

# Intrinsic disorder drives N-terminal ubiquitination by Ube2w

Vinayak Vittal<sup>1</sup>, Lei Shi<sup>1,2</sup>, Dawn M Wenzel<sup>1</sup>, K Matthew Scaglione<sup>3-5</sup>, Emily D Duncan<sup>1</sup>, Venkatesha Basrur<sup>6</sup>, Kojo S J Elenitoba-Johnson<sup>6</sup>, David Baker<sup>1,2</sup>, Henry L Paulson<sup>3</sup>, Peter S Brzovic<sup>1</sup> & Rachel E Klevit<sup>1\*</sup>

**Ubiquitination of the  $\alpha$ N-terminus of protein substrates has been reported sporadically since the early 1980s. However, the identity of an enzyme responsible for this unique ubiquitin (Ub) modification has only recently been elucidated. We show the Ub-conjugating enzyme (E2) Ube2w uses a unique mechanism to facilitate the specific ubiquitination of the  $\alpha$ -amino group of its substrates that involves recognition of backbone atoms of intrinsically disordered N termini. We present the NMR-based solution ensemble of full-length Ube2w that reveals a structural architecture unlike that of any other E2 in which its C terminus is partly disordered and flexible to accommodate variable substrate N termini. Flexibility of the substrate is critical for recognition by Ube2w, and either point mutations in or the removal of the flexible C terminus of Ube2w inhibits substrate binding and modification. Mechanistic insights reported here provide guiding principles for future efforts to define the N-terminal ubiquitome in cells.**

The attachment of Ub to cellular proteins is a highly regulated process that requires three enzyme activities. First, an E1 Ub-activating enzyme forms a thioester bond between its active site cysteine and the Ub C terminus in an ATP-dependent reaction. Second, Ub undergoes a transthiolation reaction with the active site cysteine of an E2 Ub-conjugating enzyme, forming an E2~Ub conjugate. Third, E2~Ub interacts with an E3 Ub ligase to modify protein targets via a RING-type, HECT-type or RING-between-RING-type mechanism. A distinguishing feature of RING-type mechanisms is that the E3 activates the E2~Ub conjugate to transfer Ub directly from the E2 active site to the substrate<sup>1</sup>. Thus, in RING-type mechanisms, the E2 has a direct role in interacting with substrate and dictating the final ubiquitinated product. The diversity of products depends on the enzymes involved and the biological context and may include the addition of a single Ub onto a substrate lysine or the synthesis of poly-Ub chains built from any of Ub's seven lysine residues. To generate such diversity, there are ~40 human E2s that have presumably evolved disparate functions. Some E2s are specific for a single chain type, such as the Ubc13-Mms2 complex (K63-linked chains)<sup>2</sup> or Ube2k (K48-linked chains)<sup>3</sup>, whereas others such as UbcH5c are promiscuous and can build Ub chains of multiple linkages<sup>4</sup>. Some E2s such as Ube2e1 and Ube2t add only a single Ub to their target substrate<sup>5,6</sup>. There is some evidence that certain E2s may transfer Ub to noncanonical amino acids such as serine, threonine and cysteine<sup>7,8</sup>. The E2 Ube2w was recently reported to attach mono-Ub to the  $\alpha$ N terminus of substrates rather than to the  $\epsilon$ NH<sub>2</sub> side chain group of lysine residues<sup>9,10</sup>.

Here we show that Ube2w specifically monoubiquitinates the  $\alpha$ N terminus of diverse substrates by recognizing backbone atoms of disordered N termini. The solution ensemble of Ube2w reveals a unique 'ubiquitin conjugating' (UBC) domain architecture (Supplementary Results, Supplementary Fig. 1). Though the first 118 residues adopt a canonical E2 fold, the Ube2w C-terminal region is partially unstructured and can occupy multiple positions

near the active site. Removal of the final 20 C-terminal residues or a single point mutation within this region abrogates Ube2w Ub transfer activity and affects recognition and binding of multiple substrates. Furthermore, N-terminal substrate recognition and subsequent Ub transfer catalyzed by Ube2w are intimately dependent on the noncanonical arrangement of Ube2w C-terminal residues relative to its active site.

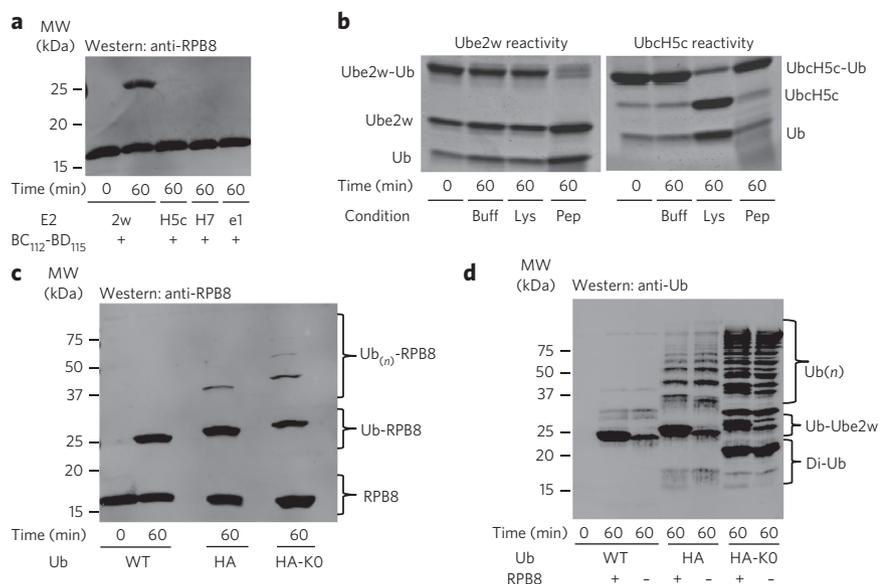
## RESULTS

### Ube2w adds mono-Ub to intrinsically disordered N termini

RNA polymerase subunit 8 (RPB8) is a highly conserved subunit of RNA polymerases I, II and III that is ubiquitinated in cells by the RING E3 ligase BRCA1-BARD1 following UV-induced DNA damage<sup>11</sup>. To identify the E2-BRCA1-BARD1 pair (or pairs) that can ubiquitinate RPB8, *in vitro* ubiquitination assays were performed using the minimal RING heterodimer of BRCA1-BARD1 (BC<sub>112</sub>-BD<sub>115</sub>) and E2s that had previously been shown to interact with BRCA1-BARD1: Ube2w, UbcH5c, UbcH7 and Ube2e1 (ref. 12). Although RPB8 contains eight lysines, only Ube2w modifies RPB8 with Ub in the presence of BC<sub>112</sub>-BD<sub>115</sub> (Fig. 1a and Supplementary Fig. 2a). Ube2w also exhibits E3-independent modification, though with substantially lower activity (Supplementary Fig. 2a, Supplementary Fig. 3). MS analysis of the monoubiquitinated RPB8 product confirmed that the Ub is attached to RPB8's  $\alpha$ N terminus (Supplementary Fig. 4). It should be noted that an initial mono-Ub attached by Ube2w can serve as a primer for poly-Ub chain synthesis by another E2 such as Ubc13-MMS2 or Ube2k<sup>10,12,13</sup>.

A direct measure of the intrinsic aminolysis activity of an E2~Ub conjugate is its reactivity toward the  $\epsilon$ NH<sub>2</sub> group of free lysine. Many E2s, such as UbcH5c, that transfer Ub to lysine side chains of protein substrates readily transfer Ub to free lysine<sup>14</sup>. The Ube2w~Ub conjugate remained intact in the presence of free lysine but reacted completely with a peptide containing a free N-terminal amino group and no lysine (NH<sub>2</sub>-Ala-Gly-Gly-Ser-Tyr-COO<sup>-</sup>) (Fig. 1b

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, Washington, USA. <sup>2</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA. <sup>3</sup>Department of Neurology, University of Michigan Medical School, Ann Arbor, Michigan, USA. <sup>4</sup>Department of Biochemistry, Medical College of Wisconsin, Milwaukee, Wisconsin, USA. <sup>5</sup>Neuroscience Research Center, Medical College of Wisconsin, Milwaukee, Wisconsin, USA. <sup>6</sup>Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA. \*e-mail: klevit@u.washington.edu



