Structural and functional characterization of ubiquitin variants and ubiquitin binding proteins

by

Noah Manczyk

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Biochemistry
University of Toronto

© Copyright by Noah Manczyk 2019
Abstract

The ubiquitin (Ub) system is an essential post-translational modification pathway found throughout eukaryotic cells and consists of proteins responsible for adding Ub to substrates (E1, E2, and E3 enzymes), removing Ub from substrates (deubiquitinases), and reading the Ub encoded signal found on substrates. The Ub system regulates cellular functions such as DNA repair and cell-cycle progression and as such, its dysregulation can contribute to disease states. Due to its prevalent nature in regulating critical cellular functions therapeutic modulation of the Ub system has become a promising area for therapeutic development. The first step towards modulating any cellular system is target identification, which requires a detailed understanding of the system’s components and how they function. Towards this end, Dr. Sidhu and his lab have developed a set of protein-based tools termed Ub variants (UbV). These reagents are engineered variants of Ub that contain substitutions that allow them to selectively and tightly bind to specific Ub binding proteins in the cell, leading to the inhibition of and in some cases activation of the targeted protein’s function. Consequently, researchers can use UbVs to understand the role of the targeted proteins in vivo and their potential as therapeutic targets. In this thesis, I aid our understanding of the Ub system through the use of biochemical and structural biology.
approaches by: (i) helping further the development of UbVs tools by identifying how they bind with high specificity and affinity to their target proteins and (ii) furthering our understanding of how the deubiquitinase USP37 uses a unique insertion containing Ub interacting motifs to cleave ubiquitin chains.
Acknowledgments

First I would like to thank my supervisor Dr. Frank Sicheri. Your passion for science and ability to distill complex concepts into the simplest and most direct forms has always amazed me. You have shaped me into the scientist and communicator I am today, for that I am forever grateful. I would also like to thank my supervisory committee members Dr. Grant Brown and Dr. Stephane Angers. Your advice and guidance over this journey has been instrumental to my success.

Next I would like to thank my collaborator Dr. Sidhu. You have made a profound impact on protein engineering and allowed me to take part in it. Additionally I would like to thank Drs. Joan Teyra, Gianluca Veggiani, Bradley Yates, and Maryna Gorelik of the Sidhu lab, you allowed me to participate in your projects and without question were always willing to help me with my own.

I would like to thank all the members of the Sicheri lab past and present. You taught me all so much and made the rollercoaster ride that is graduate school so enjoyable. I am happy to say that I did not work with colleagues but instead with friends.

I would like to thank my family. To the Lazarovitzs, thank you for providing me with a home away from home. To my uncles Michael and Chris, thank you for being the North Star on my scientific journey. To my friends who are family Daniel, Max, Seena, and JR thank you for the constant laughter and adventures. To my siblings Tamara and Aaron, thank you for your support, wisdom, and ability to immediately brighten my day. To my mother and father, thank you for your hard work and patience. I would not be who or where I am today without you. You are a source of constant inspiration and I cannot thank you enough for it.

Lastly to Gillian, your constant warmth and love encouraged me every step of the way on this journey. You are my light. Thank you.
# Table of Contents

**Acknowledgments** .................................................................................................................. iv

**Table of Contents** ..................................................................................................................... v

**List of Figures** ........................................................................................................................... ix

**Chapter 1: Introduction** ............................................................................................................. 1

1.1 Ubiquitination ......................................................................................................................... 1

1.2 SCF RING E3 Ligases ............................................................................................................... 8

1.3 Ubiquitin binding domains .................................................................................................... 12

1.4 Ubiquitin Interacting Motifs .................................................................................................. 16

1.5 Vps27 .................................................................................................................................. 19

1.6 Deubiquitination .................................................................................................................... 22

1.7 USP37 .................................................................................................................................. 32

1.8 Ubiquitin Variants .................................................................................................................. 37

1.9 Thesis Goals ........................................................................................................................... 44

**Chapter 2: Structural and functional characterization of a ubiquitin variant engineered for tight and specific binding to an alpha-helical ubiquitin interacting motif** .......................................................................................................................... 46

2.1 Introduction ............................................................................................................................ 47

2.2 Experimental Procedures ...................................................................................................... 49

2.2.1 Selection and characterization of UbV-phage binding to yUIM-1 ........................................... 49

2.2.2 Expression and purification of GST-UIM fusion proteins .................................................. 49

2.2.3 Expression and purification of UbV proteins .................................................................... 50

2.2.4 Crystallization, structure determination, refinement and analysis ...................................... 52

2.2.5 Competitive phage ELISAs .............................................................................................. 52

2.2.6 Fluorescence Polarization Binding Experiments .............................................................. 53

2.3 Results .................................................................................................................................. 54

2.3.1 UbV binders for the first UIM domain of Vps27 ................................................................. 54

2.3.2 Affinity and specificity of UbV.v27.1 ................................................................................. 56

2.3.3 Structure of UbV.v27.1 in complex with yUIM-1 .............................................................. 58

2.3.4 Site-directed mutagenesis of UbV.v27.1 and Ub.wt ............................................................ 62
2.4 Discussion ............................................................................................................................................ 64

Chapter 3: Dimerization of a ubiquitin variant leads to high affinity interactions with a ubiquitin interacting motif .................................................................................................................... 67

3.1 Introduction ........................................................................................................................................... 68

3.2 Experimental Procedures .................................................................................................................... 71

3.2.1 Protein expression and preparative scale purification ................................................................. 71

3.2.2 Peptide synthesis for analytical scale size exclusion chromatography ....................................... 71

3.2.3 Analytical scale size exclusion chromatography ......................................................................... 72

3.2.4 Refinement and structural analysis ............................................................................................ 72

3.2.5 yUIM-1 peptide synthesis for mutational analysis ......................................................................... 72

3.2.6 Fluorescence Polarization binding assay .................................................................................... 73

3.3 Results ................................................................................................................................................ 73

3.3.1 UbV.v27.1 is a dimer ....................................................................................................................... 73

3.3.2 Revised binding mode and stoichiometry of UbV.v27.1 to yUIM-1 .............................................. 76

3.4 Discussion .......................................................................................................................................... 80

Chapter 4: A Structure-based Strategy for Engineering Selective Ubiquitin Variant Inhibitors of Skp1-Cul1-F-box Ubiquitin Ligases ...................................................................................... 82

4.1 Introduction ......................................................................................................................................... 83

4.2 Experimental Procedures .................................................................................................................. 85

4.2.1 Protein expression and purification ............................................................................................. 85

4.2.2 Library construction and binding selections ................................................................................ 86

4.2.3 ELISAs ............................................................................................................................................ 86

4.2.4 Cell-based assays .......................................................................................................................... 87

4.2.5 F-box alignment ............................................................................................................................ 87

4.2.6 Crystallization and structure determination ................................................................................ 88

4.2.7 Isothermal Calorimetry .............................................................................................................. 89

4.3 Results ................................................................................................................................................. 89

4.3.1 Binding selections with an F-box targeted phage-displayed Ubv library ..................................... 89

4.3.2 Characterization of UbV binding specificity ................................................................................ 101

4.3.3 Structural Analysis of UbVs bound to Skp1-Fbl11 and Skp1-Fbl10 complexes ...................... 103

4.3.4 Effects of UbVs on the function of SCF<Fbl1 and SCF<Fbl11 in cells. .................................... 119

4.4 Discussion ....................................................................................................................................... 122
Chapter 5: The ubiquitin interacting motifs of USP37 act on the proximal Ub of a di-Ub chain to enhance catalytic efficiency ................................................................. 124

5.1 Introduction ............................................................................................................................................. 125

5.2 Experimental Procedures ......................................................................................................................... 127

5.2.1 Protein Purification ................................................................................................................................. 127

5.2.2 Size-exclusion chromatography ............................................................................................................ 128

5.2.3 Gel-based deubiquitination assays ........................................................................................................ 129

5.2.4 Fluorescence deubiquitination assays .................................................................................................... 130

5.2.5 Pulldown assays ...................................................................................................................................... 131

5.2.6 Selection of UbVs by phage display ...................................................................................................... 131

5.3 Results ....................................................................................................................................................... 132

5.3.1 Impact of UIMs on the Ub chain specificity of USP37 ........................................................................ 132

5.3.2 The effect of mutations to all three UIMs on the kinetic parameters of USP37 ........................................ 136

5.3.3 Effects of UIM mutations on USP37’s ability to cleave K11, 48 or 63-linked chains ............................ 141

5.3.4 Interrogation of proximal or distal Ub binding by the UIMs of USP37 ................................................. 142

5.3.5 Interrogation of UIM function with Ub Variants .................................................................................... 148

5.4 Discussion ................................................................................................................................................ 154

Chapter 6: Future Perspectives .......................................................................................................................... 158

6.1 Overview .................................................................................................................................................. 158

6.2 Developing a Dimeric UbV Library ............................................................................................................ 158

6.3 Crystallizing a USP37-Ub chain Complex ............................................................................................... 161

References ...................................................................................................................................................... 169
List of Tables

Table 2-1. Crystallographic data collection, processing, and refinement statistics. ...................................................... 59
Table 4-1. Summary of the F-box expression constructs, purification, and phage selection results......................... 94
Table 4-2. UbV generated in phage selections. ............................................................................................................... 96
Table 4-3. Crystallographic data collection, processing, and refinement statistics. ...................................................... 107
Table 4-4. Mutational analysis of Fbl10 and Fbl11 binding to target UbVs. ................................................................. 112
Table 5-1. Michaelis-Menten analysis of USP37<sup>wt</sup> and USP37<sub>mUIM1,2,3</sub> for K11, K48, and K63 di-Ub substrates. ... 139
List of Figures

Figure 1-1. Ubiquitination................................................................................................................. 2
Figure 1-2. Atypical Ub chains.......................................................................................................... 5
Figure 1-3. Mechanisms of Ub transfer for E3 ligases........................................................................ 7
Figure 1-4. SCF E3 ligases............................................................................................................... 9
Figure 1-5. Regulation of SCF E3 ligases.......................................................................................... 11
Figure 1-6. Classes of Ub binding domains...................................................................................... 13
Figure 1-7. Interaction surfaces of Ub............................................................................................ 15
Figure 1-8. Ub interacting motifs...................................................................................................... 17
Figure 1-9. ESCRT-0 interactions.................................................................................................... 21
Figure 1-10. Families of deubiquitinases.......................................................................................... 23
Figure 1-11. Positional nomenclature of Ub chains......................................................................... 25
Figure 1-12. DUB Ub interaction sites............................................................................................ 26
Figure 1-13. Endo and exo-peptidase activity.................................................................................. 28
Figure 1-14. Regulation of DUBs................................................................................................... 31
Figure 1-15. USP family of DUBs................................................................................................... 33
Figure 1-16. Domain architecture of USP37.................................................................................... 36
Figure 1-17. Ubiquitin Variants........................................................................................................ 38
Figure 1-18. Phase display generation of Ubiquitin Variants............................................................. 40
Figure 1-19. Binding mode of Ubiquitin Variants............................................................................. 42
Figure 2-1. Selective binding of Ub.v27.1 to yUIM-1.......................................................................... 55
Figure 2-2. Binding affinity and specificity of Ub.v27.1..................................................................... 57
Figure 2-3. The crystal structure of the Ub.v27.1-yUIM-1 complex.................................................... 60
Figure 2-4. Effects of substitutions on Ub.v27.1 and Ub.wt binding to yUIM-1................................. 63
Figure 3-1. Ub.v27.1 is an oligomeric protein in solution................................................................... 69
Figure 3-2. UbV.v27.1 is a β-strand exchanged dimer in the crystal structure................................. 75
Figure 3-3. Crystal structure of Ub.v27.1 bound to yUIM-1.............................................................. 77
Figure 3-4. Mutational and comparative analysis of the yUIM-1/Ub.v27.1 complex.......................... 79
Figure 4-1. Selection of phage-displayed UbVs targeting a panel of human F-box proteins............. 91
Figure 4-2. Influence of the β1-β2 loop insertion length on the fraction of phage-displayed UbV in the naive library. ......................................................................................................................... 93
Figure 4-3. Binding specificities of phage-displayed UbVs............................................................... 99
Figure 4-4. Phage ELISA binding specificities of 29 UbVs targeting 17 different Skp1-F-box complexes. .......................................................... 102
Figure 4-5. The crystal structure of UbV.Fl11.1 bound to Skp1-Fbl11. ........................................... 105
Figure 4-6. Comparison of the different protein complexes in the asymmetric units....................... 108
Chapter 1: Introduction

1.1 Ubiquitination

Ubiquitin (Ub) is a small protein that is at the center of a pivotal post-translational modification system conserved in all eukaryotic life, termed the Ub system. In this system, Ub is covalently attached to substrate proteins through a pathway involving a cascade of three enzyme families namely E1, E2 and E3 (Figure 1-1A). In the first step of ubiquitination, the C-terminal carboxyl group of Ub is conjugated to the sulfhydryl group of a cysteine residue in the active site of the E1 enzyme via a high-energy thioester bond. This first step is an energy consuming process and requires hydrolysis of adenosine triphosphate (ATP). Subsequently, the E1 enzyme transfers Ub onto the sulfhydryl group of a cysteine residue in the active site of the E2 enzyme via a trans-thiolation reaction. Finally, with the aid of the E3 enzyme (also termed E3 ligase), Ub is transferred onto the amine group of a substrate lysine. Substrates can be conjugated with a single Ub moiety, either on one site (mono-ubiquitination) or on multiple independent sites (multi-monoubiquitination) (Figure 1-1B). These two types of modifications play roles in processes such as protein localization and receptor endocytosis. Alternatively, Ub can be attached to a Ub already attached to substrate (poly-ubiquitination), using one of the seven internal lysines of Ub (Lys6, 11, 27, 29, 33, 48 or 63) or the amino terminus of Ub to form a chain (Figure 1-1C).
Figure 1-1. Ubiquitination.

(A) The Ub cascade involves an E1, E2 and E3 enzyme whereby subsequent “passing” of Ub between the enzymes culminates with the conjugation of the Ub moiety to a substrate protein. (B) Schematic of mono-Ubiquitination, multi-mono-Ubiquitination and poly-Ubiquitination. (C) Structure of Ub with the 7 Lys residues and the amino-terminus shown in sticks. These residues are sites for Ub chain formation.
All 8 Ub chain linkage types have been detected in cells. The best studied and most abundant of these linkages are K48 and K63 chains (termed “typical” ubiquitin chains). K48 chains were the first Ub chain type to be discovered and are most well known for their ability to recruit substrate proteins to the proteasome for degradation. K48 directed proteasomal degradation is a common mechanism for regulating cellular pathways. For example, K48 chains are formed on the fungal cyclin-dependent kinase inhibitor Sic1 leading to its degradation and promoting entry into S phase of the cell cycle. In contrast, K63 chains are not associated with proteasomal degradation but instead act as signals for recruitment of protein complexes. One of the best-studied examples of the recruitment capabilities of K63 chains are their role in recruiting DNA repair complexes to sites of DNA damage. The other 6 possible linkages (termed “atypical” ubiquitin chains) have not been as extensively characterized, but both degradative and non-degradative roles have been attributed to them. In metazoans, K11 chains are formed by the E3 ligase anaphase promoting complex (APC) on substrates such as cyclin B1 and securin. This activity enhances the degradation of substrates targeted by the APC and as a result, regulates progression through the cell cycle. Met1 chains play clearly defined non-degradative roles. This chain type is formed by the E3 ligase LUBAC on the protein NEMO to influence NF-Kβ signaling. Each of the eight different Ub chain linkages presents different epitopes to Ub binding proteins leading to differential recognition and thus different downstream functionalities for each chain type.

Further complexity and regulation of the Ub pathway can be achieved by the use of mixed chains, branched chains and post-translational modifications of Ub. In contrast to a homotypic chain that is composed of a single linkage type, mixed chains contain two or more types of linkages in a single connective chain and branched chains contain a Ub moiety with at least two or more Ub moieties attached. Collectively mixed and branched chains are referred to as heterotypic Ub chains (Figure 1-2). Both heterotypic chain types have been found to play functional roles in the cell, with K11/K63 mixed chains being formed on MHC class I molecules during endocytosis and branched K48/63 chains being used to enhance NF-KB signaling. Several post-translation modifications have been identified on Ub, including acetylation and phosphorylation. Acetylation has been detected on 6 of the 7 lysine sidechains of Ub in vivo. This modification can be used to repress Ub
chain elongation. Currently many unanswered questions regarding Ub acetylation remain, such as what proteins are responsible for (de)acetylation of Ub and what are the functional consequences of Ub acetylation. In vivo, phosphorylation has been detected on 11 sites in Ub. Currently, the only Ub phosphorylation site known to play a functional role is Ser65. When Ub is phosphorylated on Ser65 it binds to and activates the E3 ligase Parkin, which is a key regulator of mitophagy.
Figure 1-2. Atypical Ub chains.

Schematic view of homotypic, mixed (heterotypic) and branched Ub chains (heterotypic).
Of the three enzymes involved in conjugating Ub, the E3 ligase is ultimately responsible for substrate specificity. E3 ligases can be categorized into three families based on their catalytic domain: (i) Really interesting new gene (RING) and RING-like, (ii) Homologous to the E6AP carboxyl terminus (HECT), and (iii) RING-Between-RING (RBR). RING family members are characterized by the presence of a RING domain that coordinates two zinc atoms in a cross-braced fashion. The RING domain binds to both E2~Ub and substrate at the same time, bringing both in proximity for Ub transfer (Figure 1-3) \(^{13}\). Furthermore, RING domains can enhance the rate of Ub discharge from a Ub charged E2 (E2~Ub) complex by optimally positioning the C-terminal tail of Ub for nucleophilic attack \(^{14}\). RING-like domains are variations of the RING domain fold, which maintain a similar fold and functionality to RING domains but exchange coordinated zinc atoms for intra-domain sidechain interactions \(^{15}\). HECT family E3s are characterized by the presence of a bi-lobal HECT domain. Members of this family act as a catalytic intermediary between E2~Ub and substrate. The HECT domain first binds to a charged E2~Ub using its N-lobe. The Ub is then transferred to the catalytic cysteine in the C-lobe. Finally the HECT domain directly transfers Ub onto the substrate (Figure 1-3) \(^{16,17}\). Lastly, RBR E3s are characterized by a multi-domain architecture, consisting of a RING1 domain, followed by an in-between-RING (IBR) domain and finally a RING2 domain \(^{18}\). The RBR mechanism of Ub conjugation borrows from both RING and HECT E3 ligases. First, the RING1 domain binds to E2~Ub, then the Ub is transferred to a catalytic cysteine in the RING2 domain and finally the RBR ligase directly transfers the Ub moiety onto substrate (Figure 1-3) \(^{19}\).
Figure 1-3. Mechanisms of Ub transfer for E3 ligases.

RING and RING-like E3 ligases bind to both E2~Ub and substrate, allowing for Ub transfer from E2 to substrate. HECT E3 ligases bind E2~Ub with the N-lobe of the HECT domain and then transfer Ub from E2 to an internal catalytic cysteine in the C-lobe of the HECT domain. Subsequently, the HECT E3 ligase directly transfers Ub to substrate. RBR E3 ligases bind to E2~Ub via their RING1 domain and transfer Ub from E2 to a catalytic cysteine in their RING2 domain. Subsequently, the RBR E3 ligase then directly transfers Ub to substrate.
1.2 SCF RING E3 Ligases

The RING family constitutes the largest of the three families of E3 ligases and can be subdivided into: monomeric RING E3 ligases, homodimeric/heterodimeric RING E3 ligases and multi-subunit RING E3 ligases. Chapter 3 of my thesis focuses on a subset of the multi-subunit RING E3 ligases, specifically the Skp1-Cul1-F-box (SCF) E3 ligases. In the following I provide background on the structure and function of SCF ligases.

SCF E3 ligase complexes are composed of four protein subunits namely Cul1, Skp1, F-box and Rbx1. The Cul1 subunit acts as a scaffold with its N-terminal domain (NTD) interacting with the Skp1-F-box sub-complex, which is responsible for substrate recruitment, and the C-terminal domain of Cul1 (CTD) interacting with Rbx1, which is responsible for E2~Ub recruitment (Figure 1-4A)20. Cul1, Rbx1, Skp1 are common subunits of all SCF complexes, while the F-box subunit is variable. The F-box protein is the subunit responsible for substrate binding. As such the interchangeability of this subunit allows for SCF E3 ligases to recognize a diverse repertoire of substrate proteins. In the human genome there are 69 different F-box proteins and for many of these their substrate targets remain to be determined. F-box proteins contain the eponymous F-box domain (a tripartite helix) responsible for interaction with Skp1 and a protein-binding domain responsible for substrate binding (Figure 1-4B). Based on the type of substrate binding domain, F-box proteins can be subdivided into three classes: WD40 (Fbxw), LRR (Fbxl), and other (Fbxo)21.
Figure 1-4. SCF E3 ligases.

(A) Structure of a SCF E3 ligase. Cul1 acts as scaffold connecting the substrate-binding arm of the enzyme (Skp-F-box) and Ub-ligating arm of the enzyme (Rbx1) (PDBs: 1LDK and 1P22) (B) The architecture of a F-box protein consists of a F-box domain that binds to Skp1 and a substrate binding domain (PDB: 1P22).
F-box domains typically recognize substrates through short binding motifs termed degrons. Many F-box proteins employ a phospho-dependent mode of recognition of the degron. For example, the F-box Fbw7 requires its substrate cyclin E to be phosphorylated on two sites before the degron can be recognized. Phosphorylation of F-box substrates can also act in an inhibitory manner, such as phosphorylation of the Fbo11 substrate CDT2. This prevents binding of CDT2 to Fbo11. Other mechanisms employed by F-box proteins to recognize their substrate include the use of other post-translational modifications such as glycosylation, as in the case of Fbxo6 binding to the T-cell alpha chain, and direct recognition of an unmodified degron, as in the case of Fbxo4 binding to CP110.

Substrate recognition is not the only step that is regulated for SCF E3 ligase function. In particular the formation of the SCF complex and its catalytic activity are also regulated. The protein CAND1 regulates the formation of SCF complexes by directly competing with Skp-F-box complexes for an overlapping binding site on Cul1 (Figure 1-5AB). This competitive behavior also lets CAND1 act as an exchange factor for SCF complexes by allowing for different Skp1-F-box subcomplexes to engage Cul1. SCF ligase activity is also regulated by the post-translational modification termed NEDDylation. In this process, the Ub homolog Nedd8 is attached, via a cascade similar to the Ub pathway, to Cul1. This attachment of Nedd8 activates the E3 complex by conformationally increasing the flexibility of the Rbx1 RING domain (Figure 1-5C). This increase in flexibility is thought to bring the Rbx1 bound E2~Ub into closer proximity to the substrate that is remotely bound to the F-box protein. NEDDylation of Cul1 is reversible and the removal of Nedd8 is performed by the COP9 signalsome. This process of deNEDDylation reduces the activity of the SCF ligase.
Figure 1-5. Regulation of SCF E3 ligases.

(A) Structure of the exchange factor CAND1 bound to Cul1 (PDB: 1U6G). (B) CAND1 prevents Skp1 binding by competing for an overlapping interaction interface on Cul1 (PDBs: 1U6G and 3DPL). (C) NEDDylation of Cul1 causes Rbx1 to extend allowing for an increase in flexibility (PDBs: 1LDK, and 3DQV).
SCF ligases are involved in many cellular processes and consequently are dysregulated in several disease states. Thus SCF ligases represent attractive targets for therapeutic intervention. One well-studied oncogenic SCF ligase is SCFSKP2 (or SCFFbl1). This SCF complex targets the CDK inhibitor p27 for degradation and in turn promotes entry into S-phase. Several cancers overexpress Skp2, which leads to high proteolytic turnover of p27 and thus high cell cycle promoting activity from CDK-cyclin kinases. As such inhibiting the interaction of Skp1 with p27 is an attractive area for therapeutic research. On the other hand some SCF E3 ligase complexes act as tumour suppressors, with the best-studied example being SCFFbw7. This SCF complex targets several oncoproteins for degradation such as c-Myc and c-Jun. The FBW7 gene itself is mutated in 6% of all cancers, with the substrate-binding domain harboring a disproportionate number of mutations. Restoring the interaction interface between Fbw7 and its oncogenic targets through use of a small molecule could be therapeutically beneficial.

1.3 Ubiquitin binding domains

There are approximately 20 families of Ub binding domains in the human genome, responsible for “reading” the signals encoded by the different types of Ub modifications. These families can be grouped together into subfamilies including: (i) single or multiple alpha-helices, (ii) zinc fingers, (iii) pleckstrin-homology fold, and (iv) ubiquitin-conjugating-like structures (Figure 1-6).
Figure 1-6. Classes of Ub binding domains.

Depicted are the Ub-Conjugating-like (PDB:1UZX), Pleckstrin-Homology (PDB:2Z59), Zinc finger (PDB:3WWQ) and Single/Multiple Alpha Helices (PDB:2JT4) classes of Ub binding domains bound to a Ub moiety.
Ub binding domains typically bind to the hydrophobic patch of Ub, a contiguous surface centered on residues Leu8, Ile44 and Val70 (Figure 1-7A). Despite the widespread use of this surface, there are other less commonly used binding surfaces on Ub, such as those centered on Asp58 of Ub (used by the RUZ domain of Rabex) and on the C-terminal tail of Ub (used by the Zn-Ubp of USP5) (Figure 1-7B)38,39. Typically, Ub binding domains have a weak affinity for ubiquitin (>100uM) 40. Proteins that interact with Ub can increase their affinity for Ub by employing multiple Ub binding domains instead of a single Ub binding domain, by homo-oligomerizing, or by heter-oligomerizing with other proteins that bind Ub.
Figure 1-7. Interaction surfaces of Ub.

(A) Commonly, Ub binding proteins engage the hydrophobic patch of Ub, an area centered on residues Leu8, Ile44 and Val70 (PDB:1UBQ). (B) Ub binding domains can contact different interaction sites of Ub such as the hydrophobic patch used by the UBA domain (PDB:1WR1), the interaction site centered on Asp58 used by the RUZ domain (PDB:2C7N) and the interaction site centered on the C-terminal tail used by Zn-UBP (PDB:2G45).
Ub binding domains can either non-selectively bind all chain linkages or selectively bind one or a subset of linkages. Most Ub binding domains are non-selective. This is because the binding epitope of most Ub binding domains are wholly contained on a single Ub moiety. In contrast, the binding epitope for chain selective Ub binding domains is contained across two Ub moieties and can also include the (iso)peptide linkage. For example, the UBA2 domain of hHR23A binds selectively to K48 chains by contacting the hydrophobic patches of two adjacent Ub moieties and the K48 isopeptide.  

1.4 Ubiquitin Interacting Motifs

In chapter two of my thesis, I focus on a specific Ub binding domain called the Ub interacting motif (UIM). In the following, I provide an overview of the structure and function of the UIM.

UIMs are the smallest Ub binding domain, consisting of a single alpha helix of (~20 residues in length) typically conforming to the consensus sequence of e-e-x-x-φ-x-A-φ-x-(φ/e)-S-z-x-e, where e is an acidic residue, φ is a hydrophobic residue, z is a bulky hydrophobic or polar residue with high aliphatic content, A is alanine, S is serine and x is any residue. The interaction between UIMs and Ub are typically weak (>100 uM). Ub interacting motifs bind to the hydrophobic patch of Ub and involve packing of the Ala position of the UIM against residues Ile44, His68 and Val70 of Ub (Figure 1-8AB).
Figure 1-8. Ub interacting motifs.

(A) Canonical interaction between UIMs and Ub (PDB: 1Q0W). (B) The canonical UIM interaction centers about the hydrophobic patch of Ub. Involving residues Ile44, His68 and Val70, of Ub and the conserved Ala position of the UIM (PDB: 1Q0W). (C) The K63 linkage specific tandem UIMs of RAP80 contact hydrophobic patches of both Ub moieties in K63 di-Ub. The interaction helicizes the linker between the tandem UIMs (PDB: 3A1Q). (D) Non-canonical UIM-Ub interaction of the UIM1 of AIRAPL (PDB: 4XKH). This UIM interacts on the same interface of Ub as a typical UIM but with reverse polarity of the termini. (E,F) Non-canonical UIM-Ub interactions of UIM2 of AIRAPL and the double-sided UIM of Hrs. Both of these UIMs are able to contact two Ub moieties simultaneously unlike a typical UIM (PDB: 4XHK and 2D3G).
UIM containing proteins commonly include multiple copies (2 to 3) of the motif. The presence of multiple UIMs can act as a general means to increase affinity for poly-Ub chains, as previously mentioned for Ub binding domains. Alternatively, the presence of multiple UIMs can confer chain selectivity. For example, RAP80 contains two UIMs that selectively bind K63 Ub chains. The structure of the tandem UIMs of RAP80 bound to K63 di-Ub revealed that one RAP80 UIM recognizes the hydrophobic patch of one Ub moiety and the second RAP80 UIM recognizes the hydrophobic patch of the adjacent Ub moiety. Interestingly, when in the presence of K63 di-Ub the linker separating the tandem UIMs rigidifies into a helix, turning the two UIMs into a single long alpha helix (Figure 1-8C).

