Published in final edited form as:

J Mol Biol. 2016 February 22; 428(4): 679–687. doi:10.1016/j.jmb.2015.09.011.

UbSRD: The Ubiquitin Structural Relational Database

Joseph S. Harrison[1,2], **Tim M. Jacobs**[1], **Kevin Houlihan**[1], **Koenraad Van Doorslaer**[3], and **Brian Kuhlman**[1,2]

[1]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[3]DNA Tumor Virus Section, Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Abstract

The structurally defined ubiquitin-like homology fold (UBL) can engage in several unique protein–protein interactions and many of these complexes have been characterized with high-resolution techniques. Using Rosetta's structural classification tools, we have created the Ubiquitin Structural Relational Database (UbSRD), an SQL database of features for all 509 UBL-containing structures in the PDB, allowing users to browse these structures by protein–protein interaction and providing a platform for quantitative analysis of structural features. We used UbSRD to define the recognition features of ubiquitin (UBQ) and SUMO observed in the PDB and the orientation of the UBQ tail while interacting with certain types of proteins. While some of the interaction surfaces on UBQ and SUMO overlap, each molecule has distinct features that aid in molecular discrimination. Additionally, we find that the UBQ tail is malleable and can adopt a variety of conformations upon binding. UbSRD is accessible as an online resource at rosettadesign.med.unc.edu/ubsrd.

Keywords

ubiquitin; SUMO; Rosetta; structural database; protein–protein interaction

Introduction

Discrimination between the ubiquitin-like homology folds (UBL) is required for many physiological processes, most notably discrimination between ubiquitin (UBQ) and small ubiquitin-related modifier (SUMO) [1–3]. UBQ post-translation modifications can elicit a wide variety of responses (i.e., proteasome degradation, cellular trafficking, signaling transduction, and altering protein activity) and the response to SUMO is also varied; however, SUMO functions are primarily nuclear (UBQ- and SUMO-specific functions are summarized in these two references [4,5]). Several proteins and protein complexes, such as

Correspondence to Joseph S. Harrison: Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

PCNA (Proliferating Cell Nuclear Antigen), PML (Promyelocytic Leukemia Protein), PTEN (Phosphatidylinositol 3,4,5-Trisphosphate 3-Phosphatase and Dual-Specificity Protein Phosphatase), and Fanconi Anemia complex are modified by both UBQ and SUMO, regulating critical biological processes, such as DNA replication and repair, nuclear/cytoplasmic shuttling, and protein abundance [6–10]. Importantly, conjugation of UBQ and SUMO can result in disparate protein responses, as is the case for PCNA [9]. The versatility of the biological response elicited by UBLs is due in part to their ability to participate in a multitude of protein–protein interactions (PPIs), many of which are unique to the UBL family of proteins: the C-terminus can be covalently conjugated as a thioester to an active-site cysteine of activating proteins, which include E1, E2s, and E3s, or as an isopeptide bond to a lysine on a target protein (for an in-depth review of UBL conjugation, see Ref. [11]); covalently linked UBLs can be removed by specific proteases called deubiquitinases and the protease reaction coordinate often proceeds through a thioester intermediate [12]; both monomeric and geometrically distinct polymeric UBL chains can be specifically recognized by a variety of interacting partners [13]; and many UBLs are domains of larger proteins and are implicated in self-regulation or association with partners [14].

The biological importance of UBLs has resulted in extensive efforts to structurally characterize UBL interactions. These efforts have resulted in co-crystal structures of UBQ and SUMO with a number of interaction partners. However, considering that multiple lysines on virtually every cellular protein can be conjugated with an assortment of distinct UBQ chains, the potential number of conjugated UBQ interfaces is very large. Additionally, structural characterization using experimental methods is difficult due to the heterogeneous nature, weak affinities, transient lifetimes, and dynamic properties of these complexes. For these reasons, computational modeling of UBL interactions represents an attractive avenue for future structural characterization.

Computational protein modeling can be aided by implementing knowledge-based scoring metrics obtained from native protein structures. Analysis of measurable properties from large sets of molecular conformations is challenging, and to this end, structural relational databases are powerful tools [15]. The recent addition of structural analysis tools to the Rosetta molecular modeling package provides a framework to record various features for large sets of protein structures [15,16]. Here we use this methodology to create the Ubiquitin Structural Relational Database (UbSRD), which contains a manually curated set of all the UBQ homology folds in the PDB grouped by PPIs. The motivation of this study is to provide a structural dictionary of UBL-containing structures allowing rapid search of this set. As a demonstration of the capabilities of UbSRD, we explored trends of UBQ and SUMO recognition in the PDB.

UBL evolutionary relationship

The UBL fold is composed of a central β-sheet, containing five β-strands packed against a three-turn α-helix (Fig. 1a and Fig. 1 in Ref. [17]) [18]. There is an additional α-helical segment in the loop connecting β4 and β5. The UBL fold can accommodate many diverse sequences, for example, UBQ and SUMO1 have only 22% sequence similarity using the PAM-250 matrix (Fig. 1 in Ref. [17]). To cluster the UBL domains in UbSRD, we

calculated a maximum likelihood tree of all identified UBLs using a structure-based multiple sequence alignment [19,20]. Several UBLs were distantly related (ATG12, UFM1, ElonginC, URM, and UBLs from PDB codes 2BPS, 2YLM, 1WF9, and 2DZM) from the rest of the phylogenetic tree, which is clearly demarcated between SUMO-like and UBQ-like UBLs (Fig. 2 in Ref. [17]). We have named sub-families according to prominent members of each branch, and our results are generally consistent with other proposed evolutionary histories [21]. Of note, UBQ has the shortest branch length, indicating that UBQ is closest to the putative ancestor sequence. This observation is supported by the nearly identical UBQ amino acid sequence between humans and yeast and suggestive of evolutionary model where UBL domain diversification arose from UBQ gene duplications.

