# The Role of KLHDC2 in Recognizing Diglycine C-end Degron and its Therapeutic Potential

Domnita Valeria Rusnac

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Ning Zheng, Chair

Rachel E. Klevit

Richard G. Gardner

Program Authorized to Offer Degree:

Pharmacology

© Copyright 2019

Domnita Valeria Rusnac

University of Washington

## Abstract

The Role of KLHDC2 in Recognizing Diglycine C-end Degron and its Therapeutic Potential

Domnita Valeria Rusnac

Chair of the Supervisory Committee:
Ning Zheng, PhD
Pharmacology

Ubiquitin-proteasome system is crucial for controlling essential cellular processes. Eukaryotic cells conjugate the eight kDa ubiquitin polypeptide to a variety of proteins to regulate their functions or turnover. Modification of substrate proteins by ubiquitin is promoted by a three-enzyme cascade consisting of the ubiquitin activating E1 enzyme, ubiquitin conjugating E2 enzyme, and ubiquitin E3 ligase. At the final step of ubiquitin transfer, many ubiquitin E3 ligases recognize specific substrates and promote their poly-ubiquitination, which often results in proteasomal degradation. Among the three known classes of E3s, cullin-RING ubiquitin ligases (CRLs) are the largest family of multi-subunit E3s. With a modular assembly, CRLs are capable of recruiting a myriad of substrates onto a common catalytic scaffold by employing interchangeable substrate receptor subunits. Engagement between a specific substrate and its

cognate receptor is frequently mediated by a short linear amino acid sequence in the target protein, commonly known as a degron. Despite years of progress, there is still a major gap in our understanding of degron recognition by CRLs. Recent studies have identified a new class of degrons, which are localized at the C-terminal end of a panel of abnormal polypeptides. These so-called C-end degrons are specifically recognized by a subfamily of CRL2s with previously unknown functions. In my Ph.D. thesis studies, I determined three crystal structures of a novel CRL2 substrate receptor, KLHDC2, in complexes with the diglycine-ending C-end degrons of two early terminated selenoproteins and the N-terminal proteolytic fragment of USP1. KLHDC2 recognizes these C-end degron peptides with a similar coiled conformation and cradles their common C-terminal diglycine motif with a deep surface pocket. By hydrogen bonding with multiple backbone carbonyls of the peptides, KLHDC2 locks in the otherwise degenerate degrons with a compact interface and unexpected high affinities. My results reveal the structural mechanism by which KLHDC2 recognizes the simplest C-end degron and suggest a functional necessity of the E3 to tightly maintain the low abundance of its select substrates. With a virtual screen approach, I have also established KLHDC2 as a potential new platform for targeted protein degradation, which represents a new strategy in therapeutic development.

# TABLE OF CONTENTS


**CHAPTER 1. UBIQUITIN PROTEASOME SYSTEM** **1**

1.1 Introduction 1
1.2 Ubiquitin and Ubiquitin-Like Proteins 2
1.3 Proteasome as a Protein Degradation Machinery 3
1.4 The E1-E2-E3 Enzyme Cascade 4
1.5 Three Types of Ubiquitin E3 Ligases 6
1.6 Deubiquitinases 9

**CHAPTER 2. CRL UBIQUITIN LIGASES** **11**

2.1 Overall CRL Architecture 11
2.2 CRL1 Adaptor and Substrate Receptors 13
2.3 CRL2 and CRL5 16
2.4 CRL3 19
2.5 CRL4 21
2.6 Synopsis of Substrate Recognition by CRLs 25
2.7 PTM-Dependent Substrate Degron Recognition 25
2.8 Native Substrate Degron Recognition 28
2.9 Globular Substrate Protein Recognition 29
2.10 Synopsis of Regulation of CRLs by NEDD8 Modification 31
2.11 NEDD8-Modified CRLs 32
2.12 CAND1 and Cullin Cycle 35
2.13 COP9 Signalosome-CRL Interactions 38

**CHAPTER 3. CHARACTERIZATION OF KLHDC2-C-END DEGRON** **42**

3.1 Introduction 42
3.2 Mapping Key Elements in SelK C-End Degron 45
3.3 Crystal Structure of KLHDC2 Bound to SelK Degron Peptide 49
3.4 Mutational Analysis of the KLHDC2 Degron-Binding Pocket 53
3.5 Diglycine C-End Degrons from Other Substrates 56
3.6 Affinity Requirement for Substrate Degradation 58
3.7 Discussion and Concussions 61
3.8 Method Details 63
*3.8.1 Experimental Model and Subject Details* *63*
*3.8.2 Molecular Biology and Protein Purification* *63*
*3.8.3 Protein Crystallization* *64*
*3.8.4 Data Collection and Structure Determination* *65*
*3.8.5 AlphaScreen Luminescence Proximity Assay* *65*
*3.8.6 Octet BioLayer Interferometry Measurement* *67*
*3.8.7 Global Protein Stability Assay* *67*
*3.8.8 Protein Native Mass Spectrometry* *68*
*3.8.9 Affinity Pull-Down Assay* *69*

**CHAPTER 4. SMALL MOLECULE MEDIATED PROTEIN DEGRADATION** **74**

4.1 Synopsis of Targeted Protein Degradation 74
4.2 Molecular Glue 75
4.3 Proteolysis Targeting Chimeric Molecules 77
4.4 Insights Into Molecular Glue and PROTAC 81
4.5 KLHDC2 - a New Platform for Targeted Protein Degradation 83

4.6 METHOD DETAILS 90
*4.6.1 Experimental Model and Subject Details 90*
*4.6.2 Molecular Biology and Protein Purification 91*
*4.6.3 Protein Crystallization 91*
*4.6.4 Data Collection and Structure Determination 92*
*4.6.5 AlphaScreen Luminescence Proximity Assay 92*

# LIST OF FIGURES


Figure 1. Structure of ubiquitin. 2
Figure 2. Structure of tetra-ubiquitin. 2
Figure 3. Structure of 26S proteasome. 3
Figure 4. Schematic representation of the ubiquitin-proteasome system. 5
Figure 5. Schematic drawing of the three classes of E3s. 7
Figure 6. Structure of RING E3-Ub~E2 complex. 7
Figure 7. Structure of HECT E3-Ub~E2 complex. 8
Figure 8. Structure of RBR E3. 9
Figure 9. Schematic drawing of all CRLs. 11
Figure 10. Structural model of an CRL1 in complex with substrate and ubiquitin-E2. 11
Figure 11. Structure of p27-CKS1-SKP2-SKP1 complex. 14
Figure 12. Structural representation of β-catenin-βTrCP-SKP1 complex. 15
Figure 13. Structural representation of VHL-EB-EC complex. 17
Figure 14. Structure of VHL-EB-EC-CUL2 complex. 17
Figure 15. Structural representation of Vif-CBF-β-EB-EC-CUL5 complex. 18
Figure 16. Structure basis for degron recognition by dimeric SPOP. 20
Figure 17. Structure of dimeric $SPOP^{BTB}$-$CUL3^{NTD}$ complex. 20
Figure 18. Structural representation of DDB1. 22
Figure 19. Complex structure of CUL4A-RBX1-DDB1-SV5-V. 22
Figure 20. Structural mechanism of DDB1 and DDB2 interaction. 23
Figure 21. Structural basis of Cyclin E degrons to FBXW7. 26
Figure 22. Structure basis of VHL binding HIF1α degron. 27
Figure 23. Structural mechanisms of NRF2 degrons binding to KEAP1. 28
Figure 24. FAD binding to human CRY2 and the complex structure of SKP1-FBXL3-CRY2. 31
Figure 25. Structure basis of RBX1 binding to CUL1-CTD. 32
Figure 26. Two structures of RBX1 in the presence of NEDD8~CUL5-CTD. 33
Figure 27. Complex structure of CUL1-CTD-RBX1 bound to NEDD8-charged UBC12 and DCN1. 33
Figure 28. Yet another CUL1-CTD-RBX1 complex structure. 34
Figure 29. Structure basis of CAND1-CUL1-RBX1 interaction. 36
Figure 30. Structural representation of GLMN-CUL1-CTD-RBX1 complex. 37
Figure 31. The overall architecture of COP9 signalosome. 39
Figure 32. Diagram of CRL4-DDB2-COP9 complex. 40
Figure 33. GST-Pull down confirming direct interaction between KLHDC2-SelK degron. 45
Figure 34. Schematic representation of AlphaScreen-based competition assay. 46
Figure 35. Functional mapping of the SelK C-end degron. 46
Figure 36. Validation of high affinity binding between KLHDC2 and SelK 12 aa peptide via BLI. 47
Figure 37. Validation of the importance of carboxyl group and diglycine motif in degron binding. 48
Figure 38. Structure basis of SelK degron recognition by KLHDC2. 48
Figure 39. Conservation surface mapping of the KLHDC2. 49
Figure 40. Electrostatic surface potential map of KLHDC2. 50
Figure 41. Stereo view of the KLHDC2 kelch repeat domain pocket with a SelK C-end degron bound. 50
Figure 42. A stereo close-up view of the interface formed between KLHDC2 and the SelK peptide. 51
Figure 43. Ligplot diagram of the interactions between KLHDC2 and the SelK C-end degron peptide. 51
Figure 44. GST-pull down between GST-SelK 8 aa peptide and KLHDC2 mutants. 53
Figure 45. Schematic representation of the experimental design for the Global Protein Stability assay. 53
Figure 46. Assessing exogenous KLHDC2 levels using Western blot analysis. 54
Figure 47. Turnover of GFP-fused SelK or USP1-NTD monitored by GPS. 55
Figure 48. Determination of the $IC_{50}$ values for SelK, SelS and USP1 C-end degrons to KLHDC2. 56
Figure 49. Binding more of three C-end degrons to KLHDC2. 57
Figure 50. Assessment of the effect of KLHDC2 mutation on binding SelK and USP1 degrons. 59
Figure 51. The effects of KLHDC2 mutations on the degradation of GFP fused SelK and USP degrons. 60

Figure S1. Protein native mass spectrometry analysis of the KLHDC2-SelK complex. 70
Figure S2. Sequence alignment of KLHDC2 orthologs. 71
Figure S3. Binding of KLHDC2 mutants with 8 aa SelK degron fused to GST. 72
Figure 52. Structural representation of IAA7-Auxin-IP6-TIR1-ASK1 complex. 75
Figure 53. Structural representation of CK1α-lenalidomide-CRBN and GSPT1-CC-885-CRBN complexes. 76
Figure 54. Structural representation of $BRD4^{BD2}$-MZ1-CRBN complex. 79
Figure 55. Structural representation of $BRD4^{BD1}$-dBET23-CRBN complex. 80
Figure 56. The structures of $BRD4^{BD1}$- dBET23-CRBN and $BRD4^{BD1}$-dBET57-CRBN. 81
Figure 57. Activity of top virtual-screen hits tested at 300 μM. 87
Figure 58. Activity of 2-(2-hydroxyphenyl) acetic acid. 87
Figure 59. Positive Fo-Fc density for a hit compound. 88
Figure 60. Structural mechanism of 2-(2-hydroxyphenyl) acetic acid binding to KLHDC2. 88
Figure 61. Activity of M01C, M01D, and M211. 89
Figure 62. Structural basis of M01C-KLHDC2 interaction. 90

# LIST OF TABLES

Table 1. Data Collection and Refinement Statistics. 73

# LIST OF ABBREVIATIONS

AlphaScreen: Amplified luminescence proximity homogenous assay
APOBEC3: Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3
ARIH1: Adriadne-1 homolog
ASK1: Apoptosis signal-regulating kinase 1
AUX/IAA: Auxin/indole-3-acetic acid
BPA: beta-Propeller A
BPB: beta-Propeller B
BPC: beta-Propeller C
BRD4: Bromodomain-containing protein 4
BTB: Broad-Complex, Tramtrack, and Bric a brac
β-TrCP: beta-Transducin repeat Containing Protein
CAND1: Cullin associate NEDD8-dissociated protein 1
CBF-β: Core binding factor beta
CDC34: Cell division cycle 34 / Ubiquitin-conjugating enzyme E3 R1
CDK2: Cyclin-dependent kinase 2
CK1α: Casein kinase 1
CKS1: Cyclin-dependent kinases regulatory subunit 1
COI1: Coronatine-insensitive protein 1
COP9: Constitutive photomorphogenesis 9
CRBN: Cereblon
CRL: Cullin-RING ubiquitin ligases
CRY1/2: Cryptochrome circadian regulator 1/2
Cryo-EM: Cryo-electron microscopy
CSN: COP9 signalosome
CTD: C-terminal domain
CTR: C-terminal region
CUL: Cullin
DCAFs: DDB1-CUL4A-associated factors
DCN1: Defective in cullin neddylation protein 1-like protein 1
DDB1/2: DNA damage-binding protein 1/2
DesCEND: Destruction via C-end degron
DGR: Double glycine repeats
DUBs: Deubiquitinases
EB: Elongin B
EC: Elongin C
FAD: Flavin adenine dinucleotide
FBXL: F-box and leucine rich repeat proteins
FBXO: F-Box and other repeat containing proteins
FBXW: F-Box and WD repeat containing proteins
GFP: Green fluorescent protein
GLMN: Glomulin
GPS: Global Protein Stability
GSPT1: G1 to S phase transition 1
GST: Glutathione S-transferase
HBD: Helix bundle domain
HBx: Hepatitis B virus X protein
HECT: Homology to E6AP carboxyl terminus
HHARI/ARIH1: Ariadne-1 homolog
HIF1α: Hypoxia inducible factor 1 subunit alpha
IKZF1/3: IKAROS family zinc finger 1/3
IMiDs: immunomodulatory drugs
IP6: inositol hexakisphosphate

IRES: Internal ribosome entry site
IVR: Intervening region
JA: Jasmonic acid
JAZ: Jasmonate ZIM domain
$K_d$: Dissociation constant
KEAP1: Kelch-like ECH-associated protein 1
KLHDC2: Kelch domain containing protein 2
LRR: Leucine-rich repeat
MATH: Meprin and TRAF-C homology
MEIS2: Myeloid ecotropic viral integration site 1 homolog 2
MPN: MPR1/PAD1 domain
NEDD8: Neural precursor cell expressed developmentally down-regulated protein 8
NRF2: Nuclear factor erythroid 2-related factor 2
NTD: N-terminal domain
ODD: Oxygen-dependent degradation
ORF: Open reading frames (ORF)
PCI: Proteasome lid-CSN-initiation factor 3
PROTAC: Proteolysis targeting chimeric molecules
PTM: Post-translational modification
RBX1/2: RING box protein 1/2
RFP: Red fluorescent protein
RING: Really interesting new genes
RING-IBR-RING: RING in-between-ring RING
SALL4: Sal-like protein 4
SCF: SKP1, CUL1, and F-box proteins
Sec: Selenocysteine (Sec),
SelK: Selenoprotein K
SelS: Selenoprotein S
shRNA: Small hairpin RNA
SKP1/2: S-phase kinase-associated protein 1/2
SPOP: Speckle-type POZ protein
STATs: Signal transducer and activator of transcription
SUMO: Small ubiquitin-like modifier
SV5-V: Simian virus 5 V protein
TIR1: Transport inhibitor response 1
Ub~E2s: Ubiquitin-charged E2s
UBC: Ubiquitin-conjugating catalytic
UBC12/UBE2M: Ubiquitin-conjugating enzyme E2 M
UBLs: Ubiquitin-like proteins
UBR1: Ubiquitin protein ligase E3 component N-Recognin 1
UPS: Ubiquitin-proteasome system
USP1: Ubiquitin specific peptidase 1
VHL: Von Hippel-Lindau tumor suppressor
Vif: Viral infectivity factor
WHB: Winged Helix B
WHx: Woodchuck hepatitis virus X protein
Wnt: Wingless

# ACKNOWLEDGEMENTS

I often stop and think about my life, about all the people who once were in it and the people that are in it today. Every person I have bumped into up to this point in time has affected me in some way, sometimes in a major way, other times in a minor way, which I might not even be aware off. I will thank the people that I perceive to have affected me the most as I would have not been able to earn a PhD without them.

I was fortunate enough to be raised by a family of highly educated and intelligent people, who valued learning, asking questions and hard work. My parents and sisters never doubted me or my ability to accomplish anything that I set my mind to, as long as I worked hard towards my goals. And how could I not be diligent, when I saw my mother raise three children, while cooking, cleaning and being a history professor at the university. From my mother, Larisa Rusnac, I have learned how to be confident, independent, and not shy about expressing myself. She has taught me how to be myself and to rebel against the status quo and to push boundaries.

From my dear father, Valeriu Rusnac, I have learned how to ask questions about the world around us. He trained me to think in an analytical way by encouraging me to do logic puzzles. He also exposed me to ethics and philosophy at an age where I was too young to understand what he was talking about. Nevertheless, his lessons stuck with me.

My parents instilled core values in me and my sisters helped me see what I could do with those values. When I was in high school my older sister, Silvia Rusnac, helped me realize that I could pursue a career as a scientist and do research. My middle sister, Ionela Rusnac, has not dissuaded me from exploring my curiosity, even though my inquisitiveness caused literal fires. As a child, I would strike matches but was too scared to hold them, tossing them on the floor. Ionela

would simply walk by and put out the little fires that started on the linoleum floor. Those were not the last fires she put out, as she helped me go through college debt free which allowed me to pursue a PhD. Ionela put out many of my fires, but she didn't put out the fire for learning that I grew inside of me.

At college I met many incredible people who helped me grow as a person: Windom Tilton – the sunshine of my life; Stephanie Driscoll – my partner in crime; Amanda Stovall (Peterman) – the strong and fierce boss lady; Gabe Ahl – my mentor; Klara Briknarova – my Undergraduate PI; Rose Lavelle – the super chill biker; Nick Pfeil – the friend who had an open door policy whether I needed a cry or a laugh; and Brian Skeels – one of my best friends for the last 9 years who drive four hours in the middle of the winter to take me to another airport after I missed my flight, and recently took me on an amazing trip to Vegas. Sami Naboulsi, who was my partner for many years, taught me so much. He helped me learn that people don't do things to hurt me, that people live their lives and I should not take it personally. He has helped me mature, grow up, and was by my side for almost five years of graduate school.

When starting my graduate studies at the University of Washington, I was terrified about meeting my cohort, and now I am terrified of not seeing them all the time. I was incredibly lucky to have a group of amazing people to share the graduate school experience with. I can be myself around them and not be judged. We have supported each other through sad and happy times. We were together, vulnerable and raw. We talked about impostor syndrome and mental health. We talked about everything and anything. I love them all so much.

My classmates and I went on so many wonderful adventures, to name a few: Corgi Races, Puzzle Rooms, Halloween Parties, 4$^{th}$ of July, HUMP, Sam Remix, Greek festival, tulip festival, wine tasting, Seattle Restaurant Week, skiing, snowshoeing, hiking, backpacking, dancing. Our

trip to Crater Lake National Park was incredible, laying down in the middle of the road and watching the stars while talking to my dear friends, despite the bat bothering Eleanor, was something I will never forget. I will not forget the feeling I got when watching the Total Solar Eclipse as it got cold and quiet, as the time seemed to stop, and the 11-hour drive back from Salem. I am so very grateful that we could spend six friendsgivings together, with boardgames, drinks, incredible food and conversations. The truth box has brought many truths out.

I have sweet memories with all of my classmates. My dear bus buddy, Eleanor Vane, I will never forget the time you slapped the mail out of my hands, the many laughs we had, the many hikes we took where we talked for hours to make the pain less painful, and someday I will pay tribute to you. I have never seen a better meteorite shower than the one I saw with Sean F. Gagnon. I appreciate his patience when he had to drive 11 hours and his great sense of humor through it all. I am grateful that Sean and Eleanor took me to the ER when I cut myself and they waited for hours for me to get stiches while a lady meowed in the waiting room. I love and miss Gulsima Usluer. I miss our Musashi's sushi dates, followed by hour-long talks and walks until we would find a new bar to explore. How could I forget Blue Star Café and Pub? Every time I see a linden tree, I think of her. Pizza-Bobby Langan took me on my first backpacking trip in the Enchantments, despite being extremely cold, hungry, and tired, I fell in love with backpacking and I will forever be grateful for that. Waking up in the middle of the woods, having coffee by a beautiful lake in complete silence has helped me stay sane throughout the more difficult times. Lexi Walls has helped me exercise not only my mind, but also my body. I am always amazed at how her mind works, the questions she asks and the strength she has. I appreciate the many wonderful talks we had, the red and white candy story, and I was lucky to go to Banff with her and see the most beautiful lakes I have ever seen. Rachel Hutto and Sean Gillespie had the most amazing wedding.

They are both such lovely and kind people, except when we are playing boardgames together. Getting Sean's perspective on various topics has always been illuminating. Who could ever forget the Tacate – Long Island Iced Teas adventure? Seeing sweet gentle Rachel smash various objects in the Rage Room was a cathartic moment. I admire her ability to communicate scientific ideas with enthusiasm and elegance. Let's never forget the dizzy bat moment from the first year 4$^{th}$ of July party.

My friends have left such an impact on me. But graduate school wouldn't have been the same without my lab. One of the best decisions I have ever made was to join Ning Zheng's lab. He is an incredible scientist, who is not only capable of following every topic, but also can always identify key questions and directions to pursue. His ability to think, present, and work hard are inspiring. His excitement is contagious. There have been so many times I have stopped by his office to ask him a question about an experiment, and three hours later we are discussing some obscure philosophical, ethical or political topic. Ning doesn't tell me I am wrong; he just asks questions until I realize how wrong I am. He is an incredible mentor because he thinks about the people in his lab, he is caring and understanding.

Everybody in the lab has been extremely helpful and I am grateful to them for their input. The people that helped me the most are the ones that amused me. Every day Tom Hinds is in the lab it is almost certain I will laugh. Simar Singh is not only one of the funniest people I know with a dark sense of humor, but he also got me socks before the New York trip. I am excited to read his latest novel "UB Goes to College". My bay mate, Junping Fan, is sweet and caring, she is lovely and makes unexpected jokes with a straight face. I will be careful Junping! All of you made the lab so fun for me.

Outside of lab, I have made other friendships as well. Andrew Enyeart has become one of my best friends in a short period of time. He has helped me journal, meditate, seek counseling, enjoy silence and alone time. By having incredible communication skills, Andrew has given me access to his incredible brain and inner world. He is also the best driver and our trips to the Urgent Care were always so fun.

I have never been able to relate to anybody as much as I have been able to relate to my love Chloe Adams. She went from Celia's friend, who I met during Martin Luther King weekend, to my rotation student, and then to one of my best friends. We can talk openly about anything without any judgment. We have laughed and cried together too many times to count. Chloe's sense of humor is as perfect as she is. I look fondly on our celebration of friendship, our trip to Canada, the girlfriend weekend song, the art/wine walk, Friend Fridays, pH paper in the eye, and cranes. Her love and support have been essential for my wellbeing during the last seven months.

There are many people other people I have shared special moments with and I want to thank: Kimiko Lee, Ha Dang, Celia Bisbach, Inez Keiko, Ryan Kane, Ericka Garufi, Anindya Roy, Cameron Chow, George Ueda, Jorge Fallas, Ian Haydon, Lorela Paco, Hannah Baughman, Benja Basanta, Gilad Touboul, Mark Benhaim, Nicole Weston, Eddie Hodge, Matt Mellin, Alex Hammerberg, Ramon Jones, Patrick Nygren, Elena Pandres, Matt Crane, Jon Klein, Hannah Patchen, Aaron Vetter, Thomas Newbern, Eric Clute, Thomas Redd, Karena Tien, Ashley Tsue, Gurdeep Kaur. Besides people, I need to acknowledge two dogs that have helped me tremendously, Caleb and Pesto.

You all have made grad school not just bearable, but actually amazing. I truly believe it is the most fun I have ever had and that is because of the fantastic people that surround me. I am so grateful to have shared experiences with all of you. Thank you!!

# DEDICATION

To all the people that I loved, love and will love.

# Chapter 1. UBIQUITIN PROTEASOME SYSTEM

The following work has previously been published and was adapted from: Rusnac D.V., Zheng N. (2018) Overview of Protein Degradation in Plant Hormone Signaling. In: Hejátko J., Hakoshima T. (eds) Plant Structural Biology: Hormonal Regulations. Springer Nature
License Number 4720900051720

## 1.1 INTRODUCTION

Protein degradation is a proteolytic process which counteracts protein synthesis and determines the half-lives of all proteins in the cell. Although some proteins can be extremely long lived, the majority of cellular proteins have a measurable half-life, ranging from minutes to days (Toyama and Hetzer, 2013, Hershko and Ciechanover, 1998). Early studies of protein breakdown in animals and plants emphasized its roles in protein quality control and amino acid re-utilization, which help eukaryotic cells to cope with cellular and environmental stress as well as nutrient starvation. Recent advances, however, have unraveled an unexpected regulatory function of protein degradation in actively controlling the abundance of a variety of intracellular proteins, thereby, modulating their activities (Hershko and Ciechanover, 1998).