The ability of the tandem UIMs of RAP80 to selectively engage K63 chains critically depends on the linker length and sequence between tandem UIMs as when linker length was varied or residues that perturb alpha helix formation were introduced K63 binding was attenuated. A similar linker length and composition was found for another tandem UIM containing protein, namely Epsin, which also showed K63 selectivity suggesting linker helicization may be a common means to induce K63 selectivity in tandem UIMs.

Several novel variations of the canonical UIM-Ub interaction exist. One example is the first UIM (UIM1) of the human protein AIRAPL. This UIM binds to Ub in the same manner as a typical UIM but with the N and C-termini flipped (Figure 1-8D). The second UIM (UIM2) of AIRAPL also exemplifies another novel variation of the UIM. UIM2 of AIRAPL can bind two Ub moieties simultaneously by engaging the hydrophobic patch of one Ub moiety and an interface centred on Leu71 of the other Ub moiety (Figure 1-8E). Another dual Ub contacting UIM is the double-sided UIM of the human vesicle trafficking protein Hrs. This UIM essentially has two typical UIM binding surfaces spliced onto opposite faces of a single helix, allowing the double-sided UIM to contact the hydrophobic patches of two Ub moieties simultaneously (Figure 1-8F). Interestingly, the yeast ortholog of Hrs, Vps27 (Vacuolar Protein Sorting-associated protein 27), contains a tandem array of UIMs instead of a double-sided UIM, suggesting that the human ortholog evolved the tandem UIM array into a single motif.
In chapter two and three of this thesis, my research focuses on the first UIM of the yeast protein Vps27. The first UIM of Vps27 was chosen not because of its unique biological function but because this UIM represented a biochemically and structurally well-characterized UIM. Nevertheless to garner an understanding of how UIMs contribute to protein function I now provide a brief overview of the biological function of Vps27 and the role of its UIMs in Ub recognition.

Vps27 was identified in a yeast genetic screen for genes responsible for the sorting of vacuolar and Golgi proteins. Since its original identification it has become clear that Vps27 plays a role in the generation of multi-vesicle bodies, a type of late endosome involved in the process of degradation of membrane proteins, and the sorting of proteinaceous cargo into multi-vesicle bodies.

Proteinaceous cargo on the endosomal membrane meant for degradation is modified with both mono-Ub and Ub chains (typically K63). The machinery responsible for sorting this cargo and the generation of multi-vesicle bodies are a set of four protein complexes termed ESCRT 0, I, II, III. A dimeric complex of Vps27 (HRS – humans ortholog) and Hse1 (STAM1/2 - human ortholog), interacting through their GAT (GGA and Tom1) domains, make up the first complex in the pathway, namely ESCRT-0 (Figure 1-9). ESCRT-0 is recruited to the endosomal membrane through binding of its FYVE (Fab1, YOTB, Vac 1, and EEA1) domain to phosphatidylinositol 3-phosphate present on the endosomal membrane. ESCRT-0 recognizes ubiquitinated cargo through five Ub binding domains, with Vps27 containing two UIMs and one VHS (VPS-27, Hrs, and STAM) domain, and Hse1 containing one UIM and one VHS domain. The importance of the Ub binding ability of ESCRT-0 is clear from the fact that loss of this function leads to defects in cargo sorting. Interestingly, ESCRT-0 can homo-multimerize which creates subdomains on endosomal membranes that may concentrate and retain ubiquitinated cargo. The other function of ESCRT-0 is to bind and recruit the ESCRT-I complex. This function is mediated by a P(T/S)AP motif in the C-terminus of Vps27. Recruitment of ESCRT-I initiates the remaining steps of the ESCRT cascade, which culminates with membrane budding and abscission of a vesicle containing
ubiquitinated cargo into the MVB \(^5^2\). Additionally, the ESCRT machinery has roles in several other cellular processes including: cytokinetic abscission, nuclear envelop assembly and exovesicle shedding \(^5^2\).
Figure 1-9. ESCRT-0 interactions.

Schematic overview of the interactions of the yeast ESCRT-0 complex, which is composed of the proteins Vps27 and Hse1. The two proteins interact via their coiled-coil GAT domains. The FYVE domain of Vps27 recruits the complex to endosomal membranes by binding to phosphatidylinositol 3-phosphate (PI3P). The Ub binding domains of Vps27 (2 UIMs and 1 VHS domain) and Hse1 (1 UIM and 1 VHS domain) bind and sort ubiquitinated cargo into multi-vesicle bodies. The C-terminal P(T/S)AP motif of Vps27 recruits the downstream ESCRT-1 complex, setting in motion the subsequent steps of the ESCRT cascade.
1.6 Deubiquitination

Deubiquitinases (DUBs) are enzymes responsible for cleaving Ub from substrate proteins and thereby remove the signal imposed by the Ub chain. In this manner, DUBs play a pivotal role in the regulation of the Ub pathway by opposing the forward process of ubiquitination. In the human genome there are seven families of DUBs based on structural fold of the DUB domain: UCH, Otu, MPN+, Josephin, USP, MINDY, and ZUFSP (Figure 1-10). With the exception of the MPN+ family of DUBs, which are zinc metalloproteases, all DUB families are cysteine proteases. Intriguingly, the MINDY and ZUFSP families have only been recently discovered and suggest the possibility of other yet to be discovered DUB families. 

57-60.
Figure 1-10. Families of deubiquitinases.

Depicted are the structures of the 7 families of deubiquitinases based on the structural fold of their DUB domain: UCH (PDB:3TB3), Otu (PDB:2ZFY), MPN+ (PDB:5CW6), Josephin (PDB:1YZB), USP (PDB:1NB8), MINDY (PDB:5JKN) and ZUFSP (PDB:6FGE).
To understand how DUBs cleave Ub chains, a definition of the standardized nomenclature regarding the related positions of Ub molecules in a chain is required. In di-Ub, the shortest Ub chain, there is a proximal (amine donating Ub) and distal (C-terminal donating Ub) Ub (Figure 1-11A). For chains longer than 2 moieties, the labeling of proximal and distal denotes relative positions of the moieties within the chain, with the proximal Ub being the moiety closer to the substrate and the distal Ub being the moiety further away from the substrate (Figure 1-11B). All known DUBs contain a distal Ub binding site, also known as the S1 site (Figure 1-12). This minimal site is sufficient for DUBs that remove mono-Ub from proteins or that remove modifications from the C-terminus of Ub. DUBs that cleave Ub chains also contain a proximal Ub binding site, also known as the S1’ site (Figure 1-12). Additional Ub binding sites (termed S2 and S3) are present on some DUBs and engage Ub moieties in longer chains (Figure 1-12) 61.
Figure 1-11. Positional nomenclature of Ub chains.

(A) In di-Ub there is a distal (C-terminus donating) and proximal (amine donating) Ub moiety. (B) In chains longer than di-Ub, proximal and distal refer to the Ub moieties closer and farther away along the chain from the substrate, respectively.
Figure 1-12. DUB Ub interaction sites.

All known DUBs contain a distal Ub binding site (also termed S1 site). DUBs that interact with Ub chains also contain a proximal Ub binding site (also termed S1' site). Additional binding sites for more distal Ub moieties have also been found in some DUBs (S2 and S3 site).
Ub chains present multiple positions where a DUB can cleave. These positions can be categorized into 3 groups: the most distal Ub, the most proximal Ub and anywhere else within the Ub chain (Figure 1-13). DUBs that cleave at either termini are termed as having exo-peptidase activity. For example, MINDY-1 has been shown to have exo-peptidase activity for the distal termini, while USP5 has been shown to have exo-peptidase activity for the proximal termini of unanchored Ub chains \(^{57,62}\). On the other hand, DUBs that cleave internally within a Ub chain are termed as having endo-peptidase activity. Endo-peptidase DUBs are more efficient at removing Ub chains than exo-peptidase DUBs, as each cleavage event can remove more than one Ub moiety. An example of a DUB with endo-peptidase activity is the recently discovered ZUFSP protein \(^{59}\). While some DUBs can attack the chain at any of the 3 described positions, other DUBs are restricted in the positions they can cleave. What determines whether the DUB displays exo- or endo-peptidase activity partly depends on how the DUB interacts with Ub at the distal binding site. For example, the structure of Ub bound to the distal Ub binding site of USP21 revealed that while Lys63 was solvent exposed, Lys6 was buried \(^{63}\). This result suggested USP21 can bind and cleave within a K63 chain, as a chain could extend from Lys63 of Ub at the distal binding site, but can only bind and cleave the distal Ub of a K6 chain, as a chain could not extend from Lys6 of Ub at the distal binding site. Consistent with this model, USP21 has endo-peptidase activity towards K63 chains and exo-peptidase activity towards the distal end of K6 chains \(^{63,64}\).
Figure 1-13. Endo and exo-peptidase activity.

Deubiquitinases can cleave at three different positions within a Ub chain, the most proximal Ub (proximal exo-peptidase activity), the most distal Ub (distal exo-peptidase activity) or within the Ub chain (endo-peptidase activity).
DUB specificity can act at two levels, firstly at the level of the Ub chain linkage and secondly at the level of the substrate harboring the Ub modification. With regard to Ub chain specificity, non-specific DUBs predominantly recognize the distal Ub in a di-Ub chain allowing them to cleave all 8 linkages. In contrast, chain specific DUBs recognize both the distal and proximal Ub in a di-Ub chain allowing them to cleave one or a select subset of the 8 linkages. With regard to substrate specificity, DUBs must be directed to specific protein substrates to which Ub is attached to prevent off-target deubiquitination. DUB substrate specificity can be divided into direct and indirect recognition of substrates. An example of a DUB that directly recognizes its substrate is USP7, which binds peptide motifs in its substrates such as p53, MDM2, and MDMX through a TRAF domain. In the cases of indirect recognition, DUBs can associate with another protein that directly binds the substrate. One example of indirect recognition is seen for the DUBs UCH37, RPN11 and USP14, which associate with the 26S proteasome. This allows them to deubiquitinate select ubiquitinated proteins bound by the proteasome. Interestingly, in some cases of indirect recognition the protein responsible for substrate recognition is the E3 ligase that the DUB functionally antagonizes, such as in the case of the DUB GUMBY with the E3 ligase LUBAC.

Due to their critical role in the Ub pathway DUB activity is often tightly regulated. Regulation of DUB activity can be carried out by post-translational modifications or by protein cofactors that bind the DUB. In the case of post-translational modifications, one example is seen for the DUB DUBA, which has its cleavage activity enhanced by phosphorylation on Ser177. The structure of phosphorylated DUBA bound to Ub revealed that Ser177 phosphorylation structures a loop in DUBA that clamps the C-terminal tail of Ub against the core of the DUB domain and thus allows the enzyme to cleave the isopeptide bond. In a second example, oxidation was shown to regulate DUB activity. This occurs by reversible oxidation of the catalytic cysteine residue, which renders the DUB inactive. USP1 was found to have its catalytic cysteine oxidized in cells exposed to hydrogen peroxide leading to the increase in the mono-ubiquitination state of its target substrate PCNA. Most DUBs are cysteine proteases and thus may be subject to this form of regulation. DUBs can also be regulated by protein cofactors. One example of protein cofactor regulation is seen for the DUB BRCC36. The cleavage activity of BRCC36 is
dependent upon binding to its protein cofactor KIAA0157. The structures of apo-BRCC36 and BRCC36 bound to KIAA0157 revealed how KIAA0157 enhances cleavage. BRCC36 binding to KIAA0157 orients a catalytic Glu and a predicted substrate-binding loop (Ins-1) of BRCC36 to a catalytically competent conformation (Figure 1-14B). Interestingly, KIAA0157 contains the structural fold of a DUB domain (in this case a MPN(+) DUB domain) but does not have intrinsic DUB activity due to alterations in its active site. This type of protein is termed a pseudo-DUB. Other DUBs are also regulated by pseudo-DUBs such as the deubiquitinase RPN11, which is regulated by the pseudo-DUB RPN8.72,73.
Figure 1-14. Regulation of DUBs.

(A) Phosphorylation of Ser177 on DUBA enhances cleavage activity by structuring a loop that clamps the C-terminal tail of Ub to the DUB (PDB: 3TMP). (B) KIAA1057 binding to the DUB BRCC36 enhances cleavage activity by reorganizing a loop containing the catalytic Glu responsible for zinc coordination and the Ins-1 loop which is predicted to bind substrate (PDBs: 3TMP and 3TMO).
Due to their pivotal role in regulating the Ub pathway DUBs are attractive targets for therapeutic intervention. One example is USP7, which among other functions deubiquitinates and stabilizes the oncogenic E3 ligase MDM2. In this context inhibiting USP7 cleavage activity may impart a therapeutic benefit in cancer cells that overexpress MDM2. Recently, highly specific inhibitors of USP7 that act by preventing Ub chain cleavage have been developed and these were shown to inhibit tumour growth in mouse models. Generating DUB inhibitors that are not only effective but also highly selective is a requirement for drug development. However, to date only a few highly selective small-molecules have been generated despite great efforts having been made in academic labs and pharmaceutical companies.

1.7 USP37

In chapter 4, I focus on USP37 a member of the USP family of deubiquitinases. In the following, I give an overview of the USP family of DUBs and USP37 function.

The USP family represents the largest family of DUBs in the human genome with ~60 members. With only a few exceptions, the majority of USP family members do not display specificity for Ub chain linkages. The seminal structure of a USP, namely USP7, revealed the USP DUB fold, which has been likened to a hand consisting of a fingers, thumb and palm subdomain (Figure 1-15AB). The fingers subdomain forms the distal Ub binding site, which makes an extensive interaction interface with Ub including the hydrophobic patch, and the interface between the palm and thumb subdomain forms the catalytic cleft of the USP enzyme (Figure 1-15C).
Figure 1-15. USP family of DUBs.

(A) Structural fold of the USP DUB domain (PDB:1NBF). (B) The USP DUB domain can be divided into a fingers, thumb, palm subdomains (PDB:1NBF). (C) Structure of a USP DUB domain bound to Ub (PDB:1NBF). (D) USP domains contain 5 possible insertion sites (PDB:1NBF).
While the core fold of USP enzymes is only ~300 amino acids, the length of some USP domains can be significantly longer, up to 850 amino acids, due to insertions within the DUB domain at five common insertion points (Figure 1-15D) \(^78\). Some of these insertion sites contain known or predicted domains. Some of these domains can influence chain cleavage activity, such as the insertion in USP5, which contains two UBA domains that form the S2 and S3 binding sites for Ub moieties in a chain \(^61\). In contrast, some of these domains can influence protein function unrelated to chain cleavage activity, such as the insertion in CYLD, which contains a B-box domain that regulates protein complex formation and cellular localization \(^79,80\). Similarly, domains can be found either N- or C-terminal of the USP DUB domain. These can also impart functionality both related and unrelated to cleavage activity. For example the DUSP-Ubl domains of USP4 directly contribute to the catalytic efficiency of USP4 by enhancing exchange of cleaved Ub product, while the TRAF domain of USP7 is responsible for interacting with substrates such as p53 \(^65,66,81\). Despite advances in understanding how these accessory domains influence DUB cleavage activity and function, the role of many of these accessory domains remain unknown or incompletely understood.

USP37 is a member of the USP family of DUBs first identified by a bioinformatic approach aiming to identify new members of the USP family \(^82\). USP37 has been shown to regulate cellular progression into S phase by deubiquitinating and stabilizing cyclin A \(^83\). USP37 also regulates cohesion of sister chromatids during mitosis through an unclear mechanism dependent on its DUB activity \(^84\). USP37 also regulates the homologous recombination DNA repair pathway through deubiquitinating ubiquitin conjugates at DNA double-stranded breaks \(^85\). Lastly, USP37 deubiquitinites the oncogene c-Myc and the oncogenic fusion protein PLZF-RARA. These later actions raise the possibility that inhibition of USP37 DUB activity could serve a therapeutic value \(^86,87\).

USP37 has a domain architecture consisting of a PH domain, a ~200 amino acid linker and a C-terminal USP domain. The USP domain of USP37 contains a unique insertion of 283 amino acids (in human USP37) containing 3 UIMs (UIM1, UIM2, and UIM3) (Figure 1-16). This insertion has been shown to regulate USP37 cleavage activity by three means. First, the insertion has a cyclin A-CDK2 phosphorylation site (S628), which when phosphorylated increases USP37 cleavage activity by a yet to be determined mechanism \(^83\). Second, UIM2
and UIM3 but not UIM1 of USP37 enhances the cleavage activity of USP37 towards K48 and K63 chains. Third, UIM2 and UIM3 but not UIM1 play a role in binding Ub-conjugates and the cohesion regulator WAPL. In contrast to their effect on catalytic activity, the UIMs do not confer chain specificity of USP37 towards K48 over K63 chains.$^{84,88}$
Figure 1-16. Domain architecture of USP37.

USP37 consists of a N-terminal PH domain and a C-terminal USP DUB domain. Within the DUB domain of USP37 is an insertion containing 3 Ub-interacting motifs and a phosphorylation site for CyclinA-CDK2.
1.8 Ubiquitin Variants

Ubiquitin Variants (UbVs) are engineered forms of Ub that contain substitutions to allow the UbV to bind specifically and tightly to a targeted Ub binding protein. The basis of how UbVs are generated focuses on the fact that Ub forms interactions with many proteins using a common surface, namely the hydrophobic patch centered on Ile44. By substituting residues on this interaction surface of Ub it is possible to convert the typically weak and non-specific interaction between Ub and the target to a tight and specific interaction (Figure 1-17). The highly specific and tight interaction between the UbV and its target protein can competitively prevent Ub from binding the target thus allowing the UbV to act as an inhibitor. Unexpectedly, not all UbVs generated to date act as inhibitors. For example, UbVs targeting the NEDD4L E3 ligase activated the ligase activity of the enzyme, presumably by binding to a natural allosteric Ub binding site. The unique behavior of UbVs make them valuable tools to determine the biological function in vivo of a target protein (a pivotal step in drug-target identification) and to explore the molecular mechanisms of action of Ub binding proteins in vitro (as I exploit in Chapter 5).
Figure 1-17. Ubiquitin Variants.

Ub forms multiple low affinity interactions in the cell. Ub variants are engineered forms of Ub that have substitutions relative to Ub. The substitutions optimize the interaction of the UbV with a specific Ub binding protein and increase the binding affinity.
UbVs are engineered in a phage-display process developed in Dr. Sachdev Sidhu’s research group. In brief, Ub is genetically fused to the pIII outer coat protein of M13 bacteriophage (Figure 1-18A). Residues of Ub that are deemed important for interaction with a target protein of interest are then subjected to soft randomization (randomizing residues with a bias towards the wild-type sequence) using degenerate oligonucleotides creating a plasmid library of mutagenized UbV. The plasmid library is then transfected into bacterial cells and with the aid of helper phage, a phage library (library size up to $10^{10}$ different phage) is created. The phage library is then subjected to a selection process centered on binding of the phage to the Ub binding protein of interest that is coated on a surface (Figure 1-18B). In this selection process phage expressing a UbV that binds weakly to the target are washed away while stronger binders are retained. This process is repeated several times to enrich for the tightest binders. After the final washing step, the phage that bound to the Ub binding protein are amplified and sequenced. Binding between the UbV and target protein are verified using phage-based ELISAs and an orthogonal binding technique (e.g. fluorescence polarization or isothermal calorimetry). This overall process allows for the identification of UbVs against a target protein in a matter of weeks.
Figure 1-18. Phase display generation of Ubiquitin Variants.

(A) For the phage display process, Ub is fused to the pIII outer coat protein of M13 bacteriophage. (B) After the phage library is synthesized, the phage library is incubated with a surface coated target protein. The surface is washed removing weak or non-binding phage. Phage that bind the target protein stay bound and then are eluted and amplified. Binding, washing, eluting and amplifying steps are repeated to enrich for the tightest phage binders.
The first UbV library generated by Dr. Sidhu’s lab was designed to generate UbVs for the USP family of DUBs and consisted of mutagenizing ~30 residues that contacted the USP domain of USP21 in a published co-structure. This library successfully generated binders for several USP family members including USP8, USP21, and USP2a. While the library was originally created to generate UbV for USPs, it also successfully generated UbVs for other DUB families including OTUB1 and BRCC36 containing complex and other ubiquitin interacting proteins such as the HECT domain of ITCH and the E2 Cdc34. Interestingly, UbVs have also been generated against proteins not known to interact with Ub. For example, UbVs have been engineered against Skp1-F-box complexes, such as Skp1-Fbw7, which are not known to interact with Ub. Whether this interaction represents a previously unidentified Ub interface or instead represents an artifact of protein engineering is currently unknown. However it does show the versatility of the phage display technique using Ub as a scaffold.

Structural characterization of generated UbVs bound to their target has revealed that UbVs can bind their protein target in ways that are similar or dissimilar to the canonical Ub interaction. For example, UbV.21.4 bound USP21 in a manner similar to the known USP21-Ub interaction, while UbV8.2 bound USP8 in a manner dissimilar (~85° rotation and 5 Å translation) to the predicted USP8-Ub interaction (Figure 1-19). With the aid of structural information regarding how UbVs bind their target proteins, UbV libraries can be further refined to improve affinity or specificity. For example, a second generation Skp1-F-box specific UbV library was developed with an extended β1-β2 loop to improve specificity based on information from the structure of a first generation Skp1-F-box UbV bound to its target. I identify further improvements to this second generation UbV library in Chapter 4.
Figure 1-19. Binding mode of Ubiquitin Variants.

Ubiquitin Variants can bind to their target protein in modes that are similar or dissimilar to the Ub interaction. For example, a UbV engineered against USP21 (left panel) binds to USP21 in a manner similar to the Ub interaction (PDBs: 3MTN and 3I3T). While a UbV raised against USP8 (right panel) binds to USP8 in a manner that is dissimilar to the predicted Ub interaction (PDBs: 3N3K and 1NBF).
Ubiquitin variants have several uses as tools to study protein function both in vitro and in vivo. In an in vitro setting, UbVs have been used to biochemically determine and validate molecular mechanisms of action for E3 ligases and DUBs. A second example of an in vitro use of UbVs, has been their use in cryo-electron microscopy and crystallographic studies of Ub binding proteins. For structural studies formation of a stable complex is essential. The typically weak and transient nature of Ub interactions can preclude structural characterization of many Ub binding proteins and complexes. UbVs can aid in these circumstances by acting as a proxy for a Ub moiety and also trapping the interaction due to their increased binding affinity for the target protein/complex relative to Ub. For example, in a recent cryo-EM structure of the APC E3 ligase a UbV was used as a proxy for substrate-linked Ub. For in vivo studies, UbVs can be used to verify and identify the functional consequences of modulating a target protein’s Ub binding activity. Due to the typically high specificity of UbVs, in vivo studies with UbVs can typically be carried out without the confounding issue of off-target effects. For example, in an in vivo setting a UbV inhibitor for Skp1-Fbw11 displayed high specificity for its target, including an ability to discriminate between Fbw11 and its close homolog Fbw1 (81% sequence identity), as assessed by IP-mass spectrometry. This UbV also inhibited its target SCF complex, namely SCFFbw11, as assessed by western blot analysis. Another example of an in vivo use of UbVs has been their utility in screens for discovering new protein regulators of pathways. For example, a UbV screen searching for novel HECT E3 ligase regulators of cell migration was able to successfully identify a new regulator, namely NEDD4L. To perform the screen the researchers developed UbVs for 13 different HECT E3 ligases and transduced lentiviruses expressing the UbVs into human colon cancer carcinoma cells and monitored cell-migration.

UbVs are valuable tools for the identification and verification of potential therapeutic targets due to their aforementioned specificity and ability to elicit biological outcomes. UbVs provide a complementary approach with some unique advantages to genetic based techniques such as CRISPR knockout and siRNA knockdown. First, genetic manipulation acts in an inhibitory role by removing the protein from the cell. In contrast, UbVs can inhibit or activate a protein allowing one to study both modes of modulation. Second, UbVs can help identify novel modulatory sites for drug development. Third, genetic-based
techniques work at the level of DNA, while UbVs work at the level of protein. Drugs, like UbVs, also work at the level of protein and as such UbVs are more drug-like in their mechanism of action. As such, understanding how a UbV inhibits or activates its target from a structural point of view can provide a starting point for drug development. Fourth, genetic knockouts and knockdowns of the full protein perturb all functions of the target protein. In contrast, UbV inhibitors perturb only the functions of the region targeted by UbV binding. To replicate the more precise mechanism of UbVs, genetic based techniques would need to knockdown or knockout the wild-type protein and express a target protein with the regions of interest deleted. This could be difficult as it could unpredictably affect protein expression and stability.

1.9 Thesis Goals

UbVs are a recently developed tool that allows researchers to modulate the Ub pathway. Among other uses, UbVs provide a way to understand functional roles of Ub interacting proteins in vivo and thereby also identify new potential therapeutic targets. Despite advances in the development of UbVs, during my PhD several unanswered questions arose that focused on how UbVs bind tightly and specifically to their targets. The first aim of my thesis was directed at addressing these questions.

Initially, UbVs were developed against targets with large Ub interaction surfaces such as the USP-Ub interface (~2000 Å²). It was unclear if UbVs could be designed to bind tightly and specifically to the smaller Ub interaction interfaces of UIMs (~500 Å²). Our collaborators in the Sidhu lab were able to generate a tight binding and specific UbV against UIM1 of Vps27, proving that it is indeed possible to generate a specific and tight binding UbV against UIMs. However, they lacked an understanding of how this UbV bound to its cognate UIM and what substitutions were responsible for the increased binding affinity. In my chapter two and three, I address these two questions through structural, mutational, and functional studies of a specific UbV-UIM complex.

The first generation of UbVs targeting individual Skp-F-box complexes were cross-reactive for many unintended Skp-F-box complexes. The structure of a relatively non-specific UbV in complex with its intended Skp-F-box revealed that the UbV made only few contacts with
the F-box protein, the variable component of Skp-F-box complex, which would define specificity. The Sidhu lab hypothesized that they could increase specificity of their Skp-F-box binding UbVs by increasing the interaction interface between the UbV and F-box by extending the loop that made glancing contacts with the F-box in the crystal structure. Their hypothesis proved to be correct as this strategy was capable of generating UbVs that were specific for Skp1-Fbw11. Our collaborators in the Sidhu lab then proved that this strategy could be used as a general means to generate highly specific UbVs for other Skp-F-box members. However, they lacked an understanding of how the extended loop UbV engaged the F-box complex and how it caused an increase in binding specificity. In my second chapter, I addressed these questions through structural determination of UbV-Skp-F-box complexes and in collaboration with the Sidhu lab through mutational studies.

The second focus of my PhD aimed at resolving how the UIMs in USP37 increase the catalytic cleavage activity of USP37. Previous studies on USP37 shed initial light on how the UIMs influence the cleavage activity of USP37 but many questions were unresolved. In brief these previous studies revealed that UIM2 and UIM3 increase the catalytic activity and protein-binding ability of USP37 while UIM1 does not. They also revealed that the UIMs do not impart USP37’s preferential cleavage of K48 over K63 chains. Still unresolved were the questions of whether the UIMs influenced specificity towards any of the other Ub chains, exactly how the UIMs increase catalytic activity (for example at the level of Km or kcat), and whether the UIMs influence activity by binding the proximal or distal positions of a Ub chain. In the third chapter of my thesis I address these unresolved questions through a mutational biochemical analysis of USP37.
Chapter 2: Structural and functional characterization of a ubiquitin variant engineered for tight and specific binding to an alpha-helical ubiquitin interacting motif

Noah Manczyk#1,3, Bradley P. Yates#2,4, Gianluca Veggiangi2,4, Andreas Ernst5, Frank Sicheri1,2,3, Sachdev S Sidhu*,2,4

1Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada
2Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 3E1, Canada
3Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada
4Donnelly Centre for Cellular and Biomolecular Research, Banting and Best Department of Medical Research, University of Toronto, Toronto, ON M5S3E1, Canada
5Institute of Biochemistry II, Goethe University, Frankfurt am Main 60590, Germany

# Authors contributed equally to this work
* corresponding authors

A version of this chapter is published in Protein Science 2017 May;26(5)1060-1069

Attributions: Dr. Sidhu and his lab conceived the original project idea detailed in this chapter. This project started with Dr. Bradley Yates engineering, by phage display, 7 UbVs for UIM1 of Vps27 (Figure 2-1A). Bradley subsequently showed that one of the UbVs, namely UbV.v27.1, was highly specific for UIM1 of Vps27 out of a panel of 12 UIMs by phage ELISA binding assays (Figure 2-1B and 2-2). At this point, I became involved and was to understand the basis for how UbV.v27.1 bound to the UIM with high affinity. I determined the structure of UbV.v27.1 bound to the respective UIM1 by X-ray crystallography, analyzed the structure to determine binding mode and generated
hypotheses for the mechanism of high binding affinity (Figure 2-3). I then performed mutagenesis and binding experiments to test my hypotheses (Figures 2-4). Throughout Chapter 2, “I” is used where work was performed solely by me, “we” is used where I played a participating role, and the respective author(s) name such as “Bradley” where experiments were performed exclusively by others.

