Structural and Functional Characterization of Non-Canonical Ubiquitin Conjugating Enzymes

Vinayak Vittal

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Rachel E. Klevit, Chair

Ning Zheng

Dana Miller

Program Authorized to Offer Degree:

Department of Biochemistry

University of Washington

Abstract

Structural and Functional Characterization of Non-Canonical Ubiquitin Conjugating Enzymes

Vinayak Vittal

Chair of the Supervisory Committee:

Professor Rachel E. Klevit

Department of Biochemistry

Thirty-five years ago, during its initial discovery, the field of ubiquitination could not have been more aptly named. In fact, as the study of ubiquitin (Ub) expands, it is becoming clear that this small post-translational modification is far more ubiquitous than the founders of this field could have originally imagined. Initially identified as a protein degradation signal, Ub is now known to regulate vastly disparate cellular functions, including the DNA-damage response and immune signaling. This small (8.5kDa) modification achieves its diverse signaling capabilities partly through the enzymes that catalyze its attachment to cellular substrates. These enzymes (E1, E2, and E3) work in succession to identify and modify substrates with varying Ub signals. The manner in which Ub is attached (monoubiquition or 8 types of polyubiquitin chains) and the site of attachment dictate a substrate's cellular fate. Furthermore, much of the diversity in Ub signaling can be attributed to the structural characteristics of the enzymes in the ubiquitin cascade. This thesis examines the structural and biochemical features of two E2 ubiquitin-conjugating enzymes, Ube2w and Ube2h. First, we show that Ube2w specifically attaches monoubiquitin to the N-terimni of disordered protein substrates. To attach this unique Ub modification, Ube2w has evolved a non-canonical domain architecture, solved by our group using nuclear magnetic resonance (NMR) techniques. We also show, that another E2, Ube2h, contains a disordered C-terminus and that it can ubiquitinate a substrate, histone H2A/H2B, *in vitro*, in the absence of an E3. Studies presented here build the foundations to understand the cellular impact of Ube2w and Ube2h.

Table of Contents

List of Figures

List of Tables

Chapter I – Introduction

Discovery and progress in the ubiquitination field

Proteins are the molecular machines that govern nearly every functional process in our cells. The human genome contains roughly 20,000 protein-encoding genes, but due to DNA and RNA splicing events the total number of unique proteins in cells is many times more[1]. With such a diverse cast of proteins, regulation is paramount, particularly when considering proteolysis – the breakdown of proteins and peptides. The ability for cells to recycle unwanted or damaged proteins is essential and has been mechanistically uncovered through myriad of biochemical and cellular methods over the past seventy years. The concept of protein degradation was first uncovered in 1942 when Rudolf Schoenheimer made the discovery that cells outlive the proteins within them[2]. For nearly thirty-five years following this discovery it was believed that the lysosome was the primary mechanism by which proteins were degraded due to the high concentration of proteases in the organelle and because the lysosome degraded endocytosed proteins. However, in 1977 Joseph Etlinger and Alfred Goldberg discovered that rabbit reticulocytes, which lack lysosomes, were capable of degrading an abnormal form of hemoglobin in an ATP-dependent fashion, revealing an alternative method of protein degradation[3]. Following the results from Goldberg and company, Avram Hershko, Aaron Ciechanover, and Irwin Rose, through reticulocyte fractionation, discovered two components of this new degradation pathway. The first, termed APF-1 (ATP-dependent Proteolysis Factor 1), was shown to be heat-stable, and formed high molecular weight protein products in the presence of a second ATP-binding component, APF-2 (fraction II)[4,5]. Along with Arthur Haas, they went on to show that these high molecular weight products were covalently linked[6].

Nearly simultaneously, Gideon Goldstein and colleagues discovered an 8.5kDa polypeptide from bovine thymus, that *in vitro*, was capable of inducing the differentiation of both T-cells and B-cells. Their results showed that this polypeptide, provisionally named at the time as ubiquitous immunopoietic polypeptide (UBIP), was highly conserved throughout evolution and nearly universally expressed in living cells[7]. Two years after this discovery Ira Goldknopf demonstrated through tryptic digest and Edman degradation that UBIP post-translationally modified histone H2A via an isopeptide linkage on Lys119[8]. Somewhat serendipitously, in 1980 Ketih Wilkinson, Michael Urban, and Haas revealed that APF-1 and

UBIP, now termed ubiquitin (Ub), were in fact the same molecule[9]. Between 1980 and 1984 Ciechanover, Hershko, and others would go on to make seminal discoveries that Ub was successively passed among three enzymes, the E1, E2, and E3 in order to be attached to substrates[10]. They found that the E1 first adenylates the C-terminus of Ub, which is then transferred to its active site cysteine creating an E1~Ub conjugate. Ub is then passed to the active site cysteine of an E2 in a transthiolation reaction creating an E2~Ub conjugate. Finally, the E2~Ub conjugate works with an E3 to transfer Ub to a substrate lysine. Over time, it was discovered that Ub-dependent protein tagging was the major cause of proteolysis in cells[11] (**Figure 1**). The culmination of this work occurred in 2004 when Ciechanover, Hershko, and Rose were awarded the Nobel Prize in Chemistry.

In the thirty-five years since the discovery of Ub, the field has expanded at an astonishing rate. Spurring this growth was the recognition that Ub was not simply a proteolytic signal, but that it was a diverse post-translational modification that regulated different cellular processes. Much of its cell-signaling potential can be attributed to the multiple modes in which Ub can structurally be attached to its substrates. In its simplest form, a single Ub is covalently linked to a substrate lysine, resulting in monoubiquitination. A classic case is the monoubiquitination of lysine 123 on histone H2B, which acts as a positive regulator of transcription[12]. Generally, monoubiquitination events are thought to aid in protein-protein interactions. True diversity in Ub signaling, however, comes from the ability of Ub to be attached to any of its 7 lysines or its N-terminus, resulting in Ub polymers. Each polyubiquitin chain type acts in different cellular pathways. K48-linked chains are the canonical signal for proteasomal degradation while K63-linked chains predominantly signal in DNA-damage response pathways[13,14]. N-terminally linked linear Ub chains are important in mediating inflammatory responses via NF-κB signaling[15]. Although these chain types are the most thoroughly characterized it is known that in cells, Ub chains can be built from any of its other lysine residues (K6, K11, K27 K29, or K33)[16]. Adding to this diversity, polyubiquitin chains can also consist of mixed linkages where chains are branched from different lysine residues. The functional range afforded by the simple 8.5kDa Ub signal is thought to lie in both the structural signatures of each of these chain types and the machinery with which it is attached to its substrates.

Three sets of enzymes (E1, E2, and E3) drive the attachment of Ub to its substrates (**Figure 1**). The human genome encodes 2 E1s, between 35-40 E2s, and hundreds of E3s and the E2-E3 pair that

works to attach Ub to its substrate generally drives the range of modifications seen in cells. E2s dictate different Ub chain linkages and help target specific lysines on the substrate for ubiquitination. This class of enzymes will be expounded upon further in the next section. E3s are responsible for recognizing and dictating what substrates in cells are selected for ubiquitination. The hundreds of identified E3s are divided in to three classes: RING-type (Really Interesting New Gene), HECT-type (Homologous to E6AP C-terminus), and RBR-type (RING between RING) E3s. Each of these is structurally and mechanistically different. In RING-type mechanisms, a zinc-chelating RING finger domain binds an E2~Ub conjugating enzyme while a substrate-binding domain engages the substrate. From here, the RING domain allosterically activates the E2~Ub conjugate to directly transfer its Ub on to the substrate[17-19]. RING-type E3s are the most abundant E3-type in humans. In HECT-type mechanisms, the HECT domain engages the E2~Ub conjugate and a transthiolation reaction occurs, transferring Ub to the active site cysteine of the HECT. This E3~Ub conjugate then engages substrate and directly ubiquitinates it. The final class of E3s, RBRs, uses a hybrid RING-HECT mechanism where a RING-like domain engages an E2~Ub conjugate, but transfers Ub to a distal active site cysteine[20]. The RBR E3~Ub conjugate then engages substrate. A principle challenge the ubiquitination field faces today is in determining which E2-E3-substrate combinations exist in cells. Uncovering these cellular protein networks has enormous implications for understanding certain disease pathologies and potentially, towards drug discovery. Building a strong biochemical foundation upon which to understand E2s and E3s is paramount towards achieving these goals.

Diversity in E2 ubiquitin-conjugating enzymes

E2s play an integral role in the ubiquitination pathway, particularly in RING-type mechanisms. E3s bring substrate and E2~Ub conjugates into atomic proximity allowing the E2 active site and C-terminus of Ub to make physical interactions with substrates. Much of the diversity seen in Ub signaling is driven by the fact that humans have evolved E2s of varying function. Furthermore, the functional variability in E2s can be attributed to the structural properties of these enzymes. All members of the E2 family contain a catalytic ubiquitin-conjugating (UBC) domain characterized by a conserved structural architecture (**Figure 2**). This domain plays a central role in modulating interactions with the E1, E3s, Ub,

and the target ubiquitination site on the substrate. Within the E2 family, the UBC domain is ~35% conserved and has a defined structural architecture[21]. Classically UBC domains contain four α-helices, a short 3_{10}-helix near the active site, and a four-stranded β-sheet. In the view presented in the left side of Figure 2 the active site is on the bottom of the E2, sandwiched between strand 4 of the β-sheet and the 3_{10}-helix (**Figure 2**). Conserved amino acids surrounding the active site play critical roles in facilitating the transfer of Ub to substrate lysines. Ten residues preceding the active site is a loop where a sequence known as the HPN (histidine-proline-asparagine) motif is believed to play both structural and catalytic roles in facilitating Ub transfer (**Figure 3**). While the histidine is thought to stabilize this loop structurally, the asparagine is believed to facilitate the formation of a tetrahedral intermediate with the carbonyl of G76 in Ub during nucelophilic attach by an incoming substrate lysine side chain[22,23]. Mutation of these residues severely impacts Ub transfer to substrates. A second conserved position in lysine-transferring E2s is an acidic residue in loop 8. The negative charge from the side chain carboxyl group at this position is thought to suppress the pKa of an incoming lysine side chain to facilitate its deprotonation[19] (**Figure 3**). These positions are necessary for the catalytic activity of E2s, however, mutation of several residues near the active site have effects on catalysis[19,20].

Although the UBC domain is the catalytic center, structural features outside of this core region are what define the activities for many E2s. For this reason E2s are classified on the basis of N-terminal or C-terminal extensions emanating from the UBC domain. Class I E2s consist of only the catalytic UBC domain. Class II E2s harbor a C-terminal extension while class III E2s contain N-terminal extensions. E2s that have both N-terminal and C-terminal extensions are defined as class IV. Much of what we understand about E2 structure and function has come from studies on class I E2s such as UbcH7, Ubc13, and members of the UbcH5 sub-family. Studies on UbcH5 family members have uncovered mechanistic insights in to the activation of E2~Ub conjugates by RING-type E3s and insights into polyubiquitination and monoubiquitination[17-19,24] . Early studies with Ubc13 revealed a novel interaction with an adapter protein MMS2, which caused the specific catalysis of K63-linked chains[25]. Studies on UbcH7 originating in the Klevit Lab revealed this E2 to be HECT and RBR-specific and not reactive with RING-type E3s[20]. Building upon this research, the Ub field is beginning to understand how members of the other E2 classes function biochemically.

The most pertinent questions about class II, III, and IV E2s are in regards to their N and C-terminal extensions. Particularly, are these regions structured or disordered and what are their functional roles in Ub transfer activity? Studies on Ube2k and Ube2e1 have revealed distinct structural and mechanistic insights in to class II and class III E2s, respectively. Ube2k is a 200 amino acid E2 that builds K48-linked Ub chains. It contains a canonical UBC domain followed by a 47 residue C-terminal extension. Structural studies have revealed that this extension is in fact an ordered, ubiquitin-associating (UBA) domain[26]. UBAs are known to mediate noncovalent interactions with Ub. Elegant studies from Cecile Pickart's lab demonstrated that the C-terminal UBA domain of Ube2k was necessary for the formation of K48-linked chains and that the C-terminus protected the active site of Ube2k from alkylation[27]. Her group went on to show that the ability of Ube2k to build K48-linked chains was specific to the UBC and UBA domains of Ube2k as a grafting experiment where the C-terminus of Ube2k was fused to the UBC domain of another E2 failed to induce K48-linked chain building activity. These were the first studies to show that elements outside of the UBC domain were critical towards the activity of an E2.

It is clear, however, that different class II and III E2s use their extensions in very different ways. Ube2e1 is a 193 amino acid class III E2 whose first 39 N-terminal residues are disordered. Two groups revealed that Ube2e1's activity is sensitive to a number of factors, including the presence of its N-terminus. Day et al. showed that when interacting with certain RING-type E3 ligases, the N-terminus of Ube2e1 inhibits the ubiquitination activity of this E2[28]. Conversely, they went on to show that an N-terminally truncated form of this E2 had increased ubiquitination activity towards certain E3s. In a second study related to the regulation of Ube2e1, Nguyen et al. demonstrated that the N-terminal inhibition is E3-dependent and residues on the "backside" β-sheet and Ub itself are important regulators of Ube2e1 activity[29]. These results highlight the variable regulatory mechanisms of terminal extensions in class II and III E2s. The final class of E2s, class IV, is the least characterized and smallest group of E2s. Three E2s have been identified in this class, Ube2O, Ube2z, and BIRC6. The E2s in this family are between 240-4800 amino acids in length, with large N and C-terminal extensions. They also work in diverse cellular processes such as reticulocyte maturation, hematopoiesis, anti-apoptotic function, and cellular abscission. A great deal of work must be done to characterize this class of E2s biochemically and structurally.

Non-canonical sites of ubiquitination

In canonical ubiquitination mechanisms, the E1, E2, and E3 work to attach Ub to the side chain of substrate lysine residues. However, it has become clear that in cells Ub can also be attached to different side chain residues such as serine, threonine, cysteine, and even to the amino terminus of protein substrates. Attachments to these sites are considered to be non-canonical ubiquitination events. The most characterized non-lysine ubiquitination site is the amino terminus of protein substrates. As early as 1984, Avram Hershko recognized that the amino termini of protein substrates were viable sites for ubiquitination in cells[30]. Building on this work, Breitschopf et al. identified MyoD as a substrate that is specifically degraded via ubiquitination of its N-terminus[31]. Over time, the number of known substrates ubiquitinated on their N-termini has expanded and an enzyme responsible for this unique modification has been identified[32]. Ube2w is a class I E2 that specifically ubiquitinates the N-terminus of protein substrates *in vitro*. There is no confirmation of this activity yet in cells, however, it is a very likely the obvious candidate as no other E2 shares its structural and functional features. The structural and biochemical aspects of this enzyme will be expounded upon in Chapters II and III of this thesis.

Other non-canonical sites of ubiquitination include serines and threonines, which provide suitable side chain nucleophiles for Ub to form hydroxyester linkages. The strongest evidence for attachment of Ub to these non-canonical sites comes from the endoplasmic reticulum-associated degradation (ERAD) pathway. ERAD ubiquitination is an important mechanism to clear misfolded proteins via the proteasome and has important implications in cellular immunity. One known substrate targeted for degradation via serine ubiquitination is T-cell receptor α (TCR α). In its C-terminus, TCR α contains two serines that when mutated to alanine greatly decreased its ubiquitination and degradation[33]. Serine and threonine ubiquitination are also implicated in HIV immunity. Research has shown that BST-2, an important regulator of immunity, is targeted for ubiquitination on serines and threonines by the HIV protein Vpu and a human RING E3 ligase[34]. Ubiquitination of BST-2 helps HIV to evade the human immune system. Biochemically, it is thought that the Ube2j (class II) and Ube2g (class I) family of E2s are responsible for ubiquitination of serines and threonines. Finally, it is known that Ub can also be attached to substrate cysteines via a thioester linkage. Because of the heavily reducing environment of the cell it may seem counterintuitive to consider a thioester linkage as stable site of attachment, however, E1s, E2s, and E3s

all form thioester linkages to Ub themselves in cells. One example of the attachment of Ub to a substrate cysteine is the peroxisomal import factor Pex5p. Results showed that a lysine-less form of Pex5p was still ubiquitinated and that ubiquitination was lost following treatment with reducing agent[35]. Although ubiquitination of lysine side chains is the preferred site of Ub attachment in substrates, evidence for non-canonical ubiquitination events is mounting. Building a greater understanding of these new types of modifications could have large impacts on dissecting varying cellular processes.

Statement of objectives

As the Ub field diversifies it is increasingly clear that the canonical rules regarding certain biochemical and structural aspects of the Ub machinery will be broken. This is particularly true when considering three points in regards to E2 ubiquitin-conjugating enzymes: 1) expansion of studies towards class II, III, and IV E2, 2) research in to non-canonical ubiquitination events, and 3) the increasing prevalence of intrinsic disorder in the Ub system. This particular thesis takes in to account all of these considerations in the studies of two E2s, Ube2w and Ube2h. Ube2w is a class I E2 that has a non-canonical UBC domain and specifically ubiquitinates the disordered N-termini of protein substrates. Ube2h is a class II E2 with a canonical UBC domain and an unstructured C-terminus. Unexpectedly, we found that this E2 has E3-independent ubiquitination activity towards an *in vitro* substrate, histone H2A/H2B. This thesis provides structural and biochemical data on Ube2w and Ube2h that explain their unique functions. The objectives for this thesis are as follows:

1) Use biochemical techniques to understand fundamental questions about the structural architecture of the E2, Ube2w
2) Develop a method to directly assess the ubiquitination activity of Ube2w
3) Structurally characterize Ube2w and relate this to its unique function.
4) Use structural biology and biochemistry to dissect the function of the class II E2, Ube2h
5) Understand the nature of the interaction between Ube2h histone H2A/H2B.

References

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genom. *Nature*. **431**, 931-945 (2004).

2. Schoenheimer, R. The dynamic state of body constituents. *Harvard University Press*. Cambridge, MA (1942).

3. Etlinger, J.D., Goldberg, A.L. A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes. *Proc Natl Acad. Sci USA*. **74**, 54-58 (1977).

4. Ciechanover, A., Hod, Y., Hershko, A. A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. *Biochem Biophys. Res. Commun*. **81**, 1100-1105 (1978).

5. Hershko, A., Ciechanover, A., Rose, I.A. Resolution of the ATP-dependent proteolytic system from reticulocytes: a component that interacts with ATP. *Proc. Natl. Acad. Sci. USA.* **76**, 3107-3110 (1979).

6. Ciechanover, A., Heller, H., Elias, S., Haas, A.L., Hershko, A. ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc. Natl. Acad. Sci. USA.* **77**, 1365-1368 (1980).

7. Goldstein, G., Scheid, M., Hammerling, U., Schlesinger, D.H., Niall, H.D., Boyse, E. A. Isolation of a polypeptide that has lymphocyte-differentiation properties and is probably represented universally in living cells. *Proc. Natl. Acad. Sci. USA*. **72**, 11-15 (1975).

8. Goldknopf, I.L., Busch, H. Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. *Proc. Natl. Acad. Sci. USA.* **74**, 864-868 (1977).

9. Wilkinson, K.D., Urban, M.K., Haas, A.L. Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. *J. Biol. Chem*. **255**, 7529-7532 (1980).

10. Hershko, A., Heller, H., Elias, S., Ciechaover, A. Components of the ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *J. Biol. Chem*. **258**, 8206-8214 (1983).

11. Finley, D., Ciechanover, A., Varshavsky, A. Thermolability of ubiquitin activiating enzyme from the mammalian cell cycle mutant ts85. *Cell*. **37**, 43-55 (1984).

12. Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heiot, J., Boateng, M.A., et al. Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Mol. Cell*. **11**, 267-274 (2003).

13. Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K., Varshavsky, A., A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science.* **243**, 1576-1583 (1989)

14. Al-Hakim, A., Escribano-Diaz, C., Landry, M .C., O'Donnell, L, Panier, S., Szilard, R.K., Durocher, D. The ubiquitous role of ubiquitin in the DNA damage response. *DNA Repair*. **9**, 1229-1240 (2010).

15. Rahighi, S., Ikeda, F., Kawasaki, M., Akutsu, M., Suzuki, N., Kato, R., et al. Specific recognition of linear ubiquitin chains by NEMO is important for NF-κB activation. Cell 136, 1098-1109. (2009).

16. Peng, J., Schwartz, D., Elias, J.E., Thoreen, C.C., Cheng, D., Marsischky, G., et al. A proteomics approach to understanding protein ubiquitination. *Nat. Biotech*. **21**, 921-926 (2013).

17. 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).

18. 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).

19. 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).

20. 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).

21. van Wijk, S.J., Timmer, H.T. The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins. *FASEB J*. **24**, 981-993 (2010).

22. Wu, P.Y., et al. A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J*. **22**, 5241-5250 (2003).

23. 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**, 154-16 (2013).

24. Brzovic, P.S., Lissounov A., Christensen D.E., Hoyt D.W., Klevit R.E.. A UbcH5c/ubiquitin noncovalent complex is required for processive BRCA1-direct ubiquitination. *Mol Cell*. **21**, 873-880 (2006).

25. VanDermark, A.P., Hofmann, R.M., Tsui, C., Pickart, C.M., Wolberger, C. Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell*. *105*, 711-720 (2001).

26. Ko, S. et al. Structural basis of E2-25K/UBB+1 interaction leading to proteasome inhibition and neurotoxicity. *J. Biol. Chem*. **285**, 36070-36080 (2010).

27. Haldeman, M.T., Xia, G., Kasperek, E.M., Pickart, C.M. Structure and function of ubiquitin conjugating enzyme E2-25K: the tail is a core-dependent activity element. *Biochemistry*. **36**, 10526-10537 (1997).

28. Schumacher, F., Wilson, G., Day, C.L. The N-terminal Extension of UBE2E Ubiquitin-Conjugating Enzymes Limits Chain Assembly. *J. Mol. Biol*. **425**, 4099-4111 (2013).

29. Nguyen, L., Plafker, K.S., Starnes, A., Cook, M., Klevit, R.E., Plafker, S.M. The Ubiquitin-Conjugating Enzyem, UbcM2, I Restricted to Monoubiquitylation by a Two-Fold Mechanism That Involved Backside Residues of E2 and Lys48 of Ubiquitin. *Biochemistry*, **53**, 4004-4014, (2014).

30. Hershko, A., Heller, H., Eytan, E., Kaklij, G., Rose, I.A. Role of the α-amino group of protein in ubiquitin-mediated protein breakdown. *Proc. Natl. Acad Sci*. USA. **81**, 7021-7025 (1984).

31. 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).

32. Ciechanover, A. and Ben-Saadon, R. N-terminal ubiquitination: more protein substrates join in. *Trends Cell Biol*. **14**, 103-106 Review (2004).

33. Ishikura, S., Weissman, A.M., Bonifacino, J.S. Serine residues in the cytosolic tail of the T-cell antigen receptor alpha-chain mediate ubiquitination and endoplasmic reticulum-associated degradation of the unassembled protein. *J. Biol. Chem*. **285**, 23916–22394 (2010).

34. Tokarev, A.A., Munguia, J., Guatelli, J.C. Serine–threonine ubiquitination mediates downregulation of BST-2/tetherin and relief of restricted virion release by HIV-1 Vpu. *J. Virol*. **85**, 51–63 (2011).

35. Carvalho, A.F., Pinto, M.P., Grou, C.P., Alencastre, I.S., Fransen, M., Sa-Miranda, C., et al. Ubiquitination of mammalian Pex5p, the peroxisomal import receptor. *J. Biol. Chem*. **282**, 31267–31272 (2007).

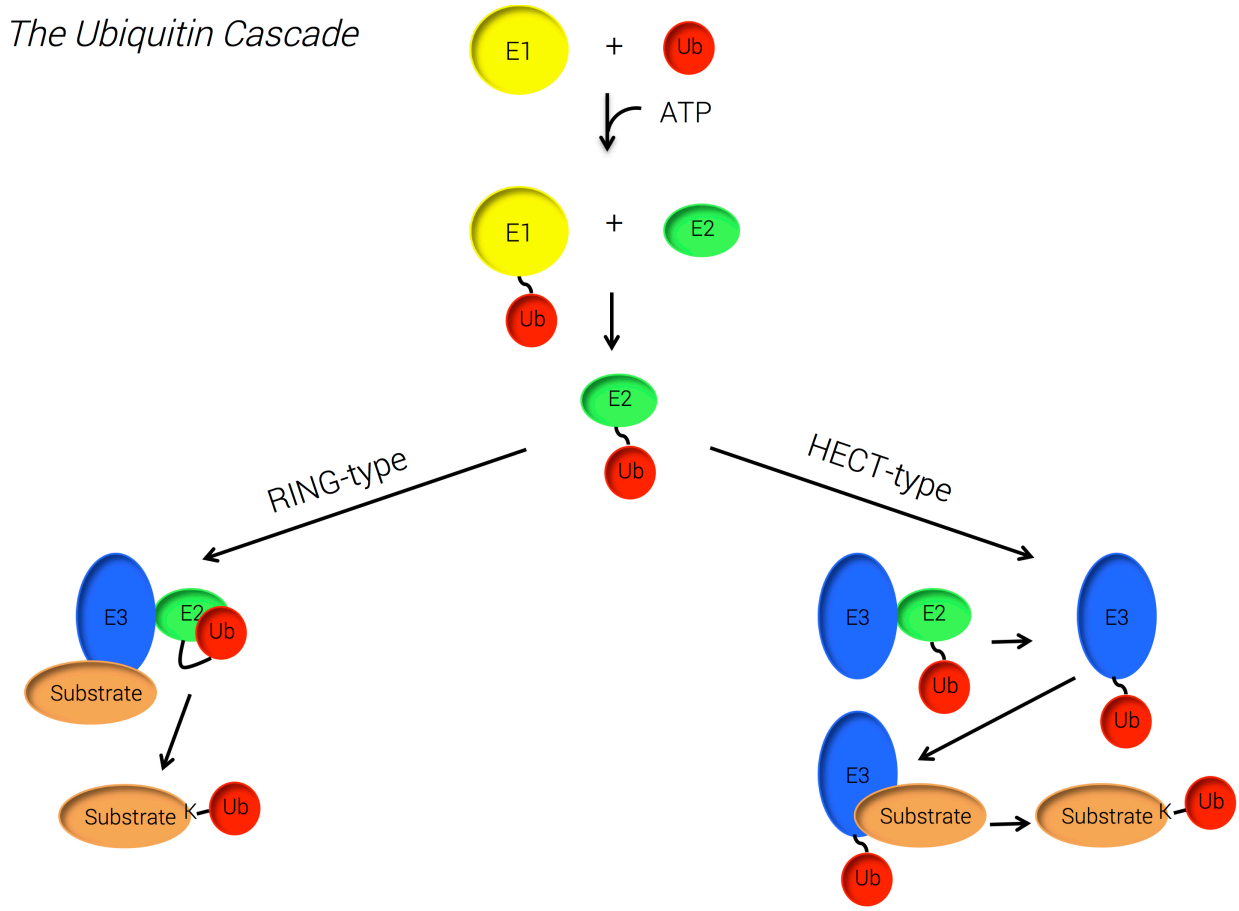


Figure 1. The ubiquitin cascade. The E1 first forms an E1~Ub conjugate at its active site cysteine in an ATP-dependent manner. The E1~Ub conjugate then engages an E2 and in a transthiolation reaction passes Ub to the active site cysteine of the E2. The newly formed E2~Ub conjugate then interacts with either a RING-type E3 or HECT-type E3 to facilitate the ubiquitination of the substrate. In the RING-type mechanism the E3 brings E2~Ub conjugates and substrate together and activates the E2 to transfer Ub to the substrate. In HECT-type mechanism a second transthiolation reaction occurs between the E2~Ub conjugate and the E3. The E3 then engages and ubiquitinates the substrate.