The ubiquitin-proteasome system (UPS) is the central player for intracellular protein degradation and is evolutionarily conserved in all eukaryotes, including plants (Vierstra, 2009, Callis, 2014). In an ATP-dependent manner, the UPS is programmed to respond to diverse cellular cues and selectively label target proteins for rapid breakdown. Thanks to the groundbreaking work by Avram Hershko, Aaron Ciechanover, Irwin Rose, Alfred Goldberg, Alexander Varshavsky, and many other pioneers in the field, most of the key components of the UPS have now been identified and biochemically characterized in great detail (Wilkinson, 2005). Our mechanistic understanding of the UPS function has also benefited tremendously from the extensive structural studies in the

past two decades. This chapter offers a brief overview of the UPS and its major constituents in eukaryotes.

## 1.2 Ubiquitin and Ubiquitin-Like Proteins

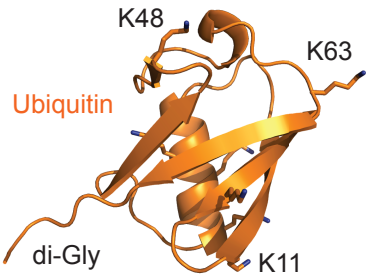



Figure 1. Structure of ubiquitin. Ubiquitin with seven lysine residues (sticks) and a C-terminal diglycine motif shown in orange cartoon representation.

Ubiquitin is a 76 amino acid protein universally found in all eukaryotic species and broadly expressed in different tissues of animals and plants. It has a highly conserved polypeptide sequence, which differs by three amino acids between the yeast and human orthologues. Ubiquitin is characterized by a compact β-grasp fold and a flexible C-terminal tail terminated by a diglycine motif after maturation (Figure 1). In the UPS, ubiquitin serves as a protein post-translational modifier, whose C-terminal carboxyl group is covalently conjugated to the ε-amino group of a substrate lysine residue via an iso-peptide bond. As ubiquitin itself also has seven lysine residues, poly-ubiquitin chains can be formed when the carboxyl terminus of one ubiquitin molecule is linked to a lysine residue of a second copy (Figure 1 and 2). Depending on which ubiquitin lysine residue is involved in chain elongation, poly-ubiquitin chains can be built with different linkages either in a homogeneous or branched fashion (Komander and Rape, 2012, Meyer and Rape, 2014). Among different types of ubiquitin chains, the Lys-48-linked tetraubiquitin chain has long been established as the minimal signal for proteasome targeting (Figure 2) (Thrower et al., 2000).

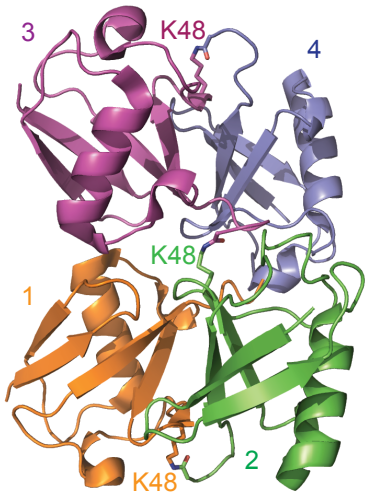



Figure 2. Structure of tetra-ubiquitin. Lys-48 linked tetra-ubiquitin chain where each ubiquitin is displayed in a different color (orange, green, purple and blue) (PDB:2O6V).

In most, if not all, eukaryotic organisms, several proteins have been found to share sequence homology with ubiquitin and adopt the same ubiquitin fold. These ubiquitin-like proteins (UBLs), exemplified by NEDD8 and SUMO, also feature a C-terminal diglycine motif after precursor processing and function as protein modifiers in diverse cellular pathways, including the UPS (van der Veen and Ploegh, 2012). In most cases, these UBLs modify substrate proteins in a monomeric form and elicit their effects by altering the structural topology, protein network, or cellular localization of the targets.

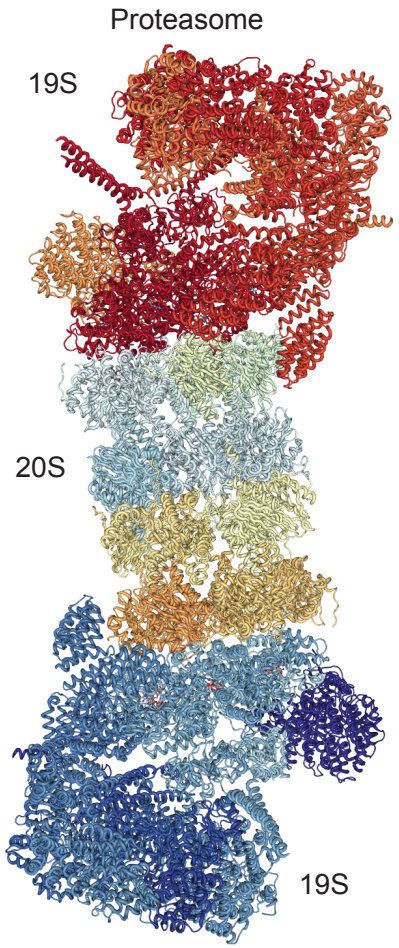



Figure 3. Structure of 26S proteasome. 26S proteasome with the 20S core particle and the 19S regulatory particle. Sub-complexes are colored differently (PDB:5GJR).

## 1.3 Proteasome as a Protein Degradation Machine

The 26S proteasome is an intracellular multi-subunit proteolytic machinery localized in both cytosolic and nuclear compartments that acts as the most downstream component of the UPS (Coux et al., 1996). Due to its protein destruction function, the 26S proteasome has evolved to safeguard its proteolytic activity at both architectural and functional levels (Tomko and Hochstrasser, 2013). To achieve tight regulation of its protease function, the 26S proteasome is composed of two parts, the 20S core particle, which carries the catalytic activities, and the 19S regulatory particle, which controls the access to the active sites hidden inside the enzymatic core (Figure 3). Crystal structures of the 20S core particle revealed a cylindrical architecture, which consists of four stacked rings sequestering a central pore (Kish-Trier and

Hill, 2013). The inner two rings are each constructed from seven β-subunits, harboring three peptidase activities with the catalytic sites buried in the interior cavity, whereas the outer two rings are each formed by seven α-subunits whose N-terminal regions converge at the center and together close up the proteolytic chamber of the core particle. By docking to the outer rings of the 20S particle, the 19S regulatory particle of the proteasome is engaged with the proteasome core on its two ends and only feeds the degradation machinery the polyubiquitinated protein substrates. Distinct from the 20S particle, the 19S particle has a highly asymmetric structure, which has historically been divided into two sub-complexes, the lid and the base (Lander et al., 2012). The base of the 19S particle contains six different ATPase subunits, which are assembled into a trimer of dimers ring structure. In addition, it also features three non-ATPase subunits, which have ubiquitin receptor functions. Together, these 19S base subunits are responsible for recognizing polyubiquitinated substrate, opening the gate of the 20S core, unfolding and translocating the linearized polypeptide into the proteolytic chamber. The 19S lid complex, which consists of ten subunits, partially covers the base ATPases and makes direct contacts with the 20S core. Besides contributing to ubiquitin recognition, one important function of the 19S lid is to catalyze the removal of ubiquitin from the substrate before it is fed into the protease core. Recent advances in cryo-electron microscopy (cryo-EM) have not only allowed near atomic resolution structural determination of the entire proteasome, including the 19S regulatory particle, but also helped reveal the protein degradation machinery in different functional states with substrate and/or nucleotides bound (Bhattacharyya et al., 2014).

## 1.4 THE E1-E2-E3 ENZYME CASCADE

Ubiquitin conjugation to a protein substrate, a process referred to as ubiquitination (or ubiquitylation), is the hallmark of ubiquitin-dependent protein degradation. Protein ubiquitination

is catalyzed by the sequential actions of three enzymes, the E1 ubiquitin-activating enzyme, the E2 ubiquitin-conjugating enzyme, and the E3 ubiquitin-protein ligase (Pickart, 2001) (Figure 4). Free ubiquitin is first activated by the E1 enzyme, which uses ATP-$Mg^{2+}$ to catalyze the acyl adenylation of ubiquitin's C-terminal carboxyl group and then captures the activated ubiquitin tail with its catalytic cysteine via a thiolester bond. Upon binding to a ubiquitin-conjugating enzyme, the ubiquitin-activating E1 enzyme subsequently transfers ubiquitin to the active site cysteine residue on E2 through a trans-thiolesterification reaction. As a highly active enzyme, E1 is responsible for constitutively charging E2 enzymes with ubiquitin. Vertebrates have two E1 genes, whose protein products, known as UBE1 and UBA6, have been found to preferentially charge different E2s (Jin et al., 2007). In *Arabidopsis thaliana*, two ubiquitin E1 enzymes, UBA1 and UBA2, have also been identified to carry out non-redundant functions (Goritschnig et al., 2007).

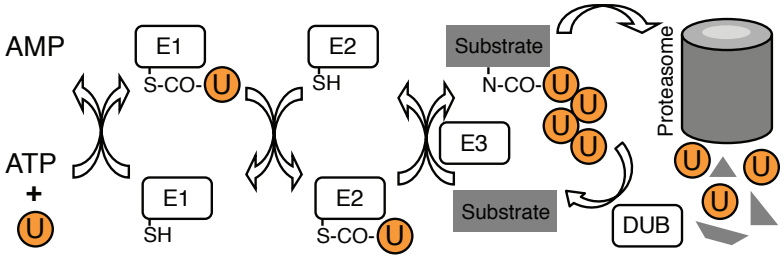



Figure 4. Schematic representation of the ubiquitin-proteasome system. Ubiquitin-proteasome system with the E1-E2-E3 enzyme cascade acting upstream of the proteasome and the counteracting deubiquitinases.

In contrast to the small number of E1 enzymes, the ubiquitin-conjugating E2 enzymes are numbered in 30 – 40 in higher eukaryotes and often act in different cellular pathways (Wenzel et al., 2011b). All E2 enzymes share a conserved ~150 amino acids catalytic core domain, which adopts a classic UBC (Ubiquitin-conjugating catalytic) fold with the active site cysteine tucked in a cleft between two loops. Certain E2s feature additional N-terminal or C-terminal extension sequences, whereas a specific sub-group of E2s contains an internal acid loop close to the active site cysteine. Although E2s were once thought to be simple ubiquitin "carriers", recent studies

have shown that they display distinct intrinsic reactivity and often play a critical role in dictating the linkage specificity of a polyubiquitin chain (Stewart et al., 2016). Because many ubiquitin-charged E2s (Ub~E2s) selectively interact and function with specific types of ubiquitin E3 ligases (see below), their active sites can have characteristic reactivity towards different attacking groups, such as the ε-amino group of the lysine side chain and the thiol group of a cysteine residue. Furthermore, with the help of extra sequence elements or binding partners, some E2s can differentiate the lysine residues on the receiver (proximal) ubiquitin, which accept the C-terminus of the incoming donor (distal) ubiquitin during chain extension. Interestingly, some E2 variants, which lack the active site cysteine, have been shown to interact with a canonical E2 to confer linkage-specific polyubiquitin chain activities.

Although the thiolester bond in the Ub~E2 conjugate is less stable than the iso-peptide bond linking ubiquitin and substrate, transfer of ubiquitin from an E2 to a substrate does not occur efficiently until an E3 ubiquitin ligase is present (Pickart, 2001). In the three-enzyme cascade, the E3 enzyme performs two critical functions to facilitate substrate ubiquitination. First, E3s stimulate the reactivity of a ubiquitin-charged E2 to accelerate ubiquitin discharge. Second, E3s provide a platform onto which a specific protein substrate and the ubiquitin-charged E2 are recruited and brought together in close proximity. Ubiquitin E3 ligases, therefore, represent an ideal class of enzymes favored by evolution for adopting novel functions that can couple protein ubiquitination and degradation with various upstream signals in diverse cellular pathways.

## 1.5 THREE TYPES OF UBIQUITIN E3 LIGASES

The functional importance and versatility of ubiquitin ligases is best manifested by the different types of E3s and their sheer number in the eukaryotic genomes in comparison to other UPS enzymes. In *Arabidopsis thaliana*, more than one thousand genes have been identified to encode

putative ubiquitin ligases (Vierstra, 2009). Although this number varies among other plant species, the prevalence of E3s and their roles in regulating plant physiology are obvious. Intriguingly, plant pathogens are known to produce effector proteins that either mimic or hijack E3 ligases to take advantage of the host UPS and benefit their infection and life cycle (Banfield, 2015). Such cross-kingdom functions further highlight the central roles played by ubiquitin ligases in the cell.

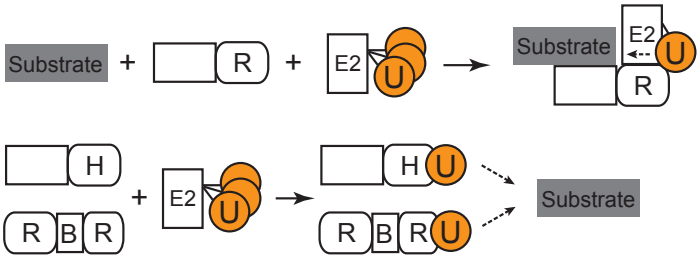



Figure 5. Schematic drawing of the three classes of E3s.
The three classes of ubiquitin E3 ligases (R=RING, H=HECT, and RBR) are shown with their different ubiquitin transfer mechanisms.

In all eukaryotes, three types of E3s have been identified, which are grouped based on their different signature sequence motifs and distinct catalytic mechanisms (Figure 5). The RING (Really Interesting New Gene) domain defines the largest family of ubiquitin ligases, known as RING-type E3s, which share a common protein fold consisting of two zinc-binding fingers with eight zinc-coordinating cysteine and histidine residue (Deshaies and Joazeiro, 2009)(Figure 6). Besides the RING domain, these E3 ligases either contain a substrate-binding domain in the same polypeptide or belong to a multi-subunit ubiquitin ligase complex, which uses another subunit for recruiting substrate. The RING-types E3s are distinguished from other E3s by catalyzing the direct transfer of ubiquitin from an E2 to the subunit. Recent structural studies have shown that, upon binding to a ubiquitin-charged E2 enzyme, the RING domain makes contacts with both the E2 and the donor ubiquitin and stabilizes the Ub~E2 conjugate in a "closed" conformation (Figure 6) (Plechanovová et al., 2011, Dou

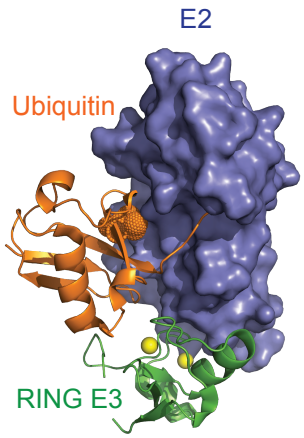



Figure 6. Structure of RING E3-Ub~E2 complex. E2 is shown using surface representation in blue, RING E3 and ubiquitin are represented in cartoon fashion in green and orange, respectively. Yellow spheres represent zinc ions. (PDB:4AP4).

et al., 2012, Pruneda et al., 2012). In doing so, a RING E3 activates the ubiquitin-charged E2 for ubiquitin transfer by presumably optimizing the geometry of the E2 active site for the nucleophilic attack by the side chain of a lysine residue in either a substrate or a receiver ubiquitin molecule.

The HECT (Homology to E6AP Carboxyl Terminus) type of E3s represents a second family of ubiquitin ligases, which are characterized by their common C-terminal catalytic domain, known as HECT domain (Rotin and Kumar, 2009)(Figure 7). With a bilobed structure, the HECT domain harbors an active site cysteine, which forms an obligate thiolester intermediate with ubiquitin to promote substrate ubiquitination (Figure 5) (Huang et al., 1999, Metzger et al., 2012). Because the first step of ubiquitin transfer mediated by the HECT E3s involves a trans-thiolesterification reaction, in which ubiquitin is passed from the active site cysteine of the E2 to that of the E3, HECT E3s only function with a specific sub-set of ubiquitin-conjugating enzymes. Akin to single polypeptide RING E3s, most known HECT E3s recognize their specific substrate through regions outside their catalytic domain. Although the human genome encodes nearly 30 HECT E3s, this family of ubiquitin ligases remains relatively small in plants (Marín, 2013).

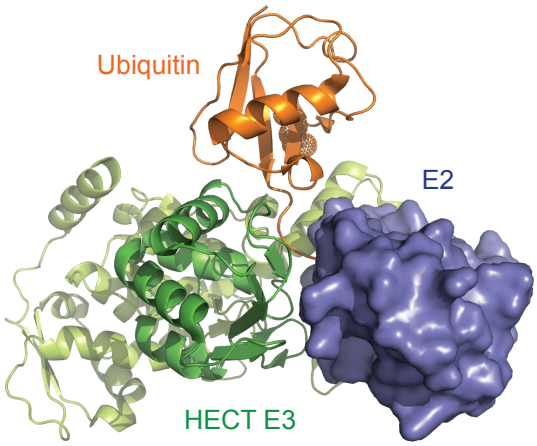


Figure 7. Structure of HECT E3-Ub~E2 complex. The surface representation of ubiquitin-conjugating enzyme E2 is in blue. HECT E3 and ubiquitin are shown in cartoon view in green and orange, respectively (PDB:3JW0).

Remarkably, recent studies have unveiled a third family of E3s, which are named RBR (RING-in-Between-ring-RING) E3s (Spratt et al., 2014, Wenzel et al., 2011a). Despite the presence of several zinc-finger-containing RING-like domains, RBR E3s are mechanistically closer to the HECT E3s than the RING E3s. While the RING1 domain of RBR E3s is responsible

for recruiting a ubiquitin-charged E2 enzyme, ligation of ubiquitin to the substrate involves the formation of a ubiquitin~E3 intermediate, which is anchored at a strictly conserved catalytic cysteine found in the RING2 domain of the E3s (Figure 5 and 8). Similar to the HECT E3s, the RBR E3s relay ubiquitin to the substrate and display strong E2 preferences. Recent structural analyses of several RBR E3s have revealed that these multi-domain ubiquitin ligases almost exclusively adopt an auto-inhibited conformation in isolation (Trempe et al., 2013, Wauer and Komander, 2013, Stieglitz et al., 2013, Lechtenberg et al., 2016) (Figure 8). Activation of these enzymes might be achieved by post-translational modifications of the E3s or upon interactions with their binding partners, which presumably recruits specific substrates. So far, RBR E3s have been poorly studied in plants (Marín, 2010). However, the potential functional connections of a RBR subfamily, Ariadne/HHARI, with the superfamily of cullin-RING E3s, as suggested by recent studies, might implicate a prominent role of the RBR E3 in plant hormone signaling (see below in Chapter 4.2) (Scott et al., 2016).

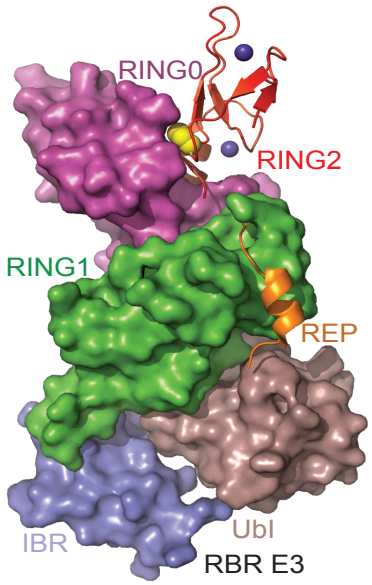


Figure 8. Structure of RBR E3. The structure of different domains from RBR E3 are displayed in cartoon and surface representation. Blue spheres represent zinc ions. (PDB:4K95)

## 1.6 DEUBIQUITINASES

Analogous to most protein post-translational modifications, protein ubiquitination is reversible and the activities of ubiquitin ligases can be counter-balanced by enzymes capable of cleaving ubiquitin-linked iso-peptide bonds (Figure 4). These iso-peptidases, also known as deubiquitinases (DUBs), can either trim various ubiquitin chains with specific linkages or catalyze the removal of ubiquitin from substrate (Komander et al., 2009). Their activities not only enable

ubiquitin recycling prior to substrate degradation by the proteasome, but also provide a mechanism for regulating protein ubiquitination in a dynamic manner. In animals, DUBs are classified into six different sub-families (USPs, UCHs, OTUs, MJDs, JAMM, and MINDYs) based on their sequence homology. The same six DUB families are also found in plants with a total of ~60 different family members in the *Arabidopsis* genome. Although little is known about their functions, it is expected that their deubiquitinase activities might be involved in fine-tuning the ubiquitination and degradation of many substrate polypeptides, including those implicated in hormone signaling.

# Chapter 2. CRL UBIQUITIN LIGASES

Accepted and soon to be published: Rusnac D.V., Zheng N. Structural Biology of CRL Ubiquitin Ligases. In: Sun, Y., Wei, W., Jin, J. (eds) Cullin-RING Ligases and Protein Neddylation: Biology and Therapeutics. Springer Nature

## 2.1 OVERALL CRL ARCHITECTURE

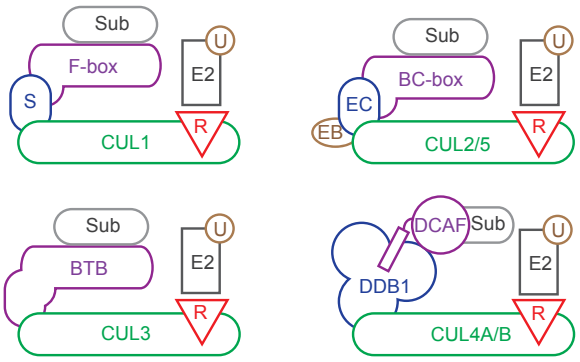



Figure 9. Schematic drawing of all CRLs. S: SKP1, R: RBX1, EB: Elongin B, EC: Elongin C, U: Ubiquitin, Sub: substrate. CULs are shown in green, adaptor proteins are in blue, substrate receptors are in purple, substrates are in gray, RBX1 is in red, E2 is in black, ubiquitin is in brown.

Cullin-RING ubiquitin Ligases (CRLs) are modular E3 ligases that utilize interchangeable substrate receptors to recruit a variety of substrates onto a common catalytic scaffold (Figure 9). All CRLs contain a scaffolding cullin protein, namely, CUL1, CUL2, CUL3, CUL4A, CUL4B, or CUL5. Most of these cullin proteins bind a specific adaptor polypeptide through their N-terminal regions, which helps engage interchangeable substrate receptors to dock their cognate substrates to the E3. The ubiquitination reaction involves the transfer of ubiquitin from an E2 ubiquitin-conjugating enzyme to the substrate. The cullin scaffold facilitates this process by using its C-terminus to house the catalytic subunit, RBX1 or RBX2, which directly interacts with the ubiquitin-conjugated E2 enzyme.

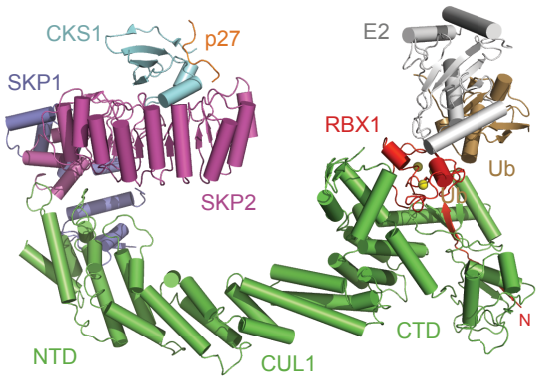



Figure 10. Structural model of an CRL1 in complex with substrate and ubiquitin-E2. A model of CRL1(SCF)$^{SKP2\text{-}CKS1}$ in complex with substrate p27 and a ubiquitin-charged E2 are shown in cartoon.

The first glimpse of the architecture of a CRL complex came from the X-ray crystal structure of human CUL1-RBX1 (Figure 10) (Zheng et al., 2002b). In the structure, CUL1

adopts a highly elongated shape with two functional domains. The stalk-like N-terminal domain of CUL1 is made of helical repeats that interact with the CRL1 adaptor protein, SKP1. The C-terminal domain of the scaffold has a more globular fold, which harbors RBX1. The N-terminal and C-terminal domains are coupled through a hydrophobic interface, which structurally affixes the two functional portions of the E3 machinery. Mutational work that introduced flexibility to CUL1 abolished its ability to promote substrate ubiquitination, but not substrate binding, reinforcing the idea that its rigidity is necessary for its E3 activity. The inflexible CUL1-RBX1 ubiquitin ligase scaffold spans more than 100 Å, allowing the E3 to accommodate substrates of various shapes and sizes. While this distance is beneficial for the docking and ubiquitination of large substrate proteins, it seems to also present a challenge for small substrates to be ubiquitinated by the E2 that is tens of ångström away, unless flexibility is introduced somewhere (see below).