2.1 Introduction

Ubiquitin (Ub) is a highly abundant 76-amino acid protein that is expressed in every eukaryotic cell, and plays roles across a diverse range of cellular processes, including protein degradation \(^{94}\), DNA damage repair \(^{95,96}\), cell cycle regulation \(^{97,98}\) and a range of cell signaling pathways \(^{99}\). Ub exerts biological effects through covalent attachment to substrate proteins, in a process known as ubiquitination, or through non-covalent interactions with other proteins. Ubiquitination requires a coordinated cascade of reactions mediated by specialized enzymes (E1 activating enzymes, E2 conjugating enzymes and E3 ligating enzymes) to ultimately create an isopeptide bond between the C-terminal glycine residue of Ub and a lysine or free amino terminus on the substrate protein. The ubiquitination cascade can also add Ub to one of seven lysines within Ub itself to produce chains of various linkages. Distinct downstream effects are mediated by ubiquitination, depending on differences in the number of Ub moieties attached to a substrate and on the type of side-chain linkages used when ubiquitin is iteratively attached to itself to form Ub chains \(^{94,100,101}\). Alternatively, Ub can interact with other proteins in a non-covalent manner through specialized Ub-binding domains (UBDs) imbedded within many proteins. These non-covalent interactions can facilitate the ubiquitination reaction itself or serve other regulatory roles \(^{40}\).

Hundreds of UBDs have been found imbedded in functionally diverse proteins, either alone or in combination with Ub enzymes or other UBDs \(^{40,102,103}\). UBDs have been classified into over 20 structurally varied protein folds, which have been divided into five subfamilies: \(\alpha\)-helices, zinc finger (ZnF) domains, Ub-conjugating like (Ubc-like) domains, Plekstrin homology (PH) domains, and “others” \(^{103,104}\).
The α-helix subfamily is the largest, and it contains several distinct structural types, including Ub-interacting motifs (UIMs), motifs interacting with Ub (MIUs), double-sided UIMs (dUIMs), Ub-associated (UBA) domains, coupling-of-Ub-conjugation-to-endoplasmic reticulum degradation (CUE) domains, GGA and TOM (GAT) domains, and Vps/Hrs/STAM (VHS) domains 40,104.

Consisting of a single α-helix of approximately 20 amino acids, UIMs represent the smallest of the helical UBDs 102, and they are widely distributed, with an estimated 12 or 60 UIMs predicted across 7 or 31 proteins in yeast and humans, respectively 105,106. Despite their diminutive size and simple structure, UIMs have been implicated in many critical cellular pathways 104, including proteasomal degradation 45, endosomal sorting 107,108, multivesicular body biogenesis 40,53,109, DNA repair 110,111 and DNA methylation 112. In most cases, UIMs recognize ubiquitinated proteins directly and many proteins that contain UIMs are themselves ubiquitinated in a UIM-dependent manner 113,114. UIMs conform to a consensus sequence of e-e-x-x-φ-x-x-A-φ-x-(φ/e)-S-z-x-e (where e is an acidic residue, φ is a hydrophobic residue, z is a bulky hydrophobic or polar residue with high aliphatic content, A is alanine, S is serine and x is any residue) 42. Numerous Ub-UIM complex structures have been elucidated, revealing the basis for interaction 42,48,115. In particular, only ~490 ± 90 Å2 or ~470 ± 70 Å2 of surface area are buried by the interaction on the Ub and the UIM α-helix, respectively (based on a comparison of the following PDB entries: 2MBH, 1Q0W, 2D3G and 1YX5), and the conserved alanine residue of the UIM packs against Ile44 of Ub at the centre of the interface. In addition to UIMs, at least 15 of the 20 families of UBDs rely on contacts with a hydrophobic patch centered on Ile44 of Ub 37,104, with the same patch also being utilized in interactions with deubiquitinases (DUBs), E2 conjugating enzymes, and E3 ligating enzymes 116-118. These interactions tend to exhibit only moderate affinities, which in the case of UBDs are typically in the high micromolar range 40,102. This observation prompted the Sidhu lab’s investigation into whether this oft-used surface of Ub could be optimized to increase affinity and selectivity for specific binding partners.

Dr. Sidhu’s lab previously designed combinatorial phage-displayed libraries of Ub variants (UbVs) in which they diversified a large surface of ~2000 Å2 that included the Ile44 hydrophobic patch 89. These libraries proved to be remarkably fruitful in providing highly
specific and potent UbVs that could target diverse Ub enzymes both in vitro and inside cells. Structural analysis of UbVs in complex with DUBs \(^{89}\), HECT E3 ligases \(^ {90}\) and SCF E3 ligases \(^ {91}\) revealed the basis for improved affinity and stability, which depended on optimized interactions across large binding surfaces on UbVs ranging from 800 to 1900 Å².

Here, Bradley Yates of the Sidhu lab extends this work to explore whether the same approach can be applied to the small binding surface on Ub that mediates low affinity interactions with UIM alpha-helices. As a model system, Bradley targeted the first of two UIMs in the yeast protein Vps27. We show the UbV derived by Bradley binds to this UIM with an affinity that is improved almost 500-fold compared with that of wild-type Ub (Ub.wt). Moreover, the UbV exhibited high specificity for the cognate UIM relative to other yeast UIMs. I characterize the interaction between the UbV and UIM biochemically and structurally to understand the basis for enhanced affinity and specificity.

### 2.2 Experimental Procedures

#### 2.2.1 Selection and characterization of UbV-phage binding to yUIM-1

A peptide containing the first UIM of Vps27 (GGGGGAEEELIRKAIELSLKEKESSSGGY) was biotinylated with N-hydroxysuccinimidyl d-biotin-15-amido-4, 7, 10, 13-tetraoxapentadecylate (NHS-PEO\(_4\)-Biotin) in accordance with the manufacturer’s instructions (Thermo Fisher Scientific). The biotinylated peptide was immobilized in 96-well Nunc-Immuno MAXISORB plates (Thermo Scientific) coated with streptavidin (New England Biolabs), and phage pools representing UbV library 2 \(^ {89}\) were cycled through five rounds of binding selections with the immobilized peptide, as described \(^ {119,120}\). UbV-phage clones that bound to the peptide but not to streptavidin or BSA were identified by clonal phage ELISAs and were subjected to DNA sequence analysis to decode the sequences of the displayed UbVs, as described \(^ {89}\).

#### 2.2.2 Expression and purification of GST-UIM fusion proteins

A comprehensive list of 12 yeast UIM domains was curated through annotated domain sites found in the SMART and Uniprot databases \(^ {105,106}\). Mutagenic oligonucleotides were
designed to insert DNA encoding for each UIM in to the pH0103(TEV) vector (a gift from Cheryl Arrowsmith, Addgene plasmid # 64660) by site-directed mutagenesis using standard methods $^{119,121,122}$ to create open reading frames encoding for the UIM fused to the carboxy terminus of 6xHis-GST.

Plasmids designed to express the GST-UIM fusion proteins were transformed into *Escherichia coli* BL21(DE3) and single colonies were used to inoculate 5 ml selective 2YT media and cultures were grown overnight with shaking at 37 °C. The cultures were used to inoculate 1 L selective 2YT media and were grown with shaking at 37 °C to mid-log phase (OD$_{600}$ ~0.8). Protein expression was induced by the addition of 100 µM IPTG (Bio Basic), the temperature was lowered to 18 °C, and cultures were incubated overnight with shaking. Bacteria were pelleted and resuspended in 20 ml Lysis Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 10 mM imidazole, 1 mg/ml Lysozyme (BioShop Canada), 0.5% Triton-X 100, 20 U/ml Benzonase (EMD Millipore), and protease inhibitor cocktail (Sigma-Aldrich)). Cells were lysed by sonication, and protein purification was performed by incubating cell lysates with 1 ml slurry Ni-NTA resin (Qiagen) and eluting proteins with an imidazol buffer gradient ranging from 30-300 mM. The purity of eluted fractions were confirmed by SDS-PAGE, and protein concentrations were determined from OD$_{280}$ measurements with extinction coefficient from ExPASy ProtParam $^{123}$.

2.2.3 Expression and purification of UbV proteins

For crystallography, the gene encoding for UbV.v27.1 was cloned into the expression vector pHH0239 (a gift from Cheryl Arrowsmith, Addgene plasmid # 51323) to produce an open reading frame encoding for UbV.v27.1 with an N-terminal 6xHis-tag and an intervening cleavage site for the Tobacco Etch Virus (TEV) protease, as well as a C-terminal GAAA motif. The plasmid was transformed into *E. coli* BL21(DE3), a single colony was used to inoculate 5 ml selective 2YT media, and the culture was grown overnight with shaking at 37 °C. The culture was used to inoculate 1 L selective 2YT media, and 6xHis-UbV.v27.1 protein was expressed and purified as described above for GST-UIM fusion proteins. Purified 6xHis-UbV.v27.1 was dialyzed into TEV Cleavage Buffer (50 mM Tris pH 8.0, 1 mM 1,4-Dithiothreitol (DTT), 0.5 mM EDTA) and digested overnight with TEV protease at a 1:100 molar ratio of protease:substrate to remove the 6xHis-tag as described $^{3391}$. 
UbV.v27.1 protein was purified by applying the reaction mixture to Ni-NTA resin and collecting the unretained fractions. Protein purity was assessed by SDS-PAGE, and UbV.v27.1 protein was concentrated and buffer-exchanged into FPLC Buffer (PBS pH 7.4, 1 mM DTT) using Amicon Ultra-4 concentrators with a 3 kDa cutoff (EMD Millipore). The protein was further purified by gel filtration using the ÄKTA system (GE Healthcare) equipped with a Hi-Load 16/60 Superdex 75 size exclusion column (GE Healthcare Life Sciences). Protein purity was verified by SDS-PAGE, and UbV.v27.1 protein was concentrated as described above. Final assessment of purity was performed by SDS-PAGE, and protein concentration was determined from OD_{280} measurement with extinction coefficient from ExPASy ProtParam.  

For fluorescence polarization experiments, UbV.v27.1 was cloned into pH0239 and Ub.wt was cloned into pProEX-HTA (Life Technology) for expression as TEV cleavable N-terminally 6xHis-tagged fusions proteins. Both proteins were expressed in E. coli BL21(DE3) as previously described. Cells were resuspended in Lysis buffer (50 mM HEPES pH 7.5, 5 mM imidazole, 500 mM NaCl, 5 mM β-mercaptoethanol, 5% glycerol) and lysed by sonication. Cell lysate was loaded onto a 1-ml HiTrap Chelating HP column (GE Healthcare) and eluted by an imidazole buffer gradient ranging from 5-300 mM. Fractions containing UbV.v27.1 or Ub.wt were pooled, dialyzed in Dialysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM β-mercaptoethanol) to remove imidazole, and incubated overnight with TEV protease to cleave the 6xHis tag. After overnight incubation, TEV protease and uncleaved UbV.v27.1 or Ub.wt were removed by applying the reaction mixtures to a 1-ml HiTrap Chelating HP column (GE Healthcare). UbV.v27.1 or Ub.wt in the flow through fractions was concentrated with a 3 kDa cutoff Amicon Ultra-4 concentrator (EMD Millipore) and resolved by size exclusion chromatography (SEC) on a Superdex 75 16/600 column (GE Healthcare) previously equilibrated with 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT. Peaks corresponding to UbV.v27.1 or Ub.wt were pooled together, concentrated and flash frozen to -80 °C. Mutants of UbV.v27.1 and Ub.wt were purified in the same manner. The quadruple UbV.v27.1 mutant eluted as a double peak during SEC. In this case, each peak was collected, concentrated, flash frozen and assayed for binding to yUIM-1 separately.
2.2.4 Crystallization, structure determination, refinement and analysis

A peptide corresponding to residues 256-278 of the Vps27 protein from *Saccharomyces cerevisiae*, with a non-native N-terminal Tyr residue to aid concentration determination by A_{280} measurement (YPEDEELIRKAIELSLKESRNSA, Genscript), was suspended in sterile water for irrigation (Braun Medical, Inc). The peptide solution was mixed in a 1:1 molar ratio with a UbV.v27.1 protein solution to a final concentration of 500 µM. 1 µl UbV.v27.1/peptide solution was mixed with 1 µl mother liquor containing 2.1 M DL-Malic acid pH 7.0, and incubated as hanging drops over lower wells containing 350 µl mother liquor. Crystals were harvested after 17 days, placed in mother liquor supplemented with 25% glycerol (vol/vol), and flash frozen in liquid nitrogen.

A single crystal dataset was collected at -180 °C on a home-source consisting of a Rigaku MicroMax-007 HF rotating anode generator coupled to a Rigaku Saturn 944 HG CCD detector, and was processed using HKL2000. The structure was solved by molecular replacement using PHASER and a search model of one molecule of yUIM-1 (PDB ID 1Q0W) and one molecule of Ub with the 5 C-terminal residues removed (PDB ID 1UBQ). The structure was refined by REFMAC and PHENIX, both using TLS parameters, and manual building in COOT. Interactions between yUIM-1 and UbV.v27.1 were analyzed using the protein interfaces, surfaces and assemblies (PISA) tool. Side chain and backbone atoms for residues Gly6 – Ala0, Thr9 and Arg77 – Ala82 of UbV.v27.1 and Lys17* – Ala23* of yUIM-1 could not be modeled. Side chain atoms other than the Cβ atoms for Gln2, Glu16, Glu18, Lys29, Ser57, Lys62, Arg74 and Ser75 of UbV.v27.1 and Glu2* of yUIM-1 could not be modeled.

Coordinate and structure factors for the UbV.v27.1-yUIM-1 complex have been deposited in the Protein Data Bank with accession number 5UCL.

2.2.5 Competitive phage ELISAs

Competitive phage ELISAs were performed as described. Briefly, GST-yUIM-1 fusion protein was immobilized in a 384-well Nunc MAXISORP plate (Thermo Scientific) and the plate was blocked with BSA. A sub-saturating concentration of UbV.v27.1-phage was
incubated for 1 hour at 25 °C with serial dilutions of GST-UIM fusion protein and the mixtures were transferred to the plate containing immobilized GST-yUIM-1. After 15 minutes of incubation, the plates were washed and bound phages were labeled with α-M13 antibody/HRP conjugate (1:5,000 dilution, GE Healthcare Life Sciences) for 30 minutes. The plate was washed and developed with 3,3',5,5'-Tetramethylbenzidine Peroxidase Substrate solution (SeraCare), and absorbance at 450 nm was measured using a PowerWave XS plate reader (BioTek). Data were plotted with the Prism 6 software (GraphPad Software Inc) and curves were fitted using a one-site binding Hill model. IC\textsubscript{50} values were determined as the concentration of solution-phase GST-UIM protein that reduced 50% binding of UbV.v27.1-phage to immobilized GST-yUIM-1.

2.2.6 Fluorescence Polarization Binding Experiments

A peptide corresponding to residues 256-278 of Vps27 (YPEDEEELIRKAIELSLKESRNSAK, Bio Basic Inc.) was suspended in 100 mM HEPES pH 7.5, 100 mM NaCl. Subsequently, pH was adjusted to 6.5 by adding 1 M NaOH and 14.8 M ammonium hydroxide to approximate final concentrations of 5 mM and 30 mM, respectively. The peptide contained a non-native N-terminal tyrosine to aid concentration determination and a non-native C-terminal lysine to allow covalent labeling with 5/6-carboxyfluorescein succinimidyl ester (NHS-FITC) (Bio Basic Inc.). The C-terminus was amidated and the N-terminus was acetylated to prevent charge effects from the termini affecting peptide binding. Peptide concentration was determined from absorbance measurement at 495 nm using the extinction coefficient of FITC (75,000 L mol\textsuperscript{-1} cm\textsuperscript{-1}). Binding measurements were performed in FP buffer (25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT, 0.01 mg/ml BSA, 0.03% BRIJ-35) by mixing in a 96-well plate 25 nM FITC-labeled yUIM-1 peptide with serial dilutions of Ub.wt and its variants ranging from 1 mM to 22 nM. For UbV.v27.1 and its variants, binding measurements were performed similarly with concentrations varying from 50 µM to 0.28 nM. Samples were equilibrated at room temperature for 30 minutes before reading plates on a HTS Multi-Mode Microplate Reader (Synergy Neo) using an excitation filter of 485 nm and an emission filter of 530 nm. Dissociation constants were determined with Prism (GraphPad Software Inc) using a one-site total binding model. For the quadruple mutant of
UbV.v27.1, which eluted as two distinct peaks by SEC, binding measurements were made on both species.

2.3 Results

2.3.1 UbV binders for the first UIM domain of Vps27

Bradley used a biotinylated synthetic peptide representing the first UIM of the yeast protein Vps27 (yUIM-1) as the bait for binding selections with “library 2”, a previously described phage-displayed library of UbVs. Library 2 was constructed by a “soft randomization” strategy whereby 27 positions across a large surface including the Ile\textsuperscript{44} hydrophobic patch were diversified in a manner such that each position contained approximately 50\% wild-type sequence and 50\% random mutations. Following five rounds of binding selections, individual phage clones were assessed for specific binding by phage ELISAs and clones that bound to the target peptide but not to streptavidin and several other negative control proteins were subjected to DNA sequence analysis, which revealed seven unique UbV sequences, all characterized by a common substitution of Val for Gly at position 10 (Figure 2-1A). Based on qualitative phage ELISAs, Bradley chose a specific UbV (named UbV.v27.1) that exhibited a strong ELISA signal for binding to peptide yUIM-1 but not to a panel of other Ub-associated proteins for detailed functional and structural characterization.
Figure 2-1. Selective binding of UbV.v27.1 to yUIM-1.

(A) Sequence alignment of UbV.v27.1 and other UbVs selected for binding to yUIM-1. The alignment shows only those positions that were diversified in the UbV library, and positions that were conserved as the wt sequence are shown as dashes. Sequences showing conservation across selected UbVs are highlighted in grey.

(B) Sequence alignment of yeast UIMs and IC₅₀ values for inhibition of UbV.v27.1 binding to immobilized yUIM-1. Residues (arbitrarily numbered from 3 to 22) that conform to the UIM consensus are highlighted in grey, and the consensus is shown below (e is an acidic residue, φ is a hydrophobic residue, z is a bulky hydrophobic or polar residue with high aliphatic content, A is alanine, S is serine, and x is any residue).

"Sequence number" denotes the position of each UIM domain in the full-length protein sequence according to the UniProt database. IC₅₀ values were defined as the concentration of solution-phase GST-UIM fusion protein that inhibited 50% of the UbV.v27.1 binding to immobilized yUIM-1. "NDI" denotes “no detectable inhibition” with 60 µM GST-UIM.
2.3.2 Affinity and specificity of UbV.v27.1

To assess the affinity and specificity of UbV.v27.1, Bradley measured affinities for all the yeast UIMs by competitive phage ELISAs. Bradley purified each of the 12 yeast UIMs as fusions with the carboxy terminus of GST and used serial dilutions of each of these fusion proteins as competitors for the interaction of immobilized yUIM-1 with solution-phase, phage-displayed UbV.v27.1. From the binding curves (Figure 2-2), Bradley was able to determine IC$_{50}$ values, which were defined as the concentration of solution-phase GST-UIM fusion protein that inhibited 50% of the UbV.v27.1-phage binding to immobilized yUIM-1 (Figure 2-1B). In this assay, UbV.v27.1 exhibited high affinity for yUIM-1 (IC$_{50}$ = 0.28 µM), moderate affinity for the third UIM of UFO1 (yUIM-5, IC$_{50}$ = 21 µM) and no detectable binding to the other 10 yeast UIMs. Thus, UbV.v27.1 exhibited high affinity and specificity for yUIM-1 relative to the set of 12 yeast UIMs.
Figure 2-2. Binding affinity and specificity of UbV.v27.1.

The specificity and affinity of UbV.v27.1-phage was assessed by analyzing the binding of the UbV to immobilized yUIM-1 in presence of varying concentrations of solution-phase GST-UIM fusion proteins as competitor. IC_{50} values were defined as the concentration of GST-UIM that inhibited the binding of UbV.v27.1-phage to yUIM-1 by 50% (mean of duplicate ±1 SD). Refer to Figure 2-1 for naming convention.
2.3.3 Structure of UbV.v27.1 in complex with yUIM-1

To understand the molecular basis for how UbV.v27.1 binds to yUIM-1 with greatly enhanced affinity and specificity, I determined the crystal structure of the complex to 2.35 Å resolution (Table 2-1 and Figure 2-3). Superposition of the complex structure with the NMR structure of the interaction of Ub.wt with yUIM-1 showed that UbV.v27.1 bound to yUIM-1 in the same orientation as did Ub.wt (Figure 2-3A, RMSD = 0.94 Å). Notably, of the total of 18 substitutions in UbV.v27.1 relative to Ub.wt, only six (Ile8, Met48, Asn66, Tyr68, Ser71, Ser72) were in close proximity to yUIM-1 (within 6 Å, Figure 2-3B), and the side-chains of Met48 and Ser71 pointed away from yUIM-1. The binding interface involved 497 Å² and 496 Å² of buried surface area on UbV.v27.1 and yUIM-1, respectively (Figure 2-3C), and was comprised of a mixture of hydrophobic and hydrophilic contacts. This was similar to but somewhat smaller than the binding interface of the native interaction, which involved 558 Å² and 552 Å² of buried surface area on Ub.wt and yUIM-1, respectively. Hydrophobic interactions were mediated by contacts between UbV.v27.1 residues Ile8, Ile44, Ala46, Tyr68 and Val70 and yUIM-1 residues Leu7*, Ile8*, Ala11*, Ile12* and Leu14* (Figure 2-3D, UIM residues are denoted by asterisks throughout this chapter). Hydrophilic interactions included a hydrogen bond between the side-chains of Glu4* and Ser72, hydrogen bonds between the side-chain of Tyr0* (a residue that is not present in the native yUIM-1 but was added to facilitate assays described in the Methods) and the backbone amide of Leu73 and side-chain of Ser72, a hydrogen bond between the side-chain of Ser15* and the backbone amide of Gly47, and lastly, a salt bridge between Glu4* and Arg42 (Figure 2-3D).
Table 2-1. Crystallographic data collection, processing, and refinement statistics.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>UbV.v27.1/yUIM-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P43212</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
</tr>
<tr>
<td>Cell dimension</td>
<td></td>
</tr>
<tr>
<td>(a, b, c) (Å)</td>
<td>44.61, 44.61, 104.34</td>
</tr>
<tr>
<td>(\alpha, \beta, \gamma) (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50 – 2.35</td>
</tr>
<tr>
<td>(R_{sym})</td>
<td>0.059 (0.302)</td>
</tr>
<tr>
<td>CC (1/2)</td>
<td>(0.96)</td>
</tr>
<tr>
<td>Total no. of observations</td>
<td>44385</td>
</tr>
<tr>
<td>Total no. unique observations</td>
<td>4788 (424)</td>
</tr>
<tr>
<td>Mean ([</td>
<td>I</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.3 (90.6)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>9.1 (4.1)</td>
</tr>
</tbody>
</table>

| Refinement       |                   |
| Resolution (Å)   | 41.0 – 2.35       |
| No. of reflections | 4521            |
| \(R_{work}/R_{free}\) | 22/25          |
| No. of atoms     |                   |
| Protein          | 711               |
| Ligand/ion       | 0                 |
| Water            | 8                 |
| Average B-factors|                   |
| Protein          | 63.8              |
| Ligand/ion       | N/A               |
| Water            | 55.65             |
| R.m.s.d          |                   |
| Bond length (Å)  | 0.0135            |
| Bond angles (°)  | 1.82              |

**Ramachandran statistics**

- Residues in favoured regions (%) 100
- Residues in allowed regions (%) 0
- Residues in disallowed regions (%) 0

Statistics in brackets are for highest resolution shell
Figure 2-3. The crystal structure of the UbV.v27.1-yUIM-1 complex.

(A) Superposition of the crystal structure of UbV.v27.1 bound to yUIM-1 with the NMR structure of Ub.wt bound to yUIM-1 (PDB entry: 1Q0W). (B) Substitutions in UbV.v27.1 relative to Ub.wt. The UbV.v27.1 backbone is shown as a green ribbon with side-chains of substitutions colored red or cyan for residues ≤6 Å or >6 Å from yUIM-1, respectively. Of note, the entirety of the side chains for Lys62 and Ser75 could not be modeled. (C) Contact surfaces at the interface between UbV.v27.1 and yUIM-1. The complex is shown in an open book view with wt and substituted contact residues colored pink or red, respectively, and non-contact residues on UbV.v27.1 and yUIM-1 colored green or orange, respectively. (D) Details of the molecular interactions between UbV.v27.1 and yUIM-1. Polar contacts are shown between Ser15* and Gly47 (left panel), between Tyr0* and Ser72 and Leu73, and between Glu4* and Arg42 and Ser72 (middle panel). Hydrophobic
interactions are shown between Ile\textsuperscript{44} and Ile\textsuperscript{8*} and Ile\textsuperscript{12*} (left panel) and between Leu\textsuperscript{7*}, Ile\textsuperscript{8*}, Ala\textsuperscript{11*}, Ile\textsuperscript{12*} and Leu\textsuperscript{14*}, and Ile\textsuperscript{9}, Ile\textsuperscript{44}, Tyr\textsuperscript{68} and Val\textsuperscript{70} (right panel). Asterisks indicate yULM-1 residues.
2.3.4 Site-directed mutagenesis of UbV.v27.1 and Ub.wt

To further dissect the basis for the improved affinity of UbV.v27.1 for yUIM-1, I subjected the UbV to site-directed mutagenesis and measured affinities for yUIM-1 by a fluorescence polarization assay. In this assay, Ub.wt bound with a $K_d$ of 114 µM, which was in reasonable agreement with a published value (277 µM) $^{24}$, and UbV.v27.1 bound with a $K_d$ of 0.26 µM, which was in good agreement with the IC$_{50}$ value determined by competitive phage ELISA (0.28 µM, Figure 2-1B). Only four of the 18 substitutions in UbV.v27.1 (Ile8, Asn66, Tyr68, Ser71) were in close proximity with their side-chains directed towards yUIM-1 (Figure 2-3B), but surprisingly, individual back mutation of each of these four residues in UbV.v27.1 to the wt sequence revealed that only the Tyr68His substitution reduced affinity appreciably (~17-fold, Figure 2-4A). To assess whether these four residues exhibit cooperativity, I generated a quadruple mutant in which the four positions in UbV.v27.1 were simultaneously reverted to the wt sequence (Ile8Leu, Asn66Thr, Tyr68His, Arg72Ser). During the purification of the quadruple UbV.v27.1 mutant, I noted two distinct peaks from a size exclusion chromatography column. Therefore, I purified each peak individually and tested each species for binding to yUIM-1. The first and second peaks, which corresponded to larger or smaller effective molecular weights, respectively, displayed decreases in affinity of 10-fold or ~35-fold, respectively, which were similar to the decrease observed for the single substitution Tyr68His mutant. Thus, amongst the four contact residues that were substituted in UbV.v27.1 relative to Ub.wt, I concluded that only Tyr68 contributed significantly to enhanced affinity.
Figure 2-4. Effects of substitutions on UbV.v27.1 and Ub.wt binding to yUIM-1.

Fluorescence polarization binding experiments are shown for the binding of yUIM-1 to UbVs harboring the indicated substitutions in the background of (A) UbV.v27.1 or (B) Ub.wt. Fluorescence polarization (y-axis) was measured for varying concentrations of UbVs (x-axis) and 25 nM yUIM-1 peptide (mean of triplicate ±1 SD). For the quadruple substitution in the UbV.v27.1 background, “large” and “small” indicate measurements for the first or second peak that eluted from a SEC column, respectively. ND indicates no detectable binding.
Since the Gly10Val substitution remote from the contact surface of UbV.v27.1 with yUIM-1 was common to all 7 UbVs selected in the phage display experiment (Figure 2-4A), I reasoned that it might also contribute to the enhanced binding affinity of UbV.v27.1 for yUIM-1. Indeed, the back mutation Val10Gly in UbV.v27.1 caused almost a complete loss of binding affinity for yUIM-1 (more so than the Tyr68His substitution). Furthermore, introduction of both Val10Gly and Tyr68His resulted in a UbV with almost no binding affinity for yUIM-1 at the concentration range tested. Thus, I concluded that the Gly10Val and His68Tyr substitutions were both key mediators of the enhanced binding affinity of UbV.v27.1 for yUIM-1.

I also examined whether individual introduction of any of the substitutions from UbV.v27.1 into Ub.wt was sufficient to enhance the affinity for yUIM-1 (Figure 2-4B). Consistent with the back mutation data described above, only one of the contact surface substitutions, namely His68Tyr, enhanced the binding affinity (~13-fold increase). Furthermore, simultaneous introduction of all four contact substitutions from UbV.v27.1 into Ub.wt resulted in a weaker enhanced affinity (only 6-fold) relative to the His68Tyr substitution alone. Taken together, these data confirm that of the four substitutions in UbV.v27.1 on the contact surface with yUIM-1, only Tyr68 contributes significantly to enhanced affinity for yUIM-1. I next tested the effect of the Gly10Val substitution in the Ub.wt background. Similar to the effect of the His68Tyr single substitution, the Gly10Val substitution enhanced binding affinity approximately 8-fold. Moreover, introduction of both Gly10Val and His68Tyr caused a greater enhancement of binding (19 fold), but the resultant affinity (Kd = 6 μM) was still significantly lower than the affinity of UbV.v27.1 for yUIM-1.