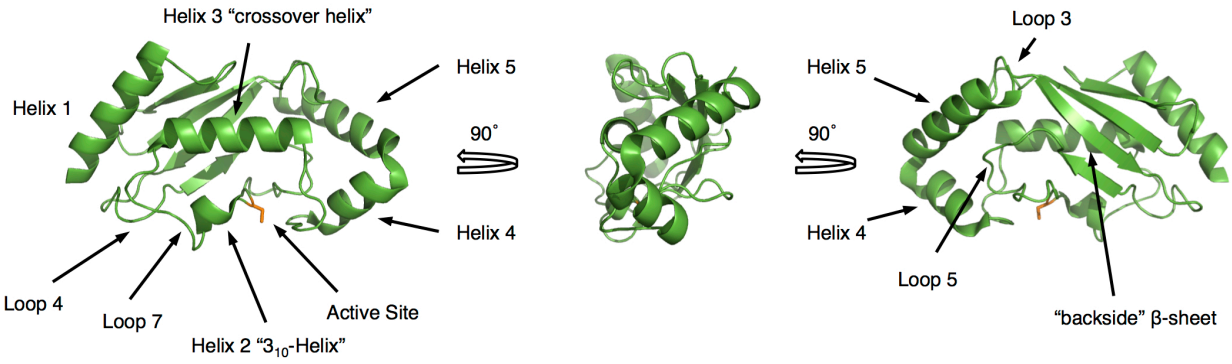


Figure 2. Architecture of an E2. ***Left,*** Structure of UbcH5c (PDB: **1X23**) facing helix-3 ("crossover helix") with structural nomenclature. ***Middle***, View of UbcH5c looking down helix-3. ***Right,*** View of UbcH5c from the "backside" β-sheet.

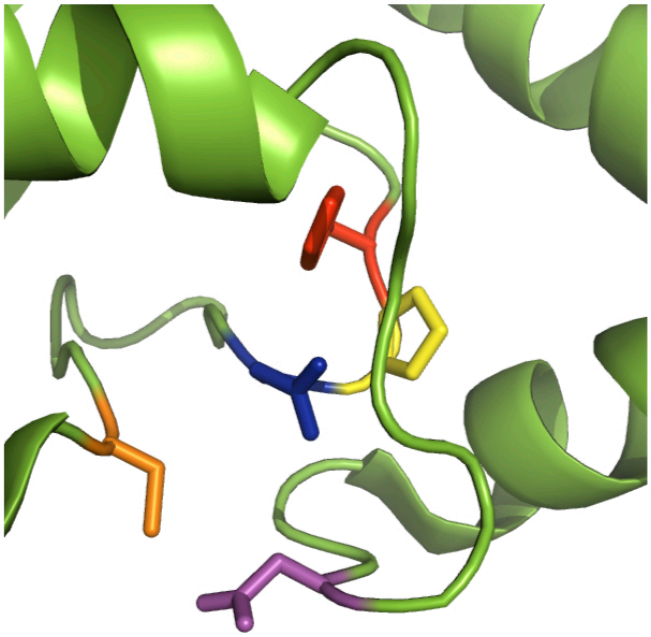


Figure 3. Active site architecture of an E2. The active site cysteine (orange) is flanked by an HPN (histidine (red), proline (yellow), asparagine (blue)) motif and an aspartate (purple).

Chapter II – Biochemical and structural characterization of the ubiquitin-conjugating enzyme Ube2w reveals formation of a noncovalent homodimer

The following chapter was published in the journal *Cell Biochemistry and Biophysics*

Reprinted with permission from

Vittal, V., Wenzel, D.M., Brzovic, P.S., Klevit, R.E. (2013) Biochemical and structural characterization of the ubiquitin-conjugating enzyme UBE2W reveals the formation of a noncovalent homodimer. *Cell Biochemistry and Biophysics*. 67(1), 103-110.

Author Contributions: V.V. performed all of the biochemical assays, NMR experiments, and generated the models. D.M.W assigned the original spectrum of Ube2w-KK and performed the initial characterization of Ube2w. V.V., P.S.B, and R.E.K conceived the experiments and wrote the manuscript. R.E.K. supervised the project.

Introduction

The attachment of ubiquitin (Ub) is a highly regulated cellular process that requires a cascade of three enzymes. The E1 Ub-activating enzyme first forms an ATP-dependent, high-energy thiolester bond with Ub through an active site cysteine. The E2 Ub-conjugating enzyme next receives Ub from the E1 in a transthiolation reaction forming an E2~Ub conjugate. Finally an E3 Ub-ligase binds E2~Ub using either a HECT domain or RING-finger motif and facilitates the transfer of Ub from the E2~Ub onto a target substrate[1]. In humans two E1s, about 35 E2s, and hundreds of E3s are responsible for attaching Ub to proteins[2]. For transfer that involves RING-type E3s, the E2 plays a role in dictating the type of Ub modification[3,4]. Depending on the required cellular cue, some substrates are modified by the addition of a single Ub molecule (mono-ubiquitination) while others are covalently attached to a chain of Ub molecules which can be linked via any of the 7 lysine residues in Ub or via its N-terminal end (poly-ubiquitination)`. At present, features and properties that determine different E2 function, especially as it relates to mono- versus poly-Ub transfer, are not well understood. Here we report biochemical and structural features of human UBE2W, a mono-ubiquitinating E2.

UBE2W is one of ten E2s that can bind to the BRCA1 RING domain of the BRCA1/BARD1 E3 complex[3]. *In vitro*, UBE2W adds only a single Ub to its target substrate, however, chain-building E2s such as the UBE2N/UBE2V2 (Ubc13/Mms2) complex can extend poly-ubiquitin chains from sites mono-ubiquitinated by UBE2W[3]. This E2 has been found to mono-ubiquitinate FANCD2, an essential step in the activation of the Fanconi anemia tumor suppressor pathway that is important in the DNA damage response[5]. More recently, Ataxin-3, a specialized deubiquitinating enzyme (DUB) important for protein quality control, has been shown to be mono-ubiquitinated by UBE2W[6]. In both cases, UBE2W can directly attach Ub to substrate lysines in the absence of an E3 in a yet unknown mechanism although ubiquitination of these substrates by UBE2W was enhanced by the presence of a RING-type E3 ligase[5,6].

To better understand the structure and function of UBE2W we began to characterize it biochemically. During size exclusion chromatography we unexpectedly observed that monomeric UBE2W is in equilibrium with a dimeric species. Here we show that homodimerization of UBE2W is driven by contacts involving its β-sheet surface, a region used by other E2s for non-covalent protein-protein interactions with Ub and non-RING E3 domains. Mutations that disrupt the dimer have only a modest

effect on the rate of Ub transfer, revealing that dimerization of UBE2W is not required for its ubiquitination activity. Surprisingly, the dimerization interface also involves residues in the C-terminal region of UBE2W. Based on the distance of the C-terminus from the β-sheet in other known E2 structures, we conclude the C-terminal region of UBE2W must adopt a non-canonical position to make this contribution. Furthermore, a C-terminally truncated version of UBE2W maintained dimeric properties similar to WT indicating the C-terminus is not required for dimerization.

Results

UBE2W exists in a monomer-dimer equilibrium.

In the process of purifying recombinant human UBE2W we observed properties that differ from other common, well-characterized ubiquitin conjugating enzymes. In particular, resolving 250uM UBE2W using a Superdex 75 size exclusion column resulted in the separation of two distinct peaks, corresponding in size to a dimeric and monomeric species (**Figure 1a**). The ability to resolve the two species by size exclusion chromatography (SEC) indicates that their interconversion is slow. SDS-PAGE analysis of the UBE2W dimer and monomer peaks under reducing (+β-mercaptoethanol) and non-reducing conditions indicated that dimerization of UBE2W is not dependent on disulfide bond formation, implying a noncovalent interaction (**Figure S1**). Re-application of the monomeric UBE2W fraction showed a re-equilibration into dimeric and monomeric forms (**Figure S2b**), and as predicted for a self-associating system, dimer formation displays a clear concentration dependence. Ten-fold and fifty-fold dilution of UBE2W prior to size exclusion chromatography shifts the equilibrium in favor of the monomeric species, consistent with an estimated affinity in the mid-micromolar range (**Figure S2**).

Other biochemically characterized, class I E2s do not show evidence for self-association at similarly high protein concentrations. For example, the well-studied E2 UBE2D3 (UbcH5c) elutes as a single peak at an elution volume consistent with its being a monomer (**Figure 1a**) as do UBE2L3 (UbcH7) and UBE2N (Ubc13) (data not shown). Thus, the observation that UBE2W forms a dimer is not due simply to our use of high protein concentrations (required for NMR spectroscopy, see later) but rather indicates that UBE2W differs from canonical class I E2s. Serendipitously, we found two residues (V30 and D67) during the course of our early studies with UBE2W and its activity with BRCA1/BARD1, that,

when mutated to lysine, disfavor the dimer and shift the equilibrium to a monomeric species (termed the UBE2W-KK mutant) (**Figure 1b**). This allowed us to 1) ask whether dimerization is required for UBE2W activity and 2) assign and compare resonances from the ^{1}H, ^{15}N-HSQC spectra of UBE2W to obtain insights regarding the structural source for dimer formation.

UBE2W transfers a single Ub to its target substrate resulting in a mono-ubiquitinated product[3,5]. In the *in vitro* assay shown, $BRCA1_{304}$/$BARD1_{327}$ functions as both the RING E3-ligase and as proxy substrate, with mono-ubiquitin products formed on $BRCA1_{304}$/$BARD1_{327}$ serving as a readout for activity. UBE2W-KK showed only a modest decrease in its ability to transfer Ub compared to wildtype, indicating dimerization of this E2 is not a strict requirement for Ub transfer activity (**Figure 2**).

Identification of the surfaces involved in UBE2W dimerization.

NMR spectroscopy was used to understand the structural consequences of the V30K/D67K double mutation and obtain insights regarding the nature of UBE2W self-association. For these experiments, we used ^{1}H, ^{15}N-HSQC NMR spectroscopy. This type of spectrum provides a distinct signal for every amino acid (except proline) that corresponds to that residue's backbone amide group. The precise position (i.e., the "chemical shift") and intensity of each peak is highly sensitive to the local environment and dynamics around the residue. Comparison of spectra for two related species allows easy identification of differences at an amino acid residue level. In this case, comparison of the ^{1}H, ^{15}N-HSQC spectra of UBE2W-WT and UBE2W-KK revealed some remarkable differences. Foremost among these, the 'KK' mutant spectrum contains roughly 120 peaks (of the 139 peaks expected) whereas the spectrum of the WT has fewer than 100 peaks (**Figure 3a**). As expected, three-dimensional NMR spectra of UBE2W-WT also contained many fewer resonances than did datasets of the monomeric mutant (unpublished data). The simplest explanation for missing peaks in the UBE2W-WT spectrum is that residues involved in the interface experience different environments in a monomer versus a dimer and are exchanging between these states on the timescale of NMR data collection. To identify residues affected by the mutation we assigned the spectrum of monomeric UBE2W-KK using conventional heteronuclear 3D-NMR approaches. We could identify most of the residues in the UBE2W-WT spectrum with high confidence by comparison with the 'KK' spectrum.

Differences observed between the ^{1}H, ^{15}N-HSQC spectra of UBE2W-KK and UBE2W-WT fell into two distinct classes: 1) signals whose positions are shifted between the WT and mutant spectra and 2) signals that appear in the mutant 'KK' spectrum that are not present in the WT (**Figure 3b**). It is generally expected that peaks arising from residues that are proximal to a site of mutation will have small chemical shift perturbations (CSPs) due to the change of side chain and, therefore, of the chemical environment due to the mutation. In the double 'KK' mutant, we expect peaks from residues near V30 or D67 in the E2 structure to be shifted (i.e., in the first class). CSPs are also expected to arise due to environmental changes in residues involved in the dimerization interface. For the second class, in the WT protein, some residues are interconverting between two states on a similar timescale to NMR data collection (~100msec); one exposed to solvent (monomer) and one buried in a protein interface (dimer). This behavior splits the intensity of a single resonance over two states causing broadening and a loss of intensity, often to a point below the level of detection. These "missing" or unobserved resonances reappear in the 'KK' spectrum because it exists predominantly as a monomer, resulting in accumulation of peak intensity at the resonance position that corresponds to the monomer.

Although a majority of human E2 structures have been determined experimentally, the only such structure for UBE2W is a crystal structure of a truncated form in which all C-terminal residues after S116 are missing (PDB:2A7L)[7]. The structure adopted in this crystal is unlike any other E2 structure solved to date. In the crystal lattice, a symmetry mate from one unit cell domain swaps its N-terminal helix with that of another UBE2W molecule in an adjacent unit cell (**Figure 4a**). This structure might imply that UBE2W dimers form using their N-terminal helices, but our NMR results are not consistent with such an inference as we do not detect perturbations in the N-terminal helix. Furthermore, because the C-terminus is absent in the crystal, it is impossible to visualize its potential role in dimerization in this context. Given the disparities between the crystal structure and our NMR observations we chose to map resonances that differ between the UBE2W-WT and 'KK' spectra onto a UBE2W sequence homology model based on the structure of UBE2D3 (PDB: 4DDG) (**Figure 5b**)[8]. This model is more consistent with our NMR data and is therefore suitable for mapping the observed NMR perturbations.

Substitution of V30 and D67 with lysine results in perturbations on two distinct surfaces on UBE2W. One surface is made of residues in Loops 4 and 7, the β-sheet ("backside") region, and the 3_{10}

helix ("surface 1" in **Figure 5**). The other surface corresponds to the C-terminal region of UBE2W ("surface 2" in **Figure 5**). Both surfaces have some resonances that are shifted from their WT position and some that are exchange broadened in the WT spectrum and appear in the 'KK' mutant. On surface 1, residues L25, D40, L56, and L57 in the β-sheet are shifted compared to the WT spectrum (**Figure 3b**). These resonances are shifted due to their proximity to the V30K mutation and/or because they are close to the dimerization interface. F60, however, is an example of a residue that is exchange broadened in the WT but appears in the 'KK' spectrum, and can therefore be assigned unambiguously as being located within the dimer interface (**Figure 3b**).

The sites of both mutations, V30 and D67 (blue), are near surface 1, implying that this region of the protein dictates dimerization of UBE2W. However, several C-terminal residues are also perturbed suggesting that residues in this region experience different environments in monomeric and dimeric states. To assess the possible contribution of the C-terminal region to self-association, a UBE2W construct truncated after residue Y131 (UBE2W-131Δ), was expressed and purified. This truncation removes the final 20 amino acids of UBE2W that are predicted to form the terminal helix and creates a construct of this E2 that is slightly longer than the previously crystallized structure [**7**]. UBE2W-131Δ is remarkably well behaved as evidenced by size exclusion chromatography and is stably folded as seen by NMR spectroscopy **(Figures 1c, 6**). Under the same conditions as UBE2W-WT, UBE2W-131Δ shows a similar monomer – dimer equilibrium, indicating that although the C-terminal region is affected by monomer-dimer exchange, it is not required for the formation of UBE2W homodimers (**Figure 1c**). Furthermore, the 'KK' mutation in the UBE2W-131Δ truncation background shifts the E2 population towards a monomeric state in the same manner as in the full-length protein (**Figure 1d**). Altogether, these results indicate that residues in surface 1 constitute the driving force for UBE2W dimerization.

Discussion

UBE2W displayed dimeric properties by SEC purification at the high concentration required for NMR studies (**Figure 1a**). This property proves problematic for detailed structural characterization by NMR spectroscopy, as multimeric proteins with mid-micromolar dissociation constants often suffer from loss of peak intensity if multiple interconverting states exist. We serendipitously discovered the monomer-

favoring ‘KK’ (V30K/D67K) double mutation in early studies aimed at creating an E2/E3 specificity-switch-pair. The mutant proved particularly useful for structural characterization by NMR spectroscopy, as the *in vitro* E2 activity was retained with a modest reduction at most (**Figure 2**).

Comparison of NMR spectra of ‘KK’ and WT UBE2W reveals two distinct surfaces that participate in homodimerization. We note that although the C-terminally truncated and domain-swapped UBE2W crystal structure lacks the residues that compose surface 2, the structure observed in the crystal does appear to use residues in surface 1 to pack neighboring UBE2W molecules. In Fig. 4b the surface 1 NMR perturbations are mapped onto the crystal structure, revealing reasonable concordance, although our surface implicates residues not entirely congruent with the current domain-swapped dimer interface. The role of the C-terminal region cannot be addressed by the crystal structure, as removal of these residues was apparently required to obtain crystals. Though the C-terminal region is not required for UBE2W dimerization, its NMR resonances are clearly perturbed upon disruption of the dimer interface (**Figure 5b**). In all experimentally determined structures of E2 UBC domains, the C-terminal helix is at least 15Å from the UBC β-sheet surface. It is therefore difficult to reconcile the observed effects in the C-terminus unless it adopts a non-canonical conformation and/or location in UBE2W to position it closer to the β-sheet and Loop 4 regions. We propose that UBE2W dimer formation through surface 1 induces ancillary contacts between surface 1 residues and C-terminal residues, which while not required for induction of the dimer, might stabilize the species once formed. Our observations are not consistent with an alternative possibility that UBE2W forms a dimer via heteromeric interactions in which the C-terminus of one protomer interacts with surface 1 residues of the other and rather supports a model in which surface 1 - to - surface 1 contacts drive homodimerization

The nature of the monomer-favoring mutations as well as the identification of the dimeric interface provides insights regarding UBE2W self-association. V30 is a hydrophobic residue on the surface of the β-sheet region of UBE2W. On its own, the V30K mutation almost completely disrupts formation of UBE2W homodimers as detected by SEC (data not shown), indicating that it is a key residue in the dimer interface. Substitution of a lysine for D67 is not as effective in disrupting UBE2W dimerization on its own but the double mutation further pushes the monomer – dimer equilibrium of UBE2W towards the monomeric state and reveals more resonances in the ^{1}H, ^{15}N-HSQC spectrum. Intriguingly, D67 is in

Loop 4, a part of the canonical E2 – E3 binding interface. This position is strongly conserved as lysine in many E2s, which is why we originally designed the D67K-UBE2W mutant. In our homology model of UBE2W, V30 rests on a turn immediately following the first β-strand in the UBC domain; the analogous position is conserved as valine in all isoforms of UBE2D (i.e., V26) as well as numerous other E2s. The surface created by the β-sheet consists of few charged residues and notably, an analogous surface on UBE2D3 interacts noncovalently with Ub, enabling formation of large multimeric assemblies of UBE2D3~Ub conjugates[9-11]. Introduction of a single charged residue into the UBE2D3 surface (S22R) is sufficient to abrogate the non-covalent E2-Ub interaction. Substitution of a positively charged lysine in place of V30 in UBE2W can disrupt the E2-E2 interaction similar to the S22R mutation in UBE2D3. Other E2s use their UBC β-sheet region in noncovalent interactions. For example, the E2s UBE2G2 and Rad6 bind non-RING portions of the E3 ligases, gp78 and Rad18, respectively, in positions very similar to the surface used by UBE2D3 to interact noncovalently with Ub[12-14]. Finally, we note that surface 1 utilized by UBE2W for self-association is closer to the active site region than the surfaces used by other E2s that use their UBC β-sheet region to form noncovalent interactions[9-14]. In this respect the UBE2W surface more closely resembles the surface used by UBE2N to interact with the E2-like protein UBE2V2 (Mms2), although this heterodimeric interaction is significantly stronger than the UBE2W homomeric E2-E2 interaction[15,16].

Dimerization of E2s *in vitro* has been reported sporadically in ubiquitination literature for more than twenty years[17-20]. Studies have also shown that crosslinking reagents enable the formation of E2 homodimers and increase ubiquitination activity, presumably by raising the local concentration of E2~Ub conjugates[21,22]. Among the putative dimer-forming E2s, the best characterized is the *S. cerevisiae* E2, Cdc34[20]. Dimerization of Cdc34~Ub conjugates has been shown to be required for E3-independent poly-ubiquitin chain formation by this E2[23]. Cdc34 also participates in E3-dependent poly-ubiquitination reactions mediated by the Skp1-CUL1-F-box (SCF) E3 ligase[24]. In this case, a model was proposed in which high local concentrations of Cdc34~Ub conjugates form dimers that are recruited to the Rbx1/Roc1 (RING) domain of the SCF complex to poly-ubiquitinate the *S. cerevisiae* substrate Sic1 *in vitro*[25]. Currently, there is no concrete evidence that E2 dimerization is a requirement for activity *in vivo*.

Cellular concentrations of UBE2W are currently unknown, however, given a K_d estimated to be in the mid-micromolar range, a small fraction of UBE2W is in the dimeric state even at sub-micromolar concentrations. Dimerization is not a requirement for activity, but may play a role in E3 recruitment or in determining E3-independent substrate specificity. It seems unlikely that dimerization of UBE2W functions similarly to *S. cerevisiae* Cdc34 so as to raise the local concentration of E2 at sites of ubiquitination, as the benefit of this property for an E2 that mono-ubiquitinates its substrates is unclear. However, weak protein-protein interactions are a hallmark of the ubiquitination machinery in the cell and regulate critical aspects of Ub transfer and signaling. With this in mind, the identified surface could play a role in mediating an as yet unknown protein-protein interaction that is important for UBE2W function in cells. The results presented here also suggest the C-terminus of UBE2W adopts a non-canonical position in UBE2W, placing it close to the β-sheet of the UBC domain. This result makes UBE2W unique amongst other characterized class I E2s. Although a functional significance for UBE2W self-association remains to be determined, there are clearly structural aspects of this protein that point to unique function and properties, the details of which await further elucidation.

Materials and Methods

Plasmids, Protein Expression, and Purification

All constructs of UBE2W (WT, KK, 131Δ, 131Δ-KK) were expressed from the pET24 vector without affinity tags. UBE2W constructs were transformed into *Escherichia coli* (BL21 DE3) cells and protein expression was induced with 0.4mM isopropyl-β-D-thio-galactoside (IPTG) at OD_{600} of 0.6, followed by growth at 16°C for 16hrs. Cells were lysed by french press in 25mM sodium phosphate (pH 7.0), 1mM EDTA for full length constructs, or 25mM 2-(*N*-morpholino)ethanesulfonic acid (pH 6.0), 1mM EDTA for 131Δ constructs. Following centrifugal clarification, the lysate was subjected to cation exchange chromatography with an elution gradient of 0-0.5M NaCl. E2-rich fractions were identified by UV absorbance, pooled, and further purified by size exclusion chromatography (SEC) on a Superdex 75 (GE Healthcare) column in 25mM sodium phosphate (pH 7.0), 150mM NaCl, the buffer used for all NMR and SEC experiments. Wheat Uba1, human Ub, and human FLAG- $BRCA1_{304}$/$BARD1_{327}$ used for ubiquitination activity assays were purified as previously described[3,26]. SEC for Figure 1 and

Supplementary Figure 2 was performed on purified samples using a Superdex 75 column (GE Healthcare). 250uL of 250uM, 25uM, or 5uM protein was applied to the column.

Autobiquitination Activity Assay

Reactions for BRCA1-directed ubiquitination assays contained 20uM Ub, 0.5uM wheat Uba1, 2uM UBE2W-WT or UBE2W-KK, 2uM FLAG-BRCA1$_{304}$/BARD1$_{327}$ (residues 1-304 or 1-327, respectively), and 10mM $MgCl_2$. Reactions were initiated by the addition of 5mM ATP and incubated at 37°C. Samples were collected at 0 and 60 min after the addition of ATP. Products were visualized by western blotting for FLAG-BRCA1.

NMR Spectroscopy

NMR spectra were collected on a 500MHz Bruker Avance II (University of Washington). Spectra were recorded at 25°C in 25mM sodium phosphate (pH 7.0), 150mM NaCl, 10% D_2O. All datasets were collected using ^{15}N-HSQC-TROSY experiments with 200uM ^{15}N-isotopically labeled protein. Data were processed using NMRPipe/NMRDraw[27] and visualized with NMRView[28]. Chemical shift perturbations were calculated using the formula $\Delta\delta j = [(\Delta\delta j\ ^{15}N/5)^2 + (\Delta\delta j\ ^{1}H)^2]^{1/2}$.

UBE2W Homology Model

The homology model for UBE2W was created using the secondary structure prediction algorithim PHYRE2[29]. The model is based on the structure of Ub-conjugated UBE2D3 in complex with the human deubiquitinating enzyme OTUB1 (PDB: 4DDG)[8].

Structure Visualization

All structural figures were created using the Pymol software.

Acknowledgments

We acknowledge D. Christensen and C. Eakin for their initial observations on the UBE2W-KK mutant. This work was supported by National Institute of General Medical Sciences grants R01 GM088055 (R.E.K.). V. V. was supported in part by the Hurd Fellowship Fund and PHS NRSA 2T32 GM007270.

References

1. Pickart, CM (2001) Mechanisms underlying ubiquitination. Annu Rev Biochem. 2001. 70, 195-201.

2. Ye, Y, and Rape, M (2009) Building ubiquitin chains: E2 enzymes at work. Nat Rev Mol Cell Biol. 2009. 10, 755-764.

3. Christensen, DE, Brzovic, PS, Klevit, RE (2007) E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. Nat Struct Mol Biol. 2007. 14, 941-948.

4. Rodrigo-Brenni, MC, Foster, SA, Morgan, DO (2010) Catalysis of lysine 48-specific ubiquitin chain assembly by residues in E2 and ubiquitin. Mol Cell. 2010. 39, 548-559.

5. Alpi, AF, Pace, PE, Babu, MM, Patel, KJ (2008) Mechanistic Insight into Site-Restricted Monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. Mol Cell. 2008. 32, 767-777.

6. Scaglione, KM, Zavodszky, E, Todi, SV, Patury, S, Xu, P, Rodríguez-Lebrón, E, et al (2011) Ube2w and Ataxin-3 Coordinately Regulate the Ubiquitin Ligase CHIP. Mol Cell. 2011. 43, 599-612.

7. Sheng, Y, Hong, JH, Doherty, R, Srikumar, T Shloush, J, Avvakumov, GV, et al (2012) A Human Ubiquitin Conjugating Enzyme (E2)-HECT E3 Ligase Structure-function Screen. Mol Cell Proteomics. 2012. 11, 329-341.

8. Juan, YC, Landry, MC, Sanches, M, Vittal, V, Leung, CC, Ceccarelli, DF, et al (2012) OTUB1 co-opts Lys48-linked ubiquitin recognition to suppress E2 enzyme function. Mol Cell. 2012. 45, 384-397.

9. Brzovic, PS, Lissounov A, Christensen DE, Hoyt DW, Klevit RE (2006) A UbcH5c/ubiquitin noncovalent complex is required for processive BRCA1-direct ubiquitination. Mol Cell. 2006. 21, 873-880.

10. Sakata E, Satoh T, Yamamoto S, Yamaguchi Y, Yagi-Utsumi M, Tanaka K, et al (2010) Crystal structure of UbcH5b~ubiquitin intermediate: insight into the formation of the self-assembled E2~Ub conjugates. Structure. 2010. 18, 138-147.

11. Page, RC, Pruneda, JN, Amick, J, Klevit, RE, Misra, S (2012) Structural insights into the conformation and oligomerization of E2~ubiquitin conjugates. Biochemistry. 2012. 51, 4175-4187.

12. Das R, Mariano J, Tsai, YC, Kalathur, RC, Kostova, Z, Li, J, et al (2009) Allosteric activation of E2-RING finger-mediate ubiquityation by a structurally defined specific E2-binding region of gp78. Mol Cell. 2009. 34, 674-685.

13. Li, W, Tu, D, Li, L, Wollert T, Ghirlando, R, Brunger, et al (2009) Mechanistic insights into active site-associated polyubiquitination by the ubiquitin-conjugating enzyme Ube2g2. Proc Natl Acad Sci USA.. 2009. 106, 3722-3727.

14. Hibbert, RG, Huang, A, Boelens, R, Sixma, TK (2011) E3 ligase Rad18 promotes monubiquitination rather than ubiquitin chain formation by E2 enzyme Rad6. Proc Natl Acad Sci USA. 2011. 108, 5590-5595.

15. VanDermark, AP, Hofmann, RM, Tsui, C, Pickart, CM, Wolberger, C (2001) Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. Cell. 2001. 105, 711-720.

16. Moraes, TF, Edwards, RA, McKenna, S, Pastushok, L, Xiao, W, Glover, JN, et al (2001) Crystal Structure of the human ubiquitin conjugating enzyme complex, hMms2-hUbc13. Nat Struct Biol. 2001. 8, 669-673.

17. Pickart, CE, Rose, IA (1985) Functional heterogeneity of ubiquitin carrier proteins. J Biol Chem. 1985. 260, 1573-1581.

18. Silver, ET, Gwozd, TJ, Ptak, C, Goebl, M, Ellison MJ (1992) A chimeric ubiquitin conjugating enzyme that combins the cell cycle properties of CDC34 (UBC3) and the DNA repair properties of RAD6 (UBC2): implications for the structure, function and evolution of the E2s. EMBO J. 1992. 11, 3091-3098.

19. Girod, PA, Vierstra, RD (1993) A major ubiquitin conjugation system in wheat germ extracts involves a 15-kDa ubiquitin-conjugating enzyme (E2) homologous to the years UBC4/UBC5 gene products. J Biol Chem. 1993. 268, 955-960.