UBQ interaction surface

The Rosetta FeaturesReporter protocol provides a framework to quantitatively record measurable properties of protein structures, and for UBQ, we analyzed the number of non-UBQ amino acid neighbors, inter-molecular hydrogen bonds, and secondary structure of interacting motifs [16]. This analysis revealed seven positions that make up the binding interface that we have named the canonical binding surface (Leu8, Arg42, Ile44, Gly47, His68, Leu70, and Leu73) (Figs. 1a and b and 3 in Ref. [17]), which includes the hydrophobic patch (Leu8, Ile44, and Val70) [22,23]. Several of these residues participate in hydrogen bonding: the Arg42 side chain is a prominent hydrogen bond donor as are the backbones of Leu8, Gly47, and Leu73 (Figs. 1a and c and 5 in Ref. [17]). The canonical binding surface has a preference to contact α-helical motifs; however, Gly47 and Leu73 have an elevated preference to interact with loops and β-sheets (Fig. 5 in Ref. [17]). Structures in UbSRD are grouped by the type of PPI, for instance, conjugated or deubiquitinase, which allowed identification of less prominent binding surfaces of UBQ, such as the alternative epitope, spanning residues 35–40, that is buried in many conjugated interfaces. A distinctive feature of this interface is a Gln40 hydrogen bond found at both HECT and RING E3 ligase interactions (Figs. 1a and c and 5 in Ref. [17]) [24,25]. Deubiquitinases can utilize two additional patches of UBQ, spanning residues 10–15 and 62–66, and the proximity of these surfaces to UBQ lysine acceptors may provide specificity for distinct UBQ chains (Fig. 3 in Ref. [17]). Recently, Ser65 phosphorylation, an important post-translation modification to UBQ, has been shown to inhibit some deubiquitinases' activity, providing an additional functional role for recognition of residues 62–66 [26]. Notably, the UBQ interaction surfaces on deubiquitinases are not α-helical and are largely composed of loops and strands (Fig. 4 in Ref. [17]). UBQ forms specific hydrogen bonds in different types of PPIs, for example, the backbone of Gly47 is a hydrogen bond donor at many noncovalent interfaces and the side chain of Gln49 is a hydrogen bond donor for some deubiquitinase interactions (Fig. 5 in Ref. [17]). However, the most distinctive feature of deubiquitinase and conjugated interfaces is extensive contacts and hydrogen bonds to the UBQ tail (71–76) that are not observed at most noncovalent interfaces (Fig. 3 in Ref. [17] and Fig. 5 in Ref. [17]). In particular, we observe the backbone of the UBQ tail participates in hydrogen bonds to interacting motifs that are loops and strands, hallmarks of inter-molecular β-strand pairing (Fig. 4 in Ref. [17] and Fig. 5 in Ref. [17]).

Surprisingly, interactions to the backside of UBQ, encompassing α1 and β2, were almost nonexistent in structures in the PDB (Fig. 1c, lower panel). This is surprising since strand pairings β2 are frequently observed for SUMO interactions and other member of the ubiquitin-like fold super-family, like RBD, PIP3K, and Ral-GDS [18]. We hypothesize this may be a primary surface for discrimination between UBQ and other UBLs. Two plausible explanations for how UBQ may avoid edge–edge β-strand pairings: (1) the side chain of Lys29 in α1 may protect the β2 strand by steric occlusion, SUMO variants encode the less bulky Ala or Ser, and (2) the flexibility of the β1–β2 loop may hinder edge–edge strand pairing [27,28]. It is important to note that the three variations between human and yeast UBQ sequences occur in β2α1 and variations to this region are viable in yeast, unlike mutations to any of the seven residues that make up the hydrophobic surface [29]. However, structures of the HOIP and HOIL domains of the E3 ligase that synthesizes linear ubiquitin chains do contact β2α1, albeit not through strand paring to β2, suggesting that this portion of the protein may aid in the formation and recognition of distinct ubiquitin polymers [30,31].

We also used UbSRD to analyze amino acid preferences for residues that interact with the canonical binding surface. We find a preference for the smaller hydrophobic residues Ile, Leu, Val, and Ala, which is most prominent for Leu8, Ile44, and Val70 (Fig. 1d). Another, general feature is that interactions with positive residues such as Arg and Lys are disfavored, likely due to the overall cationic charge on the canonical binding surface. Although not all the residue preferences are hydrophobic, Arg42 and to a lesser extent His68 have elevated preferences toward anionic residues, and Leu8 has an elevated number of Gln partners that can hydrogen bond with the Leu8 backbone. Leu73 has a surprising number of Tyr neighbors (see Leu8 for comparison), which are observed at a number of deubiquitinase and noncovalent interfaces (Fig. 1d).

SUMO interaction surface

The SUMO protein sequence, unlike UBQ, is not highly conserved between *Saccharomyces cerevisiae* and humans; humans also have four SUMO isoforms while *S. cerevisiae* has only one. SUMO has two distinct surfaces that contact partners, the hydrophobic surface and the β2α1 groove (Figs.2a and 6 in Ref. [17]). SUMO utilizes the canonical binding surface less extensively than UBQ and sites in this region have fewer neighbors per structure; in particular, positions 29 and 91 of SUMO are under-utilized relative to homologous UBQ positions (Fig. 1 in Ref. [17]). We hypothesize that this may be due to the aromatic residues found at position 91 preventing extensive contacts with the canonical binding surface. This is highlighted by the SUMO protease/SUMO interface that only has neighbors to the C-terminal half of SUMO (Fig. 6 in Ref. [17]) [32]. Additionally, the motifs that contact the canonical binding surface of SUMO primarily are not α-helical, providing another means of discrimination between SUMO and UBQ interacting proteins (Fig. 7 in Ref. [17]).