2.4 Discussion

The large and diverse array of cellular processes regulated by the Ub system make ubiquitination an attractive process for therapeutic intervention. Design and implementation of phage-displayed libraries to identify UbVs that bind with enhanced affinity to specific targets has been a boon to such research. Recent applications of this approach have been successful in producing UbVs capable of modulating particular
biological functions either as inhibitors or activators of Ub enzymes. Here, Bradley extends this body of work by applying the UbV technology to the UIM subfamily of UBDs.

UbV.v27.1 proved to bind tightly and specifically to yUIM-1 from yeast Vps27. UbV.v27.1 bound almost 500-fold more tightly to yUIM-1 than did Ub.wt (Figure 2-4), and binding was highly specific, as only one of the other 11 UIMs in the yeast proteome was recognized, and even its binding was ~100-fold weaker (Figure 2-1B). The structure of UbV.v27.1 in complex with yUIM-1 revealed that the UbV binds in the same manner as Ub.wt, and only four of the 18 substitutions relative to Ub.wt have side-chains that are in close proximity to and point towards the UIM (Figure 2-3). Unexpectedly, mutagenesis studies of these four substitutions revealed that only the His68Tyr substitution contributed significantly to enhanced binding (Figure 2-4). Further investigation revealed that the Gly10Val substitution remote from the contact surface also contributed to the improved binding affinity for yUIM-1. However, introduction of both substitutions (Gly10Val and His68Tyr) into Ub.wt was not sufficient to impart the full enhancement of affinity displayed by UbV.v27.1 for yUIM-1. These results suggest possible contributions from other substitutions remote from the contact surface to the enhanced binding affinity.

While it is easy to rationalize how the His68Tyr substitution contributes to enhanced binding affinity of UbV.v27.1 for yUIM-1, it is less clear how the Gly10Val substitution exerts its influence as this substitution is remote from the yUIM-1 contact surface. The Gly10Val substitution instead occurs in a tight turn between strands β1 and β2 of Ub.wt (Figure 2-3B). However, I note that the Thr9 residue in the middle of the turn could not be unambiguously modeled. The manner by which the Gly10Val substitution increases the binding affinity of UbV.v27.1 for yUIM-1 is the focus of Chapter 3 in this thesis.

In this study, Bradley subjected Ub to variation to find a highly specific and potent binder for yUIM-1. As there are a large number of UIMs with varying sequences and only a single conserved Ub, it would be interesting to perform the reverse experiment of subjecting a UIM to variation to find a high affinity binder to Ub. Understanding which residues of a UIM can be mutated to enhance affinity for Ub may give additional and complementary insight into the molecular basis for UIM-Ub interactions.
UbVs engineered for binding to several Ub enzymes have been shown previously to recognize their targets \textit{in vivo} \cite{89-91}. Importantly, these UbVs also affected the activity of these enzymes and consequently influenced cellular pathways in which the enzymes were involved. Mutation of yUIM-1 in Vps27 has been shown to cause defective sorting in multivesicular bodies \cite{109}, and it may be interesting to determine if UbV.v27.1 can disrupt Vps27-dependent signaling in yeast cells. Even more intriguingly, the UbV strategy could be extended to targeting the 60 UIMs in the human proteome to reveal new biological functions and potential avenues for therapeutic intervention.
Chapter 3: Dimerization of a ubiquitin variant leads to high affinity interactions with a ubiquitin interacting motif

Noah Manczyk\textsuperscript{1,2}, Gianluca Veggiani\textsuperscript{3,4}, Gerald D. Gish\textsuperscript{1}, Bradley P. Yates\textsuperscript{3,4}, Andreas Ernst\textsuperscript{5}, Sachdev S Sidhu\textsuperscript{3,4}, Frank Sicheri\textsuperscript{1,2,3}

\textbf{Affiliations}

\textsuperscript{1}Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada
\textsuperscript{2}Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada
\textsuperscript{3}Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 3E1, Canada
\textsuperscript{4}Donnelly Centre for Cellular and Biomolecular Research, Banting and Best Department of Medical Research, University of Toronto, Toronto, ON M5S3E1, Canada
\textsuperscript{5}Institute of Biochemistry II, Goethe University, Frankfurt am Main 60590, Germany

\textit{A version of this chapter is published in Protein Science 2019 February; doi: 10.1002/pro.3593}

\textbf{Attribution:} Dr. Sicheri and I conceived the original project idea, which is an extension of the work presented in Chapter 2. I performed all experiments and work presented in this chapter with the exception of peptide synthesis, which was performed either by scientists at Bio Basic Inc. or by Dr. Gish in the Sicheri lab.
3.1 Introduction

We previously reported a ubiquitin (Ub) variant (UbV), denoted UbV.v27.1 (see Figure 3-1A for sequence comparison between UbV.v27.1 and wild type Ub (Ub.wt)), that bound to the first Ub interacting motif (UIM) of the yeast protein Vacuolar protein sorting-associated protein 27 (Vps27 UIM1 (yUIM-1), see Figure 3-1B for yUIM-1 sequence) with ~500-fold higher affinity than Ub.wt. I determined the crystal structure of yUIM-1 in complex with UbV.v27.1, which revealed a 1:1 binding stoichiometry that was very similar to the canonical binding mode of Ub.wt to UIMs. Mutational and functional analysis identified two substitutions (Gly10Val/His68Tyr) in UbV.v27.1 relative to Ub.wt that were responsible in large part for the observed binding affinity of the UbV.v27.1 for yUIM-1. From the structure, I could rationalize the importance of the Tyr68 substitution in its specific contribution of a ~13-fold increase in affinity when introduced in Ub.wt, as the tyrosine side chain made extensive favorable hydrophobic contacts directly with yUIM-1. However, I could not rationalize why the Val10 substitution when introduced on its own into Ub.wt was responsible for an ~8-fold increase in binding affinity, as the Val side chain did not participate in the interface with yUIM-1. Indeed, Val10 was >7 Å away from the closest residue in yUIM-1 in my original structure. Here I have revisited my structural and functional analysis of the UbV-UIM complex, discovering that the Gly10Val substitution in UbV.v27.1 contributes to the enhanced affinity of the UbV for yUIM-1 through its unexpected ability to induce a dimer state.
Figure 3-1. UbV.v27.1 is an oligomeric protein in solution.

(A) Sequence alignment of Ub.wt, UbV.v27.1, UbV.XR, and UbV.15.D. Only positions that differ between Ub.wt and one of the UbVs are depicted. Positions that are conserved with Ub.wt are shown as dashes. UbV.15.D contains an insert of two amino acids designated as positions 10a and 10b. (B) Sequence of the yUIM-1 peptide. The non-native tyrosine residue at position 0 was included in the synthesis to aid concentration determination by absorbance at 280 nm. (C,D) Preparative scale (120-ml column volume) size exclusion chromatography elution profiles of UbV.v27.1 with the indicated back substitutions or Ub.wt with the indicated forward substitutions, respectively. The profiles of UbV.v27.1 and Ub.wt are shown in each plot as references. (E) Analytical scale (24-ml column volume) size exclusion chromatography elution profiles of Ub.wt, UbV.v27.1 and a UbV.v27.1/yUIM-1 complex. The elution volumes for molecular weight standards are

| Residue Numbers | UIM name | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|-----------------|----------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
|                 |          | F | K | L | T | G | K | T | T | I | A | K | Q | K | E | T | H | V | L | R | L | R | G | G |     |
| UbV.v27.1       |          | Q | I | V | M | R | A | - | - | M | K | R | D | N | Y | - | S | S | - | - | S | L | R | A |     |
| UbV.XR          |          | L | T | I | S | A | E | R | F | E | - | N | - | K | L | - | V | K | F | H | R | V | A | S |     |
| UbV.15.D        |          | S | A | K | F | C | G | - | L | - | - | - | - | - | - | G | Q | H | R | - | - | P | T | *   |
indicated along with the inferred molecular weights of Ub.wt, UbV.v27.1, yUIM-1, and the UbV.x27.1/yUIM-1 complex. As the molar absorptivity of yUIM-1 is much larger than UbV.x27.1 and Ub, the absorbance profiles for the presented plots were normalized to the largest peak height.
3.2 Experimental Procedures

3.2.1 Protein expression and preparative scale purification

Genes encoding UbV.v27.1 and Ub.wt were cloned into pH0239 or pProEX-HTA (Life Technologies), respectively, for expression as TEV cleavable N-terminally 6xHis-tagged fusion proteins. Protein was expressed in E. coli BL21(DE3) as previously described. Cells were resuspended in Lysis buffer (50 mM HEPES pH 7.5, 5 mM imidazole, 500 mM NaCl, 5 mM β-mercaptoethanol, 5% glycerol) and lysed by sonication. Cell lysate was loaded onto a 1-ml HiTrap Chelating HP column (GE Healthcare) and eluted by an imidazole buffer gradient from 5-300 mM. Fractions containing UbV.v27.1 or Ub.wt were pooled, dialyzed in dialysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM β-mercaptoethanol) to remove imidazole, and incubated overnight with TEV protease to cleave the 6xHis tag. After overnight incubation, TEV protease and uncleaved UbV.v27.1 or Ub.wt was removed by applying the reaction mixtures to a 1-ml HiTrap Chelating HP column (GE Healthcare). UbV.v27.1 or Ub.wt in the flow through fractions was concentrated with a 3-kDa cutoff Amicon Ultra-4 concentrator (EMD Millipore). Proteins were then injected (1 ml of ~1-9 mg/ml) on a preparative 120 ml bed volume Superdex 75 16/600 column (GE Healthcare) previously equilibrated with size exclusion buffer (25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT). Other UbVs were purified in the same manner. Absorbance at 280 nm was used to monitor eluted proteins. Eluted fractions were analysed by SDS-PAGE and pure proteins were pooled, concentrated, flash frozen in liquid nitrogen, and stored at -80 °C.

3.2.2 Peptide synthesis for analytical scale size exclusion chromatography

The yUIM-1 peptide used for size exclusion chromatography analysis was obtained from Bio Basic Inc. The sequence, YPEDEELIRKAIELSLKESRNSAK, corresponds to residues 256-278 of Vps27 with the addition of a non-native N-terminal tyrosine to aid concentration determination and a non-native C-terminal lysine to enable covalent labeling with 5/6-carboxyfluorescein succinimidyl ester (NHS-FITC). To prevent charge effects from the termini affecting peptide binding, the C-terminus was amidated and the N-terminus was acetylated. The peptide was resuspended in a buffered solution as previously described.
3.2.3 Analytical scale size exclusion chromatography

For the analyses of UbV.v.27.1 and Ub.wt, proteins were thawed and resuspended in size exclusion buffer to a volume of 500 µl and a concentration of 2 mg/ml. For the analysis of the UbV.v27.1/yUIM-1 complex, UbV.v27.1 was resuspended with yUIM-1 in size exclusion buffer to a volume of 500 µl and a final concentration of 2 mg/ml and 1 mg/ml of UbV.v27.1 and yUIM-1 (representing a 1:1.5 molar ratio), respectively. The UbV.v27.1/yUIM-1 complex was equilibrated on ice for 40 minutes. Proteins were then injected onto an Enrich SEC 70 10 x 300 mm column (Bio-Rad) column previously equilibrated with size exclusion buffer. Molecular weight standards (Bio-rad, product #1511901) including γ-globulin, ovalbumin, myoglobin, and vitamin B12 were analyzed according to the manufacturers instructions. A standard curve was generated by plotting log(molecular weight) versus $K_{av}$ ($K_{av}=(V_e-V_o)/(V_t-V_o)$, where $V_e$ is the observed elution volume, $V_o$ is the void volume (approximated as 1/3 of the total column volume), and $V_t$ is the total column volume.

3.2.4 Refinement and structural analysis

Models of the monomeric β1-β2 loop and dimeric β1 strand swap UbV.v27.1 were manually built in Coot using the previous structure (PDB:5UCL) as a template. Both models were then refined against the previously collected, integrated and scaled dataset using identical refinement strategies in Phenix. The updated coordinate file (PDB:6NJG) for PDB:5UCL has been deposited to the PDB for release upon publication. Interactions between UbV.v27.1 and yUIM-1 were analyzed using the PyMOL Molecular Graphics System (Schrodinger, LLC) and the protein interfaces, surfaces and assemblies (PISA) tool. Figures were generated using the PyMOL Molecular Graphics System.

3.2.5 yUIM-1 peptide synthesis for mutational analysis

yUIM-1 peptide (sequence corresponds to residues 256-278 of Vps27) and yUIM-1 peptides harbouring substitutions were produced by solid phase peptide synthesis in-house using 9-fluorenylmethoxycarbonyl chemistry on Rink amide MBHA resin (Novabiochem) on a Prelude peptide synthesizer (Protein Technologies, Inc.). Peptides were deprotected in the cleavage cocktail trifluoroacetic acid, phenol, water, thioanisole, 1,2-ethanediethiol (82.5%:5%:5%:5%:2.5% v/v) for 90 min at room temperature and then
precipitated in t-butyl methyl ether. Crude peptides were purified using C-18 reverse phase HPLC (Waters) and authenticity was confirmed by mass spectrometry on an Orbitrap Elite (ThermoFisher Scientific). C-terminal labelling was achieved through Cysteine derivation using 5-(Iodoacetamido) fluorescein (Sigma Aldrich). The peptide contained a non-native N-terminal tyrosine to aid concentration determination and a non-native C-terminal cysteine to allow covalent labeling with 5/6-carboxyfluorescein succinimidyl ester (NHS-FITC). Peptides were resuspended in 20 mM HEPES pH 7.0 and concentrations were determined from absorbance measurements at 495 nm using the extinction coefficient of FITC (75,000 l mol⁻¹ cm⁻¹).

3.2.6 Fluorescence Polarization binding assay

Binding measurements were performed in FP buffer (25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT, 0.1 mg/ml BSA, 0.03% BRIJ-35) by mixing, in a 96-well plate, 25 nM FITC-labeled peptides with the sequence of yUIM-1, or yUIM-1 with the indicated substitutions, with serial dilutions of UbV.v27.1 ranging from 0.28 nM to 50 μM. Samples were equilibrated at room temperature for 30 min before reading plates on a HTS Multi-Mode Microplate Reader (Synergy Neo) using an excitation filter of 485 nm and an emission filter of 530 nm. Dissociation constants were determined using Prism (GraphPad Software Inc) using a one-site total binding model.

3.3 Results

3.3.1 UbV.v27.1 is a dimer

In the process of purifying Ub, UbV.v27.1 and a series of intermediate UbVs used to probe the necessity and sufficiency of specific substitutions towards the enhancement of affinity for yUIM-1, I observed that Ub.wt eluted at a volume consistent with a smaller size than UbV.v27.1 (Figure 3-1C). I initially reasoned that the larger apparent size of UbV.v27.1 in solution reflected non-specific aggregation at high protein concentrations due to the numerous (18 in total) substitutions relative to Ub.wt.

Variants of UbV.v27.1 with individual back substitutions to the respective residue in Ub.wt (Ile8Leu, Val10Gly, Asn66Thr, Tyr68His, or Ser72Arg) also eluted as apparent oligomers with the exception of the Val10Gly back mutant. This mutant eluted at the same volume as
Ub.wt (Figure 3-1C), indicating that the Val10Gly substitution was essential for maintaining an oligomeric state. However, variants of Ub.wt with single-site substitutions to the respective residues found in UbV.v27.1 (Leu8Ile, Gly10Val, Thr66Asn, His68Tyr, or Arg72Ser) all eluted as mono-dispersed monomers, indicating that no single substitution in the Ub.wt background was sufficient to induce an oligomeric state (Figure 3-1D). In sum, these results suggested that the Gly10Val substitution is necessary but not sufficient to drive the oligomerization of UbV.v27.1, providing a first clue of how Val10 might contribute to the enhanced binding affinity of UbV.v27.1 for yUIM-1.

To gain insight into how the Gly10Val substitution might contribute to the oligomeric state of UbV.v27.1, I reanalyzed the crystal structure of UbV.v27.1 in complex with yUIM-1. In the published structure (PDB entry 5UCL), Val10 was immediately adjacent to Thr9, which was left unmodeled due to ambiguity in interpreting the connectivity of the β1-β2 loop (Figure 3-2A). With a realization that UbV.v27.1 forms an oligomer in solution, I hypothesized that the difficulty in modeling Thr9 may be due to its participation in an unexpected strand exchange mechanism between two crystallographic symmetry related protomers rather than a conventional β1-β2 loop observed in the Ub.wt monomer (Figure 3-2Bi top left panel versus top right panel). In the strand exchange configuration, residues 8-10 of UbV.v27.1 adopt a linear conformation that seamlessly links strand β1 from one protomer to strand β2 of the second protomer. I termed this extended β-strand, β1’. The correctness of the strand exchange model was confirmed by the improved fit of 2Fo-Fc and Fo-Fc electron density maps and by the refinement statistics (Figure 3-2BC).
Figure 3-2. UbV.v27.1 is a β-strand exchanged dimer in the crystal structure.

(A) Architecture of UbV.v27.1 (UbV) and a crystal symmetry related UbV* (PDB entry: 5UCL). Zoom in window depicts the 2Fo-Fc map of model in PDB:5UCL around the unmodeled Thr9 position. (B) Comparison of β1-β2 loop and β1-swap models for the UbV.v27.1 structure. The panels depict: (i) ribbon cartoons of the affected region for each model, (ii) the respective 2Fo-Fc map, (iii) the respective Fo-Fc map, (iv) the R_work/R_free factors for each model after refinement using Phenix. (C) Data collection and refinement statistics for the β1-swap UbV.v27.1-yUIM-1 crystal structure.
To investigate the actual size of the UbV.v27.1 oligomer in solution, I reanalyzed UbV.v27.1 using an analytical scale size exclusion column calibrated with molecular weight markers. Surprisingly, UbV.v27.1 eluted at a volume corresponding to a molecular weight (50 kDa) much larger than the theoretical molecular weight of a dimer (19.5 kDa) (Figure 3-1E). I reasoned that this discrepancy might be due to further oligomerization of UbV.v27.1 in the absence of a yUIM-1 binding partner. Indeed, when UbV.v27.1 was mixed with a molar excess of yUIM-1, its apparent molecular weight down-shifted to 30 kDa (Figure 3-1E), which closely coincided with the expected molecular weight of the UbV.v27.1:y-UIM complex (26 kDa for a 2:2 binding stoichiometry) observed in the crystal structure. These observations supported the notion that UbV.v27.1 was in fact a functional dimer both in solution and in the crystal environment.

3.3.2 Revised binding mode and stoichiometry of UbV.v27.1 to yUIM-1

The strand exchange structure of UbV.v27.1 resulted in an unexpected 2:2 binding stoichiometry for yUIM-1, unlike the 1:1 binding stoichiometry of the previous model (Figure 3-3AB). Notably, with the UbV.v27.1 dimer, each UIM now makes both canonical contacts (buried surface on yUIM-1 = 537 Å²), typical of Ub-UIM complexes 42, and non-canonical contacts (buried surface area on yUIM-1 = 297 Å²) not previously observed in Ub-UIM complexes. I previously overlooked the latter contacts as reflecting irrelevant crystal packing interactions. The canonical contacts involving each UIM are essentially as described previously, with the notable exception that the contact surface on the UbV is now composed of sequence elements from both UbV protomers in the strand exchange dimer (Figure 3-3CD). The non-canonical contacts involving each UIM are also composed of sequence elements from both protomers in the UbV dimer. Included are hydrophobic interactions between Leu7 of yUIM-1 and Ile8 of the first UbV protomer and Val10* of the second UbV* protomer, and hydrophilic interactions between Asp3 of yUIM-1 and Gln6 of the first UbV protomer and Tyr68* of the second UbV* protomer (Figure 3-3E).
Figure 3-3. Crystal structure of UbV.v27.1 bound to yUIM-1.

(A) Overview of the dimeric UbV.v27.1-yUIM-1 structure. “*” or “#” denotes the symmetry related UbV or yUIM-1, respectively. (B) Overview of the monomeric UbV.v27.1-yUIM-1 structure. (C, D) Details of the molecular interactions on the canonical interface. An asterisk “*” indicates residues from UbV*. (E) Details of the molecular interactions on the non-canonical interface. (F) Details of the molecular interactions on the non-canonical interface. A hash sign “#” indicates residues from yUIM-1#. 
Interestingly, the 2:2 binding interaction between yUIM-1 and the UbV.v27.1 dimer juxtaposes the two UIMs in an anti-parallel orientation with a buried surface area on each UIM of 276 Å² (Figure 3-3A). Contacts include hydrophobic interactions between Leu residues at positions 7 and 14 of one protomer and the equivalent position of the second protomer (Figure 3-3F) and reciprocal salt bridges between Glu and Lys residues at positions 6 and 10, respectively (Figure 3-3F). I hypothesized that these inter-UIM contacts might also contribute to the enhanced affinity of yUIM-1 binding to UbV.v27.1. In this regard, the salt bridge between Glu6 and Lys10# provided a useful test case as neither residue participates in direct contacts with UbV.v27.1. However, the individual substitutions Glu6Lys or Lys10Glu, which were expected to abrogate the salt interaction and the double substitution Glu6Lys/Lys10Glu, which was expected to restore the salt interaction with opposite polarity, did not appreciably change the affinity of yUIM-1 for UbV.v27.1 (Figure 3-4A). Since all other contacting residues between the two UIMs also participate in direct contacts with the UbV, I could not explore this issue unambiguously further.
Figure 3-4. Mutational and comparative analysis of the yUIM-1/UbV.v27.1 complex.

(A) Effects of substitutions on yUIM-1 binding to UbV.v27.1. Fluorescence polarization binding experiments are shown for the binding of UbV.v.27.1 to yUIM-1 peptides harboring the indicated substitutions. Values represent mean of readings done in triplicate ± SD. (B,C) Comparison of dimeric UbVs. (B) Side by side view of the dimeric UbVs, UbV.v27.1, UbV.XR (PDB:506T), and UbV.15.D (PDB:6DJ9). (C) Superposition of UbV.v27.1, UbV.XR, and UbV.15.D.
3.4 Discussion

I have discovered that UbV.v27.1 forms a strand exchange dimer allowing it to bind yUIM-1 with 2:2 stoichiometry. Oligomerization of UbV.v27.1 in solution is critically dependent on the Gly10Val substitution, as the back substitution to Gly converts UbV.v27.1 to a monomer.

How precisely does Val10 help to promote the dimer state? I surmised Val10 might act by destabilizing the monomer by preventing formation of the tight β-turn between strands β1 and β2. While Val10 in the strand exchange dimer adopts different Ramachandran Phi and Psi values than Gly10 in Ub.wt monomer (77°,17° versus -113°, 123° degrees, respectively, PDB entry 1UBQ), the latter conformation is fully accessible to Val (i.e. it is not a disallowed conformation in the Ramachandran plot), suggesting that this may not be the case. My finding that the Gly10Val single-site substitution in Ub.wt behaves as a monomer in SEC analysis proves that Valine at position 10 can support the tight β-turn promoting conformation required in the monomer state. Thus the exact reason why Val at position 10 promotes dimerization remains an open question.

The finding that the Gly10Val substitution is not sufficient to drive dimerization indicates that other substitutions in UbV.v27.1 relative to the observed positions in Ub.wt help to promote the strand exchange mechanism. Other substituted positions in UbV.v27.1 that may collaborate with the Gly10Val substitution include residues Gln6, Ile8, Met11, Arg12, and Ala14, which lie in close proximity to the strand exchange junction, and Leu76, which participates in UbV.27.1 dimer contacts elsewhere in the structure (see Figure 3-1A for sequence comparison between Ub.wt and UbV.v27.1). Resolving how precisely Val10 helps to promote UbV dimerization may require an understanding of how these substitutions cooperate.

The structure of UbV.v27.1 bound to yUIM-1 allows for a more comprehensive rationalization of my previous mutational studies. The position of Tyr68 on the canonical contact surface between UbV.v27.1 and yUIM-1 readily explains its strong contribution to binding. As described above, it creates favorable interactions with yUIM-1, which directly enhance affinity. The contribution of the substituted Val10, which was difficult to
rationalize previously, is now also readily apparent. First, Val10 participates in direct favourable hydrophobic interactions with yUIM-1 through its position on the non-canonical UIM binding surface. Second, by promoting strand exchange dimerization, Val10 indirectly supports the formation of the non-canonical contact surface with yUIM-1 that affords additional opportunities for favourable interactions. Third, by enabling a 2:2 binding mode with yUIM-1, Val10 also provides opportunities for favorable contacts between the two UIMs. I note that in addition to contributing to the enhanced binding affinity of UbV.v27.1 for yUIM-1, the non-canonical contacts between UbV and yUIM-1, and possibly between the two UIMs, could in principle contribute to the binding specificity of UbV.v27.1, which Bradley Yates a member of the Sidhu lab, showed was highly specific for yUIM-1 over a panel of 11 other yeast UIMs.

The dimerization of UbV.v27.1 through a strand exchange mechanism is reminiscent of the behavior of two other dimeric UbVs that have been reported recently, namely UbV.XR and UbV.15.D, which bind to the dimeric RING E3 ligase XIAP and the DUSP domain of USP15, respectively.\textsuperscript{92,134,135} Similarities between the three UbVs include similar strand exchange topology involving strand β1 that creates an extended β1-β2 strand fusion (Figure 3-4B,C - note that in the case of UbV.15.D, the extended β1-β2 strand fusion is partially interrupted by a helical kink). Moreover, dimerization of UbV.v27.1 and UbV.XR is also dependent on substitutions at position 10. However, whereas dimerization of UbV.v27.1 is dependent on a Val at position 10, UbV.XR is dependent on an Ala (Figure 3-1A) at that position\textsuperscript{92}. UbV.15.D also differs from Ub.wt at position 10 (Figure 3-1A) but whether its Gly10Phe substitution is required for UbV.15.D dimerization remains to be determined. I posit that systematic sampling of the residues at position 10 might prove to be a useful and general means for generating dimeric UbVs with enhanced affinities and specificities for specific ligands.
Chapter 4: A Structure-based Strategy for Engineering Selective Ubiquitin Variant Inhibitors of Skp1-Cul1-F-box Ubiquitin Ligases

Maryna Gorelik\textsuperscript{1,2,y}, Noah Manczyk\textsuperscript{3,4,y}, Alia Pavlenco\textsuperscript{1}, Igor Kurinov\textsuperscript{5}, Sachdev S. Sidhu\textsuperscript{1,2,*}, & Frank Sicheri\textsuperscript{2,3,4,*}

Affiliations
\textsuperscript{1} Banting and Best Department of Medical Research, Terrence Donnelly Center for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada, M5S 3E1
\textsuperscript{2} Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada, M5S 3E1
\textsuperscript{3} Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, M5G 1X5
\textsuperscript{4} Department of Biochemistry, University of Toronto, Toronto, ON, Canada, M5S 3E1
\textsuperscript{5} Department of Chemistry and Chemical Biology, Cornell University, Argonne, IL, USA, 60439

Author list footnotes
Y - authors contributed equally to this work
* - corresponding author

A version of this chapter is published in Structure 2018 September; 4(26)1226-1236.

Attributions: Dr. Sidhu and his lab conceived the original project idea detailed in this chapter. This project started with Dr. Maryna Gorelik engineering, by phage display, UbV binders for 21 different Skp1-F-box complexes and assaying their binding specificity by phage ELISA (Figure 4-1 and 4-3). I then became involved and was to understand the basis for how select UbVs bound to their respective Skp1-F-box complexes and how exemplar UbV’s achieved high substrate specificity. I determined the structure of UbV.Fl11.1 bound to Skp1-Fbl11, UbV.Fl10.1 bound to Skp1-Fbl10, and UbV.Fl10.1 bound to Skp1-Fbl11 by X-
ray crystallography. I then analyzed the structures to determine the binding mode of the UbVs and to generate hypotheses for binding specificity (Figure 4-5 and 4-9). Maryna then tested my binding specificity hypotheses by mutagenesis and protein ELISAs (Table 4-4). Finally, Maryna tested two exemplar UbVs for their ability to inhibit their target Skp1-F-box complexes in vivo (Figure 4-11). Throughout Chapter 4, “I” is used where work was performed solely by me, “we” is used where I played a participatory role, and the respective author(s) name such as “Maryna” where experiments were performed exclusively by others.

4.1 Introduction

Ubiquitin (Ub), a highly conserved 76-residue protein, is attached through the E1-E2-E3 enzymatic cascade in a variety of topologies to mark proteins for degradation or to alter their activity. Not surprisingly, the E3 ligases that control specificity of ubiquitination constitute the largest group with ~600 members in humans. Due to their large number and importance in controlling cellular fate, E3 ligases are considered important targets for therapeutic intervention.