The globular domain of CUL1 does not only foster RBX1, but also intercalates the catalytic subunit. RBX1 is a RING-type zinc-finger, which consists of a N-terminal β-strand and a C-terminal core domain that coordinates three zinc ions. The RBX1 β-strand inserts itself between five β-strands of CUL1's globular domain to create a stable intermolecular β-sheet. At the same time, the RING domain of RBX1 is hosted by the rest of CUL1 C-terminal region through a seemingly loose interface. Overall, the two proteins appear to exist in a "symbiotic" relationship, where they stay together throughout their life cycles. This interaction between CUL1 and RBX1, which involves an anchored end and a relaxed interface, allows for the RING domain of RBX1 to pop out upon the post-translational modification (PTM) of CUL1 by NEDD8, a ubiquitin-like molecule. Such an intermolecular topological change provides flexibility to CRL1 that helps reposition the ubiquitin-conjugated E2 to potentially approach the substrate. Further discussion about the regulation of CRLs by neddylation will be in Section 2.11.

CUL1 uses two α-helices to bind SKP1. The majority of the residues clustered at this interface are strictly conserved between CUL1 orthologues. Interestingly, these residues are not conserved across CUL1 paralogues, i.e. CUL1-CUL5. But they are conserved within orthologues of each cullin family member (Zheng et al., 2002b), suggesting that the binding mode between CUL1 and SKP1 is not unique to CRL1, but common to all cullins and their adaptors. This has been confirmed by subsequent structural studies of other CRLs. Sharing sequence homology with SKP1, Elongin C (EC) serves as the adaptor for CUL2 and CUL5 together with Elongin B (EB) to recruit BC-box substrate receptors. Despite the lack of sequence homology with SKP1, the CUL3 adaptor/substrate receptor, BTB proteins, share structure homology with SKP1 within their BTB domains. On one hand, it is not unexpected that the same two N-terminal helices of CUL1, CUL2, and CUL3 are used for the recruitment of their adaptor proteins. On the other hand, it is surprising that CUL4 uses the exact same region as CUL1–3 to interact with its adaptor, DDB1, which is neither structurally nor sequence wise similar to SKP1, Elongin C, or the BTB domain.

## 2.2 CRL1 Adaptor and Substrate Receptors

CRL1, the prototype of CRL E3s, is also known as SCF, which stands for SKP1, CUL1, and F-box proteins. The substrate recruitment function of CRL1 is conducted by a family of proteins that share an N-terminal ~40 residues F-box motif that constitutively interacts with SKP1 (Skowyra et al., 1997). These F-box proteins feature characteristic C-terminal protein-protein interaction domains, which are responsible for binding specific substrates. Based on their predicted tertiary structures, the 69 human F-box proteins have been classified in three groups: FBXL for leucine-rich repeats (LRR)-containing proteins, FBXW for the ones containing WD40 repeats, and FBXO for F-box protein with other folds (Jin et al., 2004). The diverse protein-protein interaction domains used by F-box proteins enable them to recognize different substrates with high specificity and

optimally orient these substrates to receive ubiquitin from the RBX1-bound E2 enzyme. In most, if not all, structures of F-box proteins in complex with SKP1, the N-terminal domain of the substrate receptors appears to be structurally coupled to the SKP1-F-box module, which displays little structural variation. Together with the stable SKP1-CUL1 interface, the entire CUL1-SKP1-F-box protein complex has been postulated to play a role in spatially positioning the bound substrate relative to the catalytic end of the E3 platform (Zheng et al., 2002b).

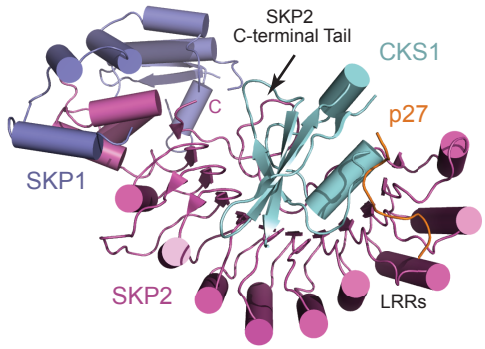



Figure 11. Structure of p27-CKS1-SKP2-SKP1 complex. An orthogonal view of the SKP1-SKP2-CKS1-p27 complex, where each protein is displayed in cartoon fashion in blue, magenta, cyan, and orange, respectively (PDB: 1LDK, 2AST).

The first structure of an F-box protein, SKP2, was determined almost 20 years ago, which revealed the binding mode of the F-box motif to SKP1 (Figure 10, 11) (Schulman et al., 2000). $CRL1^{SKP2}$ is a key regulator of mammalian cell cycle progression. The overall structure of the SKP1-SKP2 complex resembles a sickle. The handle of the sickle is made out of the SKP1-F-box structural module, while the variable linker and leucine-rich repeats of SKP2 constitute the blade. SKP1 and the F-box motif of SKP2 interact via an extensive interdigitated interface that consists of four alternating layers from each protein. Although part of the interface is mediated by residues that are not strictly conserved among all F-box motifs, it is generally believed, and subsequently validated, that all F-box proteins engage SKP1 in a similar binding mode.

The C-terminal domain of SKP2 is composed of ten LRRs, each of which is made of an α-helix and a β-strand. These LRRs pack in tandem and give rise to an arc-shaped structure, which is characterized by a concave side formed by parallel β-strands and a convex surface presented by α-helices. In many known non-E3 LRR proteins, the concave surface is involved in protein-protein

interaction. The FBXLs, including SKP2, are no exception. Remarkably, the FBXL subfamily of F-box proteins have evolved different numbers of LRRs, which give rise to curved structures with different diameters, arc lengths, and pitches. These features presumably allow the FBXLs to hold diverse substrates. Interestingly, SKP2 also features a long C-terminal tail, which wraps back to the concave surface of the LRR domain and provides an additional structural element for substrate recruitment.

Unlike many F-box proteins, which directly recruit CRL1 substrates, SKP2 requires yet another adaptor protein, CKS1, to bind and ubiquitinate its substrate $p27^{Kip1}$. The crystal structure of the SKP1-SKP2-CKS1-p27 complex shows that CKS1 is anchored to the concave surface of SKP2 LRR domain and is supported underneath by the SKP2 C-terminal tail (Figure 11) (Hao et al., 2005). The residues that are involved in the CKS1-SKP2 interaction are conserved in animal orthologs on both sides of the interface, underlying the functional importance of CKS1. Furthermore, CKS1 has been found to form a stable complex with CDK2-Cyclin A, which might contribute to the binding of $p27^{Kip1}$ to the CRL1 machinery.

The crystal structure of SKP1-βTrCP offered the first sight of an FBXW-type F-box protein, which plays a major role in the Wnt signaling pathway by catalyzing the polyubiquitination and degradation of β-catenin (Figure 12) (Wu et al., 2003). The SKP1-βTrCP complex adopts a bi-lobal structure, displaying the substrate at the opposite side of SKP1. The F-box motif of βTrCP interacts with SKP1 in a similar fashion as SKP2. The substrate binding function of the

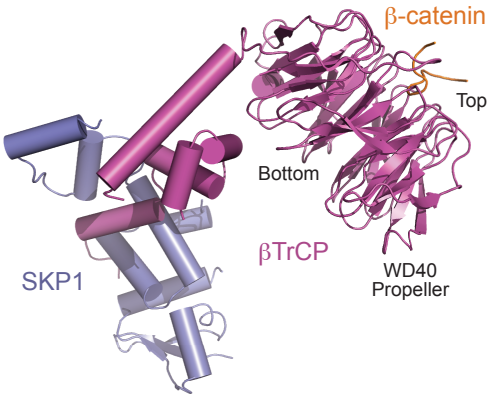



Figure 12. Structural representation of β-catenin-βTrCP-SKP1 complex. Crystal structure of SKP1 in complex with the WD40-repeat domain containing F-box protein, β-TrCP, which interacts with the degron of the substrate protein, β-catenin. SKP1 is displayed in blue, β-TrCP in magenta and β-catenin in orange (PDB: 1P22).

F-box protein is performed by its C-terminal WD40-repeat domain, whose name is derived from the ~40 amino acid sequence repeat that contains structurally essential tryptophan (W) and aspartic acid (D) amino acids. WD40 repeats are known to fold into a β-propeller structure, which is usually made of seven β-sheets arranged in a circular manner around a central narrow channel. In βTrCP and other FBXW proteins with known structures, the amino acids located on the "top" surface are involved in substrate recognition. Through a rigid linker helix connecting to the SKP1-F-box module, βTrCP is thought to position its substrate toward the E2 for ubiquitin transfer.

## 2.3 CRL2 AND CRL5

CRL2 is organized in a similar manner as CRL1, with CUL2 serving as the scaffolding protein, EB-EC assisting as adaptor proteins, and members from the BC-box family of proteins acting as substrate receptors (Mahrour et al., 2008). CUL2 shares 38% of sequence identity with CUL1. As predicted by homology models and subsequently confirmed by crystal structure, CUL2 adopts a similar elongated structure as CUL1 (Nguyen et al., 2015, Cardote et al., 2017). As the adaptor protein of CUL2, EC shares about 30% sequence identity with SKP1 and has a BTB core fold. Unlike SKP1, EC does not act alone to bridge the substrate receptor BC-box proteins to CUL2 and mandates its association with the ubiquitin-like molecule EB. It is not clear yet if EB with the ubiquitin-like fold plays any particular role in CRL2 besides stabilizing EC. In comparison to the F-box proteins, the BC-box proteins contain structural elements that do not only interact with the adaptor proteins, but also make direct contacts with their cognate cullin scaffolds. Besides a short BC-box motif, some BC-box proteins feature a CUL2-box motif that specifically recognizes CUL2. Interestingly, these two motifs are always consecutively localized in the BC-box protein sequences but could be found at different positions relative to the substrate-binding domain in the

polypeptide. The best studied member from this family of proteins is the VHL tumor suppressor protein that promotes the ubiquitination of HIF1α under normoxia (Kaelin, 2005). When its substrate-binding domain or BC-box motif is mutated, VHL fails to target HIF1α for ubiquitination and degradation, which leads to von Hippel-Lindau disorder, a hereditary predisposition to develop tumors in a variety of organs.

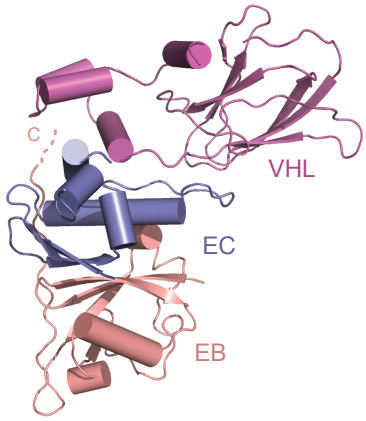



Figure 13. Structural representation of VHL-EB-EC complex. Crystal structure of the BC-box protein, VHL, in complex with Elongin B and Elongin C. The proteins are shown in cartoon representation with VHL in magenta, EC in blue, and EB in peach (PDB: 1VCB).

The structure of VHL-EB-EC unveils the portion of VHL that is involved in substrate recognition as a beta sandwich (Figure 13) (Stebbins et al., 1999). The BC-box and CUL2-binding motifs of VHL together form a three α-helices structural module that is very similar to the structure of the F-box motif. Moreover, EC can be superimposed almost perfectly on part of SKP1. As predicted, the other adaptor protein, EB, adopts a ubiquitin-like α/β roll structure. In this crystal structure, the C-terminal tail of EB interacts with EC and points towards VHL with the last 20 amino acids disordered. Snapshots of VHL-EB-EC in complex with the first N-terminal repeat of CUL2 or the full-length protein reveal that the C-terminal tail of EB adopts an ordered structure interacting with VHL (Nguyen et al., 2015, Cardote et al., 2017). In addition, an internal loop of EC that is disordered in the VHL-EB-EC complex becomes structured upon its interaction with CUL2 (Figure

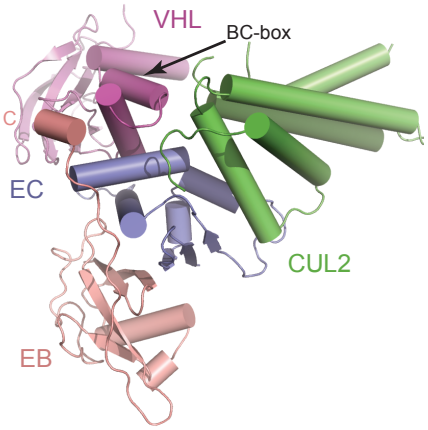



Figure 14. Structure of VHL-EB-EC-CUL2 complex. A different view of VHL-EB-EC complex bound to CUL2 with the stabilized C-terminus of EB labeled with "C". The proteins are shown in cartoon representation with VHL in magenta, EC in blue, EB in peach, and CUL2 in green (PDB: 4WQO).

14). Remarkably, the CUL2-box of VHL makes polar interactions with part of the cullin scaffold, which hints at how different BC-box proteins could differentiate between CUL2 and CUL5, even if they use the same adaptor proteins. A myriad of BC-box proteins with various folds have been identified by biochemical and bioinformatical approaches (Mahrour et al., 2008). A variety of substrate-binding domains have been mapped for the ones that bind CUL2, such as leucine-rich repeats, ankyrin repeats, tetratricopeptide repeats, armadillo repeats, kelch repeats, or SWIM zinc fingers.

Similar to CUL2, CUL5 employs the same adaptor proteins EB and EC to associate with a distinct subfamily of the BC-box proteins (Mahrour et al., 2008). The substrate receptors that bind CUL5 can have one of the following folds: Src homology 2 phosphotyrosine binding domain, ankyrin repeats, SP1a and ryanodine receptor domain, WD40 repeats, Rab-like GTPase domain, protein-L-isoaspartate carboxymethyltransferase, and transcription factor SII-like domain. Unexpectedly, not only cellular BC-box proteins have been found to bind CUL5-EB-EC. A very interesting study has demonstrated that HIV-1 virion is able to hijack CRL5 via an accessory protein, Vif, to promote the degradation of APOBEC3, the host restriction factor that blocks the replication of the virus (Yu et al., 2003). Recent studies showed that this function of Vif entails yet another host protein, the transcription factor CBF-β that plays a role in APOBEC3 expression (Kim et al., 2013). The structure of the Vif-CBF-β-CUL5$_N$-EB-EC

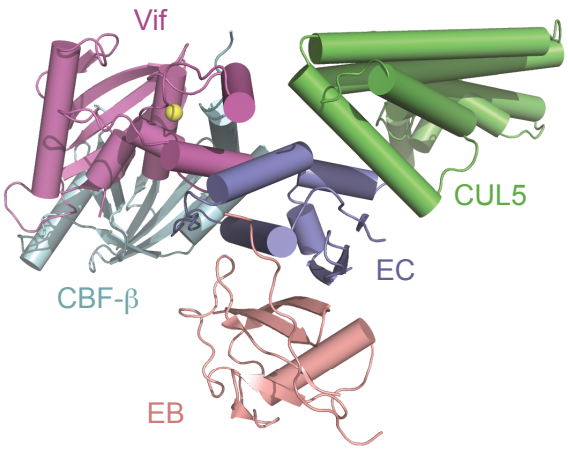



Figure 15. Structural representation of Vif-CBF-β-EB-EC-CUL5 complex. A crystal structure of the HIV accessary protein Vif in complex with EB, EC, CUL5 and a cellular factor, CBF-β, where the proteins are shown in cartoon representation and colored in magenta, peach, blue, green, and cyan respectively. Yellow spheres represent zinc ions (PDB: 4N9F).

complex reveals an overall U-shape, with Vif-CBF-β and $CUL5_N$ making up the arms and EB-EC sticking out at the bottom (Figure 15) (Guo et al., 2014). Akin to cellular BC-box proteins, Vif uses its BC-box and CUL5-box to bridge EC and CUL5, mimicking the positioning of SKP1 and SKP2 in the context of CRL1. This structure provides further evidence supporting the notion that there is a conserved assembly mode for most CRLs.

## 2.4 CRL3

CUL1, CUL2, CUL4, and CUL5 all use adaptor proteins that recruit specific substrate receptors. CUL3, on the other hand, recruits BTB domain-containing proteins that combine the adaptor and substrate receptor functions into a single polypeptide (Pintard et al., 2003, Xu et al., 2003, Geyer et al., 2003). Without sharing sequence homology to SKP1 or EC, the BTB domain adopts a structural fold analogous to these two CRL adaptors and anchors itself to CUL3 in a similar manner. In addition, the BTB domain is known to dimerize, which facilitates the formation of homodimeric CRL3 complexes with two copies of each components of the E3 machinery. This property is reminiscent of select F-box proteins, such as FBXW7 and βTrCP, which also contain a dimerization domain on the N-terminal side of the F-box motif. Similar to the CRL1 and CRL2/5 substrate receptors, the BTB family of proteins feature different protein-protein interaction domains to recruit substrates. These include, but are not limited to, the well characterized MATH and kelch repeats domains that are present in SPOP and KEAP1, respectively.

KEAP1 is one of the best studied CUL3 substrate specific adaptors, as it controls the degradation of NRF2 transcription factor involved in oxidative stress response pathway (Yamamoto et al., 2018). KEAP1 consists of four functional domains: BTB, intervening region (IVR), double glycine repeats (DGR, a.k.a. kelch repeats), and C-terminal region (CTR). While the DGR and CTR give rise to the kelch repeat β-propeller, the IVR domain contains two reactive

cysteine residues that can be modified by toxic electrophiles and subsequently alter the spatial configuration of other domains of the BTB protein. Despite extensive studies, it is still unclear how these chemical-induced structural changes affect the overall architecture of the KEAP1 dimer, thereby affecting NRF2 polyubiquitination and degradation.

SPOP is a BTB-domain protein that is frequently mutated in human cancer. The dimerization and substrate binding mode of SPOP has been determined in its near full-length form (Figure 16) (Zhuang et al., 2009). The MATH domain of SPOP forms an antiparallel β-sandwich, with a central shallow grove that is used by substrates to dock. Dimerization of the CRL3 substrate receptor takes place through a hydrophobic interface between the BTB domains. Mutational analysis indicates that BTB domain dimerization is not necessary for CUL3 binding, which is consistent with its spatial separation from the BTB-CUL3 interface (Figure 17). Nonetheless, defects in BTB dimerization negatively impact the polyubiquitination of the SPOP substrates, suggesting that the dimeric architecture of the CRL3 complex is critical for productive substrate ubiquitination. Interestingly,

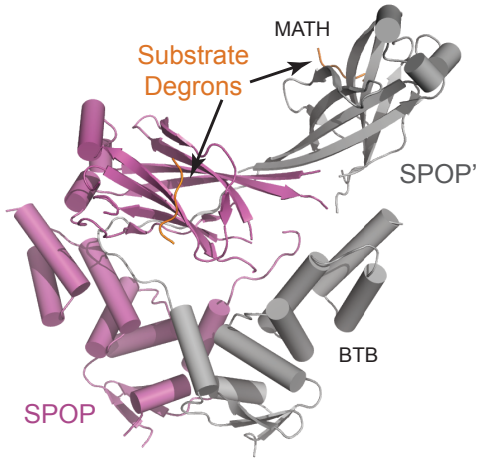



Figure 16. Structure basis for degron recognition by dimeric SPOP. An asymmetric dimeric structure of the BTB-domain protein, SPOP, in complex with two substrate degron peptides. The proteins are shown in cartoon representation. Each SPOP monomer is colored differently in either magenta or gray, while the degron peptides are shown in orange (PDB: 3HQI).

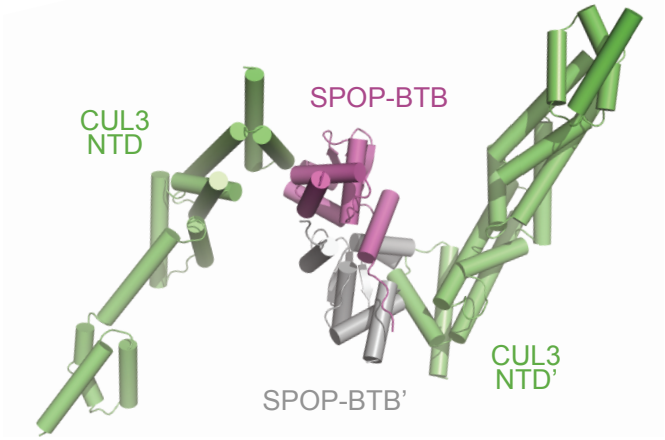



Figure 17. Structure of dimeric $SPOP^{BTB}$-$CUL3^{NTD}$ complex. The dimeric SPOP BTB domain in complex with CUL3-NTD are shown in cartoon representation and colored in magenta or gray and green, respectively (PDB: 4EOZ).

the two MATH domains in the crystal structure are asymmetrically arranged with one MATH domain cradled by the groove between the BTB domains and the other one pushed away from its BTB. Based on different crystal structures obtained for the same complex, slight topological differences have been observed, suggesting that the linker between the two functional domains of SPOP confers structural flexibility to the E3 complex. This structural plasticity has not been previously observed for other CRL family members. The MATH domain of SPOP can recognize a host of substrate degron motifs. It is possible that MATH-domain containing BTB proteins, including SPOP, have evolved a mechanism of engaging substrates with high affinity by simultaneously recognizing two low affinity degron motifs.

Although the BTB domains of KEAP1 and SPOP are responsible for CUL3 binding, not all BTB domain-containing proteins function as CRL3 substrate adaptors. An interesting finding from the SPOP structure is a pair of C-terminal helices that have structural equivalence in the F-box or CUL2-Box. Because the function of the BTB domain was thought to bridge CUL3 and the MATH domain of SPOP, it seemed unnecessary to have such a vestigial element. Surprisingly, this helix pair of the SPOP BTB domain has been proven to be crucial for CUL3 interaction. Because this structural element, named 3-box, is found in some, but not all, BTB proteins, it could be the structural determinant that allows a subset of BTB proteins to function as CRL3 substrate adaptors.

## 2.5 CRL4

While most CRL adaptors adopt a BTB or BTB-like fold, CRL4 employs a 127 kDa protein, DDB1, to dock substrate receptors. DDB1 is a multi-domain protein, made of three β-propellers, labeled as BPA, BPB, and BPC, and a C-terminal helical domain (CTD) (Figure 18) (Li et al., 2006). Remarkably, the three β-propellers are not folded in a linear fashion within the polypeptide.

Instead, two of the β-propellers, BPA and BPB, are inserted into two internal loops of BPC. Together, the three propellers adopt a compact tri-star structure with the CTD housed in the middle. The BPB propeller binds the N-terminal domain (NTD) of CUL4 via two interfaces, one resembling SKP1 binding to CUL1 and the other involving an N-terminal conserved sequence of CUL4 cradling DDB1 (Figure 19) (Angers et al., 2006). The BPA and BPC propellers, on the other hand, pack against each other to create an open clam-shaped structure, which is responsible for holding CRL4 substrate receptors.

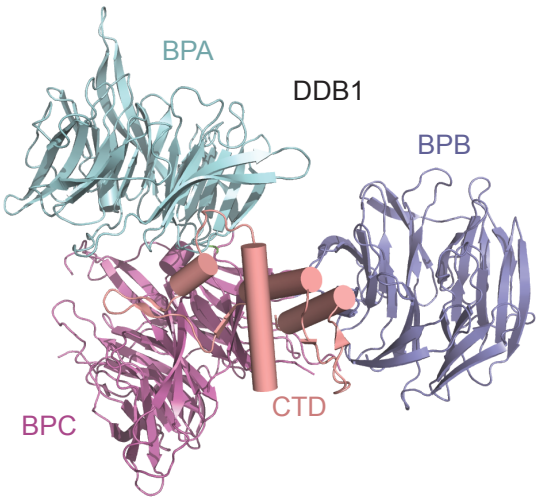



Figure 18. Structural representation of DDB1. Domain architecture of DDB1 with three propellers (BPA, BPB, and BPC) and a helical CTD, which are shown in cartoon representation and colored in cyan, blue, magenta, and peach, respectively (PDB: 2B5M).