**Figure 1 | Ube2w has distinct E2 activity.** (a) Ube2w (2w) transfers a single Ub to RPB8 *in vitro*, whereas other BRCA1-interacting Ub conjugating enzymes UbcH5c (H5c), UbcH7 (H7) and Ube2e1 (e1) do not (**Supplementary Fig. 2a**). MW, molecular weight. (b) Left, a nucleophile reactivity assay reveals that Ube2w has intrinsic activity with  $\alpha\text{NH}_2$  groups of a peptide with a free  $\text{NH}_2$  group at its N terminus ( $\text{NH}_2\text{-A-G-G-S-Y-COO}^-$ ; 50 mM) but not the  $\epsilon\text{NH}_2$  groups of lysine. Buff, buffer; Pep, peptide. Right, identical reactions with UbcH5c-Ub conjugates confirm the previously reported lysine reactivity of UbcH5c and reveal it to be unreactive toward the peptide (**Supplementary Fig. 2b**). (c) Products generated on RPB8 depend on the Ub species in the reaction. Lanes 1 and 2: a single Ub is attached to RPB8 in a reaction with WT Ub. Lane 3: attachment of an additional Ub is detected in a reaction with HA-Ub, which contains a 13-residue tag at the N-terminal end of Ub. Lane 4: the reaction carried out with lysine-free HA-Ub (HA-Ub(KO)) confirms that Ube2w builds linear Ub chains (i.e., attaches the C terminus of one Ub to the N terminus of another) on RPB8 with HA-Ub (**Supplementary Fig. 2d**). (d) Reactions shown in **c** were blotted for Ub, revealing that Ube2w builds linear poly-Ub chains only when Ub harbors an N-terminal HA tag (**Supplementary Fig. 2e**).

and **Supplementary Figs. 2b,c** and 5). In contrast, UbcH5c~Ub reacted completely with free lysine but did not react with the peptide substrate (**Fig. 1b**). Thus, Ube2w's intrinsic aminolysis reactivity is limited to  $\alpha\text{NH}_2$  groups in the context of a polypeptide as it did not transfer Ub to the  $\alpha\text{NH}_2$  group of a free amino acid, whereas UbcH5c's intrinsic aminolysis reactivity is limited to  $\epsilon\text{NH}_2$  groups of lysine residues.

In addition to RPB8, full-length carboxy terminus of HSP70-interacting protein (CHIP), the minimal U-box domain of CHIP, small Ub-like modifier 2 (SUMO-2), tau, ataxin-3, FANCL, FANCD2 and Ube2w have been reported to be ubiquitinated by Ube2w<sup>9,10,15,16</sup>. In all cases, a single mono-Ub is attached, implying that the N terminus of the attached Ub cannot also serve as a Ube2w substrate. Serendipitously, we discovered that Ub harboring a 13-residue N-terminal human influenza hemagglutinin (HA) tag is a robust Ube2w substrate *in vitro*. Unlike with wild-type (WT) Ub, Ube2w adds multiple HA-Ubs onto RPB8, and this activity is retained with a lysine-free version, HA-Ub(KO), consistent with each attachment occurring through the  $\alpha\text{NH}_2$  group (**Fig. 1c** and **Supplementary Fig. 2d**). Notably, the major species formed under the reaction conditions used here (i.e., HA-Ub in excess over RPB8) are free poly-HA-Ub chains, as observed in an anti-Ub immunoblot of the same reaction (**Fig. 1d** and **Supplementary Fig. 2e**). <sup>1</sup>H-<sup>15</sup>N NOE (heteronuclear NOE (hetNOE)) values of HA-Ub, which are sensitive to high-frequency motions on the pico- to nanosecond timescales, confirm that the additional residues in the HA tag are highly mobile, as shown by their small and/or negative hetNOE values (**Fig. 2a** and **Supplementary Fig. 6**). In contrast,

hetNOE values for Ub residues 1–72 show that the N terminus of WT Ub is well ordered<sup>17</sup>, as expected from Ub crystal structures (**Fig. 2a,b**). The ability of HA-Ub but not WT Ub to serve as substrate illustrates that addition of a disordered segment to its natively structured N terminus is sufficient to convert Ub into a Ube2w substrate. We note that all of the currently identified targets of Ube2w-dependent ubiquitination (described above) have or are predicted to have a disordered segment at their N termini<sup>18–20</sup> (**Supplementary Fig. 7**).

### Ube2w recognizes the backbone atoms of its substrates

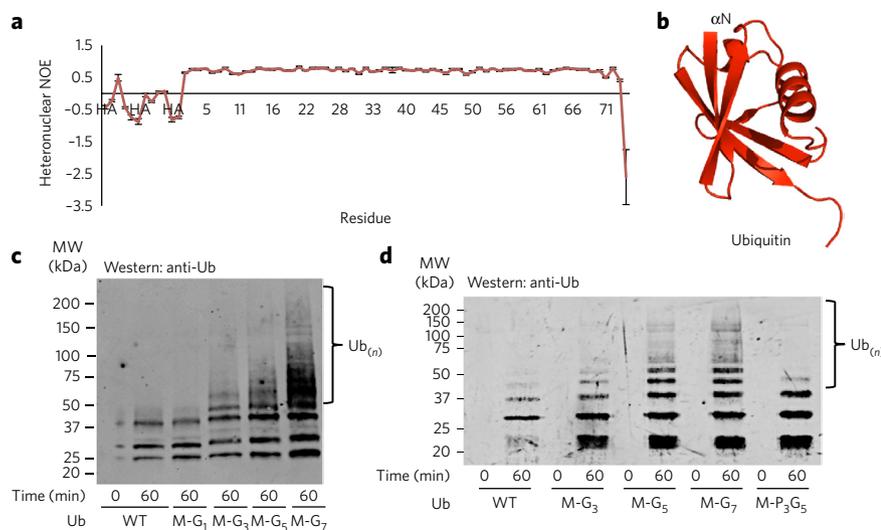
Diverse N-terminal sequences in protein substrates suggest two mechanistic possibilities: (i) Ube2w preferentially recognizes backbone atoms over amino acid side chains and (ii) regular N-terminal secondary structure elements ( $\alpha$ -helices and  $\beta$ -sheets) would inhibit necessary contacts with Ube2w. To further test these hypotheses, we chose to add glycine residues (lacking side chains) to the N terminus of Ub. Addition of two residues (Met-Gly-Ub) did not result in Ube2w-dependent modification, whereas addition of four N-terminal residues (Met-Gly<sub>3</sub>-Ub) results in modest activity, and additional glycine residues (Met-Gly<sub>5</sub>- or Met-Gly<sub>7</sub>-Ub) result in a further increase in Ube2w activity (**Fig. 2c** and **Supplementary Fig. 2f**). The results show that, *in vitro*, a disordered polypeptide chain composed of a methionine and three glycine residues is sufficient for Ube2w to recognize a protein as a potential 'substrate' and that Ube2w activity increases as additional disordered N-terminal glycines are present. Replacement of glycine residues with

prolines disrupts the ability of Ube2w to use N-terminally tagged Ub as a substrate. Ub harboring three N-terminal prolines (Met-Pro<sub>3</sub>-Gly<sub>5</sub>-Ub) is incapable of forming large poly-Ub chains and forms similar products to WT Ub, indicating that amide groups at positions 2 through 4 are necessary for Ube2w-dependent N-terminal ubiquitination (**Fig. 2d** and **Supplementary Fig. 2g**). Furthermore, a methionine at position 1 is not necessary for N-terminal ubiquitination by Ube2w (**Supplementary Table 1**). Altogether, the results are consistent with a model whereby Ube2w recognizes substrates through backbone carbonyl and amide groups rather than side chain atoms.

Ub transfer via aminolysis most likely has differing requirements for an  $\alpha$ -amino group as opposed to the  $\epsilon$ -amino group of a lysine side chain<sup>21</sup>. The  $\text{pK}_a$  of an N-terminal amino group is in the range of  $7.7 \pm 0.5$ , whereas lysine side chain  $\text{pK}_a$  values range around  $10.5 \pm 1.1$  (ref. 22). Thus, at physiological pH, a larger proportion of  $\alpha\text{N}$  termini will be deprotonated, bypassing a need to deprotonate the incoming nucleophile. The  $\text{pK}_a$  of an N-terminal amino group depends on the identity of the side chain at position 1. However, our results show that Ube2w is capable of ubiquitinating N termini over a wide  $\text{pK}_a$  range ( $\text{pK}_a = \sim 7.3$  to  $\sim 9.1$ ), indicating that the nucleophile's  $\text{pK}_a$  is not the determining factor for Ube2w's special reactivity (**Supplementary Figs. 2h** and 8)

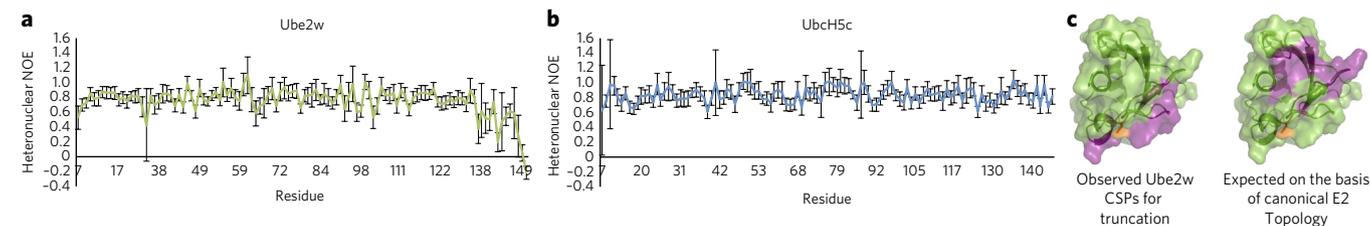
### Ube2w has a noncanonical UBC domain

For insights into Ube2w's unique Ub transfer specificity, we characterized the dynamic and structural properties of the E2 using NMR. Though a crystal structure exists (Protein Data Bank (PDB) code