Multisubunit Cullin RING Ligases (CRLs) constitute the largest family of E3 ligases with ~250 members and are characterized by the presence of a Cullin subunit responsible for tethering a substrate receptor and a RING protein that in turn recruits E2-Ub. Seven distinct Cullin subunits (Cul1, Cul2, Cul3, Cul4A, Cul4B, Cul5 and Cul7) define seven CRL subfamilies, which also differ in terms of the type of substrate receptor (Skp1-F-box, EloBC-VHL, BTB, DDB1-DCAF or EloBC-SOCS-box) and RING protein (Rbx1 and Rbx2) that are tethered. Skp1-Cul1-F-box (SCF) E3 ligases represent the best characterized CRL family, and they have attracted great attention because of their established and inferred roles in regulating diverse biological processes. In humans, 69 F-box proteins act as substrate receptors for SCF E3 ligases and are bound to the Cul1 subunit through the adaptor Skp1. Binding to Skp1 is mediated by a small F-box domain of 50-residues that is similar in all F-box proteins. F-box family members are divided into three subfamilies denoted Fbw, Fbl and Fbo, based on the presence of an additional WD40, LRR or “other” domain, respectively, which functions to recruit substrates.
Small molecule inhibitors of individual SCF E3 ligases have been developed to disrupt substrate binding \cite{139-142} or the interaction between F-box and Skp1 \cite{143,144}. However, effective inhibitors are lacking for the vast majority of F-box proteins, and thus, the biological effects and therapeutic potential of inhibiting their E3 ligase functions remain unexplored.

Protein-based modulators of enzymes can aid in the development of small molecule therapeutics by assisting in target validation, by serving as probes in displacement screens, and by serving as leads for structure-based inhibitor design. The Ub scaffold is particularly amenable for protein engineering of variants that function as either inhibitors or activators. Ub variants (UbVs) that bind to a variety of target proteins have been generated through several different approaches \cite{145-148}. Research by Dr. Sidhu’s group using a phage display method to generate UbVs has been successful in targeting diverse components of the Ub proteasome system (UPS) including deubiquitinases \cite{89}, UIMs \cite{133}, E2 enzymes \cite{89}, HECT E3 ligases \cite{90}, RING and U-box E3 ligases \cite{92}, the Anaphase Promoting Complex/Cyclosome APC/C complex \cite{93}, and SCF E3 ligases \cite{91}.

Since many UPS enzymes use weak Ub-binding sites for their normal biological functions, it is not surprising that most UbVs are generated by the strengthening of these natural interactions. However, in the case of SCF E3 ligases, Maryna Gorelik of the Sidhu lab and Stephen Orlicky of the Sicheri lab identified a UbV that bound to a composite surface on the Skp1-Fbw7 subcomplex not previously known to interact with Ub. The UbV-binding surface overlapped extensively with the binding surface on Skp1 for Cul1. Consequently, the UbV functioned as an inhibitor of SCF\textsuperscript{Fbw7} activity by disrupting Cul1 binding. This finding revealed a previously uncharacterized inhibitory site in SCF E3 ligases \cite{91}.

One attractive feature of targeting the Cul1 interacting surface of the Skp1-F-box subcomplex is that it affords a potentially general means to inhibit an SCF complex without any knowledge of substrates and substrate binding mechanisms of the F-box subunit, which remains poorly understood for most F-box proteins. However, the issue of specificity remains a potential liability to this strategy because the Cul1 binding surface is largely contained on the common Skp1 subunit. Indeed, in generating UbVs that bound to Skp1-Fbw7, Maryna Gorelik observed that the model UbV displayed cross-reactivity to several
other Skp1-F-box complexes. By analyzing the co-structure of a UbV bound to Skp1-Fbw7, Stephen Orlicky confirmed that cross-reactivity stemmed from the UbV forming the majority of its contacts with the common Skp1 component while forming only a minority of contacts, notably involving the β1-β2 loop of the UbV, with the variable F-box component (Fbw7). Thus Maryna hypothesized that specificity could be improved by increasing the contacts between the F-box and the β1-β2 loop. To this end, Maryna increased the length of the β1-β2 loop and subjected it to random diversification. In doing so, Maryna isolated a UbV with an extended β1-β2 loop that bound with high specificity to Fbw11, discriminating even against the highly related protein Fbw1.

Here Maryna sought to determine if the β1-β2 loop diversification strategy is generalizable for generating specific UbV binders for a larger set of Skp1-F-box complexes, including members of Fbw, Fbl and Fbo sub-families. Maryna generated UbV binders targeting 21 different Skp1-F-box complexes, and amongst these, showed that UbVs targeting 17 complexes were highly specific. Furthermore, I solved structures of a representative high specificity UbV generated against Skp1-Fbl11 and a less specific UbV generated against Skp1-Fbl10 in complex with their cognate targets. These structures describe the binding mechanism employed by UbVs with extended β1-β2 loops and provide insight into the specificity determinants of UbV-F-box interactions. Together, our findings demonstrate that a F-box targeted UbV library can be exploited to generate specific binders against F-box proteins on a large scale and suggest possible modifications to the library design for further improvements.

4.2 Experimental Procedures

4.2.1 Protein expression and purification

The Skp1-F-box, UbV, and Cul1 proteins employed in phage binding selections and/or ELISAs were expressed and purified as described. Domain boundaries of F-box constructs are listed in Table 4-1. The Skp1-F-box-UbV complexes used in crystallization were prepared as described previously for the Skp1-Fbw7-UbV.Fw7.1 complex.
4.2.2 Library construction and binding selections

The UbV-phage library was constructed as described previously for Library 3, except that 10-12 residues were inserted in the β1-β2 loop. The diversity of the constructed library was \(1.8 \times 10^{10}\). For binding selections, GST-tagged Skp1-F-box complexes were coated on 96-well MaxiSorp plates (Thermoscientific 12565135) by adding 100 μL of 50 nM protein solution and incubating overnight at 4 °C. Four rounds of binding selections with phage library pools were performed with immobilized proteins as described. To eliminate UbV-phage that bound nonspecifically, input phage pools were pre-incubated on plates coated with non-target Skp1-F-box proteins (Table 4-1) and the input phage pools were also mixed with 1 μM of the same non-target Skp1-F-box complex during the positive selection step. Several targets (Fbo24, Fbo28, Fbl6), which failed to generate specific UbV-phage in the initial binding selections, were subjected to a second binding selection using GST as the negative selection target.

For each Skp1-F-box target, 24-192 UbV-phage clones were screened by phage ELISA for binding to target Skp1-F-box, a non-target Skp1-F-box and with competition from Cul1. UbV-phage that exhibited >2-fold higher binding signal to target versus GST or Skp1-Fbw7, and at least 20% reduction in binding in the presence of Cul1, were further characterized. Up to 24 positive clones were sequenced from each selection (Table 4-2).

4.2.3 ELISAs

For phage ELISAs, target Skp1-F-box complexes were immobilized on 384-well MaxiSorp plates (Thermoscientific 12665347) by adding 30 μL of 50 nM protein solutions and incubating overnight at 4 °C. Phage ELISAs were performed as described, except that three washes were performed for all wash steps and volumes were scaled down from 100 μL to 30 μL to accommodate the 384-well format. Binding of phage was detected using anti-M13-HRP antibody at 1:5000 dilution. For testing whether the binding of UbV-phage was inhibited by Cul1 protein, either 20 μL of PBS or 20 μL of 200 nM Cul1 were added for 10 min prior to addition of 10 μL UbV-phage. Reduction of phage-displayed UbV binding by >20 % was considered to signify competition with Cul1 for binding. For protein ELISAs, Skp1-F-box complexes were immobilized by adding 30 μL of 200 nM protein solutions to
384-well MaxiSorp plates (Thermoscientific 12665347) and incubating overnight at 4 °C. UbV.Fl11.1 or UbV.Fl101.1 at sub-saturating concentration were mixed with the serial dilutions of the tested Skp1-F-box complexes and applied to the plate containing immobilized Skp1-Fbl11 or Skp1-Fbl10 complexes respectively. Binding of FLAG-tagged UbV was detected using anti-FLAG-HRP antibody at 1:5000 dilution. Competitive ELISA IC50 values were derived from the binding curves and corresponded to the concentration of the tested Skp1-F-box complexes at which 50% of the UbV binding was observed. IC50 values were calculated by fitting the obtained binding curves to four parameter logistic nonlinear regression model using GraphPad Prism software. Data represent mean ± SEM of four binding curves.

4.2.4 Cell-based assays

On day 0, 6-well plates were seeded with 4 x 10^5 HEK293T cells. On day 1, cells were transfected with 2 µg of plasmid DNA to express FLAG-UbV using the X-tremeGENE transfection reagent according to manufacturer's protocol. For Snai1 stability assays, the cells were simultaneously transfected with 0.1 µg of plasmid DNA to express HA-tagged Snai1. On day 3, cycloheximide (100 µg/ml for p27 stability assays or 20 µg/ml for Snai1 stability assays) was added for 0-6 hours. Cells were lysed in lysis buffer (Cell Signalling 9803) and cell lysates were subjected to western blot analysis.

4.2.5 F-box alignment

The alignment of F-box domains was performed using Cn3D. An initial structure based model of the F-box domain was performed using the F-box structures from Fbl3 (PDB:4IGJ), Fbw1 (PDB:1P22), Skp2 (PDB:2ASS), Fbw7 (PDB:2OVP), Cdc4 (PDB:3MKS), and Fbl10 (PDB:5JH5). All other human F-box sequences were then imported into Cn3D and aligned using the “block align single” function. Further refinement of the alignment was done manually. Conservation was calculated using AL2C0 1.0 in Chimera.
4.2.6 Crystallization and structure determination

The UbV.Fbl11.1-Skp1-Fbl11 complex was crystallized in sitting drops at 20 °C by mixing 0.2 µL protein complex (17 mg/mL) with 0.2 µL mother liquor (0.1 M PCTP buffer pH 6 and 25% w/v PEG 1500). Crystals were transferred to a cryoprotectant solution (mother liquor with 20% ethylene Glycol) and flash frozen in liquid nitrogen. A single crystal dataset was collected at -180 °C on a home-source consisting of a Rigaku MicroMax-007 HF rotating anode generator coupled to a Rigaku Saturn 944 HG CCD detector. Data was processed by HKL-2000 153. The structure was solved by molecular replacement using Phaser125 and search models of Ub with the 5 C-terminal residues removed (PDB:1UBQ), Skp1 (PDB:5IBK) and Fbl10 residues 1066-1104 (PDB:5JH5). The structure was refined by PHENIX 127 using TLS parameters 128 and manual building in Coot 129.

The UbV.Fbl10.1-Skp1-Fbl10 complex was crystallized in hanging drops at 20 °C by mixing 1 µL protein complex (19 mg/mL) with 1 µL mother liquor (0.1 M Hepes buffer pH 7.5, 0.2 M CaCl2 and 18% w/v PEG 6000). Crystals were transferred to a cryoprotectant solution (mother liquor with 20% ethylene glycol) and flash frozen in liquid nitrogen. A single crystal dataset was collected at -180 °C on a PILATUS 6M-F detector at station 24-ID-C, NE CAT beamline, Advanced Photon Source (APS) and processed using HKL-2000153. The structure was solved by molecular replacement using Phaser 125 and search models of ubiquitin with the 5 C-terminal residues removed (PDB:1UBQ) and Skp1-Fbl11 from the UbV.Fbl11.1-Skp1-Fbl11 structure. The structure was refined by PHENIX 127 using TLS parameters 128 and manual building in Coot 129.

The UbV.Fbl10.1-Skp1-Fbl11 complex was crystallized in hanging drops at 20 °C by mixing 1 µL protein complex (18 mg/mL) with 1 µL mother liquor (0.1 M Malic Acid pH 4.5, 0.15 M NaCl, 27% w/v PEG 3350). Crystals were transferred to a cryoprotectant solution (mother liquor with 20% ethylene glycol) and flash frozen in liquid nitrogen. A single crystal dataset was collected at -180 °C on an EIGER 16M detector at station 24-ID-C, NE CAT beamline, Advanced Photon Source (APS) and processed using XDS 154. The structure was solved by molecular replacement using Phaser 125 and search models of Ub with the 5 C-terminal residues removed (PDB:1UBQ) and Skp1-Fbl10 from the UbV.Fbl10.1-Skp1-Fbl10 structure. The structure was refined by PHENIX 127 and manual building in Coot 129.
Interactions for structures were analyzed manually using The PyMOL Molecular Graphics System (Schrodinger, LLC) and using the protein interfaces, surfaces and assemblies (PISA) tool\textsuperscript{130}. Alignments were performed using Coot. Rotation and translation of structures were analyzed by the Pymol script ‘RotationAxis’ available on the Pymol wiki webpage. Structure representations were performed using the PyMOL molecular graphics system.

4.2.7 Isothermal Calorimetry

Calorimetric titrations were performed on a Malvern MicroCal Auto-iTC200 (SBC Facility at The Hospital for Sick Children) at 25 °C. Protein samples were dialyzed in 300 mM NaCl, 50 mM Hepes pH 7.5, 5% glycerol, 4 mM β-mercaptoethanol. To measure UbV.L11.1 binding to Skp1-Fbl11, 150 µM UbV.L11.1 in the syringe was titrated into 20 µM Skp1-Fbl11 complex in the lower cell. To measure binding between UbV.L10.1 and Skp1-Fbl10 or Skp1-Fbl11, 167 µM or 150 µM UbV.L10.1 in the syringe was titrated into 20 µM Skp1-Fbl10 or Skp1-Fbl11 in the lower cell. Sixteen 2.49 µL injections were performed with an interval of 120 seconds. Experiments were carried out in triplicate. Analysis was performed by nonlinear curve-fitting of the corrected data to a model with one site using ORIGIN software. Reported values represent the mean ± SD.

4.3 Results

4.3.1 Binding selections with an F-box targeted phage-displayed UbV library

Towards developing improved F-box specific UbVs, Maryna designed a phage-display library based on one that was employed in the previous study. In particular, Maryna left the predicted Skp1 contacting residues in the UbV template fixed to those observed in a previously characterized Skp1-Fbw7 binder, UbV.Fw7.5, and varied the size and sequence of the β1-β2 loop. Maryna’s strategy for varying the β1-β2 loop differed slightly from the previous UbV library\textsuperscript{91} by inserting 6-8 residues (rather than 7-9 residues) and by randomizing 10-12 residues (rather 11-13 residues) (Figure 4-1A). Maryna found that by slightly decreasing the number of inserted residues in the β1-β2 loop, the fraction of UbVs in the naïve library that were displayed at high levels on phage was increased (Figure 4-2).
Maryna chose 32 human F-box family members as targets for UbV selections, including some with well-characterized biological functions and connections to disease (e.g. Fbl1 (Skp2), Fbw1 (β-TrcpA), Fbo5 (Emi1)) and others with completely uncharacterized functions. Maryna expressed each isolated F-box domain (see Table 4-1 for domain boundaries) in complex with Skp1. Of 32 F-box domains, 23 could be purified in complex with Skp1 for use in 4 rounds of binding selections with the F-box targeted phage-displayed UbV library (Figure 4-1B). Following selections, the binding of individual UbV-phage clones was tested by phage enzyme-linked immunosorbent assays (ELISAs). Only those UbVs which met the following criteria were further characterized: 1) Confirmed as binders with greater than 2 fold higher signal towards the target Skp1-F-box relative to GST; 2) Showed some measure of specificity with greater than 2-fold higher signal towards the target Skp1-F-box relative to the reference standard Skp1-Fbw7; 3) Binding inhibited by Cul1 with greater than 20% reduction in binding signal in the presence of Cul1. Maryna expected that these criteria would narrow the population of UbV binders to those that conformed to the design strategy and specifically targeted the Cul1 binding surface of a particular Skp1-F-box complex. Two Skp1-F-box complexes (Fbl13 and Fbl18) did not meet the first selection criterion and four complexes (Fbl6, Fbo24, Fbo28 and Fbl6) did not meet the second selection criterion, and thus, 6 of the 23 purified complexes failed to generate UbVs selective for the target Skp1-F-box (Figure 4-1B).
Figure 4-1. Selection of phage-displayed UbVs targeting a panel of human F-box proteins.

(A) F-box targeted UbV library design. The sequence of Ub is shown and UbV positions conserved as wt are indicated by dashes. Positions that were fixed in UbV.Fw7.1 or the library but differ from wt are shown. Each position that was diversified in the library is indicated by an “X”, which represents all 20 amino acids encoded by a degenerate NNK codon (N = A/G/C/T, K = G/T). Three sub-libraries were made with insertions of 6, 7 or 8 codons between positions 11 and 12, and these were combined to generate the final library used for selections. (B) Summary of Skp1-F-box protein purification and UbV selection results. The targeted F-box
proteins are shown as follows: phage selections generated UbVs against the F-box target, "Selection successful"; generated UbVs demonstrated preference for the target Skp1-F-box complex (>2-fold higher signal to target Skp1-F-box versus Skp1-Fbw7 in phage ELISA), "Target selective"; generated UbVs showed no preference for the target Skp1-F-box complex (<2-fold higher signal to target Skp1-F-box versus Skp1-Fbw7 in phage ELISA), "Target non-selective"; phage selections failed to produce UbVs against the F-box target, "Selection failed"; Skp1-F-box complexes could be purified at sufficient quantities for the use in phage selections and ELISAs, "Purification successful"; Skp1-F-box complexes could not be purified, "Purification failed". (C) Number of unique binding UbVs generated for each Skp1-F-box complex. Black bars indicated targets for which UbVs exhibited a clear consensus sequence in the diversified β1-β2 loop. (D) Sequence logos for the diversified residues in the β1-β2 loop of unique UbVs selected for binding to the indicated Skp1-F-box complexes.
Figure 4-2. Influence of the β1-β2 loop insertion length on the fraction of phage-displayed UbV in the naïve library.

Binding of phage displaying FLAG-tagged UbV with anti-FLAG antibody was assessed by phage ELISA. For each insert length, 94 individual phage clones from naïve sub-libraries were tested and the fractions exhibiting strong binding signals (OD$_{450} > 0.5$) are depicted.
Table 4-1. Summary of the F-box expression constructs, purification, and phage selection results.

<table>
<thead>
<tr>
<th>F-box</th>
<th>Expression construct(^a)</th>
<th>Purification(^b)</th>
<th>Negative Selection(^c)</th>
<th>Phage Selection(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fbw1</td>
<td>175-288</td>
<td>√ [18]</td>
<td>Fbw7</td>
<td>√</td>
</tr>
<tr>
<td>Fbw11</td>
<td>114-225</td>
<td>√ [18]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbw7</td>
<td>183-287</td>
<td>√ [18]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbw5</td>
<td>1-79</td>
<td>√ [18]</td>
<td></td>
<td>Not tested</td>
</tr>
<tr>
<td>Fbw12</td>
<td>1-49</td>
<td>√ [18]</td>
<td></td>
<td>Not tested</td>
</tr>
<tr>
<td>Fbo1</td>
<td>95-139</td>
<td>√ [18]</td>
<td>Fbw7</td>
<td>√</td>
</tr>
<tr>
<td>Fbo6</td>
<td>110-159</td>
<td>√</td>
<td>Fbl7, GST</td>
<td>X, X</td>
</tr>
<tr>
<td>Fbo7</td>
<td>112-156</td>
<td>√</td>
<td>Fbl6</td>
<td>√</td>
</tr>
<tr>
<td>Fbo10</td>
<td>1060-1104</td>
<td>√</td>
<td>Fbw7</td>
<td>√</td>
</tr>
<tr>
<td>Fbo11</td>
<td>887-931</td>
<td>√</td>
<td>Fbw7</td>
<td>√</td>
</tr>
<tr>
<td>Fbo13</td>
<td>152-197</td>
<td>√</td>
<td>Fbo24</td>
<td>X</td>
</tr>
<tr>
<td>Fbo15</td>
<td>22-64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbo16</td>
<td>87-136</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbo18</td>
<td>25-71</td>
<td>√</td>
<td>Fbo36</td>
<td>X</td>
</tr>
<tr>
<td>Fbo20</td>
<td>22-67</td>
<td>√</td>
<td>Fbo10</td>
<td>√</td>
</tr>
<tr>
<td>Fbo3</td>
<td>10-55</td>
<td>√</td>
<td>Fbw7</td>
<td>X</td>
</tr>
<tr>
<td>Fbo4</td>
<td>57-100</td>
<td>√</td>
<td>Fbw7</td>
<td>√</td>
</tr>
<tr>
<td>Fbo5</td>
<td>242-293</td>
<td>√</td>
<td>Fbw7</td>
<td>√</td>
</tr>
<tr>
<td>Fbo6</td>
<td>11-56</td>
<td>√</td>
<td>Fbo4</td>
<td>√</td>
</tr>
<tr>
<td>Fbo7</td>
<td>328-373</td>
<td>√</td>
<td>Fbw7</td>
<td>√</td>
</tr>
<tr>
<td>Fbo8</td>
<td>67-110</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbo9</td>
<td>168-217</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbo10</td>
<td>1-47</td>
<td>√</td>
<td>Fbl20</td>
<td>√</td>
</tr>
<tr>
<td>Fbo11</td>
<td>156-201</td>
<td>√</td>
<td>Fbo34</td>
<td>√</td>
</tr>
<tr>
<td>Fbo17</td>
<td>15-61</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbo22</td>
<td>18-63</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbo24</td>
<td>36-81</td>
<td>√</td>
<td>Fbo28, GST</td>
<td>X, X</td>
</tr>
<tr>
<td>Fbo28</td>
<td>62-105</td>
<td>√</td>
<td>Fbo24, GST</td>
<td>X, X</td>
</tr>
<tr>
<td>Fbo34</td>
<td>573-617</td>
<td>√</td>
<td>Fbo11</td>
<td>√</td>
</tr>
<tr>
<td>Fbo36</td>
<td>92-136</td>
<td>√</td>
<td>Fbl18</td>
<td>√</td>
</tr>
<tr>
<td>Fbo38</td>
<td>28-72</td>
<td>√</td>
<td>Fbw7</td>
<td>√</td>
</tr>
<tr>
<td>Fbo39</td>
<td>14-57</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbo43</td>
<td>493-545</td>
<td>√</td>
<td>Fbw7</td>
<td>√</td>
</tr>
<tr>
<td>Fbo44</td>
<td>4-49</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbo45</td>
<td>34-81</td>
<td>√</td>
<td>Fbo4</td>
<td>√</td>
</tr>
<tr>
<td>Fbo46</td>
<td>297-341</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) All F-box domains were expressed with an N-terminal His tag, in complex with GST-tagged Skp1 as previously described.\(^b\)

\(^b\) Skp1-F-box complexes were purified as previously described, with the presence and absence of expected protein product indicated by “√” and “X” respectively.

\(^c\) Skp1-F-box proteins used in negative selection are indicated.

\(^d\) Phage selections which were successful and which failed to generate UbV binders with specific preference for the target are indicated by “√” and “X” respectively. Phage selections against Skp1-F-box\(^b\)bw7 and Skp1-F-box\(^b\)bw11 complexes were carried out in the previous study.
Using the above criteria Maryna succeeded in generating UbV binders with at least some evidence of specificity for 17 of the 23 purified Skp1-F-box complexes (Figure 4-1B, Table 4-1), representing a success rate of over 70%. The number of unique UbV binders obtained for each Skp1-F-box complex varied considerably (Figure 4-1C, Table 4-2). Fbl7, Fbl10, Fbo4, Fbo11, Fbo34, Fbo43, and Fbo45 yielded 6-18 unique UbVs each, which allowed Maryna to define consensus motifs for the diversified positions (Figure 4-1D). Other selections yielded either fewer unique UbVs (<5) or highly variable UbV sequences, which precluded determination of consensus motifs.
Table 4-2. UbV generated in phage selections.

<table>
<thead>
<tr>
<th>F-box</th>
<th>Ubv&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Ubv sequence (Region1)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Number of clones&lt;sup&gt;c&lt;/sup&gt;</th>
<th>F-box</th>
<th>Ubv</th>
<th>Ubv sequence (Region1)</th>
<th>Number of clones&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fbw1</td>
<td>Fw1.1</td>
<td>GNYFKRRKSG</td>
<td>(16/22)</td>
<td>Fbl7</td>
<td>F11.1</td>
<td>AFPVLGVPLE</td>
<td>(2/14)</td>
</tr>
<tr>
<td></td>
<td>Fw1.2</td>
<td>PYVVCDAIGCF</td>
<td></td>
<td></td>
<td></td>
<td>SFPGMGVTLETL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HKWAYWDYGD</td>
<td></td>
<td></td>
<td></td>
<td>IFTSSPMWRSDD</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>YDIYDIEVEI</td>
<td></td>
<td></td>
<td></td>
<td>LLPVLGIPFS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WVEYRWGKPR</td>
<td>(2/22)</td>
<td></td>
<td></td>
<td>PCLRVCLGRRLM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WHDGYFRRVG</td>
<td></td>
<td></td>
<td></td>
<td>ILNPYRFILSG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GQARPRWQGVG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SIRPVRLSVE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LLRSAVGLGN</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LLWSSVMRTGGN</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>YFTLNHFKSVPG</td>
<td></td>
</tr>
<tr>
<td>Fbl7</td>
<td>F17.1</td>
<td>LLLSDRPSTGH</td>
<td>(2/21)</td>
<td>Fbl10</td>
<td>F110.3</td>
<td>LDRLGLIPKT</td>
<td>(2/23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LDLRSRPLGE</td>
<td></td>
<td></td>
<td></td>
<td>FRDRLRNLPAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LQIDMPAFGE</td>
<td></td>
<td></td>
<td></td>
<td>FRDRLRNRPAT</td>
<td>(7/23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LDLYKSPLRGE</td>
<td></td>
<td></td>
<td></td>
<td>FRGRLRNRPAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MDLNNSPRFYRT</td>
<td>(2/21)</td>
<td></td>
<td></td>
<td>FRDRLRNLLPQ</td>
<td>(8/23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LSSSRPTGT</td>
<td></td>
<td></td>
<td></td>
<td>WEQRSSWGGRP</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LMRASAPSSGH</td>
<td></td>
<td></td>
<td></td>
<td>RTSWHTGFQIQNI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LDLKNSRPLGE</td>
<td></td>
<td></td>
<td></td>
<td>PFIHRAVGYVYS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LDLKSKPLQGS</td>
<td>(2/21)</td>
<td></td>
<td></td>
<td>YIFAISLH.DFK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MNSAPFPQAKS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LLDDLPLPTFTDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LSLERPAVGN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CIVELWACRRY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WGWIREARDG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MNTAPFPAAPQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>VQKGWRAPPEW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LHRTLSPLKGV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LHLHNSPGNGVS</td>
<td>(219/20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbl11</td>
<td>F111.1</td>
<td>RHSYKHIHGIENS</td>
<td>(19/20)</td>
<td>Fbl20</td>
<td>F120.1</td>
<td>AARTWLPAGG</td>
<td>(16/22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RLNRYPYIordinator</td>
<td></td>
<td></td>
<td></td>
<td>ILSPRMSWPAGS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FTRTRFGYVNDT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WGAWRRPEWW</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSRTWSLPVGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HVRSKYGFHVKA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LGPWVRSYFQ</td>
<td></td>
</tr>
<tr>
<td>Fbo4</td>
<td>F04.2</td>
<td>RAASWTVGDWR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EHKYRHILSG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PDRYRRLLMSP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PPHYRRLLMPD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHKYRRLMKN</td>
<td>(5/24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EHKYRHRLLKN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DRRYRQLVSG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHYRRHLKN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PDRYRHLST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MKKYRHLING</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHRYRLLLSG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PPOYRLLMQP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EHKYRHRLLPN</td>
<td>(2/24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DRRYRLLVSG</td>
<td>(2/24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PQKYRYLLHE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLRYRLLINP</td>
<td>(3/24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbo5</td>
<td>F05.1</td>
<td>LGWDKGKGVRG</td>
<td>(7/13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LGWDTSNGVRK</td>
<td>(4/13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LGWDVGPVSR</td>
<td>(2/13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-box</td>
<td>Ubv</td>
<td>Ubv sequence (Region1)</td>
<td>Number of clones&lt;sup&gt;a&lt;/sup&gt;</td>
<td>F-box</td>
<td>Ubv</td>
<td>Ubv sequence (Region1)</td>
<td>Number of clones&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>------------------------</td>
<td>----------------------------</td>
<td>-------</td>
<td>------</td>
<td>------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Fbo43</td>
<td>Fo43.1</td>
<td>LGWDKGGVRG LGWDVGGWSR LGWDITGGRHR AGFDRGGVAE LGYDTGKVRG LGWDITGKVRG HGFDKAGVRG</td>
<td>(14/23)</td>
<td>Fbo6</td>
<td>Fo6.1</td>
<td>RGRLPSWVEQV CWSMKCKFSSM</td>
<td>(17/24)</td>
</tr>
<tr>
<td></td>
<td>Fo7.1</td>
<td>SWEGLGLGPLRW VRYGTLGQGRAV YFLLFNTLPL FRGLSLATFPWSW YIYLSIFIWMPK GMYMENIYLPLF</td>
<td>(6/21)</td>
<td>Fbo10</td>
<td>Fo10.1</td>
<td>LGYWERRLLP</td>
<td>(24/24)</td>
</tr>
<tr>
<td></td>
<td>Fo7.2</td>
<td></td>
<td>(2/21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fo11.2</td>
<td>QLKYRHLRP PERYPRLMTLP PMYRLLLPP DPRYLLMSP PRKRYYLLSS PSRYRLLMPD GGSRQVAIHS PSRYRHLME ASKRYRHISQ PRKYYYYLLPP DAKYRHILMP SGWSWRLRGG APKYRLLRNNE AARYRHILITA</td>
<td>(6/23)</td>
<td>Fbo34</td>
<td>Fo34.1</td>
<td>SLARPEGRGTR PWARPEGVGR T PWARPEGWGT PLATAEGPGTR PWPAERRGAR PYAKPEASGVR</td>
<td>(4/17)</td>
</tr>
<tr>
<td></td>
<td>Fo11.1</td>
<td>QLKYRHLRP PERYPRLMTLP PMYRLLLPP DPRYLLMSP PRKRYYLLSS PSRYRLLMPD GGSRQVAIHS PSRYRHLME ASKRYRHISQ PRKYYYYLLPP DAKYRHILMP SGWSWRLRGG APKYRLLRNNE AARYRHILITA</td>
<td>(3/23)</td>
<td>Fbo38</td>
<td>Fo38.1</td>
<td>HSTLARLLALPS</td>
<td>(7/7)</td>
</tr>
<tr>
<td></td>
<td>Fo36.1</td>
<td>LFVHSAGSKYHI FCLNVCLVPAR PNLATLSFSTR LNYFCWGDDCL CGAWWLCAPGQ LSFSYFNPRP LCYIYGFCAPR</td>
<td>(2/11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fo36.2</td>
<td>LFVHSAGSKYHI FCLNVCLVPAR PNLATLSFSTR LNYFCWGDDCL CGAWWLCAPGQ LSFSYFNPRP LCYIYGFCAPR</td>
<td>(2/11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fo45.1</td>
<td>WVVMMGRAREE PSRCVGTFCGDL WGWWGRQAREE SIKCGVFCA PG PWVCVGFCPG AWNCVGVFRCPR WGWWGRLVCE SYRCVGLYCATL PWACVWHFCAPG PWWWVGHFCCPG SINCIGVFCA PG SIGCMGVVCPG</td>
<td>(2/23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fo45.2</td>
<td>WVVMMGRAREE PSRCVGTFCGDL WGWWGRQAREE SIKCGVFCA PG PWVCVGFCPG AWNCVGVFRCPR WGWWGRLVCE SYRCVGLYCATL PWACVWHFCAPG PWWWVGHFCCPG SINCIGVFCA PG SIGCMGVVCPG</td>
<td>(2/23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>The name is indicated for the UbVs selected for further characterization (Figure 4-3).