SUMO variants frequently have interactions with the β2α1 groove at conjugated and noncovalent interfaces, but not with SUMO proteases (Fig. 2a) [33,34]. These interactions are characterized by backbone hydrogen bonds to positions 33, 35, and 37 to primarily β-strand motifs (Figs. 2b and c, 7, and 8 in Ref. [17]). Also, residues 42, 46, and 50 of α1 contact interacting proteins and the side chains of these residues participate in hydrogen

bonds (Fig. 2b and c and 8 in Ref. [17]). As mentioned above, this surface of UBQ has not been characterized interacting in this manner, suggesting that the β2α1 groove may provide specificity between the two folds.

We also observe extensive contacts to the SUMO tail at both conjugated and deubiquitinase interfaces, but not at noncovalent interfaces (Fig. 6 in Ref. [17]). As we observed for UBQ interfaces, there are extensive backbone hydrogen bonds to the tail indicating that this mode of interaction is conserved between the two UBLs, implicating that the general mode of tail recognition is similar for UBQ and SUMO despite the difference in amino acid identity (Figs. 1 and 8 in Ref. [17]). Several other sites of SUMO can participate in hydrogen bonds, most notably residues 60, 67, and 70. In particular, the side chains of several residues in the C-terminal portion of SUMO function as hydrogen bond acceptors at many interfaces, indicating that electrostatic differences between UBQ and SUMO aid in discrimination.

Plasticity of the UBQ tail

To explore the conformational freedom of the UBQ tail, we constructed Ramachandran plots for all the deubiquitinase and conjugated complexes. We find that the ϕ/φ angles of the UBQ tail primarily reside within the β-strand boundary; however, the backbone conformational freedom increases toward the C-terminus (Fig. 3a). Beginning with Arg72, the UBQ tail residues can adopt a diverse set ϕ/φ angles found within the βs, βp, γ-turns, and α-helical conformations [35]. This trend is particularly evident for Leu73 where the UBQ tail can adopt both L-handed and R-handed α-helices. Recently, it has been discovered that deubiquitinases cannot cleave conjugated UBQ variants containing Leu73Pro substitutions and this may be due in part to restrictions that Pro would impose on the backbone conformation of the UBQ tail [36].

We also classified the UBQ tail conformation by the family of interacting proteins (see the methods section). We observe that families of interacting proteins tend to recognize similar tail conformations, although a variety of tail conformations can exist between families of proteins. Leu73 exemplifies this observation: all four characterized E2/RING E3 complexes have Ramachandran angles in the β-region, but for HECT E3 ligases, the backbone torsion angles are within the α-helical boundary. For OTU, Leu73 ϕ/φ resides in both the R-handed and L-handed α-helices, and for UCH UBQ complexes, Leu73 is found within the γ-turn region of Ramachandran space. The differences in the UBQ tail backbone conformations when coordinated to deubiquitinase and conjugated proteins provide an additional layer of specificity.

UbSRD, the Web resource

One of the motivating factors of this study was to create a tool broadly accessible to researchers that can be easily updated as the PDB expands. To this end, we created a Web site for simple browsing of UbSRD. Users can browse the structures by phylogeny, PPI, UBL type, UBQ chain linkage, and the partner identity (for conjugated and deubiquitylase interfaces). Also users can search for amino acid partners to UBQ and SUMO residues. The entire SQL database, the Newick format of the phylogenetic tree, and the multiple sequence alignment used to construct the phylogram are available for download. When creating

UbSRD, we collected many more features than we included in our analysis, such as crystallographic waters, side-chain rotamers, and global data about protein interfaces and there are example queries available to help users get started with querying UbSRD. Moreover, the customizable nature of the Rosetta features reporter framework allows for the simple generation of additional features[†].

Discussion

Here we present UbSRD, an SQL database of Rosetta-derived structural features from all UBL domains we identified in the PDB. Each UBL-containing structure is classified by type of protein interface, which allowed us to examine features of UBQ and SUMO recognition while engaged in different types of interactions. The most notable difference is that interactions with the β2α1 groove of UBQ are rare; conversely, for SUMO, this region is a primary recognition surface [37]. UBQ interactions rely heavily on the canonical binding surface, which includes Leu8 of the β1–β2 loop; SUMO infrequently uses this loop for binding. The UBQ canonical binding surface is extensive and we were surprised by the extent that Leu73 was utilized in all types of PPIs compared to other residues in the UBQ tail. Moreover, the quantity of inter-molecular hydrogen bonds to the canonical binding surface is also surprising, as Leu8, Arg42, Gly47, and Leu73 are all hydrogen bond sites; the hydrophobic patch is often discussed as the primary determinant of UBQ binding in the literature. A commonality in both UBQ and SUMO binding is the role of Arg42/Arg63. Interestingly, Arg42/63 and Gly47/68 are the only identical residues in the canonical binding surfaces of UBQ and SUMO, but unlike Arg42/Arg63, Gly47/Gly68 is utilized differently between these two molecules; we did not observe hydrogen bonds to position 68 of SUMO but instead observe them to the preceding residue at deubiquitinase and some conjugated interfaces. This distinction may aid in discrimination of β3 between the two molecules.

A striking feature of UBQ interactions is the conformational plasticity of the UBQ tail while participating in different types of interactions. Available hydrogen bond sites on interacting proteins likely direct these conformations; however, this feature will require further examination. Leu73 and Arg74 can occupy a variety of regions of Ramachandran space demonstrating a challenge associated with modeling tethered UBQ; however, the insights gained from UbSRD will aid in restricting the degrees of backbone freedom allowed during computational modeling of the UBQ tail.