**Figure 2 | Ube2w transfers Ub to flexible/disordered N termini.** (a) Negative  $^1\text{H}$ - $^{15}\text{N}$  hetNOE values for residues derived from the HA tag are indicative of highly flexible amino acids. Error bars represent s.e.m. Resonances from the HA tag are not assigned and are not plotted sequentially; they are labeled simply as 'HA'. (b) Consistent with the  $^1\text{H}$ - $^{15}\text{N}$  hetNOE data, the crystal structure of Ub (PDB code 1UBQ) is ordered at its  $\alpha\text{N}$  terminus and immediately forms a  $\beta$ -strand with residue Met1. (c) Ub to which two N-terminal amino acids have been added at the N terminus (Met-Gly-Ub (M-G<sub>2</sub>)) is not incorporated into chains by Ube2w and displays similar activity to WT Ub. Four N-terminal residues (Met-Gly<sub>3</sub>-Ub (M-G<sub>3</sub>)) are sufficient to induce Ube2w activity toward Ub. Addition of six (Met-Gly<sub>5</sub>-Ub) or eight (Met-Gly<sub>7</sub>-Ub) residues increases Ube2w N-terminal ubiquitination activity (note: bands below 37 kDa are consistent with autoubiquitinated E2 and E3) (**Supplementary Fig. 2f**). MW, molecular weight. (d) N-terminal backbone amide groups are necessary for Ube2w-dependent ubiquitination. Lanes 1–8: Ube2w shows increased activity with the addition of disordered N-terminal amino acids. Lanes 9 and 10: proline (P) at positions 2–4 (M-P<sub>3</sub>G<sub>5</sub>) inhibits Ube2w chain-building activity to levels similar to WT Ub (**Supplementary Fig. 2g**).

2A7L), it contains a truncated Ube2w that lacks residues Ser117–Cys151 (ref. 23). A similar truncated version of Ube2w can form an activated thioester with Ub (**Supplementary Figs. 2i and 9**) but does not transfer Ub onto a substrate (data not shown) and thus lacks the structural features needed to understand Ube2w substrate selectivity.  $^1\text{H}$ - $^{15}\text{N}$  hetNOE values indicate that the final three residues of full-length Ube2w are highly flexible (negative hNOE values) and that residues 135–151 undergo higher-frequency motions (small positive hetNOE values) than the rest of the protein (**Fig. 3a**). An identical experiment conducted on UbcH5c reveals that its C terminus undergoes motions consistent with the core of the protein (**Fig. 3b**).

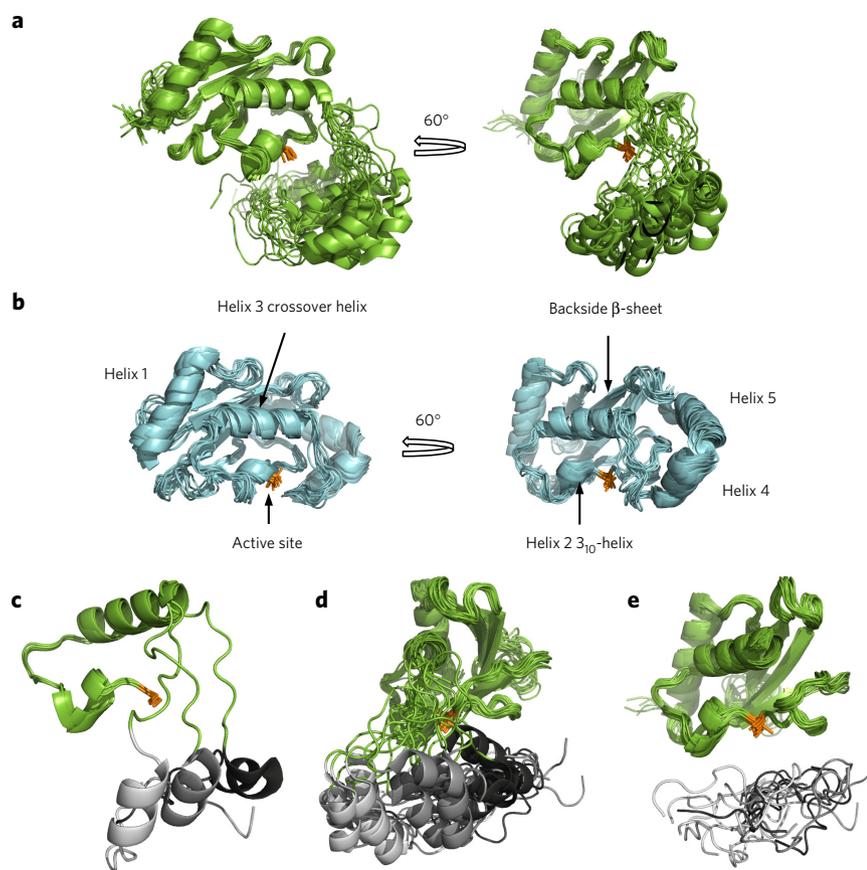


**Figure 3 | The Ube2w C terminus is flexible and occupies a noncanonical position.** (a) Residues 7–132 of Ube2w have generally uniform and positive  $^1\text{H}$ - $^{15}\text{N}$  hetNOE values. Beginning at residue 137, values decrease and ultimately become negative at the extreme C terminus, consistent with a region that undergoes motions at higher frequencies than the core of the protein. Error bars represent s.e.m. (b) For comparison, UbcH5c has positive hetNOE values throughout its entire protein sequence, even at the far C terminus. Error bars represent s.e.m. (c) Left, experimental CSP data based on a comparison of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TROSY spectra of Ube2w-KK and Ube2w-131 $\Delta$ -KK reveals that removal of the C terminus perturbs residues near the active site, in the  $3_{10}$  helix and on the backside  $\beta$ -sheet (purple). Right, if C-terminal helices were to reside in their canonical positions in Ube2w, a surface consisting of loops 3 and 5 would be perturbed by removal of residues 132–151. Residues depicted to be perturbed are colored in purple, demonstrating that the C-terminal region of Ube2w is different from other E2s.

To understand the unique structural properties of Ube2w, further NMR experiments were conducted. NMR data were collected on a monomeric Ube2w mutant (Ube2w<sup>V30K D67K</sup>, termed 'KK') that reduces self-association at high concentrations but retains Ubc2w activity<sup>24</sup>. We compared the  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC-transverse relaxation-optimized spectroscopy (TROSY) spectra of full-length Ube2w-KK with a fragment missing the C-terminal residues following Tyr131 (Ube2w-131 $\Delta$ -KK). Ube2w resonances for residues S86, N87 and T96 (near the active site); in the  $3_{10}$  helix; and on the 'backside'  $\beta$ -sheet experience large chemical shift perturbations (CSPs) as a consequence of removal of the C terminus (**Fig. 3c** and **Supplementary Fig. 10a,b**). These observations were unexpected in that comparisons with canonical E2 structures, such as UbcH5c, predict that a different set of CSPs would be observed, particularly with respect to the E2 active site and loops 3 and 5 (**Fig. 3c** and **Supplementary Fig. 1**)<sup>25</sup>.

On the basis of these results, we sought to characterize the structure of full-length Ube2w. A conventional *de novo* NOE-based structure determination was precluded by a paucity of NOE cross-peaks (**Supplementary Fig. 11**). Therefore, we pursued an NMR-driven solution structure using alternative parameters. NMR chemical shifts (HN, N, C, C $\alpha$ , and C $\beta$ ) for 137 residues, 109 residual dipolar coupling (RDC) values for NH pairs and the CSPs between full-length and truncated Ube2w were initially input into the Rosetta-based algorithm, Chemical-Shift Rosetta (CS-Rosetta), a program that uses chemical shift values and other experimentally driven NMR restraints to generate solution ensembles<sup>26,27</sup> (**Supplementary Tables 2 and 3, Supplementary Data Set 1 and Supplementary Figs. 10c and 12**). This initial computational stage produced 16,000 structures that were further filtered using two spin label positions (C91 and C135) and small-angle X-ray scattering (SAXS) to generate the final ensemble.