<sup>b</sup>The variable region (Region 1) is shown for the Ubvs generated against the targeted Skp1-F-box complex. Sequences are arranged from the strongest Ubv binder according to phage ELISA (top) to the weakest Ubv binder (bottom).

<sup>c</sup>The number of times a UbV sequence was observed out of the total number of sequences (if more than 1).
Comparison of the unique UbV sequences for each Skp1-F-box complex revealed that only two UbVs (out of 128 total) were selected by more than one (in this case two in total) F-box proteins. In one case, UbV.Fo5/43.1 was selected as one of 3 or one of 7 unique UbVs by Fbo5 (Emi1) or Fbo43 (Emi2), respectively (Table 4-2). While these two F-boxes exhibit 44% sequence identity (Figure 4-3A) and thus represent two of the more similar proteins in our study, two even more similar F-boxes, namely Fbl10 (Kdm2b) and Fbl11 (Kdm2a) (67% identity; Figure 4-3A), selected distinct UbV sequences. Specifically, Fbl10 selected multiple UbVs that shared a motif (Figure 4-1D), whereas Fbl11 strongly selected (95% of all clones sequenced) a distinct UbV (Table 4-2). This result suggested that sequence diversity across the F-box family may be sufficient to generate distinct binding preferences, an issue explored further below. In the second case, two of the more dissimilar F-boxes, Fbo4 and Fbo11 (36% identity; Figure 4-3A), both selected UbV.Fo11.1 (Table 4-2) and families of UbVs with similar consensus motifs (Figure 4-1D).
Figure 4-3. Binding specificities of phage-displayed UbVs.

(A) Sequence alignment of the 69 human F-box domains. The F-box domains were aligned based on the available F-box structures and conservation calculated at each position (see experimental procedures). Residues with >75% conservation and >50% conservation are shaded dark grey or light grey, respectively. Only 24 F-box domains included in the UbV binding specificity analysis are shown. Asterisks (*) indicate F-box domain positions involved in UbV interactions based on the structures of UbVs in complex with Skp1-Fbl11 or Skp1-Fbl10. (B) Phage ELISA binding specificities of the 17 most specific UbVs targeting 17 different Skp1-F-box complexes. For the Skp1-F-box targets with more than one characterized UbV, the UbV with the highest binding specificity is shown.
highest specificity is shown. See **Figure 4-4** for the specificity profiles of all 29 characterized UbVs. Binding of phage-displayed UbV to immobilized Skp1-F-box complexes was detected spectrophotometrically and the absorbance at 450 nM is shown normalized to the signal for cognate Skp1-F-box complex. Grey boxes indicate interactions that were not assayed. The UbV.Fw7.5 and UbV.Fw11.2, targeting Fbw7 or Fbw11 respectively, were described previously\(^9\). UbVs classified as specific (normalized binding to all non-cognate Skp1-F-box complexes <0.1) are shaded green and the remaining UbVs with different degrees of cross-reactive behavior are shaded red.
4.3.2 Characterization of UbV binding specificity

To investigate specificity in greater detail, Maryna assessed binding of the selected UbVs for the various Skp1-F-box complexes by phage ELISAs. For each Skp1-F-box complex, Maryna chose up to three UbVs for characterization based on prioritization of binding strength, clone abundance in the selected pool, and sequence diversity. Maryna measured the binding signals of 29 unique UbV-phage across a panel of 19-24 immobilized Skp1-F-box complexes. Figure 4-3B shows the most specific UbV for each of the Skp1-F-box complexes, while Figure 4-4 shows specificity data for all 29 UbVs that were tested. Providing corroboration that the selection process does impart significant specificity, 27 of 29 UbVs bound to their cognate target with the highest binding signal (Figure 4-3B, Figure 4-4). The two exceptions were UbVs selected for binding to Fbw1, which exhibited higher binding signals to the closely related Fbw11 (89% identity; Figure 4-3B, Figure 4-4).
**Figure 4-4. Phage ELISA binding specificities of 29 UbVs targeting 17 different Skp1-F-box complexes.**

Binding of UbV-phage to immobilized Skp1-F-box complexes was detected spectrophotometrically and absorbance at 450 nM is shown normalized to the cognate Skp1-F-box complex. Grey boxes indicate interactions that were not assayed. The UbV.Fw7.5 and UbV.Fw11.2, targeting Fbw7 or Fbw11 respectively, were described previously. UbVs classified as specific (normalized binding to all non-cognate Skp1-F-box complexes <0.1) are shaded green and other UbVs with different degrees of cross-reactive behavior are shaded red.
Maryna classified UbVs as monospecific if they displayed >10-fold higher binding signal towards their target relative to every other Skp1-F-box tested. Based on this definition, Maryna generated monospecific UbVs for 11 of the 17 Skp1-F-box complexes that yielded specific binders (Figure 4-3B). UbV.Fw11.2, which targets Fbw11, was previously shown to be highly selective when tested against 6 other Skp1-F-box complexes (including the highly related Fbw1, 89% identity). When assayed here for binding to 21 Skp1-F-box complexes, it exhibited absolute specificity for Fbw11. Monospecificity was also exhibited by UbV.Fl11.1, which was selected for binding to Skp1-Fbl11 and did not bind appreciably even to the closely related Skp1-Fbl10 (67% identity in the F-box domains). UbVs exhibiting monospecificity were also isolated for Skp1 in complex with Fbl1, Fbl7, Fbl20, Fbo6, Fbo10, Fbo34, Fbo36, Fbo38 or Fbo45.

For two pairs of Skp1-F-box complexes, the selections yielded bispecific UbVs. UbVs selected for binding to Skp1-Fbo4 or Skp1-Fbo11 exhibited similar sequences (Figure 4-1D) and recognized both complexes. In the case of the Fbo5 and Fbo43 complexes, which share high sequence homology (44% identity in the F-box domain) and biological functions, the same UbV.Fo5/43.1 was the most specific and it recognized both complexes but no others. UbV.Fbl10.1, which was selected for binding to Fbl10, bound most strongly to its target Skp1-Fbl10 and to the similar Skp1-Fbl11 (67% identity in the F-box domain) but also bound weakly to several other Skp1-F-box complexes. Finally, the most specific UbVs selected for binding to Skp1-Fbw1 (UbV.Fw1.1), Skp1-Fbw7 (UbV.Fw7.5) or Skp1-Fbo7 (UbV.Fo7.1) exhibited significant binding to several complexes beyond the expected targets (Figure 4-3B). Taken together, these results demonstrate that extension and diversification of the β1-β2 loop enabled the selection of absolutely or highly selective UbVs for the majority of Skp1-F-box complexes that were targeted.

4.3.3 Structural Analysis of UbVs bound to Skp1-Fbl11 and Skp1-Fbl10 complexes

Stephen Orlicky of the Sicheri lab previously elucidated the structure of UbV.Fw7.1 bound to Skp1-Fbw7 (PDB:5IBK). However, we lacked a structural understanding of the mechanism of action of UbVs bearing extended β1-β2 loops designed to favour particular
Skp1-F-box complexes. Thus, to better understand the molecular basis for specificity and to confirm the design strategy, we chose representative examples of highly specific (UbV.Fl11.1) and less specific (UbV.Fl10.1) UbVs for structural analysis.

I solved the structure of UbV.Fl11.1 bound to Skp1-Fbl11 to 2.61 Å resolution by molecular replacement (Figure 4-5A, see Table 4-3 for data collection and structure refinement statistics). The crystal asymmetric unit consisted of 3 highly similar (average RMSD = 0.28 Å) trimeric complexes, and as such I focused my description on a single complex (chains B,F,H) (Figure 4-6A). The binding mode of UbV.Fl11.1 to Skp1-Fbl11 was similar in some respects to the binding mode of UbV.Fw7.1 to Skp1-Fbw7 (Figure 4-5B). UbV.Fl11.1 targeted a very similar Cul1 binding surface on the Skp1-F-box complex and the total surface areas buried by the UbVs were similar (1913 Å² for UbV.Fl11.1 and 1738 Å² for UbV.Fw7.1). Despite sharing a common binding surface on the Fbl11-Skp1 complex, UbV.Fl11.1 bound in a strikingly different orientation relative to UbV.Fw7.1 (a 70° rotation and a 3 Å translation of the UbVs) (Figure 4-5B). The change in binding orientation of UbV.Fl11.1 relative to UbV.Fw7.1 may be caused by one or a combination of the following factors: 1) the 8 residue insertion within the β1-β2 loop of UbV.Fl11.1; 2) 9 substitutions outside the β1-β2 loop that UbV.Fl11.1 contains relative to UbV.Fw7.1, which were incorporated to improve Skp1 binding (A12T, I42V, L49R, H62Q, K63R, I72V, R74G, G75R, G76R, substitutions are written as UbV.Fw7.1 to UbV.Fl11.1); 3) the presence of Skp1 acidic loops in the Skp1-Fbl11-UbV.Fl11.1 complex that were shortened in the Skp1-Fbw7-UbV.Fw7.1 complex 155; 4) the two extra C-terminal residues on UbV.Fw7.1 relative to UbV.Fl11.1. It is not readily apparent which of these changes are responsible for the altered binding modes of the two UbVs.
Figure 4-5. The crystal structure of UbV.Fl11.1 bound to Skp1-Fbl11.

A) Overall architecture of the UbV.Fl11.1-Skp1-Fbl11 complex. B) The structures of UbV.Fl11.1-Skp1-Fbl11 and UbV.Fw7.1-Skp1-Fbw7 superimposed by their Skp1 subunits. The rotation axis, rotation angle and translation required to superimpose UbV.Fw7.1 onto UbV.Fl11.1 are shown. Translation is occurring outwards from page. C) Open book view of the contact surfaces between UbV.Fl11.1 and Skp1 and Fbl11. Skp1 and Fbl11 residues engaged by the β1-β2 loop of UbV.L11.1 are colored yellow and those engaged by UbV.L11.1 residues that were fixed in the library are colored orange. Residues that contact both the β1-β2 loop and the fixed residues are colored based on which part of UbV.L11.1 they form the majority of their contacts with. Unengaged surfaces of Skp1 are colored blue and unengaged surfaces of Fbl11 are colored red. The UbV.L11.1 β1-β2 loop and core residues engaged in the interaction with the Skp1-Fbl11 complex are
colored yellow or orange, respectively. Unengaged surfaces on UbV.L11 are colored green. D) Details of the molecular interactions between UbV.L11.1 and Fbl11. Dashed lines indicate hydrogen bonds.
Table 4-3. Crystallographic data collection, processing, and refinement statistics.

<table>
<thead>
<tr>
<th></th>
<th>UbV.F11.1-Skp1-Fbl11</th>
<th>UbV.F10.1-Skp1-Fbl10</th>
<th>UbV.F10.1-Skp1-Fbl11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
<td>0.979</td>
<td>0.979</td>
</tr>
<tr>
<td>Cell dimension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a,b,c (Å)</td>
<td>60.6, 118.2, 146.3</td>
<td>65.4, 81.8, 128.7</td>
<td>38.2, 119.6, 63.7</td>
</tr>
<tr>
<td>α,β,γ (°)</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
<td>90, 98.5, 90</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50 – 2.61</td>
<td>50 – 2.66</td>
<td>19.82 – 3.27</td>
</tr>
<tr>
<td>R_index</td>
<td>0.102 (1.00)</td>
<td>0.174 (1.07)</td>
<td>0.179 (1.07)</td>
</tr>
<tr>
<td>CC (1/2)</td>
<td>(0.864)</td>
<td>(0.781)</td>
<td>(0.547)</td>
</tr>
<tr>
<td>Total no. of observations</td>
<td>218630</td>
<td>109832</td>
<td>31378</td>
</tr>
<tr>
<td>Total no. unique observations</td>
<td>32211 (2735)</td>
<td>20716 (2028)</td>
<td>8718 (1811)</td>
</tr>
<tr>
<td>Mean [(I)/σ(I)]</td>
<td>24.7 (1.7)</td>
<td>16.5 (2.15)</td>
<td>5.2 (1.1)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.5 (85.4)</td>
<td>99.8 (99.9)</td>
<td>98.9 (99.3)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>6.8 (5.2)</td>
<td>5.3 (4.6)</td>
<td>3.6 (3.5)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>43.4 – 2.61</td>
<td>47.5 – 2.66</td>
<td>19.82 – 3.27</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>31608</td>
<td>20406</td>
<td>8658</td>
</tr>
<tr>
<td>R_index/R_free</td>
<td>23.7/29.4</td>
<td>21.6/26.3</td>
<td>26.0/31.7</td>
</tr>
<tr>
<td>No. of atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>6166</td>
<td>4071</td>
<td>3529</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Water</td>
<td>70</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>Average B-factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>66.88</td>
<td>69.54</td>
<td>91.84</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Water</td>
<td>51.17</td>
<td>49.33</td>
<td>N/A</td>
</tr>
<tr>
<td>R.m.s.d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>0.006</td>
<td>0.006</td>
<td>0.003</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.799</td>
<td>0.881</td>
<td>0.615</td>
</tr>
<tr>
<td>Ramachandran statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residues in favoured regions (%)</td>
<td>95.21</td>
<td>96.39</td>
<td>95.06</td>
</tr>
<tr>
<td>Residues in allowed regions (%)</td>
<td>4.66</td>
<td>3.61</td>
<td>4.94</td>
</tr>
<tr>
<td>Residues in disallowed regions (%)</td>
<td>0.13</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Statistics in brackets are for highest resolution shell
Figure 4-6. Comparison of the different protein complexes in the asymmetric units.

A) Superposition of the three unique complexes in the UbV.Fbl11.1-Skp1-Fbl11 crystal structure. B) Superposition of the two unique complexes in the UbV.Fbl10.1-Skp1-Fbl10 crystal structure. C) Superposition of the two unique complexes in the UbV.Fbl10.1-Skp1-Fbl11 crystal structure in crystal asymmetric units.
Consistent with the library design, the extended β1-β2 loop of UbV.Fl11.1 engaged the F-box domain, and furthermore, the surface area buried between UbV.Fl11.1 and the Fbl11 F-box subunit (Figure 4-5C) was increased greatly relative to that between UbV.Fw7.1 and the Fbw7 F-box (727 Å² and 322 Å², respectively). In particular, the β1-β2 loop of UbV.Fl11.1 formed the majority of the contacts with Fbl11 (~75% of the buried surface area), but other residues that were fixed in the library design also contributed to the interaction. Contacts with the F-box mediated in part by the extended β1-β2 loop included both hydrophobic and hydrogen bond interactions, and provided a likely basis for the specificity of the UbV. Specifically, hydrophobic residues Trp892 and Met893 on Fbl11 interacted with UbV residues Tyr11*, Leu71* and Phe73*, while Met899 on Fbl11 interacted with Leu11d* (UbV residues are indicated by “*” and insertions relative to the Ub sequence are indicated by lower case letters) (Figure 4-5D, left panel and detailed stereo view in Figure 4-7, top panel). Hydrogen bond interactions involved the side chain of Trp892 with the side chain of Gln40* and backbone oxygen of Val72* (Figure 4-5D, left panel), the side chain of Arg895 with the main chain oxygen atoms of Ser10*, Tyr11* and His11b*, the side chain of Arg928 with the main chain oxygen of His11b* (Figure 4-5D, center panel), and the side chain of Glu896 with the side chain of Ser10* (Figure 4-5D, right panel).
Figure 4-7. Stereo view of interaction between UbV.Fbl11.1 and Skp1-Fbl11.

Top panel depicts interaction interface centered on Fbl11 residue Trp892, which was shown to be a key determinant for binding. Bottom panel depicts a valine at the corresponding position of Trp892. This substitution reduces the ability of UbV.Fbl11.1 to bind Skp1-Fbl11 ~ 9-fold and correlates with a loss of H-bonding interactions and extensive hydrophobic Van der Waal interactions with UbV.Fbl11.1. Dashed lines indicate hydrogen bonds.
Notably, amongst all the contact residues on Fbl11, only Trp892 (Val1064 in Fbl10) differed from the corresponding residues in Fbl10, but nonetheless, Fbl10 exhibited only weak binding to UbV.Fl11.1. Trp892 makes contacts with residues both within and outside of the β1-β2 loop (Figure 4-5D left panel and Figure 4-7 top panel). The smaller side chain of Val1064 in Fbl10 relative to Trp892 in Fbl11 would likely make suboptimal packing contacts with a loss of hydrogen bond interactions (Figure 4-7 bottom panel). Thus, it appears that Trp892 is a key determinant of the specificity of UbV.Fl11.1 for Skp1-Fbl11 over Skp1-Fbl10. To test my hypothesis Maryna substituted Val for Trp at position 892 in Fbl11 and Trp for Val at position 1064 in Fbl10 and tested each for binding to UbV.Fl11.1 (Table 4-4) using a competitive protein ELISA. In this assay Skp1-Fbl11 displayed an IC$_{50}$ of 79 nM against UbV.Fl11.1, similar to the K$_d$ of 25 nM measured by ITC (Figure 4-8, left panel). Consistent with the preference for Fbl11 demonstrated by phage ELISA (Figure 4-3B), UbV.Fl11.1 bound Fbl11 with ~10-fold greater affinity relative to Fbl10 (Table 4-4). Confirming my prediction that Trp892 specifies preferential binding of UbV.Fl11.1 to Fbl11, the Fbl11 variant with a Trp892Val substitution displayed reduced binding comparable to that of Fbl10, and conversely, the Fbl10 variant with a Val1064Trp substitution displayed increased binding comparable to that of Fbl11 (Table 4-4).
Table 4-4. Mutational analysis of Fbl10 and Fbl11 binding to target UbVs.

<table>
<thead>
<tr>
<th></th>
<th>Fbl11</th>
<th>Fbl11</th>
<th>Fbl10</th>
<th>Fbl10</th>
<th>Fbl10</th>
<th>Fbl10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IC$_{50}$ (nM)$^a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ubv.Fl11.1</td>
<td>$79 \pm 11$</td>
<td>$690 \pm 110$</td>
<td>$870 \pm 80$</td>
<td>$71 \pm 10$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ubv.Fl10.1</td>
<td>$430 \pm 30$</td>
<td></td>
<td>$140 \pm 11$</td>
<td></td>
<td>$220 \pm 30$</td>
<td>$22 \pm 1.4$</td>
</tr>
</tbody>
</table>

$^a$ IC$_{50}$ values were calculated by competitive ELISA as the concentration of Skp1-F-box in solution that blocked 50% of UbV binding to immobilized Skp1-F-box complexes.
Figure 4-8. Isothermal titration calorimetry analysis of UbV.Fbl11.1 binding to Skp1-Fbl11, UbV.Fbl10.1 binding to Skp1-Fbl10, and UbV.Fbl10.1 binding to Skp1-Fbl11.

Experiment performed in triplicate. Values represent mean ± SD.
I also used molecular replacement to solve the structure of UbV.Fl10.1 bound to each of its two strongest binding partners, Skp1-Fbl10 and Skp1-Fbl11, at 2.66 Å and 3.27 Å resolution, respectively (Figure 4-9A; see Table 4-3 for data collection and structure refinement statistics). Both crystal asymmetric units consisted of 2 highly similar trimeric complexes (RMSD = 0.35 Å for UbV.Fl10.1-Skp1-Fbl10 and RMSD = 0.95 Å for UbV.Fl10.1-Skp1-Fbl11), and as such, I focused my descriptions on a single complex for each structure (chains A,C,F and chains A,C,H respectively; Figure 4-6B,C). The overall binding modes of UbV.Fl10.1 to Skp1-Fbl10 and to Skp1-Fbl11 were highly similar to each other and to the binding mode of UbV.Fl11.1 to Skp-Fbl11, with only small differences in center of mass positions and rotations (Figure 4-10). Thus, in all three UbV structures with extended β1-β2 loops, the UbVs bound to Skp1-F-box complexes with very similar orientations that were distinct from that of UbV.Fw7.1 with a short β1-β2 loop.
Figure 4-9. The crystal structures of UbV.FI10.1 bound to Skp1-Fbl10 or Skp1-Fbl11.
A) Overall architecture of the UbV.Fl10.1-Skp1-Fbl10 and UbV.Fl10.1-Skp1-Fbl11 complexes. B) Open book views of contact surfaces between UbV.L10.1 and Skp1-Fbl10 or Skp1-Fbl11. Coloring is the same as Figure 4-5, except for unengaged surfaces of Fbl10, which are colored pink. C) Details of the molecular interactions between UbV.L10.1 and Skp1-Fbl10 or Skp1-Fbl11. Dashed lines indicate hydrogen bonds.
Figure 4-10. Comparison of the crystal structures of the Ub.V.Fl11.1-Skp1-Fbl11, Ub.V.Fl10.1-Skp1-Fbl10, and Ub.V.Fl10.1-Skp1-Fbl11 complexes.

A) The structures of Ub.V.Fl10.1-Skp1-Fbl10 and Ub.V.Fl10.1-Skp1-Fbl11 are superimposed by their Skp1 subunits. The rotation axis, rotation angle and translation required to superimpose Ub.V.Fl10.1 bound to Skp1-Fbl10 onto Ub.V.Fl10.1 bound to Skp1-Fbl11 are shown. B) The structures of Ub.V.Fl11.1-Skp1-Fbl11 and
UbV.Fl10.1-Skp1-Fbl10 are superimposed by their Skp1 subunits. The rotation axis, rotation angle and translation required to superimpose UbV.Fl10.1 onto UbV.Fl11.1 are shown. Translation is occurring outwards of page. C) The structures of UbV.Fl11.1-Skp1-Fbl11 and UbV.Fl10.1-Skp1-Fbl11 complexes are superimposed by their Skp1 subunits. The rotation axis, rotation angle and translation required to superimpose UbV.Fl10.1 onto UbV.Fl11.1 are shown.
The surfaces between UbV.Fbl10.1 and Skp1-Fbl10 or Skp1-Fbl11 were highly overlapping with contact areas of 2078 Å² and 1724 Å², respectively (Figure 4-9). Consistent with the lower specificity of UbV.Fbl10.1 relative to UbV.Fl11.1 (Figure 4-3B), UbV.Fbl10.1 had a smaller contact surface with Fbl10 (336 Å²) or Fbl11 (344 Å²) than did UbV.Fbl11.1 with Fbl11 (727 Å²).

Among F-box residues that contact UbV.Fbl10.1 and thus may contribute to binding specificity, Val1069 and Ala1072 in Fbl10 and the corresponding Val897 and Ser900 in Fbl11, differ in other F-box proteins that bind weakly to UbV.Fbl10.1 (Figure 4-9C, Figure 4-3A). These side chains in the F-box appear to allow efficient packing against Leu11a* in UbV.Fbl10.1 (Figure 4-9C). In contrast, the weak binding F-box proteins have bulkier residues at the position corresponding to Ala1072/Ser900 and/or contain a residue other than valine at the position corresponding to Val1069/Val897. To test whether my hypothesis of these two positions dictating the preference of Ub.V.Fl10.1 for Fbl10 and Fbl11 over other F-box proteins, Maryna assessed the effect of V1069A (Ala representing an amino acid different than the conserved Val) and A1072Y (Tyr representing a bulkier sidechain compared to Ala/Ser) substitutions in Fbl10 on Ub.V.Fl10.1 binding using the competitive protein ELISA. In this assay, Skp1-Fbl10 exhibited an IC₅₀ of 140 nM against Ub.V.Fl10.1 and Skp1-Fbl11 exhibited an IC₅₀ of 430 nM against Skp1-Fbl11 (Table 4-4), similar to the K_d of 91 nM or 421 nM, respectively, I measured by ITC (Figure 4-8, middle and right panels). Contrary to my hypothesis, the Fbl10 V1069A substitution did not decrease binding considerably, while the Fbl10 A1072Y substitution actually increased binding (Table 4-4). Thus the basis for the specificity of Ub.V.Fbl10.1 for Skp1-Fbl10 and Skp1-Fbl11 remains an open question.

4.3.4 Effects of UbVs on the function of SCF°Fbl1 and SCF°Fbo11 in cells.

To test whether UbVs are able to inhibit SCF E3 ligases in cells, Maryna chose Ub.V.L1.1 and Ub.V.011.1 targeting SCF°Fbl1 and SCF°Fbo11, respectively, for testing in HEK293T cells. SCF°Fbl1 (Skp2) is one of the best characterized SCF E3 ligases and is considered an important oncogene with numerous studies attempting to engineer SCF°Fbl1 inhibitors for cancer treatment 20. SCF°Fbo11 is also well characterized and, along with SCF°Fbl1, has been shown to be among the most abundant SCF ligases in HEK293T cells 156. In order to determine
whether cellular expression of UbV.Fl1.1 or UbV.Fo11.1 could inhibit their cognate ligases, Maryna analyzed the stability of ligase substrates. Consistent with anticipated inhibitory effects, UbV.Fl1.1 stabilized the SCFFbl1 substrate p27, whereas UbV.O11.1 stabilized the SCFFbo11 substrate Snail1 (Figure 4-11AB) 157,158. The observed effect was specific, as no stabilization was observed for substrates targeted by other SCF E3 ligases including Cyclin E (Fbw7 substrate) and Cry2 (Fbxl3 substrate) in the case of UbV.L1.1 and Cyclin E, Cry2 and c-Myc (Fbw7 and other E3 ligase substrate) in the case of UbV.O11.1 (Figure 4-11AB). Taken together with the results of the previous study 91, which demonstrated inhibition of cellular SCFFbw7 and SCFFbw11 by UbVs, these results confirm that UbVs can target Cul1 binding sites on SCF E3 ligases to inhibit their functions in cells.
Figure 4-11. Biological activity of UbVs in HEK293T cells.

A) Expression of UbV.Fl1.1 stabilizes the SCFFbl1 substrate p27, but has no effect on stability of the SCFFbw7 substrate Cyclin E and the SCFFbx13 substrate Cry2. Cells were transiently transfected with either empty vector or vector encoding FLAG-UbV.Fl1.1. Cells were treated with 100 µg/ml Cycloheximide (CHX) for the indicated time points and cell lysates were probed with antibodies against endogeneous p27, Cyclin E, Cry2, GADPH (loading control) or FLAG (UbV levels). B) Expression of UbV.O11.1 stabilizes the SCFFbo11 substrate Snai1 but has no effect on the stability of the SCFFbw7 substrate Cyclin E, the SCFFbx13 substrate Cry and the SCFFbw7,Skp2 substrate c-Myc. Cells were transiently transfected with HA-tagged Snai1 and either empty vector or vector encoding FLAG-UbV.Fo11.1. Cells were treated with 20 µg/ml CHX for the indicated time points and cell lysates were probed with antibodies against Snai1, Cyclin E, Cry2, c-Myc, GADPH (loading control) or FLAG (UbV levels).
4.4 Discussion

Our research groups previously reported a strategy to develop specific UbV inhibitors of SCF E3 ligases using the well-characterized F-box proteins Fbw7 and Fbw11 as test cases. Here Maryna applied this approach to approximately half of the human F-box family by attempting to target 32 of the 69 members (Figure 4-1B). Maryna successfully purified 23 of the 32 F-box domains in complex with Skp1, and 17 of these UbV binders displayed at least some level of specificity. Our results demonstrate that this strategy can generate fairly specific UbV inhibitors for a large majority of stable Skp1-F-box complexes that can be purified by recombinant expression in bacteria. Further improvements in this success rate should be achievable by optimization of the bacterial Skp1-F-box expression system or by using eukaryotic expression systems that are better able to support the correct folding of Skp1-F-box complexes.