To date, a number of DDB1 structures have been documented. Strikingly, the linker between the BPB domain and the BPA-BPC double propeller appears to have a large degree of plasticity, which enables the two functional modules of DDB1 to adopt different orientations relative to each other (Figure 19 and 20). This feature might allow CRL4 to accommodate

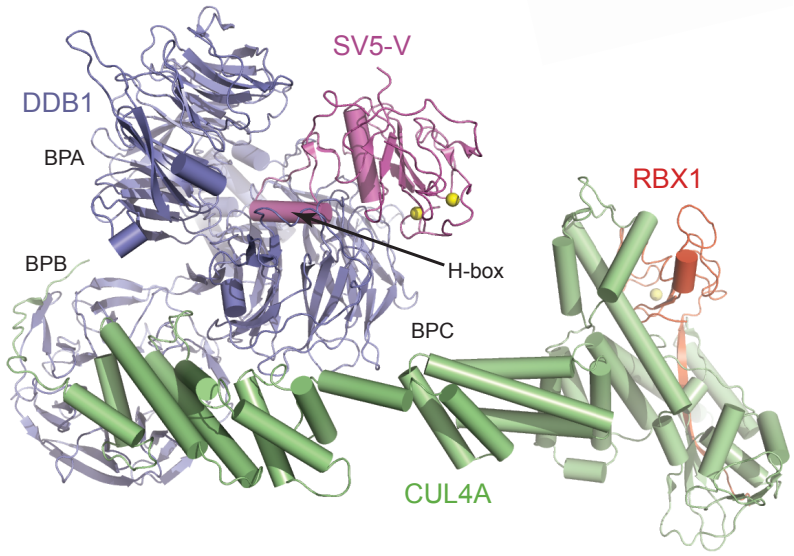



Figure 19. Complex structure of CUL4A-RBX1-DDB1-SV5-V. The H-box motif of SV5-V responsible for binding DDB1 is labeled. The proteins are shown in cartoon representation. RBX1 is in red, CUL4A is in green, DDB1 is in blue, whole SV5-V is in magenta. Yellow spheres represent zinc ions (PDB:2HYE).

and polyubiquitinate substrates of different shapes and sizes. In addition, the structural flexibility of DDB1 enables CUL4 to rotate up to 150° around the substrate receptor, thereby creating a ubiquitination zone that could help detect various lysines on a substrate and promote their ubiquitination. The precise role of the structural flexibility within CRL4 remains to be elucidated.

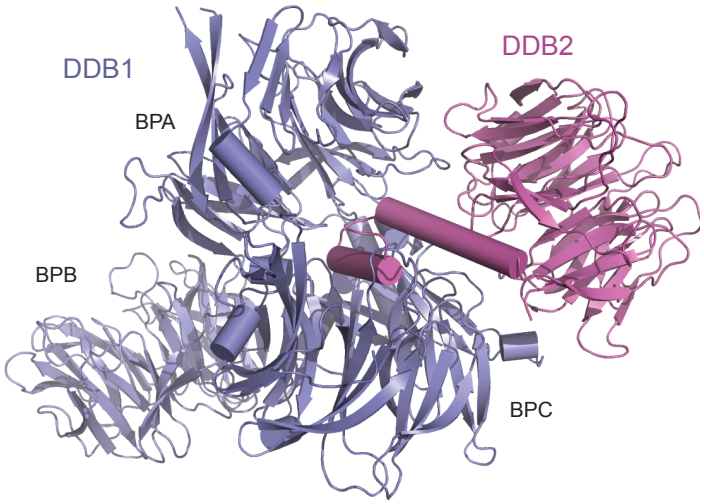



Figure 20. Structural mechanism of DDB1 and DDB2 interaction. Crystal structure of DDB1 in complex with a DCAF protein, DDB2, shown in cartoon representation in blue and magenta, respectively (PDB: 3EI3).

Similar to CRL5, CRL4 is also known to be hijacked by viruses. The SV5-V protein encoded by paramyxovirus has been shown to functionally mimic the CRL4 substrate receptors to mediate the polyubiquitination and degradation of the otherwise stable STAT proteins in the interferon pathway (Horvath, 2004). SV5-V anchors itself to DDB1 by inserting its N-terminal helix into the opening of the BPA-BPC double propeller, interacting predominantly with the "top" surface of BPC (Figure 19) (Li et al., 2006). The C-terminal region of SV5-V folds into unique globular structure featuring a bowl-shaped depression with many conserved hydrophobic and nonpolar residues. This surface region of the viral protein is critical for the recruitment of STATs. Perhaps due to its rich structural features, DDB1 appears to be a frequent target for viral hijacking. Besides SV5-V, Hepatitis B virus X protein (HBx) and woodchuck hepatitis virus X protein (WHx) have also been reported to reprogram the CRL4 adaptor to degrade host factors (Decorsière et al., 2016). Peptide motifs from these proteins have been mapped and crystalized in complex with DDB1. Despite their divergent sequences, these motifs form a common three-turn α-helix

(termed H-box) that anchors to the "top" surface of BPC in a similar fashion as the SV5-V N-terminal helix (Figure 19) (Li et al., 2010).

The structural insights obtained from the DDB1-viral hijacker complexes prompted multiple proteomics studies that were aimed at identifying possible cellular CRL4 substrate receptors (Angers et al., 2006, Jin et al., 2006). Multiple DDB1-CUL4A-associated factors (DCAFs), most of which contain a WD40-repeat domain, have been classified as subunits of CRL4 E3 complexes for recruiting substrates. Interestingly, DDB2, which was originally identified together with DDB1 as a UV damaged-DNA binding protein, anchors to DDB1 like a canonical DCAF protein, but functions to recognize DNA abduct in the nuclear excision repair pathway (Figure 20) (Fischer et al., 2011). With the structural knowledge gathered from the DDB1-hijacking viral proteins, the H-box motif has been found in a number of DCAFs and validated by crystallography. A natural question arises as to whether the H-box motif exists in other DCAFs. However, the lack of an obvious consensus sequence for the H-box motif has made it challenging to find the answer to this question.

Distinct from most DCAFs, Cereblon (CRBN) does not contain a WD-40-repeat domain. Instead, it consists of a seven-stranded β-sheet NTD, a bundle domain composed of seven α-helices (HBD), and an eight-stranded β-sheet CTD (Petzold et al., 2016, Matyskiela et al., 2016). Unlike DCAFs, CRBN uses its HBD with its α-helices to bind DDB1 in the cavity between BPA and BPC. Its CTD domain is involved in the recruitment of native substrates like MEIS2 via a conserved binding pocket. Among all DCAFs, CRBN stands out by being the target of thalidomide, which has a notorious history in biomedicine, but is now repurposed to treat multiple myeloma (see Chapter 4).

The combination of structural, biochemical, and proteomic approaches has helped delineate the composition and architecture of the CRL E3 superfamily in great details. Although there might be outliers unknown to us, most CRL complexes are expected to assemble following the structural principles described above. In contrast to the common architecture shared among different CRL E3s, the mechanisms by which these ubiquitin ligases recognize their specific substrates in response to different cellular cues are incredibly diverse.

## 2.6 Synopsis of Substrate Recognition by CRLs

To achieve spatially and temporally controlled degradation, a substrate needs to be recognized by its cognate E3 with high specificity. This interaction is often mediated by a short linear sequence motif, termed degron, present on the substrate. It has been challenging to identify substrates for different E3 ligases and to fully understand how their recognition is regulated. To date, some substrates have been found to have unmodified degrons that are recognized by the E3 either directly or with the help of a small molecule (Chapter 4.2), while others utilize post-translationally modified degrons to engage their E3 ligases. In some emerging cases, the whole globular domain of a substrate protein contributes to the specific interaction. Moving beyond the know degrons, it is expected that more regulatory mechanisms will be revealed in the future.

## 2.7 PTM-Dependent Substrate Degron Recognition

Protein phosphorylation is one of the most common forms of post-translational modification and is involved in regulating essentially all cellular processes. Many degrons are under the control of this PTM. One classic example is the cell cycle regulatory protein, Cyclin E, which has not just one, but two phosphorylated degrons. These phospho-degrons are specifically recognized by FBXW7, a substrate receptor for CRL1 that is responsible for the degradation of cyclin E (Welcker

and Clurman, 2008). The FBXW7 F-box protein contains a D-box dimerization domain N-terminal to its F-box motif and a C-terminal eight β-sheets WD40 propeller. The FBXW7 D-box promotes the dimerization of the F-box protein, which enables simultaneous binding of two different phosphorylated degrons that independently dock to the "top" surface of the β-propeller in an extended conformation (Figure 21).

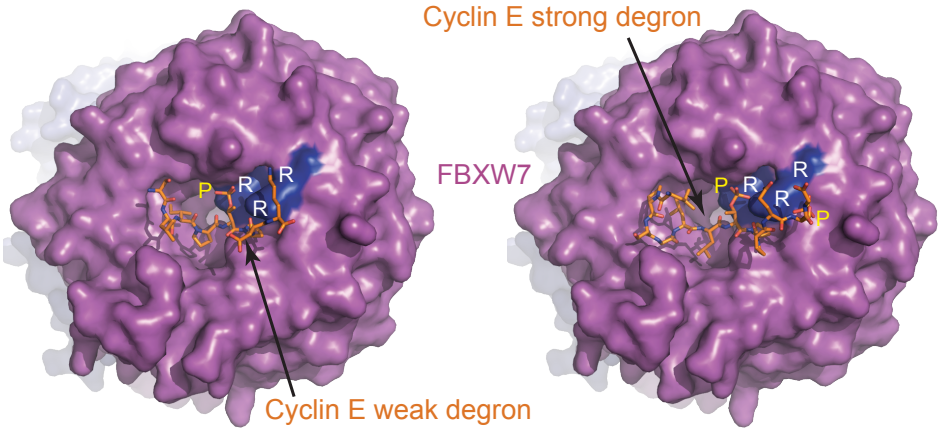



Figure 21. Structural basis of Cyclin E degrons docking to FBXW7. The strong and weak Cyclin E degrons bind FBXW7. R: Arg, P: phosphate. Phosphate-binding arginine residues of FBXW7 are colored in blue, while the degron is in orange sticks and the F-box is displayed in surface representation in magenta (PDB: 2OVR, 2OVQ).

The two Cyclin E degrons share similarities, but also differ in certain aspects. Both phosphorylated degrons have been co-crystalized with FBXW7 (Hao et al., 2007). The degron located at the C-terminus of Cyclin E has three phosphorylated residues and is characterized as a strong degron due to its nanomolar affinity to FBXW7. This tight interaction can be easily explained by the numerous contacts that the degron makes with the substrate receptor, including both polar interactions and van der Waals packings. It is notable that two out of the three phosphate groups present on the degron are recognized by three arginine residues that are strictly conserved among FBXW7 orthologues. By contrast, the N-terminal degron of Cyclin E features only one phosphorylated threonine residue and makes fewer interactions with FBXW7, which leads to weaker binding. Interestingly, the three conserved FBXW7 arginine residues are frequently

mutated in human cancers, rendering the E3 incapable of promoting the degradation of Cyclin E and possibly other substrates. Despite the well elucidated mechanism of the FBXW7-Cyclin E interaction, the necessity of having two degrons on Cyclin E for its productive degradation remains unclear. A cell-based study suggests that the strong degron alone is sufficient for substrate degradation (Welcker and Clurman, 2008).

Proline hydroxylation is another well-known, albeit less common, form of PTM that is involved in degron regulation. Hypoxia-Inducible Factor 1-alpha (HIF1α) contains a critical proline in its oxygen-dependent degradation domain (ODD) that gets hydroxylated in the presence of oxygen (Ivan et al., 2001). This normoxia-associated modification allows VHL, a BC-box protein, to bind, ubiquitinate and target HIF1α for degradation. Under low oxygen conditions, HIF1α is spared from degradation and functions as a transcription factor activating angiogenic gene expression. The crystal structure of VHL-EB-EC in complex with a partial ODD peptide from HIF1α elucidated how the post-translational modification dictates interaction (Figure 22) (Min et al., 2002). The elongated HIF1α peptide adopts a β-strand-like conformation and interacts with the β-domain of VHL in a bipartite manner. Within the N-terminal segment of the HIF1α peptide, the hydroxyproline is deeply embedded at the interface, forming multiple van der Waals contacts and hydrogen bonds with highly conserved residues of the BC-box protein. This interaction is substantiated by the backbones of the amino acids neighboring the proline, which interact with VHL via additional hydrogen bonds. At the C-terminal segment, there are a few interactions that

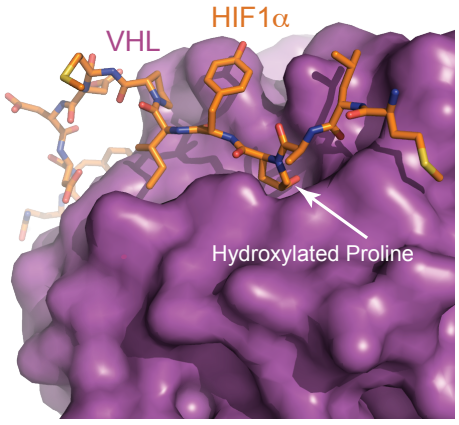



Figure 22. Structure basis of VHL binding HIF1α degron. Recognition of HIF1α degron with a hydroxylated proline by VHL. The degron binding pocket of VHL is shown in magenta surface representation, while the HIF1α peptide is displayed in orange sticks (PDB: 1LM8).

do not seem essential for binding to VHL. The importance of hydroxyproline interface is highlighted by the clustering of cancer inducing mutations involved in von Hippel Lindau disorder around this site.

## 2.8 Native Substrate Degron Recognition

Although many degrons are post-translationally modified, substrate-E3 interaction can be regulated not only on the substrate side, but also on the E3 side. Regulation on the E3 side is best exemplified by KEAP1, a CUL3 substrate receptor (Yamamoto et al., 2018). Under normal cellular conditions, KEAP1 seizes the native transcription factor NRF2 in the cytoplasm and mediates its constitutive ubiquitination and degradation without PTM. When the cell goes through oxidative stress, two reactive cysteines on the IVR domain of KEAP1 are modified, which alters the topological configuration of the substrate receptor domains. The resulting E3 is inactivated and can no longer bind and ubiquitinate NRF2. This allows the transcription factor to translocate into the nucleus and upregulate the expression of cytoprotective genes in response to oxidative insults.

The proper positioning of the two kelch repeat domains from the KEAP1 dimer is essential

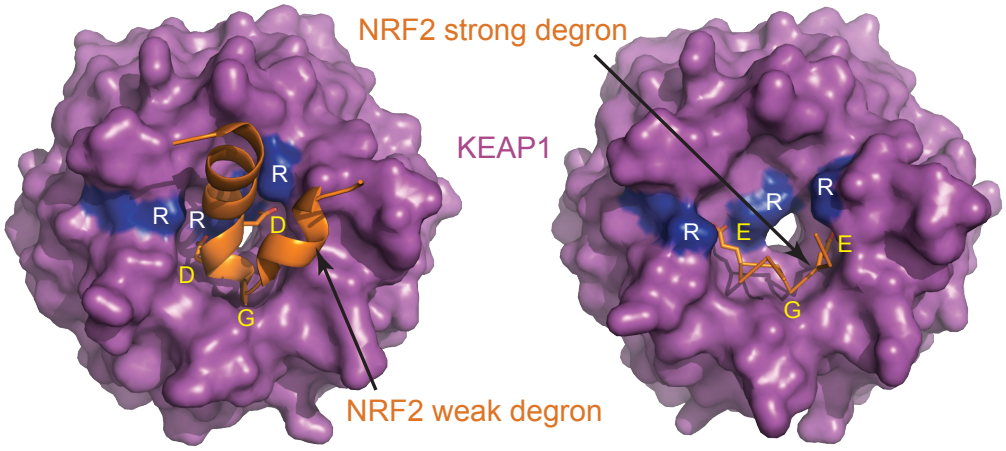



Figure 23. Structural mechanisms of NRF2 degrons binding to KEAP1. Both strong and weak degrons of NRF2 are shown. R: Arg, D: Asp, E: Glu, G: Gly. Arginine residues of KEAP1 responsible for binding negatively charged NRF2 degron residues are colored in blue (PDB: 3WN7, 1X2R).

for productive binding of NRF2, which holds two degron motifs, one with a low affinity and the other with a high affinity to the E3. The structure of the KEAP1 kelch repeat domain has been determined in complexes with both degrons (Figure 23) (Padmanabhan et al., 2006, Fukutomi et al., 2014). As expected, the two degrons bind to the same "top" surface of KEAP1 kelch repeat propeller, although they seem to adopt different conformations. The weak degron contains 35 amino acids and forms a three-helices structure, whereas the strong degron is only 9 amino acids long and folds into a β-hairpin upon binding to the E3. The only common feature shared by both degrons is a central glycine residue preceded by a negatively charged amino acid. These two residues are located at the tip of both degron structures, anchoring themselves deep into the substrate-binding pocket of the E3. Importantly, KEAP1 employs multiple conserved arginine residues to stabilize the two degrons, two of which are dedicated to interact with the negatively charged amino acid preceding the central glycine residue.

Multiple cancer related loss-of-function mutations have been found in NRF2, many of which are localized within the low affinity degron. These mutations correlate with either disruption of the three-helix structure or steric hindrance introduced by a bulky side chain. It is still unclear how the full-length NRF2 substrate interacts with KEAP1 dimer at the structural level. Depicting how NRF2 is positioned relative to the E3 and determining which lysine residues are ubiquitinated would clarify a significant mystery in the field.

## 2.9 GLOBULAR SUBSTRATE PROTEIN RECOGNITION

Although degron-mediated substrate-E3 interaction has become a widely accepted dogma, an increasing number of studies have revealed an alternative strategy for certain CRL E3s to recognize their cognate substrates with high specificity. Perhaps the best example comes from the mammalian cryptochrome proteins, CRY1 and CRY2, which are central components of the

circadian clock in mammals (Takahashi, 2017). The mammalian circadian rhythm is an internal timing system that synchronizes physiological processes to the ~24 hour solar day. In all mammalian cells, the circadian clock is driven by a transcription-translation negative feedback loop, in which the CRY1/2 and PERIOD proteins heterodimerize and suppress their own gene expression. Protein degradation plays an important role in oscillating the clock by periodically removing both proteins, thereby, alleviating their inhibitory effects. While the PERIOD proteins are polyubiquitinated by $CRL1^{\beta\text{-}TrCP}$, which recognizes their phosphorylated degrons, the CRY1/2 protein are destabilized by $CRL1^{FBXL3}$ without an obvious degron (Shirogane et al., 2005, Busino et al., 2007, Siepka et al., 2007, Godinho et al., 2007). Similar to their orthologs in insects and plants, mammalian CRY1/2 adopt a large globular fold with a deep binding pocket for flavin adenine dinucleotide (FAD) (Xing et al., 2013). The structure of the mammalian CRY2 in complex with FAD shows a partially solvent-exposed pocket, which differs from the closed pocket seen in its plant and insect orthologues (Figure 24). In the crystal structure of the CRY2-FBXL3-SKP1 complex, the LRR domain of FBXL3 adopts an expected arch-shaped structure, whose concave surface wraps around the CRY2 globular domain, burying many residues that are only spatially but not sequence-wise connected. A surprising and crucial element of the FBXL3-CRY2 interaction involves the C-terminal tail of the F-box protein, which inserts into the FAD-binding pocket of CRY2. This unexpected interface strongly suggests that FAD might be able to compete with the ubiquitin ligase and protect CRY2 from polyubiquitination. Moreover, the surface area of CRY2 involved in binding FBXL3 overlaps with PERIOD2-binding interface, indicating that the PERIOD proteins might also play an antagonistic role in keeping the E3 ligase in check (Nangle et al., 2013, Schmalen et al., 2014). Because the cellular circadian clock can be entrained by many signals, such as metabolism and hormones, the complex binding mode of FBXL3-CRY2 might

have evolved to allow the single substrate-E3 interacting pair to be regulated through multifaceted mechanisms. As more substrate-E3 interactions are mechanistically interrogated, it is expected that a wider variety of regulatory and structural factors will be revealed beyond the simple degron.

## 2.10 Synopsis of Regulation of CRLs by NEDD8 Modification

As the central ubiquitin ligase machineries regulating diverse cellular pathways, CRLs rely on a multitude of substrate receptors to recognize and recruit their specific substrates. How do these interchangeable substrate receptor subunits share the common cullin scaffolds without interfering with each other's function? How is the ubiquitin ligase activity of the resulting E3 complexes modulated in the cell? Since CRLs were discovered, a battery of cullin-interacting proteins has been identified as important cellular factors that coordinate CRL complex assembly and control their ubiquitin ligase functions. The structural biology approach has not only helped establish the structural framework for investigating the regulation of CRL E3s, but also revealed the detailed mechanisms for several key steps.

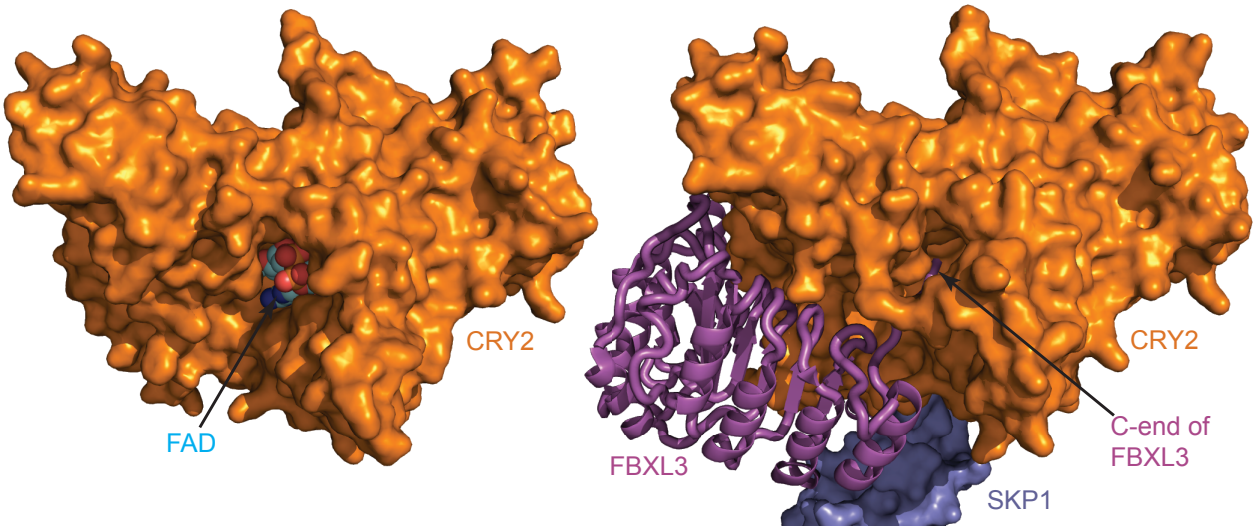



Figure 24. FAD binding to human CRY2 and the complex structure of SKP1-FBXL3-CRY2. CRY2 and SKP1 are shown in surface representation in orange and blue, while FBXL3 is displayed in magenta cartoon fashion, and the FAD is represented in cyan spheres, colored by element (PDB: 4I6G, 4I6J).

## 2.11 NEDD8-MODIFIED CRLS

Many cellular enzymes catalyzing a form of post-translational modification are themselves subject to the same modification. For example, protein kinases are often activated by phosphorylation. CRL E3s follow this trend with a slight variation. All cullins can be modified by the ubiquitin-like molecule, NEDD8, at a specific lysine residue in their C-terminal WHB domain, which is close to the RBX1/2 binding site (Hori et al., 1999). This form of cullin modification, often referred to as neddylation, is conserved from fungi to humans and plays a role in stimulating the E3 activity of CRLs. Although cullin neddylation is not essential in budding yeast, it has been shown to alleviate the autoinhibition of CRLs through augmenting CRL-E2 interaction, closing the gap between the CRL-bound substrate and RBX1-bound E2, and promoting the amide bond formation at the E2 active site (Saha and Deshaies, 2008, Yamoah et al., 2008).

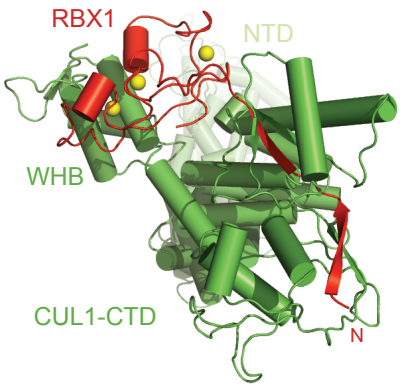



Figure 25. Structure basis of RBX1 binding to CUL1-CTD. Interaction between CUL1-CTD (in green) and RBX1 (in red) viewed 90 degree from Figure 26 (PDB: 1LDJ). The N-terminus of RBX1 is indicated by "N". Yellow spheres represent zinc ions.

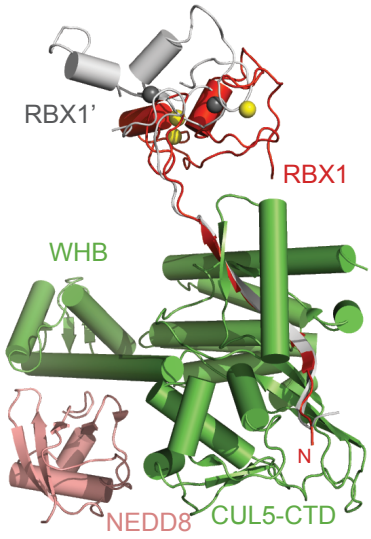



Figure 26. Two structures of RBX1 in the presence of NEDD8~CUL5-CTD. Dislodging of RBX1 RING domain from CUL5-CTD pocket upon cullin neddylation. Two orientations of the RBX1 RING domain captured in the crystal structure are shown. NEDD8 is shown in peach, CUL5-CTD is in green, the different conformations of RBX1 are shown in red and gray. Yellow spheres represent zinc ions (PDB:3DQV).