An ensemble of the 20 lowest energy structures is shown in **Figure 4a**. Ube2w has a well-defined core that closely resembles canonical UBC domains, such as that of UbcH5c (**Fig. 4b**). The average pairwise r.m.s. deviation for residues R7–S118 for the 20



**Figure 4 | NMR ensemble of Ube2w reveals a new E2 architecture.** (a) Solution ensemble of Ube2w derived from NMR restraints (backbone chemical shifts, CSPs, RDCs, paramagnetic spin-label data and SAXS) calculated with CS-Rosetta. The 20 lowest-energy members of the ensemble are shown and reveal a well-defined core with high structural similarity to canonical E2s. The C-terminal region is partially disordered and occupies multiple positions near the Ube2w active site C91 (orange). (b) Similar views of a representative canonical UBC domain structure (UbcH5c; PDB code 2FUH). (c) Helix 4 (the penultimate helix) has distinct positions in Ube2w (three representatives of the 20-member ensemble are shown for clarity). A flexible loop emanating from helix 3 leads away from the protein core. Helix 4 is clustered in three distinct positions in the ensemble (light gray, cluster 1; gray, cluster 2; dark gray, cluster 3). (d) Side view of the full Ube2w ensemble looking down the helix 3 axis reveals the three clusters. (e) In all 20 members of the Ube2w ensemble, residues N136–W145 occupy positions beneath the active site, C91 (orange). Residues 119–135 are not shown for clarity. No clustering is evident for this region, and the C $\beta$  atom of every residue is, on average, 14.5–17.5 Å away from the active site.

members of the ensemble is 1.35 Å, indicating that the available experimental observations used are sufficient to define the structure. Furthermore, the 20-member ensemble reveals favorable Ramachandran statistics (Supplementary Table 2). The average pairwise r.m.s. deviation for all of the backbone atoms across the Ube2w ensemble is 4.1 Å, consistent with  $^1\text{H}$ – $^{15}\text{N}$  hetNOE values that reveal a highly flexible C terminus. Features conserved among canonical E2s in the Ube2w solution structure include helix 1, the four-stranded backside  $\beta$ -sheet, the structural architecture of the active site and helix 3 (‘crossover’ helix; Fig. 4a,b).

The first seven residues of Ube2w are not observed in our NMR spectra, indicating that they undergo conformational exchange, most likely a helix-to-coil transition. However, owing to the component of CS-Rosetta that uses SPARTA-based selection of protein fragments from the PDB, our ensemble contains an ordered N terminus based on homology modeling. The most distinctive feature of the Ube2w ensemble is its C-terminal region (residues 127–151), which adopts multiple orientations near the active site. In the

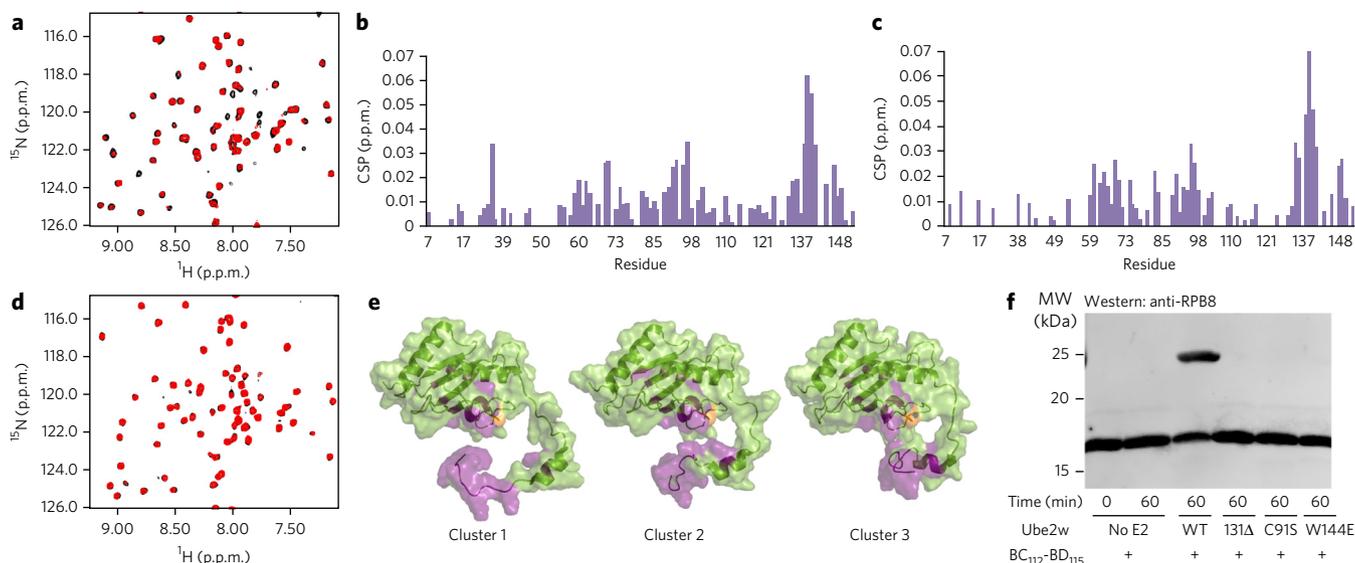
ensemble, C-terminal positions are determined by the CSP restraints, SAXS and the spin label effects at residues C91 and C135.

In Ube2w, a long disordered loop following the crossover helix leads away from the  $\beta$ -sheet and is followed by a single helix formed by residues 127–135. This helix does not seem to adopt a unique position but is located within 7–19 Å of the active site, C91, in all members of the ensemble (Fig. 4c). By contrast, in canonical UBC domains, a loop leads from the end of the crossover helix toward the protein core (near the backside  $\beta$ -sheet) and is followed by two C-terminal helices, helix 4 (15.5 Å from active site) and helix 5 (Fig. 4b). Three distinct C-terminal clusters are evident in the ensemble, in which helix 4 occupies positions facing (i) closer to the  $3_{10}$ -10 helix, (ii) adjacent to the active site and (iii) closer to the backside  $\beta$ -sheet (Fig. 4c,d). In place of the final C-terminal helix present in canonical UBC folds, residues N136–W145 form a disordered region that occupies positions directly beneath the active site in all states of the ensemble, as revealed by the spin label attached to active site C91 (Fig. 4e and Supplementary Fig. 13a). The final six amino acids are completely disordered and are not constrained to a particular region, consistent with  $^1\text{H}$ – $^{15}\text{N}$  hetNOE results.

The correlation between the experimental and back-calculated SAXS curves shows some disagreement (Supplementary Fig. 12f). Possible sources for the discrepancies are (i) presence of some aggregated material, (ii) presence of low concentrations of dimeric Ube2w even with the dimer-disrupting mutation V30K D67K and (iii) the aforementioned dynamics of the N terminus. Nevertheless, the ensemble was generated from a combination of different experimental restraints, and we believe it is an accurate representation of the predominant species in solution. In this respect, a recent crystal structure of an E2 from the fungi *Agrocybe aegerita* that shares 50% identity with Ube2w has both a disordered C terminus and a noncanonical position for helix 4 that falls within our Ube2w ensemble (PDB code 3WE5)<sup>18</sup>. The average r.m.s. deviation for all backbone atoms to the closest member of the Ube2w ensemble and this crystal structure is 2.55 Å over the entire protein sequence.

### Ube2w C terminus mediates substrate interactions

Multiple observations suggest that Ube2w has a predominant role in mediating interactions with substrates. First, Ube2w shows robust Ub transfer activity (Fig. 1a) in the presence of a minimal RING construct that offers no substrate binding functionality. Second, Ube2w adds Ub to the  $\alpha\text{N}$  terminus of substrates in the absence of an E3, albeit at a slower rate (Supplementary Fig. 3)<sup>9</sup>. RING-type E3s enhance the activity of other E2s by promoting closed E2~Ub conformations that promote Ub transfer<sup>21,28,29</sup>. Our results also demonstrate that Ube2w~Ub is allosterically activated by RING-type E3s to form closed conformations (Supplementary Figs. 2j and 13).  $^1\text{H}$ – $^{15}\text{N}$  HSQC-TROSY experiments capable of detecting low-affinity binding interactions reveal peak broadening and CSPs for a subset of [ $^{15}\text{N}$ ]Ube2w-KK resonances upon addition



**Figure 5 | The Ube2w C terminus is required to interact with substrates.** (a)  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TROSY spectrum of Ube2w-KK in the absence (black spectrum) and presence of 1 molar equivalent of RPB8 (red spectrum). Evidence for binding is seen as peak broadening (loss of intensity) and CSPs of specific peaks in the Ube2w NMR spectrum. (b) A histogram showing CSPs upon addition of 1 molar equivalent of RPB8 into Ube2w-KK. (c) Titration of 1 molar equivalent of tau into Ube2w-KK reveals very similar CSPs to addition of RPB8. (d)  $^1\text{H}$ - $^{15}\text{N}$  HSQC-TROSY spectrum of Ube2w-131Δ-KK in the absence (black spectrum) and presence of 1 molar equivalent of RPB8 (red spectrum). Truncated Ube2w shows no interaction with RPB8. (e) Residues whose resonances have considerable intensity losses and/or CSPs ( $>1$  s.d.) are mapped in purple onto members of the Ube2w ensemble (one representative from each cluster). (f) In an *in vitro* ubiquitination assay, Ube2w-131Δ does not transfer Ub to RPB8 after 1 h. Mutation of a single residue in the C-terminal region, W144E, also abrogates detectable activity. The loss of activity associated with the C-terminal region is equivalent to a mutant in which the active site is dead (C91S) (**Supplementary Fig. 2k**). MW, molecular weight.

of the substrates RPB8 and tau (**Fig. 5a–c** and **Supplementary Fig. 14a**). On the basis of the magnitude of the observed perturbations, these interactions are highly transient. Residues near the active site and within the C-terminal region are the most extensively perturbed upon addition of RPB8: Y131, K137, N138, K140, K141, K143, W144 and W145 in the C terminus and S93, I94, L95, T96 and E97 in the  $3_{10}$  helix that immediately follows active site C91 ( $>1$  s.d. below or above mean for intensity loss or shifting, respectively). Notably, all of the residues in the unstructured region positioned directly beneath the active site in the Ube2w ensemble (K137–W145) are appreciably perturbed (**Fig. 4e**). C-terminally truncated Ube2w-131Δ-KK shows neither CSPs in NMR binding experiments with RPB8 nor Ub transfer activity, indicating that the C terminus is essential for substrate recognition (**Fig. 5d–f** and **Supplementary Fig. 2k**). Finally, addition of tau, a Ube2w substrate with a different N-terminal sequence, perturbs a nearly identical set of residues in [ $^{15}\text{N}$ ]Ube2w-KK, as does RPB8, suggesting that side chain identity has, at most, a minimal role in Ube2w substrate recognition (**Fig. 5c** and **Supplementary Fig. 14b,c**).