A survey of binding specificities across a large panel of Skp1-F-box proteins (Figure 4-3B) revealed that many UbVs display high specificity for their cognate targets. Notably, for 11 Skp1-F-box complexes, Maryna identified at least one monospecific UbV, and for 4 others, Maryna identified bispecific UbVs. Maryna designed the F-box targeted UbV library with the expectation that extension and diversification of the β1-β2 loop would yield UbV binders with improved specificity through expanded and optimized interactions with the F-box domain. Consistent with the design, the β1-β2 loop of the highly specific UbV.Fl11.1 forms extensive interactions with the Fbl11 F-box domain (Figure 4-5CD). My structures reported in this study of UbVs bound to Skp1-F-box complexes reveal that Gln40*, Leu71*, Val 72* and Phe73*, which were fixed in the library design, make contacts with the F-box domain (Figure 4-5D, Figure 4-9C) that were prohibited by distance constraints in the original structure of UbV.Fw7.1 bound to Skp1-Fbw7. We posit that these positions may be fertile ground for additional sequence diversification to further optimize specificity and affinity.

Although Maryna focused her efforts on developing UbV inhibitors for SCF E3 ligases, targeting other CRL families is also of great interest. While the approach described here is not directly applicable to Cul4A and Cul4B families, since their substrate receptors do not
interact directly with Cullin subunits\textsuperscript{159}, it could likely be applied to the other three CRL subfamilies. Namely EloBC-VHL-Cul2, EloBC-SOCs-box-Cul5, and BTB-Cul3 families not only share a common architecture with the SCF E3 ligases, they also share a structurally conserved interaction interface with Cul1\textsuperscript{160}. As in the case of SCF E3 ligases, UbV inhibitors of these CRL families could be generated by targeting a defined and relatively small component that is responsible for the interaction with a Cullin subunit, without any knowledge of the overall structure or function. This strategy may be most successful for the BTB-Cul3 family, which is distinct in that BTB proteins interact with Cul3 directly, limiting the issue of specificity that arises when a common component such as Skp1 contributes to the targeted interaction surface. Most importantly, previous UbV inhibitors of Fbw7 and Fbw11\textsuperscript{91} and the UbVs described here can serve as useful tools to interrogate functions of SCF E3 ligases. Thus, we are well positioned to apply the UbV strategy to systematically develop specific inhibitors to a significant fraction of the large CRL family of E3 ligases.
Chapter 5: The ubiquitin interacting motifs of USP37 act on the proximal Ub of a di-Ub chain to enhance catalytic efficiency

Noah Manczyk\textsuperscript{1,2}, Gianluca Veggiani\textsuperscript{3,4}, Joan Teyra\textsuperscript{4}, Amy Strilchuk\textsuperscript{2}, Sachdev S Sidhu\textsuperscript{3,4,*}, Frank Sicheri\textsuperscript{1,2,3,*}

\textbf{Affiliations}
1 Department of Biochemistry, University of Toronto, Toronto, ON, Canada, M5S 3E1
2 Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, M5G 1X5
3 Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada, M5S 3E1
4 Banting and Best Department of Medical Research, Terrence Donnelly Center for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada, M5S 3E1

\textbf{Author list footnotes}
* corresponding authors

\textit{A version of this chapter is published in Scientific Reports 2019 March;9(1)4119.}

\textbf{Attributions:} Dr. Sicheri and I conceived the original project idea detailed in this chapter. I performed all the experiments and work presented in the chapter with the exceptions of (1) UbV generation which was performed by Dr. Gianluca Veggiani and Dr. Joan Teyra in the Sidhu lab and (2) purification of UbV.core and UbV.UIM1 which was performed by Amy Strilchuk a project student in the Sicheri lab working under my supervision. Throughout Chapter 4, “I” is used where work was performed solely by me, “we” is used where I played an assisting role, and the respective author(s) name such as “Gianluca” where experiments were performed exclusively by others.
5.1 Introduction

Ubiquitination is the process whereby the small protein ubiquitin (Ub) is covalently attached to a substrate protein via a cascade of three (E1-E2-E3) enzymes. Multiple Ub moieties can be covalently linked together to form chains on the substrate via one of the seven lysine side chains or the free amino terminus of Ub. The eight Ub chain types can lead to a variety of cellular outcomes for a modified target protein, the best characterized of which is protein degradation through the 26S proteasome in the case of K48-linked chains\(^1\). Ubiquitination plays a key regulatory role in diverse cellular processes including DNA repair and cell-cycle progression, and as such the process is tightly regulated\(^{161,162}\).

The deconstruction of Ub chains is carried out by deubiquitinases (DUBs). In human, there are 7 families of DUBs based on domain structure, two of which have been recently discovered\(^{57-60}\), and among these, the Ub Specific Protease (USP) family represents the largest with more than 50 members. In contrast to the OTU family of DUBs, USP family members are notable for their generally poor ability to discriminate between different Ub chain linkages\(^ {163}\). USPs commonly possess auxiliary domains, either N or C-terminal flanking, or internal to their catalytic domains at one of 5 possible insertion sites\(^ {78,117}\).

These domains function in different capacities, for example, binding Ub chains in the case of the UBA (Ub associated) domains of USP5, promoting protein complex formation in the case of the B-box domain in CYLD, or enhancing catalytic efficiency in the case of the DUSP-Ubl domains in USP4\(^ {61,79,81}\).

USP37 is a USP family member implicated in the regulation of multiple critical cellular processes. For example, USP37 deubiquitinates cyclin A during G1 phase causing cyclin A stabilization and timely entry into S phase\(^ {83}\). Additionally, USP37 facilitates the resolution of sister chromatids during prophase in a manner dependent on its DUB catalytic function\(^ {84}\). In relation to its role maintaining chromosomal integrity, USP37 localizes to double-strand breaks and promotes BRCA1 inclusion into complexes responsible for homologous recombination\(^ {85}\). Lastly, USP37 deubiquitinates and stabilizes the proto-oncogene c-Myc and the oncogenic fusion PLZF-RARA, suggesting that inhibition of USP37 DUB activity could have therapeutic potential\(^ {86,87}\).
USP37 has a distinctive domain architecture consisting of an N-terminal PH domain, an interdomain linker and a C-terminal catalytic domain. Located within the catalytic domain is a large insertion of 284 amino acids (human USP37) containing three Ub-interacting motifs (UIMs) embedded at a site approximately 30 Å from the catalytic cleft. UIMs are single alpha-helical elements that bind to Ub with modest affinity (0.1-2 mM). UIMs conform to the consensus sequence e-e-x-x-φ-x-A-φ-x-(φ/e)-S-z-x-e, where e is an acidic residue, φ is a hydrophobic residue, z is a bulky hydrophobic or polar residue with high aliphatic content, A is alanine, S is serine and x is any residue. UIM binding to Ub is routinely disabled by mutation of the consensus alanine position to glycine or the consensus serine position to alanine. In other DUBs, UIMs have been shown to confer cleavage preference for specific Ub chain types, such as K63-linked chains in the case of OTUD1 and Ataxin-3 or K48-linked Ub chains in the case of USP25. Additionally, the UIMs of USP25 and Ataxin-3 have been shown to increase the ubiquitination state of the DUB itself, although the precise mechanism by which this is achieved remains unknown.

Previous studies have shown that the UIMs of USP37 play an essential role in Ub cleavage activity and substrate binding properties of USP37. Specifically, A814G and S818A mutations to UIM2 and/or A836G and S840A mutations to UIM3 impaired the ability of USP37 to cleave K48- and K63-linked chains, while V712G and S716A mutations to UIM1 had no discernable effect. Furthermore, while combined mutation of all three UIMs had a marked effect on DUB activity, it had no effect on the cleavage specificity of USP37 towards K48-linked chains over K63-linked chains. Lastly, mutation of UIM2 and/or UIM3 perturbed USP37 binding to the cohesin regulator WAPL and to endogenous Ub-protein conjugates. While these observations made clear that UIM2 and UIM3 play an important functional role in supporting the DUB activity of USP37, several questions remain unresolved, including: 1) do the individual UIMs impact the ability of USP37 to cleave the 6 Ub chain types not previously tested, 2) does mutation of the UIMs in USP37 selectively affect the $k_{cat}$ or $K_M$ of the enzyme, and 3) do the UIMs of USP37 act by engaging the proximal or distal position Ub moiety in a Ub chain. To address these questions, I have performed detailed mutational, biochemical and enzymatic characterization of the USP37 UIMs.
5.2 Experimental Procedures

5.2.1 Protein Purification

For biochemical experiments, a codon optimized gene (GeneArt) encoding for Danio rerio USP37\textsuperscript{wt} (312-927) containing a N-terminal GST tag was expressed in Escherichia coli BL21(DE3)-RIL cells. Cells were lysed by homogenization in lysis buffer (50 mM HEPES pH 8, 500 mM NaCl, 5% glycerol, 5 mM DTT) with 0.5 mM phenylmethane sulfonfyl fluoride (PMSF). After clarification via centrifugation, supernatant was incubated via gravity column with glutathione resin and washed with lysis buffer. Protein bound to glutathione resin was incubated overnight in presence of TEV protease at 4° C. Protein was eluted with lysis buffer and concentrated to ~4 mg/mL and subsequently flash diluted to ~0.2 mg/mL in low salt buffer (25 mM HEPES pH 8, 50 mM NaCl, 5 mM DTT) and loaded onto an HiTrap Q HP anion exchange column (GE Healthcare). Protein was eluted via a linear gradient with low and high salt buffer (25 mM HEPES pH 8, 500 mM NaCl, 5 mM DTT). Fractions containing USP37\textsuperscript{wt} were pooled, concentrated and injected onto a Superdex 200 column, pre-equilibrated with sizing buffer (25 mM HEPES pH 8, 150 mM NaCl, 5 mM DTT). Fractions containing USP37\textsuperscript{wt} were pooled, concentrated and stored at -80 ° C. USP37 mutants (V680G and S684A for UIM1, A774G and S778A for UIM2, and A796G and S800A for UIM3) were cloned from the WT construct using the QuikChange method (Stratagene) and expressed and purified in similar fashion.

UbVs containing a 6xHis (UbV.UIM1), a 6xHis-GST (UbV.core) and 6xHis-MBP (UbV.UIM*) tag were expressed in E. coli BL21(DE3)-RIL cells. Cells were lysed by sonication in low imidazole buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 5mM imidazole, 4 mM b-mercaptoethanol (b-mer), 0.5 mM PMSF). After clarification via centrifugation, supernatant was loaded onto a Hitrap IMAC HP column (GE Healthcare). Protein was eluted using a linear gradient (5-300 mM imidazole) and dialyzed overnight in low imidazole buffer in the presence of TEV protease. A subtractive nickel column affinity purification was performed to remove TEV protease. Flow-through was collected, concentrated and injected onto a Superdex 75 column pre-equilibrated with sizing buffer (25 mM HEPES pH 8, 150 mM NaCl, 5 mM DTT).
7.5, 100 mM NaCl, 2 mM DTT). Protein fractions were pooled, concentrated and stored at -80° C.

For phage display selections, USP37 protein fragments were expressed either as a GST-fusion (USP37Δloop (Danio rerio 312-927 Δ600-884)), His-GST-fusion (human UIM2 (806-825)) or His-MBP fusion (UIM1 (Danio rerio 670-693)). Purification of GST-USP37Δloop followed the same strategy as USP37wt with the exception that GST-USP37Δloop was eluted from glutathione resin using lysis buffer with 10 mM reduced glutathione rather than by cleavage with TEV protease. His-GST-UIM2, cells were lysed by sonication in low imidazole buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, 20 U/ml Benzonase (EMD Millipore), protease inhibitor cocktail (Sigma-Aldrich)). After clarification by centrifugation, supernatant was incubated on Ni-NTA resin (Qiagen). Protein was eluted with an imidazole buffer gradient (30-300 mM), fractions containing UIM2 were pooled, and the buffer was exchanged by dialysis into PBS pH 7.4 at 4 °C. The protein was then concentrated and stored at -80 °C. Protein was concentrated and stored at -80 ° C. His-MBP-UIM1, cells were lysed by sonication in low imidazole buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole, 5% glycerol, 4 mM b-mer, 0.5 mM PMSF). After clarification by centrifugation, supernatant was loaded onto a Hitrap IMAC HP column (GE Healthcare). His-MBP-UIM1 was eluted using a linear imidazole gradient (5-300 mM). Protein was dialyzed overnight into PBS pH 7.4 at 4 °C, concentrated and stored at -80 °C.

5.2.2 Size-exclusion chromatography

USP37wt was purified by glutathione resin and anion exchange chromatography as detailed in the protein purification methods section. USP37wt was concentrated and injected (1 mL of ~2 mg/mL) onto a Superdex S200 (GE Healthcare) previously equilibrated with 25 mM HEPES pH 8, 150 mM NaCl, 5 mM DTT. Absorbance at 280nm was used to monitor the elution of the protein and determination of effective molecular weight was performed using a formula derived from protein standards.
5.2.3 Gel-based deubiquitination assays

Gel-based deubiquitination assays were performed in 25 mM Hepes pH 8, 150 mM NaCl, 5 mM DTT, 0.1 mg/ml BSA. For data in Figure 5-1B, 1 nM USP37 enzyme was incubated with 9 µM di-Ub substrate at 20 °C. Di-Ub chains were enzymatically produced by myself (K11, K48, and K63) or Boston Biochem (K6, K27, K29, and K33), with the exception of Met1 di-Ub chains, which were produced as a tandem fusion directly in E. coli. Met1-linked Ub chains were produced according to published methods 169. K48- and K63-linked Ub chains were produced as described170, with the use of a K48R and K63R Ub mutant (for distal Ub) and with the variation of using a Ub with the last two Gly residues deleted (Ub∆GG, for the proximal Ub). K11-linked Ub chains were produced similarly (K11R Ub for distal Ub and Ub∆GG for proximal Ub), with the exception of using 30 µM Ube2S as the E2 conjugating enzyme according to published methods 171. K6-, K27-, K29- and K33-linked Ub chains were obtained from a commercial source (Boston Biochem, catalog #: UC-11B, UC-61B, UC-81B, and UC-101B, respectively). Time points were analyzed by SDS-PAGE and staining with gel-code blue (Thermofisher, catalog # 24590). Images were taken on a Bio-Rad Chemidoc MP Imaging System. Images were cropped using Image Lab (Bio-Rad).

For data in Figure 5-2, 10 nM USP37 enzyme was incubated with 6 µM tetra-Ub at 20 °C. K48 and K63 tetra-Ub chains were obtained from a commercial source (Boston Biochem, catalog #: UC-210B and UC-310B, respectively). Time points were analyzed by SDS-PAGE and staining with Coomassie blue. Images were taken on a Bio-Rad Chemidoc MP Imaging System. Images were cropped using Image Lab (Bio-Rad).

For data in Figure 5-4, 1 nM USP37 enzyme was incubated with 9 µM of the indicated di-Ub substrate at 20 °C. Time points were analyzed by SDS-PAGE. Gels were fluorescently scanned using a Bio-Rad Chemidoc MP Imaging System and Green Epi illumination with a 605/50 filter. Gels were then stained with Coomassie blue and imaged on a Bio-Rad Chemidoc MP Imaging System. Images were cropped using Image Lab (Bio-Rad).

For data in Figure 5-7 and Figure 5-8, 5 nM USP37 enzyme was incubated with 18 µM di-Ub substrate at 20 °C. Mutant K48-linked di-Ubs were produced with the same method as WT K48-linked di-Ub using K48R Ub (for distal Ub) and I44A/W Ub∆GG (for proximal Ub).
Time points were analyzed by SDS-PAGE and staining with Coomassie blue. Images were taken on a Bio-Rad ChemiDoc MP Imaging System. Images were cropped using Image Lab (Bio-Rad).

5.2.4 Fluorescence deubiquitination assays

Fluorescent deubiquitination assays were performed at 30 °C in 20 µL DUB buffer (25 mM Hepes pH 8, 150 mM NaCl, 5 mM DTT, 0.1 mg/ml BSA, 0.03% BRIJ-35) in 384-well black flat-bottom plates (Corning 3573). Measurements were taken each minute in a BioTek Synergy Neo plate reader machine using excitation and emission wavelengths of 540 and 580 nM respectively for assays using internally-quench-fluorescent (IQF) probes and excitation and emission wavelengths of 345 and 445 nM respectively for assays using Ub-AMC.

For data in Figure 5-3A, substrate protein was a mixture of 1 part IQF K11-(BostonBiochem, catalog #: UF-440), K48- (Lifesensors, catalog # DU4802) or K63-linked (Lifesensors, catalog # DU6303) di-Ub and 4 parts of the respective unlabeled chain. 25 nM USP37 (wild type or mutant form) was added to the indicated concentrations of K11 substrate. 1.25 nM USP37 enzyme was added to the indicated concentrations of K48 substrate. 12.5 nM USP37 enzyme was added to the indicated concentrations of K63 substrate. V0 was calculated from the linear portion of the progress curve. V0 was converted from fluorescent units/min to µM/min using a Ub-TAMRA standard curve. V0 was plotted against di-Ub concentration and kinetic parameters (Km and kcat) were calculated using GraphPad Prism 5.

For data in Figure 5-3B, 25 nM (for K11), 1.25 nM (for K48) or 12.5 nM for (K63) of the indicated USP37 proteins were mixed with 250 nM of IQF K11, K48 or K63 di-Ub chains. For data in Figure 5-6C, 0.5 nM USP37wt or USP37mut1,2,3 was incubated with 250 nM Ub-AMC (Boston Biochem, catalog # U-550). For data in Figure 5-12, 1 nM USP37wt (10 µL) was incubated with 5 µL of varying concentrations of the respective UbV (0.008-250 µM) for 20 mins at 20 °C. 1 µM Ub-AMC or IQF K48 (5 µL) was then added, and measurements were taken. V0 was calculated after initial noise dissipated and linear portion of the
progress curve appeared. $V_0$ was plotted against UbV concentration and IC$_{50}$ values were calculated using GraphPad Prism 5.

### 5.2.5 Pulldown assays

GST-USP37 (Danio rerio 312-927), GST-USP37$^{\Delta$loop} (Danio rerio 312-927 Δ600-884), and GST-hsUSP37 (Homo sapiens 332-969) were expressed in E. coli BL21(DE3)-RIL cells that were lysed by homogenization in lysis buffer (25 mM Hepes pH 8, 500 mM NaCl, 5% Glycerol, 5 mM DTT), whereas GST, HisMBP, HisMBP-UIM1 (Danio rerio 670-693), HisMBP-UIM2 (Danio rerio 764-787) and HisMBP-UIM3 (Danio rerio 786-807) were expressed in E. coli BL21(DE3)-RIL that were lysed by sonication in lysis buffer. Proteins were incubated with their respective affinity resin (Glutathione or Maltose) for 1 hour for batch binding. Unbound proteins were washed away with lysis buffer. Beads were then equilibrated in binding buffer (25 mM Hepes pH 8, 150 mM NaCl, 5 mM DTT, 0.1 mg/ml BSA, 0.03% BRIJ-35). 30 µg of UbV.core, UbV.UIM1 or UbV.UIM$^*$ were then incubated with the beads for 1 hour and washed twice with binding buffer (wash volume = 25 times bed volume). Samples were resolved and analyzed via SDS-PAGE with Coomassie blue staining. Images were taken on a Bio-Rad ChemiDoc MP Imaging System.

### 5.2.6 Selection of UbVs by phage display

The phage-displayed UbV library 2 (Ernst et al., 2013) was used to perform binding selections for USP37$^{\Delta$loop} and UIM1. For selection of UbVs for binding to UIM2, Gianluca of the Sidhu lab used a new library (Veggiani and Sidhu) that was designed by inspection of the structure of a UbV bound to a yeast UIM (Manczyk et al., 2017). Selections were performed as described (Ernst et al., 2013). Briefly, library phage pools were cycled through rounds of binding selections with target proteins immobilized on 96-well Maxisorp Immunoplates (Nunc). After five rounds, phage from individual colonies were assessed for binding to immobilized proteins by phage ELISA (Persson et al., 2013), and clones that bound to target protein but not to negative control proteins were subjected to DNA sequencing to decode the sequences of the displayed UbVs.
5.3 Results

5.3.1 Impact of UIMs on the Ub chain specificity of USP37

I expressed the isolated catalytic domain of USP37 from *Danio rerio* (residues 312-927, denoted USP37<sub>wt</sub>), which unlike USP37 from human, mouse and chicken, could be expressed in bacteria in stable form for biochemical and enzymatic studies. *Danio rerio* and *Homo sapiens* USP37 share identical domain architectures and strong sequence similarity (72% and 77% similarity overall and over the catalytic domain, respectively) suggesting that this ortholog would serve as a good model for understanding the mechanism of action of the human enzyme (*Figure 5-1A*). To determine comprehensively if the UIMs of USP37 contribute to chain linkage cleavage specificity, I tested the ability of USP37<sub>wt</sub> to cleave all 8 possible di-Ub chain types including the 5 Ub chain types not previously tested. USP37<sub>wt</sub> displayed a preference for K6-, 11-, 33-, 48- and 63-linked di-Ub chains (*Figure 5-1B*, top panel). Strikingly, a mutant form of USP37 (USP37<sub>mUIM1,2,3</sub>) bearing mutations in all three UIMs predicted to abolish Ub binding (see experimental procedures for details) displayed no change in chain specificity. Indeed, USP37<sub>mUIM1,2,3</sub> still displayed a strong preference for K6-, K11-, K33-, K48- and K63-linked di-Ub chains, although with reduced activity overall relative to USP37<sub>wt</sub> (*Figure 5-1B*, bottom panel).
Figure 5-1. Di-Ub cleavage specificity of USP37<sup>wt</sup> and USP37<sup>mUIM1,2,3</sup>.

A) Domain architecture of human and zebrafish USP37. B) Deubiquitination time course for USP37<sup>wt</sup> (top panel) and USP37<sup>mUIM1,2,3</sup> (bottom panel) towards each of the 8 di-Ub linkages. Assay was performed in three independent experiments and representative gels are shown. USP37<sup>wt</sup> and USP37<sup>mUIM1,2,3</sup> were used at a concentration of 1 nM.
I next performed experiments using tetra Ub chains and observed a similar overall trend (Figure 5-2). However, zebrafish USP37 displayed a greater preference for K48 over K63 chains, which more closely matched the reported chain specificity of the human enzyme. Mutation of UIMs drastically decreased cleavage efficiency of USP37 against K48 and K63 tetra Ub chains without changing its preference for K48 tetra Ub chains. These findings further support the notion that the UIMs of USP37 do not confer specificity towards specific Ub chain types but are instead required for full enzymatic activity.
**Figure 5.2.** K48 and K63 tetra-Ub cleavage specificity of USP37\textsuperscript{wt} and USP37\textsuperscript{mUIM1,2,3}.

Deubiquitination time course for USP37\textsuperscript{wt} (top panel) and USP37\textsuperscript{mUIM1,2,3} (bottom panel) towards K48 and K63 tetra-Ub. Assay was performed in three independent experiments and representative gels are shown. USP37\textsuperscript{wt} and USP37\textsuperscript{mUIM1,2,3} were used at a concentration of 10 nM.
5.3.2 The effect of mutations to all three UIMs on the kinetic parameters of USP37

To determine how precisely the UIMs impact on the catalytic efficiency of USP37, I determined the effect of UIM mutations on the $k_{\text{cat}}$ and $K_m$ parameters of the enzyme. To this end, I performed kinetic experiments with USP37$^{\text{wt}}$ and USP37$^{\text{mUIM1,2,3}}$ against the three most preferred Ub chain substrates, namely K11-, 48-, and 63-linked chains, bearing internally quenched fluorescent (IQF) di-Ub probes (Figure 5-3A). In this assay format, cleavage of di-Ub liberates the fluorophore from the quencher, yielding an increase in fluorescence proportional to the amount of substrate cleaved. To estimate how well the IQF substrate mimics the unmodified substrates, I performed a gel-based cleavage assay comparing USP37$^{\text{wt}}$ cleavage of unmodified K48 and K63 di-Ub with IQF labeled K48 and K63 di-Ub (Figure 5-4). I observed that IQF modifications adversely impacted the cleavage efficiency of substrate with the negative effect on K63 chains being more pronounced than K48 chains. As such, when using IQF substrates I restricted my comparisons to reactions using the same substrate type. For K48-linked chains, I observed a large reduction in $k_{\text{cat}}$ (decreased $\sim$15-fold) and a smaller perturbation in $K_m$ (increased $\sim$1.8-fold) (Figure 5-3A left panel, Table 5-1 – see also Figure 5-5 for a zoom-in view of the USP37$mUIM1,2,3$ cleavage profile) in response to all three UIM mutations. For K11- and K63-linked chains, I observed more modest reductions in both $k_{\text{cat}}$ (decreased $\sim$2.6-fold for K11 and $\sim$3.4-fold for K63) and $K_m$ (decreased $\sim$1.7-fold for K11 and $\sim$2.1-fold for K63) (Figure 5-3A middle and right panel, Table 5-1). Thus, the ability of UIMs to bind Ub plays differential roles on the kinetic parameters of USP37 depending on the Ub chain type tested.
Figure 5-3. Activity of USP37\textsuperscript{wt} and USP37 mutants towards IQF K11, K48, and K63 probes.

A) Michaelis-Menten kinetic analysis of USP37\textsuperscript{wt} and USP37\textsuperscript{mUIM1,2,3} for the substrates K11-, K48- and K63-linked IQF di-Ubs. Curves represent measurements from three independent experiments each measured in duplicate. Values reported are mean ± SD. B) Progress curves for USP37\textsuperscript{wt} and USP37 mutants towards IQF K11-, K48- and K63-linked di-Ubs. Assay was performed in duplicate in three independent experiments and representative curves are shown. For both experiments, USP37 was used at 25, 1.25 or 12.5 nM for K11, K48 or K63 di-Ub chains, respectively.
Figure 5-4. USP37<sup>wt</sup> activity towards unmodified and IQF modified K48 and K63 di-Ub chains.

Time course for USP37<sup>wt</sup> deubiquitination of K48 (top set of panels) and K63 (bottom set of panels) with unmodified and IQF modified di-Ub chains. Gels were scanned for fluorescence to detect cleavage of IQF modified di-Ub chains and stained with Coomassie to detect cleavage of both unmodified and IQF modified di-Ub chains. Assay was performed in three independent experiments and representative gels are shown. USP37<sup>wt</sup> and USP37<sup>mUIM1,2,3</sup> were used at a concentration of 1 nM.
Table 5-1. Michaelis-Menten analysis of USP37<sup>wt</sup> and USP37<sup>UIM1,2,3</sup> for K11, K48, and K63 di-Ub substrates.

<table>
<thead>
<tr>
<th>Substrate Enzyme</th>
<th>K11</th>
<th>K48</th>
<th>K63</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>USP37&lt;sup&gt;wt&lt;/sup&gt;</td>
<td>USP37&lt;sup&gt;UIM1,2,3&lt;/sup&gt;</td>
<td>USP37&lt;sup&gt;wt&lt;/sup&gt;</td>
</tr>
<tr>
<td>$[E]$ (nM)</td>
<td>25</td>
<td>25</td>
<td>1.25</td>
</tr>
<tr>
<td>$V_{\text{max}}$ (nM/min)</td>
<td>285 ± 22</td>
<td>108 ± 12</td>
<td>17 ± 2.1</td>
</tr>
<tr>
<td>$k_{\text{cat}}$ (s&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>0.2</td>
<td>0.07</td>
<td>0.2</td>
</tr>
<tr>
<td>$K_M$ (µM)</td>
<td>11 ± 2.3</td>
<td>19 ± 1.2</td>
<td>0.40 ± 0.09</td>
</tr>
</tbody>
</table>
Figure 5-5. Michaelis-Menten analysis of $\text{USP37}^{\text{mUIM1,2,3}}$ for K48-linked IQF di-Ub.

Curve represents measurements from three independent experiments measured in duplicate. Enzyme was used at 1.25 nM. The data for $\text{USP37}^{\text{mUIM1,2,3}}$ was reproduced here from the K48 subpanel of Figure 5-3A for clarity.
5.3.3 Effects of UIM mutations on USP37’s ability to cleave K11, 48 or 63-linked chains

To determine if any particular UIM is strongly associated with cleavage of a specific chain linkage, I generated UIM mutations predicted to abolish Ub binding in all possible single, double and triple combinations. I then tested each protein for its ability to cleave K11-, K48- or K63-linked di-Ub chains. In characterizing the activity of the mutants against these 3 linkage types, I noted the following strong patterns of behavior (Figure 5-3B).

