 2 or 0.2 μg GST-Uba5 or GST-Uba5$^{C250A}$, and 4 μg GST-Ufc1 or GST-Ufc1$^{C116A}$. Reactions were incubated for 30 min at 25°C and stopped by the addition of SDS-containing loading buffer either lacking reducing agent or containing 100 mM DTT, followed by a 10 min incubation at 37°C, SDS–PAGE (4–12% acrylamide gradient) and Coomassie brilliant blue staining.

### *Protein identification by LC–MS/MS analysis*

The Uba5-associated complexes were digested with *Achromobacter* protease I and the resulting peptides were analyzed using a nanoscale LC–MS/MS system as described previously (Natsume *et al*, 2002). The peptide mixture was applied to a Mightysil-PR-18 (1 μm particle, Kanto Chemical) frit-less column (45 mm × 0.150 mm ID) and separated using a 0–40% gradient of acetonitrile containing 0.1% formic acid over 30 min at a flow rate of 50 nl/min. Eluted peptides were sprayed directly into a quadropole time-of-flight hybrid mass spectrometer (Q-Tof *Ultima*, Micromass, Manchester, UK). MS and MS/MS spectra were obtained in a data-dependent mode. Up to four precursor ions above an intensity threshold of 10 counts/s were selected for MS/MS analyses from each survey scan. All MS/MS spectra were searched against protein sequences of Swiss Prot and RefSeq (NCBI) using batch processes of Mascot software package (Matrix Science, London, UK). The criteria for match acceptance were the following: (1) When the match score was 10 over each threshold, identification was accepted without further consideration. (2) When the difference of score and threshold was lower than 10, or when proteins were identified based on a single matched MS/MS spectrum, we manually confirmed the raw data prior to acceptance. (3) Peptides assigned by less than three y series ions and peptides with +4 charge state were all eliminated regardless of their scores.

### *Immunofluorescence*

HeLa cells grown on glass coverslips were fixed in 4% paraformaldehyde (PFA) in PBS for 15 min, and permeabilized with 0.2% (vol/vol) Triton X-100 in PBS for 30 min. After permeabilization, the cells were blocked for 30 min with 5% (vol/vol) normal goat serum in PBS, incubated for 1 h at 37°C with anti-Ufm1 serum or preimmune serum, washed with PBS, and incubated for 30 min with Alexa 488 nm anti-rabbit antibodies (Molecular Probes). The coverslips were washed and mounted on slides. Fluorescence images were obtained using a fluorescence microscope (DMIRE2; Leica) equipped with a cooled charge-coupled device camera (CTR MIC; Leica). Pictures were taken using Leica Qfluoro software (Leica).

## Acknowledgements

We thank T Mizushima (Nagoya University) for the computer-assisted structural modeling of human Ufm1. This work was supported in part by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## References

Bonifacino JS, Weissman AM (1998) Ubiquitin and the control of protein fate in the secretory and endocytic pathways. *Annu Rev Cell Dev Biol* **14:** 19–57

Cort JR, Chiang Y, Zheng D, Montelione GT, Kennedy MA (2002) NMR structure of conserved eukaryotic protein ZK652.3 from *C. elegans*: a ubiquitin-like fold. *Proteins* **48:** 733–736

D'Cunha J, Knight Jr E, Haas AL, Truitt RL, Borden EC (1996) Immunoregulatory properties of ISG15, an interferon-induced cytokine. *Proc Natl Acad Sci USA* **93:** 211–215

Furukawa K, Mizushima N, Noda T, Ohsumi Y (2000) A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes. *J Biol Chem* **275:** 7462–7465

Glickman MH, Ciechanover A (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev* **82:** 373–428

Harding HP, Zhang Y, Zeng H, Novoa I, Lu PD, Calfon M, Sadri N, Yun C, Popko B, Paules R, Stojdl DF, Bell JC, Hettmann T, Leiden JM, Ron D (2003) An integrated stress response regulates amino

©2004 European Molecular Biology Organization

acid metabolism and resistance to oxidative stress. *Mol Cell* **11:** 619–633

Hershko A, Ciechanover A (1998) The ubiquitin system. *Annu Rev Biochem* **67:** 425–479

Hodgins RR, Ellison KS, Ellison MJ (1992) Expression of a ubiquitin derivative that conjugates to protein irreversibly produces phenotypes consistent with a ubiquitin deficiency. *J Biol Chem* **267:** 8807–8812

Ichimura Y, Kirisako T, Takao T, Satomi Y, Shimonishi Y, Ishihara N, Mizushima N, Tanida I, Kominami E, Ohsumi M, Noda T, Ohsumi Y (2000) A ubiquitin-like system mediates protein lipidation. *Nature* **408:** 488–492

Jentsch S, Pyrowolakis G (2000) Ubiquitin and its kin: how close are the family ties? *Trends Cell Biol* **10:** 335–342

Johnson ES, Blobel G (1997) Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. *J Biol Chem* **272:** 26799–26802

Johnson ES, Schwienhorst I, Dohmen RJ, Blobel G (1997) The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. *EMBO J* **16:** 5509–5519

Kabeya Y, Mizushima N, Ueno T, Yamamoto A, Kirisako T, Noda T, Kominami E, Ohsumi Y, Yoshimori T (2000) LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *EMBO J* **19:** 5720–5728

Kamitani T, Nguyen HP, Yeh ET (1997) Preferential modification of nuclear proteins by a novel ubiquitin-like molecule. *J Biol Chem* **272:** 14001–14004

Klionsky DJ, Cregg JM, Dunn Jr WA, Emr SD, Sakai Y, Sandoval IV, Sibirny A, Subramani S, Thumm M, Veenhuis M, Ohsumi Y (2003) A unified nomenclature for yeast autophagy-related genes. *Dev Cell* **5:** 539–545

Klionsky DJ, Emr SD (2000) Autophagy as a regulated pathway of cellular degradation. *Science* **290:** 1717–1721

Komatsu M, Tanida I, Ueno T, Ohsumi M, Ohsumi Y, Kominami E (2001) The C-terminal region of an Apg7p/Cvt2p is required for homodimerization and is essential for its E1 activity and E1–E2 complex formation. *J Biol Chem* **276:** 9846–9854

Lammer D, Mathias N, Laplaza JM, Jiang W, Liu Y, Callis J, Goebl M, Estelle M (1998) Modification of yeast Cdc53p by the ubiquitin-related protein rub1p affects function of the SCFCdc4 complex. *Genes Dev* **12:** 914–926

Liakopoulos D, Doenges G, Matuschewski K, Jentsch S (1998) A novel protein modification pathway related to the ubiquitin system. *EMBO J* **17:** 2208–2214

Liu YC, Pan J, Zhang C, Fan W, Collinge M, Bender JR, Weissman SM (1999) A MHC-encoded ubiquitin-like protein (FAT10) binds noncovalently to the spindle assembly checkpoint protein MAD2. *Proc Natl Acad Sci USA* **96:** 4313–4318

Mizushima N, Noda T, Yoshimori T, Tanaka Y, Ishii T, George MD, Klionsky DJ, Ohsumi M, Ohsumi Y (1998) A protein conjugation system essential for autophagy. *Nature* **395:** 395–398

Nakamura M, Xavier RM, Tsunematsu T, Tanigawa Y (1995) Molecular cloning and characterization of a cDNA encoding monoclonal nonspecific suppressor factor. *Proc Natl Acad Sci USA* **92:** 3463–3467

Natsume T, Yamauchi Y, Nakayama H, Shinkawa T, Yanagida M, Takahashi N, Isobe T (2002) A direct nanoflow liquid chromatography–tandem mass spectrometry system for interaction proteomics. *Anal Chem* **74:** 4725–4733

Ohsumi Y (2001) Molecular dissection of autophagy: two ubiquitin-like systems. *Nat Rev Mol Cell Biol* **2:** 211–216

Osaka F, Kawasaki H, Aida N, Saeki M, Chiba T, Kawashima S, Tanaka K, Kato S (1998) A new NEDD8-ligating system for cullin-4A. *Genes Dev* **12:** 2263–2268

Pickart CM (2001) Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70:** 503–533

Schwartz DC, Hochstrasser M (2003) A superfamily of protein tags: ubiquitin, SUMO and related modifiers. *Trends Biochem Sci* **28:** 321–328

Shintani T, Mizushima N, Ogawa Y, Matsuura A, Noda T, Ohsumi Y (1999) Apg10p, a novel protein-conjugating enzyme essential for autophagy in yeast. *EMBO J* **18:** 5234–5241

Tanaka K, Suzuki T, Chiba T (1998) The ligation systems for ubiquitin and ubiquitin-like proteins. *Mol Cells* **8:** 503–512

Tanida I, Komatsu M, Ueno T, Kominami E (2003) GATE-16 and GABARAP are authentic modifiers mediated by Apg7 and Apg3. *Biochem Biophys Res Commun* **300:** 637–644