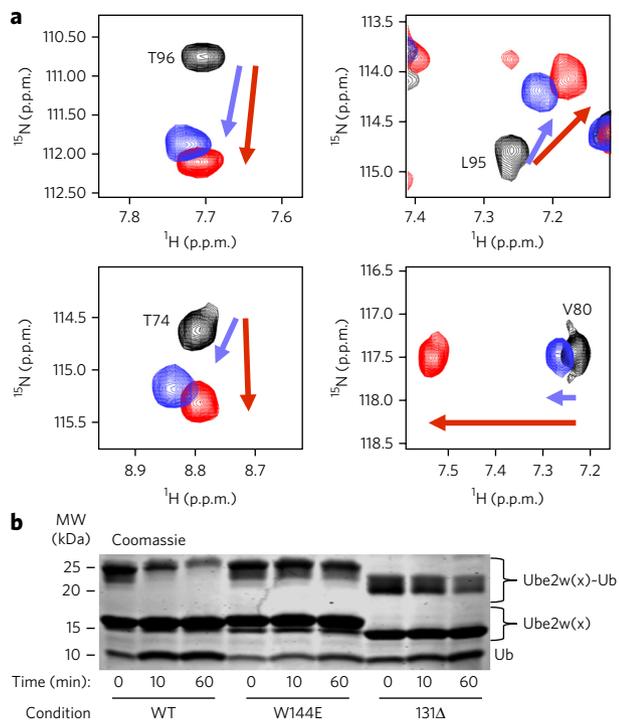
Although multiple sequence alignment of the Ube2w C terminus with other E2s shows considerable divergence, the C-terminal region is strongly conserved among Ube2w orthologs. Even the slime mold *Dictyostelium discoideum* has strong conservation of C-terminal amino acids (R133, N136, K137, P139, W144, H147, D148 and D149; **Supplementary Fig. 15a**). Our NMR experiments show that residues N136 through W145 represent an important substrate-binding region. W144, which occupies positions as close as 7 Å (average distance is 16.7 Å) to the active site in the NMR ensemble of Ube2w, is one of the most highly conserved amino acids among Ube2w orthologs, is positioned in a highly disordered region of the Ube2w ensemble (**Fig. 4e** and **Supplementary Fig. 15a,b**) and shows large perturbations upon substrate binding (**Fig. 5b,c,e**). Substituting a glutamate (the corresponding residue in the UbcH5c sequence) for Ube2w W144 (W144E) completely abrogates Ube2w activity toward four different substrates: RPB8, ataxin-3, tau and CHIP (**Fig. 5f** and

**Supplementary Fig. 16a–d**). The mutation has a twofold effect. First, substrate binding is inhibited, as evidenced by decreased peak broadening upon addition of RPB8 to the mutant protein in an NMR binding experiment (**Supplementary Fig. 16e**). Second, the W144E mutation generates NMR CSPs in Ube2w active site resonances that resemble those seen in the truncated Ube2w-131Δ mutant (**Fig. 6a**). To probe the role of the Ube2w C terminus in N-terminal ubiquitination function, we used our intrinsic reactivity assay. Ube2w~Ub readily transfers its Ub to the minimal peptide substrate in a 1-h reaction, but Ube2w-131Δ and Ube2w-W144E show almost no transfer to the peptide (**Fig. 6b** and **Supplementary Fig. 2l**). The slow loss of the Ube2w-131Δ~Ub conjugate during the experiment is due to hydrolysis of Ub and not aminolysis with the substrate (**Supplementary Figs. 2m** and **15c**). Thus, without an intact C terminus, Ube2w does not transfer Ub to a free  $\alpha$ -amino group.

## DISCUSSION

To date, Ube2w is the only E2 demonstrated to attach Ub directly and specifically to the N terminus of proteins. Although it shares certain mechanistic features with lysine-reactive E2s, our results indicate that Ube2w is uniquely adapted to facilitate selective  $\alpha$ -amino ubiquitination. As with UbcH5c, binding to a RING E3 shifts the population of Ube2w~Ub toward a closed conformation that facilitates aminolysis of the E2~Ub thioester. However, a highly flexible C-terminal region allows Ube2w to bind and ubiquitinate a diverse set of disordered N termini. The Ube2w C-terminal region adopts multiple orientations in proximity to the E2 active site cysteine.

Most E2s, including those that transfer SUMO or Nedd8 to lysine side chains, have a conserved N residue (N77 in UbcH5) in the loop immediately preceding the active site that is proposed to have both catalytic and structural roles for transfer to substrate lysines. N77 is thought to stabilize the oxyanion intermediate formed following lysine attack of the E2~Ub thioester<sup>30</sup> and most likely performs a structural role in helping to stabilize the loop preceding the active site<sup>31</sup>. Notably, Ube2w contains a histidine at this position,



**Figure 6 | The Ube2w C terminus facilitates  $\alpha$ -amino group reactivity.**

(a) Selected  $^1\text{H}$ - $^{15}\text{N}$  resonances of residues near the Ube2w active site are compared in the spectra of full-length WT (black), Ube2w<sup>W144E</sup> (blue) and Ube2w-131 $\Delta$  (red). Resonances move along similar trajectories as a result of the W144 mutation or C-terminal ablation, indicating similar chemical environments for the affected residues. (b) Mutation or ablation of the C terminus affects the intrinsic aminolysis activity of Ube2w. In a 1-h reaction, WT Ube2w-Ub shows robust transfer activity toward peptide (NH<sub>2</sub>-A-G-G-S-Y-COO<sup>-</sup>; 30 mM), as seen by increased amounts of free Ube2w and free Ub. Ube2w<sup>W144E</sup>-Ub and the Ube2w-131 $\Delta$ -Ub mutants show almost no Ub transfer activity to this minimal substrate over the same time period (Supplementary Fig. 21). MW, molecular weight.

H83. Mutation of H83 to asparagine (H83N or H94N in human Ube2w isoform 2) results in a marked decrease in the E3-enhanced N-terminal ubiquitination activity of Ube2w<sup>9</sup>, highlighting structural adaptations for N-terminal ubiquitination. Notably, Ube2w is unique among the ~40 human E2s as it contains a histidine at this key position, strongly implicating it as the sole E2 with N-terminal ubiquitination activity.

Our work shows that Ube2w's noncanonical, flexible C-terminal structure provides a platform for recognition of diverse substrates. We propose a mechanism in which Ube2w recognizes and binds backbone groups in substrates, explaining the observed requirement for N-terminal disorder in substrates. This mechanism is supported by observations that Ube2w can (i) transfer Ub to the  $\alpha\text{NH}_2$  group of a 5-amino-acid peptide but not to the  $\alpha\text{NH}_2$  group of a free amino acid and (ii) create linear Ub chains using Ub that has at least four unstructured residues (Met-Gly<sub>3</sub>-Ub) appended to its N terminus but not if those residues lack amide groups (Met-Pro<sub>3</sub>-Gly<sub>5</sub>-Ub) (Figs. 1b and 2e). Polyglycine lacks a side chain that could serve as a basis for recognition, implicating interaction via backbone atoms. The N-terminal sequences of proteins targeted by Ube2w are diverse, and we have not yet identified any Ube2w sequence preferences, but all of the sequences are either known to be or predicted to be disordered<sup>18–20</sup> (Supplementary Fig. 7). A mechanism that involves substrate backbone atoms would require those groups to be available; adoption of secondary structure such as  $\alpha$ -helices and  $\beta$ -strands by N-terminal residues would place these

groups in intramolecular hydrogen bonds, making them unavailable for formation of necessary contacts for Ube2w-dependent N-terminal ubiquitination.