A major breakthrough in our understanding of the effect of cullin neddylation came from the crystal structure of a NEDD8 modified CUL5-CTD-RBX1 complex (Figure 25 and 26) (Duda et al., 2008). Upon NEDD8 conjugation, the C-terminal portion of the CUL5-CTD undergoes a large degree

of rotation, which reorients the WHB domain relative to the rest of the cullin scaffold. Because the WHB domain and its preceding long α-helix are responsible for cradling and stabilizing the globular RING-domain of RBX1 in the unmodified form of cullin, this neddylation-induced conformational change releases the RING-domain of the catalytic subunit from the cullin CTD. Due to the stable intermolecular β-sheet formed between the N-terminal β-strand of RBX1 and the α/β sub-domain of CUL5-CTD, RBX1 remains bound to the cullin scaffold with its RING-domain gaining a significant degree of freedom to move around. This topological change of the cullin-RBX1 complex is thought to help bring the RBX1-bound E2 closer to the substrate anchored on the substrate receptor.

Just like protein ubiquitination, cullin neddylation requires the actions of the NEDD8-specific E1, E2 (UBC12), and E3 (DCN1 and DCN1 paralogues) enzymes (Liakopoulos et al., 1998, Osaka et al., 1998, Kurz et al., 2005). Similar to all ubiquitin-specific E2s, UBC12 (a.k.a. UBE2M) features a canonical E2 catalytic core domain, harboring an active site cysteine residue that can form a thioester bond with NEDD8 after it is activated by the E1 enzyme. Distinct from most ubiquitin-specific E2s, however, UBC12 contains an N-terminal extension sequence, whose extreme N-terminus has been shown to be acetylated. Remarkably, the acetylated UBC12 N-terminal extension adopts an α-helical conformation and specifically interacts with the

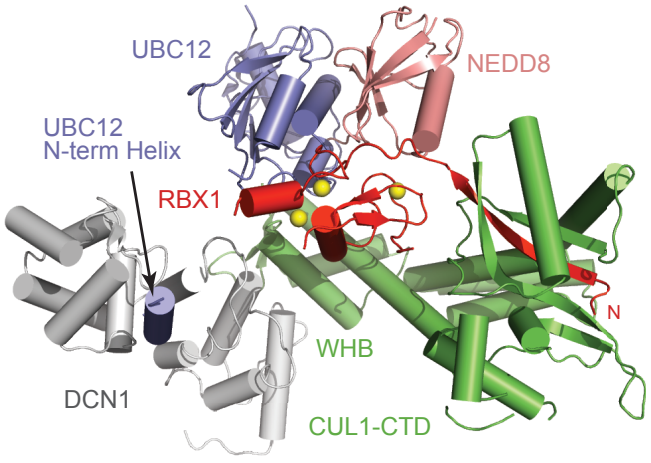



Figure 27. Complex structure of CUL1-CTD-RBX1 bound to NEDD8-charged UBC12 and DCN1. The linker between the N-terminal helix and catalytic domain of UBC12 is disordered. CUL1-CTD is colored in green, RXB1 is in red, NEDD8 is in peach, DNC1 is in gray, and UBC12 is in blue. Yellow spheres represent zinc ions (PDB:4P5O).

neddylation E3 protein, DCN1 (Scott et al., 2011). With two EF hand-like subdomains juxtaposed together, DCN1 is a compact all helical protein with a slightly elongated shape (Figure 27). At the center of the protein is a hydrophobic pocket, which can specifically recognize the acetyl group of UBC12 N-terminus and the first methionine residue. Separate from this pocket, DCN1 also features a surface area that is able to engage cullin C-terminal WHB domain. Together, these interactions represent major interfaces through which DCN1 recruits the NEDD8-specific E2, UBC12, to catalyze the NEDD8 transfer reaction.

Strictly speaking, the NEDD8 E3 ligase function is performed by DCN1 in conjugation with RBX1, which plays a critical role in docking and activating the NEDD8-charged UBC12 catalytic core for cullin neddylation. Given the structural similarity of the catalytic domain between UBC12 and ubiquitin-specific E2s, the NEDD8 E2 is expected to interact with RBX1 RING domain in a similar fashion as ubiquitin-specific E2s to RING E3s. A simple modeling of a UBC12-RBX1 complex in the context of the unmodified cullin-RBX1 structures, however, readily reveals a long distance between the catalytic cysteine residue of UBC12 and the cullin neddylation site. For the cullin's lysine residue to attack the thioester bond formed between the UBC12 active site cysteine and the carboxyl terminus of NEDD8, the two residues have to be close to each other. This geometrical requirement strongly suggests that the RBX1 RING domain has to be re-oriented before cullin neddylation can take place. Indeed, the crystal structure of an isolated CUL1-CTD-RBX1 complex revealed that the RING-domain of RBX1 can be disengaged from its binding site on the CUL1-

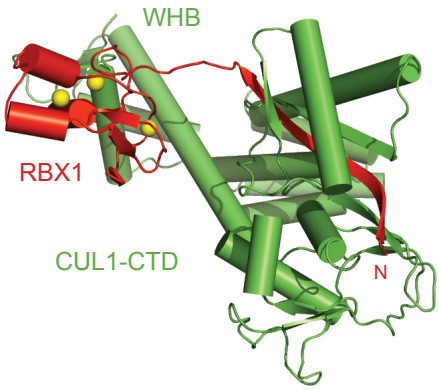



Figure 28. Yet another CUL1-CTD-RBX1 complex structure. A new position of RBX1 RING domain revealed by a CUL1-CTD-RBX1 complex structure. In cartoon representations, CUL1-CTD is shown in green and RBX1 is displayed in red. Yellow spheres represent zinc ions (PDB:3RTR).

CTD in the absence of neddylation (Figure 28) (Calabrese et al., 2011). When UBC12 was modeled onto RBX1 in this structure, the gap between the UBC12 catalytic site and the cullin neddylation site is mostly closed.

The final picture of cullin neddylation has been depicted by the crystal structure of CUL1-CTD-RBX1 in complex with DCN1 and NEDD8-charged UBC12 (Figure 28) (Scott et al., 2014). In this structure, the CUL1 C-terminal WHB domain and its preceding long α-helix is shifted away from the rest of the CUL1-CTD, which allows the RING domain of RBX1 to adopt yet another orientation. Resembling the previously reported docking model of ubiquitin-charged E2 to RING E3s, NEDD8-charged UBC12 is anchored to the RBX1 RING domain and their compact structure is stabilized by a "linchpin" arginine residue unique to RBX1. Importantly, the NEDD8 molecule conjugated to the E2 also makes contacts with the linker sequence that connects the N-terminal β-strand of RBX1 to its RING domain, thereby, optimally positioning the NEDD8 transfer module so that the catalytic site of the NEDD8 E2 is placed right next to the CUL1 neddylation site. Consistent with this notion, the interface between UBC12 catalytic domain and CUL1-WHB, which harbors the neddylation site, is kept minimal. Although DCN1 is also co-crystallized with the complex, it does not make direct interacts with the UBC12 catalytic domain (Figure 27). A flexible linker between the catalytic domain of the NEDD8 E2 and its N-terminal extension, which stably binds DCN1, is thought to accommodate the movement of the NEDD8 transfer module formed between RBX1 and NEDD8-charged UBC12 catalytic domain relative to the cullin scaffold.

## 2.12 CAND1 AND CULLIN CYCLE

Cullin-associated and neddylation-disassociated protein 1 (CAND1) was the first cullin-binding protein identified that does not belong to the basal subunits of CRLs (Zheng et al., 2002a, Liu et

al., 2002). It is a 120 kDa HEAT-repeat proteins that can form a stable complex with native, but not neddylated, cullin-RBX1 catalytic core. Interestingly, CAND1 binding seems to inhibit CUL1 from binding SKP1 and the substrate receptor F-box proteins, suggesting that CAND1 and SKP1-F-box proteins are mutually exclusive on the CRL1 scaffold. The crystal structure of a CAND1-CUL1-RBX1 complex unveiled the structural basis of all these biochemical activities of CAND1 (Figure 29) (Goldenberg et al., 2004). The 120 kDa protein adopts a super-helical structure with 27 consecutively stacked HEAT repeats that together form a long but highly sinuous fold. By curving around the entire CUL1-RBX1 structure, CAND1 grasps onto CUL1 like a two-pronged clamp. Importantly, CAND1 sports a β-hairpin projecting out of one of its HEAT repeats and reaching to the SKP1-binding site of the cullin scaffold. In doing so, CAND1 is able to compete with SKP1 for binding to the N-terminal end of CUL1. At the opposite end, the first two HEAT repeats of CAND1 closely pack against the WHB domain of CUL1, burying the neddylation site lysine residue. This suggests that CUL1 neddylation would sterically block CAND1 from binding.

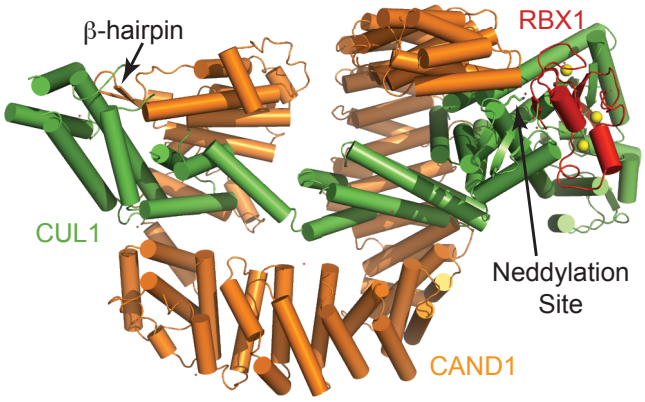



Figure 29. Structure basis of CAND1-CUL1-RBX1 interaction. CAND1 wraps around CUL1-RBX1, burying the cullin neddylation site lysine residue and blocking SKP1-binding site with a β-hairpin. CUL1, RBX1, and CAND1 are shown in cartoon representation in green, red, and orange respectively. Yellow spheres represent zinc ions (PDB: 1U6G).

Because SKP1-F-box proteins are responsible for recruiting substrates and CUL1 neddylation is thought to activate the E3 complex, the binding mode and biochemical properties of CAND1 seem to suggest that it acts as an inhibitor of CRL1. However, genetic studies indicate that CAND1 plays a positive role in regulating substrate ubiquitination and degradation by the E3

machinery. A growing body of evidence has helped raise an interesting model designating CAND1 as an exchange factor of CRL substrate receptors (Pierce et al., 2013, Reitsma et al., 2017, Liu et al., 2018, Wu et al., 2013, Zemla et al., 2013). In this model, CAND1 can promote the disassembly of a SKP1-F-box protein complex from CUL1-RBX1, thereby, allowing another SKP1-F-box protein complex to engage the cullin scaffold. The structure of the CAND1-CUL1-RBX1 complex supports the notion that NEDD8 conjugation not only stimulates the activity of the E3 complex, but also prevents CAND1 from dislodging an existing SKP1-F-box protein from the cullin scaffold.

In addition to CAND1, two other cellular factors have been documented to regulate a subset of CRLs, an α-helical protein known as Glomulin (GLMN) and a RING-IBR-RING (RBR) protein, HHARI (a.k.a. ARIH1). The gene encoding GLMN is mutated in the hereditary diseases glomuvenous malformations, which are characterized by venous lesions involving glomus cells. GLMN was initially identified as a protein that binds the C-terminus of CUL7, a distinct family member of CRLs (Arai et al., 2003). It was later shown to directly interact with the RBX1 RING domain and block its E3 ubiquitin ligase activity (Tron et al., 2012). The crystal structure of GLMN in complex with RBX1 bound to a fragment of CUL1-CTD revealed that GLMN contains two HEAT-repeats-like sub-domains, which show structural similarity to each other (Figure 30) (Duda et al., 2012). One side of the GLMN C-terminal domain forms an extensive interface with the RING domain of RBX1, masking its E2-binding site. Although the CUL1-CTD

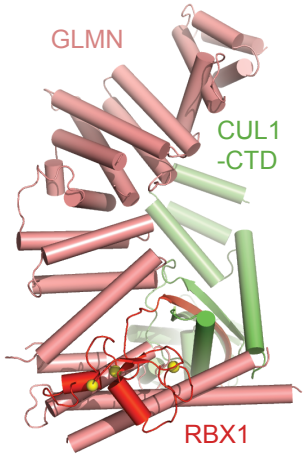



Figure 30. Structural representation of GLMN-CUL1-CTD-RBX1 complex. GLMN binds and blocks the E2-binding surface of RBX1 RING domain, which is flexibly tethered to the CUL1-CTD via an N-terminal β-strand. The complex is shown in cartoon representation with CUL1-CTD in green, RBX1 in red, and GLMN I peach. Yellow spheres represent zinc ions (PDB: 4F52).

fragment was present in the crystal and contacts GLMN C-terminal domain, it plays minimal role in stabilizing the complex formation. Owing to the orientational flexibility of RBX1 RING domain relative to the rest of the CRL1, GLMN binding is compatible with CRL assembly both at the RBX1-cullin interface and the cullin-adaptor-substrate receptor site. Even cullin neddylation showed no effect to GLMN-RBX1 interaction. Overall, GLMN appears to be an RBX1-specific inhibitor. Nonetheless, GLMN only binds a small subset of CRLs in human cells, suggesting that an unknown mechanism is involved in selectively controlling GLMN-RBX1 interaction in the context of the CRL functions.

RBR E3s represent a distinct class of ubiquitin ligases, which are a hybrid of the canonical RING-type and the HECT-type E3s (Zheng and Shabek, 2017). RBRs are characterized by multiple RING domains and thioester intermediates they form with ubiquitin before the modifier is transferred to a substrate. ARH1, a member of the RBR E3s, has recently been identified to be preferentially associated with NEDD8-modified CRL1-CRL3 (but not CRL4) (Scott et al., 2016). Interestingly, it catalyzes mono-ubiquitination of representative substrates of these CRL E3s, which can be further polyubiquitinated by CDC34, the cognate E2 for RBX1. The precise mechanism by which ARH1 coordinates with neddylated CRLs to mediate the ubiquitin transfer reaction awaits future structural studies.

## 2.13 COP9 SIGNALOSOME-CRL INTERACTIONS

Just like protein ubiquitination, cullin neddylation is reversible. Deconjugation of NEDD8 from cullins is catalyzed by an evolutionarily conserved eight-subunits protein complex, known as the COP9 signalosome (CSN). CSN was first identified in plants based on mutants that showed a *constitutive photomorphogenesis* (*COP*) phenotype (Wei et al., 1994). These plant mutants turned out to carry mutations in eight genes, whose protein products form a stable complex with each

subunit sharing sequence homology with one component of the eight-subunit lid complex of the 19S proteasome (Chamovitz et al., 1996, Wei et al., 1994). Among the eight CSN subunits, CSN5 is a zinc-containing metalloprotease that is responsible for cleaving the iso-peptide bond between cullins and NEDD8 (Cope et al., 2002).

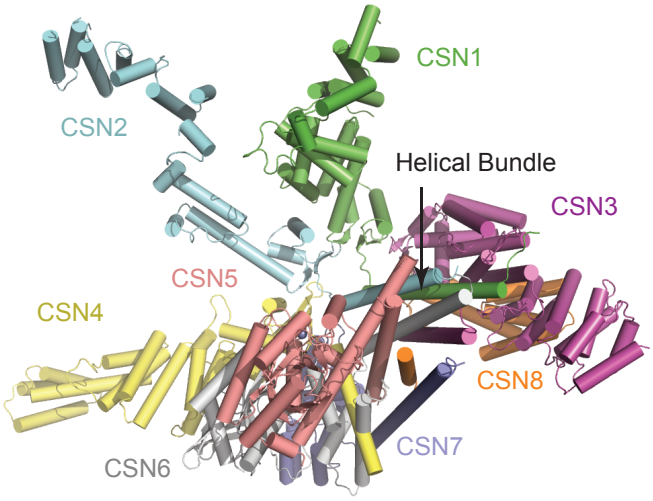



Figure 31. The overall architecture of COP9 signalosome. Each subunit is colored differently (PDB: 4D10).

The assembly mechanism of the COP9 signalosome was first revealed in the crystal structure of the human CSN complex (Figure 31) (Lingaraju et al., 2014). Each of the CSN subunits employs one or two α-helices to build a super-helical bundle, which contributes to the stable assembly of the deneddylase complex. Meanwhile, six of the CSN subunits encircle a horse shoe-shaped ring structure with elongated α-helical PCI (proteasome lid-CSN-initiation factor 3) domains projecting away from the center. CSN5 and CSN6 share a common MPN (MPR1/PAD1) domain with a metalloprotease fold and together form a heterodimer. With their C-terminal regions integrated into the super-helical bundle, these two subunits anchor themselves onto one side of the CSN ring structure. Interestingly, the catalytic site of CSN5 in the CSN holoenzyme was found in an auto-inhibited state, suggesting that CSN has to be activated upon binding to its substrate.

Recent advances in cryo-EM technology have enabled several studies that have shed light on how CSN interacts with different CRL complexes. Despite limited resolution, single particle analysis of CSN bound to neddylated CUL1-RBX1 in complexes with SKP1-SKP2-CKS1 and monomeric SKP1-FBXW7 offered the first glimpse of the CSN-CRL interaction (Enchev et al.,

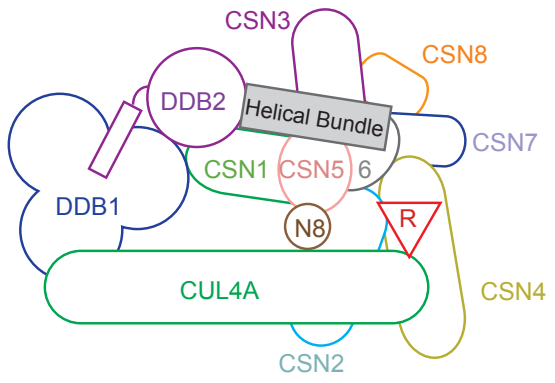



Figure 32. Diagram of CRL4-DDB2-COP9 complex. A schematic drawing of NEDD8 modified CUL4A-RBX1 in complex with DDB1-DDB2 and COP9 signalosome. R: RBX1. 6: CSN6. N8: NEDD8.

2012). In the structural models derived from the EM density maps, CSN2 appears to make major contacts with CUL1-CTD, whereas the distal end of the two F-box proteins are located close to CSN1 and CSN3. Structural modeling and biochemical analysis indicated that CSN competes with both substrates and E2 for binding the E3 platform, thereby, raising the possibility that substrate-loaded CRL1 might protect neddylated CUL1 by preventing CSN from accessing the E3. This notion was subsequently supported by the cryo-EM structure of CSN bound to the neddylated CUL4A-RBX1-DDB1-DDB2 complex (Figure 32) (Cavadini et al., 2016). In this super-assembly, CSN2 not only interacts with CUL4A-CTD, but also sandwiches RBX1 RING domain together with CSN4, thereby, preventing the E3 scaffold from recruiting an E2 molecule. While CSN1 makes specific interactions with DDB1, DDB2 is snugly situated in between DDB1 and the CSN helical bundle. In comparison to the CSN holoenzyme structure, the CSN helical bundle is repositioned to accommodate the CUL4A-DDB1 substrate receptor. Interestingly, despite its topological flexibility, the CSN helical bundle cannot accommodate additional cellular factors that interact with DDB2, corroborating the idea that substrate binding to CRL4 will introduce steric hindrance for CSN binding. Nevertheless, a question remains as to whether CSN has the ability to differentiate variations of CRL substrate receptor in size and shape from the binding of a small degron as part of a flexible substrate polypeptide.

Upon binding to CRL4, CSN undergoes multiple conformational changes to not only adapt to the landscape of its substrate, but also alleviate its auto-inhibition. These changes include

movements of the PCI domains of CSN2 and CSN4 for clamping down to CUL4A-CTD and RBX1 and the translocation of the CSN5-CSN6 dimer to approach NEDD8. Although CRL4 is significantly different from other CRLs, similar structural changes have also been observed in the EM structure of CSN in complex with neddylated Cul1 with SKP1-SKP2-CKS1 bound. Although biochemical analyses have helped identify several structural elements that relay these structural changes to the alleviation of CSN autoinhibition, the detailed structural mechanism underlying CSN5 activation requires structural analysis at a higher resolution (Cavadini et al., 2016, Mosadeghi et al., 2016).

# Chapter 3. CHARACTERIZATION OF KLHDC2-C-END DEGRON

The following work has previously been published and was adapted from: Rusnac, D.-V., Lin, H.-C., Canzani, D., Tien, K. X., Hinds, T. R., Tsue, A. F., Bush, M. F., Yen, H.-C. S., Zheng, N. (2018). Recognition of the Diglycine C-End Degron by $CRL2^{KLHDC2}$ Ubiquitin Ligase. Molecular Cell 72, 813 - 822.

## 3.1 INTRODUCTION

The ubiquitin-proteasome system (UPS) removes aberrant proteins and regulates diverse cellular functions by promoting the turnover of numerous protein substrates (Hershko and Ciechanover, 1998, Goldberg, 2003). The high capacity and selectivity of UPS is conferred by a plethora of ubiquitin E3 ligases, which number in the hundreds for humans (Zheng and Shabek, 2017). Acting at the final step of a three-enzyme cascade, many ubiquitin E3 ligases recognize their cognate substrates through a short linear sequence motif, known as a degron (Lucas and Ciulli, 2017, Mészáros et al., 2017, Guharoy et al., 2016). To achieve high specificity, eukaryotic cells have evolved a variety of mechanisms governing degron-E3 interactions. As discussed in detail in the previous chapter, for many degrons, proper forms of post-translational modifications, such as serine/threonine phosphorylation and proline hydroxylation, have been shown to play a key role in enabling E3 binding (Hao et al., 2005, Hao et al., 2007, Wu et al., 2003, Orlicky et al., 2003, Min et al., 2002). On the other hand, an increasing number of degrons have also been identified to function in a modification-independent manner (Padmanabhan et al., 2006, da Fonseca et al., 2011, Kraft et al., 2005, Zhuang et al., 2009, Uljon et al., 2016).

In the classical N-end rule pathway, both aforementioned degrons have been discovered at the extreme N-terminus of proteins (Tasaki et al., 2012, Varshavsky, 2011). Some of these N-degrons are generated by proteolytic cleavage, whereas others are created by N-terminal modifications. While certain resulting N-terminal amino acids can stabilize a protein, other ones, such as arginine, have been shown to substantially reduce the half-life of a substrate. Previous

studies have identified a family of UBR-box-containing proteins, classified as N-recognins, responsible for recognizing the N-terminal degrons (Tasaki et al., 2009, Tasaki et al., 2005, Xia et al., 2008). The crystal structure of the UBR-box domain from yeast UBR1 has been determined in complex with the N-degron peptide of the cohesion subunit Scc1 (Choi et al., 2010). In this structure, the E3 recognizes the N-end degron of the proteolytically generated Scc1 C-terminal fragment through its free backbone amino group, the side chain of the leading arginine, and the penultimate residue. A similar recognition interface has also been revealed for the mammalian UBR1 ortholog (Matta-Camacho et al., 2010). By recognizing the N-end degrons exposed by endopeptidase cleavage, the N-end rule not only dictates the half-life of full-length proteins, but also participates in protein quality control by eliminating proteolytic products. Like the C-terminal polypeptide fragments generated by proteolysis, the N-terminal segments and early terminated protein products could also impose a threat to the normal functions of the cell.

Recent studies have unraveled a cohort of ubiquitin ligases from the cullin-RING superfamily, which are implicated in the recognition of specific sequence elements embedded in the extreme C-terminus of these truncated polypeptides (Lin et al., 2018, Koren et al., 2018). In 2008, Hsueh-Chi Sherry Yen and Stephen Elledge developed a Global Protein Stability (GPS) assay that allows real-time high-throughput tracking of protein turnover under various conditions in mammalian cells. The GPS reporter system was based on the co-expression of green fluorescent protein (GFP) and red fluorescent protein (RFP) from a single transcript enabled by an internal ribosome entry site (IRES). In this assay, GFP is fused with thousands (~8000-15,483) of human open reading frames (ORFs), while RFP serves as a non-degradable internal control. The GFP/RFP ratio thus indicates the stability of the GFP-fused constructs and is analyzed by flow cytometry.