20. Ptak C, Prendergast, JA, Hodgins, R, Kay, CM, Chau, V, Ellison, MJ (1994) Functional and physical characterization of the cell cycle ubiquitin-conjugating enzyme CDC34 (UBC3). Identification of a functional determinant within the tail that facilitates CDC34 self-association. J Biol Chem. 1994. 269, 26539-26545.

21. Haldeman, MT, Xia, G, Kasperek, EM, Pickart, CM (1997) Structure and function of ubiquitin conjugating enzyme E2-25K: the tail is a core-dependent activity element. Biochemistry. 1997. 36, 10526-10537.

22. Gazdoiu S, Yamoah, K, Wu, K, Escalante, CR, Tappin, I, Bermudez, V, et al (2005) Proximity-induced activation of human Cdc34 through heterologous dimerization. Proc Natl Acad Sci USA. 102, 2005. 15053-15058.

23. Varelas, X, Ptak, C, Ellison, MJ (2003) Cdc34 self-association is facilitated by ubiquitin thiolester formation and is required for its catalytic activity. Mol Cell Biol. 2003. 23, 5388-5400.

24. Skowyra, D, Graig, KL, Tyers, M, Elledge, SJ, Harper, JW (1997) F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. Cell. 1997 91, 209-219.

25. Deffenbaugh, AE, Scaglione, KM, Zhang, L, Moore, JM, Buranda, T, Sklar, LA, et al. (2003) Release of ubiquitin-charged Cdc34-S – Ub from the RING domain is essential for ubiquitination of the SCF(Cdc4)-bound substrate Sic1. Cell. 2003, 114, 611-622.

26. Pickart, CM, Raasi, S (2005) Controlled synthesis of polyubiquitin chains. Methods Enyzmol. 2005. 399, 21-36.

27. Delaglio, F, Grzesiek, S, Vuister, GW, Zhu, G, Pfeifer, J, Bax, A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995. 6, 277-293.

28. Johnson, BA, Blevins, RA (1994) NMR View: a computer program for the visualization and analysis of NMR data. J Biomol NMR. 1994. 4, 603-614.

29. Kelley, LA, Sternberg, MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. Nat Protoc. 2009. 4, 363-371.

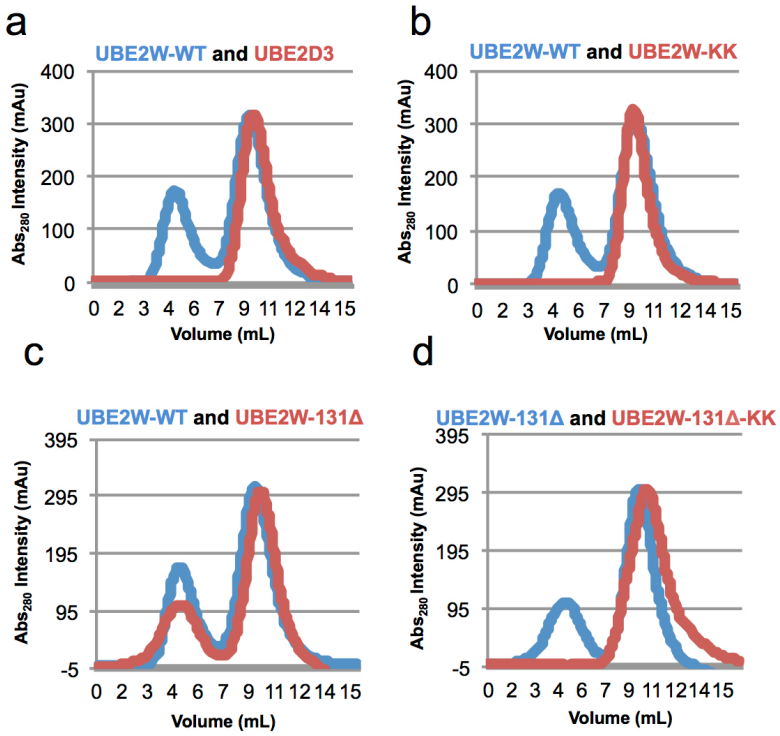


Figure 1. UBE2W exists in a monomer-dimer equilibrium. **(a)** Size exclusion elution profiles for UBE2W-WT (blue) and UBE2D3 (red) detect both a monomer and dimer for UBE2W, but only a monomer of UBE2D3. **(b)** Mutation of V30 and D67 to lysine in UBE2W (UBE2W-KK) (red) shifts the population of E2 towards a single monomeric peak as compared to UBE2W-WT (blue) **(c)** A C-terminal truncation mutant (UBE2W-131Δ) (red) elutes as both a monomer and dimer, similar to UBE2W-WT (blue). **(d)** The 'KK' mutation in the UBE2W-131Δ (UBE2W-131Δ-KK) (red) truncation background shifts the E2 population to a single monomeric peak when compared to the UBE2W-131Δ (blue). All size exclusion injections were performed on a column with a relative void volume of -5mL.

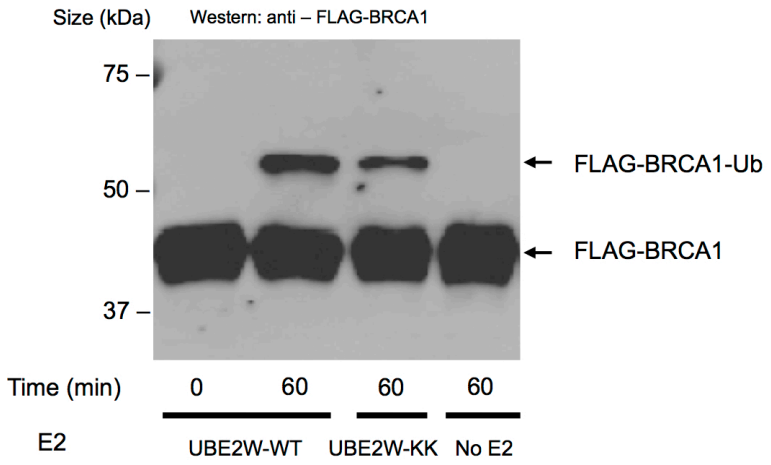


Figure 2. Disruption of the UBE2W dimer does not abrogate its ubiquitin transfer activity. Autoubiquitination of BRCA1 by UBE2W-WT and UBE2W-KK (1 hr reaction) is detected in a western blot for FLAG-BRCA1.

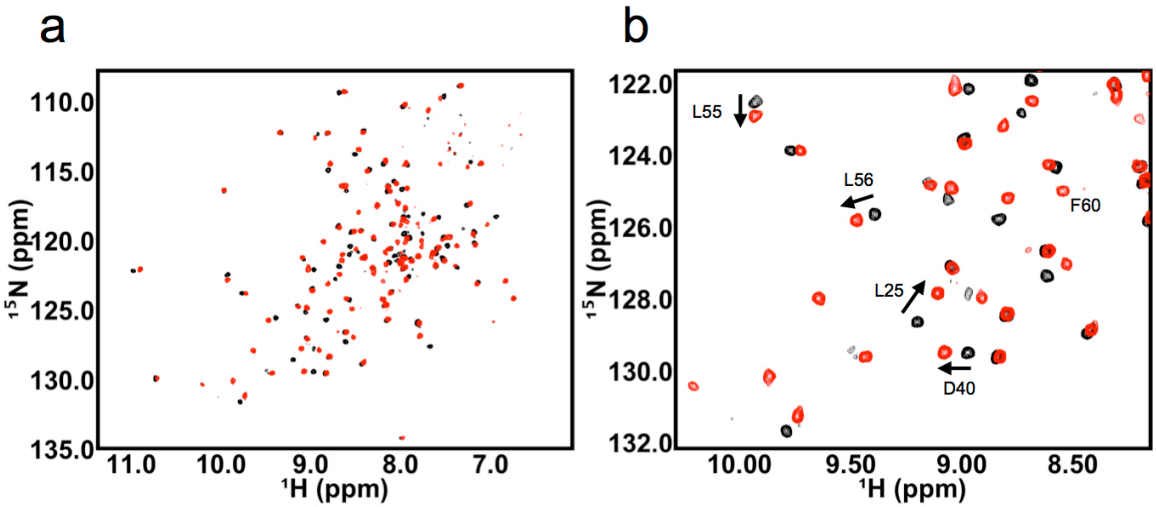


Figure 3. Comparison of the NMR spectra of UBE2W-WT and UEB2W-KK. **(a)** Full ^{1}H, ^{15}N-HSQC spectrum of UBE2W-WT (black) overlayed with the ^{1}H, ^{15}N-HSQC spectrum of UBE2W-KK (red). **(b)** Region of ^{1}H, ^{15}N-HSQC spectrum showing representative resonances shifted between UBE2W-WT (black) and UBE2W-KK (red) (i.e., L25, D40, L55, L56) and new resonances in the UBE2W-KK spectrum (i.e., F60).

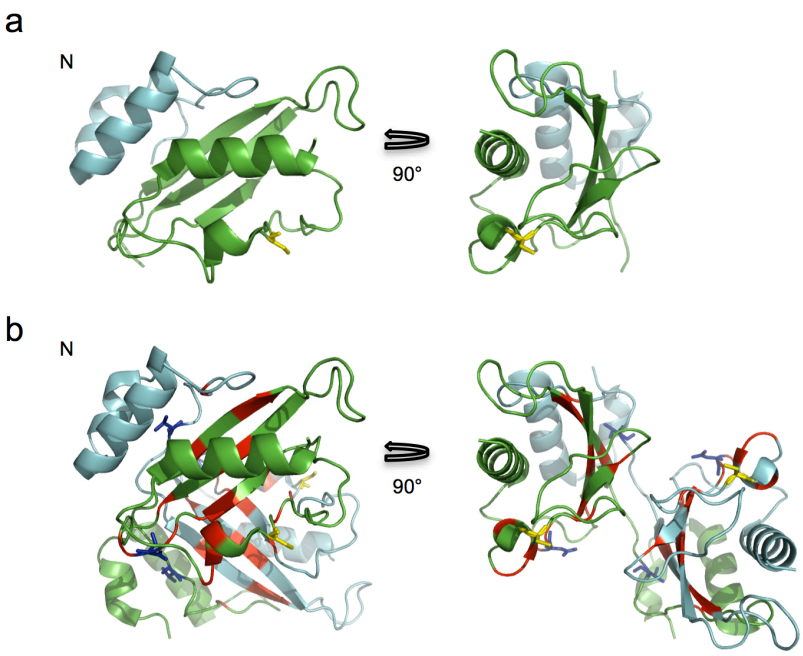


Figure 4. Surface 1 on UBE2W is involved in crystal contacts. **(a)** Domain-swapped UBE2W (truncated at S116) crystal structure (2A7L). The N-terminal helix from one UBE2W subunit (green) non-covalently interacts with the UBC domain of another subunit (cyan), and vice versus. Only one pair of the dimer is shown for clarity. **(b)** Surface 1 resonances mapped onto the dimer-swapped structure (red). In blue are positions V30 and D67, and in yellow is the active site cysteine (C91).

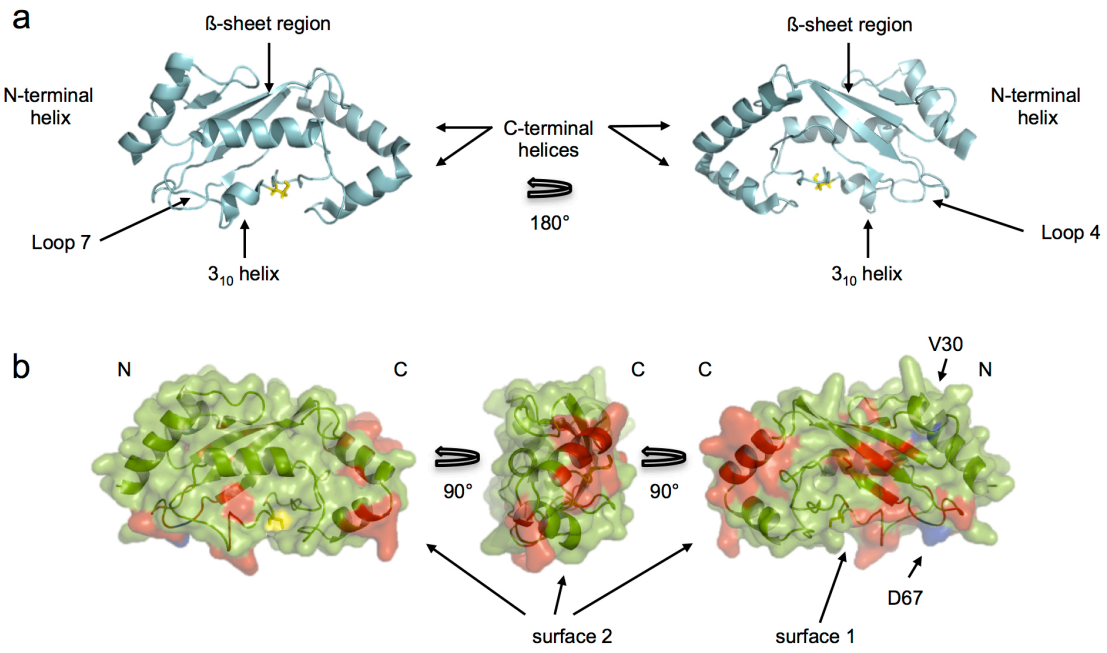


Figure 5. Domain architecture and nomenclature of an E2 (a) Structure of UBE2D3, the active site C85 is colored in yellow (PDB:4DDG)**. (b)** Observed NMR perturbations identify two surfaces on UBE2W. UBE2W homology model based on the structure of UBE2D3 (PDB:2FUH) is shown. Positions V30 and D67 (mutated to lysine in the 'KK' mutant background) are colored in blue, while red represents both significantly shifted (top 10% of shifted resonances) resonances between WT and 'KK' spectra and new peaks that appear in the 'KK' spectrum. Surface 1 includes Loops 4, and 7, the β-sheet region, and the 3_{10} helix. Surface 2 is comprised entirely of the C-terminal region. The active site cysteine (C91) is colored yellow. For perturbed residues, only those resonances with unambiguous assignments are colored red.

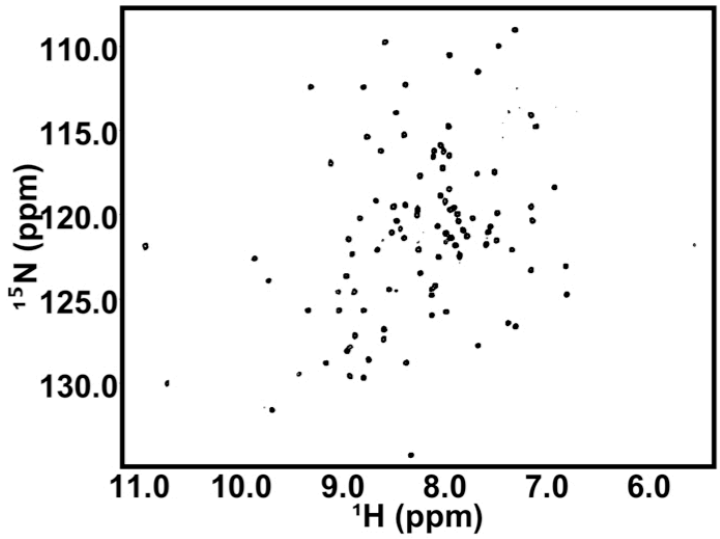


Figure 6. UBE2W lacking its C-terminal region is stable and retains its ability to form a dimer. Dispersion of resonances in the spectrum reveals a folded protein.

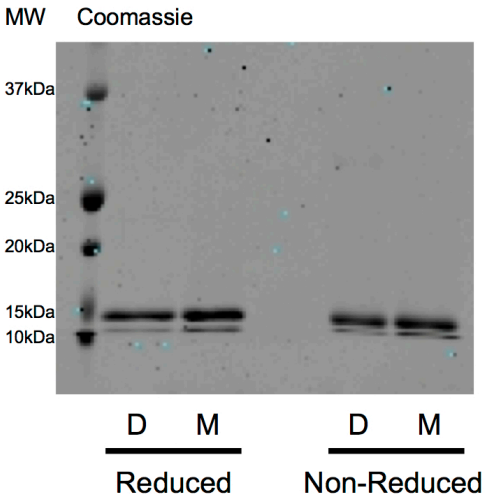


Figure S1. Coomassie stained gel shows dimer and monomer peaks from UBE2W-WT Size exclusion chromatography under reducing (+ β-mercaptoethanol) and non-reducing (- β-mercaptoethanol) conditions for dimeric (D) and monomeric (M) peaks.

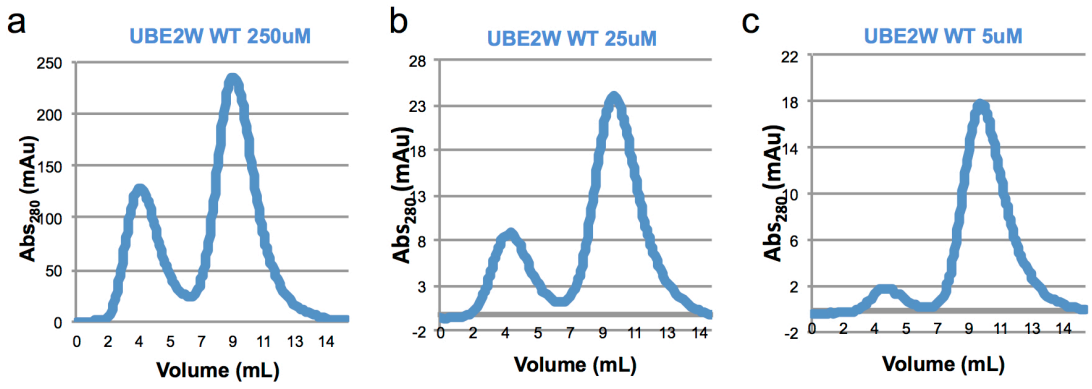


Figure S2. Size exclusion chromatography of UBE2W at decreasing concentrations.
Decreased concentrations shifts the ratio of dimer/monomer towards monomeric conformations.

Chapter III – Intrinsic disorder drives N-terminal ubiquitination by Ube2w

The following chapter was published in the journal *Nature Chemical Biology*

Vittal, V., Shi, L., Wenzel, D.M., Scalgione, K.M., Duncan, E.D., Basrur, V., Elenitoba-Johnson, K.S., Baker, D., Paulson, H.L., Brzovic, P.S., Klevit, R.E. (2015) Intrinsic disorder drives N-terminal ubiquitination by Ube2w. *Nature Chemical Biology*. 11(1), 83-9.

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.

Introduction

The attachment of ubiquitin (Ub) to cellular proteins is a highly regulated process that requires three enzyme activities. First, an E1 ubiquitin-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 ubiquitin-conjugating enzyme forming an E2~Ub conjugate. Third, E2~Ub interacts with an E3 ubiquitin ligase to modify protein targets via a RING-type, HECT-type, or RING-between-RING-type (RBR-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[1]. Thus, in RING-type mechanisms, the E2 plays 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-ubiquitin 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)[2] or Ube2k (K48-linked chains)[3], while others such as UbcH5c are promiscuous and can build Ub chains of multiple linkages [4]. Some E2s such as Ube2e1 and Ube2t add only a single Ub to their target substrate[5,6]. There is some evidence that certain E2s may transfer Ub to non-canonical amino acids such as serine, threonine, and cysteine[7,8]. The E2 Ube2w was recently reported to attach mono-Ub to the aN-terminus of substrates rather than to the εNH_2 side chain group of lysine residues[9,10].

Here we show that Ube2w specifically mono-ubiquitinates the αN-terminus of diverse substrates by recognizing backbone atoms of disordered N-termini. The solution ensemble of Ube2w reveals a novel UBC ("Ubiquitin Conjugating") domain architecture (**Figure S1**). 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 twenty C-terminal residues or a single point mutation within this region abrogates Ube2w ubiquitin transfer activity and impacts recognition and binding of multiple substrates. Furthermore, N-terminal substrate recognition and subsequent Ub transfer catalyzed by Ube2w are intimately dependent on the non-canonical 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[11]. To identify the E2 – BRCA1/BARD1 pair(s) that can ubiquitinate RPB8, *in vitro* ubiquitination assays were performed using the minimal RING heterodimer of BRCA1/BARD1 (BC_{112}/BD_{115}), and E2s that had previously been shown to interact with BRCA1/BARD1: Ube2w, UbcH5c, UbcH7, and Ube2e1[12]. Although RPB8 contains eight lysines, only Ube2w modifies RPB8 with Ub in the presence of BC_{112}/BD_{115} (**Figure 1, Figure S2a**). Ube2w also exhibits E3-independent modification, though with substantially lower activity (**Figure S2a**, **Figure S3**). Mass spectrometry analysis of the mono-ubiquitinated RPB8 product confirmed that the Ub is attached to RPB8's αN-terminus (**Figure S4**). 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[10,12,13].

A direct measure of the intrinsic aminolysis activity of an E2~Ub conjugate is its reactivity towards the εNH_2 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[14]. As shown in Fig. 1b (**left panel**), 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_2-Ala-Gly-Gly-Ser-Tyr-COO^-) **(Figure 1b left, Figures S2b,S2c,S5**). In contrast, UbcH5c~Ub reacted completely with free lysine but did not react with the peptide substrate (**Figure 1b right**). Thus, Ube2w's intrinsic aminolysis reactivity is limited to aNH_2 groups in the context of a polypeptide, as it did not transfer Ub to the αNH_2 group of a free amino acid, while UbcH5's intrinsic aminolysis reactivity is limited to εNH_2 groups of lysine residues.

In addition to RPB8, full-length CHIP (Carboxy terminus of HSP70-interacting protein), the minimal U-box domain of CHIP, small ubiquitin-like modifier 2 (SUMO-2), tau, ataxin-3, FANCL, FANCD2, and Ube2w have been reported to be ubiquitinated by Ube2w[9,10,15,16]. In all cases, a single mono-Ub is attached, implying 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-tag (HA-tag) is a robust Ube2w substrate *in vitro*. Unlike with WT-Ub, Ube2w adds multiple HA-Ubs onto RPB8 and this activity is retained with a lysine-less version, HA-Ub(K0), consistent with each attachment occurring through the aNH_2 group (**Figure 1c, Figure S2d**). 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 (**Figure 1d, Figure S2e)**. $\{^{1}H - ^{15}N\}$ NOE (hetNOE) values of HA-Ub, which are sensitive to high frequency motions on the pico to nano-second timescales, confirm that the additional residues in the HA tag are highly mobile as shown by their small and/or negative hetNOE values (**Figure 2a, Figure S6**). In contrast, hetNOE values for Ub residues 1-72 show that the N-terminus of WT-Ub is well ordered[17] as expected from Ub crystal structures (**Figure 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 currently identified targets of Ube2w-dependent ubiquitination (see above) have or are predicted to have a disordered segment at their N-termini[18-20] (**Figure S7)**.

Ube2w recognizes the backbone atoms of its substrates

Diverse N-terminal sequences in protein substrates suggest two mechanistic possibilities: 1) Ube2w preferentially recognizes backbone atoms over amino acid side chains, and 2) regular N-terminal secondary structure elements (a-helices and b-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_3-Ub) results in modest activity, and additional glycine residues (Met-Gly_5- or Met-Gly_7-Ub) result in a further increase in Ube2w activity (**Figure 2c, Figure S2f**). 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_3-Gly_5-Ub) is incapable of forming large poly-Ub chains and forms similar products to WT-Ub indicating amide groups at positions two through four are necessary for Ube2w-dependent N-terminal ubiquitination (**Figure 2e, Figure S2g).** Furthermore, a methionine at position 1 is not necessary for N-terminal ubiquitination by Ube2w (**Table S1)**. 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 likely has differing requirements for an α-amino group as opposed to the ε-amino group of a lysine side chain[21]. The pK_a of an N-terminal amino group is 7.7 ± 0.5 whereas lysine side chains have pK_a values around 10.5 ± 1.1[22]. Thus, at physiological pH, a larger proportion of αN-termini will be deprotonated, bypassing a need to deprotonate the incoming nucleophile. The pK_a of an N-terminal amino group depends on the identity of the sidechain at position 1. However, our results indicate that Ube2w is capable of ubiquitinating N-termini over a wide pKa range (pKa = ~7.3 to ~9.1), indicating that the nucleophile's pKa is not the determining factor for Ube2w's special reactivity (**Figures S2h,S8**)

Ube2w has a non-canonical 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 (PDB: **2A7L**), it is for a truncation of Ube2w that lacks residues Ser117- Cys151[23]. A similar truncated version of Ube2w can form an activated thioester with Ub (**Figures S2i,S9**), but does not transfer Ub onto a substrate (data not shown) and thus lacks the structural features needed to understand Ube2w substrate selectivity. {^{1}H – ^{15}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 (**Figure 3a).** An identical experiment conducted on UbcH5c reveals its C-terminus undergoes motions consistent with the core of the protein (**Figure 3b**).

To understand the unique structural properties of Ube2w, further NMR experiments were conducted. NMR data were collected on a monomeric Ube2w mutant (V30K/D67K-Ube2w termed "KK")

that reduces self-association at high concentrations, but retains Ube2w activity[24]. We compared the {^{1}H, ^{15}N} – HSQC-TROSY spectra of full-length Ube2w-KK with a fragment that is missing the C-terminal residues following Tyr131 (Ube2w-131-KK). Ube2w resonances for residues S86, N87, and T96 (near the active site), in the 3_{10}-helix, and on the 'backside' b-sheet experience large chemical shift perturbations (CSPs) as a consequence of removal of the C-terminus (**Figure 3c left, Figures S10a,Sb**). These observations are surprising in that comparisons with canonical E2 structures, such as UbcH5c, predict a different set of CSPs would be observed, particularly with respect to the E2 active site and loops 3 and 5 (**Figure 3c right and Figure S1 for nomenclature**)[25].

Based on 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 crosspeaks (**Figure S11**). Therefore, we pursued an NMR-driven solution structure using alternative parameters. NMR chemical shifts (HN, N, C, Ca, Cb **)** for 137 residues, 109 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 utilizes chemical shift values and other experimentally-driven NMR restraints to generate solution ensembles[26,27] (**Tables S2,S3, Figures S10c,S12**). This initial computational stage produced 16,000 structures that were further filtered using two spin label positions (C91 and C135) and SAXS to generate the final ensemble.

An ensemble of the twenty lowest energy structures is shown in Fig. 4a. Ube2w has a well-defined core that closely resembles canonical UBC domains, such as UbcH5c (**Figure 4b**). The average pairwise RMSD for residues R7–S118 for the twenty members of the ensemble is 1.35 Å, indicating that the available experimental observations used are sufficient to define the structure. Furthermore, the twenty-member ensemble reveals favorable Ramachandran statistics (**Table S2**). The average pairwise RMSD for all backbone atoms across the Ube2w ensemble is 4.1 Å, consistent with {^{1}H – ^{15}N} hetNOE values that reveal a highly flexible C-terminus. Features conserved amongst canonical E2s in the Ube2w solution structure include helix-1, the 4-stranded 'backside' b-sheet, the structural architecture of the active site, and helix-3 ("crossover" helix) (**Figures 4a,b**).

The first seven residues of Ube2w are not observed in our NMR spectra, indicating they undergo conformational exchange, most likely a helix-to-coil transition. However, due to the component of CS-

Rosetta that utilizes 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, the spin label effects at residues C91 and C135, and SAXS.

In Ube2w, a long disordered loop following the crossover helix leads away from the b-sheet and is followed by a single helix formed by residues 127-135. This helix does not appear to adopt a unique position but is located within 7-19 Å of the active site, C91, in all members of the ensemble **(Figure 4c**). By contrast, in canonical UBC domains, a loop leads from the end of the crossover helix toward the protein core (near the 'backside' b-sheet) and is followed by two C-terminal helices, helix-4 (15.5 Å from active site) and helix-5 (**Figure 4b**). Three distinct C-terminal clusters are evident in the ensemble, in which helix-4 occupies positions facing 1) closer to the 3_{10}-10 helix, 2) adjacent to the active site, and 3) closer to the 'backside' b-sheet (**Figures 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 (**Figure 4e, Figure S13a**). The final six amino acids are completely disordered and are not constrained to a particular region, consistent with $\{^{1}H - ^{15}N\}$ hetNOE results.