Tanida I, Mizushima N, Kiyooka M, Ohsumi M, Ueno T, Ohsumi Y, Kominami E (1999) Apg7p/Cvt2p: a novel protein-activating enzyme essential for autophagy. *Mol Biol Cell* **10:** 1367–1379

Varshavsky A (1997) The ubiquitin system. *Trends Biochem Sci* **22:** 383–387

Walden H, Podgorski MS, Huang DT, Miller DW, Howard RJ, Minor Jr DL, Holton JM, Schulman BA (2003a) The structure of the APPBP1–UBA3–NEDD8–ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. *Mol Cell* **12:** 1427–1437

Walden H, Podgorski MS, Schulman BA (2003b) Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. *Nature* **422:** 330–334

Yeh ET, Gong L, Kamitani T (2000) Ubiquitin-like proteins: new wines in new bottles. *Gene* **248:** 1–14



©2004 European Molecular Biology Organization

nature
structural &
molecular biology

# Structural basis of sugar-recognizing ubiquitin ligase

Tsunehiro Mizushima[1,2], Takeshi Hirao[3], Yukiko Yoshida[2,4], Soo Jae Lee[5], Tomoki Chiba[2], Kazuhiro Iwai[4,6], Yoshiki Yamaguchi[3,4], Koichi Kato[3,4,7], Tomitake Tsukihara[5] & Keiji Tanaka[2]

Reprinted from nature structural & molecular biology, volume 11, april 2004

nature
structural &
molecular biology

# Structural basis of sugar-recognizing ubiquitin ligase

Tsunehiro Mizushima[1,2], Takeshi Hirao[3], Yukiko Yoshida[2,4], Soo Jae Lee[5], Tomoki Chiba[2], Kazuhiro Iwai[4,6], Yoshiki Yamaguchi[3,4], Koichi Kato[3,4,7], Tomitake Tsukihara[5] & Keiji Tanaka[2]

**$SCF^{Fbs1}$ is a ubiquitin ligase that functions in the endoplasmic reticulum (ER)-associated degradation pathway. Fbs1/Fbx2, a member of the F-box proteins, recognizes high-mannose oligosaccharides. Efficient binding to an N-glycan requires di-*N*-acetylchitobiose (chitobiose). Here we report the crystal structures of the sugar-binding domain (SBD) of Fbs1 alone and in complex with chitobiose. The SBD is composed of a ten-stranded antiparallel β-sandwich. The structure of the SBD–chitobiose complex includes hydrogen bonds between Fbs1 and chitobiose and insertion of the methyl group of chitobiose into a small hydrophobic pocket of Fbs1. Moreover, NMR spectroscopy has demonstrated that the amino acid residues adjoining the chitobiose-binding site interact with the outer branches of the carbohydrate moiety. Considering that the innermost chitobiose moieties in N-glycans are usually involved in intramolecular interactions with the polypeptide moieties, we propose that Fbs1 interacts with the chitobiose in unfolded N-glycoprotein, pointing the protein moiety toward E2 for ubiquitination.**

So far, numerous studies have emphasized the physiological importance of the ubiquitin- and proteasome-mediated proteolytic pathway[1]. The ubiquitination reaction is catalyzed by an elaborate cascade system, consisting of activating (E1), conjugating (E2) and ligating (E3) enzymes[1,2]. Of these enzymes, E3 enzymes are considered to exist as molecules with a large diversity and to have a principal role in the selection of target proteins for ubiquitination in a temporally and spatially regulated fashion[3].

One of the best-characterized E3 enzymes is the SCF complex (composed of Skp1, Cul1, Roc1 (also called Rbx1) and an F-box protein), which regulates degradation of a broad range of cellular proteins[4]. The F-box proteins consist of an F-box domain that binds to Skp1, and various C-terminal substrate recognition regions, which are subclassified into a family of proteins named Fbw and Fbl that contain WD40-repeat and leucine-rich repeat (LRR) domains, respectively[5]. In addition, the remaining groups have been provisionally classified as Fbx proteins, which show no homology to any other known proteins[5]. However, we recently discovered a third category of the F-box protein family named Fbs (F-box sugar recognition)/FBG[6], consisting of at least five structurally related proteins including Fbs1 (named originally as Fbx2)[7] and Fbs2/Fbx6b (ref. 8). Fbw and Fbl proteins usually recognize the phosphorylation status of the substrate, and the tertiary structures of some of these proteins, such as Fbw1/βTrCP, Fbw7/Cdc4 and Fbl1/Skp2, have been determined by X-ray crystallography, providing valuable information for determining the molecular recognition mechanisms of target proteins[9–12]. However, the molecular basis underlying the ability of Fbs proteins to recognize target glycoproteins remains to be clarified.

Eukaryotic cells have an abundant and diverse repertoire of N-linked oligosaccharide structures, but the role of N-glycosylation of the proteins remains largely unknown. N-glycans have recently been shown to have an important role in glycoprotein transport and sorting[13,14]. N-glycoproteins are also subjected to 'quality control,' in which aberrant proteins are distinguished from properly folded proteins and retained in the ER[15]. When the improperly folded or incompletely assembled proteins fail to restore their functional states, they are degraded by the ER-associated degradation (ERAD) system, which involves retrograde transfer of proteins from the ER to the cytosol and subsequent degradation mediated by ubiquitin and proteasomes[16]. Recently, we identified the $SCF^{Fbs1}$ as an E3 ubiquitin–ligase complex that ubiquitinates N-linked glycoproteins, serving to clear these glycoproteins in the cytosol of the cell[7]. Fbs1 recognizes N-linked high-mannose oligosaccharides, especially the internal diacetylchitobiose structure[8]. However, the molecular mechanism of the recognition of N-glycans by Fbs1 is unknown at present. To understand the molecular basis of the interaction between Fbs1 and N-glycans, we conducted crystal structural analyses of the SBD of Fbs1 and its complex with chitobiose.

## RESULTS

### Overall structure of the SBD in Fbs1

The structure of the SBD of Fbs1, as determined at a resolution of 2.0 Å (Table 1), is an ellipsoid composed of a ten-stranded antiparallel β-sandwich with two α-helices (Fig. 1a,b). This structure is

[1]Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. [2]Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113-8613, Japan. [3]Department of Structural Biology and Biomolecular Engineering, Graduate School of Pharmaceutical Sciences, Nagoya City University, 3-1 Tanabe-dori, Mizuho-ku, Nagoya 467-8603, Japan. [4]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Saitama 332-0012, Japan. [5]Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan. [6]Department of Molecular Cell Biology, Graduate School of Medicine, Osaka City University, Osaka 545-8585, Japan. [7]Genomic Sciences Center, RIKEN Yokohama Institute, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. Correspondence should be addressed to K.T. (tanakak@rinshoken.or.jp).