The relative levels of GFP-fussed proteins, but not RFP control, are expected to change under certain experimental conditions.

To determine the substrates of various CRLs, dominant-negative cullins were expressed in the GPS assay. Intriguingly, Yen et al. found 102 substrates for CRL2 ubiquitin ligase, five of which were early terminated selenoproteins. All selenoproteins contain at least one selenocysteine (Sec), which is encoded by the stop codon UGA (Ambrogelly et al., 2007). With help from a Sec insertion sequence element in the 3' untranslated region, this terminal signal in the mRNA is translated into Sec, allowing the production of the full-length protein (Driscoll and Copeland, 2003). When selenium is scarce, the lack of Sec-transfer RNA causes premature termination of translation. Recent studies have shown that the resulting five early terminated selenoproteins bear C-terminal degrons, which triggers their CRL2-mediated proteasomal degradation (Lin et al., 2015).

This novel protein degradation mechanism, named DesCEND (destruction via C-end degron), relies on select amino acids at key positions within the degrons, which are usually less than ten residues long. Interestingly, one of these C-end degrons is characterized by a strikingly simple diglycine motif at the extreme C-terminus. With a highly degenerate N-terminal region, this diglycine-containing sequence is thought to be recognized by a BC-box protein, KLHDC2, which functions as a substrate receptor of the CRL2 E3 complex. Remarkably, the diglycine C-end degron has been found in multiple classes of DesCEND substrates, including early terminated selenoproteins (SelK and SelS), the N-terminal proteolytic product of a deubiquitinating enzyme (USP1), and a number of full-length proteins (Lin et al., 2018, Koren et al., 2018). *Klhdc2* homozygous knockout mice display an embryonic lethal phenotype (Dickinson et al., 2016).

KLHDC2 has thus emerged as an essential E3 ligase with a multi-faceted role in protein homeostasis.

Similar to other known degrons, the C-end degrons are self-sufficient and can induce proteasomal degradation when fused to otherwise stable proteins. To reveal the mechanism by which a prototypical C-terminal degron is decoded in the DesCEND pathway, I determined the crystal structures of KLHDC2 in complex with three different diglycine C-degron peptides and performed detailed analyses of the interaction interface. Our studies have not only mapped the essential elements dictating the specific interactions between the C-end degron and its cognate E3, but also provided a detailed quantitative understanding of the system and shed light on a unique role of DesCEND that is distinct from the N-end rule pathway.

## 3.2 Mapping Key Elements in SelK C-End Degron

Under certain cellular conditions, such as selenium deficiency, a C-terminal diglycine degron of SelK is generated due to early termination in translation that leads to $CRL2^{KLHDC2}$-mediated proteasomal degradation (Lin et al., 2015). The BC-box protein KLHDC2 contains a C-terminal region responsible for CUL2-Elongin-B/C binding and an N-terminal kelch repeat β-propeller domain, which is commonly used by CRL substrate receptor subunits to engage substrates (Mahrour et al., 2008, Zimmerman et al., 2010, Duda et al., 2011). To formally establish the direct interaction between SelK C-end degron and KLHDC2, I purified the recombinant KLHDC2 kelch domain (hereafter referred to as KLHDC2) and performed GST pull-down assays using GST-fused SelK C-end degron that are 8 or 12 aa long. KLHDC2 displayed robust interaction with both degron peptides, but not

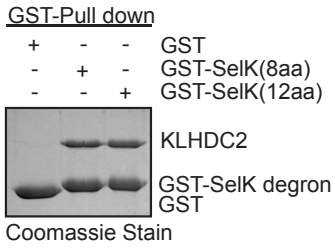



Figure 33. GST-Pull down confirming direct interaction between KLHDC2-SelK degron. Validation of direct interaction between purified KLHDC2 kelch repeat domain and an 8 aa and a 12 aa SelK C-end degron fused to GST in a GST-pull down assay.

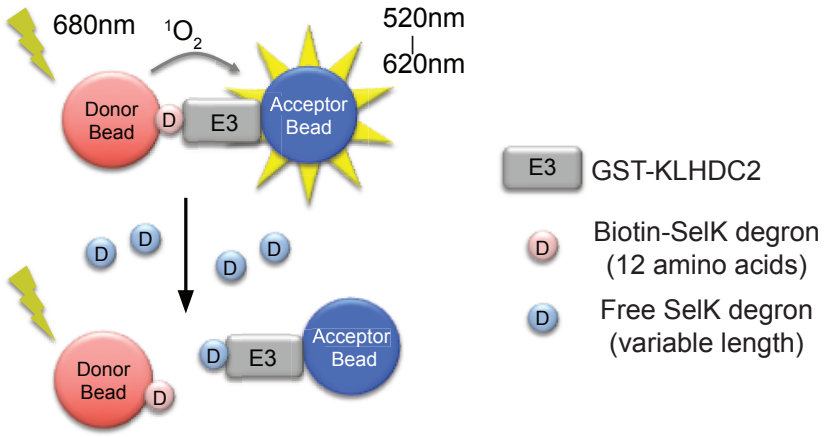
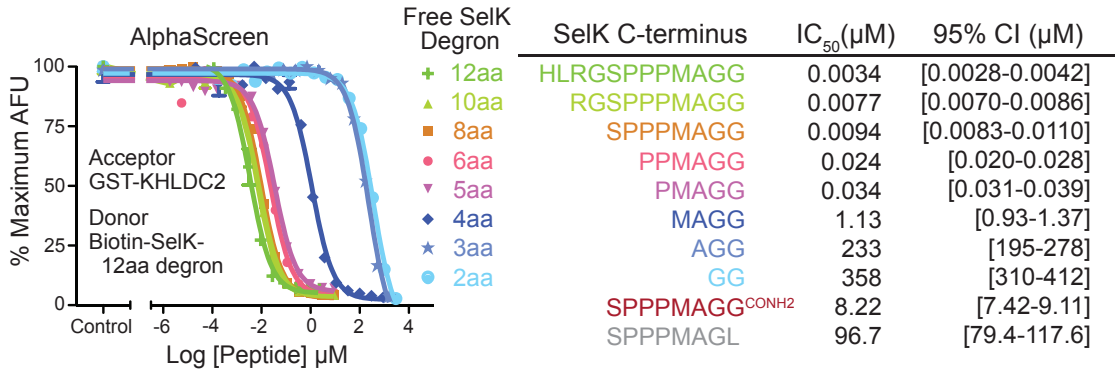



Figure 34. Schematic representation of AlphaScreen-based competition assay. The AlphaScreen-based competition assay designed for assessing the affinity of SelK C-end degron peptides with KLHDC2.

GST alone (Figure 33). This result indicates that the BC-box protein can directly recognize the SelK C-end degron and that the kelch repeat domain of KLHDC2 is sufficient for this interaction.

To map the minimal length of the SelK C-end degron for high affinity binding to KLHDC2, I established the degron peptide-E3 interaction in an Amplified Luminescence Proximity Homogenous Assay (AlphaScreen) with a biotinylated 12 aa SelK degron peptide and GST-fused KLHDC2 immobilized on the donor and acceptor beads, respectively (Figure 34). I then carried out competition experiments with label-free SelK degron peptides of varying lengths ranging from 12 to 2 amino acids. These peptides share the critical extreme C-terminal diglycine motif with variably shortened N-terminal ends. By titrating the concentration of the competing peptides, I was




| SelK C-terminus | $IC_{50}$(μM) | 95% CI (μM) |
|---|---|---|
| HLRGSPPPMAGG | 0.0034 | [0.0028-0.0042] |
| RGSPPPMAGG | 0.0077 | [0.0070-0.0086] |
| SPPPMAGG | 0.0094 | [0.0083-0.0110] |
| PPMAGG | 0.024 | [0.020-0.028] |
| PMAGG | 0.034 | [0.031-0.039] |
| MAGG | 1.13 | [0.93-1.37] |
| AGG | 233 | [195-278] |
| GG | 358 | [310-412] |
| SPPPMAGG$^{CONH2}$ | 8.22 | [7.42-9.11] |
| SPPPMAGL | 96.7 | [79.4-117.6] |

Figure 35. Functional mapping of the SelK C-end degron. AlphaScreen competition assay for assessing the affinity of SelK C-end degron peptides with variable lengths to KLHDC2. The dose response curves of the peptides ranging from 2aa to 12aa are colored in the same scheme as the peptide labels. The $IC_{50}$ value and the 95% confidence interval for each peptide is listed together with its amino acid sequence at the bottom table. AFU: arbitrary fluorescence units. Data are measured in triplicates and represented as mean ± SEM.

able to derive dose response curves and obtain the $IC_{50}$ values for each peptide. Due to the extremely low concentration of the immobilized components, the $IC_{50}$ value determined from our competition assay can be approximated to $K_d$. To our amazement, the 12 aa SelK degron peptide abolished the AlphasScreen signal with an $IC_{50}$ of ~3 nM (Figure 35). Decreasing the length of the degron peptide from 12 to 5 aa caused a gradual and slight decrease in affinity. The 5-residue degron, nevertheless, retained an affinity toward the E3 ligase in the low-nanomolar range. I observed a dramatic 30-fold increase in $K_d$ when the peptide length is reduced to 4 amino acids, indicating the loss of a critical contact between the degron and the E3. Interestingly, even diglycine was able to bind KLHDC2 and compete with the biotinylated 12 aa degron that is immobilized on the AlphaScreen donor beads (Figure 35). The extreme C-terminal diglycine motif, therefore, contributes a substantial amount of binding energy to the degron-E3 interaction. Together, these results revealed a strong correlation between the length of the SelK degron peptide and its strength in binding KLHDC2, and established the minimal degron length of 5 amino acids for high affinity interaction. The remarkably tight binding between the 12 amino acids SelK degron and KLHDC2 was further confirmed by the dissociation constant of their direct interaction measured with Octet

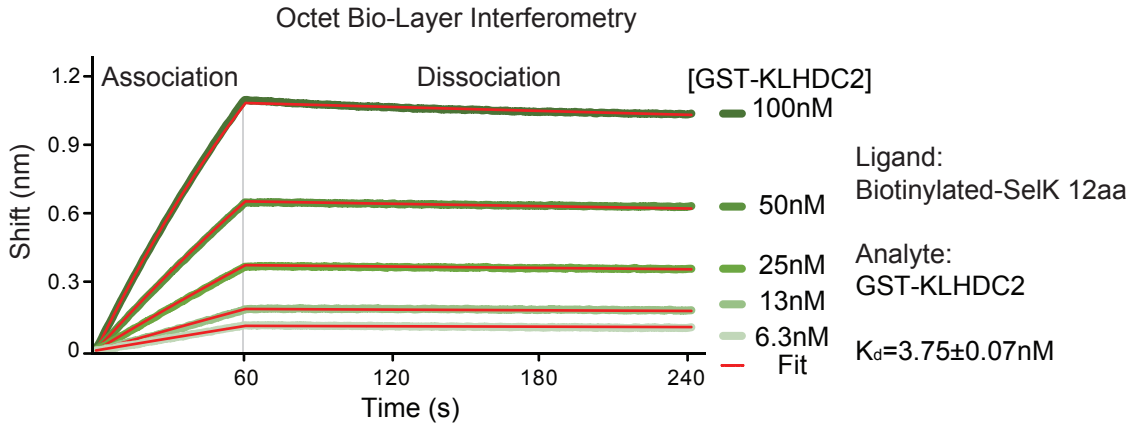



Figure 36. Validation of high affinity binding between KLHDC2 and SelK 12 aa peptide via BLI. Affinity determination of KLHDC2 binding to the 12 amino acids SelK degron peptide by Octet BioLayer Interferometry. Green lines represent the kinetics of association and dissociation of GST-KLHDC2 to biotinylated SelK peptide, which is immobilized on the probe. Red lines represent global fit of the binding curves with a 1:1 ligand model. The dissociation constant ($K_d$) was calculated from the $k_{on}$ and $k_{dis}$ values.

BioLayer Interferometry, which yielded a $K_d$ value of ~4 nM (Figure 36).

The full-length SelK protein contains three additional amino acids beyond the diglycine motif, which allows the polypeptide to evade DesCEND. Recent studies have revealed a requirement for the digycline motif to terminate a polypeptide in order to give rise to a functional C-end degron (Lin et al., 2015). hypothesized that the free extreme C-terminal backbone carboxyl group, along with the tandem glycine residues, must play a role in E3 binding. Using our AlphaScreen-based competition assay, I compared the affinity of the 8 aa SelK degron peptide that was C-terminally amidated to that of the unmodified peptide. Consistent with our hypothesis, this simple modification of the C-terminal carboxyl group profoundly decreased the affinity of the degron peptide by nearly a 1000-fold (Figure 37). Using an 8 aa peptide with the last glycine residue

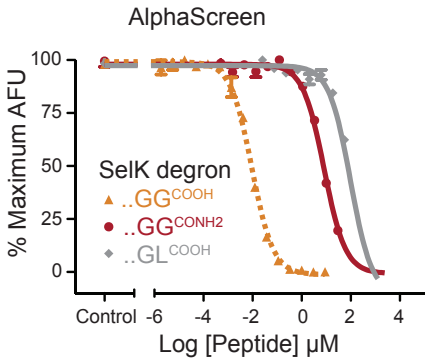



Figure 37. Validation of the importance of carboxyl group and diglycine motif in degron binding. AlphaScreen competition assay for assessing the affinity of the 8 aa WT (…$GG^{COOH}$), C-terminally amidated (…$GG^{CONH2}$), and mutant (...$GL^{COOH}$) SelK C-end degron peptides to KLHDC2. The $IC_{50}$ values and the 95% confidence intervals for all three peptides are listed together with their amino acid sequence at the bottom table in (Figure 35).

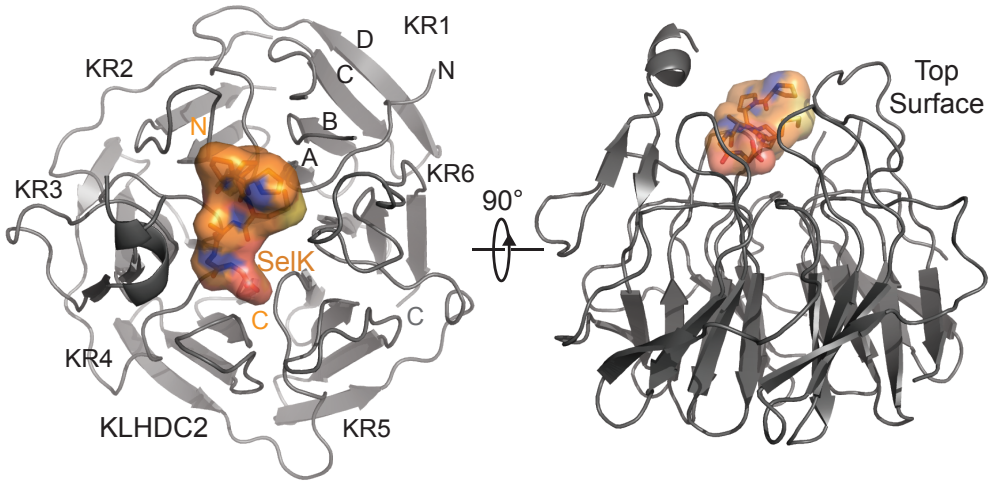



Figure 38. Structure basis of SelK degron recognition by KLHDC2. Two orthogonal views of KLHDC2 kelch repeat domain in complex with a SelK C-end degron peptide. The β-propeller domain of KLHDC2 is shown in grey ribbon with its six blades labeled "KR1" to "KR6". The four anti-parallel β-strands in repeat 1 are labeled "A" to "D". The SelK C-end degron peptide is shown in orange sticks and surface representation. The N- and C-termini of the two polypeptides are labeled "N" and "C" (PDB:6DO3).

mutated to leucine, I further characterized and validated the importance of the diglycine residues, which has been previously established in cell-based assays (Lin et al., 2018). In contrast to the single digit nM affinity of the wild type SelK degron peptide, the glycine to leucine mutation increased the $IC_{50}$ value to ~100 μM (Figure 37).

## 3.3 Crystal Structure of KLHDC2 Bound to SelK Degron Peptide

In agreement with the high affinity binding determined from the AlphaScreen assay, KLHDC2 forms a stable complex with the SelK degron peptide as detected by native protein mass spectrometry (Figure S1). I crystallized and determined the structure of the KLHDC2 kelch repeat domain in complex with the 8 aa SelK C-end degron peptide (Table 1). Akin to most kelch repeats, KLHDC2 adopts a six-bladed β-propeller fold with each blade composed of four antiparallel β-strands (named A to D from inner to outer position) (Figure 38). Conventionally, the side of a β-propeller fold constructed by the loops linking strands A and D and strands B and C is referred to as the "top" surface (Sprague et al., 2000). These loops are highly variable between different β-propeller proteins and are frequently involved in protein interaction or enzymatic catalysis. In KLHDC2, these loops form a deep binding pocket, which is responsible for recognizing the C-end

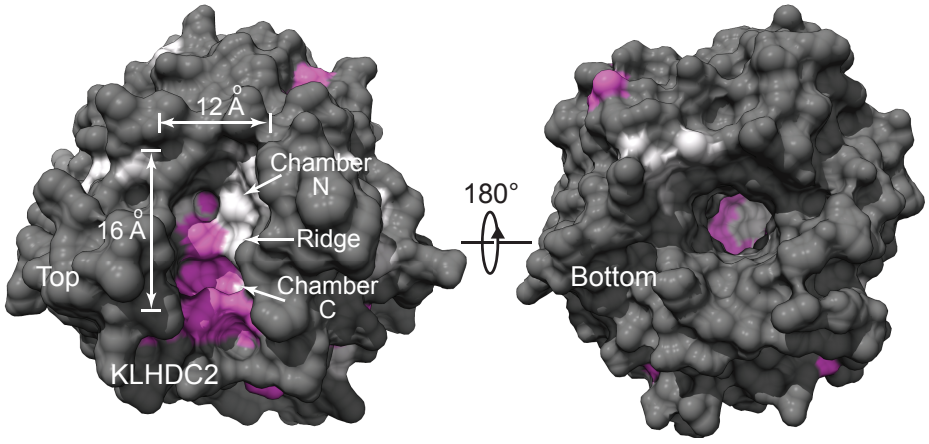



Figure 39. Conservation surface mapping of the KLHDC2. The kelch repeat domain is in its top and bottom views. Residues that are strictly (100%) and highly (80-100%) conserved are colored in magenta and light-grey, respectively. The rest of the molecule is colored in dark grey. The C-end degron-binding pocket is annotated for its dimension, two separate chambers, and the middle ridge (PDB:6DO3).

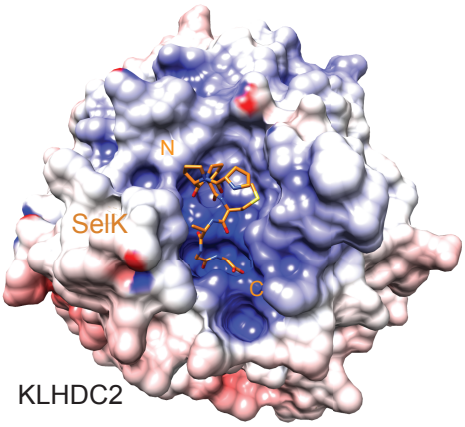



Figure 40. Electrostatic surface potential map of KLHDC2. The KLHDC2 kelch repeat domain is in the same view as shown in Figure 39 and colored based on electrostatic protentional, positive in blue and negative in red. The SelK C-end degron is shown in orange sticks (PDB:6DO3).

degron.

KLHDC2 is evolutionarily conserved from amoeba to humans (Figure S2). Surface conservation mapping of its kelch repeat domain reveals a group of solvent-exposed and highly conserved residues clustered at the bottom and one side of the top surface pocket. This special feature of the binding pocket highlights its functional importance (Figure 39). The KLHDC2 pocket is ~16 Å long and ~12 Å wide. It can be divided into two chambers, which are separated by a low ridge in the middle. Interestingly, one chamber is more conserved than the other. Moreover, electrostatic potential mapping of KLHDC2 unveils a basic patch at the bottom of the pocket, which is co-localized with the highly conserved surface area (Figure 40).

The SelK C-end degron peptide is anchored to the KLHDC2 top surface pocket in a highly coiled conformation. Its C-terminal end is deeply embedded in the more conserved pocket chamber, hereafter referred to as "chamber C". By contrast, the N-terminal end of the degron peptide is accommodated by "chamber N" and more exposed to

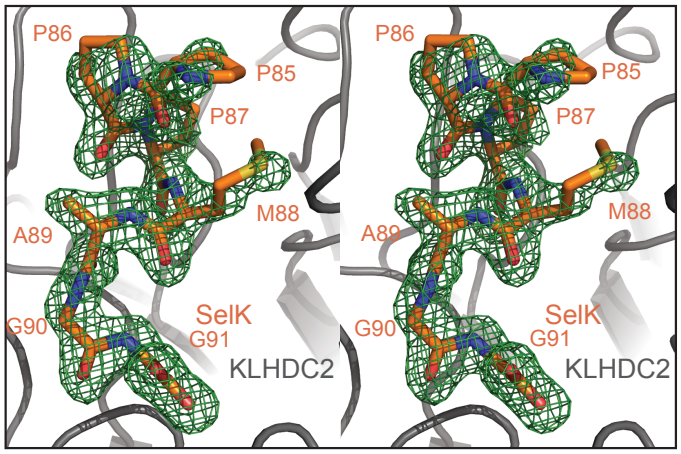



Figure 41. Stereo view of the KLHDC2 kelch repeat domain pocket with a SelK C-end degron bound. KLHDC2 (gray) is shown in cartoon. SelK C-end degron (orange) is shown in sticks together with its positive Fo-Fc electron density (forest green) calculated and contoured at 3σ before it was built into the complex model (PDB:6DO3).

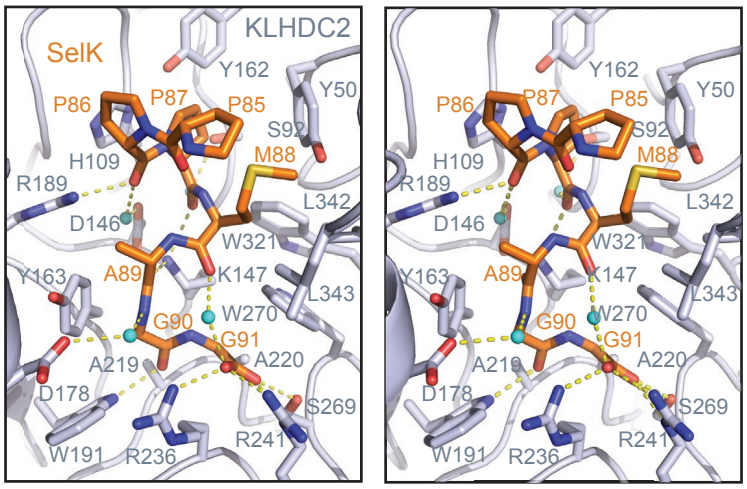



Figure 42. A stereo close-up view of the interface formed between KLHDC2 and the SelK peptide. KLHDC2 is colored in light grey with its SelK-interacting residues shown in sticks. The SelK peptide is shown in orange sticks. Hydrogen bonds and salt bridges are indicated by yellow dashed lines. Water molecules are shown as cyan spheres (PDB:6DO3).

solvent (Figure 39 and 41). The complex structure unmasks an intimate and compact interface between the E3 ligase and the degron peptide, which is dominated by inter-molecular hydrogen bond networks and salt bridges. The extreme C-terminal carboxyl group simultaneously forms two hydrogen bonds and two salt bridges with three highly conserved KLHDC2 residues, Ser269, Arg241, and Arg236, which demarcate one side of chamber C (Figure 39, 41, and 42). The tandem glycine resides are buttressed by two KLHDC2 alanine residues, Ala219 and Ala220, from the bottom, and flanked by three KLHDC2 aromatic residues, Tyr163, Trp191, and Trp270, on the sides. The backbone carbonyl group of the penultimate glycine residue is further locked in through a hydrogen bond donated by the side chain of KLHDC2 Trp191 (Figure 42 and 43). Overall, the strict requirement of a glycine residue at the penultimate position can be explained by two factors. First, the backbone dihedral angles (φ ≅

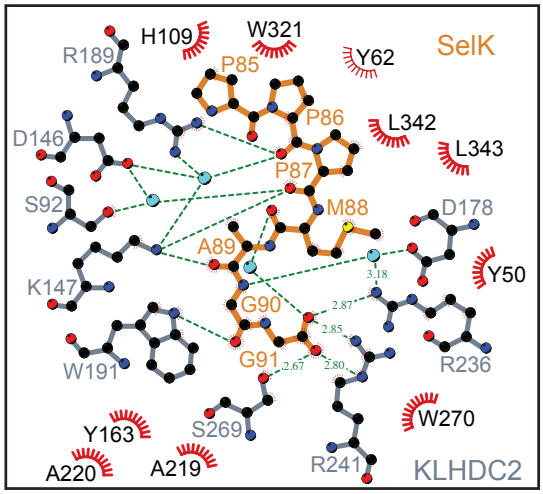



Figure 43. Ligplot diagram of the interactions between KLHDC2 and the SelK C-end degron peptide. The SelK peptide is shown in orange and the KLHDC2 residues forming hydrogen bonds with SelK or water molecules are shown in grey. Residues in KLHDC2 involved in van der Waals packing are shown in black with 1/3 circle eyelash shape in maroon. Water molecules are shown as cyan spheres.