Ube2w has evolved a unique C terminus whose relative orientation to active site residues seems to be crucial for Ube2w  $\alpha\text{NH}_2$  ubiquitination function, as its removal (Ube2w-131 $\Delta$ ) or alteration (Ube2w<sup>W144E</sup>) reduces substrate binding toward disordered N termini, affects the environment of active site residues, decreases intrinsic aminolysis activity and, consequently, inhibits ubiquitination of substrates (Figs. 5d,f and 6 and Supplementary Fig. 16). As shown by our ensemble, each individual member shows clear active site accessibility for an incoming substrate. Remarkably, C-terminal residues N136–W145, which can occupy multiple positions directly beneath the site of catalysis, represent the primary substrate recognition surface. Because Ube2w must ubiquitinate diverse substrate sequences, this region may have evolved to adopt multiple conformations to sterically accommodate unstructured N termini that harbor unique side chain identities.

There are numerous reports implicating N-terminal ubiquitination of specific proteins in cells. Perhaps the best-studied example is myogenic transcriptional switch protein (MyoD). Mutation of all lysine residues to arginine did not abrogate the protein's ubiquitination or its degradation in COS-7 cells, whereas specific chemical modification of the  $\alpha$ -amino group through carbamylation<sup>32</sup> inhibited ubiquitination of MyoD<sup>33</sup>. Other substrates identified by similar strategies are human papillomavirus 16 oncoprotein, E7 (E7-16), latent membrane protein 1 and inhibitor of differentiation 2 (ref. 34). In each case, truncation of the N-terminal region inhibited ubiquitination and degradation, implying a role for mobility or flexibility in the N terminus<sup>34</sup>. The extracellular signal-regulated kinase 3 (ERK3) has been identified as a target for N-terminal ubiquitination in HEK-293 cells, and an N-terminal Ub on ERK3 can be further modified to create Ub chains that signal for its degradation<sup>35</sup>. Bulky tags such as a Myc<sub>6</sub> or EGFP on the N terminus of WT ERK3 (which contains 45 lysines) inhibited its degradation in a proteasome-dependent manner, whereas smaller tags, such as HA or His<sub>6</sub>, had no such effect<sup>35</sup>. A crystal structure of ERK3 indicates that approximately the first ten residues are disordered (PDB code 216L). The N-terminal regions of all the above proteins are either known to be or are predicted to be disordered, consistent with the notion that they could be cellular Ube2w substrates<sup>18–20</sup>.

Identification of cellular Ube2w substrates is a critical step toward understanding the function of N-terminal ubiquitination. In eukaryotic cells, protein N termini seem to be an important site of regulation. N-terminal acetylation is a prevalent post-translational modification that most likely has a competing role with N-terminal ubiquitination, as acetylation is irreversible and would preclude further modification by Ub. Reports suggest that 60–90% of cytosolic proteins may harbor an acetyl moiety on their N termini<sup>36</sup>. We note that roughly 25% of proteins in the human proteome are predicted to have at least ten N-terminal unstructured residues. Additionally, post-translational proteolysis may expose new disordered N termini in protein substrates. An intriguing feature of the Ube2w C terminus is that it harbors a nuclear localization signal suggesting that it could have a role in the regulation of nuclear proteins<sup>37</sup>. Indeed, a majority of N-terminally ubiquitinated proteins identified thus far are nuclear proteins. Another notable possibility for Ube2w substrates includes nascent polypeptide chains on stalled ribosomes that may not contain any lysine residues.

Identification of N-terminally ubiquitinated substrates using existing proteomics approaches poses challenges. Current methods to enrich for ubiquitinated proteins that use the antibody that recognizes the Gly-Gly-Lys fragment generated by attachment of Ub via lysine side chains do not recognize substrates ubiquitinated at their N termini. N-terminal processing of proteins in cells means that an antibody derived toward a single Ub-amino acid linkage

(for example, Gly-Gly-Met) will not be sufficient to recognize all potential substrates. Therefore, new methods to elucidate the N-terminal ubiquitome are required. As a caveat, future studies focused on Ube2w should avoid the use of N-terminally tagged Ub or substrates that introduce N-terminal disorder as these can induce non-native Ube2w activity. In sum, Ube2w is unique among members of the Ub-conjugating enzyme family in both its structural and biochemical properties. Insights revealed here can guide future efforts to identify bona fide *in vivo* substrates for Ube2w and to further elucidate its distinct cellular function.

Received 7 April 2014; accepted 9 October 2014;  
published online 1 December 2014

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** PDB. The solution ensemble of full-length Ube2w was deposited under the code **2MT6**. Biological Magnetic Resource Bank. NMR assignments were deposited under the accession code **25150**.

## References

- Pickart, C.M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
- Deng, L. *et al.* Activation of the I $\kappa$ B kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103**, 351–361 (2000).
- Chen, Z. & Pickart, C.M. A 25-kilodalton ubiquitin carrier protein (E2) catalyzes multi-ubiquitin chain synthesis via lysine 48 of ubiquitin. *J. Biol. Chem.* **265**, 21835–21842 (1990).
- Brzovic, P.S. & Klevit, R.E. Ubiquitin transfer from the E2 perspective: why is UbcH5c so promiscuous? *Cell Cycle* **5**, 2867–2873 (2006).
- Nuber, U., Schwarz, S., Kaiser, P., Schneider, R. & Scheffner, M. Cloning of human ubiquitin-conjugating enzymes UbcH6 and UbcH7 (E2-F1) and characterization of their interaction E6-AP and RSP5. *J. Biol. Chem.* **271**, 2795–2800 (1996).
- Machida, Y.J. *et al.* UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation. *Mol. Cell* **23**, 589–596 (2006).
- McDowell, G.S., Kucerova, R. & Philpott, A. Non-canonical ubiquitylation of the proneural protein Ngn2 occurs in both *Xenopus* embryos and mammalian cells. *Biochem. Biophys. Res. Commun.* **400**, 655–660 (2010).
- Vosper, J.M. *et al.* Ubiquitylation on canonical and non-canonical sites targets the transcription factor neurogenin for ubiquitin-mediated proteolysis. *J. Biol. Chem.* **284**, 15458–15468 (2009).
- Scaglione, K.M. *et al.* The ubiquitin-conjugating enzyme (E2) Ube2w ubiquitinates the N terminus of substrates. *J. Biol. Chem.* **288**, 18784–18788 (2013).
- Tatham, M.H., Plechanovová, A., Jaffray, E.G., Salmen, H. & Hay, R.T. Ube2W conjugates ubiquitin to  $\alpha$ -amino groups of protein N-termini. *Biochem. J.* **453**, 137–145 (2013).
- Wu, W. *et al.* BRCA1 ubiquitinates RPB8 in response to DNA damage. *Cancer Res.* **67**, 951–958 (2007).
- Christensen, D.E., Brzovic, P.S. & Klevit, R.E. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nat. Struct. Mol. Biol.* **14**, 941–948 (2007).
- Guzzo, C.M. *et al.* RNF4-dependent hybrid SUMO-ubiquitin chains are signals for RAP80 and thereby mediate the recruitment of BRCA1 to site of DNA damage. *Sci. Signal.* **5**, ra88 (2012).
- Wenzel, D.M., Lissounov, A., Brzovic, P.S. & Klevit, R.E. UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature* **474**, 105–108 (2011).
- Zhang, Y. *et al.* UBE2W interacts with FANCL and regulates the monoubiquitination of Fanconi anemia protein FANCD2. *Mol. Cells* **31**, 113–122 (2011).
- Alpi, A.F., Pace, P.E., Babu, M.M. & Patel, K.J. Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Mol. Cell* **32**, 767–777 (2008).
- Wand, A.J., Urbauer, J.L., McEvoy, R.P. & Beiber, R.J. Internal dynamics of human ubiquitin revealed by <sup>13</sup>C-relaxation studies of randomly fractionally labeled protein. *Biochemistry* **35**, 6116–6125 (1996).
- Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