UbSRD is designed to provide a searchable structural dictionary of UBL-containing structures, allowing users to query the structures by features of the set. UBLs engage in many thousands of diverse interactions and the biological effects of mono-UBQ have been shown to vary between targets [4,5]. Computational protein modeling is an attractive method for probing the structural details of UBL interactions and can serve as a valuable starting point in delineating potential mechanisms of UBL regulation. The insights described here will serve as a general tool to better understand UBQ and SUMO recognition and guide future computational modeling efforts.

[†]UbSRD is available at rosettadesign.med.unc.edu/ubsrd.

Methods

Populating UbSRD

To populate UbSRD with all UBLs in the PDB, we used seven iterative rounds of delta-PSI-BLAST using the UBQ sequence as a query [38]. Additionally, we used the yeast SUMO homolog, Smt3, since we could not identify these sequences with the UBQ sequence. These results were classified into six groups based upon the mode of PPI. The interaction types are monomeric UBL (MU), free ubiquitin chains (UC), noncovalent (NC), conjugated (CJ), deubiquitinase (DB), and chimeric fused (FU). The structures were also grouped by the type of UBL: small ubiquitin-like modifier 1–4 (SUMO1–4), neural precursor cell expressed developmentally down-regulated and the yeast homolog Rub1 (NEDD), autophagy-related protein (ATG), interferon-simulated gene (ISG), ElonginB (ELOB), ubiquitin-like domain (UBLD), ubiquitin mutants (UBQ_M), ubiquitin-like domain mutant UBQ-like 3 (UBL3), UBQ-like 5 and the yeast Hub1 (UBL5), and UBQ fold modifier (UFM). The conjugated and deubiquitinase interfaces were further classified by the type of interacting protein. For conjugated complexes, these groups include E1, E2, HECT E3, E2/RINGE3, E2/DB, and substrates (Sub). We included E1 in the conjugated set due to the thioester-linked intermediate formed as UBLs are transferred to E2 enzymes despite several structures co-crystallized with the enzymatic substrates. For deubiquitinase complexes, these groups include UBQ-specific peptidase (USP), UBQ C-terminal hydrolase (UCH), ovarian tumor (OTU), Josephin (Jose), JAMM, ubiquitin-like deubiquitinases (ULD), and viral deubiquitinases that have homology to papain proteases (Pap).

Renumbering UBQ and SUMO chains and executing the Rosetta feature reporter

To prepare the PDBs for database analysis, it was first necessary to renumber the UBL chains; even for proteins with canonical numberings such as UBQ, alternative numbering schemes are prevalent in some PDBs. For SUMO, all variants were converted into the human SUMO1 numbering scheme (for alignment, see Fig. 1 in Ref. [17]). We created a python script to renumber the PDBs that (1) scores each chain for similarity to a target sequence, (2) extracts chains above a certain threshold, (3) uses MUSCLE [39] to create a multiple sequence alignment of the selected chains, (4) creates renumbered PDBs to the canonical chain numbering, and (5) creates a table of the chains in the SQL database (this script is available to download on the example analysis page). These renumbered PDBs were used for feature extraction with Rosetta feature reporter. The necessary flags and the executable are found in the accompanying data in brief article [17]. Additional several manually generated tables were also added to the SQL database: interaction_type that classifies each PDB by the type of PPI; cj_type and db_type that have detailed information about the interacting proteins and the type of chemical linkage that connects the tail; ubq_chains and sumo_chains that store the chain ids of UBQ and SUMO chains; ubq_numbering and sumo_numbering that are a numbering key for the different schemes used in the database; and ubl_id, ppi_id, partner_id, and chain_id tables that are foreign keys for each classification. The UbSRD SQL database can be downloaded on the UbSRD Web site.

Phylogenetic groupings

Amino acid sequences were extracted from individual PDB files and aligned using PROMALS3D [20]. Maximum likelihood phylogenetic tree reconstruction was performed using PhyML, while implementing the WAG + I + G model (as selected by the AIC criterion as implemented within ProtTest 3 [19,40]). The phylogenetic tree was rooted at the midpoint.

Neighbor counts

We used a 6-Å cutoff to define amino acid neighbors. These measurements are from the action coordinate, which is an approximation of the average geometric center of the side chain. For all analyses, we removed self-neighbors and neighbors with other UBQ chains, and for each PDB, we divided the total number of neighbors by the number of UBL chains found. To define the secondary structure of the neighbors, we used a simplified DSSP code that only includes loops (L), helix (H), and strands (E). Hydrogen bonds were identified using the Rosetta score function and all hydrogen bonds detected by Rosetta were included regardless of the weight assigned to them. We used a python script to select only unique hydrogen bonds found in PDB structures. We did not attempt to address redundancy in our data set. We reasoned that redundant sequences may be involved in many different conformations with UBQ, the many different structures of UBQ with Ubc5 exemplifies this dilemma.

Acknowledgements

J.S.H. was supported though a National Institutes of Health (NIH) postdoctoral training grant from the Lineberger Comprehensive Cancer Center at University of North Carolina (T32-CA009156). This research was supported in part by the NIH, the National Institute of Allergy and Infectious Diseases intramural program, and from an NIH research grant (GM073960).