The correlation between the experimental and back-calculated SAXS curves shows some disagreement (**Figure S12f**). Possible sources for the discrepancies are 1) presence of some aggregated material, 2) presence of low concentrations of dimeric Ube2w even with the dimer-disrupting mutation V30K/D67K, and 3) 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 non-canonical position for helix-4 that falls within our Ube2w ensemble and a disordered C-terminus (PDB: **3WE5**)[18]. The average RMSD 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 plays a predominant role in mediating interactions with substrates. First, Ube2w shows robust Ub transfer activity (**Figure 1a left panel**) in the presence of a minimal RING construct that offers no substrate binding functionality. Second, Ube2w adds Ub to the αN-terminus of substrates, in the absence of an E3, albeit at a slower rate (**Figure S3)**[9]. RING-type E3s enhance the activity of other E2s by promoting closed E2~Ub conformations that promote Ub transfer[21,28,29]. Our results also demonstrate that Ube2w~Ub is allosterically activated by RING-type E3s to form closed conformations (**Figure S2j,S13**). {^{1}H – ^{15}N} – HSQC-TROSY experiments capable of detecting low affinity binding interactions reveal peak broadening and chemical shift perturbations (CSPs) for a subset of ^{15}N-Ube2w-KK resonances upon addition of the substrates RPB8 and tau (**Figure 5a,b,c Figure S14a**). Based on 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 significantly 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 standard deviation above/below mean for intensity loss/shifting, respectively). Intriguingly, all residues in the unstructured region positioned directly beneath the active site in the Ube2w ensemble (K137-W145) are significantly perturbed (**Figure 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 (**Figure 5d,e,f, Figure S2k**). Finally, addition of tau, a Ube2w substrate with a different N-terminal sequence, perturbs a nearly identical set of residues in ^{15}N-Ube2w-KK as does RPB8, suggesting side-chain identity plays, at most, a minimal role in Ube2w substrate recognition (**Figure 5c, Figures S14b,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* displays significant conservation of C-terminal amino acids (R133, N136, K137, P139, W144, H147, D148, D149) (**Supplementary Fig. 15a**). Our NMR experiments show 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 (**Figure 4e, Figures S15a,b**), and shows significant perturbations upon substrate binding (**Figures 5b,c,e**). Substituting a glutamate (the corresponding residue in the UbcH5c sequence) for Ube2w Trp144 (W144E) completely abrogates Ube2w activity towards four different substrates: RBP8, ataxin-3, tau, and CHIP (**Figures 5f, Figures S16a-d**). The mutation has a two-fold effect. First, substrate binding is inhibited, evidenced by decreased peak broadening upon addition of RPB8 to the mutant protein in an NMR binding experiment (**Figures S16e**). Second, the W144E mutation generates NMR CSPs in Ube2w active site resonances that resemble those seen in the truncated Ube2w-131Δ mutant (**Figure 6a**). To probe the role of the Ube2w C-terminus in N-terminal ubiquitination function, we employed our intrinsic reactivity assay. Ube2w~Ub readily transfers its Ub to the minimal peptide substrate in a 1hr reaction but Ube2w-131Δ and Ube2w-W144E show almost no transfer to the peptide (**Figure 6b, Figure S2l**). The slow loss of the Ube2w-131Δ~Ub conjugate during the experiment is due to hydrolysis of Ub and not aminolysis with the substrate (**Figures S2m,S15c**). Thus, without an intact C-terminus, Ube2w does not transfer Ub to a free α-amino group.

Discussion

To date, Ube2w is the only E2 demonstrated to attach Ub directly and specifically to the N-terminus of proteins. While it shares certain mechanistic features with lysine-reactive E2s, our results indicate that Ube2w is uniquely adapted to facilitate selective α-amino ubiquitination. Similar to 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 sidechains have a conserved Asn residue (N77 in UbcH5) in the loop immediately preceding the active site that is proposed to play both catalytic and structural roles for transfer to substrate lysines. Asn77 is thought to stabilize the oxyanion intermediate formed following lysine attack of the E2~Ub thioester[30] and likely performs a structural role in helping to stabilize the loop preceding the

active site[31]. Notably, Ube2w contains a histidine at this position, H83. Mutation of H83 to Asn (H83N or H94N in human Ube2w isoform 2) results in a marked decrease in the E3-enhanced N-terminal ubiquitination activity of Ube2w[9], highlighting structural adaptations for N-terminal ubiquitination. Notably, Ube2w is unique among the ~40 human E2s by containing a histidine at this key position, strongly implicating it as the sole E2 with N-terminal ubiquitination activity.

Our work shows that Ube2w's non-canonical, flexible C-terminal structure provides a platform for recognition of diverse substrates. We propose a mechanism in which Ube2w recognizes and binds to backbone groups in substrates, explaining the observed requirement for N-terminal disorder in substrates. This mechanism is supported by observations that Ube2w can 1) transfer Ub to the αNH_2 group of a 5 amino acid peptide but not to the aNH_2 group of a free amino acid and 2) create linear Ub chains using Ub that has at least four unstructured residues (Met-Gly_3-Ub) appended to its N-terminus but not if those residues lack amide groups (Met-Pro_3-Gly_5-Ub). (**Figure 1b left, Figure 2e**). Poly-glycine 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 the sequences are either known to be or predicted to be disordered[18-20] (**Figure S7**). A mechanism that involves substrate backbone atoms would require those groups to be available; adoption of secondary structure such as α-helices and β-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 appears to be crucial for Ube2w aNH_2 ubiquitination function, as its removal (Ube2w-131Δ) or alteration (Ube2w-W144E) reduces substrate binding towards disordered N-termini, affects the environment of active site residues, decreases intrinsic aminolysis activity, and consequently, inhibits ubiquitination of substrates (**Figures 5d,f,6** and **Figure S16**). Based on 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, while specific chemical modification of the a-amino group through carbamylation[32] inhibited ubiquitination of MyoD[33]. Other substrates identified by similar strategies are human papillomavirus 16 oncoprotein, E7 (E7-16), latent membrane protein 1, and inhibitor of differentiation 2[34]. In each case, truncation of the N-terminal region inhibited ubiquitination and degradation implying a role for mobility/flexibility in the N-terminus[34]**.** 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[35]. Bulky tags such as a Myc_6 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_6-, had no such effect[35]. A crystal structure of ERK3 indicates that the first ~10 residues are disordered (PDB: **2I6L**). 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[18-20].

Identification of cellular Ube2w substrates is a critical step towards understanding the function of N-terminal ubiquitination. In eukaryotic cells, protein N-termini appear to be an important site of regulation. N-terminal acetylation is a prevalent post-translational modification that likely plays 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[36]. 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 play a role in the regulation of nuclear proteins[37]. Indeed, a majority of N-terminally-ubiquitinated proteins identified thus far are nuclear proteins. Another

intriguing possibility for Ube2w substrates include 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 Ub antibody enrichment methods that utilize the diGly-antibody do not recognize substrates ubiquitinated at their N-termini. N-terminal processing of proteins in cells means that an antibody derived towards 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 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 ubiquitin-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.

Materials and Methods

Plasmids, Protein Expression

Ube2w constructs (WT, V30K/D67K (KK), 131Δ, 131Δ-KK, W144E, W144E-KK L110Q, C91S, C91S-KK, 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 0.5mM iso-propyl-β -D-thio-galactoside (IPTG) at OD_{600} of 0.6, followed by growth for 16hr at 16˚C. Cells were lysed by French press in 25mM sodium phosphate (pH 7.0), 1mM EDTA for full-length constructs and 25mM 2-(*N*-morpholino) ethanesulfonic acid (pH 6.0), 1mM EDTA for 131Δ constructs. Following centrifugal clarification, Ube2w was applied to a cation exchange column with an elution gradient of 0-0.5M 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 25mM sodium phosphate (pH 7.0), 150mM NaCl. This buffer was used for all NMR experiements.

RPB8 (1-150) was expressed from the pET28a vector with a thrombin cleavable HIS_6-tag. RPB8 plasmids were transformed into Escherichia coli (BL21 DE3) cells and protein expression was induced with 0.5mM

iso-propyl-β -D-thio-galactoside (IPTG) at OD_{600} of 0.6, followed by growth for 16hr at 16˚C. Cells were lysed by French press in 25mM Tris-HCl (pH 7.6), 200mM NaCl, 10mM imidazole and applied to a Ni^{2+}-NTA gravity flow column (Invitrogen) equilibrated in the same buffer. The column was washed with 5 column volumes of 25mM Tris-HCl (pH 7.6), 200mM NaCl, 50mM imidazole and eluted with 25mM Tris-HCl (pH 7.6), 200mM NaCl, 500mM imidazole. The elution was subject to cleavage at 4˚C overnight in 25mM Tris-HCl (pH 7.6), 200mM NaCl with 1mg thrombin from bovine serum (Sigma-Aldrich). Following cleavage the dialyzed sample was re-applied to a Ni^{2+}-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 25mM sodium phosphate (pH 7.0), 150mM NaCl.

Wheat Uba1 and Ub were cloned, expressed, and purified as previously described[38]. Identical purification protocols for WT-Ub were followed for, Met-Gly-Ub, Met-Gly_3-Ub, Met-Gly_5-Ub, Met-Gly_7-Ub, Ub-I44A, HA-Ub and HA-Ub(K0).

BC_{112}/BD_{115} was expressed from the pET28N and pCOT7N expression systems (generous gift of Dr. M. Wittekind, Bristol-Myers Squibb). Proteins were purified as previously described[39].
E2's UbcH5c, Ube2e1, and UbcH7 were cloned, expressed, and purified as previously described[12].

Ataxin-3 was cloned, expressed and purified as previously described[40]. Tau protein was cloned, expressed, and purified as previously described[41].

Full length CHIP and CHIP U-box were cloned, expressed, and purified as previously described[42,43].

***In vitro* ubiquitination assays**

Ubiquitination assays involving RPB8 or HA-Ub were conducted at 37˚C for 1hr in 25mM sodium phosphate (pH 7.0), 150mM NaCl. Protein concentrations were as follows: 1uM Uba1, 4uM E2 enzyme, 4uM RPB8, 2uM $BC_{112}BD_{115}$, 30uM Ub or HA-Ub, and 10mM $MgCl_2$. Reactions were initiated by addition

of 5mM ATP. Samples were boiled in SDS buffer, loaded onto an SDS-PAGE, and visualized by western blotting with appropriate antibodies; anti-RPB8 (Abnova), anti-Ub (Santa Cruz).

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 (50mM Tris [pH 7.5], 50mM KCl, 0.2mM DTT), Ub^{mix} (2.5mM ATP, 5mM $MgCl_2$, 50nM Ube1, and 250µM ubiquitin), 1µM indicated E2, 1µM CHIP, and 1µM of indicated substrate. For CHIP independent ubiquitination reactions CHIP was omitted from the reaction, substrate concentration was increased to 20uM, and E2 concentration was 9uM 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 25mM sodium phosphate (pH 7.0), 150mM NaCl. Reactions contained 1uM Uba1, 20uM Ub, 20uM E2, 10mM $MgCl_2$. Reactions were initiated with 5mM ATP and allowed for form E2~Ub conjugates for 30min. Nucleophile reactivity was induced by addition of 50mM free lysine or 50mM 5aa peptide (NH_2-A-A-G-S-Y-COO^-) and allowed to react for 1hr. Samples were collected in non-reducing SDS buffer and loaded onto a SDS-PAGE gel. Results were visualized by coommasie staining. Peptides were purchased (United Biosystems Inc).

Protein Identification by LC-MS/MS

Protein identification and ubiquitination sites on RPB8 were conducted based on previously described protocols[9]. Briefly, protein bands corresponding ubiquitinated substrate (as indicated) 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 (Picofrit 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 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 25mM sodium phosphate (pH 7.0), 150mM NaCl using either 90% H_2O/D_2O or 100% D_2O. Samples for Ube2w-KK utilized either uniformly ^{15}N or ^{15}N, ^{13}C-labeled protein at concentrations from 400uM to 200uM. Titration experiments involving ^{15}N – Ube2w-KK were performed by equimolar addition of unlabeled RPB8 or tau. The magnitude of chemical shift perturbations for each resonance was quantified in Hz according to the equation $\Delta\delta_j=((\Delta\delta_j\ ^{15}N)^2 + (\Delta\delta_j\ ^{1}H)^2)^{1/2}$. Data collection for resonance assignments utilized standard three-dimensional NMR techniques[44] collected on INOVA 600 and 800 MHz spectrometers (Varian) at Pacific Northwest National Labs (PNNL). All other NMR-based experiments (substrate titrations, $\{^1H – ^{15}N\}$ hetNOEs, T_1,T_2 relaxation experiments, $\{^1H – ^{15}N\}$ RDC measurements, 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[45] and visualized with NMRView[46].

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 2hrs at 30˚C. Reaction yields were quantified by MALDI-MS. Only those that reacted to completion (>95%) were utilized for spin-label experiments. Unreacted probe was cleared by dialysis overnight at 4˚C. Identical $\{^1H –^{15}N\}$ – HSQC-TROSY experiments were conducted in the presence and absence of the probe reducing agent, ascorbate.

Residual Dipolar Couplings (RDCs)

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

Small-Angle X ray Scattering (SAXS)

SAXS data were collected at Stanford Synchrotron Radiation Lightsource beamline 4-2. Data were collected for Ube2w at concentrations of 10, 5, and 0.5 mg/mL in 25mM sodium phosphate (pH 7.0), 150mM NaCl, and 2mM DTT at 25˚C. SAXS statistics were calculated using the EMLB CRYSOL[47] 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[24,25,48]. Interatomic restraints are derived from alignments of Ube2w sequence to templates[49]. 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 Cβ 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 penalty term related to the homologous structural and experimental N-H RDC restraints. Backbone chemical shift perturbation 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, and 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 are used to rank the top models. Pearson correlation is calculated between experimental data and distance from paramagnetic center to HN atom. Distances were calculated by explicitly adding spin label to each model at residue 135 or 91. The top 20 models with 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)[50]. 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.

Q factor for RDCs reflects the agreement between calculated Q_{calc} and experimental Q_{obs} dipolar couplings as: $Q = \mathrm{rms}(D^{calc} - D^{obs})/\{D_a^2[4 + 3R^2]/5\}^{1/2}$, where D_a and R refer to the magnitude and rhombicity of the alignment tensor, respectively[51].

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[47].

Accession Codes

PDB. The solution ensemble of full-length Ube2w was deposited under the code 2MT6. The NMR assignments were deposited to the BMRB under the accession code 25150.

References:

1. Pickart C.M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 195-201 (2001).

2. Deng, L., et al. Activation of the IκB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103**, 351-361 (2000).

3. Chen, Z. and 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).

4. Brzovic, P.S. and Klevit, R.E. Ubiquitin Transfer from the E2 Perspective: why is UbcH5c so promiscuous? *Cell Cycle* **24**, 2867-2873 (2006).

5. 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).

6. Machida, Y.J., et al. UBE2T Is the E2 in the Fanconi Anemia Pathway and Undergoes Negative Autoregulation. *Mol. Cell.* **23**, 589-596 (2006).

7. 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).

8. Vosper, J.M., et al. Ubiquitylation on canonical and non-canonical sties targets the transcription factor neurogenin for ubiquitin-mediate proteolysis. *J. Biol. Chem*. **284**, 15458-15468 (2009).

9. Scaglione, K.M., et al. The ubiquitin-conjugating enzyme (E2) Ube2w ubiquitinates the N terminus of substrates. *J. Biol. Chem*. **288**, 18784-18788 (2013).

10. Tatham, M.H., Plechanovová, A., Jaffray, E.G., Salmen, H., Hay, R.T. Ube2W conjugates ubiquitin to α-amino groups of protein N-termini. *Biochem. J*. **453**, 137-145 (2013).

11. Wu, W., et al. BRCA1 ubiquitinates RPB8 in response to DNA damage. *Cancer Res*. **67**, 951-958 (2007).

12. 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).

13. 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**, 1-7 (2012).

14. 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).

15. Zhang, Y., et al. UBE2W interacts with FANCL and regulates the monoubiquitination of Fanconi anemia protein FANCD2. *Mol. Cells*. **31**, 113-122 (2011).

16. Alpi, A.F., Pace, P.E. ,Babu, M.M., Patel, K.J. Mechanistic Insight into Site-Resictricted Monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Mol. Cell*. **32**, 767-777 (2008).

17. Wand, A.J., Urbauer, J.L., McEvoy, R.P., Beiber, R.J. Internal dynamics of human ubiquitin revealed by 13C-relaxation studies of randomly fractionally labeled protein. *Biochemistry*. **35**, 6116-6125 (1996).

18. Berman, H.M., et al. The Protein Data Bank. *Nuc. Acids Res*. **28**, 235-242 (2000).

19. Kelley, L.A. and Sternberg, M.J.E. Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.* 4, 363-371 (2009).

20. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol*. **292**, 195-202 (1999).

21. 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).

22. 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).

23. Sheng, Y., et al. A human ubiquitin conjugating enzyme (E2)-HECT E3 ligase structure-function screen. *Mol Cell Proteomics*. **11**, 329-341 (2012).

24. 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).

25. Laskowski R.A. PDBsum new things. *Nuc. Acids. Res*. **37**, D355-D359 (2009).

26. Shen, Y. et al. Consistent blind protein structure generation from NMR chemical shit data. *Proc. Natl. Acad. Sci USA*. **105**, 4685-4690 (2008).

27. 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).

28. 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).

29. 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).

30. Wu, P.Y., et al. A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J*. **22**, 5241-5250 (2003).

31. 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**, 154-16 (2013).

32. Hershko, A., Heller, H., Eytan, E., Kaklij, G., Rose, I.A. Role of the α-amino group of protein in ubiquitin-mediated protein breakdown. *Proc. Natl. Acad Sci*. USA. **81**, 7021-7025 (1984).

33. 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).

34. Ciechanover, A. and Ben-Saadon, R. N-terminal ubiquitination: more protein substrates join in. *Trends Cell Biol*. **14**, 103-106 Review (2004).

35. 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).

36. Dormeyer, W., Mohammed, S., Breukelen, B., Krijgsveld J., Heck, A.J. Targeted analysis of protein termini. *J. Proteome Res*. **6**, 4634-4645 (2007).

37. 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).

38. Pickart, C.M. and Raasi, S. Controlled synthesis of polyubiquitin chains. *Methods Enyzmol.* **399**, 21-36 (2005).

39. Brzovic, P.S., et al. Binding and recognition oin 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. Amyloid Proteins: Methods and Protocols. *Methods Mol. Biol*. **299**, 35-51 (2004).

42. Winborn, B.J., et al. The deubiquitinating denzyme 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 experiements for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nuc. Mag. Res. Spec*. **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. and 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., and Koch, M.H.J. CRYSOL: 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., and Gaetano, M.T. Evaluating protein structures determined by structural genomics consortia. *Proteins*. **66**, 778-795.

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.

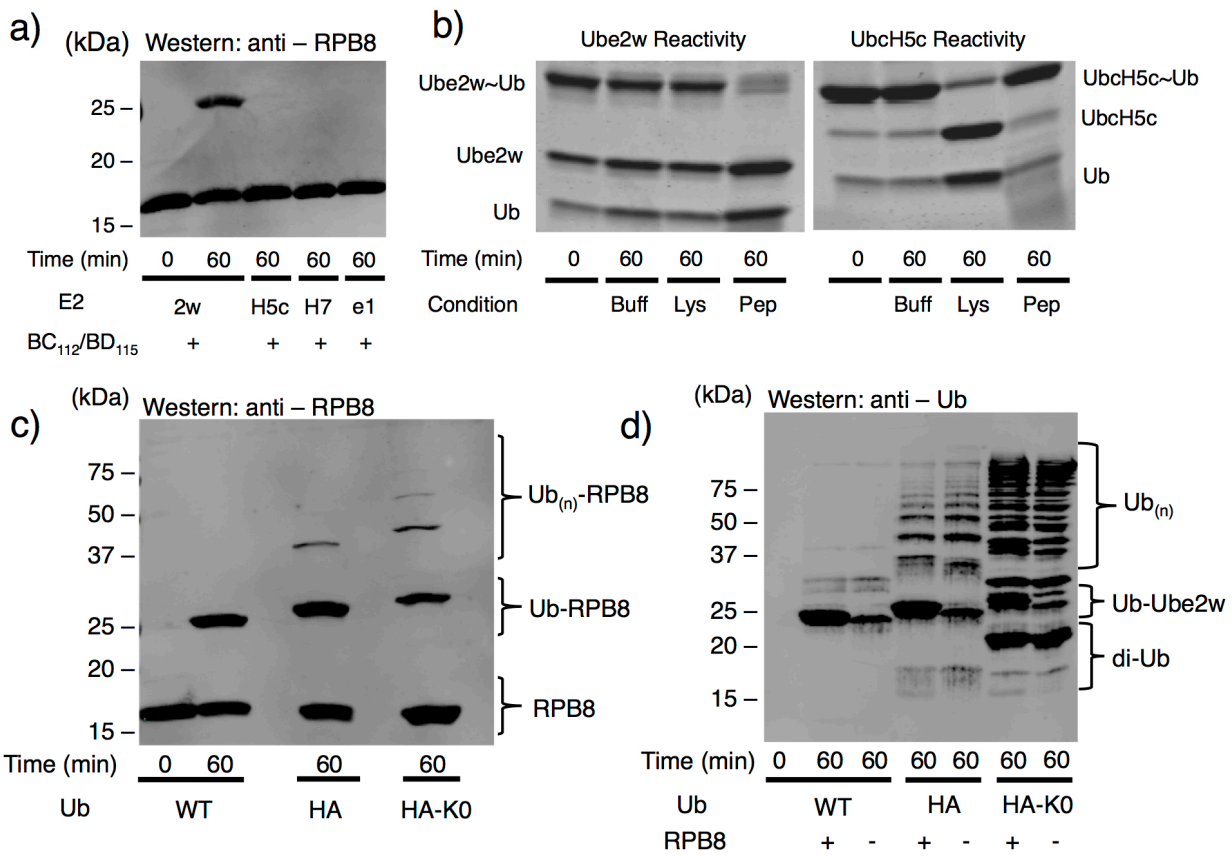


Figure 1. **Ube2w has distinct E2 activity**. **(a)** Ube2w transfers a single Ub to RPB8 *in vitro* while other BRCA1-interacting ubiquitin conjugating enzymes UbcH5c ("H5c"), UbcH7 ("H7"), and Ube2e1 ("e1") do not (**Supplementary Fig. 2a**). **(b)** ***Left*** A nucleophile reactivity assay reveals Ube2w has intrinsic activity with αNH_2 groups of a peptide with a free NH_2 group at its N-terminus (NH_2-A-G-G-S-Y-COO^-; 50 mM) but not the εNH_2 groups of lysine. ***Right,*** Identical reactions with UbcH5c~Ub conjugates confirm the previously reported lysine reactivity of UbcH5c and reveal it to be unreactive towards 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 HA-Ub, which contains a 13-residue tag at the N-terminal end of Ub; Lane 4: Reaction carried out with lysine-less HA-Ub (HA-Ub(K0)) 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 Panel (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**).

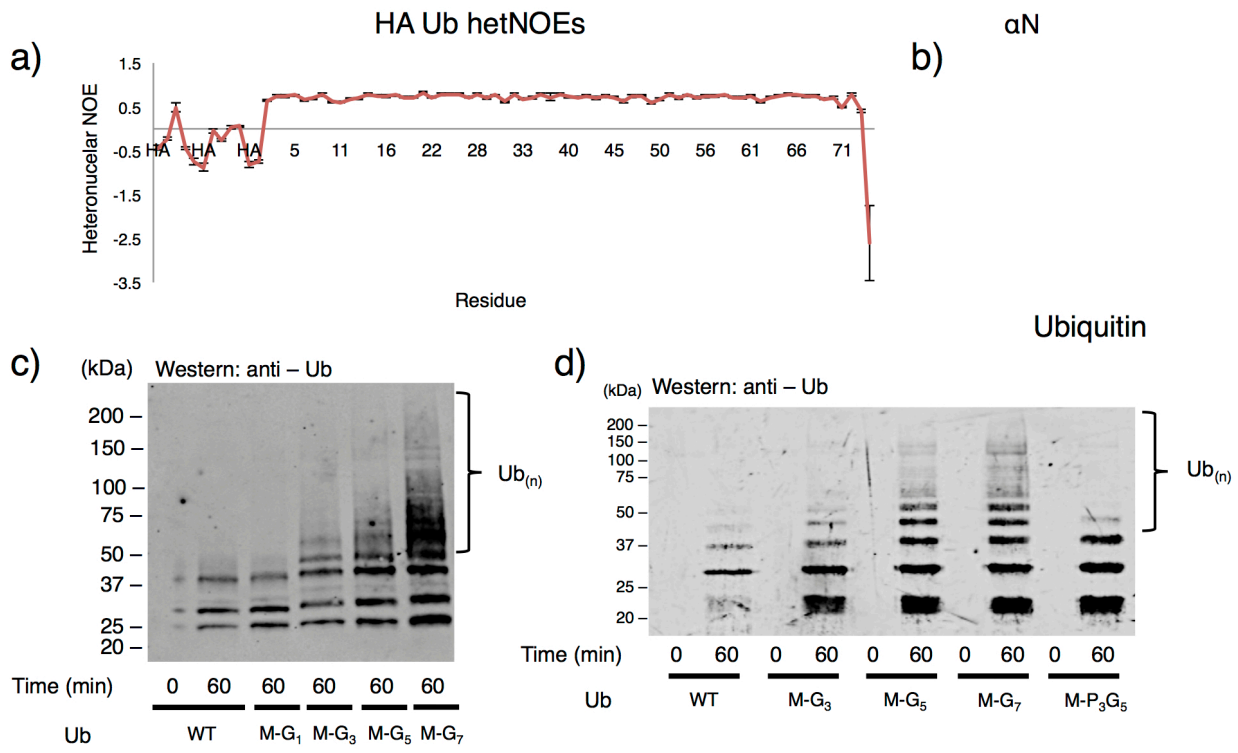


Figure 2. **Ube2w transfers Ub to flexible/disordered N-termini**. **(a)** Negative {^{1}H – ^{15}N} hetNOE values for residues derived from the HA-tag are indicative of highly flexible amino acids. Errors bars represent the standard error from the mean. (Resonances from this tag are not assigned and are not plotted sequentially. They are labeled simply as "HA"). **(b)** Consistent with the {^{1}H – ^{15}N} hetNOE data, the crystal structure of Ub (PDB: **1UBQ**) is ordered at its αN-terminus and immediately forms a β-strand with residue Met1. **(c)** Ub to which two N-terminal amino acids have been added at the N-terminus (Met-Gly-Ub) is not incorporated into chains by Ube2w and displays similar activity to WT Ub. Four N-terminal residues (Met-Gly_3-Ub) are sufficient to induce Ube2w activity towards Ub. Addition of six (Met-Gly_5-Ub) or eight (Met-Gly_7-Ub) residues increases Ube2w N-terminal ubiquitination activity *(Note: bands below 37kDa are consistent with auto-ubiquitinated E2 and E3)* (**Supplementary Fig. 2f**). **(d)** N-terminal backbone amide groups are necessary for Ube2w-dependent ubiquitination. Ube2w shows increased activity with the addition of disordered N-terminal amino acids, (Lanes 1-8). Proline at positions 2-4 (Lanes 9,10) inhibits Ube2w chain-building activity to levels similar to WT-Ub (**Supplementary Fig. 2g**).

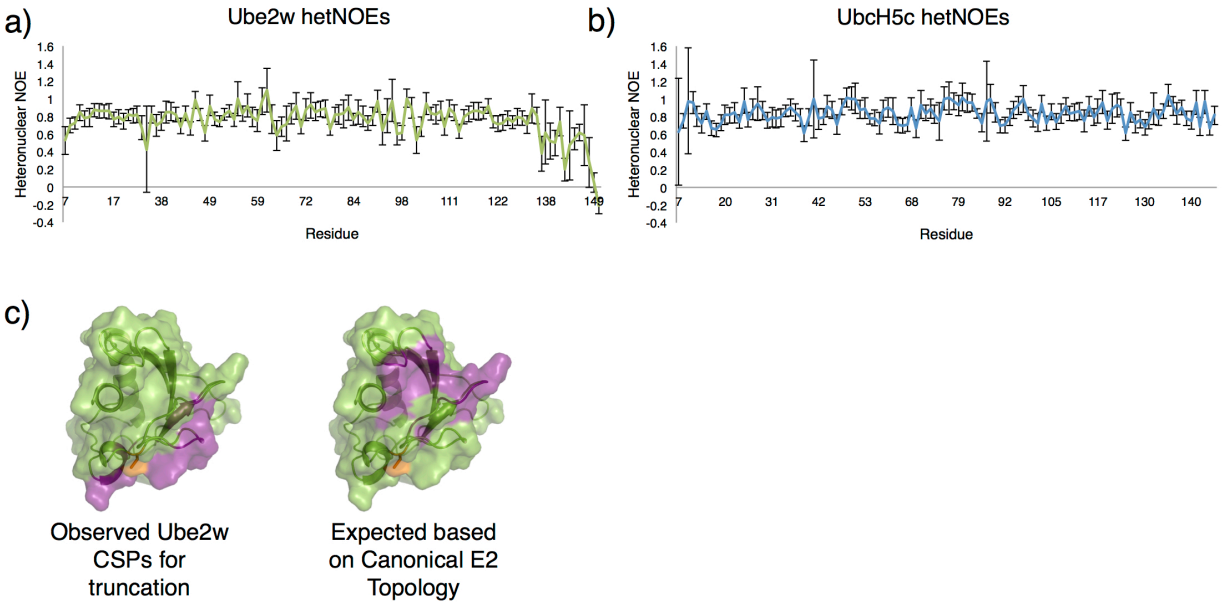


Figure 3. The Ube2w C-terminus is flexible and occupies a non-canonical position. **(a)** Residues 7-132 of Ube2w have generally uniform and positive {^{1}H – ^{15}N} hetNOE values. Beginning at residue 137, values decrease and ultimately become negative at the extreme C-terminus, consistent with a region that undergoes motions at higher frequencies than the core of the protein. Errors bars represent the standard error from the mean. **(b)** For comparison, UbcH5c has positive hetNOE values throughout its entire protein sequence, even at the far C-terminus. **(c)** ***Left,*** Experimental CSP data based on comparing the (^{1}H – ^{15}N) – HSQC-TROSY spectra of Ube2w-KK and Ube2w-131Δ-KK reveals that removal of the C-terminus perturbs residues near the active site, in the 3_{10}-helix, and on the 'backside' β-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.