Published online 29 February 2004; doi:10.1038/nsmb732

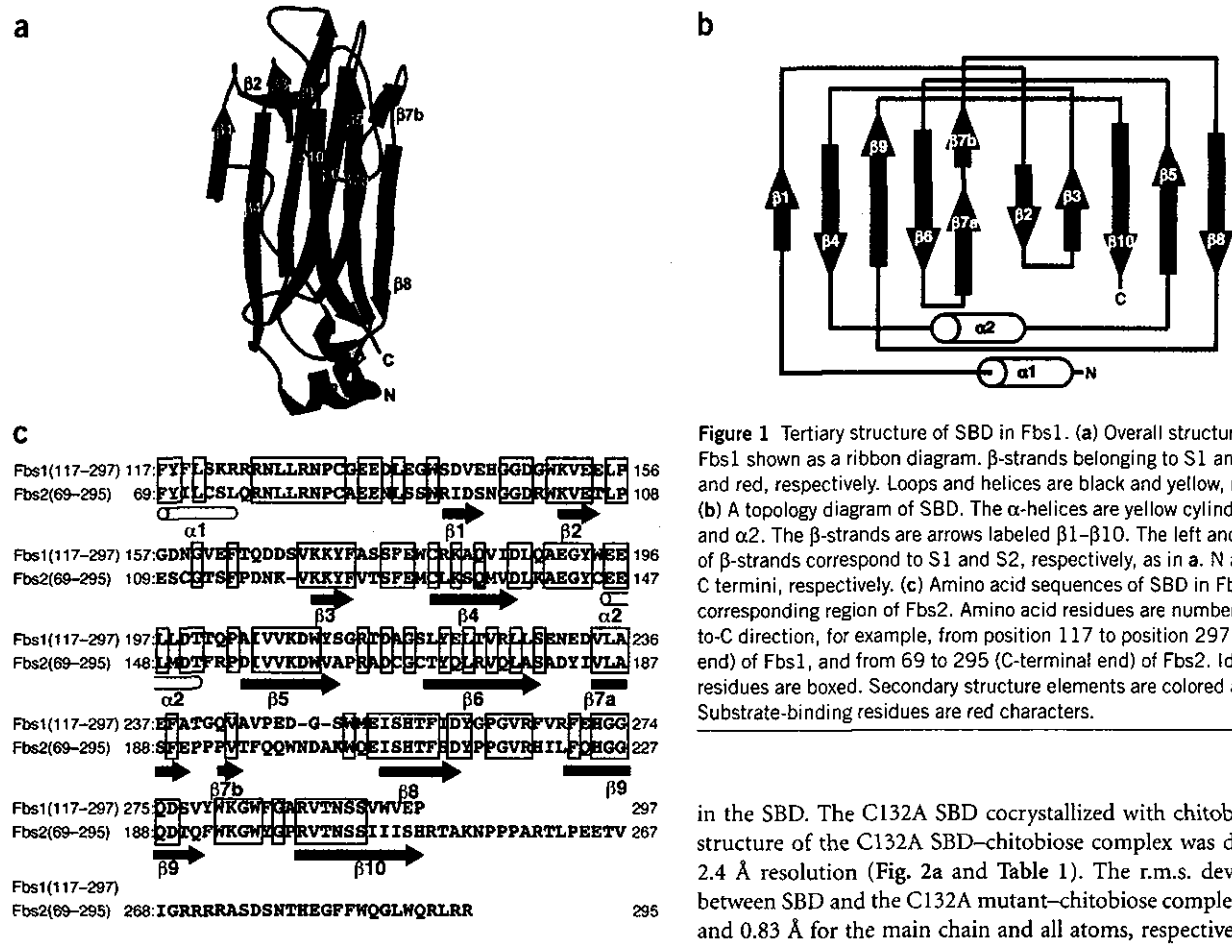



**Figure 1** Tertiary structure of SBD in Fbs1. (**a**) Overall structure of SBD of Fbs1 shown as a ribbon diagram. β-strands belonging to S1 and S2 are blue and red, respectively. Loops and helices are black and yellow, respectively. (**b**) A topology diagram of SBD. The α-helices are yellow cylinders labeled α1 and α2. The β-strands are arrows labeled β1–β10. The left and right forms of β-strands correspond to S1 and S2, respectively, as in **a**. N and C, N and C termini, respectively. (**c**) Amino acid sequences of SBD in Fbs1 and corresponding region of Fbs2. Amino acid residues are numbered in the N-to-C direction, for example, from position 117 to position 297 (C-terminal end) of Fbs1, and from 69 to 295 (C-terminal end) of Fbs2. Identical residues are boxed. Secondary structure elements are colored as **a**. Substrate-binding residues are red characters.

completely different from the folds of substrate-binding regions of F-box domains so far reported, including the WD40-repeat domain of CDC4 (ref. 11) and the LRR domain of Skp2 (ref. 12). Strands β1, β4, β6, β7 and β9 form one β-sheet, whereas the other β-sheet consists of strands β2, β3, β5, β8 and β10. These two sheets are named the S1 and S2 sheets, respectively (Fig. 1a–c). Strand β7, which is located at one edge of the S1 sheet, is composed of two segments (β7a and β7b) separated by a bent structure. The two α-helices (α1 and α2) lie at one end of the β-sandwich; the α1 helix is at the N terminus and α2 helix is in the loop between β4 and β5. Comparisons of the SBD structure described here with the Protein Data Bank using the Dali server[17] revealed that the SBD is structurally similar to certain lectins, such as the galectin-3 carbohydrate recognition domain[18] and second family 4 carbohydrate-binding modules of xylanase 10A (ref. 19), with r.m.s. deviation values of 2.8 and 2.8 Å, respectively. Indeed, both xylanase and galectin-3 domains are composed of 11-stranded antiparallel β-sandwiches, consisting of 5- and 6-stranded β-sheets, respectively, the overall structures of which resemble that of the SBD, although, in contrast to the SBD, they lack α-helices. Comparing their primary structures, the SBD shows considerable homology to the carbohydrate-binding domain of xylanase 10A, exhibiting amino acid identity of ~20%, whereas no obvious sequence homology was found between SBD and the carbohydrate recognition of galectin-3.

## The sugar-binding site

Next, we analyzed the structure of the SBD in complex with chitobiose. We could not obtain SBD crystals in the presence of chitobiose that diffracted to high resolution, so we introduced a C132A mutation in the SBD. The C132A SBD cocrystallized with chitobiose, and the structure of the C132A SBD–chitobiose complex was determined at 2.4 Å resolution (Fig. 2a and Table 1). The r.m.s. deviation values between SBD and the C132A mutant–chitobiose complex were 0.53 Å and 0.83 Å for the main chain and all atoms, respectively, indicating that the cysteine mutation hardly perturbs the conformation of the wild type SBD. It was also confirmed that this mutation did not alter the interaction and ubiquitination activities of glycoproteins (see Fig. 3b).

Chitobiose was clearly located in the difference electron density map (Fig. 2a), with *B*-factors as low as those of protein atoms around the chitobiose. The chitobiose in the SBD–chitobiose complex exhibited a *trans* conformation with respect to the *N*-acetyl groups, similar to the structures of a large number of N-glycan-binding proteins[20]. The bound chitobiose formed an intramolecular hydrogen bond between O6 of one GlcNAc(A) residue and O3 of the other GlcNAc(B) residue (Fig. 2b,c). The sugar-binding surface consists of two loops: L1 connects strands β3 and β4, and L2 is between strands β9 and β10. The GlcNAc(A) residue stacks on the aromatic ring of Trp280, as is often found in protein-carbohydrate interactions[21,22]. The GlcNAc-Trp280 stacking is stabilized by hydrogen bonds mediated by a water molecule between the O7 of the GlcNAc(A) and Nε1 of Trp280 (Fig. 2c), as well as a hydrogen bond between O6 of the GlcNAc(A) and the carbonyl oxygen atom of Lys281. The other GlcNAc(B) residue inserts the methyl group of its *N*-acetyl moiety into a small hydrophobic pocket surrounded by side chains of Phe177, Tyr279 and Lys281 (Fig. 2a,b) and forms a hydrogen bond between its hydroxyl and the main chain N atom of Lys281 (Fig. 2c). The orientation of the phenyl group of Tyr279 is stabilized by a hydrogen bond with the carboxyl group of Glu178. Upon N-glycoprotein uptake, the SBD swings the side chain of Lys281 (data not shown) and shields the methyl group from the molecular surface. The hydroxyl groups of the chitobiose almost exactly replace two binding site water molecules, which form a hydrogen bond to the backbone O and N of Lys281, respectively (Fig. 2b,c).

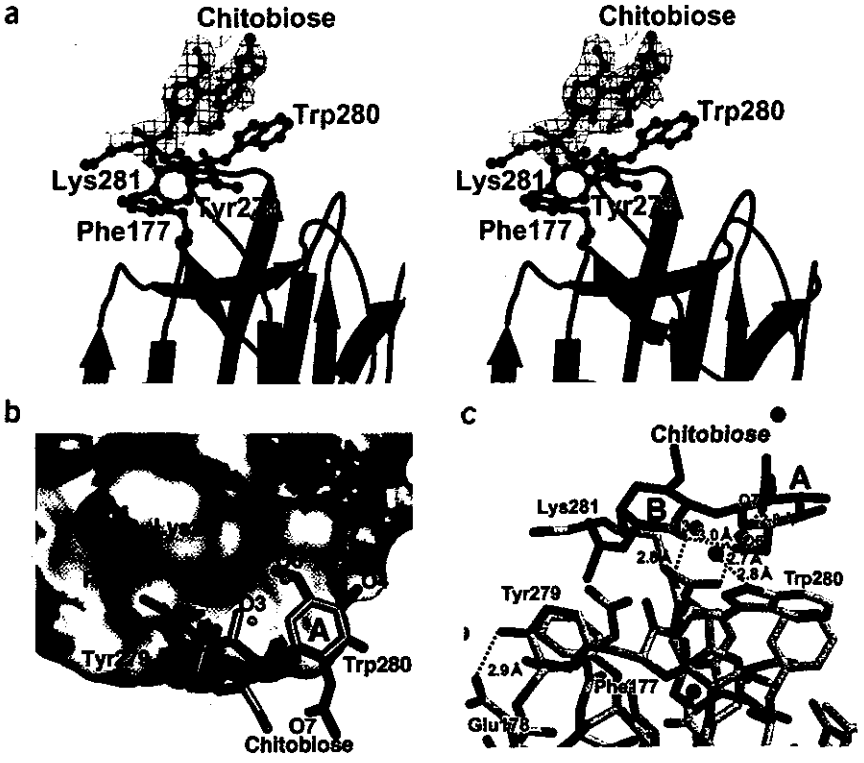



**Figure 2** Structure of SBD in complex with chitobiose. (**a**) Stereo view of the difference-density map ($F_o - F_c$ with phase from the Fbs1 model) of binding chitobiose, contoured at 2.1 σ, modeled into the electron density. β-strands belonging to S1 and S2 are blue and red, respectively. Loops are black. The bound chitobiose is orange, and the residues involved in the substrate binding (FYWK, see **Fig. 1c**) are green. (**b**) Molecular surface representation of the chitobiose-binding region. The bound chitobiose is shown in ball-and-stick representation. Two GlcNAc residues are represented by A and B. Cyan spheres are two water molecules of wild type SBD that are fixed on the molecular surface through hydrogen bonds with the backbone N and O of Lys281, respectively. These water molecules are replaced by O3 and O6 of the chitobiose upon formation of the SBD–chitobiose complex. (**c**) Stick representation of the amino acids involved in binding. Hydrogen bonds are dashed lines. Oxygen and nitrogen are red and blue, respectively. Symbols of two water molecules are as in **b**.