Abbreviations used

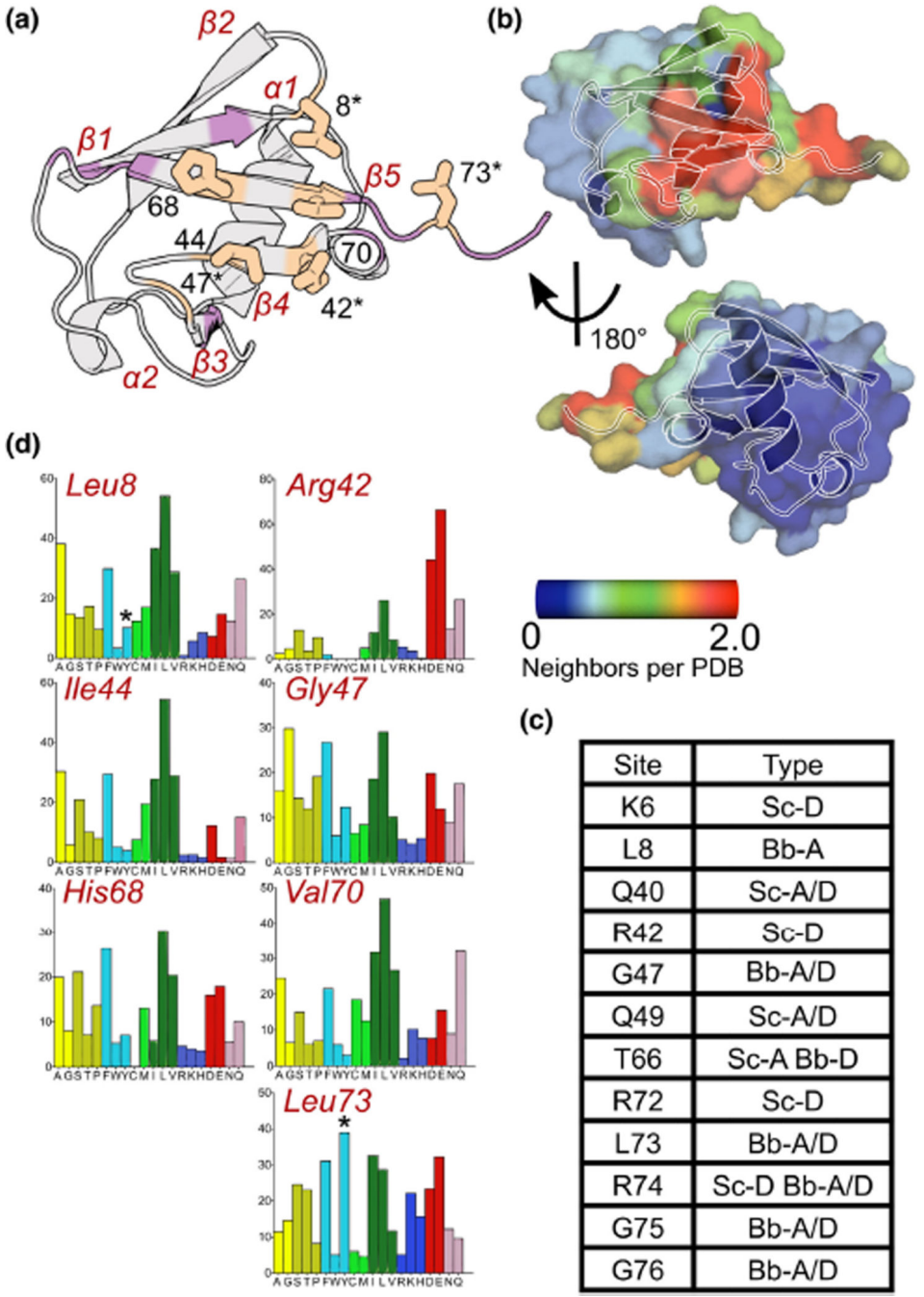
PPI	protein–protein interaction
NIH	National Institutes of Health

References

1. Rodríguez JA. Interplay between nuclear transport and ubiquitin/SUMO modifications in the regulation of cancer-related proteins. Semin. Cancer Biol. 2014; 27:11–19. http://dx.doi.org/10.1016/j.semcancer.2014.03.005. [PubMed: 24704338]
2. Sriramachandran AM, Dohmen RJ. SUMO-targeted ubiquitin ligases. Biochim. Biophys. Acta. 1843; 2014:75–85. http://dx.doi.org/10.1016/j.bbamcr.2013.08.022.
3. Bergink S, Jentsch S. Principles of ubiquitin and SUMO modifications in DNA repair. Nature. 2009; 458:461–467. http://dx.doi.org/10.1038/nature07963. [PubMed: 19325626]
4. Komander D. The emerging complexity of protein ubiquitination. Biochem. Soc. Trans. 2009; 37:937–953. http://dx.doi.org/10.1042/BST0370937. [PubMed: 19754430]
5. Geiss-Friedlander R, Melchior F. Concepts in sumoylation: A decade on. Nat. Rev. Mol. Cell Biol. 2007; 8:947–956. http://dx.doi.org/10.1038/nrm2293. [PubMed: 18000527]