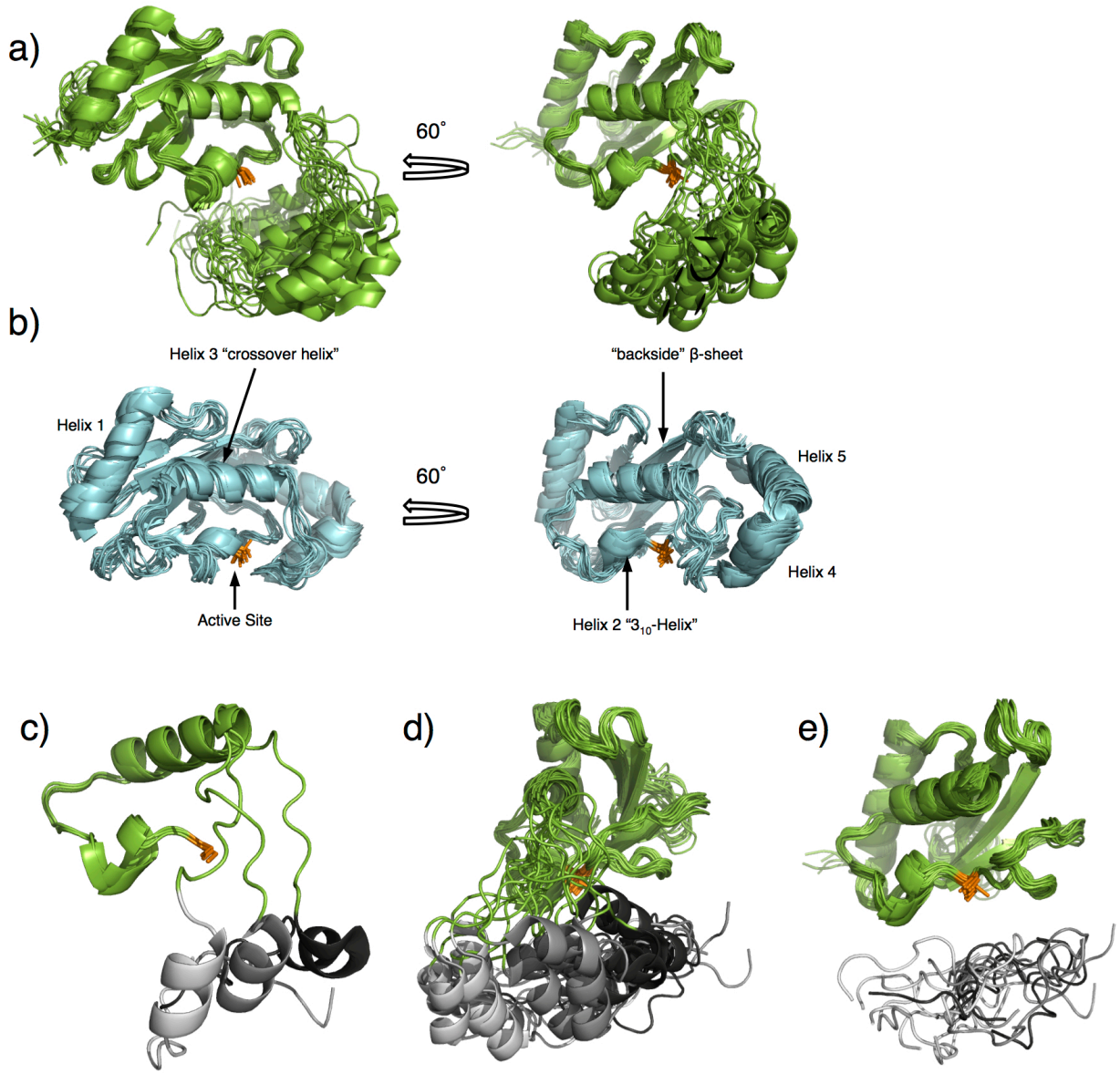


Figure 4. **NMR ensemble of Ube2w reveals a novel E2 architecture**. **(a)** Solution ensemble of Ube2w derived from NMR restraints (backbone chemical shifts, CSPs, residual dipolar couplings (RDCs), paramagnetic spin-label data, and small-angle X-ray scattering (SAXS)) calculated with CS-Rosetta. The twenty 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 **2FUH**). **(c)** Helix-4 (penultimate helix) is in distinct positions in Ube2w (3 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 (Cluster 1, light gray; Cluster 2, gray; Cluster 3, dark gray). **(d)** Side-view of the full Ube2w ensemble looking down the helix-3 axis reveals the three clusters. **(e)** In all twenty 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 Cβ atom of every residue is on average 14.5-17.5 Å away from the active site.

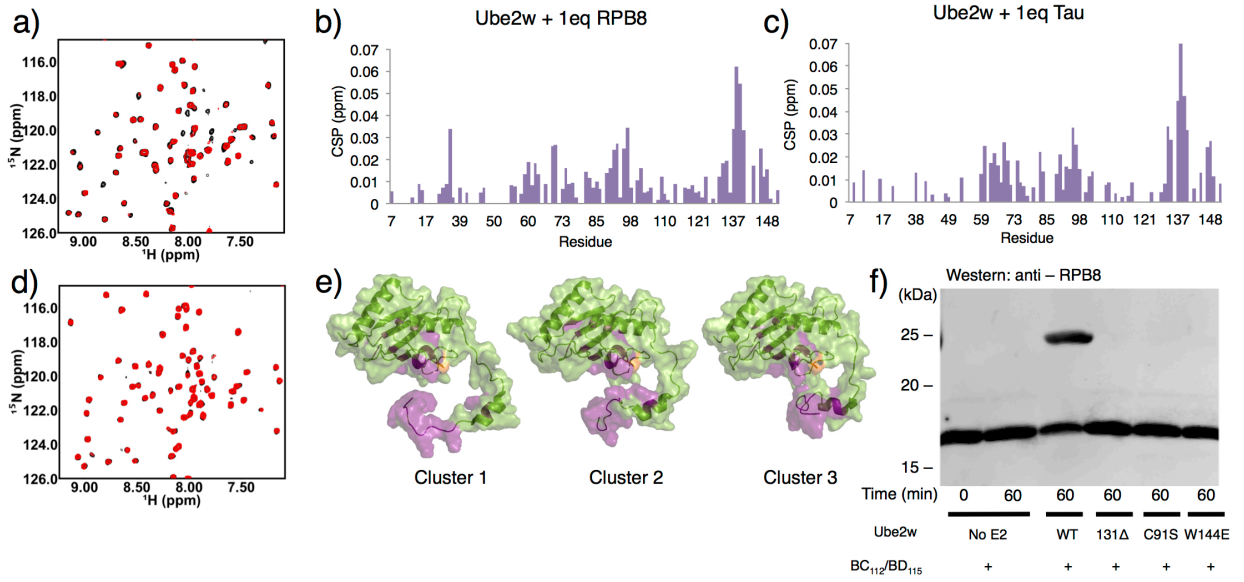


Figure 5. The Ube2w C-terminus is required to interact with substrates. **(a)** {^{1}H – ^{15}N} – HSQC-TROSY spectrum of Ube2w-KK in the absence (black spectrum) and presence of 1 molar equivalent of RPB8 (red spectrum). Evidence for binding is seen as peak broadening (loss of intensity) and chemical shift perturbations of specific peaks in the Ube2w NMR spectrum. **(b)** A histogram showing chemical shift perturbations (CSPs) upon 1 molar equivalent of RP8 into Ube2w-KK. **(c)** Titration of 1 molar equivalent of tau into Ube2w-KK reveals very similar CSPs to addition of RPB8. **(d)** {^{1}H – ^{15}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 significant intensity losses and/or CSPs (> 1 standard deviation) 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 1hr. 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 an active site-dead mutant, C91S (**Supplementary Fig. 2k**).

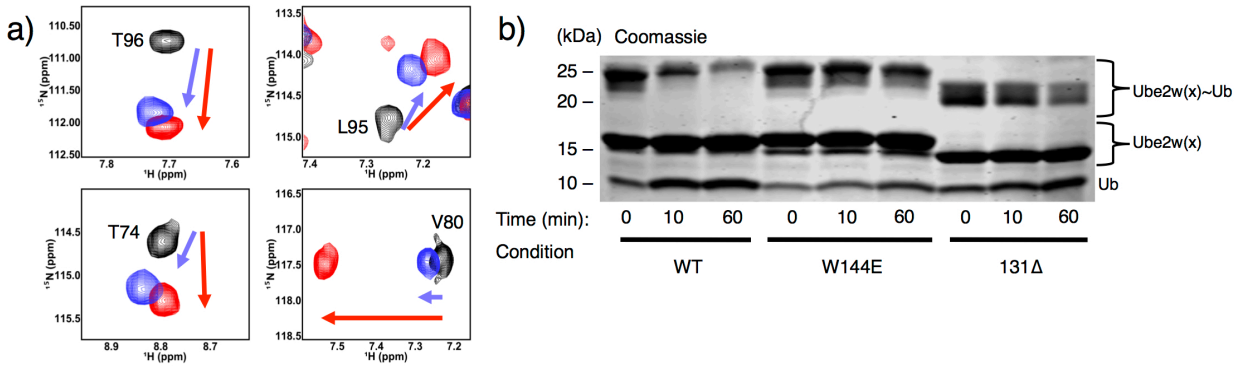


Figure 6. The Ube2w C-terminus facilitates α-amino group reactivity. **(a)** Selected {^{1}H – ^{15}N} resonances of residues near the Ube2w active site are compared in the spectra of full-length WT (black), W144E-Ube2w, (purple), and Ube2w-131Δ (red). Resonances move along similar trajectories as a 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 1hr reaction WT Ube2w~Ub shows robust transfer activity towards peptide (NH_2-A-G-G-S-Y-COO^-; 30 mM) as seen by increased amounts of free Ube2w and free Ub. Ube2w-W144E~Ub and the Ube2w-131Δ~Ub mutants show almost no Ub transfer activity to this minimal substrate over the same time period (**Supplementary Fig. 2l**).

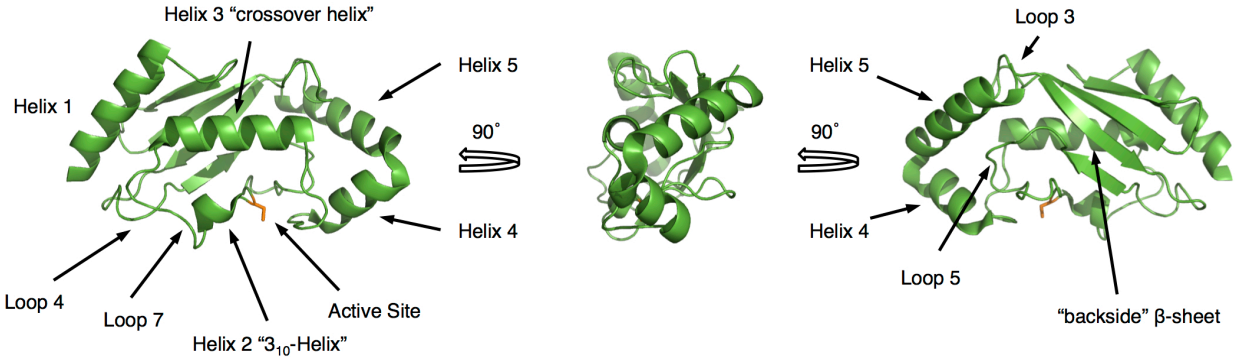


Figure S1. Structural architecture of canonical E2 ubiquitin-conjugating enzymes. ***Left,*** Structure of UbcH5c (PDB: **1X23**) facing helix-3 ("crossover helix") with structural nomenclature. ***Middle***, View of UbcH5c looking down helix-3. ***Right,*** View of UbcH5c from the "backside" β-sheet.

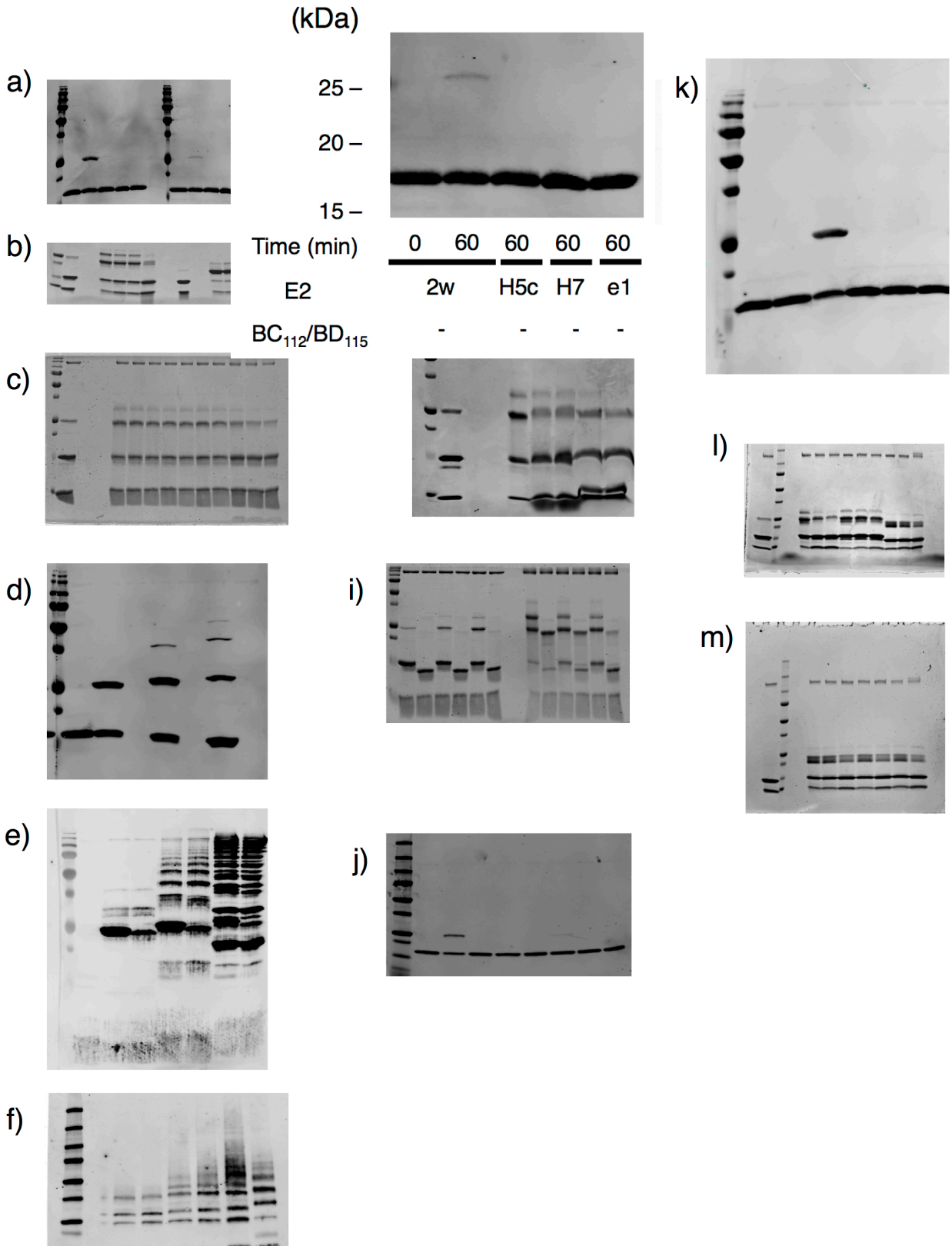


Figure S2. Un-cropped gel and western images.

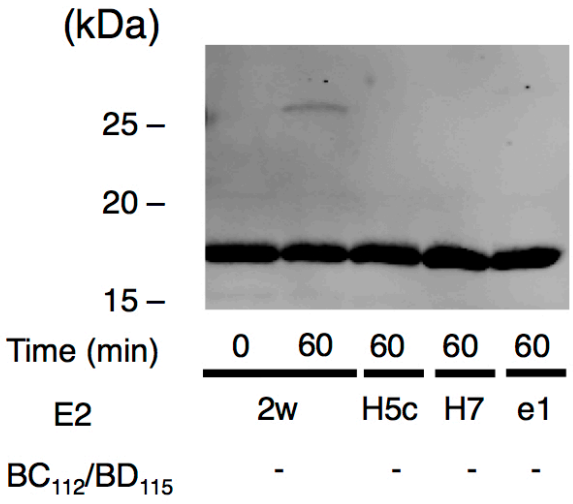
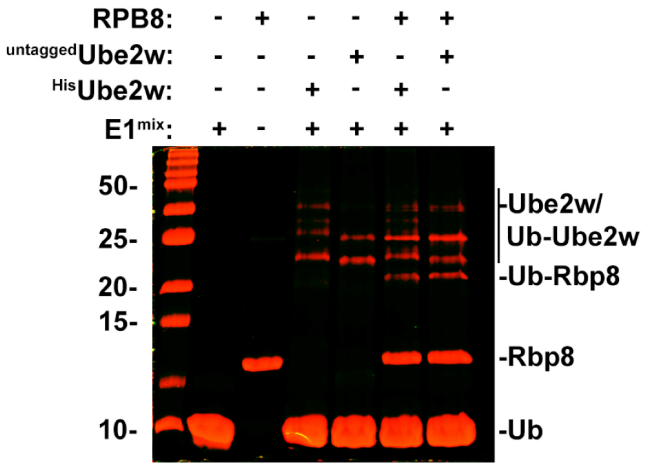
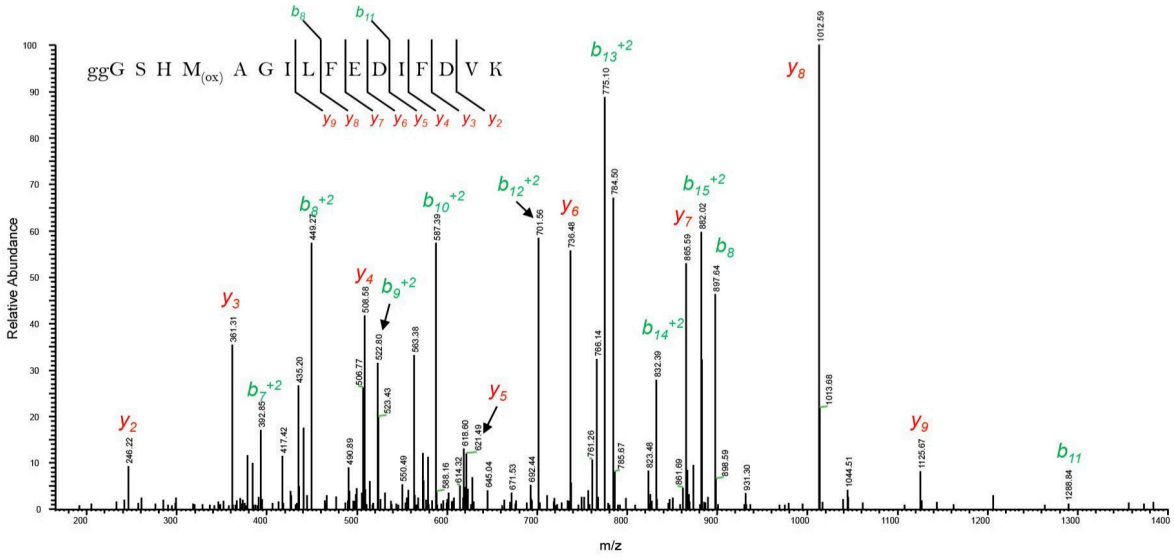


Figure S3. Ube2w shows low E3-independent ubiquitin transfer to RPB8. Ube2w adds mono-ubiquitin to RPB8 in the absence of the minimal RING domain of BRCA1/BARD1 ($BRCA1_{112}$/$BARD1_{115}$) in a 1hr reaction, as evidenced by the new band that appears in a western blot against RPB8. (**Figure S2a**).

a)



b)



Figures S4. Ube2w modifies the αN-terminus of RPB8. **(a)** RPB8 was ubiquitinated *in vitro* by Ube2w, run on SDS-PAGE. After in-gel digestion with trypsin, peptides were subjected to LC-MS/MS analysis using an orbitrap mass spectrometer. **(b)** The resulting MS/MS spectra were searched against the Swiss-Prot human protein database appended with synthetic RPB8 using the x!Tandem/TPP software suite. The majority of PSMs assigned to N-terminal peptide were ubiquitinated, indicating modification of the RPB8 αN-terminus.

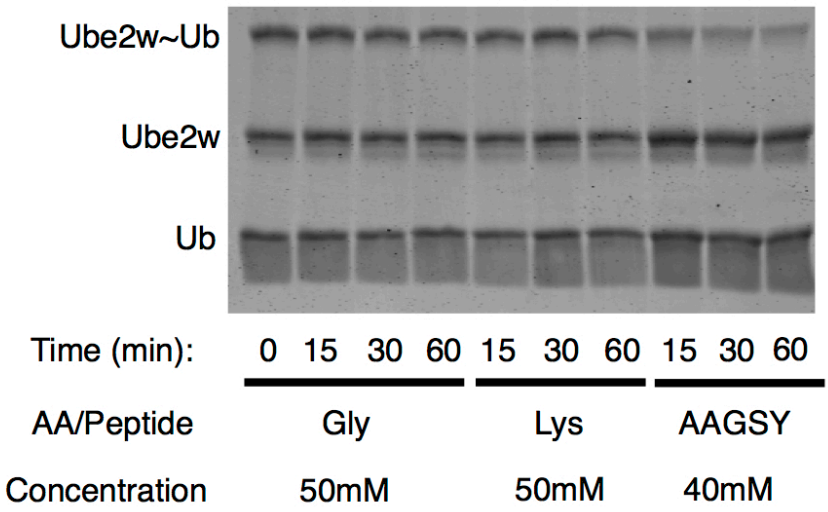


Figure S5. **Time courses of nucleophile reactivity assays**. In nucleophile reactivity assays Ube2w~Ub conjugates are unreactive with the free amino acids glycine and lysine (50 mM) over a 30-minute time course (coomassie-stained gel). In contrast, the conjugate has reacted significantly after 15min with a peptide containing a free αNH_2 group (NH_2-A-G-G-S-Y-COO^-; 40mM). The results indicate that Ube2w~Ub conjugates react specifically with αNH_2 group in the context of a peptide but not with the αNH_2 group in the context of a free amino acid or with εNH_2 groups (**Figure S2c**).

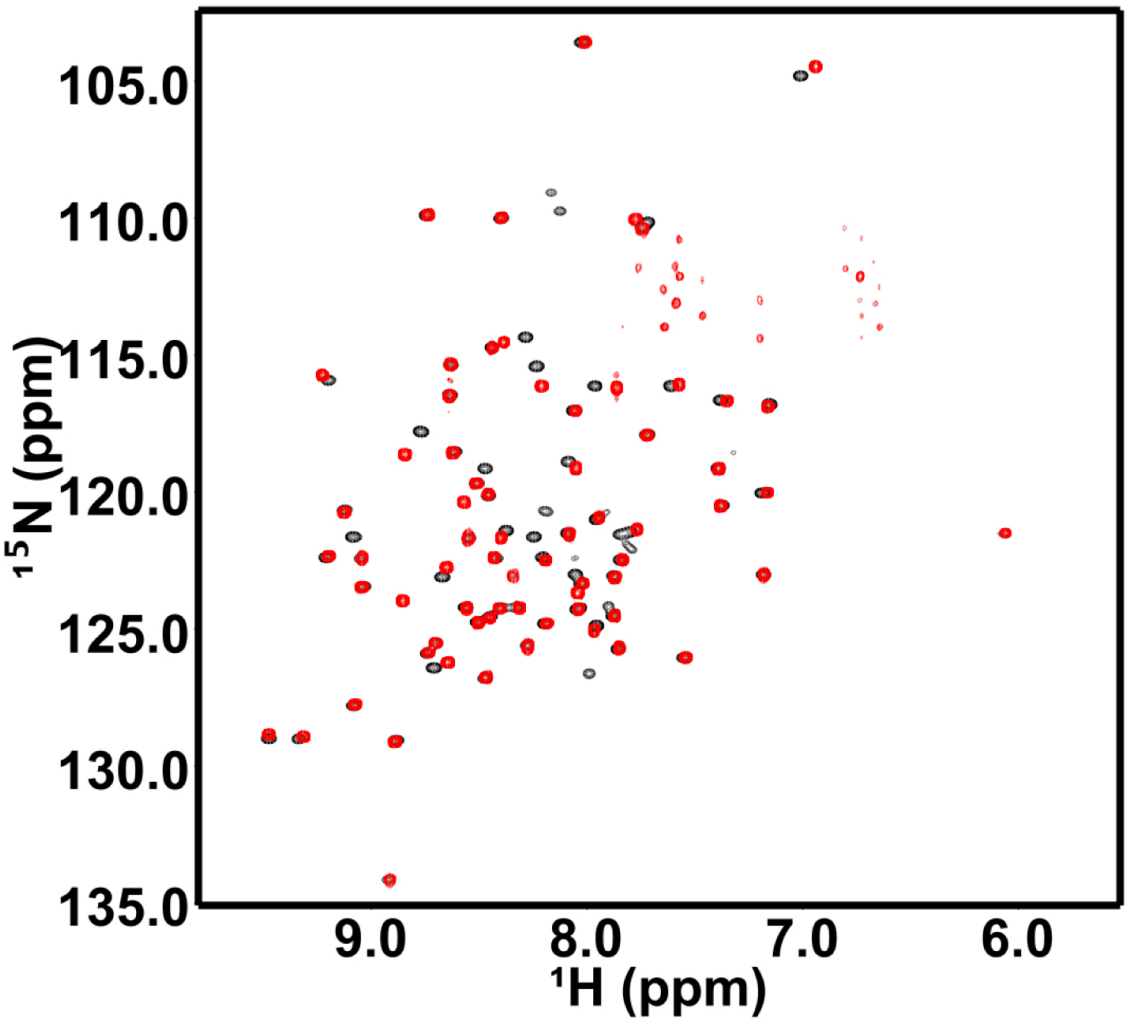


Figure S6. **Overlay of {^{1}H – ^{15}N} – HSQC-TROSY spectra of WT Ub and HA-Ub**. {^{1}H – ^{15}N} – HSQC-TROSY spectrum of HA-Ub (black) overlaid with the {^{1}H – ^{15}N} – HSQC-TROSY of WT Ub (red). Black with no red peaks overlapping are associated with the N-terminal HA-tag but have not been assigned to specific residues in the tag.