UIM1 appeared completely dispensable for the proteolytic activity of USP37 against all tested Ub chain types. Specifically, introduction of disabling mutations within UIM1 had no adverse effect on enzyme activity regardless of the presence or absence of mutations in the other UIMs (compare enzyme pairs against the three di-Ub substrates: USP37<sup>wt</sup> (black) to USP37<sup>mUIM1</sup> (pink), USP37<sup>mUIM2</sup> (cyan) to USP37<sup>mUIM1,2</sup> (purple), USP37<sup>mUIM3</sup> (green) to USP37<sup>mUIM1,3</sup> (red) and USP37<sup>mUIM2,3</sup> (brown) to USP37<sup>mUIM1,2,3</sup> (blue)). These results are consistent with prior findings 88, which tested UIM disabling mutations in the single (USP37<sup>mUIM1</sup>, USP37<sup>mUIM2</sup>, USP37<sup>mUIM3</sup>) and triple combinations (USP37<sup>mUIM1,2,3</sup>) in human USP37 against K48- and K63-linked poly-Ub chains and found no difference in cleavage activity between human USP37<sup>wt</sup> and USP37<sup>mUIM1</sup> towards K48- and K63-linked poly-Ub chains.

Both UIM2 and UIM3 appeared important for the ability of USP37 to cleave all Ub chain types but to varying degrees. In the case of K11- and K63-linked substrates, individually disabling UIM2 (USP37<sup>mUIM2</sup>) or UIM3 (USP37<sup>mUIM3</sup>) caused partial impairment of activity, while simultaneously disabling UIM2 and UIM3 (USP37<sup>mUIM2,3</sup>) resulted in an additive impairment of activity. These results are consistent with prior findings 88, which showed that human USP37 with disabling mutations in all three UIMs had a larger decrease in activity towards K48- and K63-linked poly-Ub chains as compared to USP37 with disabling mutations in UIM2 or UIM3 alone.

In the case of the K48-linked Ub substrate, UIM3 was clearly more vital than UIM2 for the activity of USP37 (Figure 5-3B). Disabling UIM3 alone (USP37<sup>mUIM3</sup>) resulted in a large decrease in activity, while disabling UIM2 alone (USP37<sup>mUIM2</sup>) had no observable effect on
activity. However, the importance of UIM2 was revealed when UIM2 and UIM3 were disabled together, resulting in an enzyme (USP37\textsuperscript{mUIM2,3}) in which activity was almost completely abolished relative to WT (~4% activity remaining based on V0). Such a drastic effect was not observed for the same double mutant when cleaving K11- or K63-linked chains, where USP37\textsuperscript{mUIM2,3} retained ~20% or ~18% of WT activity (based on V0), respectively. These results suggested that UIM1 does not contribute to the activity of USP37 towards K11-, K48- and K63-linked chains, whereas UIM2 and UIM3 do contribute to activity towards these chains, but to varying degrees that are dependent on the specific type of chain linkage. As the disabling mutations in the UIMs had the strongest effects on the ability of USP37 to cleave K48-linked Ub chains, I chose to focus subsequent functional characterizations on this substrate type.

5.3.4 Interrogation of proximal or distal Ub binding by the UIMs of USP37

To determine if the UIMs of USP37 specifically engage the proximal Ub (primary amine donating) or the distal Ub (carboxy terminus donating) in a K48 di-Ub chain substrate to exert their effect on cleavage activity, I first made use of the model substrate Ub-AMC in which the proximal Ub moiety is replaced with a cleavable fluorescent molecule (AMC, 7-Amino-4-Methylcoumarin) (Figure 5-6A).
Figure 5-6. USP37<sup>wt</sup> and USP37<sup>mUIM1,2,3</sup> activity towards Ub-AMC.

A) Cartoon models of di-Ub and Ub-AMC illustrating how di-Ub contains both a distal and proximal Ub, while Ub-AMC contains only a distal Ub. B) Cartoon model of the two possibilities of how the UIMs of USP37 can
interact with Ub-AMC and the expected effects of each on Ub-AMC hydrolysis. The cartoon depicts a model in which the UIMs bind to the same di-Ub chain displayed across the enzyme active site. C) Progress curve for USP37\textsuperscript{wt} and USP37\textsubscript{UIM1,2,3} activity against Ub-AMC. Assays were performed in duplicate in three independent experiments and representative curves are shown. Enzymes were used at a concentration of 0.5 nM.
The proximal Ub is absent in Ub-AMC, and thus, I posited that if the UIMs act on the proximal Ub, I would observe no difference in cleavage by USP37\textsuperscript{wt} or USP37\textsuperscript{mUIM1,2,3} (Figure 5-6B). In contrast, if the UIMs act by binding to the distal Ub moiety of a di-Ub substrate, then I would expect a major difference in the cleavage of Ub-AMC by USP37\textsuperscript{wt} or USP37\textsuperscript{mUIM1,2,3}, as the distal Ub is fully available for binding (Figure 5-6B). In comparing the activity of USP37\textsuperscript{wt} and USP37\textsuperscript{mUIM1,2,3} against Ub-AMC, I observed no difference in DUB activity (Figure 5-6C). This result is consistent with the UIMs of USP37 acting by engaging the proximal Ub (Figure 5-6C).

To further verify that the UIMs of USP37 act by engaging the proximal Ub of K48 di-Ub, I used a mutant form of K48-linked di-Ub chains in which substitutions were incorporated into the proximal Ub moiety at position 44 (Ile44Ala or Ile44Trp). The Ile44 side chain of Ub comprises part of the canonical hydrophobic binding surface for UIMs, and substitutions at this position are likely to perturb UIM binding. If the UIMs of USP37 bind to the proximal Ub of a di-Ub substrate, then di-Ub proteins with substitutions at position 44 of the proximal Ub should be poorer substrates relative to WT di-Ub, in the case of the USP37\textsuperscript{wt} but not in the case of USP37\textsuperscript{mUIM1,2,3} in which all UIMs are disabled (Figure 5-7A). Indeed this is what I observed. Compared with cleavage of WT K48-linked di-Ub, USP37\textsuperscript{wt} displayed a reduced ability to cleave di-Ub substrates containing an I44A/W substitution (Figure 5-7B top panel). In contrast, USP37\textsuperscript{mUIM1,2,3} displayed similar activity for cleavage of all three substrates, albeit at reduced rates relative to the WT enzyme and substrate forms (Figure 5-7B, bottom panel). I also tested the three single site UIM mutants of USP37 for their ability to cleave K48 di-Ub substrates bearing the I44A/W substitution in the proximal Ub position. All three single site mutant proteins displayed a reduced ability to cleave the mutant substrates relative to wild type K48 di-Ub, similar to the behavior of USP37\textsuperscript{wt}. This expected result is consistent with a degree of functional redundancy between UIMs (Figure 5-8). Together, these results indicate that the UIMs of USP37 act to support enzymatic activity by selectively engaging the proximal Ub of K48 di-Ub.
Figure 5-7. USP37wt and USP37mUIM1,2,3 activity towards proximally mutated K48 di-Ub.

A) Cartoon model for how USP37wt and USP37mUIM1,2,3 are expected to cleave K48 and proximally mutated K48 chains. The cartoon depicts a model in which the UIMs bind to the same di-Ub chain displayed across the enzyme active site. B) Deubiquitination time course for USP37wt (top panel) and USP37mUIM1,2,3 (bottom panels) towards K48, K48I44A, and K48I44W di-Ub. Assay was performed in three independent experiments and representative gels are shown. USP37 enzymes were used at a concentration of 5 nM.
Figure 5-8. Activity of USP37<sup>wt</sup> and USP37 mutants towards proximally mutated K48 di-Ub.

Deubiquitination time course for USP37<sup>wt</sup>, USP37<sup>mUIM1</sup>, USP37<sup>mUIM2</sup>, USP37<sup>mUIM3</sup>, and USP37<sup>mUIM1,2,3</sup> towards K48, K48<sup>I44A</sup>, and K48<sup>I44W</sup> di-Ub. Assay was performed in three independent experiments and representative gels are shown. USP37 enzymes were used at a concentration of 5 nM. The data for USP37<sup>wt</sup> and USP37<sup>mUIM1,2,3</sup> was reproduced here from Figure 5-7 for clarity.
5.3.5 Interrogation of UIM function with Ub Variants

Ub variants (UbVs) are engineered forms of Ub that bind specifically and tightly to a target protein of interest 89. A UbV usually binds with higher affinity to preexisting Ub-binding site in a manner similar to the native interaction89,90,133,172. UbV reagents have been used previously to interrogate the mechanism of action and biological function of Ub-related enzymes both in vitro and in vivo 89-93,172-175.

To further verify my model that UIM2 and UIM3 act by selectively binding to the proximal Ub of K48 di-Ub, Joan Teyra and Gianluca Veggiani of the Sidhu lab used phage-display to derive one UbV that bound to the core catalytic domain lacking the insert containing the UIMs (UbV.core), another UbV that bound to the first UIM domain of USP37 (UbV.UIM1), and a third UbV (UbV.UIM*) that bound all three the UIM domains of the enzyme (Figure 5-9). I assessed the binding specificity of each UbV using GST/MBP pulldown experiments with the UbVs as prey and each isolated UIM (His-MBP-UIM1, His-MBP-UIM2, His-MBP-UIM3), the catalytic domain (GST-USP37) or the catalytic domain lacking all three UIMs (GST-USP37ΔIIM3) as bait (Figure 5-10). I also assessed the binding of the three UbVs with human USP37 and observed coincidentally that UbV.UIM* and UbV.UIM1 but not UbV.core were retained by GST-hUSP37 in pulldowns (Figure 5-11).
Figure 5-9. Sequence alignment of UbVs.

Regions that were mutated in the phage-displayed libraries are shown and *dashes* denote the WT sequence.
Figure 5-10. Binding specificity of UbVs.

UbV.core, UbV.UIM1 and UbV.UIM* were incubated with the indicated bait proteins. After washing, resulting fraction still bound to beads was analyzed by SDS-PAGE and Coomassie blue staining.
Figure 5-11. UbV binding ability to human USP37.

UbV.core, UbV.UIM1 and UbV.UIM* were incubated with human USP37 fused at the N-terminus to GST (GST-hsUSP37) and GST. After washing, resulting fraction still bound to beads was analyzed by SDS-PAGE and Coomassie blue staining.
I next determined the ability of each of the three UbVs to inhibit hydrolysis of Ub-AMC by zebrafish USP37<sup>wt</sup>. Mutation of UIM1 had no effect on enzyme function, and as expected, UbV.UIM1 did not affect USP37<sup>wt</sup> activity (Figure 5-12A). In my working model, UIM2 and UIM3 engage the proximal Ub moiety of K48 di-Ub to influence catalytic activity. As Ub-AMC lacks a proximal Ub, I was not surprised to see that UbV.UIM* had little effect on hydrolysis of this substrate (Figure 5-12A). In contrast UbV.core potently inhibited Ub-AMC cleavage (IC<sub>50</sub> ~ 9 nM). This result is consistent with the likelihood that UbV.core directly competed for binding to the distal Ub-binding site of USP37 (Figure 5-12A).
Figure 5-12. Analysis of inhibitory effects of UbVs on USP37 activity.

A) Inhibition of USP37\textsuperscript{wt} cleavage of Ub-AMC by UbV.core, UbV.UIM1, and UbV.UIM*. Curves represents measurements from three independent experiments. Values reported are mean ± SD. B) Inhibition of USP37\textsuperscript{wt} cleavage of IQF K48-linked di-Ub by UbV.core, UbV.UIM1, and UbV.UIM*. Curves represents measurements from three independent experiments. Values reported are mean ± SD. For both experiments USP37 was used at a concentration of 0.5 nM.
I next determined the ability of each of the three UbVs to inhibit the activity of zebrafish USP37 for cleavage of the K48-linked di-Ub substrate bearing an internally quenched fluorophore (IQF). UbV.UIM1 had no effect on this activity (Figure 5-12B), consistent with my mutational data for UIM1 (Figure 5-3B). In contrast, UbV.UIM* potently inhibited cleavage of the K48-linked di-Ub substrate, as did UbV.core (Figure 5-12B). Taken together, these results support my model in which UIM2 and UIM3 engage the proximal Ub of K48 di-Ub to enhance cleavage by USP37, whereas UIM1 is not involved in substrate recognition.

5.4 Discussion

My results are consistent with a model whereby UIM2 and UIM3 of USP37 bind the proximal Ub of K48 di-Ub to increase catalytic efficiency. In the case of K48-linked substrates, binding of the UIMs to the proximal Ub increased catalytic efficiency of USP37 largely through $k_{cat}$ and to a lesser degree through $K_M$. This result was somewhat counter-intuitive since I expected that binding of the UIMs of USP37 to Ub to have a more pronounced effect on $K_M$ (a proxy for binding affinity) since others have shown that UIM mutations perturbed the ability to pull down Ub conjugates 88. Furthermore, I observed that in the case of K48-linked di-Ub substrate, UIM3 was more vital than UIM2 for the proteolytic activity of USP37. This behavior was not noted in the same previous study 88. A possible explanation for my findings relates to the main type of Ub conjugates investigated, which involved chains of variable length greater than 2 moieties 88 versus Ub chains of fixed length equal to 2 Ub moieties used here. Characterizing the kinetic parameters of cleavage of longer Ub chains by USP37 could help to address this point. The issue of how the binding of the UIMs of USP37 to the proximal Ub of K48 di-Ub enhances activity remains an open question. I envision a model whereby the UIMs optimally position a di-Ub chain across the active site of the USP domain for more efficient catalysis of the inter-Ub isopeptide bond by selectively binding the proximal Ub position. I expect the UIMs to act upon the same substrate engaged by the USP domain active site (i.e. in cis) as zebrafish USP37 displays no tendency to multimerize in solution as assessed by size exclusion chromatography (Figure 5-13). Whether this model is generalizable to other Ub chain types, requires further verification. Furthermore, as my experiments were largely focused
on di-Ub substrates, it is possible that the UIMs could function in additional capacities for longer chains types. A co-structure of USP37 with a Ub chain substrate would be highly informative for resolving the remaining mechanistic questions.
**Figure 5-13.** USP37\textsuperscript{wt} is a monomeric protein.

Size exclusion chromatography elution profile for USP37\textsuperscript{wt}.
Joan and Gianluca developed UbVs that bound the core catalytic domain or the UIMs of USP37 and functioned as potent inhibitors of catalytic activity in vitro. The UbVs targeting the UIMs represent the first report of UbV DUB inhibitors that act by targeting domains distinct from the catalytic domain. Notably, UbVs have been shown to inhibit DUBs in cells, and we posit that the UbVs developed here may prove useful as reagents for future studies interrogating the in vitro biochemical functions and in vivo biological functions of USP37.

While inhibitors of USP enzymes have been developed, specificity has been a recurring problem. This may be due to the lack of differentiating features between the catalytic sites of USP family members. Only recently have groups demonstrated effective specific inhibitors against a USP enzyme, namely USP7. Interestingly, one of these inhibitors did not target the conserved active site directly, but rather, acted by targeting a USP7 loop distinct from other USP family members. In an analogous manner, we exploited the unique UIMs of USP37 to derive a specific inhibitor of USP37.

As noted, USP37 stabilizes the proto-oncogene c-MYC and the oncogenic fusion PLZF/RARA in cells. The demonstration that knockdown of USP37 decreases the stability of these oncogenic proteins, raises the possibility that specific and potent inhibitors of USP37 may provide a viable therapeutic strategy to treat human cancers. To this end, targeting UIM2 and UIM3 may provide the basis for potent and specific inhibition of USP37 DUB activity.
Chapter 6: Future Perspectives

6.1 Overview

The research presented in this thesis has advanced (1) our understanding of how UbVs target their binding partners and (2) our understanding of how DUBs use accessory domains to augment cleavage mechanisms. In chapters two, three, and four I revealed using X-ray crystallography and functional studies how UbVs recognize substrates with high specificity and affinity. The work in these three chapters serves as a basis for how to improve the engineering of UbV libraries geared towards binding to UIMs and Skp1-F-box complexes. The most interesting result from my work in these chapters was the finding, in chapter 3, that UbV.v27.1 binds its target UIM in a dimeric fashion. This UbV, along with two other dimeric UbVs engineered to date by our collaborators in the Sidhu lab, suggests the possibility of a general way to engineer dimeric UbVs. How to develop a UbV library with a propensity to dimerize is the focus of my first future direction. In chapter five I revealed, using biochemical experiments, how the UIMs of USP37 function to influence the enzyme’s cleavage mechanism. However to understand on a molecular level how the UIMs function in detail, a structure of USP37 bound to a Ub substrate is required. Methods to obtain such a structure are the focus of my second future direction.

6.2 Developing a Dimeric UbV Library

Dimeric UbVs have two unique characteristics compared to their monomeric counterparts, which makes developing a dimeric UbV library a worthwhile pursuit. First, dimeric UbVs can provide a larger interaction interface than monomeric UbVs, which can in principal be exploited for tighter binding and higher specificity. Second, by virtue of the UbV being dimeric it can in principal dimerize its binding target. As some Ub binding proteins are regulated by dimerization, we could in theory modulate these proteins through the use of dimeric UbVs. For example, the E3 ligase activity of Cbl-b is activated by dimerization. This protein also contains a UBA domain, which functions by binding Ub. Thus, a dimeric UbV raised against the UBA domain of Cbl-b could potentially provide a means to activate the E3 ligase activity of Cbl-b (Figure 6-1). In the following, I propose two methods to generate a UbV library with a propensity to form dimers.
Figure 6-1. Dimeric UbV activation of the E3 ligase Cbl-b.

The E3 ligase Cbl-b contains a RING domain (for Ub chain formation) and a UBA domain (for Ub binding). Upon dimerization the E3 ligase activity of Cbl-b is activated. A dimeric UbV that binds the UBA domain of Cbl-b could force the enzyme into a dimeric state and activate the enzyme activity.
I would create a UbV library in which the background sequence used for randomization contains and holds fixed a set of substitutions that drive UbV dimerization. To define this set of substitutions, I would first determine what residues are required for dimerization of the three known dimeric UbVs. In chapter 3, I identified 4 substitutions that were not required for dimerization (Leu8I, Thr66Asn, His68Tyr, and Ser72Arg) and identified 1 substitution that was required for dimerization (Gly10Val) of UbV.v27.1. I would next test the remaining substitutions in UbV.v27.1 for their necessity in dimerization by using the mutagenesis and size-exclusion chromatography strategy employed in chapter 3. After defining which substitutions are required for UbV.v27.1 dimerization, I would test if introduction of this set of substitutions into Ub is sufficient to drive oligomerization. If this strategy identifies a minimal set of dimer-driving substitutions, I would then create a UbV library where the background sequence for randomization contains this set of dimer driving substitutions. To increase the odds of defining a set of dimer driving substitutions, in parallel, I would perform the same procedure for UbV.XR and UbV.15.D, the other known dimeric UbVs. One pitfall of this strategy is that the set of substitutions required for dimerization may not be the same as the set of substitutions sufficient for dimerization. Thus figuring out which substitutions are sufficient for dimerization might require a great deal of sampling, for example UbV.v27.1 contains 18 substitutions.

For the second method, I would create a UbV library that randomizes the typical binding positions to all amino acids but randomizes the Gly10 position only to amino acids known to be required for dimerization. Currently, we know two substitutions at the Gly10 position, namely valine and alanine, which are required for dimerizing the UbVs UbV.v27.1 and UbV.XR, respectively. A third substitution at Gly10 that may be required for dimerization is phenylalanine, which is present at the Gly10 position in UbV.15.D. However the dependency of UbV.15.D dimerization on Phe10 needs to be tested first. If confirmed, I would randomize Gly10 between alanine, valine, and phenylalanine. Overall this method would be simpler and more tractable to execute than the first method described. However, a caveat of this method is that it only biases the library to dimeric UbVs and does not exclusively produce dimeric UbVs. This is because UbVs in this library will be guaranteed to contain a substitution at the Gly10 position that is required for dimerization but not guaranteed to contain substitutions sufficient to drive dimerization. To ensure that it is at least possible for UbVs in this library to contain a set of substitution sufficient to drive dimerization, I would randomize all positions, except Gly10, that
were randomized in the libraries engineered to produce the dimeric UbVs UbV.v27.1, UbV.XR, and UbV.15.D. Thus ensuring the possibility of minimally re-discovering related dimeric UbVs. To quickly ascertain if generated UbVs are monomeric or a dimeric a pull-down assay could be employed. With the libraries I have detailed here, dimeric UbVs would be readily produced allowing for improved modulation of the Ub system.

6.3 Crystallizing a USP37-Ub chain Complex

The work presented in chapter 5 shows that UIM2 and UIM3 engage the proximal Ub of a di-Ub substrate to increase USP37 cleavage activity. However to answer the questions of (1) how do UIM2 and UIM3 functionally complement one another and (2) how do the UIMs orient the Ub chain to increase catalytic efficiency, a crystal structure of USP37 bound to a Ub chain is required.

During my PhD, I attempted to crystallize a catalytically dead (C331A) version of the zebrafish USP37 DUB domain with K48 di-Ub. However, the strategy did not lead to any crystal hits. I recognized that one possible hindrance to USP37 crystallization was the UIM insertion itself, as it is partially unstructured based on a trypsin digest I performed. Unstructured regions are known to hinder protein crystallization because the conformational heterogeneity and flexibility can obstruct crystal packing. To address this potential problem, I would create USP37 constructs with the unstructured regions removed. To date I have created three constructs of the zebrafish USP37 DUB domain that have progressively larger regions of the insertion removed (regions that I predict to be unstructured and nonfunctional) (Figure 6-2). All three constructs maintained cleavage activity similar to wild-type USP37 and thus were good candidates to model the wild-type USP37 (Figure 6-2). I attempted to crystallize the catalytically dead versions of these 3 loop deletion constructs with K48 di-Ub. Despite this revised strategy no crystal hits were obtained.
Figure 6-2. USP37 loop deletion constructs.

(A) Schematic cartoon of USP37 and the three loop deletion USP37 constructs. (B) Progress curves of USP37 and the three loop deletion USP37 constructs towards IQF K48 di-Ub. Reaction conditions are the same as used in Figure 5-3B K48 panel.
One additional parameter that may hinder USP37 crystallization is the functional redundancy of UIM2 and UIM3. The biochemical data presented in chapter 5 showed that both UIM2 and UIM3 can bind the proximal Ub to increase catalytic activity. This suggests that the USP37-K48 di-Ub complex used to screen crystallization conditions could be heterogeneous, in terms of one population of USP37 having UIM2 bound to the proximal Ub and the other population of USP37 having UIM3 bound to the proximal Ub (Figure 6-3). As stated before conformational heterogeneity can perturb crystal formation. To resolve this potential problem, I would disable one of UIM2 or UIM3 (with the UIM mutations used in chapter 5) in the crystallization constructs (Figure 6-3). To further increase the chances of crystallization, I would also perform this strategy with the three loop deletion constructs and also use K11 and K63 di-Ub, USP37’s two other most preferred di-Ub substrates.
USP37<sup>wt</sup>

**UIM2 or UIM3 can bind the proximal Ub**

USP37<sup>mUIM2</sup>

**Only UIM3 can bind the proximal Ub**

USP37<sup>mUIM3</sup>

**Only UIM2 can bind the proximal Ub**

**x** UIM disabling mutations

*Figure 6-3. Functional redundancy between UIM2 and UIM3 can cause heterogeneity during crystallization trials.*

USP37<sup>wt</sup> can have either UIM2 or UIM3 bound to the proximal Ub. In contrast, USP37<sup>mUIM2</sup> or USP37<sup>mUIM3</sup> can only have UIM3 or UIM2, respectively, bound to the proximal Ub thereby removing heterogeneity.
Another approach to obtain a crystal structure of USP37 in action would be through the use of UbVs. As stated in Chapter 1, UbVs can aid structure determination in two ways. First, the tight binding of UbVs to their targets can increase the stability of the complex and thereby aid crystallization. Second, structures of Ub binding proteins bound to UbVs can provide insights into the molecular mechanism of action of the respective protein.

For a UbV/USP37 complex to model a di-Ub/USP37 complex I would need to identify both a UbV that binds to the core DUB domain and a UbV that binds to UIM2 or UIM3 (Figure 6-4). Recall from Chapter 5 that Joan Teyra in the Sidhu lab engineered a UbV that targets the DUB domain (UbV.core) that would meet my needs. Also Gianluca Veggiani in the Sidhu lab generated a UbV, namely UbV.UIM*, that targets UIM2 and UIM3. However, UbV.UIM* would not be an ideal choice for use in crystallization for two reasons. First, UbV.UIM* binds to all three UIMs in USP37 as seen in my GST-pulldown in chapter 5 and thus heterogeneity within the USP37-UbV.UIM* complex would hinder crystallization. Second, UbV.UIM* may be a dimeric UbV as it contains the Gly10Val substitution that is required for UbV.v27.1 dimerization and also elutes as an oligomer on a size exclusion chromatography column. While each half of a dimeric UbV mimics the fold of Ub, dimeric UbVs would not make for an ideal mimic of the proximal Ub in a di-Ub chain due to their larger size and shape. For these reasons, I propose engineering new UbVs for UIM2 and UIM3 that are both specific for their intended UIM and monomeric. To this end, I would make a modified version of the recently developed UIM UbV library developed by Gianluca, that was used to generate UbV.UIM*. In its present form this UIM UbV library randomizes Gly10, which I have shown is a pivotal position for dimerization. To prevent engineering dimeric UbVs, I would create a modified version of the UIM UbV library where the Gly10 position is not randomized. Using this modified library I would screen zebrafish UIM2 and zebrafish UIM3. I would assess any UbV hits for their UIM specificity by the GST-pulldown assay performed in Chapter 5 using UIM1, UIM2, and UIM3 as my test panel. If this screen produced no UbVs that were specific for their intended UIM, then I would screen UIM2 and UIM3 using the other available UbV libraries.
USP37 bound to a UbV that targets the USP core (i.e. distal binding site) and a UbV that targets UIM2 or UIM3 (i.e. UIM binding site) may mimic USP37 bound to a di-Ub substrate.
As the UbVs are not linked together, it is possible that the UbVs that bind to the distal and proximal Ub binding sites of USP37 would not spatially mimic the actual di-Ub substrate interaction. To increase the chances that the two UbV’s bind in a manner similar to di-Ub, I would also try to enzymatically link the proximal and distal UbVs together through a K11, K48, or K63 isopeptide linkage (Figure 6-5). I would choose these three linkage types as they are the preferred substrates of zebrafish USP37. Also as the protocols for the enzymatic synthesis of these types of chains are well-defined and I have experience producing them. Before attempting an enzymatic synthesis of a di-UbV, first some minimal sequences must be added to the UbVs. First Ub contains a di-Gly at its C-terminus, a feature that is required for efficient Ub conjugation of the distal Ub to the proximal Ub (Pickart et al 1994). In its present form UbV.core (the distal UbV) does not contain a Gly-Gly C-terminus. As such, I would mutate the C-terminus of UbV.core to include the di-Gly motif. I would ensure that this modification does not abrogate UbV.core binding to USP37 by measuring the interaction of the two proteins via a GST-pulldown. For K11, K48, or K63 Ub chain synthesis a lysine is required on the proximal Ub at positions 11, 48, and 63, respectively. Any of the UbVs for UIM2 and UIM3 engineered in the previous section that do not contain a lysine at these positions would be mutated to include them. Again I would ensure that the modifications did not abrogate UbV binding to USP37 by assessing the interaction between UbV and USP37 via a GST-pulldown. Once UbV.core and the UbVs for UIM2 and UIM3 contain the required sequences, I would set up small-scale di-Ub conjugation reactions for all possible combinations of UbVs and linkage types using the E2s Ube2s (K11 specific E2), Ube2k (K48 specific E2), and Ubc13/MMS2 (K63 specific E2 complex). Trying all possible combinations is necessary as the substitutions present in some UbVs may prohibit them from working with the E1 and E2 enzymes. If any reactions are capable of producing a di-UbV, I would then upscale, purify the di-UbV, test its ability to inhibit USP37 cleavage of its corresponding di-Ub, and compare this ability to the untethered UbVs. Any di-UbV that inhibited cleavage of its corresponding di-Ub better than the untethered UbVs would be assumed to mimic di-Ub binding and thus would be screened with USP37 for crystallization. Together with the work presented in chapter 5, a crystal structure of USP37 with Ub substrate would provide the complete picture of how the UIMs of USP37 influence catalytic activity.
Unlinked UbVs may bind to the proximal and distal Ub binding sites of USP37 in a manner similar to how di-Ub binds USP37 or the unlinked UbVs may bind in a manner that does not mimic the way di-Ub binds USP37. Whereas, linking the two UbVs with an isopeptide linkage (ex. K11, K48, or K63) to form a di-UbV may increase the likelihood of the UbV binding mode being similar to a di-Ub substrate.
References