To verify that the crystal structure accurately represents the complex formed in solution, we introduced point mutations into the residues in the pocket, and examined the *in vitro* activities in binding the ribonuclease B (RNase B) carrying a high-mannose oligosaccharide (**Fig. 3a**). Indeed, F177A, Y279A and W280A mutations reduced binding to the RNase B, whereas the K281A mutation had no effect on the binding (**Fig. 3a**, left panel). We next tested the *in vivo* activities of these mutants in binding the precursor of integrin β1, one of the *in vivo* Fbs1 targets[7] that contains high-mannose oligosaccharides. Consistent with the *in vitro* results, F177A, Y279A and W280A, but not K281A, failed to bind integrin β1 (**Fig. 3a**, right panel). These results suggest that Phe177, Tyr279 and Trp280, located in the hydrophobic pocket at the edge of the β-sandwich, are important for interaction with chitobiose in the high-mannose oligosaccharides. In contrast, although comparison of the structure of the SBD alone and that of the SBD–chitobiose complex indicated that the side chain of Lys281 underwent a conformational change upon ligand binding, the *in vivo* and *in vitro* binding studies suggested that this conformational change was not essential for the recognition of oligosaccharides. Moreover, we examined the impact of these mutants on the ubiquitinating activities of the SCF onto GlcNAc-terminated fetuin (GTF) *in vitro* (**Fig. 3b**). SCF (Fbs1-W280A), which could not bind to N-glycans, failed to ubiquitinate GTF, whereas the ubiquitinating activities of the K281A mutant were retained. Taken together, these results indicate that the hydrophobic interactions between GlcNAc(A) residue and Trp280, and of GlcNAc(B) residue with the small hydrophobic pocket, are required for substrate recognition. In addition, the hydrogen bonds between the chitobiose and Fbs1 atoms (Nε1 of Trp280 and the carbonyl oxygen atom of Lys281) are involved in selective binding to chitobiose.

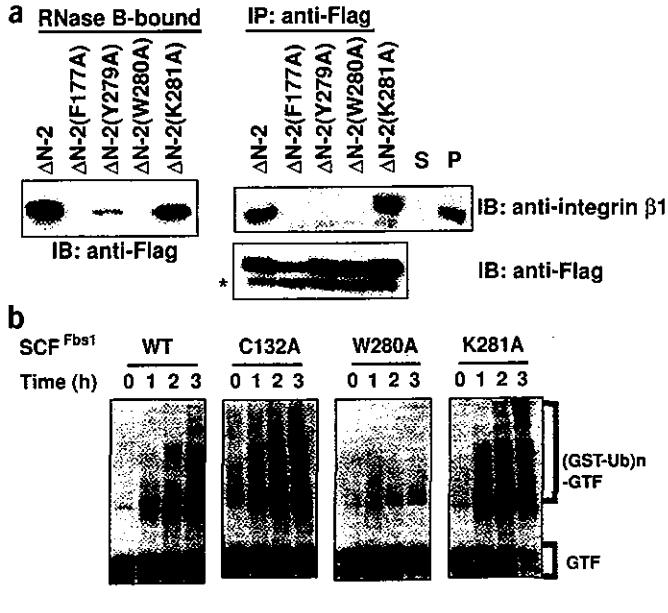



**Figure 3** Residues required for interaction of Fbs1 and glycoproteins with high-mannose oligosaccharides. (**a**) Neuro2a cells were transfected with Flag-tagged Fbs1 (ΔN-2) and its listed derivatives. In pull-down assay (RNase B-bound), each ΔN-2–expressing WCE was incubated with RNase B–immobilized beads; bound proteins were eluted by 0.1 M of chitobiose and then analyzed by immunoblotting with an antibody to Flag (anti-Flag). In immunoprecipitation with anti-Flag (IP: anti-Flag), the Fbs1-binding proteins in the immune complex were analyzed by immunoblotting using the anti-Flag or anti-integrin β1 antibody. Asterisks indicate the light chain of IgG. (**b**) *In vitro* ubiquitination of GTF by the $SCF^{Fbs1}$ E3–ligase system. The high-molecular-mass ubiquitinated GTF [(GST-Ub)n-GTF] was detected by immunoblotting with an antibody to fetuin.

### NMR analyses of the SBD-sugar interactions

We have previously reported that Fbs1 shows higher affinity to $Man_{3-9}GlcNAc_2$ glycans than to chitobiose, and the number of mannose residues did not influence the affinity[8]. We conducted NMR spectroscopic analysis to determine the contribution of the outer

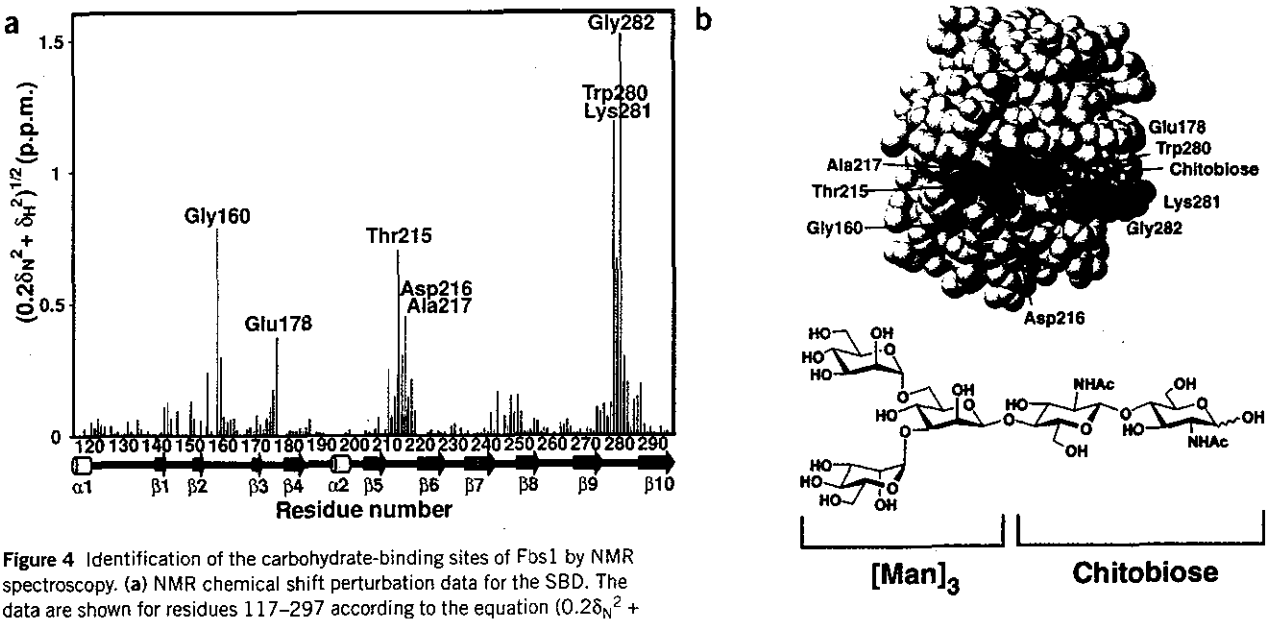



**Figure 4** Identification of the carbohydrate-binding sites of Fbs1 by NMR spectroscopy. (**a**) NMR chemical shift perturbation data for the SBD. The data are shown for residues 117–297 according to the equation $(0.2\delta_N^2 + \delta_H^2)^{1/2}$, where $\delta_N$ and $\delta_H$ represent the change in nitrogen and proton chemical shifts on addition of chitobiose (blue) and $Man_3GlcNAc_2$ (red). Secondary structure elements for SBD are shown below the plot. (**b**) Mapping the perturbed residues of the SBD(97–297) ($(0.2\delta_N^2 + \delta_H^2)^{1/2} > 0.3$) upon binding to chitobiose (blue) and $Man_3GlcNAc_2$ (blue and red). The chitobiose molecule bound to SBD is yellow. The $Man_3GlcNAc_2$ structure is represented in the right panel.