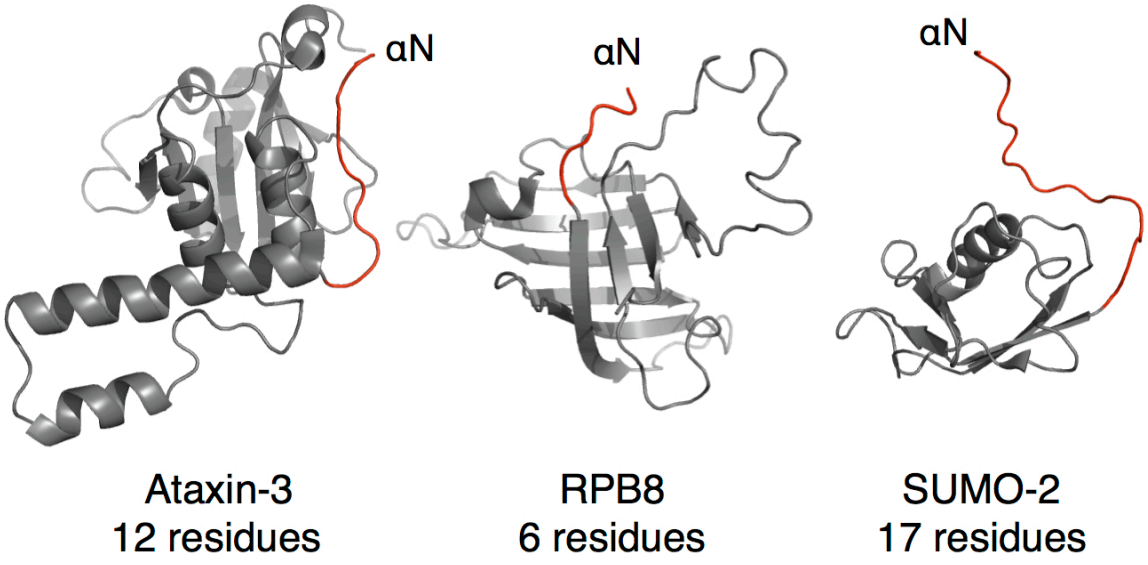


Figure S7. Structures of proteins shown to be ubiquitinated on their N-termini by Ube2w; ataxin-3 (PDB: **1YZB**), RPB8 (PDB: **2F3I**), and SUMO-2 (PDB: **1WZ0**) all have unstructured N-termini (red).

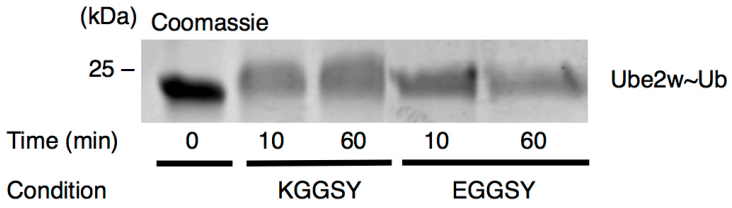


Figure S8. Ube2w can ubiquitinate N-termini over a wide range of pKa values. Two different peptides (NH_2-K-G-G-S-Y-COO^- and NH_2-E-G-G-S-Y-COO^-) are similarly reactive as substrate in a peptide reactivity assay at pH 7.0. Based on calculated values for protein N-termini, a lysine at position 1 is expected to lower the a-amino pKa to ~7.3, while glutamate in position 1 will raise the pKa to ~9.1. Ube2w is able to transfer Ub to either peptide at similar levels at pH 7.0, demonstrating it is capable of reacting with N-termini having diverse pKa values (**Figure S2h**).

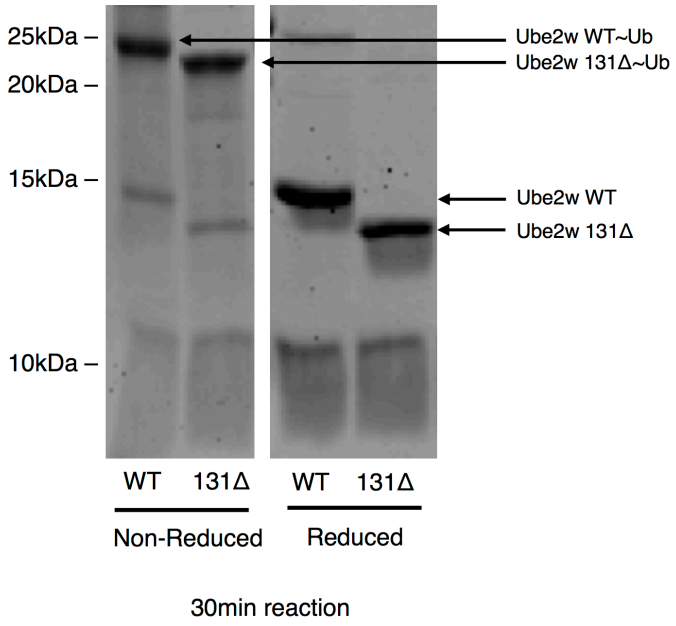


Figure S9. C-terminally truncated Ube2w can form an E2~Ub conjugate. In a ubiquitin loading assay, Ube2w-131Δ forms a Ub-conjugated species at levels similar to the full-length Ube2w (left panel). Addition of reducing agent shows that the observed species is formed through a thioester bond (right panel). Reaction times, 30 minutes (**Figure S2i)**

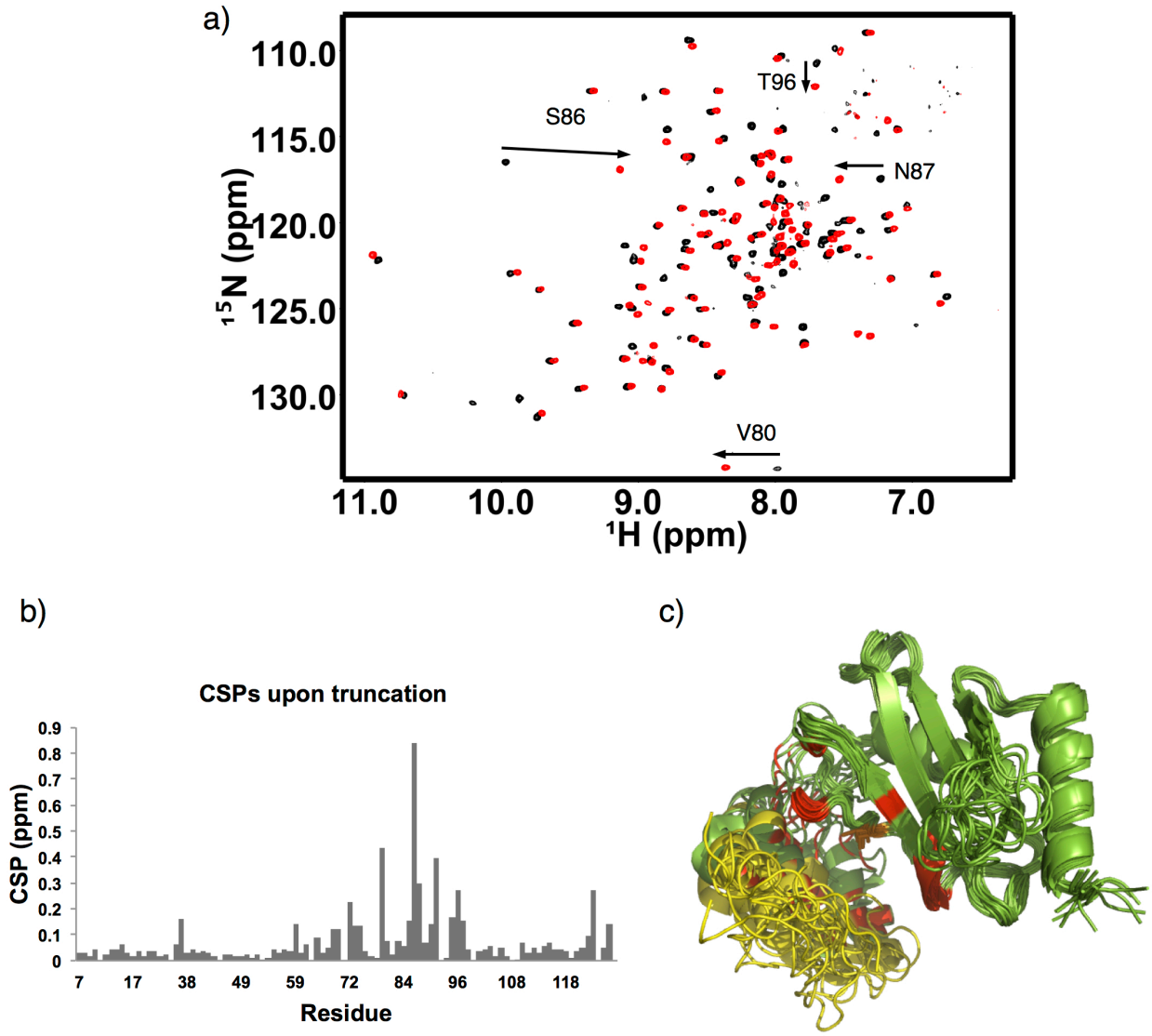


Figure S10. (**a**) **Comparison of the {^{1}H – ^{15}N} – HSQC-TROSY Ube2w-KK (black) spectrum with the {^{1}H – ^{15}N} – HSQC-TROSY Ube2w-131Δ-KK (red) spectrum**. Removal of residues 132-151 of Ube2w creates large chemical shift perturbations (CSPs) in the NMR spectrum. (**b**) A histogram of Ube2w-131Δ shows that CSPs upon truncation of the C-terminus cluster around the active site. **(c)** The Ube2w ensemble (green) depicting the region that was truncated (yellow), V132-C151. CSPs > .15ppm (M72, I78, Y85, S86, N87, L95, T96, E97, R124, and Y131) used as restraints in calculating the ensemble are mapped onto the final ensemble (red).

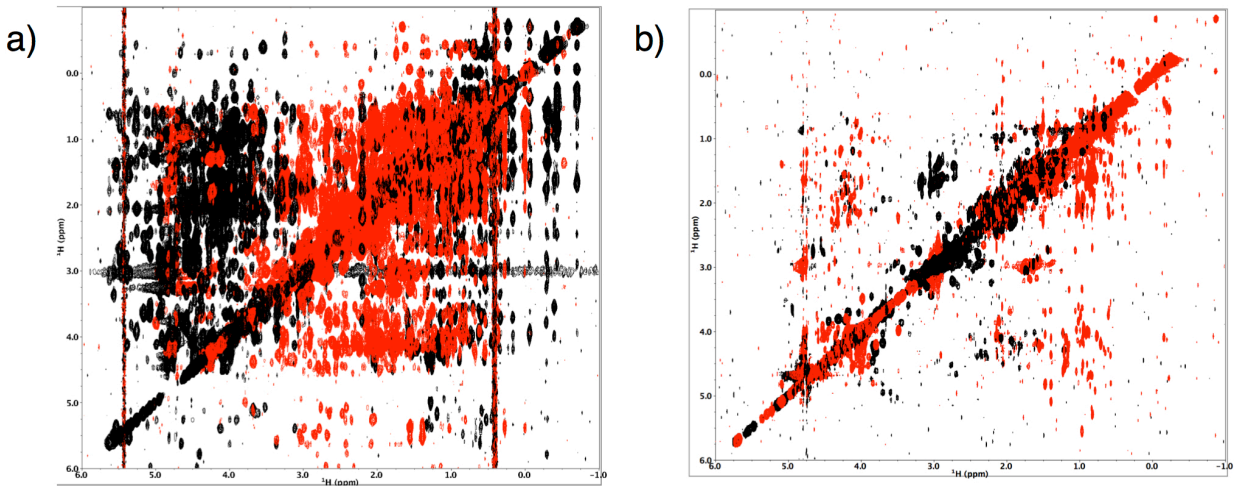


Figure S11. Ube2w ^{13}C-NOESY spectra shows a paucity of signal. (**a**) The 2D projection of a ^{13}C-HSQC-NOESY spectrum from UbcH5c shows robust off-diagonal NOE peaks, providing many distance constraints with which to calculate a *de novo* solution structure. (**b**) An identical experiment collected on Ube2w shows sparse and weak NOE signals, precluding an NOE-based structure calculation.

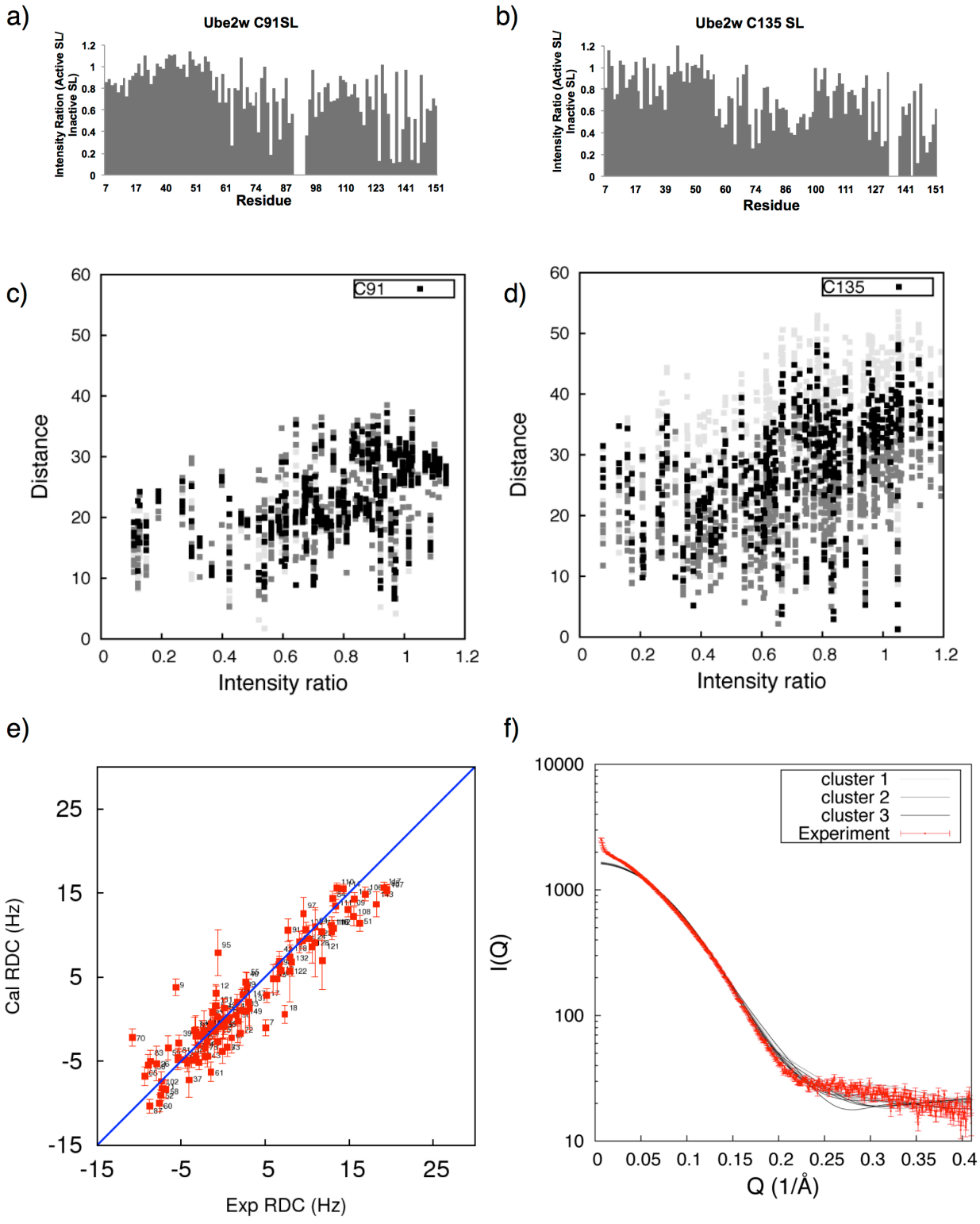

a)
Ube2w C91SL
Intensity Ration (Active SL/ Inactive SL)
Residue
b)
Ube2w C135 SL
Intensity Ratio (Active SL/ Inactive SL)
Residue
c)
C91
Distance
Intensity ratio
d)
C135
Distance
Intensity ratio
e)
Cal RDC (Hz)
Exp RDC (Hz)
f)
cluster 1
cluster 2
cluster 3
Experiment
I(Q)
Q (1/Å)

Figure S12. Correlation data for the Ube2w ensemble. **(a)** Histogram shows paramagnetic relaxation enhancement (i.e., peak broadening/intensity loss) upon activation of spin label at position C91 in Ube2w. Values are a ratio of the intensity of each resonance in the spin-label-active {^{1}H – ^{15}N} – HSQC-TROSY spectrum (Active SL) relative to the intensity of that resonance in the spin-label-inactive {^{1}H – ^{15}N} – HSQC-TROSY spectrum (Inactive SL) **(b)** Histogram shows paramagnetic relaxation enhancement (i.e., peak broadening/intensity loss) upon activation of spin label at position C135 in Ube2w, calculated as in (a) **(c)** Correlation of paramagnetic spin label data and the Ube2w ensemble. Each square in this plot represents a single amino acid from each of the 20 members in the final Ube2w ensemble, colored based on the cluster from which they are derived. On the x-axis is the experimentally observed intensity ratio based on the activation of the spin label at position C91. On the y-axis is the observed distance (Å) between the backbone amide hydrogen of that amino acid and an explicitly modeled SL at position C91. Based on these plots we were able to calculate Pearson correlation coefficients for individual members of the ensemble, each cluster, and the ensemble as a whole. **(d)** An identical plot with the spin label positioned at C135 **(e)** Scatter plot comparing experimentally measured RDCs versus RDCs back-calculated from the final structure ensemble. Each red square represents an amide in the Ube2w sequence. Error bars depict the range of back-calculated RDCs for each amino acid in the 20-member ensemble. (**f**) Experimental SAXS curve (red) is compared to back-calculated SAXS curves calculated from the final ensemble (gray).

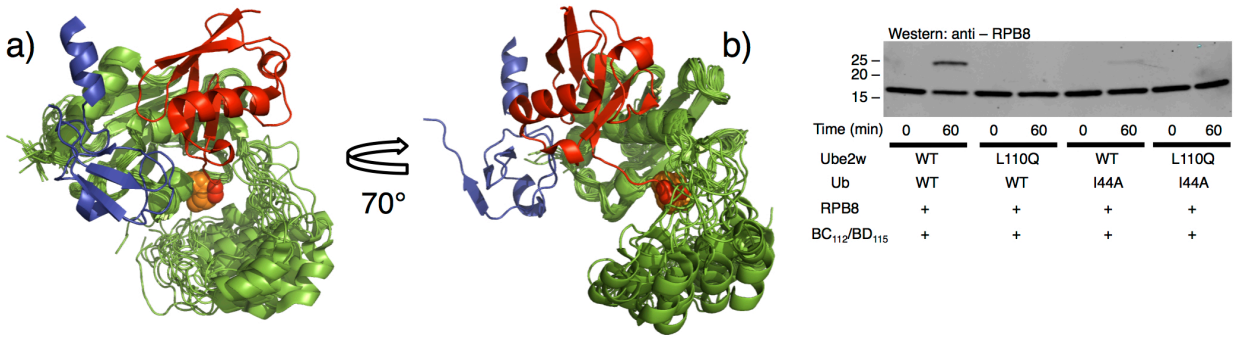


Figure S13. Ube2w is activated via a closed Ube2w~Ub state, similar to lysine-reactive E2s. **(a)** Model of Ube2w~Ub conjugate. The Ube2w ensemble was superimposed onto UbcH5b in the crystal structure of a RING E3/E2~Ub (BRIC7/UbcH5b~Ub) ternary complex[47] (PDB: **4AUQ**). The active site, C91, is colored in orange. **(b)** Mutations analogous to those that inhibit closed conformations in RING-type activation of lysine ubiquitination also inhibit N-terminal ubiquitination by Ube2w. An L110Q mutation in the crossover helix of Ube2w that inhibits closed E2~Ub conformations shows greatly decreased ability to ubiquitinate RPB8. Use of I44A-Ub in place of WT-Ub in the assay also inhibits N-terminal ubiquitination of RPB8. Substrate ubiquitination if further inhibited when both mutant proteins are used (**Figure S2j**).

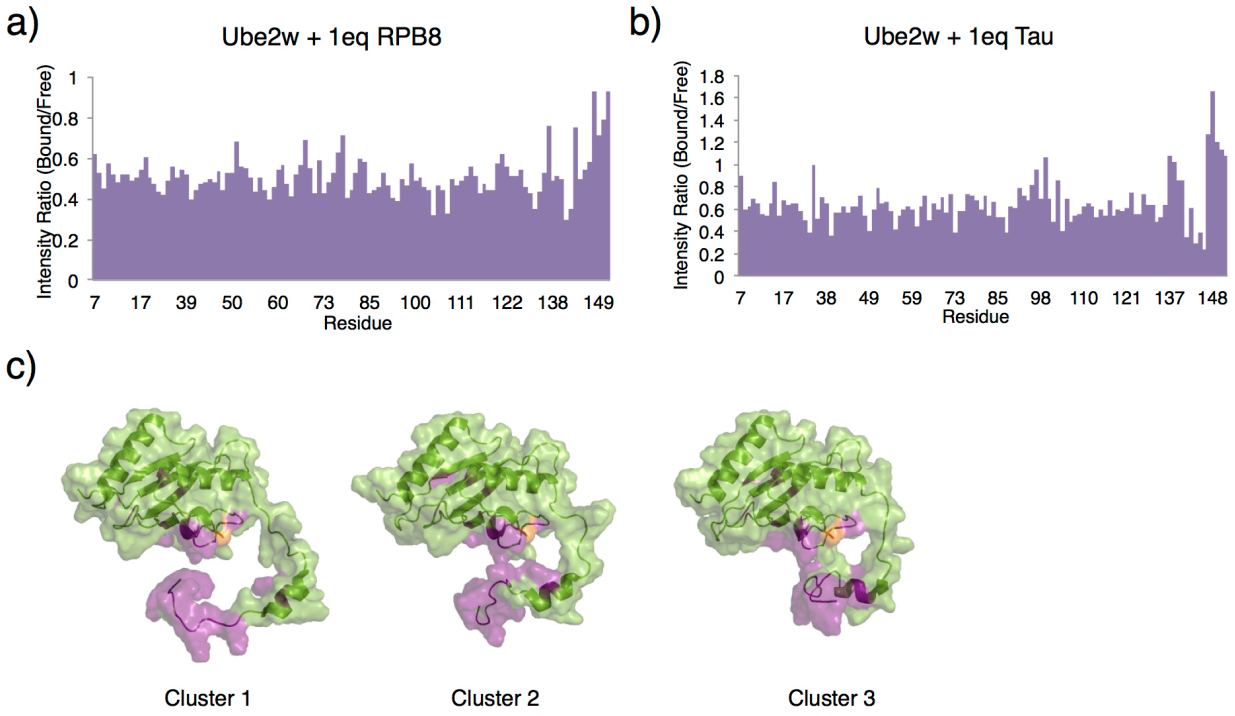


Figure S14. Perturbed residues in Ube2w upon addition of tau (a) Histogram shows peak broadening (relative peak intensities) observed upon addition of 1 molar equivalent of RPB8 to Ube2w-KK. Peaks with the most significant intensity loss are from residues in the C-terminal region. **(b)** Addition of 1 molar equivalent of tau reveals a similar peak broadening profile to RPB8. **(c)** NMR titration experiments with tau reveals a nearly identical interaction surface (red) as RPB8 on Ube2w.

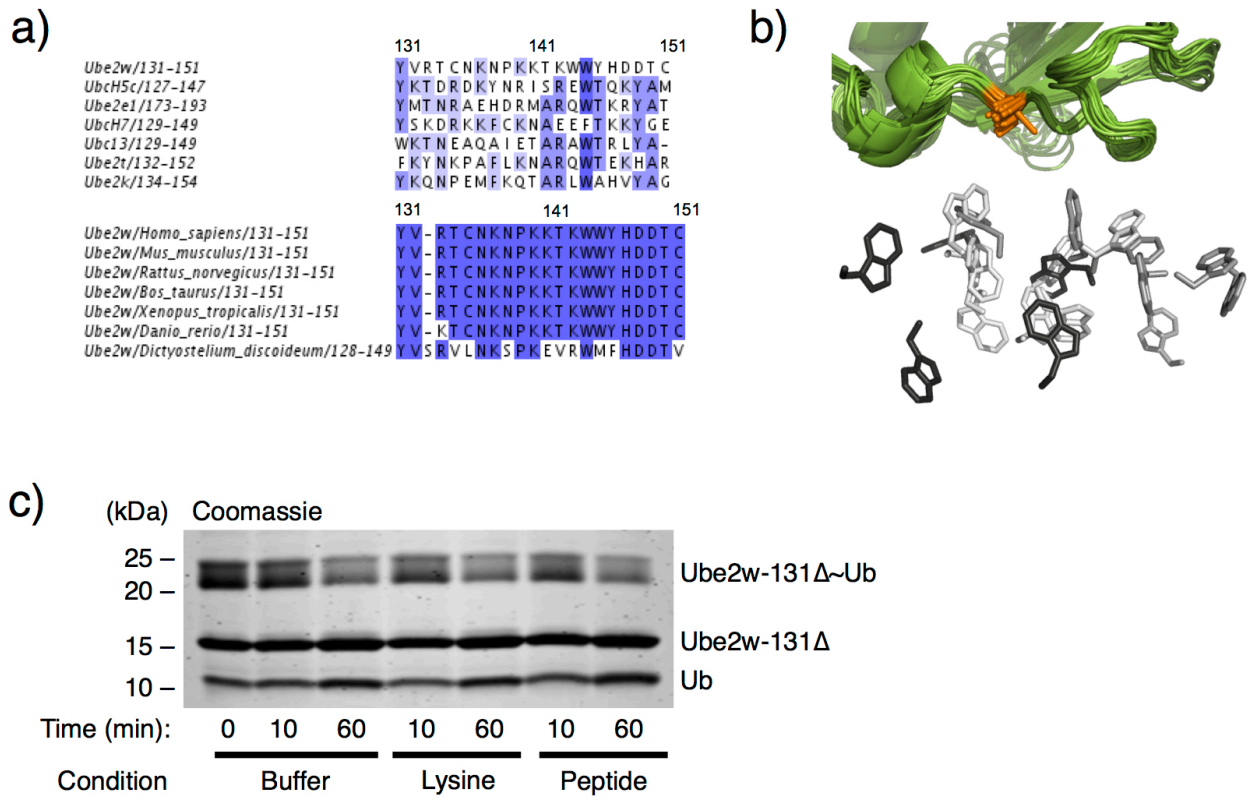


Figure S15. C-terminal contribution to Ube2w ubiquitination activity a) ***Top,*** Sequence alignment of the C-terminal region of Ube2w with other human E2s reveals low sequence similarity. ***Bottom,*** The Ube2w C-terminus is highly conserved throughout evolution, with even the slime mold *Dictyostelium discoideum* displaying high identity at amino acid positions that diverge from other E2s (R133, P139, W144, D148, D149). **(b)** Side chain positions of W144 in the twenty members of the Ube2w structure ensemble**.** W144 occupies positions as close as 7 Å (average 16.7 Å) away from the active site, C91. **(c)** Reactivity of the Ube2w-131Δ~Ub conjugate is due to hydrolysis and not aminolysis. After 1hr, Ube2w-131Δ~Ub reacts with buffer, lysine, or peptide identically, implying that the observed reactivity is due to hydrolysis (**Figure S2m**).