90º, ψ ≅ -5º) are limited to glycine. Second, the space between the Cα atom and the surrounding KLHDC2 residues can only accommodate a residue without a side chain. Although glycine is strongly preferred at the last position of the diglycine degron, alanine is also allowed based on previous findings (Lin et al., 2018). The backbone geometry of the terminating residue is less constrained than internal amino acids, but the limited space between its Cα atom and the surrounding KLHDC2 residues would only accommodate a small side chain.

Beyond the C-terminal diglycine motif, the SelK degron makes three additional hydrogen bonds with the E3 ligase, all through its backbone carbonyls. Pointing up from the bottom ridge of the degron-binding pocket, Lys147 of KLHDC2 stabilizes the SelK peptide by hydrogen bonding with the backbone carbonyl groups at the -3 and -5 positions (Figure 42 and 43). Arg189 of the E3, meanwhile, holds the N-terminal end of the SelK degron in place by interacting with the -6 carbonyl. This hydrogen bond network is further substantiated by four stably bound water molecules, which either bridge the degron and the E3 or stabilize the conformation of the peptide itself. In contrast to these extensive polar interactions, the side chains of the SelK degron peptide only make limited van der Waals packing against KLHDC2. Out of the seven degron residues observed in our structure, only two at -4 and -5 positions use their side chains to make contacts with the boundary of chamber N (SelK-M88 with Y50, L342, L343 in KLHDC2, and SelK-P87 with Y62, H109, and W321 in KLHDC2). Collectively, the KLHDC2-SelK degron interface is predominantly mediated by backbone groups on the degron side and the side chains of amino acids lining the E3 pocket. The physical size of the E3 pocket, which can only accommodate 7 amino acids from the degron, limits the impact of a longer degron sequence on binding affinity.

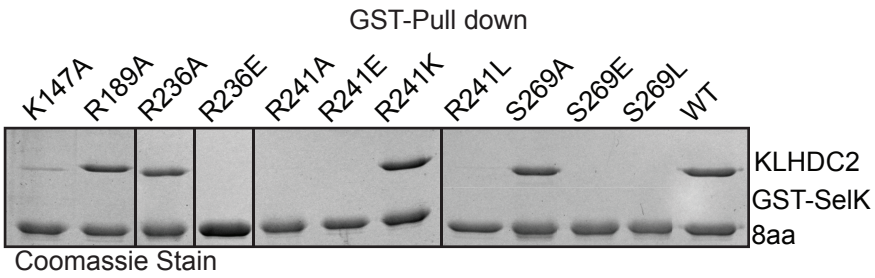



Figure 44. GST-pull down between GST-SelK 8 aa peptide and KLHDC2 mutants. Interactions between purified KLHDC2 mutants and SelK C-end degron fused with GST detected by GST pull-down assay and visualized on SDS-PAGE with Coomassie stain.

Meanwhile, the close and continuous contacts the pocket makes with the C-terminal 5 amino acids of the degron explain its minimum length requirement for high affinity interactions.

## 3.4 Mutational Analysis of the KLHDC2 Degron-Binding Pocket

To dissect the functional roles of the degron-binding residues in KLHDC2, I engineered a series of single amino acid mutants and tested their activities both *in vitro* and *in vivo*. I first purified individual KLHDC2 mutants and assessed their ability to bind GST-fused SelK degron in a pull-down assay (Figure 44 and S3). In parallel, we generated a GPS reporter cell line expressing GFP-fused SelK C-end degron with endogenous KLHDC2 knocked down by shRNA. We then

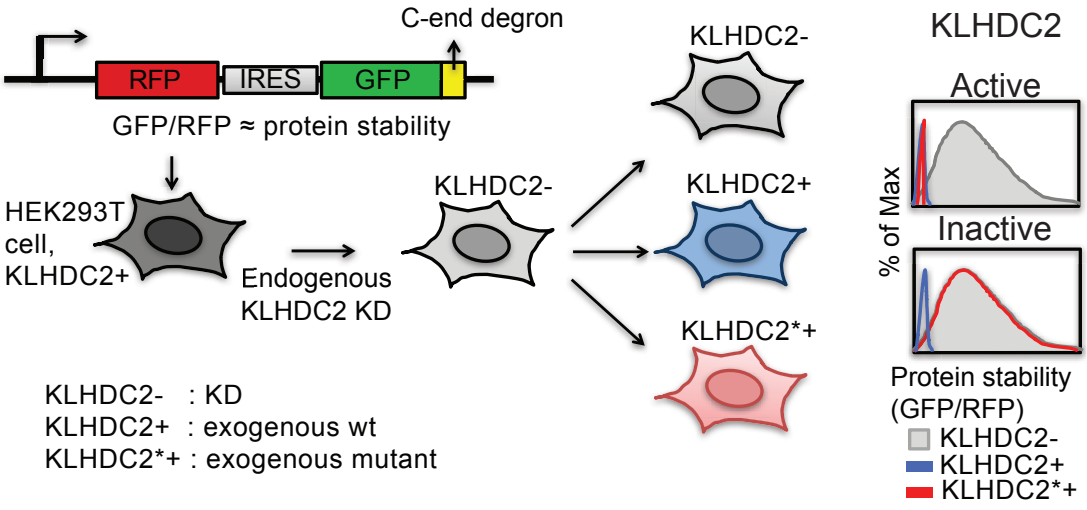



Figure 45. Schematic representation of the experimental design for the Global Protein Stability assay. GPS experimental design for assessing the effects of wild type (KLHDC2+) and mutant KLHDC2 (KLHDC2*+) on the stability of GFP fused to the SelK or USP1-NTD degron. The GFP/RFP ratio is used to indicate the stability of GFP fused with a C-end degron and was analyzed by flow cytometry and presented in histogram plots. KD: knockdown (KLHDC2-).

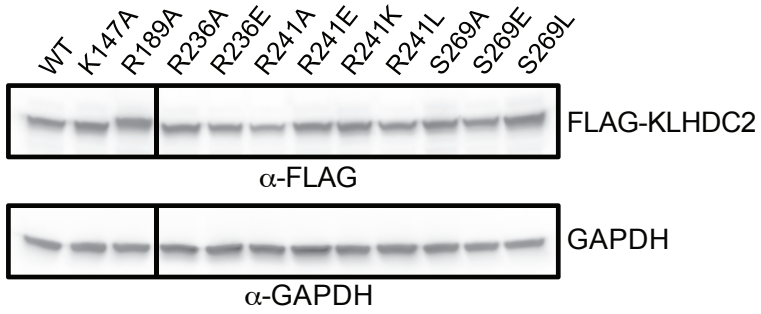



Figure 46. Assessing exogenous KLHDC2 levels using Western blot analysis. Western blot analysis on the expression of exogenous wild type (WT) and KLHDC2 mutants from HEK293T cells with endogenous KLHDC2 knocked down in the GPS assay.

individually introduced wild type KLHDC2 or its single point mutants and compared their abilities to promote the degradation of SelK degron (Figure 45 and 46). In this cell-based system, degradation of the GFP-fused substrate protein can be visualized by a lower steady-state abundance of GFP-substrate to RFP among the entire cell population (peak shift from right to left). Among the three KLHDC2 residues directly interacting with the C-terminal carboxyl group, alanine mutation of Arg241, but not the other two amino acids, Arg236 and Ser269, abrogated the binding of the SelK degron and failed to induce the degradation of the substrate (Figure 44 and 47). These results accentuate the importance of the C-terminal carboxyl group in the SelK degron

and indicate a central role of Arg241. This strictly conserved and positively charged residue makes a bidentate interaction with the peptide-terminating group via a salt bridge and a hydrogen bond. Intriguingly, when Arg241 is mutated to lysine, KLHDC2 retained its ability to bind and destabilize SelK, highlighting the "hot-spot" nature of the inter-molecular salt bridge. In support of this notion, removal or reversal of the positive charge at this position (R241L and R241E), or neutralization of its positive charge by mutating the nearby R236 residue to a glutamate (R236E), prevented KLHDC2 from binding and degrading the substrate. Noticeably, altering Ser269, which lies underneath the carboxyl group, to a bulkier hydrophobic or negatively charged amino acid also abolished the E3-degron engagement. This result implies that the C-terminal carboxyl group has to be precisely positioned for productive interaction.

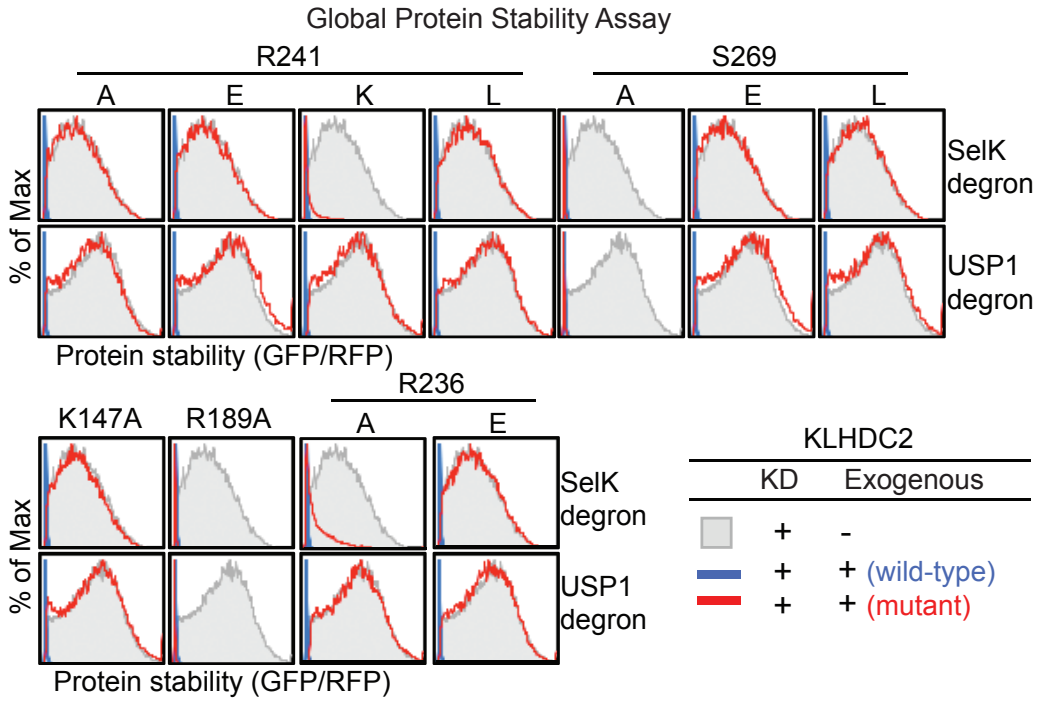


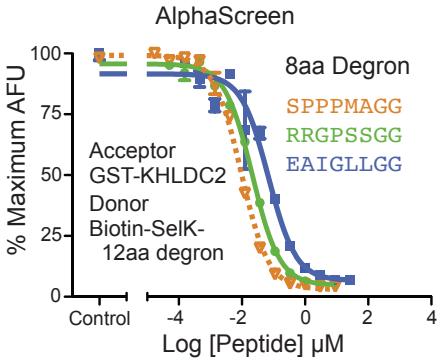

Figure 47. Turnover of GFP-fused SelK or USP1-NTD monitored by GPS. Stability of GFP-fused SelK or USP1-NTD C-end degrons were monitored by global protein stability assay with endogenous KLHDC2 knocked down by shRNA and complemented by exogenously expressed KLHDC2 wild type and mutant proteins. Destabilization of the target protein by the wild type (black line) or mutant (red line) KLHDC2 is indicated by a sharp peak at the left side of each panel.

Due to their potential roles in stabilizing the kelch repeat fold, I chose not to mutate the three aromatic residues sandwiching the diglycine motif. Although the two nearby alanine residues, A219 and A220, at the bottom of the KLHDC2 pocket are solvent-exposed and become buried by the tandem glycine residues in the complex, mutating either one to leucine altered the solution behavior of KLHDC2 as detected by size exclusion chromatography (data now shown). The integrity of this portion of the pocket, therefore, is important to the proper folding of the β-propeller. Moving to the middle of the pocket, I next mutated the ridge residue, Lys147, which donates hydrogen bonds to the -3 and -5 backbone carbonyl groups. The KLHDC2 K147A mutant has a severely impaired degron-binding activity and loses its function in downregulating SelK. This data echoes the importance of the -5 amino acid of the SelK degron in maintaining high affinity binding, as observed in our AlphaScreen competition assay. In chamber N of the KLHDC2

pocket, Arg189 makes the only direct hydrogen bond interaction with the degron backbone at the -6 position. Mutating this positively charged residue to alanine had little to no effect on degron binding and SelK stability, in accordance with the subtle change in $IC_{50}$ when the degron peptide length is reduced from 6 to 5 amino acids. As anticipated, degron binding determined by our *in vitro* pull-down experiments is strongly correlated with protein stability measured by the cell-based GPS assay.

## 3.5 Diglycine C-End Degrons from Other Substrates

In addition to SelK, the extreme C-terminal diglycine degron has been found in a number of proteins, either in their full length, early terminated, or proteolytically processed form (Lin et al., 2018, Koren et al., 2018). These degrons do not share any consensus sequence except the C-terminal diglycine motif. To understand how degenerate N-terminal sequences impact KLHDC2 binding and how they are accommodated by the E3, I chose to focus on the diglycine degrons from two additional proteins, SelS, and USP1-NTD, which have been previously suggested to be KLHDC2 substrates. In our AlphaScreen-based competition assay, the 8 aa SelS degron peptide displayed an $IC_{50}$ value slightly above SelK, whereas the potency of the USP1-NTD degron peptide is about 7 folds lower (Figure 48). The N-terminal sequences of USP1 and SelS degron peptides are distinct from SelK, with the exception of a single




| | $IC_{50}$(µM) | 95% CI (µM) |
|---|---|---|
| SelK | 0.0094 | [0.0083-0.0107] |
| SelS | 0.021 | [0.019-0.024] |
| USP1 | 0.0695 | [0.0458-0.1053] |

Figure 48. Determination of the $IC_{50}$ values for SelK, SelS and USP1 C-end degrons to KLHDC2. Titration curves from the AlphaScreen competition assay were measured for quantifying the affinity of the SelS (green) and USP1 (blue) C-end degron peptides in comparison to the SelK 8 aa degron peptide (orange, dashed line). The $IC_{50}$ value and the 95% confidence interval for each peptide is listed in the table at the bottom. The dose response curve for the SelK 8 aa degron peptide represented in a dashed line is identical to the data shown in Figure 35. Data are measured in triplicates and represented as mean ± SEM.

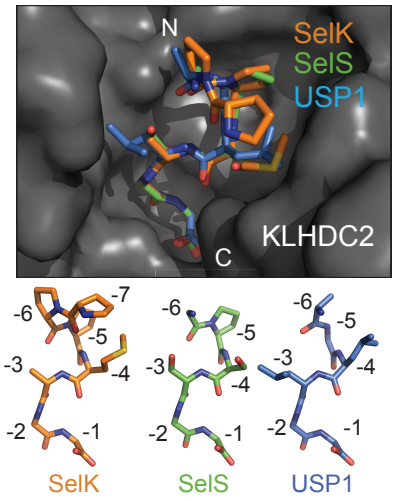



Figure 49. Binding more of three C-end degrons to KLHDC2. Superposition analysis of the three complexes formed between KLHDC2 and the three C-end degron peptides included in this study. The surface representation of the KLHDC2 pocket is shown in grey. SelK, SelS, and USP1 degrons are colored in orange, green, and blue, respectively. Superposition of the three structures was performed using the entire KLHDC2-peptide complex. The conformation of each individual peptide bound to the E3 is illustrated at the bottom with the amino acid positions from the C-terminal end labeled (PDB:6DO3, 6DO4, 6DO5).

proline residue at the -5 position in SelS. Some variation in this portion of the degron, therefore, can have detectable, albeit minor, effects on the degron-KLHDC2 interaction.

To compare the binding modes of the two degrons to SelK, I determined their crystal structures in complex with KLHDC2. The 8 amino acids SelS and USP1 degron peptides dock to the KLHDC2 top surface pocket in a nearly identical topology to SelK. The superposition of the three structures reveals a perfectly aligned C-terminal diglycine motif and a small degree of variation between the backbone atoms of the N-terminal residues (Figure 49). From the distal end of chamber C to the opposite end in chamber N, the KLHDC2 pocket widens, thus harboring side chains of different sizes decorating the N-terminal end of the three degrons. For instance, at the -3 position of the degron, the side chain can extend from a single methyl group to an isobutyl group, while at the -4 position, the KLHDC2 pocket is able to house a degron residue ranging from a serine to a methionine. Among the three degron peptides, SelK has the highest affinity toward KLHDC2 and displays a clear electron density for its residue at the -7 position. These properties could be attributed to the rigidity provided by the three sequential prolines at its N-terminal end. By contrast, USP1 features a glycine at the -5 position, which could destabilize the conformation of the degron and contribute to its weaker binding.

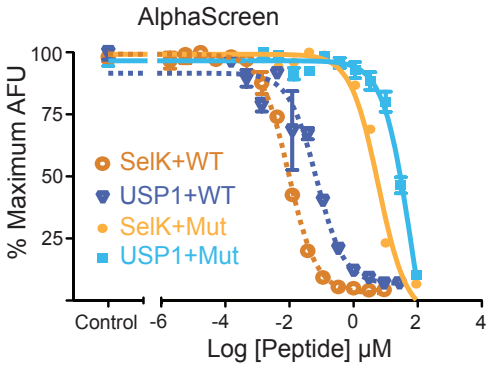
We next subjected the C-end degron of USP1-NTD to the GPS-based stability assay in conjunction with the series of KLHDC2 mutants previously described. As expected, the majority of the KLHDC2 mutations located in the degron-binding pocket, such as R241A, K147A, and R189A, elicited the same effects on USP1-NTD degron stability, as they did to SelK (Figure 47). To our surprise, though, two KLHDC2 mutants yielded opposite results for the two substrates. While the KLHDC2 R241K mutant maintained its wild type ability to destabilize SelK, the same mutant failed to promote the degradation of USP1-NTD. This differential effect was also observed for the KLHDC2 R236A mutant. Overall, the USP1-KLHDC2 interface, particularly at the position where the C-terminal carboxyl group of the degron is recognized, seems to be less tolerant to perturbation.

## 3.6 Affinity Requirement for Substrate Degradation

The opposite effects of the KLHDC2 arginine mutations on the stability of SelK and USP1 degrons is in stark contrast to the structural similarity of their E3-bound forms. Why would the KLHDC2 R241K mutant degrade one substrate but spare the other? Multiple factors, such as E3 binding, the nature of E2, and the availability of ubiquitin-conjugating sites, could influence the efficiency of substrate ubiquitination. In our case, most of these factors are constant and can be ruled out. Moreover, USP1 can be destabilized by the wild type E3, and the KLHDC2 R241K mutant is able to degrade SelK. The defective degradation of USP1 by the mutated E3, therefore, is not due to loss of function in either the degron or the KLHDC2 mutant. I postulated that the answer to the question above lies in the altered affinity of the two substrates toward the kelch repeat mutant and their cellular concentrations, which dictate their complex formation.

To test our hypothesis, I first sought to quantify the binding of the substrate degrons to the KLHDC2 R241K mutant. In the crystal structures of wild type KLHDC2 bound to SelK and USP1

degron peptides, Arg241 interacts with the extreme C-terminus of both peptides via a salt bridge and a hydrogen bond. The R241K mutation is predicted to weaken the binding by losing the hydrogen bond while retaining the salt bridge. I replaced wild type KLHDC2 with the R241K mutant in the AlphaScreen assay and re-established the degron-E3 binding with the biotinylated 12 aa SelK degron peptide. Next, individually applied the label-free 8 aa SelK and USP1 degron peptide at different concentrations and derived their $IC_{50}$ values. The SelK and USP1 peptides showed $IC_{50}$ of 5.9 μM and 62.6 μM for the KLHDC2 R241K mutant, which represents a 630- and 900-fold increase over the wild type E3, respectively (Figure 50). Consistent with a ~7-fold lower affinity of USP1 vs. SelK for the wild type ubiquitin ligase (9.4 nM vs. 69.5 nM), USP1 binds the mutant E3 ~10-fold weaker than SelK. The similar degree of affinity changes for the two degron peptides when binding to the mutant KLHDC2 can be explained by the free energy loss association with the missing hydrogen bond at -3.9 to -4.1kcal/mol.




| KLHDC2-R241K | $IC_{50}$(μM) | 95% CI (μM) |
|---|---|---|
| SelK | 5.94 | [4.23-8.35] |
| USP1 | 62.6 | [40.7-96.3] |

Figure 50. Assessment of the effect of KLHDC2 mutation on binding SelK and USP1 degrons. Titration curves from the AlphaScreen competition assay were measured for quantifying the affinity of the SelK and USP1 C-end degron peptides towards KLHDC2-R241K mutant in comparison to the wild type E3 protein. The dose response curves of SelK against wild type and mutant KLHDC2 are displayed in orange dashed line and tangerine solid line, respectively. The dose response curves of USP1 against wild type and mutant KLHDC2 are displayed in blue dashed line and cyan solid line, respectively. The orange and blue dashed lines are identical to the dose response curves shown in Figure 48. Data are measured in triplicates and represented as mean ± SEM.

The diminished affinities of the two degrons and their consistent difference, in and of themselves, cannot unmask the reason for the differential effects of the E3 mutation on the stability of SelK and USP1. However, once the cellular concentrations of the two proteins and the E3 mutant are considered, a possible explanation for the disparity in their degradation arises. Using

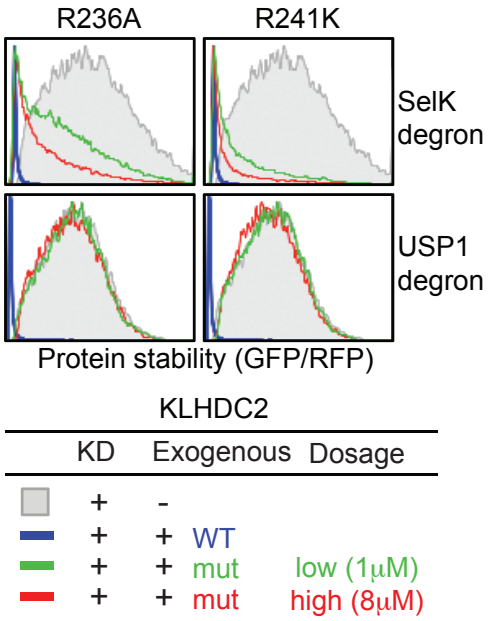



Figure 51. The effects of KLHDC2 mutations on the degradation of GFP fused SelK and USP-NTD degrons. Stability of GFP-fused SelK or USP1-NTD C-end degron monitored by global protein stability assay with endogenous KLHDC2 knocked down by shRNA and complemented by two exogenously expressed KLHDC2 mutant proteins (R236A and R241K).

purified KLHDC2 as standard, we performed quantitative western blot analysis and determined the cellular concentration of the KLHDC2 R241K mutant expressed in our GPS assay to be ~1μM. Through a similar analysis with purified GFP as standard, the concentration of GFP-fused USP1-NTD degron was calculated to be 11.2 μM, 6-fold below the affinity of its degron toward the E3 mutant ($K_d$ ~62.6 μM). These results suggest that only a small fraction of the GFP-USP1-NTD degron fusion protein will be complexed with KLHDC2 at any time. The E3 mutant, therefore, will not be able to degrade USP1 efficiently. By contrast, the cellular concentration of GFP-fused SelK degron was measured to be ~8.8 μM, which is above the concentration at which 50% of the substrate is bound to the E3 ($K_d$ ~5.9 μM) and favors complex formation. Thus, GFP-fused SelK degron can be effectively destabilized by the KLHDC2 R241K mutant. In support of this reasoning, when we increased the dosage of the exogenous KLHDC2 mutants to reach a concentration of ~8 μM, we observed noticeable enhancement in GFP-SelK degron degradation, but only marginal effect on USP1 (Figure 51). Besides the inefficiency in complex formation, the low affinity between USP1-NTD and KLHDC2 R241K mutant might also prevent the ubiquitin transfer reaction itself from happening. The fast $k_{off}$ associated with weak binding will particularly impact polyubiquitin chain assembly by impeding the attachment of the first ubiquitin to the CRL substrates (Pierce et al., 2009).