mannose branches to the interaction with Fbs1, because it is generally not feasible to crystallize or to interpret the electron density of complexes of lectins with larger oligosaccharides. To identify the oligosaccharide-binding site of the SBD in solution, we compared the pattern of the chemical shift perturbation between chitobiose and$Man_3GlcNAc_2$ (Man$\alpha$1→3[Man$\alpha$1→6]Man$\beta$1→4GlcNAc$\beta$1→4GlcNAc) using the isotopically labeled SBD (Fig. 4 and see **Supplementary Fig. 1** online). Chitobiose binding resulted in marked chemical shift perturbation for Glu178 in the L1 loop and for Trp280, Lys281 and Gly282 in the L2 loop; this is consistent with the X-ray structure of the C132A SBD–chitobiose complex. When $Man_3GlcNAc_2$ was used as a ligand, chemical shift perturbations were observed for Gly160, Thr215, Asp216 and Ala217 in addition to the residues perturbed by chitobiose, indicating that the outer branches of the carbohydrate moiety interact with the $\beta$5–$\beta$6 loop. In the crystal structure of the SBD–chitobiose complex, O4 of GlcNAc(A) is oriented toward these amino acid residues. Thus, the orientation of the sugar chain deduced from the NMR data is consistent with the expected orientation on inspection of the crystal structure. Although mannose residues of $Man_3GlcNAc_2$ seem to interact with the $\beta$5–$\beta$6 loop of Fbs1, almost no chemical shift perturbation of any consequence was observed upon addition of mannotriose (Man$\alpha$1→3[Man$\alpha$1→6]Man) alone (data not shown), suggesting that the affinity of $Man_3GlcNAc_2$ is dominated by the interaction with its chitobiose portion.

## DISCUSSION

In the present study, we determined the SBD structure of Fbs1 and its complex with chitobiose. In general, most lectins bind nonreducing terminal sugar groups in the concave surface, which consists of several strands of $\beta$-sheets. The substrate-binding site of galectin-3 that forms a complex with *N*-acetyl-lactosamine is formed by $\beta$-strands[18]. The amino acids in these $\beta$-strands interact with the bound substrate through direct and water-mediated hydrogen bonds or through van der Waals contacts. In contrast, Fbs1 recognizes the inner chitobiose of N-linked high-mannose oligosaccharides by a specific binding surface located at one tip of the $\beta$-sandwich. To our knowledge, this is the first report of the sugar-binding mode of lectins, that is, interaction with the innermost portion of the carbohydrate moieties of glycoproteins. This is in marked contrast to the lectin chaperones in the ER, namely calnexin and calreticulin, which recognize nonreducing terminal glucose molecules of the high-mannose oligosaccharides expressed on their target glycoproteins.

In crystal structures of glycoproteins, electron densities of carbohydrate moieties are generally ambiguous because the carbohydrate moieties attached to the crystallized proteins do not necessarily exhibit a uniform chemical structure and may possess freedom of internal motion. In many cases, however, the innermost GlcNAc residue does provide unambiguous electron density because it is involved in interactions with the polypeptide moieties. It seems that those intramolecular interactions hamper the binding of Fbs1 to the chitobiose portions of glycoproteins as a result of steric hindrance in their native states. We propose that the novel sugar-binding mode embodied by Fbs1 is suitable for recognition of unfolded glycoproteins targeted in the ERAD system. RNase B used in the *in vitro* binding assay reveals an oligosaccharide that does not contact the polypeptide chain except at the covalent attachment point[23]. This glycoprotein could interact with Fbs1 even in the native form in the *in vitro* binding experiment, probably as a result of the exceptional freedom of the chitobiose portion of its carbohydrate moiety.

On inspection of the X-ray crystallographic and mutagenesis data, we conclude that the hydrophobic interaction between the GlcNAc(A) residue and Trp280, the hydrophobic interaction between the GlcNAc(B) residue and the small hydrophobic pocket composed of Phe177 and Tyr279, and the hydrogen bonds between the chitobiose and Fbs1 atoms, are essential for selective binding to this disaccharide moiety. Recently, we reported that Fbs2 also binds high-mannose oligosaccharides in a chitobiose-dependent manner, but the strength

**Table 1 Data collection, phasing and refinement statistics**

| **Data collection** | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Native | PCMBS | $NaAuCl_4$ | $SmCl_3$ | $OsCl_3$ | $HgNO_3$ | Complex |
| Space group | $P3_221$ | | | | | | $P4_32_12$ |
| Resolution (Å) | 2.0 | 2.5 | 2.7 | 2.5 | 2.7 | 2.2 | 2.4 |
| Observations | 192,305 | 100,271 | 80,395 | 59,131 | 79,742 | 96,045 | 93,981 |
| Unique reflections | 18,483 | 9,577 | 7,689 | 9,633 | 7,695 | 13,725 | 12,611 |
| Completeness (%)[a] | 99.9 (99.9) | 99.8 (99.8) | 99.8 (99.8) | 99.8 (99.8) | 99.8 (99.8) | 98.6 (98.6) | 99.8 (99.8) |
| Redundancy[a] | 10.4 (10.1) | 10.5 (10.6) | 10.5 (10.6) | 6.1 (6.2) | 10.4 (10.5) | 7.0 (6.9) | 7.5 (7.2) |
| $R_{sym}$ (%)[a,b] | 9.5 (27.8) | 12.5 (30.6) | 9.9 (24.0) | 13.5 (22.7) | 14.3 (28.0) | 14.4 (26.4) | 8.5 (18.4) |
| $I / \sigma$[a] | 4.4 (2.5) | 4.4 (2.3) | 6.0 (2.9) | 3.9 (3.0) | 3.7 (2.3) | 3.2 (2.6) | 6.4 (3.6) |
| **MIRAS phasing** | | | | | | | |
| Resolution (Å) | 2.5 | 2.7 | 2.5 | 2.7 | 2.2 | | |
| Heavy atom sites | 3 | 2 | 2 | 1 | 2 | | |
| Phasing power | 1.37 | 0.89 | 0.76 | 1.71 | 3.31 | | |
| $R_{cullis}$[c] | 0.8 | 0.84 | 0.88 | 0.69 | 0.46 | | |
| **Refinement statistics** | | | | | | | |
| | SBD | SBD C132A–chitobiose | | | | | |
| Resolution (Å) | 2.0 | 2.4 | | | | | |
| Reflections | 17,470 | 11,950 | | | | | |
| Total atoms | 1,602 | 1,562 | | | | | |
| *R*-factor (%) | 16.2 | 20.0 | | | | | |
| $R_{free}$ (%) | 19.9 | 26.3 | | | | | |
| R.m.s. deviations | | | | | | | |
| Bond length (Å) | 0.020 | 0.038 | | | | | |
| Bond angle (°) | 1.8 | 2.9 | | | | | |

[a]Values in parentheses are for the highest-resolution shell. [b]$R_{sym} = \Sigma_h \Sigma_j |I_{hj} - \langle I_h \rangle| / \Sigma_h \Sigma_j |I_{hj}|$, where $h$ represents a unique reflection and $j$ represents symmetry-equivalent indices. $I$ is the observed intensity and $\langle I \rangle$ is the mean value of $I$. [c]$R_{cullis} = \Sigma ||F_{PH}| \pm |F_P|| - |F_H|| / \Sigma |F_{PH}| \pm |F_P|$.

of the glycoprotein-binding ability is weaker than that of Fbs1 (ref. 8). In Fbs2, the positions corresponding to Phe177, Tyr279 and Trp280 in Fbs1 are occupied by phenylalanine, phenylalanine, and tryptophan, respectively (**Fig. 1c**). The conservation of these residues suggests that the chitobiose-binding mode of Fbs2 is similar to that of Fbs1, and the reduced binding ability of Fbs2 may be attributed to the Tyr→Phe substitution in the chitobiose-binding pocket. Fbs1 interacts not only with chitobiose at the L1 and L2 loops but also with the outer mannose residues at the β5–β6 loop. Although binding of mannose residues to the β5–β6 loop seems to strengthen the binding affinity, mannose did not bind alone. This inner chitobiose-dependent interaction mode further restricts Fbs1 to interacting with native proteins carrying high-mannose oligosaccharides. The positions corresponding to Thr215 and Ala217, which are involved in the interaction with the mannose residues, are occupied by alanine and cysteine, respectively, in Fbs2 (**Fig. 1c**). This might be associated with the distinct affinity of Fbs1 and Fbs2 to N-glycans (ref. 8).