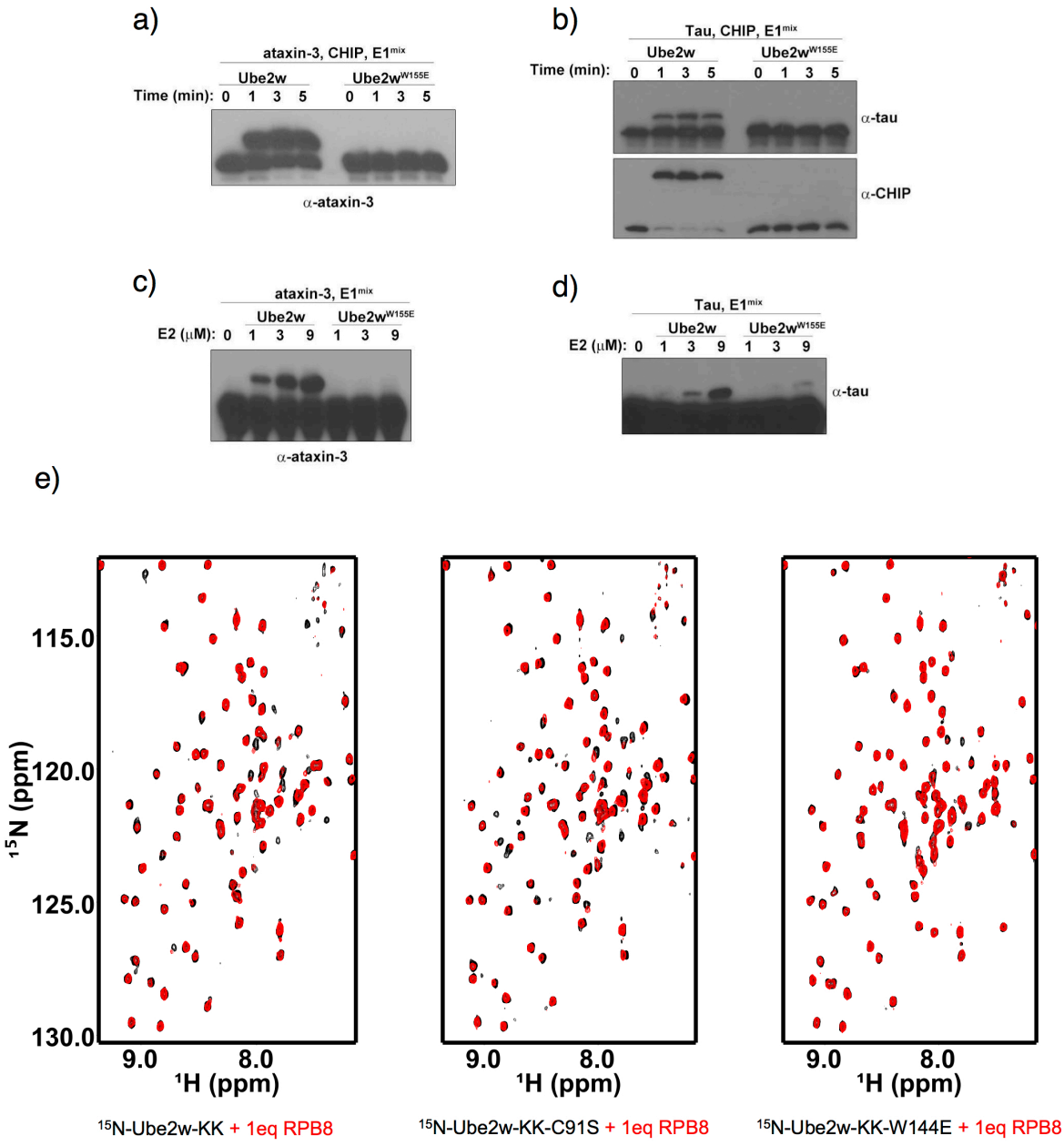

a)
ataxin-3, CHIP, E1mix
Ube2w
Ube2wW155E
Time (min): 0 1 3 5 0 1 3 5
α-ataxin-3
b)
Tau, CHIP, E1mix
Ube2w
Ube2wW155E
Time (min): 0 1 3 5 0 1 3 5
α-tau
α-CHIP
c)
ataxin-3, E1mix
Ube2w
Ube2wW155E
E2 (μM): 0 1 3 9 1 3 9
α-ataxin-3
d)
Tau, E1mix
Ube2w
Ube2wW155E
E2 (μM): 0 1 3 9 1 3 9
α-tau
e)
115.0
120.0
125.0
130.0
15N (ppm)
9.0
8.0
1H (ppm)
15N-Ube2w-KK + 1eq RPB8
15N-Ube2w-KK-C91S + 1eq RPB8
15N-Ube2w-KK-W144E + 1eq RPB8

Figure S16. The W144E mutation has effects in diverse substrates. *In vitro* ubiquitination assays were performed on previously reported substrates, ataxin-3 (panels (**a**) and (**c**)) and tau (panels (**b**) and (**d**)). One hour reactions were carried out as a function of increasing Ube2w concentration, in the presence (Panels (**a**) and (**b**)) or absence (Panels (**c**) and (**d**)) of the E3, CHIP, with WT-Ube2w (Lanes 1-4 on each blot) or with the W155E-Ube2w mutant (Lanes 5-7 on each blot). While Ube2w robustly adds a single ubiquitin to ataxin-3, Ube2w-W155E is incapable of ubiquitinating ataxin-3 or tau. (**e**) The W144E mutation affects the ability of Ube2w to interact with substrate. Overlay of {^{1}H – ^{15}N} – HSQC-TROSY spectrum of Ube2w-KK species in the absence (black spectrum) and presence of 1 equivalent of RPB8 (red spectrum). Specific peak broadening and shifting are observed for Ube2w-KK (***Left***) and the active site-dead mutant, Ube2w-C91S-KK (***Middle)***. Titration of a 1 molar equivalent of RPB8 into ^{15}N-Ube2w-W144E-KK shows less peak broadening and shifting compared to Ube2w-KK, indicative of a decrease in binding interaction (***Right***).

Construct	N-terminal sequence
RPB8	GSHMAGF
tau	MGSHHH
ataxin-3	GPLSIN
CHIP Full Length	GPLGSM
CHIP U-box	GKRDIP
HIS-tag	MGSHHHHHH
HA-tag	MYPYDVPDYAAGGS
T7-tag	GSHMASMTGG

Table S1. N-terminal sequences ubiquitinated by Ube2w. Ube2w ubiquitinates diverse N-terminal sequences with no discernable sequence homology. Furthermore a methionine at position 1 is not a requirement for Ube2w-dependent ubiquitination.

Experimental NMR restraints and refinement statistics for Ube2w solution structure ensemble.

NMR restraints	**Number of unique restraints used to generate CS-Rosetta Ube2w structure**
Backbone chemical shift restraints	
Total	618
HN	121
N	130
CA	137
C	104
CB	126
RDC (H-N)	109
SAXS	543
Spin label C91	104
Spin label C135	109
Total Restraints	1483
Structure statistics	
Deviations from idealized geometry	
Bond lengths (Å)	0.009
Bond angles (°)	0.4
Average pairwise r.m.s. deviation** (Å)	
All backbone	4.1
All heavy	4.6
RDC Q-factor	0.38 ± 0.02
Ramachandran statistics	
Residues in most favored region	88.9
Residues in additionally allowed region	10.6
Residues in generously allowed region	0.5
Residues in disallowed region	0.0

** 20 structures used in RMSD calculations. Pairwise r.m.s. deviation was calculated among 20 refined structures.

Table S2. **Experimental restraints and structure statistics**. Type and number of unique experimental restraints used to generate solution structure ensemble are listed. Structural statistics for the final structure ensemble were calculated using the Protein Structure Validation Software (PSVS) suite and X^2 values were calculated using the European Molecular Biology Laboratory (EMBL) CRYSOL software.

Model	SAXS [X^2]	RDC [Q]	C91 [Pearson Corr]	C135 [Pearson Corr]
c1_m01	10.68	0.3997	0.522393743	0.459397189
c3_m02	11.73	0.3882	0.470840305	0.47815034
c1_m03	10.04	0.38911	0.56797393	0.51784523
c2_m04	10.33	0.3754	0.456744167	0.459344746
c1_m05	11.53	0.39088	0.518720743	0.474355803
c3_m06	11.29	0.39014	0.478353895	0.470661834
c2_m07	9.03	0.41292	0.516314824	0.506287194
c1_m08	10.31	0.4036	0.485138411	0.495520513
c1_m09	8.62	0.32618	0.520148728	0.454588322
c1_m10	8.68	0.38784	0.530484121	0.512382162
c1_m11	7.09	0.34324	0.490705528	0.443691154
c3_m12	8.95	0.37718	0.514574079	0.471484177
c1_m13	9.54	0.38669	0.504495797	0.483148155
c3_m14	9.53	0.35767	0.500839375	0.489409108
c2_m15	11.58	0.38797	0.469305258	0.56439435
c2_m16	11.43	0.38409	0.451672504	0.458219614
c2_m17	11.58	0.3786	0.450581865	0.480244929
c2_m18	9.24	0.39843	0.538941933	0.504266527
c3_m19	10.32	0.41105	0.420683561	0.530363488
c2_m20	8.6	0.37991	0.505141012	0.401250264
Ensemble	10.005	0.38344	0.495702689	0.482750255
Error	1.2964506	0.02107	0.035532669	0.03501417

Cluster	SAXS [X^2]	RDC [Q]	C91 [Pearson Corr]	C135 [Pearson Corr]
Cluster 1	9.56125	0.3784	0.517507625	0.480116066
Cluster 1 Error	1.3983402	0.02798	0.025864992	0.027129251
Cluster 2	10.255714	0.38819	0.484100223	0.482001089
Cluster 2 error	1.3017534	0.01327	0.035661348	0.050788552
Cluster 3	10.364	0.38485	0.477058243	0.488013789
Cluster 3 error	1.164079	0.01951	0.036025777	0.024834703

Table S3. **Back-calculated values for SAXS, RDC, and spin label parameters for the Ube2w ensemble**. Included are X^2 (SAXS), Q-factors (RDCs), and Pearson correlations (spin label experiments) for each member of the ensemble, an average over the entire ensemble, and for members of each cluster.

Chapter IV – Biochemical characterization of the ubiquitin-conjugating enzyme Ube2h

Vittal, V., Stewart, M.D., Kim, J., Vane, E., Ueda, G., Baughman, H., Brzovic, P.S., Klevit, R.E.

Author Contributions: V.V. performed all of the biochemical assays, NMR experiments. M.D.S supervised NMR assignments of the Ube2h core and C-terminus. J.K. and E.V. assigned the Ube2h core domain. G.U assigned the Ube2h C-terminus, H.B., performed biochemical assays. V.V., P.S.B, and R.E.K. conceived the experiments. R.E.K. supervised the project.

Introduction

In RING-type E3 mechanisms, the E2~Ub conjugate plays a significant role in dictating the final Ub product on the substrate. Canonically, the RING E3 interacts simultaneously with both the E2~Ub conjugate and a substrate using different protein interfaces. The E3 works to allosterically activate the E2~Ub conjugate for transfer, and subsequently Ub is transferred directly to the substrate from the active site of the E2[1-3]. Mutations that disrupt the allostery between the E3 and E2 are as detrimental to the Ub transfer process as mutations that abrogate interactions between E3 and the substrate. Intriguingly, there are known cases of E2s that can bypass this traditional mechanism and directly ubiquitinate substrates in an E3-independent manner, albeit at a significantly lower level. Examples of this type of activity are E2s that building poly-ubiquitin chains. Ube2k, which builds K48-linked chains, contains two functionally distinct domains. The core UBC domain, which contains the active site cysteine, makes a covalent linkage with ubiquitin after a transthiolation event with E1. However, a structured C-terminal ubiquitin associate (UBA) domain forms noncovalent interactions with an acceptor Ub that presents K48 to the active site of the UBC domain. Successive iterations of this process produces K48-linked Ub chains.

There are other instances*, in vitro*, where E2's can bypass the canonical RING-type mechanism to ubiquitinate non-Ub substrates. Nearly twenty years ago it was shown that human Ube2h was capable of ubiquitinating bovine histone H2A in an E3-independent fashion *in vitro*[5]. Intriguingly, the authors showed that Ube2h contains two functionally distinct domains, a catalytic UBC domain and a C-terminal substrate-binding domain. They demonstrated that removal or truncation of the Ube2h C-terminus decreased ubiquitination of the substrate, but did not affect the loading of ubiquitin to its active site. Finally, the authors demonstrated that the C-terminal substrate-binding domain could be grafted to a different E2 enzyme that originally had no intrinsic activity towards histone H2A and convert it in to a histone ubiquitinating E2[5]. Although Ube2h possesses this E3-independent activity *in vitro*, in a cellular context, it is likely that it pairs with one or several cognate RING-type E3 ligases. Nevertheless, these results interested our laboratory and we pursued further biochemical characterization of Ube2h.

The results in this chapter expand upon the work from Kaiser et al. and demonstrate that Ube2h does contain two functionally and structurally distinct domains[5]. Ube2h contains a canonical N-terminal UBC domain, however the C-terminal substrate-binding domain is intrinsically disordered. We confirm that this disordered region is indeed necessary to ubiquitinate histone H2A/H2B dimer from *X. laevis*, and

identify sites of interaction in the C-terminus using NMR. Further, we present initial studies on the Ube2h~Ub conjugate.

Results

Ube2h has an intrinsically disordered C-terminus

To characterize Ube2h biochemically and structurally we turned to NMR spectroscopy. A ^{1}H, ^{15}N-HSQC spectrum of full-length Ube2h revealed high intensity resonances clustered between 8.0 and 8.5 ppm in the HN dimension of the spectrum (**Figure 1**). Generally, these types of clustered, high intensity resonances are indicative of an intrinsically disordered region. Following results from Kaiser et al. we tested if this disordered region was derived from the functionally independent C-terminal region of Ube2h. We created a construct of Ube2h that truncated the final 31 residues after E151 (Ube2h-151Δ), the predicted end of the UBC core domain. An identical ^{1}H, ^{15}N-HSQC spectrum collected on Ube2h-151Δ does not contain these high intensity resonances, confirming that the C-terminus of full-length Ube2h is in fact, disordered. Because of exchange processes related to the disordered C-terminus, full-length Ube2h proved to be a challenging species to collect three-dimensional NMR datasets for spectral assignments. Therefore, we chose to pursue characterizing Ube2h-151Δ and assigned its ^{1}H, ^{15}N-HSQC spectrum.

The Ube2h C-terminus is required for histone H2A/H2B ubiquitination

As show in Kaiser et al., Ube2h is able to ubiquitinate bovine histone H2A in the absence of an E3 ubiquitin ligase[5]. We wanted to ask what role the disordered C-terminus of Ube2h plays in this ubiquitination event. In an *in vitro* reaction containing human Uba1, full-length Ube2h, *X. laevis* histone H2A/H2B dimer (generously provided by Dr. Lisa Gloss, Washington State University), and HA-Ub, full-length Ube2h is capable of robustly monoubiquitinating the H2A/H2B dimer as visualized by a western blot for anti-HA-Ub (**Figure 2a, panel 1**). When H2A/H2B dimer is omitted from the reaction several ubiquitinated species can be distinguished, including the di-HA-Ub and auto-ubiquitinated Ube2h products, however, the ubiquitinated H2A/H2B band is lost (**Figure 2a, panel 2**). When full-length Ube2h

is replaced with the C-terminally truncated Ube2h-151Δ mutant, H2A/H2B monoubiquitination is nearly lost, indicating the C-terminus is important for mediating the Ube2h – H2A/H2B interaction (**Figure 2a, panel 3**). Omission of H2A/H2B from the reaction reveals that the intrinsic activity of Ube2h-151Δ is retained, as it is still capable of forming di-HA-Ub and its own auto-ubiquitinated products (**Figure 2a, panel 4**). To assure that the ubiquitination activity of full-length Ube2h towards histone H2A/H2B was E2-specific, we conducted a similar experiment with the E2, UbcH5c. Under identical reaction conditions with UbcH5c, no ubiquitinated H2A/H2B band is observed indicating this E2 is not reactive towards this substrate. Together, these results confirm that the histone H2A/H2B ubiquitination is specific towards Ube2h and that its C-terminus is important for mediating substrate interactions but not necessary for the intrinsic activity of the enzyme.

The Ube2h C-terminus mediates substrate interactions with histone H2A/H2B

To obtain residue specific information regarding the interaction surface between Ube2h and histone H2A/H2B we conducted ^{1}H, ^{15}N-HSQC titrations experiments. Upon the addition of .1 molar equivalents of H2A/H2B into full-length Ube2h global peak broadening effects were immediately observed making it difficult to pinpoint an interaction surface (**Figure 3**). Addition of more than .1 molar equivalents caused increased broadening effects and caused protein precipitation. A possible cause for the precipitation seen for this complex is the extreme difference in pI between Ube2h and histone H2A/H2B. Full length Ube2h has a theoretical pI of 4.55, while the C-terminal region between E152 and L183 has an even more extreme pI of 3.46. Conversely, histone H2A and H2B have very basic pI's of 10.55 and 10.31, respectively. To overcome the strong ionic interaction between this protein complex the salt concentration of the buffer was increased, yet, this had only minimal affects on decreasing precipitation. Buffer conditions must be explored further in order to stabilize this protein complex.

To directly observe interactions between the C-terminus and histone H2A/H2B an isolated construct of the final 31 residues of Ube2h was created (Ube2h-152-183). A ^{1}H, ^{15}N-HSQC of Ube2h-151-183 confirmed previous results that this region is disordered (**Figure 1, Figure 4a**). Three-dimensional NMR datasets were acquired for this construct and the ^{1}H, ^{15}N-HSQC was assigned. We then attempted

titration experiments by addition of histone H2A/H2B. Similar to the full-length version of Ube2h, addition of small equivalencies of H2A/H2B caused broadening in the Ube2h-152-183 spectrum, however, in solution, the complex did not precipitate (**Figure 4a**). Analysis of the most broadened residues revealed that two acidic patches in the C-terminus are sites of interaction with histone H2A/H2B. The first interaction site consists of only three acidic residues, E153, E158, and D163. Resonance A154 to Q157 and E159 to G162 show only minimal interaction with histone H2A/H2B (**Figure S1**). The second site is a major interacting area and consists of residues S168 to the C-terminal L183. Within this region, residues M170 to E177 have the greatest intensity loss (**Figure S1**). This region also has a stretch of three consecutive acidic residues, E175, D176, and E177. Based on the fact that histones H2A and H2B have very basic pI's, mutating these residues to alanine was a logical choice to potentially disrupt the interaction between the C-terminus of Ube2h and the H2A/H2B dimer. A ^{1}H, ^{15}N-HSQC of the triple mutant Ube2h-152-183-EDE-AAA reveals a similar spectrum to Ube2h-152-183, however, upon addition of .25 molar equivalents of histone H2A/H2B this mutant form interacts at a much lower level and the specific broadening seen in the titration of Ube2h-152-183 with H2A/H2B is lost (**Figure 4b**). These preliminary results must be tested further in biochemical assays that incorporate the EDE to AAA mutation into the full-length Ube2h to understand its effects on substrate ubiquitination.

The C-terminus of Ube2h interacts with residues in its UBC core domain

An overlay of the ^{1}H, ^{15}N-HSQC of full-length Ube2h and Ube2h-151Δ reveals that most residues in the UBC domain overlay quite well, indicating they are in similar chemical environments (**Figure 1**). However, upon closer examination, some resonances experience chemical shifts upon deletion of the Ube2h C-terminus (**Figure 5a**). These small shifts are indicative of changes in the chemical environment of backbone amides, in this case, a protein-protein interaction between the disordered C-terminus of Ube2h and residues within its core UBC domain. Mapping of the most shifted residues onto the crystal structure of the Ube2h UBC domain (PDB: **2Z5D**) reveals two surfaces[6]. The first, unsurprisingly, is near the site of truncation and extends back towards helix-5. However, a second surface is also extended away from the C-terminus of the core domain and wraps along the "backside" β-sheet of Ube2h (**Figure**

5b). Residues in this region include K37, G48, G49, K52, M71, and N72. Although it is difficult to speculate on whether this secondary interaction is functionally relevant, based on the previously addressed role of the C-terminus in substrate binding, this result is noteworthy and mutations must be made.

Initial characterization of the Ube2h-151Δ~Ub conjugate

Although studying E2s as individual entities can reveal intriguing insights, E2~Ub conjugates are more relevant, as they are the species that makes direct contacts with substrates. In order to study the conjugate using NMR, a simple cysteine to serine mutation must be made at the active site in order to form a more stable oxyester attachment between the active site of the E2 and the C-terminus of Ub. The natural thioester bond formed between the E2 active site and Ub proves to be labile and not suitable for NMR. Initial Ub loading assays with human E1 enzyme and WT Ub demonstrated that Ube2h-151Δ_{C87S} is capable of forming a robust Ube2h-151Δ-O~Ub species after 4hrs of incubation at 37˚C (**Figure 6a**).

An initial ^{1}H, ^{15}N-HSQC spectrum of the Ube2h-151Δ-O~Ub revealed several changes compared to the spectrum of Ube2h-151Δ (**Figure 6b**). Although this new spectrum has not been assigned by traditional three-dimensional techniques some inferences about the relative orientation of Ub to the Ube2h core domain can be made based on transferring assignments from the Ube2h-151Δ spectrum. Mapping of the most perturbed resonances following conjugation of Ub to the Ube2h-151Δ_{C87S} active site revealed several points of interaction (**Figure 6c, Figure S2**). Consistent with attachment of Ub to the active site serine, several residues surrounding this area are significantly perturbed. Particularly, H76 and N78 of the canonical HPN motif in lysine transferring E2s as well as residues I79, D80, S83, G84, T85, and V86 that immediately precede the active site. A second site of interaction, although less contiguous, emanates from the 3-turn helix immediately after the active site serine and extends towards the first 2 turns of the “crossover helix”, helix 3. This surface although not identical, more closely resembles the active conformation for canonical lysine transferring E2s, where the hydrophobic I44 surface of Ub forms a noncovalent interaction with helix 3. Finally, upon conjugation, a third set of perturbations occurs on the “backside” β-sheet of Ube2h (**Figure 6c**). The surface extends through a contiguous region that begins in

strand 2 and ends in strand 4, close to the Ube2h active site. This particular site of interaction is intriguing and may represent an important conformational state of the Ube2h-151Δ~Ub conjugate. Complete assignments for the conjugate must be made for this species to confirm these preliminary results.

Discussion

In this chapter we confirmed previously published results that Ube2h ubiquitinates histone H2A/H2B in the absence of an E3[5]. Furthermore, we demonstrate that there are two functionally distinct domains and that the C-terminal region of Ube2h, which is intrinsically disordered, is responsible for mediating interactions with the substrate. Acidic residues in this region, particularly, E153, E158, D163, E175, D176, and E177, drive the interaction between Ube2h and the histone H2A/H2B dimer (**Figure 4a, Figure S1**). Mutation of three of these residues, E175, D176, and E177 to alanine greatly diminishes binding between the C-terminus of Ube2h and histone H2A/H2B (**Figure 4b**). Our results also show that the disordered C-terminus interacts with the UBC domain of Ube2h on the "backside" β-sheet. The functionality of this interaction is yet to be defined. Finally, upon conjugation to Ub, Ube2h-151Δ shows distinct sites of interaction in the 3-turn helix immediately following the active site, the "crossover helix", helix-3, and intriguingly, on the "backside" β-sheet.

Although we have a basic understanding of the biochemistry of Ube2h, several structural and functional questions remain regarding this E2. Further biochemical and structural work needs to be done to fully understand the interaction between the C-terminus of Ube2h and the histone H2A/H2B dimer. NMR buffer conditions must be tested to circumvent protein precipitation issues. Furthermore, structural studies must be applied to biochemical assays to confirm their affect on ubiquitination activity. Another important aspect to consider is the relevance of this interaction in cells. To date, there have been no *in vivo* experiments to confirm that Ube2h is indeed an E2 that interacts with histone H2A/H2B or that it is localized to the nucleus. Studies presented in this chapter build a foundation to understand the capabilities of this E2 functionally and structurally, but do not address basic questions regarding its role in cells. In support of its relevance, however, Ube2h is partially conserved in both *S. cerevisiae* and *C. elegans*, particularly in the C-terminal region between the region of S168 and D176 there may be some functional relevance to the interaction seen with histone H2A/H2B. Clearly it is interesting that Ube2h is

capable of robustly ubiquitinating a substrate in the absence of an E3 *in vitro*, yet it is known that in a cellular context, Ube2h does form interactions with cognate RING-type E3s. Multiple published reports suggest that in cells the RING E3, mitsugumin-53 (MG53) is a Ube2h-interacting E3[7,8].

MG53 is a 53kDa tripartite motif-containing protein (TRIM72) that contains an N-terminal RING domain followed by a second zinc-chelating B-Box domain, which is thought to be a protein-protein interaction motif. This E3 ligase is muscle specific and is known to be important in wound repair and has recently been implicated as an important negative regulator of myogenesis[7-9]. Currently, there are two known substrates, insulin receptor substrate 1 (IRS-1) and focal adhesion kinase (FAK) that are ubiquitinated in an MG53/Ube2h-dependent manner[7,8]. IRS-1 is a positive regulator of myogenesis and FAK is a scaffolding tyrosine-kinase that also regulates skeletal myogenesis. These substrates were confirmed as Ube2h targets as siRNA knockdown of the E2 resulted in decreased levels of substrate ubiquitination in HEK-293 cells[7,8]. Initially, the interaction between Ube2h and the RING domain of MG53 should be explored, as Ube2h is not activated by canonical RING-type E3s such as BRCA1/BARD1. In canonical RINGs a conserved basic residue allosterically activates E2~Ub conjugates to transfer ubiquitin to substrate. MG53 lacks this basic residue and instead has a glutamine at this position (**Figure S3**). Studying this E2-E3 pair could provide new mechanistic insights into RING-type activation of E2s. Interactions between Ube2h and the substrates IRS-1 and FAK should also be explored, as these are bona fide cellular substrates. Understanding if the C-terminus of Ube2h plays a similar role to histone H2A/H2B substrate recruitment with IRS-1 and FAK could prove fruitful.

Studies of the Ube2h-151Δ~Ub conjugate should also be pursued. Comparison of conjugated Ube2h-151Δ and unconjugated Ube2h-151Δ reveal some features consistent with canonical E2~Ub conjugates. For example, perturbations near the active site and in the “crossover helix”, helix 3 are consistent with previously published reports of E2~Ub conjugates, however a third surface on the “backside” β-sheet is rather unexpected[1-3]. Although this surface could be do to an intermolecular interaction, the result may also suggest that when attached to the active site of Ube2h-151Δ, Ub samples conformations that interact with the “backside”, particularly in β-strands 2, 3, and 4. The implications of Ub sampling conformations towards this region are unknown, but could be functionally relevant, as Ube2h is not activated by canonical RING-type ligases. The Ube2h-151Δ should be assigned and small angle X-

ray scattering (SAXS) experiments should be considered in order to understand the flexible nature of the conjugate. Finally, studies of full-length Ube2h~Ub should also pursued. There is potential that the C-terminus of Ube2h may make noncovalent interactions with Ub when it is attached to the active site. Because the disordered C-terminus is 31 residues, theoretically it could form interactions with a conjugated Ub. Clearly, this must be tested before any conclusions can be made.

Currently, there is a paucity of published biochemical and structural work on Ube2h, but with the emergence of confirmed cellular protein-protein interaction partners, this E2 represents an exciting avenue of research. Ube2h possesses non-canonical behavior on two fronts. First, from the perspective of understanding how this E2 interacts with its cognate E3s, initial evidence suggests Ube2h may harbor a unique activation mechanism. Second, the potential of Ube2h to directly interact and ubiquitinate substrates is unconventional. Biochemical and structural insights garnered from the *in vitro* substrate histone H2A/H2B should be expanded towards IRS-1 and FAK.

Materials and Methods

Plasmids, Protein Expression, and Purification

All Ube2h constructs (full-length Ube2h, Ube2h-151Δ, and Ube2h-152-183) wer expressed from the pET28a (Novagen) vector with an N-terminal thrombin cleavable HIS_6 tag. Ube2h constructs were transformed into *Escherichia coli* (BL21 DE3) cells and protein expression was induced with 0.4mM isopropyl-β-D-thio-galactoside (IPTG) at OD_{600} of 0.6, followed by growth at 16°C for 16hrs. Cells were lysed by French press in 25mM Tris-HCl (pH 7.6), 200mM NaCl, 10mM imidazole and applied to a Ni^{2+}-NTA gravity flow column (Invitrogen) equilibrated in the same buffer. The column was washed with 5 column volumes of 25mM Tris-HCl (pH 7.6), 200mM NaCl, 50mM imidazole and eluted with 25mM Tris-HCl (pH 7.6), 200mM NaCl, 500mM imidazole. The elution was subject to cleavage at 4˚C overnight in 25mM Tris-HCl (pH 7.6), 200mM NaCl with 1mg thrombin from bovine serum (Sigma-Aldrich). Following cleavage the dialyzed sample was re-applied to a Ni^{2+}-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 25mM sodium phosphate (pH 7.0), 150mM NaCl. Wheat Uba1, human Ub, human HA-Ub, and *X. laevis* histone H2A/H2B were purified as previously described[10,11].

Histone Ubiquitination Assays

Reactions for histone H2A/H2B ubiquitination assays contained 20uM HA-Ub, 0.5uM wheat Uba1, 2uM Ube2h or Ube2h-151Δ, 4uM histone H2A/H2B, and 10mM $MgCl_2$. Reactions were initiated by the addition of 5mM ATP and incubated at 37°C. Samples were collected at 0 and 60 min after the addition of ATP. Products were visualized by western blotting for HA-Ub.

NMR Spectroscopy

All two-dimensional NMR spectra were collected on a 500MHz Bruker Avance II (University of Washington). All three-dimensional spectra were collected on a Bruker Avance III spectrometer equipped with a cryoprobe and operating at 600 MHz (University of Washington). All spectra for full length Ube2h and Ube2h-151Δ were recorded at 25°C in 25mM sodium phosphate (pH 7.0), 150mM NaCl, 10% D_2O. Spectra for Ube2h-152-183 were collected at 15°C in 25mM sodium phosphate (pH 6.2), 250mM NaCl, 10% D_2O. Three-dimensional data sets contained 600uM of ^{15}N – ^{13}C-isotopically labeled protein while two-dimensional data sets for titrations contained 100-200uM of ^{15}N-labeled protein. Data were processed using NMRPipe/NMRDraw[12] and visualized with NMRView[13]. Chemical shift perturbations were calculated using the formula $\Delta\delta j = [(\Delta\delta j\ ^{15}N/5)^2 + (\Delta\delta j\ ^{1}H)^2]^{1/2}$.