- Kelley, L.A. & Sternberg, M.J.E. Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371 (2009).
- Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
- Plechanovová, A., Jaffray, E.G., Tatham, M.H., Naismith, J.H. & Hay, R.T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115–120 (2012).
- Grimsley, G.R., Scholtz, J.M. & Pace, C.N. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.* **18**, 247–251 (2009).
- Sheng, Y. *et al.* A human ubiquitin conjugating enzyme (E2)-HECT E3 ligase structure-function screen. *Mol. Cell. Proteomics* **11**, 329–341 (2012).
- Vittal, V., Wenzel, D.M., Brzovic, P.S. & Klevit, R.E. Biochemical and structural characterization of the ubiquitin-conjugating enzyme UBE2W reveals the formation of a noncovalent homodimer. *Cell Biochem. Biophys.* **67**, 103–110 (2013).
- Laskowski, R.A. PDBsum new things. *Nucleic Acids Res.* **37**, D355–D359 (2009).
- Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA* **105**, 4685–4690 (2008).
- Shen, Y., Vernon, R., Baker, D. & Bax, A. *De novo* protein structure generation from incomplete chemical shift assignments. *J. Biol. NMR* **43**, 63–78 (2009).
- Pruneda, J.N. *et al.* Structure of an E3:E2-Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol. Cell* **47**, 933–942 (2012).
- Dou, H., Buetow, L., Sibbet, G.J., Cameron, K. & Huang, D.T. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat. Struct. Mol. Biol.* **19**, 876–883 (2012).
- Wu, P.Y. *et al.* A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J.* **22**, 5241–5250 (2003); erratum **23**, 4876 (2004); erratum **26**, 4051 (2007).
- Berndsen, C.E., Wiener, R., Yu, I.W., Ringel, A.E. & Wolberger, C. A conserved asparagine has a structural role in ubiquitin-conjugating enzyme. *Nat. Chem. Biol.* **9**, 154–156 (2013).
- Hershko, A., Heller, H., Eytan, E., Kaklij, G. & Rose, I.A. Role of the  $\alpha$ -amino group of protein in ubiquitin-mediated protein breakdown. *Proc. Natl. Acad. Sci. USA* **81**, 7021–7025 (1984).
- Breitschopf, K., Bengal, E., Ziz, T., Admon, A. & Ciechanover, A. A novel site of ubiquitination: the N-terminal residue, and not internal lysines of MyoD, is essential for conjugation and degradation of the protein. *EMBO J.* **17**, 5964–5973 (1998).
- Ciechanover, A. & Ben-Saadon, R. N-terminal ubiquitination: more protein substrates join in. *Trends Cell Biol.* **14**, 103–106 (2004).
- Coulombe, P., Rodier, G., Bonneil, E., Thibault, P. & Meloche, S. N-terminal ubiquitination of extracellular signal-regulated kinase 3 and p21 directs their degradation by the proteasome. *Mol. Cell. Biol.* **24**, 6140–6150 (2004).
- Dormeyer, W., Mohammed, S., Breukelen, B., Krijgsveld, J. & Heck, A.J. Targeted analysis of protein termini. *J. Proteome Res.* **6**, 4634–4645 (2007).
- Yin, G. *et al.* Cloning, characterization and subcellular localization of a gene encoding a human ubiquitin-conjugating enzyme (E2) homologous to the *Arabidopsis thaliana* UBC-16 gene product. *Front. Biosci.* **11**, 1500–1507 (2006).

## Acknowledgments

We acknowledge D. Christensen and C. Eakin for their initial observations on Ube2w and J.N. Pruneda for collecting SAXS data on Ube2w. This work was supported by National Institute of General Medical Sciences grants R01 GM088055 (REK), R01 GM098503 (P.S.B.), K99 NS073936 (K.M.S.) and R01 AG034228 (H.L.P.) and the University of Washington Hurd Fellowship Fund and Public Health Service National Research Service Award 2T32 GM007270 (to V.V.).

## Author contributions

V.V., P.S.B. and R.E.K. conceived the experiments and wrote the manuscript. V.V. performed the biochemical and structural experiments with help from K.M.S. and E.D.D. D.M.W. performed the initial characterization of Ube2w. V.B. and K.S.J.E.-J. performed MS. L.S. and D.B. performed the structure calculations. H.L.P. provided guidance. R.E.K. supervised the project.

## Competing financial interests

The authors declare no competing financial interests.

## Additional information

Supplementary information and chemical compound information is available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to R.E.K.

## ONLINE METHODS

**Plasmids and protein expression.** Ube2w constructs (WT, V30K/D67K (KK), 131Δ, 131Δ-KK, W144E, W144E-KK L110Q, C91S, C91S-KK and C91S/C119S/C151S/KK) were expressed from the pET24 vector without affinity tags. Ube2w plasmids were transformed into *Escherichia coli* (BL21 DE3) cells, and protein expression was induced with 0.5 mM iso-propyl-β-D-thio-galactoside (IPTG) at an OD<sub>600</sub> of 0.6, followed by growth for 16 h at 16 °C. Cells were lysed by French press in 25 mM sodium phosphate (pH 7.0) and 1 mM EDTA for full-length constructs and 25 mM 2-(*N*-morpholino) ethanesulfonic acid (pH 6.0) and 1 mM EDTA for 131Δ constructs. Following centrifugal clarification, Ube2w was applied to a cation exchange column with an elution gradient of 0–0.5 M NaCl. E2-rich fractions were identified by UV absorbance, concentrated, pooled and further purified by size exclusion chromatography on a Superdex 75 (GE Healthcare) column equilibrated in 25 mM sodium phosphate (pH 7.0), 150 mM NaCl. This buffer was used for all NMR experiments.

RPB8 (1–150) was expressed from the pET28a vector with a thrombin cleavable His<sub>6</sub> tag. RPB8 plasmids were transformed into *E. coli* (BL21 DE3) cells, and protein expression was induced with 0.5 mM iso-propyl-β-D-thio-galactoside (IPTG) at OD<sub>600</sub> of 0.6, followed by growth for 16 h at 16 °C. Cells were lysed in a French press in 25 mM Tris-HCl (pH 7.6), 200 mM NaCl and 10 mM imidazole and were applied to a Ni<sup>2+</sup>-NTA gravity flow column (Invitrogen) equilibrated in the same buffer. The column was washed with five column volumes of 25 mM Tris-HCl (pH 7.6), 200 mM NaCl and 50 mM imidazole and eluted with 25 mM Tris-HCl (pH 7.6), 200 mM NaCl and 500 mM imidazole. The elution was subject to cleavage at 4 °C overnight in 25 mM Tris-HCl (pH 7.6), 200 mM NaCl with 1 mg thrombin from bovine serum (Sigma-Aldrich). Following cleavage, the dialyzed sample was reappplied to a Ni<sup>2+</sup>-NTA gravity flow column. The flow-through was collected, concentrated and further purified by size exclusion chromatography on a Superdex 75 (GE Healthcare) column equilibrated in 25 mM sodium phosphate (pH 7.0) and 150 mM NaCl. Uba1 and Ub were cloned, expressed and purified as previously described<sup>38</sup>. Identical purification protocols for WT Ub were followed for Met-Gly-Ub, Met-Gly<sub>3</sub>-Ub, Met-Gly<sub>5</sub>-Ub, Met-Gly<sub>7</sub>-Ub, Ub-I44A, HA-Ub and HA-Ub(K0). BC<sub>112</sub>-BD<sub>115</sub> was expressed from the pET28N and pCOT7N expression systems (generous gift of M. Wittekind, Bristol-Myers Squibb). Proteins were purified as previously described<sup>39</sup>. The E2s UbcH5c, Ube2e1 and UbcH7 were cloned, expressed and purified as previously described<sup>12</sup>. Ataxin-3 was cloned, expressed and purified as previously described<sup>40</sup>. Tau protein was cloned, expressed and purified as previously described<sup>41</sup>. Full-length CHIP and CHIP U-box were cloned, expressed and purified as previously described<sup>42,43</sup>.

**In vitro ubiquitination assays.** Ubiquitination assays involving RPB8 or HA-Ub were conducted at 37 °C for 1 h in 25 mM sodium phosphate (pH 7.0) and 150 mM NaCl. Protein concentrations were as follows: 1 μM Uba1, 4 μM E2 enzyme, 4 μM RPB8, 2 μM BC<sub>112</sub>-BD<sub>115</sub>, 30 μM Ub or HA-Ub and 10 mM MgCl<sub>2</sub>. Reactions were initiated by addition of 5 mM ATP. Samples were boiled in SDS buffer, loaded onto an SDS-PAGE and visualized by western blotting with the appropriate antibodies, anti-RPB8 (Abnova H00005437-A01) and anti-Ub (Santa Cruz P4D1). The dilution used for both antibodies was 1:5,000.

Ubiquitination assays for tau, ataxin-3 and CHIP were typically performed for 1–5 min at 37 °C in 10-μl mixtures containing buffer A (50 mM Tris (pH 7.5), 50 mM KCl and 0.2 mM DTT), Ub<sup>mix</sup> (2.5 mM ATP, 5 mM MgCl<sub>2</sub>, 50 nM Ube1 and 250 μM Ub), 1 μM of the indicated E2, 1 μM CHIP and 1 μM of either tau or ataxin-3. For CHIP-independent ubiquitination reactions, CHIP was omitted from the reaction, the substrate concentration was increased to 20 μM and the E2 concentration was 9 μM unless otherwise indicated. Reactions were stopped by addition of SDS-Laemmli buffer and boiling, followed by separation of proteins by SDS-PAGE and visualization by western blotting with appropriate antibodies.

**Nucleophile reactivity assays.** Nucleophile reactivity assays were performed at 37 °C in buffer containing 25 mM sodium phosphate (pH 7.0) and 150 mM NaCl. Reactions contained 1 μM Uba1, 20 μM Ub, 20 μM E2 and 10 mM MgCl<sub>2</sub>. Reactions were initiated with 5 mM ATP and allowed to form E2-Ub conjugates for 30 min. Nucleophile reactivity was induced by addition of 50 mM free lysine or 50 mM 5-aa peptide (NH<sub>2</sub>-A-A-G-S-Y-COO<sup>-</sup>) and allowed to react for 1 h. Samples were collected in nonreducing SDS buffer and

loaded onto a SDS-PAGE gel. Results were visualized by Coomassie staining. Peptides were purchased from United Biosystems Inc.