We attempted to model the full-length Fbs1 protein and its assembly in the SCF-E2 complex by positioning the N terminus of the SBD adjacent to the F-box domain of the reported $SCF^{Skp2}$ complex[9]. This model places the chitobiose >50 Å from the expected position of the E2 protein that donates ubiquitin to the glycoprotein, consistent with modeling of other SCF complexes, such as $SCF^{Skp2}$ (ref. 9) and $SCF^{Cdc4}$ (ref. 11).

Fbs1 is a functionally unique molecule that recognizes the innermost position of N-glycans as a signal for denatured glycoproteins. Our results confirmed structurally that N-glycans act as a ubiquitination signal, thus providing new insights into the biological roles of sugar chains coupled to proteins within cells.

## METHODS

**Crystallization and structure determination.** The SBD of murine Fbs1(117–297), with a molecular mass of 20.6 kDa, and its mutant protein C132A SBD were cloned into pET15b and expressed in *Escherichia coli*. Both proteins were purified by nickel affinity and gel filtration chromatography and concentrated to 20 mg $ml^{-1}$. All mutations were constructed by QuikChange mutagenesis kit (Stratagene), using synthetic oligonucleotides, and the sequences were verified in their entirety.

Crystals of SBD were grown at 25 °C by the hanging-drop vapor diffusion method. SBD crystals were grown from 1.7 M ammonium sulfate, 0.01 M nickel chloride, 0.1% (v/v) PEG400 and 0.1 M Tris-HCl, pH 8.5. The C132A SBD was cocrystallized with chitobiose (Seikagaku). C132A SBD crystals were prepared using 1.4 M sodium chloride, 1.7 M ammonium sulfate, 0.1 M PIPES, pH 7.0, and 30 mM chitobiose. SBD and its cysteine mutant crystals belonged to $P3_221$ and $P4_32_12$ space groups, with cell dimensions of $a = b = 62.4$ Å, $c = 117.2$ Å, and $a = b = 63.8$ Å, $c = 147.8$ Å, respectively. Heavy-atom soaks were done in crystallization buffer with saturated *p*-chloromercuribenzene sulfonate (PCMBS) (for 3 h), 10 mM $NaAuCl_4$ (15 h), 10 mM $SmCl_3$ (15 h), 10 mM $OsCl_3$ (15 h) and 10 mM $HgNO_3$ (15 h). Intensity data sets were collected on a Rigaku R-AXIS IV nickel-filtered double-mirror focused CuKα-radiation detector. A Rigaku RU-200 rotating-anode X-ray generator was operated at 40 kV and 100 mA. Data were processed with MOSFLM[24] and SCALA[25,26]. The structure of the SBD was determined by the multiple isomorphous replacement anomalous scattering (MIRAS) method. Phases were calculated with MLPHARE[26] to 2.2 Å. The initial electron density map was then refined by solvent flattening[27] and histogram mapping[28] using DM[26]. The initial model was constructed with Arp/WArp[29] and O[30]. The model was refined at a resolution of 2.0 Å with REFMAC[31]. Residues 43–55 of the SBD have high temperature factors and are presumably partially disordered.

Intensity data sets of the mutant protein were collected at 100 K. For cryoprotection, 20% (v/v) glycerol was added to the crystallizing solution. The

structure of C132A was determined by the molecular replacement technique using AMoRe[32] with the refined model of SBD. Structure refinement of the mutant SBD was guided by referring to the structure of the wild type. Refinement statistics of both structures are summarized in Table 1.

**Binding assay.** Neuro2a cells were transfected with Flag-tagged murine Fbs1 ΔN-2(95–297) and its derivatives that lack the region from the N terminus to the F-box domain[7] by lipofection (Lipofectamine Plus; Gibco BRL). At 48 h after transfection, the whole-cell extracts (WCEs), solubilized with TBS containing 0.5% (v/v) NP-40, were used for immunoprecipitation using mouse monoclonal anti-Flag (M2; Sigma), or pull-down assay using RNase B–immobilized resin, as described[7]. The (co-)immunoprecipitated proteins were detected by immunoblotting using mouse monoclonal antibodies anti-Flag (M2) and anti-integrin β1 (Transduction Laboratories).

***In vitro* ubiquitination assay.** Recombinant His-tagged human Ubc4 was produced in *E. coli*. Recombinant His-tagged mouse E1 (Uba1), His-tagged rat Fbs1 (ΔN-2) and each $SCF^{Fbs1}$ (Flag-tagged human Skp1, human Cul1-HA/His-tagged rat Fbs1 derivatives, T7-tagged human Roc1) were produced by baculovirus-infected HiFive insect cells. Each $SCF^{Fbs1}$ complex was obtained by simultaneously infecting four baculoviruses. These proteins were affinity-purified by a HiTrap HP column (Amersham Pharmacia Biotech), as described[7]. Each 1 μg of GTF was incubated in 50 μl of the reaction mixture containing ATP-regenerating system, 0.5 μg E1, 1 μg Ubc4 (E2), 2 μg $SCF^{Fbs1}$, 6.5 μg recombinant GST-ubiquitin and NEDD8 system[33] at 30 °C. After the reaction was terminated by adding 25 μl of 3× SDS-PAGE sample buffer, the proteins in 8 μl of the boiled supernatants were separated with 5–20% (w/v) SDS-PAGE, and the high-molecular-mass ubiquitinated proteins were detected by immunoblotting with anti-fetuin (Chemicon International).

**NMR spectroscopy.** The DNA fragment encoding residues 97–297 of Fbs1was inserted into the pGEX-6P-1 plasmid vector (Amersham Biosciences) with a N-terminal GST moiety. The protein was expressed in *E. coli* BL21(DE3) CodonPlus strain (Stratagene) in M9 minimal medium with appropriate [$^{15}$N]$NH_4Cl$ and [$^{13}C_6$]glucose. GST-fusion protein was purified from cell lysates on a glutathione-Sepharose column. The fusion protein was cleaved by incubation with PreScission protease (Amersham Biosciences), and GST was removed by loading a second glutathione-Sepharose column. The protein was further purified using a Superose12 gel filtration column.

NMR samples contained 0.1–1.0 mM SBD(97–297) in 10 mM $NaH_2PO_4/Na_2HPO_4$, pH 6.5, 50 mM KCl and 10 mM DTT. For chemical shift perturbation experiments, a ten-fold molar excess of chitobiose, ten-fold molar excess of Manα1'3(Manα1'6)Man (Calbiochem) or one molar equivalent of $Man_3GlcNAc_2$ (Sigma) was added to the protein solution. NMR spectra were acquired at 30 °C on Bruker DMX500 and DRX800 spectrometers. The $^1$H, $^{15}$N and $^{13}$C resonances of the backbone were assigned using a standard set of double- and triple-resonance experiments[34].

**Coordinates.** The atomic coordinates of the SBD and C132A SBD–chitobiose complex have been deposited in the Protein Data Bank (accession codes 1UMH and 1UMI, respectively).

*Note: Supplementary information is available on the Nature Structural & Molecular Biology website.*

ACKNOWLEDGMENTS

We thank Y. Wada for her help in the preparation of isotopically labeled recombinant proteins and E. Adachi for expert technical assistance. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science and Culture of Japan (K.T. and K.K.), Japan Society for the Promotion of Science and Mizutani Foundation for Glycoscience (K.K.), and in part by the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sport, Science and Technology of Japan (T.T.).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 6 October 2003; accepted 20 January 2004
Published online at http://www.nature.com/natstructmolbiol/