References

1. 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).

2. 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).

3. 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).

4. Haldeman, M.T., Xia, G., Kasperek, E.M., Pickart, C.M. Structure and function of ubiquitin conjugating enzyme E2-25K: the tail is a core-dependent activity element. *Biochemistry*. **36**, 10526-10537 (1997).

5. Kaiser, P., Mandl, S., Schweiger, M., Schneider, R. Characterization of functionally independent domains in the human ubiquitin conjugating enzyme UbcH2. *FEBS Lett*. **377**, 193-196.

6. Sheng, Y, Hong, JH, Doherty, R, Srikumar, T Shloush, J, Avvakumov, GV, et al (2012) A Human Ubiquitin Conjugating Enzyme (E2)-HECT E3 Ligase Structure-function Screen. *Mol Cell Proteomics*. 2012. **11**, 329-341.

7. Yi, J.S., et al. MG53-induced IRS-1 ubiquitination negatively regulates skeletal myogenesis and insulin signaling. *Nat. Commun*. 2013. **4**, 2354.

8. Mitsugumin 53 (MG53) ligase ubiquitinates focal adhesion kinase during skeletal myogenesis. *J. Biol. Chem*., 2014, **289**, 3209-3216.

9. Jia, Y., et al. Treatment of acute lung injury by targeting MG53-mediated cell membrane repair. *Nat Commun.* 2014, **5**, 4387.

10. Pickart, C.M. and Raasi, S. Controlled synthesis of polyubiquitin chains. *Methods Enyzmol.* **399**, 21-36 (2005).

11. Gloss, L.M., Placek, B.J. The effect of salts on the stability of the H2A-H2B histone dimer. *Biochemistry*. 2002, **41**,14951–14959

12. Delaglio, F, Grzesiek, S, Vuister, GW, Zhu, G, Pfeifer, J, Bax, A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 1995. **6**, 277-293.

13. Johnson, BA, Blevins, RA (1994) NMR View: a computer program for the visualization and analysis of NMR data. *J Biomol NMR*. 1994. **4**, 603-614

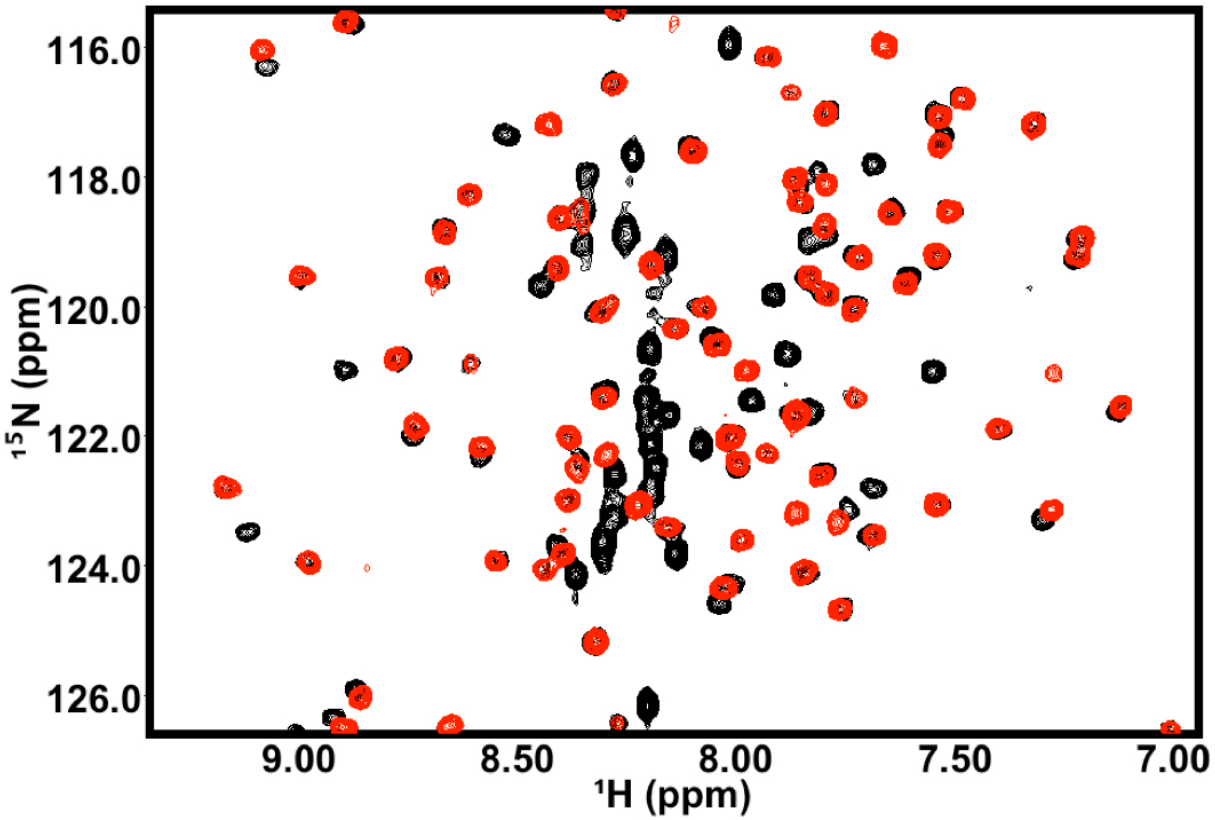


Figure 1. The C-terminus of Ube2h is intrinsically disordered. The ^{1}H, ^{15}N-HSQC spectrum of full length Ube2h (black) contains high intensity, clustered peaks between 8.0 and 8.5ppm in the hydrogen dimension that are characteristic of intrinsically disordered regions. A ^{1}H, ^{15}N-HSQC spectrum of truncated Ube2h-151Δ does not contain this peaks, indicating the disordered region of Ube2h is between residues 152 and 183.

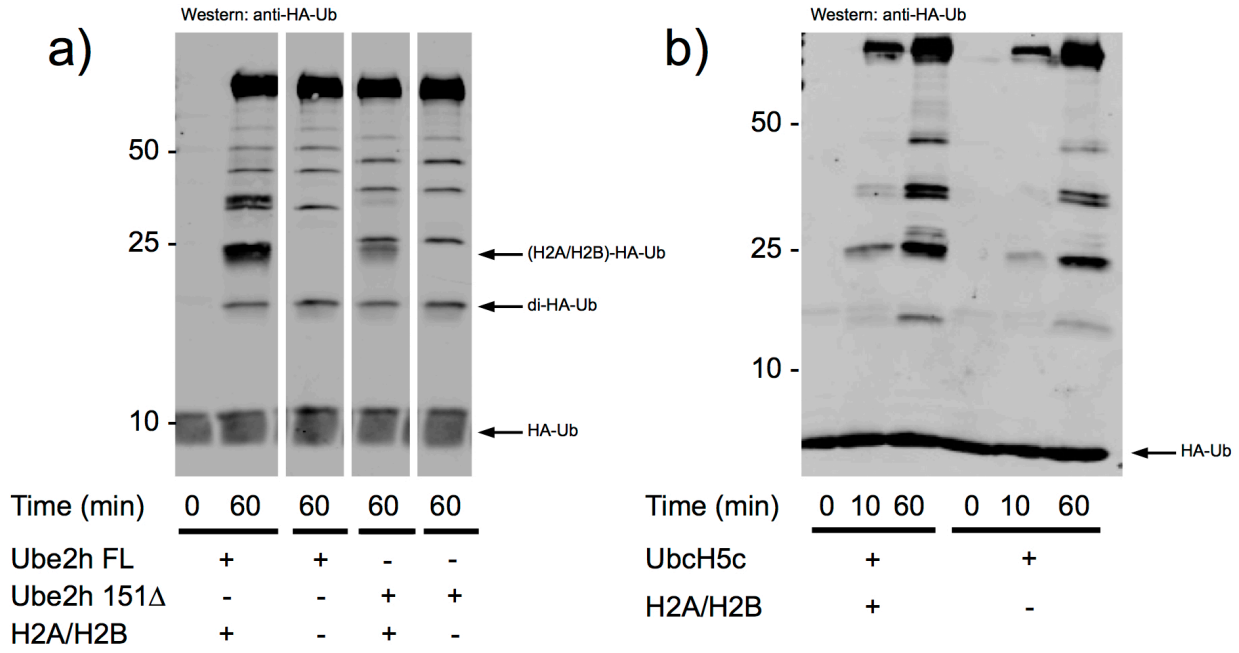


Figure 2. Histone H2A/H2B is ubiquitinated by Ube2h dependent on its C-terminus. (a) In a 1hr reaction, full length Ube2h is capable of robustly ubiquitinating Ube2h as visualized by a western blot for HA-Ub. Panel 1 shows a robust band below 25kDa that is consistent with either HA-Ub-H2A or HA-Ub-H2B. This band disappears when histone H2A/H2B is omitted from the reaction (panel 2). When the mutant Ube2h-151Δ is substituted in the reaction H2A/H2B ubiquitination is severely decreased (panel 3). Omitting H2A/H2B from this reaction reveals that Ube2h-151Δ is still intrinsically reactive (panel 4). (**b)** Substituting UbcH5c in this reaction for Ube2h demonstrates that H2A/H2B ubiquitination is specific to Ube2h as no ubiquitinated forms of the substrate are detected.

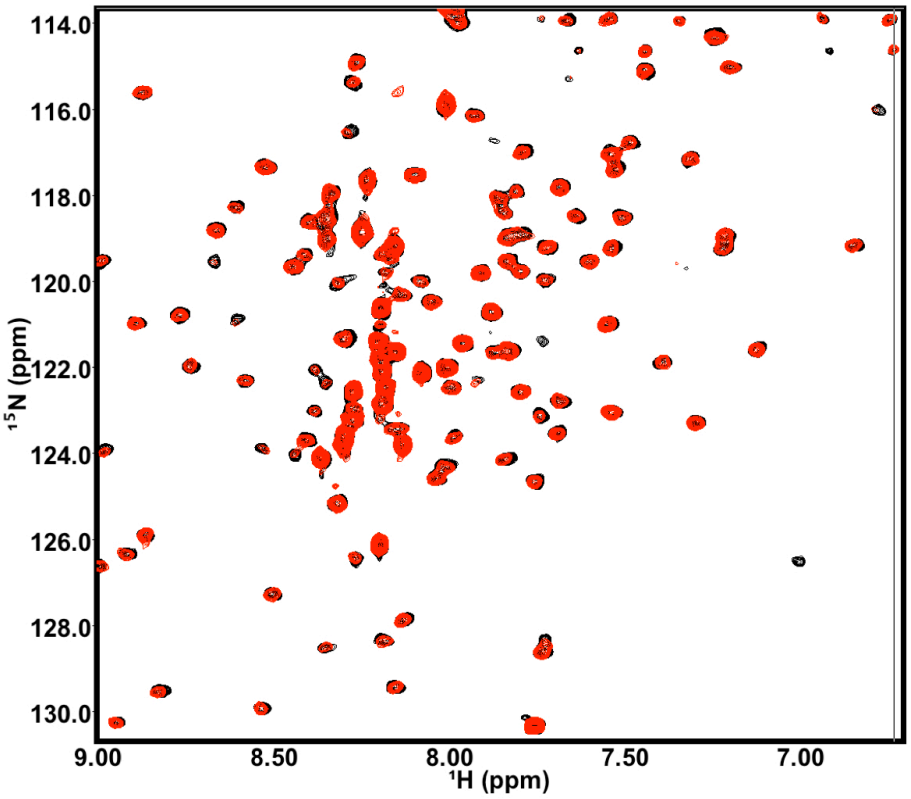


Figure 3. Addition of histone H2A/H2B to full length Ube2h causes global broadening. Upon the addition of .1 molar equivalent of histone H2A/H2B (red) the ^{1}H, ^{15}N-HSQC spectrum of full length Ube2h (black) shows global peak broadening effects.

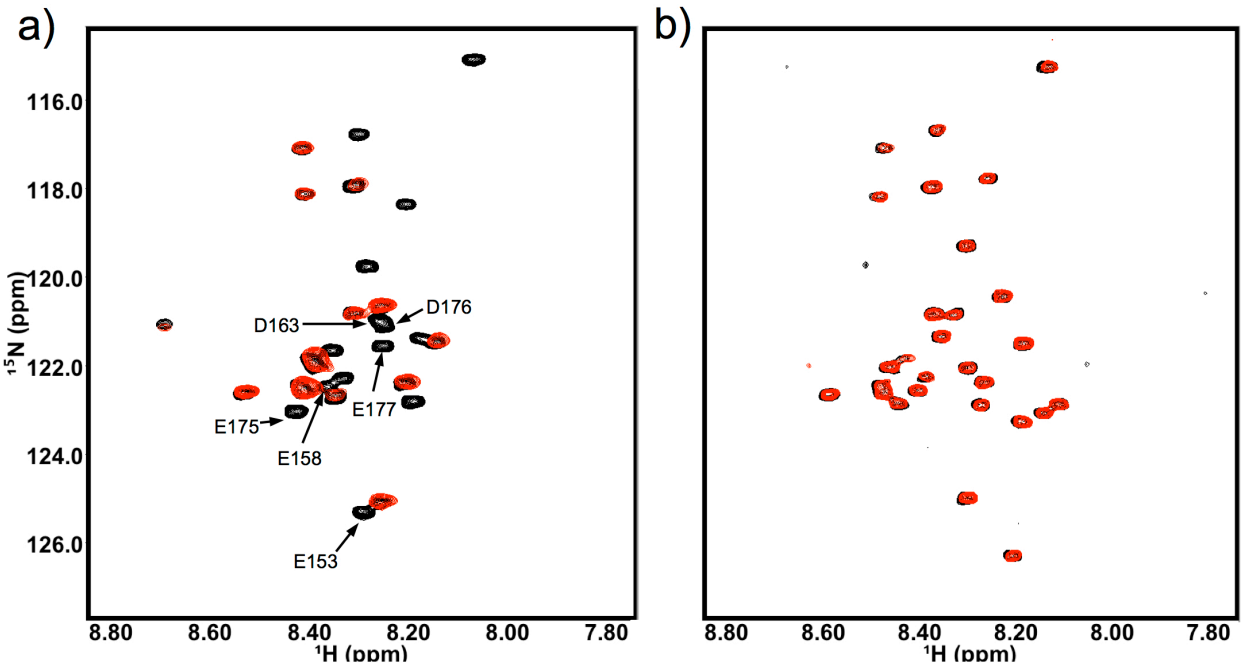


Figure 4. Histone H2A/H2B interacts with the C-terminus of Ube2h. (a) Spectrum of the ^{1}H, ^{15}N-HSQC of Ube2h-152-183 (black) overlaid with the spectrum of Ube2h-152-183 with the addition of .25 molar equivalents of histone H2A/H2B (red). Specific broadening of acidic residues is detected. **(b)** ^{1}H, ^{15}N-HSQC spectrum of Ube2h-152-183-EDE-AAA (black) overlaid with the spectrum of Ube2h-152-183-EDE-AAA following the addition of .25 molar equivalents of histone H2A/H2B. Specific broadening of acidic residues is decreased.

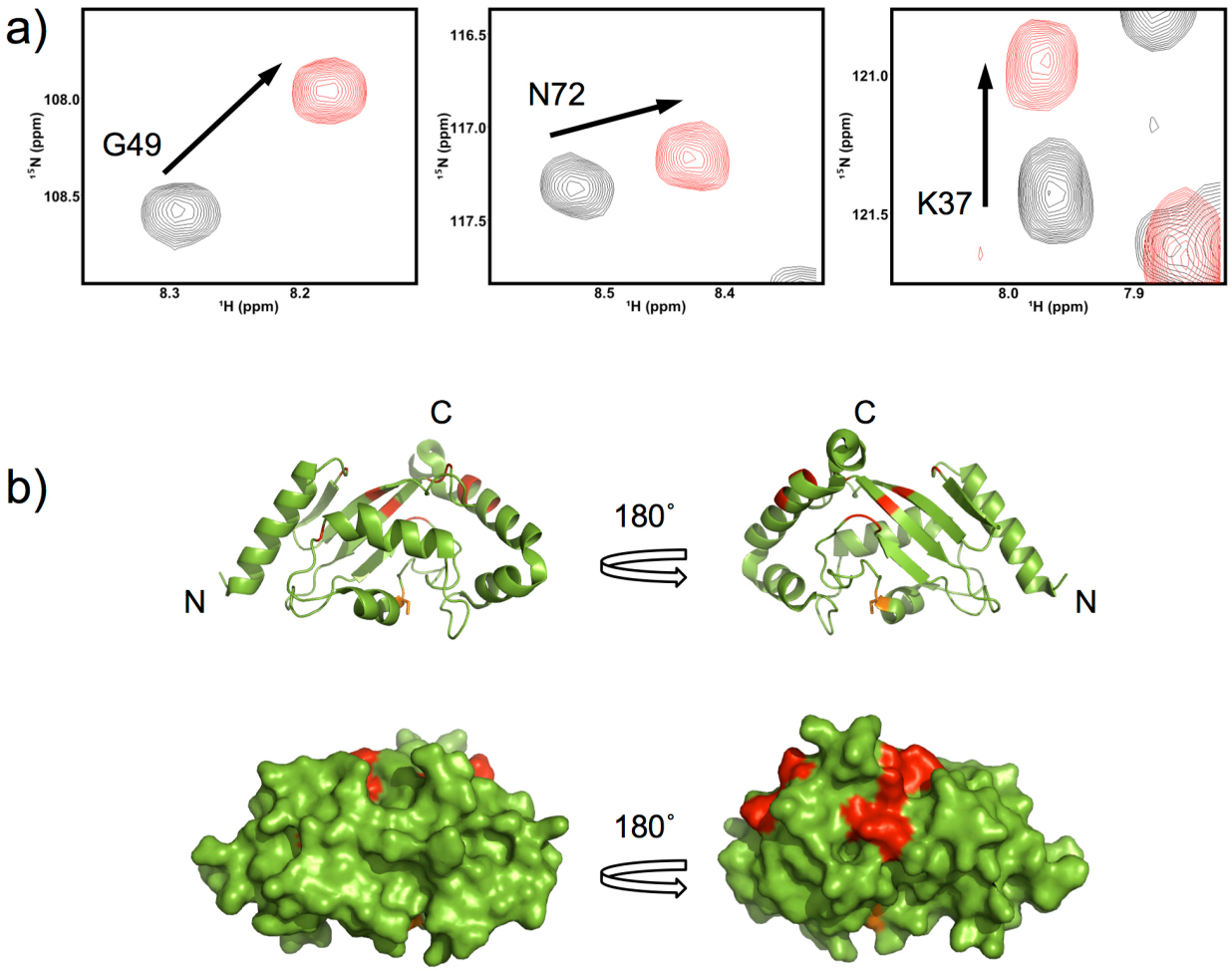


Figure 5. The Ube2h C-terminus forms interactions with the core UBC domain. (a) Selected residues in the ^{1}H, ^{15}N-HSQC spectrum of Ube2h-151Δ (red) that shift following truncation of full length Ube2h (black) **(b)** Crystal structure of the Ube2h UBC domain (green) with the interaction surface of the C-terminus colored in red. Two main sites of interaction are observed: 1) helix 5 near the site of truncation and 2) the "backside" β-sheet.

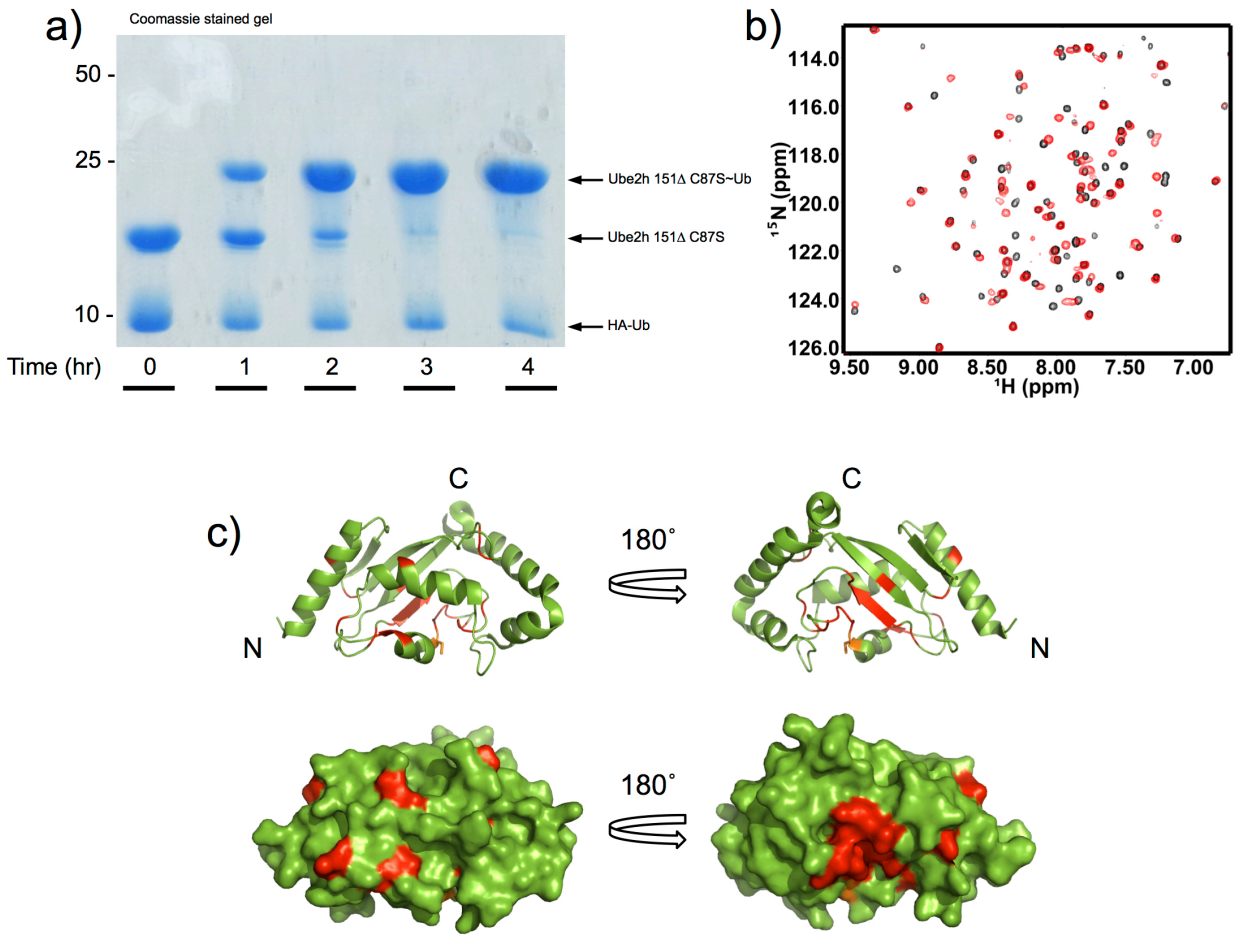


Figure 6. Understanding the Ube2h-151Δ-O~Ub conjugate. (a) Ub loading assay for Ube2h-151Δ_{C87S}. Following 1hr of incubation Ube2h-151Δ_{C87S} begins to form an oxyester intermediate with Ub. After 4hrs, almost 100% loading of the E2 is observed. **(b)** Overlay of the ^{1}H, ^{15}N-HSQC spectrum of Ube2h-151Δ (black) with the ^{1}H, ^{15}N-HSQC spectrum of the Ube2h-151Δ-O~Ub conjugate (red). Several spectral changes are observed. **(c)** Mapping of chemical shifts following Ube2h-151Δ_{C87S} conjugation to ubiquitin. Perturbations are observed near the active site and on the "crossover helix". There are also distinct perturbation on the "backside" β-sheet, near the active site.

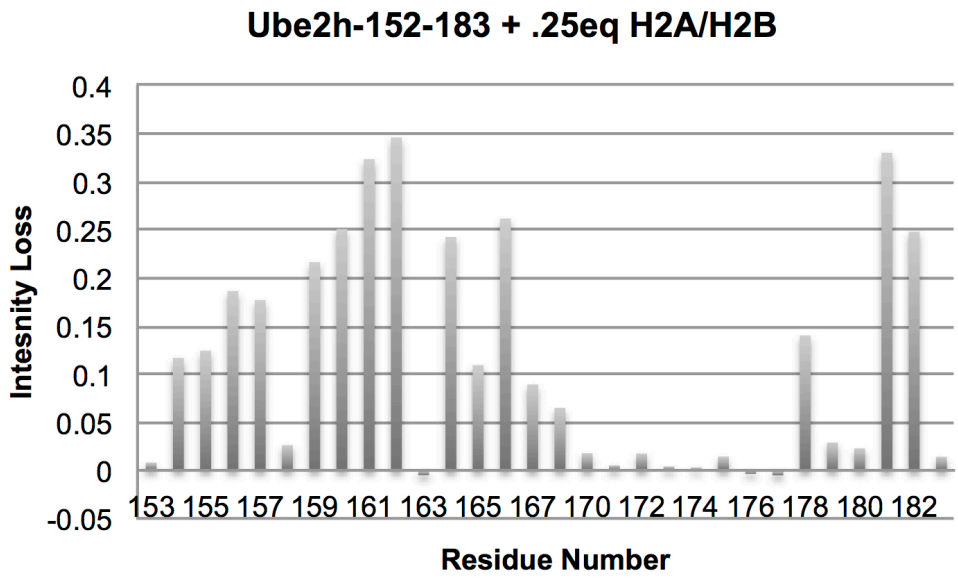


Figure S1. Histone H2A/H2B – Ube2h-152-183 titration analysis. Intensity loss analysis following the addition of .25 molar equivalents of histone H2A/H2B into ^{15}N-Ube2h-152-183.

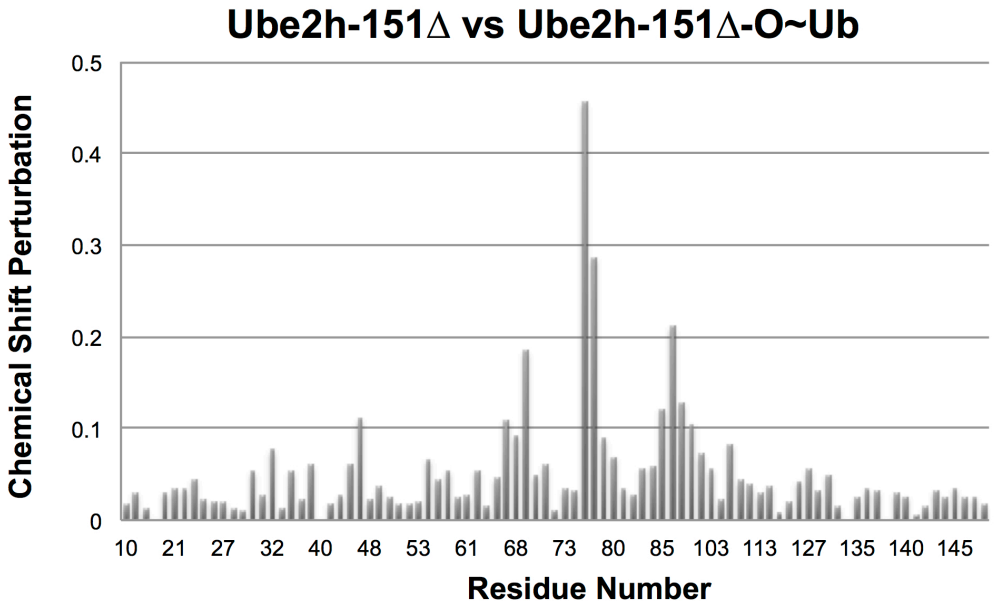


Figure S2. Comparing Ube2h-151Δ and the Ube2h-151Δ-O~Ub conjugate. Chemical shift perturbations when comparing the Ube2h-151Δ and the Ube2h-151Δ-O~Ub conjugate are mapped.

```
MG53            MSAAPGLLHQELSCPLCLQLFDAPVTAECGHSFCRACLGRVAGEPAADGTVLCPCCQ
BRCA1           VQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQ--KKGPSQCPLCK
                :. . . ::: *.**:**:*:. **:::*.* **: *: :: .:   .*.  ** *:
```

Figure S3. Sequence alignment of the RING domain of MG53 and BRCA1. Sequence alignment of the MG53 RING domain with the BRCA1 RING domain shows that MG53 does not contain a conserved basic residue (last residue of sequence).