6. Guo L, Giasson BI, Glavis-Bloom A, Brewer MD, Shorter J, Gitler AD, et al. A cellular system that degrades misfolded proteins and protects against neurodegeneration. Mol. Cell. 2014; 55:15–30. http://dx.doi.org/10.1016/j.molcel.2014.04.030. [PubMed: 24882209]
7. Huang J, Yan J, Zhang J, Zhu S, Wang Y, Shi T, et al. SUMO1 modification of PTEN regulates tumorigenesis by controlling its association with the plasma membrane. Nat. Commun. 2012; 3:911. http://dx.doi.org/10.1038/ncomms1919. [PubMed: 22713753]
8. Trotman LC, Wang X, Alimonti A, Chen Z, Teruya-Feldstein J, Yang H, et al. Ubiquitination regulates PTEN nuclear import and tumor suppression. Cell. 2007; 128:141–156. http://dx.doi.org/10.1016/j.cell.2006.11.040. [PubMed: 17218261]
9. Tsutakawa SE, Yan C, Xu X, Weinacht CP, Freudenthal BD, Yang K, et al. Structurally distinct ubiquitin- and Sumo-modified PCNA: Implications for their distinct roles in the DNA damage response. Structure. 2015 http://dx.doi.org/10.1016/j.str.2015.02.008.
10. Gibbs-Seymour I, Oka Y, Rajendra E, Weinert BT, Passmore LA, Patel KJ, et al. Ubiquitin-SUMO circuitry controls activated Fanconi Anemia ID complex dosage in response to DNA damage. Mol. Cell. 2014; 57:150–164. http://dx.doi.org/10.1016/j.molcel.2014.12.001. [PubMed: 25557546]
11. Schulman BA. Twists and turns in ubiquitin-like protein conjugation cascades. Protein Sci. 2011; 20:1941–1954. http://dx.doi.org/10.1002/pro.750. [PubMed: 22012881]
12. Komander D. Mechanism, specificity and structure of the deubiquitinases. Subcell. Biochem. 2010; 54:69–87. http://dx.doi.org/10.1007/978-1-4419-6676-6_6. [PubMed: 21222274]
13. Husnjak K, Dikic I. Ubiquitin-binding proteins: Decoders of ubiquitin-mediated cellular functions. Annu. Rev. Biochem. 2012; 81:291–322. http://dx.doi.org/10.1146/annurev-biochem-051810-094654. [PubMed: 22482907]
14. Faesen AC, Luna-Vargas MPA, Sixma TK. The role of UBL domains in ubiquitin-specific proteases. Biochem. Soc. Trans. 2012; 40:539–545. http://dx.doi.org/10.1042/BST20120004. [PubMed: 22616864]
15. O'Meara MJ, Leaver-Fay A, Tyka M, Stein A, Houlihan K, DiMaio F, et al. A combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. J. Chem. Theory Comput. 2015; 11(2):609–622. [PubMed: 25866491]
16. Leaver-Fay A, O'Meara MJ, Tyka M, Jacak R, Song Y, Kellogg EH, et al. Scientific benchmarks for guiding macromolecular energy function improvement. Methods Enzymol. 2013; 523:109–143. http://dx.doi.org/10.1016/B978-0-12-394292-0.00006-0. [PubMed: 23422428]
17. Harrison JS, Houlihan K, Jacobs TM, Van Doorslaer K, Kuhlman BA. Data in support of UbSRD: The Ubiquitin Structural Relational Database. Data Br. 2015
18. Kiel C, Serrano L. The ubiquitin domain superfold: Structure-based sequence alignments and characterization of binding epitopes. J. Mol. Biol. 2006; 355:821–844. http://dx.doi.org/10.1016/j.jmb.2005.10.010. [PubMed: 16310215]
19. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. 2010; 59:307–321. http://dx.doi.org/10.1093/sysbio/syq010. [PubMed: 20525638]
20. Pei J, Grishin NV. PROMALS3D: Multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol. Biol. 2014; 1079:263–271. http://dx.doi.org/10.1007/978-1-62703-646-7_17. [PubMed: 24170408]
21. Burroughs AM, Iyer LM, Aravind L. Structure and evolution of ubiquitin and ubiquitin-related domains. Methods Mol. Biol. 2012; 832:15–63. http://dx.doi.org/10.1007/978-1-61779-474-2_2. [PubMed: 22350875]
22. Perica T, Chothia C. Ubiquitin—Molecular mechanisms for recognition of different structures. Curr. Opin. Struct. Biol. 2010; 20:367–376. http://dx.doi.org/10.1016/j.sbi.2010.03.007. [PubMed: 20456943]
23. Winget J, Mayor T. The diversity of ubiquitin recognition: Hot spots and varied specificity. Mol. Cell. 2010; 38:627–635. http://dx.doi.org/10.1016/j.molcel.2010.05.003. [PubMed: 20541996]
24. Kamadurai HB, Souphron J, Scott DC, Duda DM, Miller DJ, Stringer D, et al. Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. Mol. Cell. 2009; 36:1095–1102. http://dx.doi.org/10.1016/j.molcel.2009.11.010. [PubMed: 20064473]

25. Plechanovová A, Jaffray EG, Tatham MH, Naismith JH, Hay RT. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. Nature. 2012; 489:115–120. http://dx.doi.org/10.1038/nature11376. [PubMed: 22842904]
26. Wauer T, Swatek KN, Wagstaff JL, Gladkova C, Pruneda JN, Michel MA, et al. Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. EMBO J. 2015; 34:307–325. http://dx.doi.org/10.15252/embj.201489847. [PubMed: 25527291]
27. Richardson JS, Richardson DC. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc. Natl. Acad. Sci. U. S. A. 2002; 99:2754–2759. http://dx.doi.org/10.1073/pnas.052706099. [PubMed: 11880627]
28. Lange OF, Lakomek N-A, Farès C, Schröder GF, Walter KFA, Becker S, et al. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science. 2008; 320:1471–1475. http://dx.doi.org/10.1126/science.1157092. [PubMed: 18556554]
29. Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DNA. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J. Mol. Biol. 2013; 425:1363–1377. http://dx.doi.org/10.1016/j.jmb.2013.01.032. [PubMed: 23376099]
30. Sato Y, Fujita H, Yoshikawa A, Yamashita M, Yamagata A, Kaiser SE, et al. Specific recognition of linear ubiquitin chains by the Npl4 zinc finger (NZF) domain of the HOIL-1L subunit of the linear ubiquitin chain assembly complex. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:20520–20525. http://dx.doi.org/10.1073/pnas.1109088108. [PubMed: 22139374]
31. Stieglitz B, Rana RR, Koliopoulos MG, Morris-Davies AC, Schaeffer V, Christodoulou E, et al. Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. Nature. 2013; 503:422–426. http://dx.doi.org/10.1038/nature12638. [PubMed: 24141947]
32. Mossessova E, Lima CD. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol. Cell. 2000; 5:865–876. [PubMed: 10882122]
33. Shen L, Tatham MH, Dong C, Zagórska A, Naismith JH, Hay RT. SUMO protease SENP1 induces isomerization of the scissile peptide bond. Nat. Struct. Mol. Biol. 2006; 13:1069–1077. http://dx.doi.org/10.1038/nsmb1172. [PubMed: 17099698]
34. Xu Y, Plechanovová A, Simpson P, Marchant J, Leidecker O, Kraatz S, et al. Structural insight into SUMO chain recognition and manipulation by the ubiquitin ligase RNF4. Nat. Commun. 2014; 5:4217. http://dx.doi.org/10.1038/ncomms5217. [PubMed: 24969970]
35. Ho BK, Brasseur R. The Ramachandran plots of glycine and pre-proline. BMC Struct. Biol. 2005; 5:14. http://dx.doi.org/10.1186/1472-6807-5-14. [PubMed: 16105172]
36. Békés M, Okamoto K, Crist SB, Jones MJ, Chapman JR, Brasher BB, et al. DUB-resistant ubiquitin to survey ubiquitination switches in mammalian cells. Cell Rep. 2013; 5:826–838. http://dx.doi.org/10.1016/j.celrep.2013.10.008. [PubMed: 24210823]
37. Gareau JR, Reverter D, Lima CD. Determinants of small ubiquitin-like modifier 1 (SUMO1) protein specificity, E3 ligase, and SUMO-RanGAP1 binding activities of nucleoporin RanBP2. J. Biol. Chem. 2012; 287:4740–4751. http://dx.doi.org/10.1074/jbc.M111.321141. [PubMed: 22194619]
38. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. Biol. Direct. 2012; 7:12. http://dx.doi.org/10.1186/1745-6150-7-12. [PubMed: 22510480]
39. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–1797. http://dx.doi.org/10.1093/nar/gkh340. [PubMed: 15034147]
40. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: Fast selection of best-fit models of protein evolution. Bioinformatics. 2011; 27:1164–1165. http://dx.doi.org/10.1093/bioinformatics/btr088. [PubMed: 21335321]