1. Hershko, A., Ciechanover, A. & Varshavsky, A. Basic Medical Research Award. The ubiquitin system. *Nat. Med.* **6**, 1073–1081 (2000).
2. Pickart, C.M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
3. Weissman, A.M. Themes and variations on ubiquitylation. *Nat. Rev. Mol. Cell. Biol.* **2**, 169–178 (2001).
4. Deshaies, R.J. SCF and Cullin/Ring H2–based ubiquitin ligases. *Annu. Rev. Cell. Dev. Biol.* **15**, 435–467 (1999).
5. Winston, J.T., Koepp, D.M., Zhu, C., Elledge, S.J. & Harper, J.W. A family of mammalian F-box proteins. *Curr. Biol.* **9**, 1180–1182 (1999).
6. Ilyin, G.P. *et al.* A new subfamily of structurally related human F-box proteins. *Gene* **296**, 11–20 (2002).
7. Yoshida, Y. *et al.* E3 ubiquitin ligase that recognizes sugar chains. *Nature* **418**, 438–442 (2002).
8. Yoshida, Y. *et al.* Fbs2 is a new member of the E3 ubiquitin ligase family that recognizes sugar chains. *J. Biol. Chem.* **278**, 43877–43884 (2003).
9. Zheng, N. *et al.* Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**, 703–709 (2002).
10. Wu, G. *et al.* Structure of a β-TrCP1-Skp1-β-catenin complex: destruction motif binding and lysine specificity of the SCF(β-TrCP1) ubiquitin ligase. *Mol. Cell* **11**, 1445–1456 (2003).
11. Orlicky, S., Tang, X., Willems, A., Tyers, M. & Sicheri, F. Structural basis for phosphodependent substrate selection and orientation by the $SCF^{Cdc4}$ ubiquitin ligase. *Cell* **112**, 243–256 (2003).
12. Schulman, B.A. *et al.* Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* **408**, 381–386 (2000).
13. Plemper, R.K. & Wolf, D.H. Retrograde protein translocation: ERADication of secretory proteins in health and disease. *Trends Biochem. Sci.* **24**, 266–270 (1999).
14. Fiedler, K. & Simons, K. The role of *N*-glycans in the secretory pathway. *Cell* **81**, 309–312 (1995).
15. Helenius, A. & Aebi, M. Intracellular functions of *N*-linked glycans. *Science* **291**, 2364–2369 (2001).
16. Eilgaard, L., Molinari, M. & Helenius, A. Setting the standards: quality control in the secretory pathway. *Science* **286**, 1882–1888 (1999).
17. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
18. Seetharaman, J. *et al.* X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-Å resolution. *J. Biol. Chem.* **273**, 13047–13052 (1998).
19. Simpson, P.J. *et al.* The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. *Biochemistry* **41**, 5712–5719 (2002).
20. Petrescu, A.J., Petrescu, S.M., Dwek, R.A. & Wormald, M.R. A statistical analysis of *N*- and *O*-glycan linkage conformations from crystallographic data. *Glycobiology* **9**, 343–352 (1999)
21. Vyas, N.K. Atomic features of protein–carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1**, 732–740 (1991).
22. Quiocho, F.A. Carbohydrate-binding proteins: tertiary structures and protein–sugar interactions. *Annu. Rev. Biochem.* **55**, 287–315 (1986).
23. Williams, R.L., Greene, S.M. & McPherson, A. The crystal structure of ribonuclease B at 2.5-Å resolution. *J. Biol. Chem.* **262**, 16020–16031 (1987).
24. Leslie, A.G.W. Molecular data processing. *Crystallographic Computing* **5**, 50–61 (1991).
25. Kabsch, W. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.* **21**, 916–924 (1988).
26. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
27. Wang, B.C. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112 (1985).
28. Zhang, K.Y.J. & Main P. The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. *Acta Crystallogr. A* **46**, 377–381 (1990).
29. Perrakis, A., Morris, R. & Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463 (1999).
30. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47** (Part 2), 110–119 (1991).
31. Murshudov, B.W. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
32. Navaza, J. An automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
33. Kawakami, T. *et al.* NEDD8 recruits E2-ubiquitin to SCF E3 ligase. *EMBO J.* **20**, 4003–4012 (2001).
34. Clore, G.M. & Gronenborn, A.M. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* **239**, 349–363 (1994).

# Structural basis for distinct roles of Lys63- and Lys48-linked polyubiquitin chains

Takeshi Tenno[1,3], Kenichiro Fujiwara[3], Hidehito Tochio[3], Kazuhiro Iwai[4], E. Hayato Morita[5], Hidenori Hayashi[2], Shigeo Murata[6], Hidekazu Hiroaki[3], Mamoru Sato[3], Keiji Tanaka[6] and Masahiro Shirakawa[3,*]

[1]*Graduate School of Science and Engineering, Ehime University, Matsuyama, Ehime, Japan;*
[2]*Graduate School of Integrated Science, Yokohama City University, Yokohama, Kanagawa, Japan;*
[3]*Graduate School of Medicine, Osaka City University, Abeno-ku, Osaka, Japan;*
[4]*Integrated Center for Sciences, Ehime University, Matsuyama, Ehime, Japan;*
[5]*Cell-Free Science and Technology Research Center, Ehime University, Matsuyama, Ehime, Japan;*
[6]*Department of Molecular Oncology, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo, Japan*

**Ubiquitination, a modification in which single or multiple ubiquitin molecules are attached to a protein, serves as a signalling function that controls a wide variety of cellular processes. To date, two major forms of polyubiquitin chain have been functionally characterized, in which the isopeptide bond linkages involve Lys48 or Lys63. Lys48-linked polyubiquitin tagging is mostly used to target proteins for degradation by the proteasome, whereas Lys63-linked polyubiquitination has been linked to numerous cellular events that do not rely on degradative signalling via the proteasome. Apparently linkage-specific conformations of polyubiquitin chains are important for these cellular functions, but the structural bases distinguishing Lys48- and Lys63-linked chains remain elusive. Here, we report NMR and small-angle X-ray scattering (SAXS) studies on the intersubunit interfaces and conformations of Lys63- and Lys48-linked di- and tetraubiquitin chains. Our results indicate that, in marked contrast to Lys48-linked chains, Lys63-linked chains are elongated molecules with no stable non-covalent intersubunit interfaces and thus adopt a radically different conformation from that of Lys48-linked chains.**

## Introduction

Chains or single molecules of ubiquitin, a 76-amino acid protein that is highly conserved throughout eukaryotes, can be covalently attached to various cellular proteins (Pickart 2000; Weissman 2001; Marx 2002). Chains of ubiquitin molecules are generated through isopeptide bonds formed between the C terminus of one ubiquitin molecule and a specific lysine residue in the next. Ubiquitin has seven Lys residues, which are all used for polymerization (Peng *et al.* 2003), but polyubiquitin chains formed via Lys at 48 or 63 have been extensively studied so far. These two types of chain share a common specific enzymatic cascade consisting of E1 (ubiquitin-activating enzyme), E2 (ubiquitin-conjugating enzyme) and E3 (ubiquitin-protein ligase), of which E2 seems to define the topology of the chains. Whereas several E2 enzymes have been identified for Lys48-linked chains, the only known E2 enzymes responsible for conjugating Lys63-linked chains are the yeast Mms2/Ubc13 heterodimer and its human counterpart Uev1A/Ubc13 (Deng *et al.* 2000; Hofmann & Pickart 2001). Mms2 and Uev1A, a ubiquitin E2 variant (UEV) protein, resemble E2 enzymes for Lys48-linked chains, but lack the active cysteine residue that forms a thioester bond with the C terminus of ubiquitin. Structural studies have shown that UEV proteins act as an adaptor that directs Lys63 of the acceptor ubiquitin to an active cysteine on Ubc13 (Moraes *et al.* 2001; VanDemark *et al.* 2001; McKenna *et al.* 2003).

A polyubiquitin chain formed via Lys48 functions mainly as a marker for proteolytic attack by the 26S proteasome (a eukaryotic ATP-dependent proteolytic

*Communicated by*: Kozo Kaibuchi
**Correspondence*: Email: shirakawa@tsurumi.yokohama-cu.ac.jp

DOI: 10.1111/j.1365-2443.2004.00780.x
© Blackwell Publishing Limited

complex) (Coux *et al.* 1996; Baumeister *et al.* 1998). In contrast, Lys63-linked polyubiquitin tagging is involved in a variety of cellular events that do not rely on degradative signalling via the proteasome (Weissman 2001; Marx 2002). Of these, the best-characterized is post-replicative DNA repair, which restarts replication machinery that is stalled at a damaged template DNA by repairing the damaged site (Friedberg *et al.* 1995). In *Saccharomyces cerevisiae*, this repair process is carried out through the *RAD6* pathway, which involves several proteins, including the E2 complex for Lys63-linked chains, Mms2/Ubc13. Thus, the modification of nuclear factors by Lys63-linked polyubiquitin has been implicated in post-replicative DNA repair (Ulrich & Jentsch 2000). Recently, proliferating cell nuclear antigen (PCNA), a DNA polymerase sliding clamp that is required for replication and several DNA repair pathways, has been shown to be a substrate of Mms2/Ubc13-mediated Lys63-linked polyubiquitination (Hoege *et al.* 2002). The same lysine residue was found to be the target of modification by monoubiquitin or a small ubiquitin-related modifier (SUMO), and it has been suggested that these different modifications switch the function of PCNA (Hoege *et al.* 2002; Stelter & Ulrich 2003). Although the Lys63-linked chain was shown to signal substrate degradation *in vitro*, proteasomal degradation seems not to be required for error-free post-replicative repair (Hofmann & Pickar 2001). Lys63-linked polyubiquitination has also been shown to be involved in activation of IκB kinase, ribosome function, endocytosis, inheritance of mitochondrial DNA, transcriptional regulation of HIV-1 and the mitotic checkpoint (Galan & Haguenauer-Tsapis 1997; Fisk & Yaffe 1999; Springael *et al.* 1999; Deng *et al.* 2000; Spence *et al.* 2000; Bothos *et al.* 2003; Bres *et al.* 2003).