**Protein identification by LC/MS/MS.** Protein identification and ubiquitination sites on RPB8 were conducted on the basis of previously described protocols<sup>9</sup>. Briefly, protein bands corresponding to ubiquitinated substrate (Supplementary Fig. 4a) were excised and destained with 30% methanol for 4 h. Upon reduction (10 mM DTT) and alkylation (65 mM 2-chloroacetamide or iodoacetamide, with similar results) of the cysteines, proteins were digested overnight with sequencing-grade modified trypsin (Promega). Resulting peptides were resolved on a nano-capillary reverse phase column (Pico frit column, New Objective) using a 1% acetic acid/acetonitrile gradient at 300 nl/min and directly introduced into a linear ion-trap mass spectrometer (LTQ Orbitrap XL, Thermo Fisher). Data-dependent MS/MS spectra on the five most intense ions from each full MS scan were collected (relative collision energy 35%). Proteins were identified by searching the data against the Human International Protein Index database (version 3.5) appended with decoy (reverse) sequences using the X!Tandem/Trans-Proteomic Pipeline (TPP) software suite. All peptides and proteins with a PeptideProphet and ProteinProphet probability score of >0.9 (false discovery rate <2%) were considered positive identifications and manually verified.

**NMR spectroscopy.** All NMR samples were prepared in 25 mM sodium phosphate (pH 7.0), 150 mM NaCl using either 90% H<sub>2</sub>O/D<sub>2</sub>O or 100% D<sub>2</sub>O. Samples for Ube2w-KK used either uniformly <sup>15</sup>N- or <sup>15</sup>N, <sup>13</sup>C-labeled protein at concentrations from 400 μM to 200 μM. Titration experiments involving [<sup>15</sup>N]Ube2w-KK were performed by equimolar addition of unlabeled RPB8 or tau. The magnitude of CSPs for each resonance was quantified in Hz according to the equation  $\Delta\delta_j = ((\Delta\delta_j^{15N/5})^2 + (\Delta\delta_j^{1H})^2)^{1/2}$ . Data collection for resonance assignments used standard three-dimensional NMR techniques<sup>44</sup> collected on INOVA 600 and 800 MHz spectrometers (Varian) at Pacific Northwest National Labs (PNNL). All other NMR-based experiments (substrate titrations; <sup>1</sup>H-<sup>15</sup>N hetNOEs; T<sub>1</sub>, T<sub>2</sub> relaxation experiments; <sup>1</sup>H-<sup>15</sup>N RDC measurements; and paramagnetic spin labeling experiments) were collected on a 500 MHz Bruker Avance II (University of Washington). All spectra were collected at 25 °C. Data were processed using NMR-Pipe/NMRDRaw<sup>45</sup> and visualized with NMRView<sup>46</sup>.

**Spin-label modification.** A mutant form of Ube2w (C91S, C119S, C151S, V30K, D67K) was generated by site-directed mutagenesis to incorporate a single cysteine chemical modification at position C135. A similar mutant (C119S, C135S, C151S, V30K, D67K) was created for the C91 modification. The thiol-reactive relaxation probe 4-(2-iodoacetamido)-TEMPO (Sigma-Aldrich) was mixed at a 1:5 Ube2w/TEMPO molar ratio for 2 h at 30 °C. Reaction yields were quantified by MALDI-MS. Only those that reacted to completion (>95%) were used for spin-label experiments. Unreacted probe was cleared by dialysis overnight at 4 °C. Identical <sup>1</sup>H-<sup>15</sup>N HSQC-TROSY experiments were conducted in the presence and absence of the probe reducing agent, ascorbate.

**RDCs.** Pf1-phage (ALSA Biotech) was added to 20% in a solution containing 250 μM [<sup>15</sup>N]Ube2w-KK in 25 mM sodium phosphate (pH 7.0), 150 mM NaCl. This resulted in precipitation of the protein around the phage. An additional 200 mM NaCl was added to dissolve the precipitant. The sample was centrifuged to remove bubbles. The final solution contained 10 mg/ml phage and 350 μM NaCl.

**SAXS.** SAXS data were collected at Stanford Synchrotron Radiation Lightsource beamline 4-2. Data were collected for Ube2w at concentrations of 10 mg/ml, 5 mg/ml and 0.5 mg/ml in 25 mM sodium phosphate (pH 7.0), 150 mM NaCl and 2 mM DTT at 25 °C. SAXS statistics were calculated using the EMLB CRY SOL<sup>47</sup> server.

**CS-Rosetta structure calculation for Ube2w.** Monomeric structures of Ube2w were generated from NMR data in combination with homologous structural information. The standard CS-Rosetta method is used to derive fragments from the sequence profile, predicted secondary structure and backbone and Cβ chemical shift data<sup>24,25,48</sup>. Interatomic restraints are derived from alignments of Ube2w sequence to templates<sup>49</sup>. The computational protocol includes a low-resolution stage and a high-resolution stage. In the low-resolution step, side chains are represented using a single, residue-specific pseudo-atom, positioned

at the C $\beta$  carbon. From an extended chain, structures are assembled by fragment insertion under a force field that favors compactness and formation of secondary structures. In the high-resolution stage, side chains and hydrogen atoms are explicitly represented, and the structures are optimized using the Rosetta full-atom energy function. In both stages of conformational sampling, Rosetta energy function is augmented with a penalty term related to the homologous structural and experimental N-H RDC restraints. Backbone CSP data (residues 72, 78, 85, 86, 87, 95, 96, 97, 124 and 131) derived upon deletion of residues 132–151 was also included as ambiguous distance constraints during sampling and refinement. A total of 16,000 models are generated. Models with top 10% lowest Rosetta energy, 25% lowest homologous constraint energy and 25% lowest RDC energy were selected for further analysis. Paramagnetic broadening data from spin labels at residues C91 and C135 and SAXS were used to rank the top models. Pearson correlation was calculated between experimental data and distance from paramagnetic center to the HN atom. Distances were calculated by explicitly adding the spin label to each model at residue 135 or 91. The top 20 models with the best correlation with paramagnetic quenching and SAXS data were chosen as the final ensemble. Structural statistics were calculated using the Protein Structure Validation Server suite (PSVS)<sup>50</sup>. Favorable Ramachandran statistics were observed, with 88.9% of residues in most favored regions, 10.6% in additionally allowed regions, 0.5% in generously allowed regions and 0% in disallowed regions.

**Statistical analysis.** Pearson correlations were calculated by standard methods. The Pearson correlation is the standard product-moment correlation coefficient that describes the linear correlation between two variables. The Q factor for RDCs reflects the agreement between calculated  $Q_{\text{calc}}$  and experimental  $Q_{\text{obs}}$  dipolar couplings as:  $Q = \text{r.m.s.} (D^{\text{calc}} - D^{\text{obs}}) / (D_a^2 [4 + 3R^2] / 5)^{1/2}$ , where  $D_a$  and  $R$  refer to the magnitude and rhombicity of the alignment tensor, respectively<sup>51</sup>. The SAXS agreement is calculated using the EMBL Crysol server, which evaluates X-ray solution scattering curves from atomic models.  $\chi$  describes the discrepancy between theoretical and experimental curves<sup>47</sup>.

38. Pickart, C.M. & Raasi, S. Controlled synthesis of polyubiquitin chains. *Methods Enzymol.* **399**, 21–36 (2005).
39. Brzovic, P.S. *et al.* Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proc. Natl. Acad. Sci. USA* **100**, 5646–5651 (2003).
40. Todi, S.V. *et al.* Cellular turnover of the polyglutamine disease protein ataxin-3 is regulated by its catalytic activity. *J. Biol. Chem.* **282**, 29348–29358 (2007).
41. Barghorn, S., Biernat, J. & Mandelkow, E. Purification of recombinant tau protein and preparation of Alzheimer-paired helical filaments *in vitro*. *Methods Mol. Biol.* **299**, 35–51 (2005).
42. Winborn, B.J. *et al.* The deubiquitinating enzyme ataxin-3, a polyglutamine disease protein, edits Lys63 linkages in mixed linkage ubiquitin chains. *J. Biol. Chem.* **283**, 26436–26443 (2008).
43. Xu, Z. *et al.* Structure and interactions of the helical and U-box domains of CHIP, the C terminus of HSP70 interacting protein. *Biochemistry* **45**, 4749–4759 (2006).
44. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
45. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
46. Johnson, B.A. & Blevins, R.A. NMR View: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
47. Svergun, D., Barberato, C. & Koch, M.H.J. CRYSOLE: A program to evaluate X-ray solution scatter of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
48. Raman, S. *et al.* NMR structure determination for larger proteins using backbone-only data. *Science* **327**, 1014–1018 (2010).
49. Thompson, J.M. *et al.* Accurate protein structure modeling using sparse NMR data and homologous structure information. *Proc. Natl. Acad. Sci. USA* **109**, 9875–9880 (2012).
50. Bhattacharya, A., Tejero, R. & Gaetano, M.T. Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778–795 (2007).
51. Ulmer, T.S., Ramirez, B.E., Delaglio, F. & Bax, A. Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *J. Am. Chem. Soc.* **125**, 9179–9191 (2003).