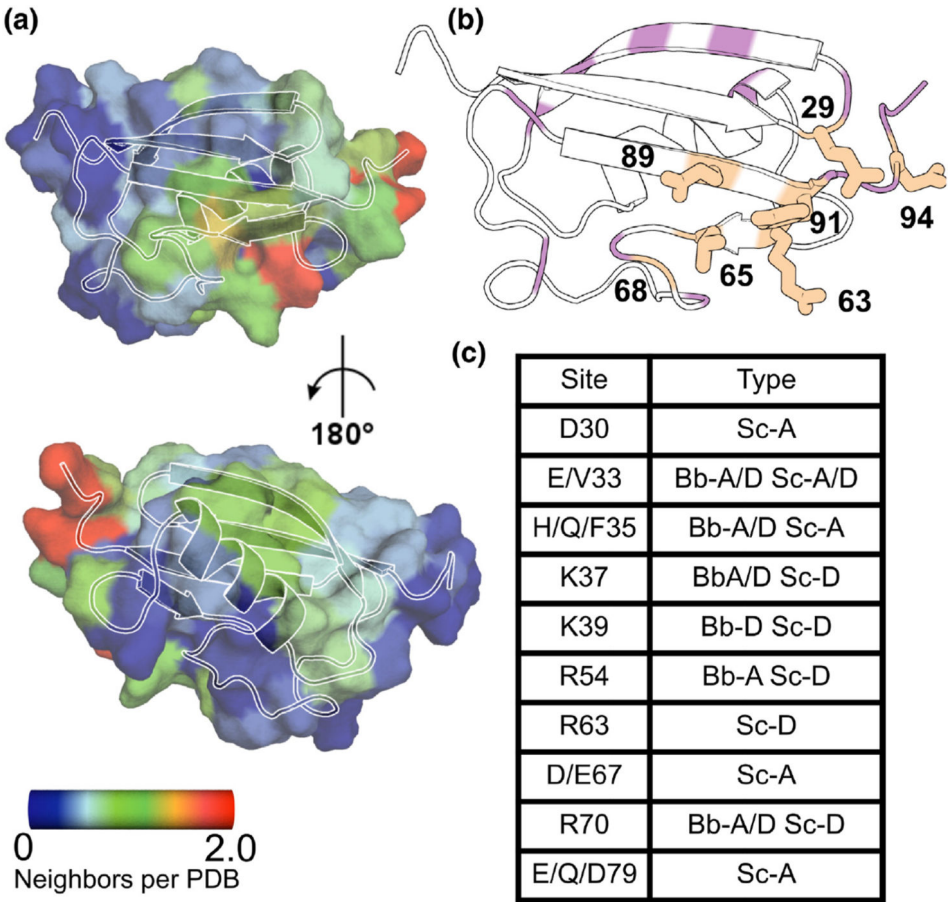
(c)

Site	Type
K6	Sc-D
L8	Bb-A
Q40	Sc-A/D
R42	Sc-D
G47	Bb-A/D
Q49	Sc-A/D
T66	Sc-A Bb-D
R72	Sc-D
L73	Bb-A/D
R74	Sc-D Bb-A/D
G75	Bb-A/D
G76	Bb-A/D

Fig. 1.

(a) Cartoon representation of UBQ. Positions that belong to the canonical binding surface are colored yellow while sites of inter-molecular hydrogen bond are colored purple. Residues in the hydrophobic face that can form hydrogen bonds are denoted with an asterisk. (b) UBQ surface heat map is colored by average number of neighbors per PDB using a 6-Å cutoff. (c) Table of hydrogen bond sites on UBQ. Hydrogen bond sites are labeled by the following classifications: ***Bb***, backbone; ***Sc***, side chain; ***D***, donor; ***A***, acceptor. (d) Amino acid partner preference for the residue in the hydrophobic surface. We grouped

and colored amino acids by properties AG (yellow), STP (gold), FWY (cyan), CM (green), ILV (hunter green), RKH (blue), DE (red), and NQ (pink). The asterisk denotes the difference in Tyr neighbors for Leu8 and Leu73.