Structural information on polyubiquitin chains is needed to understand the signals that polyubiquitin tags provide, because their linkage-specific conformations present different surfaces to interacting proteins that mediate the signals. Structural studies on Lys48-linked di- and tetraubiquitins have been reported. The crystal structure of the diubiquitin shows that a hydrophobic patch, composed of Leu8, Ile44 and Val70, in each ubiquitin molecule comprises the interface between its two subunits (Cook *et al.* 1992). Consistent with this, NMR studies indicate that the diubiquitin adopts a closed conformation at neutral pH in which the hydrophobic patch is sequestered at the interface, but undergoes a conformational change to an open form at lower pH (Varadan *et al.* 2002). For Lys48-linked tetraubiquitin, two crystal forms have been reported in which the conformations are significantly different, suggesting that the polyubiquitin chains take a conformation depending on crystal packing and possess a high degree of conformational flexibility (Cook *et al.* 1994; Phillips *et al.* 2001). Such a dynamic property has also been suggested for Lys48-linked diubiquitin (Varadan *et al.* 2002). NMR chemical shift perturbation experiments suggest that the distal two subunits of Lys48-linked tetraubiquitin may adopt a closed conformation in solution, which is similar to that of the Lys48-linked diubiquitin (Varadan *et al.* 2002).

In contrast to the Lys48-linked form, structural studies of Lys63-linked chains have not been reported so far. Here, we have carried out NMR and small-angle X-ray scattering (SAXS) studies of the intersubunit interfaces of the Lys63-linked di- and tetraubiquitin chains and have compared them with those of Lys48-linked chains. Our results indicate that Lys63-linked di- and tetraubiquitins are much more elongated than Lys48-linked di- and tetraubiquitins, and thus the intersubunit interfaces of Lys63-linked polyubiquitin chains are distinct from those of Lys48-linked chains, suggesting that these forms of polyubiquitin chain adopt radically different conformations.

## Results

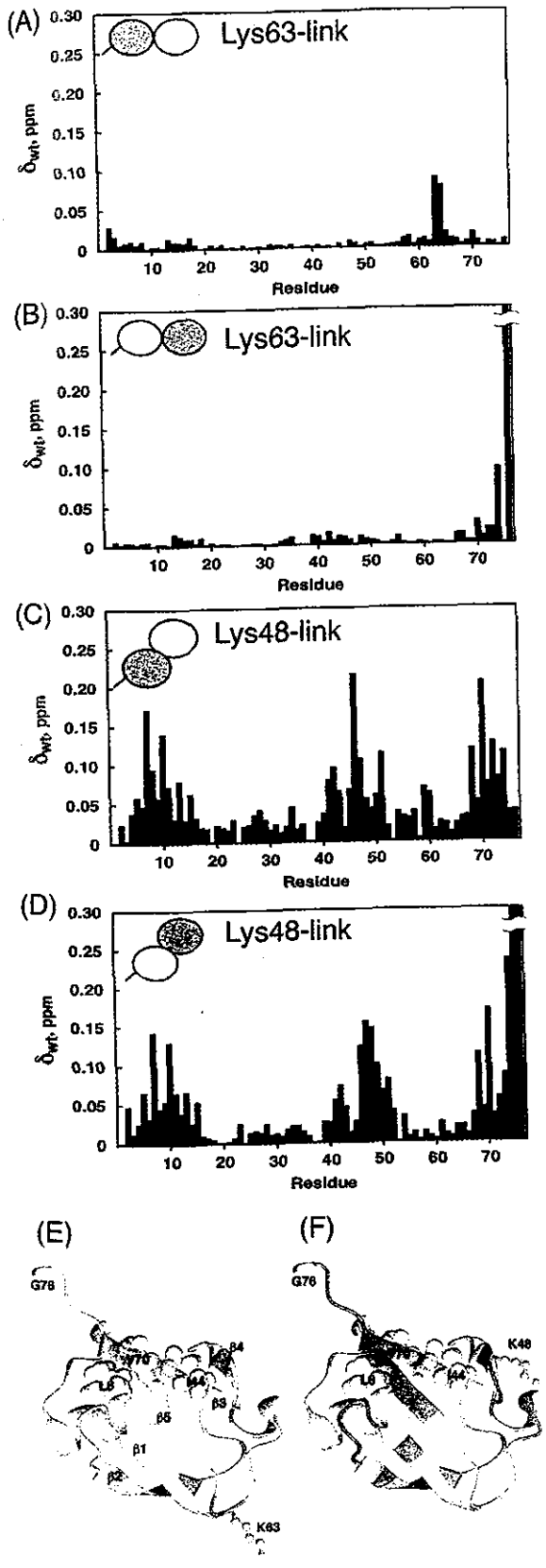
### Chemical shift mapping of the intersubunit interfaces of Lys63- and Lys48-linked ubiquitin chains

To identify the intersubunit interface of Lys63- and Lys48-linked diubiquitin, we compared the $^{15}N$ and $^{1}H$ chemical shifts of the main-chain amide groups of these diubiquitins with those of the unlinked ubiquitin derivatives (see Experimental procedures). Such chemical shift perturbation experiments have been frequently used to identify protein–protein interfaces. For the analyses, we labelled only one subunit of diubiquitin with $^{15}N$ nuclei (see Experimental procedures) so that a $[^{1}H,^{15}N]$-correlation spectrum of the subunit-specific labelled sample would give cross-peaks attributable to the amide nuclei of the labelled subunit. Hereafter, the subunits of the polyubiquitin chains are numbered according to the convention of Thrower *et al.* (2000), that is, the subunit that contains a free main-chain carboxyl terminus is denoted subunit 1; the one whose C-terminal Gly76 is linked to a lysine of subunit 1 through the isopeptide bond is denoted subunit 2; and so on.

The two subunits of Lys48-linked diubiquitin displayed similar patterns of chemical shift perturbation at pH 6.8, except for the C-terminal three residues of subunit 2, which contain the isopeptide bonding site, Gly76 (Fig. 1C,D). Residues in strands β1–β5 (β1: residues 1–6, β2: 12–17, β3: 41–45, β4: 48–49, β5: 66–71), and



© Blackwell Publishing Limited




loops between strands β1 and β2, and between strands β3 and β4 of both ubiquitin subunits showed significant differences in their chemical shifts when compared with the unlinked ubiquitin derivatives. These residues are all confined to a surface of ubiquitin that centres on a hydrophobic patch composed of the side chains of Leu8, Ile44 and Val70 on the β-sheet (Fig. 1F). This observation suggests that a similar protein surface on each subunit interacts with each other at the interface.

We then analysed the residues at the intersubunit interface of Lys63-linked diubiquitin by chemical shift perturbation using subunit 1- and subunit 2-selective [$^{15}$N]-labelled diubiquitin samples. Chemical shift perturbation of either subunit of Lys63-linked diubiquitin was much less than that of either subunit of Lys48-linked diubiquitin (Fig. 1A,B). A significant chemical shift perturbation was observed only for the following residues, which are located at or adjacent to the Gly76–Lys63 isopeptide bond: Lys63 and Glu64 of subunit 1, and Arg74 and Gly76 of subunit 2 (Fig. 1E). Changes in these residues can be attributed simply to the chemical modification associated with formation of the isopeptide bond, and unlikely to non-covalent intersubunit contacts. Apart from these residues, the weighted chemical shift differences, $\delta_{wt}$, associated with the isopeptide bond formation were less than 0.03 p.p.m. This is in clear contrast to Lys48-linked diubiquitin, whose 44 residues of subunit 1 and 32 of subunit 2 exhibited weighted chemical shift differences larger than 0.03 p.p.m. upon isopeptide bond formation. These results indicate that Lys63-linked diubiquitin does not have a well-defined non-covalent intersubunit interface distal from the linkage site. Furthermore, the chemical shift values were essentially invariant from pH 4.5 to pH 7.5, suggesting the

**Figure 1** Chemical shift perturbations of Lys63- and Lys48-linked diubiquitin chains compared with those of the unlinked ubiquitin derivatives. (A, B) Lys63-linked diubiquitin; (C, D) Lys48-linked diubiquitin. A and C correspond to $^{15}$N-labelled (indicated in grey) subunit 1, B and D to $^{15}$N-labelled subunit 2 (see text for subunit naming convention). The weighted chemical shift differences, $\delta_{wt}$, are shown (see Experimental procedures). (E, F) Ribbon representation of ubiquitin (PDB code: 1D3Z), with colour coding indicating the chemical shift perturbations observed for Lys63-linked diubiquitin (E) and Lys48-linked diubiquitin (F). Residues that exhibited weighted chemical shift differences ($\delta_{wt}$) of more than 0.05 p.p.m. are coloured blue if they showed this difference only in subunit 1, green if only in subunit 2 and red if they showed this difference in both subunits. The side chains of the linkage-site residues in the Lys48- and Lys63-linked polyubiquitin chains and the hydrophobic patch-forming residues are also indicated.

© Blackwell Publishing Limited