(c)

Site	Type
D30	Sc-A
E/V33	Bb-A/D Sc-A/D
H/Q/F35	Bb-A/D Sc-A
K37	BbA/D Sc-D
K39	Bb-D Sc-D
R54	Bb-A Sc-D
R63	Sc-D
D/E67	Sc-A
R70	Bb-A/D Sc-D
E/Q/D79	Sc-A
D86	Sc-A
D89	Sc-A
D/E91	Sc-A
D92	Sc-A
Q93	Sc-A
Q94	Bb-A/D Sc-D
Q95	Bb-A/D
G96	Bb-A/D
G97	Bb-A/D

Fig. 2.

(a) SUMO surface heat map is colored by average number of neighbors per PDB using a 6-Å cutoff. (b) Cartoon representation of SUMO. Positions that belong to the hydrophobic surface are colored yellow while sites of inter-molecular hydrogen bonds are colored purple. Residues in the hydrophobic face that can form hydrogen bonds are denoted with an asterisk. (c) Table of hydrogen bond sites on SUMO. Hydrogen bond sites are labeled by the following classifications: ***Bb***, backbone; ***Sc***, side chain; ***D***, donor; ***A***, acceptor.

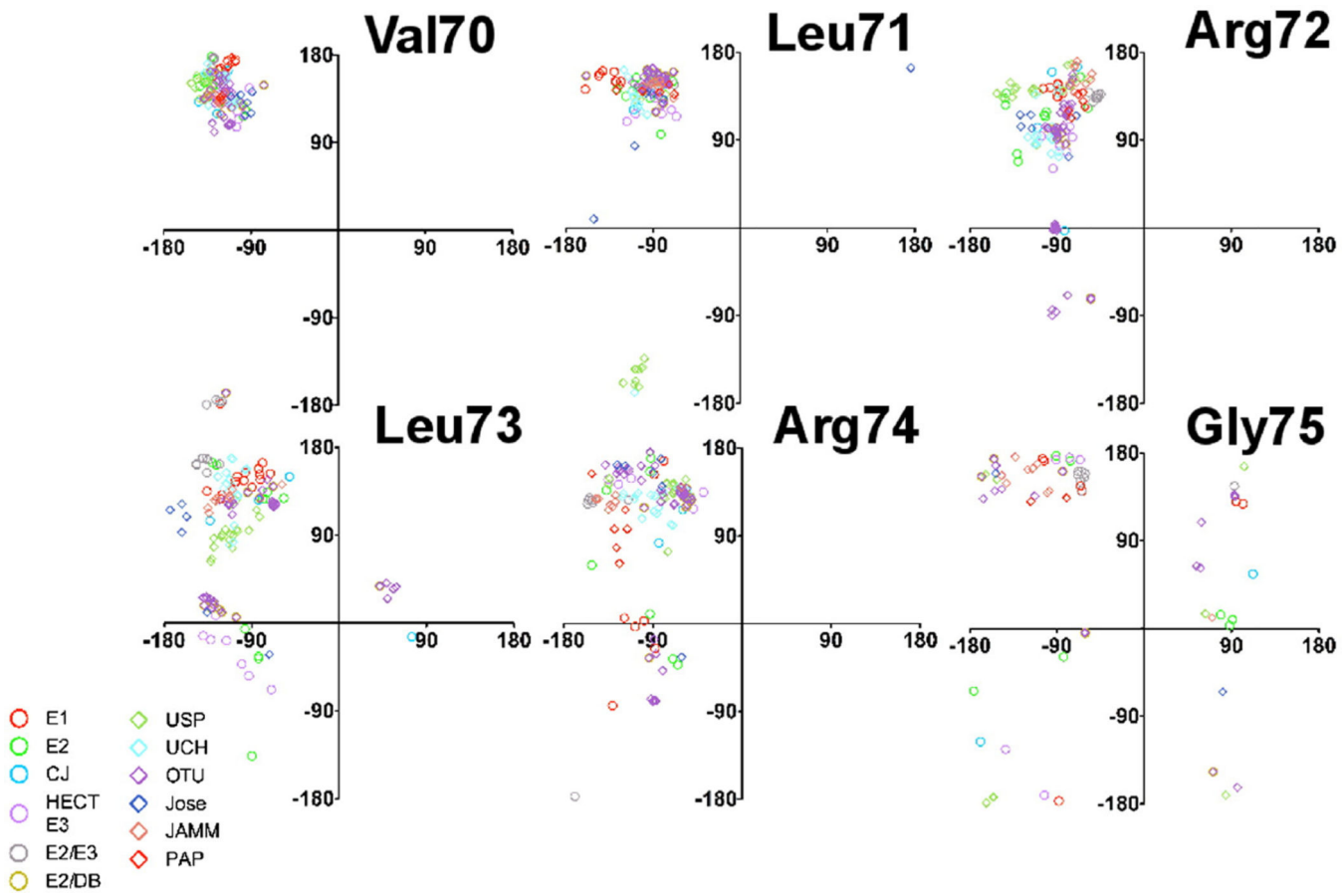


Fig. 3.

Ramachandran plots of the residues in the UBQ tail found at conjugated and deubiquitinase interfaces. We have grouped each UBQ tail by the interacting partner. For conjugated interfaces, these groups include E1, E2, HECT E3, E2/RING E3 (E2/E3), E2/deubiquitinase (E2/DB), and substrates (Sub). For deubiquitinase interfaces, these groups include UBQ-specific peptidase (USP), UBQ C-terminal hydrolase (UCH), ovarian tumor (OTU), Josephin (Jose), JAMM, ubiquitin-like deubiquitinases (ULD), and viral deubiquitinases that have homology to papain proteases (Pap).