## 3.7 Discussion and Concussions

Here I elucidate for the first time how a C-end degron is recognized by its cognate E3 ligase. Among the different classes of C-end degradation signals, the diglycine degron is distinguished from others by the simplicity of its consensus sequence that lacks any side chains. The crystal structures of KLHDC2 in complex with three diglycine C-end degrons reveal a surprisingly condensed degron-E3 interface, which is cemented by a network of inter-molecular polar interactions. Upon docking to the E3, the degron adopts a compact conformation, which allows its carboxyl terminus and backbone carbonyl groups to make interactions with a cluster of conserved E3 side chains. This binding mode licenses the overall promiscuous diglycine degron with a minimum of five amino acids to achieve high affinity KLHDC2 binding. Previous studies have shown that the SelK degron, when fused to an otherwise stable protein, has to be seven amino acids long in order to induce proteasomal degradation (Lin et al., 2015). In light of our structural results, this requirement can be explained by the potential steric clash between the globular domain of the substrate and the opening edge of the degron-binding pocket of the kelch repeat protein. Consistent with this notion, the 6 aa C-terminal tail of ubiquitin, which is characterized by its terminating diglycine motif, has to be extended by a linker to become a C-end degron (Lin et al., 2018).

Despite their functional similarity and positional symmetry within a polypeptide, the N-end and C-end degrons have different features, implicating them in distinct regulatory processes. The most striking difference lies in the affinity of the two classes of degrons toward their cognate E3s. The dissociation constants of the complexes formed between the classical Arg/N-end degrons and the UBR box of N-recognins have been documented to be ~4 μM or higher (Choi et al., 2010, Matta-Camacho et al., 2010). Similarly, the recently identified Pro/N-end degron binds to its E3

ligase, GID4, with 2.5 μM affinity (Chen et al., 2017, Dong et al., 2018). By contrast, the three diglycine C-end degron peptides included in this study display affinity toward KLHDC2 in the low nanomolar range. Remarkably, the binding constant of the SelK degron for the E3 was determined to be below 10 nM. Such a strong interaction is reminiscent of the binding of several known CRL substrates, such as Nrf2 and CyclinE, to their E3s (Suzuki and Yamamoto, 2015, Hao et al., 2007). These cellular regulatory proteins are characterized by the tight regulation of their protein abundance in key signaling pathways. The high affinity between the diglycine C-end degron and KLHDC2 suggests that this specific branch of the DesCEND pathway could have evolved to degrade substrates of low abundance and/or eliminate substrates quickly in response to certain cellular events. It is conceivable that a high affinity might also be required to keep substrates at a very low level due to their potential toxicity.

The diglycine degron belongs to a small cohort of C-end degrons characterized by a preferred C-terminal glycine residue (Lin et al., 2018, Koren et al., 2018). The majority of these glycine-terminating C-end degrons are recognized by kelch domain-containing proteins, including KLHDC2, KLHDC3 and KLHDC10. With a common kelch repeat β-propeller fold, KLHDC3 and KLHDC10 are predicted to interact with their specific degrons via a similar top surface pocket as the one revealed in our structure. The particular residues involved in degron binding most likely vary among these kelch domain proteins. Nevertheless, the carboxyl terminus and backbone carbonyls of these glycine-ending degrons are expected to directly interact with interface residues conserved within the KLHDC3 and KLHDC10 orthologs. The compact substrate-binding pocket of KLHDC2 renders the E3 ligase a potential druggable target. In our competition assay, a peptide as small as a diglycine motif with a molecular weight of 132 Da was able to bind to the KLHDC2 pocket with an estimated affinity of ~360 μM. It is conceivable that the KLHDC2 pocket can be

exploited by degron-mimicking small molecule compounds to reprogram the E3 ligase for degrading disease-related targets (Raina and Crews, 2017, Zheng and Shabek, 2017).

## 3.8 METHOD DETAILS

### 3.8.1 *Experimental Model and Subject Details*

For DNA extraction, *E.coli* DH5α was grown for 16hr at 37°C. For bacmid production, *E.coli* DH10Bac was grown for 16hr at 37°C. For baculovirus production and amplification, Sf9 insect cells were grown for 2-3 days at 26°C. For protein expression, both *E.coli* BL21(DE3) (grown for 4hr at 37°C) and HighFive insect cells (grown for 2-3 at 26°C, 105RPM) were used. LB Broth Miller (Fisher BioReagents) was used for *E.coli*. Sf9 insect cells were maintained in Grace's Insect Medium (Gibco) supplemented with 7% FBS (Gibco) and 1% Penicillin-Streptomycin (HyClone) solution. Suspension Hi5 cells were grown in EXPRESS$^{TM}$ FIVE SFM(Gibco) supplemented with 5% L-Glutamine 200mM (HyClone) and 1% Penicillin-Streptomycin (HyClone) solution. HEK293T cells were maintained in DMEM with 10% FBS and antibiotics at 37°C in a 6% CO2 atmosphere. Tissue culture media and supplements were from GIBCO Life Technologies (Carlsbad, CA, USA).

### 3.8.2 *Molecular Biology and Protein Purification*

The kelch repeat domain of human KLHDC2 (amino acid 1–362) was subcloned into the pFastBac vector with an N-terminally fused glutathione-S-transferase (GST) and a TEV-cleavage site. A recombinant baculovirus was produced and amplified three times in Sf9 monolayer cells to produce P4. The P4 virus was used to infect Hi5 suspension insect cell cultures to produce the recombinant GST-KLHDC2 protein. The cells were harvested 2-3 days post-infection, re-suspended and lysed in lysis buffer (20 mM Tris, pH 8.0, 200 mM NaCl, 5 mM DTT) in the

presence of protease inhibitors (1µg/ml Leupeptin, 1µg/ml Pepstatin and 100µM PMSF) using a microfluidizer. The GST-KLHDC2 protein was isolated from the soluble cell lysate by Pierce™ Glutathione Agarose (Thermo Scientific). For AlphaScreen competition assays, the GST tagged KLHDC2 was further purified by Q Sepharose High Performance resin (GE Healthcare). The NaCl eluates were subjected to Superdex-200 size exclusion chromatography column (GE Healthcare). For crystallization and GST-pull down assays, the same purification steps were employed, with a GST tag removal step following affinity purification. The samples used for crystallization were concentrated by ultrafiltration to 19–29 mg $mL^{-1}$. All single amino acids KLHDC2 mutants were purified by glutathione agarose resin, cleaved with TEV, and further purified by ion exchange chromatography following the same procedure as described for the wild type protein. GST-fused SelK degron proteins were overexpressed and purified from BL21 (DE3) *E. coli* cells. Bacterial cells transformed with the pGEX-based expression plasmid for GST, GST-8 aa and -12 aa SelK degron were grown in LB broth at 37 °C to an $OD_{600}$ of 0.8 and induced with 0.1 mM IPTG for 4 hrs. Cells were harvested, re-suspended and lysed in lysis buffer. The proteins were isolated from soluble cell lysate by glutathione agarose resin. All protein samples were flash frozen in liquid nitrogen for future use.

### 3.8.3 *Protein Crystallization*

The crystals of KLHDC2 in complex with the 8 aa peptides from SelK and USP1-NTD C-end degrons (Bio-Synthesis, Inc) were grown at 25 °C by the hanging-drop vapor diffusion method with 0.150 µL protein complex sample mixed with 0.075 µL volume of reservoir solution containing 0.03 M $MgCl_2*6H_2O$, 0.03 M $CaCl_2*2H_2O$, 0.05 M imidazole, 0.05 M 2-(N-morpholino)ethanesulfonic acid (MES) monohydrate, 12.5% (v/v) 2-methyl-2,4-pentanediol (MPD), 12.5% (w/v) PEG1000, 12.5% (w/v) PEG3350, pH 6.5. KLHDC2 in complex with the 8

aa SelS C-end degron peptide (Bio-Synthesis, Inc) was crystallized in 0.03 M MgCl2*6H2O, 0.03 M CaCl2*2H2O, 0.05 M Imidazole, 0.05 M MES monohydrate, 13.25% (v/v) MPD, 13.25% (w/v) PEG1000, 13.25% (w/v) PEG3350, pH 6.8. Crystals of maximal sizes were obtained and harvested after a few days. Cryoprotection was provided by the crystallization condition.

### 3.8.4 *Data Collection and Structure Determination*

After collecting a native dataset at Advanced Light Source Beamline 8.2.1, X-ray diffraction data were integrated and scaled with HKL2000 package (Otwinowski and Minor, 1997). The structure was solved by molecular replacement with ensembler and phaser from the PHENIX suite of programs. Ensembler was used to superimpose multiple kelch domain containing structures (PDB: 2VPJ, 5A10, 1ZGK, 4CHB, 5CGJ, and 3II7). The output ensembler search model was used for molecular replacement in phaser. Initial structural models were built, refined and rebuilt using COOT (Emsley et al., 2010) and PHENIX (Adams et al., 2002). All structural models were manually built, refined, and rebuilt with PHENIX and COOT. PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC), CHIMERA (Pettersen et al., 2004), and LIGPLOT (Laskowski and Swindells, 2011) were used to generate figures. Protein sequence alignment was performed with CLC Sequence Viewer 7.

### 3.8.5 *AlphaScreen Luminescence Proximity Assay*

AlphaScreen assays for determining and measuring protein-protein interactions were performed using EnSpire reader (PerkinElmer). GST-tagged KLHDC2 (WT or R241K) was attached to anti-GST AlphaScreen acceptor beads. Synthetic biotinylated 12 aa SelK degron peptide (Bio-Synthesis, Inc.) was immobilized to streptavidin-coated AlphaScreen donor beads. The donor and acceptor beads were brought into proximity by the interactions between the SelK peptide and

KLHDC2. Excitation of the donor beads by a laser beam of 680 nm promotes the formation of singlet oxygen. When an acceptor bead is close proximity, the singlet oxygen reacts with thioxene derivatives in the acceptor beads and causes the emission of 520-620 nm photons, which are detected as the binding signal. If the beads are not in close proximity to each other, the oxygen will return to its ground state and the acceptor beads will not emit light. Competition assays were performed in the presence of numerous peptides which were titrated at various concentrations. Different lengths of the SelK peptides were used (12, 10, 8, 6, 5, 4, 3, and 2 amino acids) in order to determine the minimal length of a high affinity binding degron. In order to dissect the role of the extreme C-terminal diglycine motif, two 8 aa SelK peptides were used for competition assays, one with the C-terminal glycine mutated to leucine and the other with an amidated C-terminus. Lastly, to assess the binding of degrons of different substrate proteins to KLHDC2, I used 8 aa peptides from the extreme C-terminus of early terminated SelS and USP1-NTD.

The experiments were conducted with 0.12 nM of GST-KLHDC2 and 1.7 nM biotinylated 12 aa SelK peptide or with 11 nM GST-KLHDC2-R241K and 11 nM biotinylated 12 aa SelK peptide in the presence of 5 μg/ml donor and acceptor beads in a buffer of 25 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM TCEP, 0.1% Tween-20, and 0.05 mg/ml Bovine Serum Albumin. The concentrations of the peptides used in competition assays ranged from 2 nM to 3 mM. The precise concentrations of the stock peptide were determined by amino acid analysis (TAMU Protein Chemistry Lab, Texas A&M University). The experiments were done in triplicates. $IC_{50}$ values were determined using non-linear curve fitting of the dose response curves generated with Prism 4 (GraphPad).

### 3.8.6 *Octet BioLayer Interferometry Measurement*

Binding affinity of biotinylated 12 aa SelK peptide with GST-KLHDC2 was measured using the Octet Red 96 (ForteBio, Pall Life Sciences) following the manufacturer's procedures in quadruplicates. The optical probes were coated with streptavidin, loaded with 200 nM biotinylated 12 aa SelK peptide as ligand and quenched with 0.1 mM biocytin prior to kinetic binding analysis. The reaction was carried out in black 96 well plates maintained at 30°C. The reaction volume was 200 µL in each well. The binding buffer contained 20 mM Tris-HCl, 200 mM NaCl, 5 mM DTT and 0.1% BSA, pH 8.0. The concentrations of GST-KLHDC2 as the analyte in the binding buffer were 100, 50, 25, 13, and 6.3 nM. There was no binding of the analyte to the unloaded probes. Binding kinetics of the analyte at all five concentrations were measured simultaneously using instrumental defaults. The data was analyzed by the Octet data analysis software. The association and dissociation curves were globally fit with a 1:1 ligand model. The $k_{on}$ and $k_{dis}$ values were used to calculate the dissociation constant, $K_d$, with kinetic analysis of the direct binding.

### 3.8.7 *Global Protein Stability Assay*

The GPS assays using SelK and USP1-NTD reporter cell lines were performed (Lin et al., 2018, Emanuele et al., 2011). The GPS reporter system was based on the co-expression of GFP and RFP from a single transcript enabled by an internal ribosome entry site (IRES). GFP was fused with the C-end degron of SelK or USP1-NTD while RFP served as a non-degradable internal control. The GFP/RFP ratio thus indicated the stability GFP-fused C-end degron and was analyzed by flow cytometry. To testing the activity of KLHDC2 mutants, endogenous KLHDC2 was knocked down by shRNA and exogenous KLHDC2 with indicated single point mutations was introduced in GPS reporter cells. Targeting sequence of KLHDC2 shRNA is CTTGGTGTCTGGGTATATA.

To calculate the volume of a cell, ~10^7 cells were collected and resuspended with PBS buffer with known volume and the final volume was measured. By subtracting the initial volume of PBS from the final volume and dividing the value by number of cells, the volume of a single cell was estimated to be 7950 fL. To obtain the absolute abundance of the GFP-tagged SELK and USP1 proteins, purified Flag-GFP protein was quantified and used as a standard for quantitative western blot analysis.

### 3.8.8 *Protein Native Mass Spectrometry*

The KLHDC2-SelK complex was prepared for native mass spectrometry by exchanging the solution prepared for crystallography into aqueous 200 mM ammonium acetate at pH 7 using four cycles of dilution and re-concentration with a 10K MWCO Corning Spin-X UF centrifugal concentrator (14,000 g, 1 hr, 4 °C). The final concentration of the KLHDC2-SelK complex used for experiments was approximately 20 μM. That solution was then ionized using electrokinetic nanoelectrospray ionization. The ionization source used in these experiments consisted of a platinum wire electrode and a borosilicate glass capillary emitter. A micropipette puller (Sutter Instruments Model P-97) was used to fabricate the emitter from a borosilicate capillary. Approximately 3 uL of sample was added into the tip of the capillary. The platinum wire electrode was placed into the capillary and in direct contact with the sample. Between 0.5 and 1.0 kV was applied to the wire to achieve ionization (Xing et al., 2013). Those ions were then analyzed using a hybrid electrospray/quadrupole/ion-mobility/time-of-flight mass spectrometer (Waters Synapt G2 HDMS) (Allen et al., 2016), in which the original traveling-wave ion mobility cell was replaced by a radio-frequency confining drift cell. Briefly, the native mass spectrum was acquired using a 45 V bias between the sampling and extraction cones in the atmospheric-pressure interface (operated at room temperature) as well as using a 3 V bias between the quadrupole mass filter

(operated as an RF-only ion guide) and the trap collision cell (containing ~20 mTorr of argon gas). Activation in the trap cell was performed with the identical conditions used to acquire the native mass spectrum, but the bias between the quadrupole and trap collision cell was increased to 45 V. In-source activation was performed by increasing the bias between the sampling and extraction cones to 120 V. Tandem mass spectrometry was performed using in-source activation and by increasing the bias between the quadrupole mass filter (now used to isolate the released $[SelK+H]^+$ ions) and the trap collision cell to 30 V. Mass spectra were calibrated externally using electrospray generated ions from a 30 mg/ml solution of CsI.

### 3.8.9 *Affinity Pull-Down Assay*

Pull-down assay was preformed using ~200 μg of purified GST or GST-peptides as the bait and ~ 250-300 μg of KLHDC2 WT or mutant proteins. Reaction mixtures (140 μl) were incubated with 50 μl GST beads (Thermo Scientific) at 4 °C for 1 hr in the binding buffer with 20 mM Tris-HCl, pH 8, 200 mM NaCl, and 5 mM DTT. After extensive wash with 0.5 ml binding buffer three times and 140 μl binding buffer once, the protein complexes on the beads were eluted by 3 column volumes of 5 mM glutathione. SDS-PAGE loading buffer was added to the samples and proteins were analyzed by Coomassie staining. Inputs samples represent 5-10% of the total reaction.

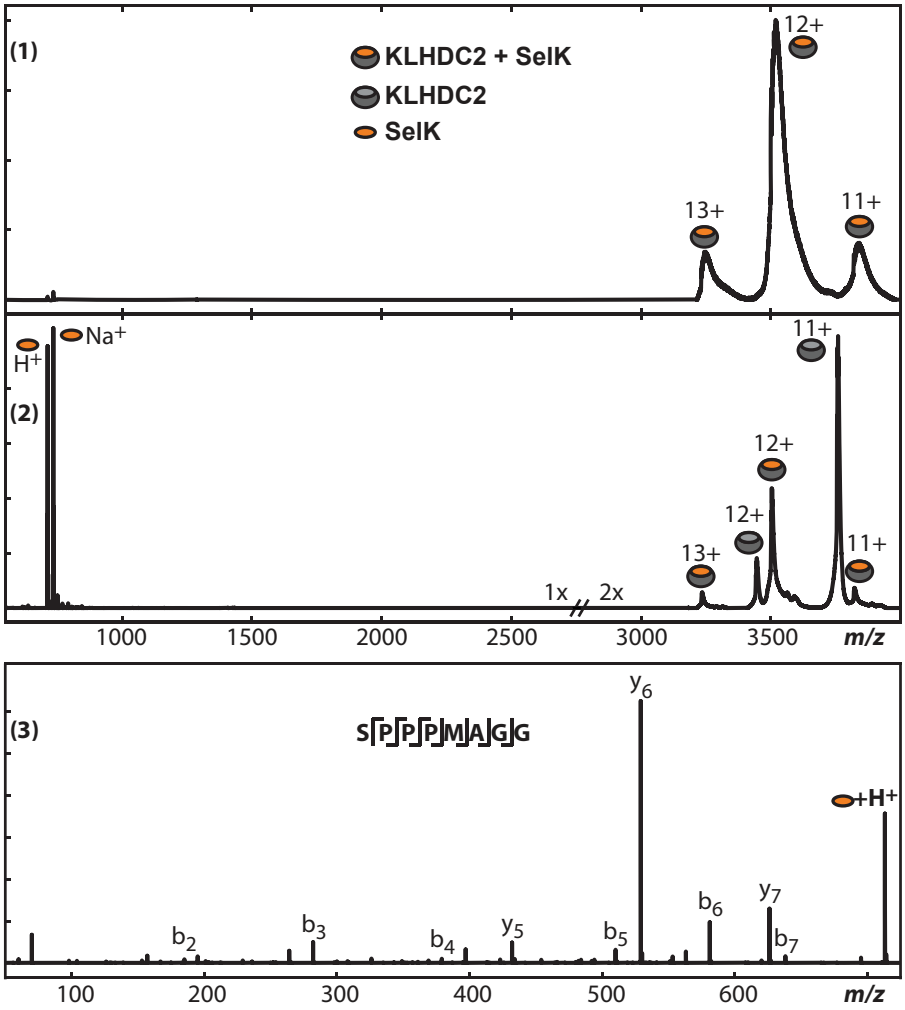



Figure S1. Protein native mass spectrometry analysis of the KLHDC2-SelK complex (1) Native mass spectrum of the KLHDC2-SelK complex, which exhibits no evidence for apo KLHDC2. (2) Mass spectrum obtained using conditions similar to that in (1), but with in-source activation. Under these conditions, $[SelK^{+}H]^{+}$ or $[SelK^{+}Na]^{+}$ are released from some of the KLHDC2-SelK complexes. (3) $[SelK^{+}H]^{+}$ released from the complex was quadrupole selected and subjected to collision-induced dissociation. This resulted in an information-rich fragmentation spectrum, confirming the assignment of the released peptide cation. Additional details of these experiments are discussed in the Methods section.

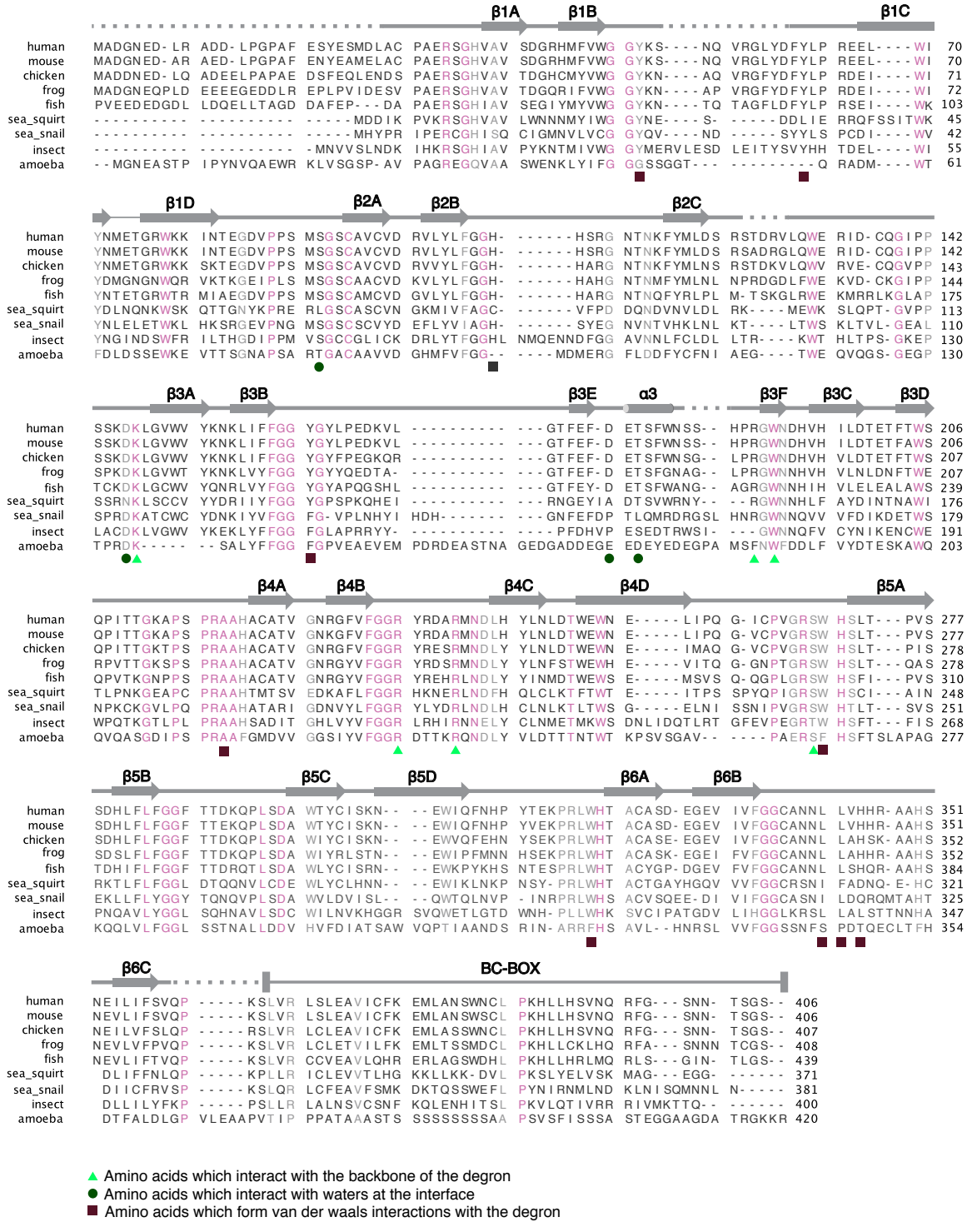



Figure S2. Sequence alignment of KLHDC2 orthologs. Sequence alignment of KLHDC2 orthologs from different species, including human (Homo sapiens), mouse (Mus musculus), chicken (Gallus gallus), frog (Xenopus lavis), fish (Scleropages formosus), sea squirt (Ciona Intestinalis), sea snail (Lottia gigantea), insect (Clastoptera arizonana), and amoeba (Acanthamoeba castellanii str. Neff). The N-terminal sequence of fish KLHDC2 and the C-terminal sequence of sea squirt and amoeba KLHDC2 orthologs are omitted for clarity. Strictly conserved residues (100%) are colored in magenta. Highly conserved residues (80-100%) are colored in light grey. Second structure elements including a-helices and b-strands are indicated by cylinders and arrows, respectively.

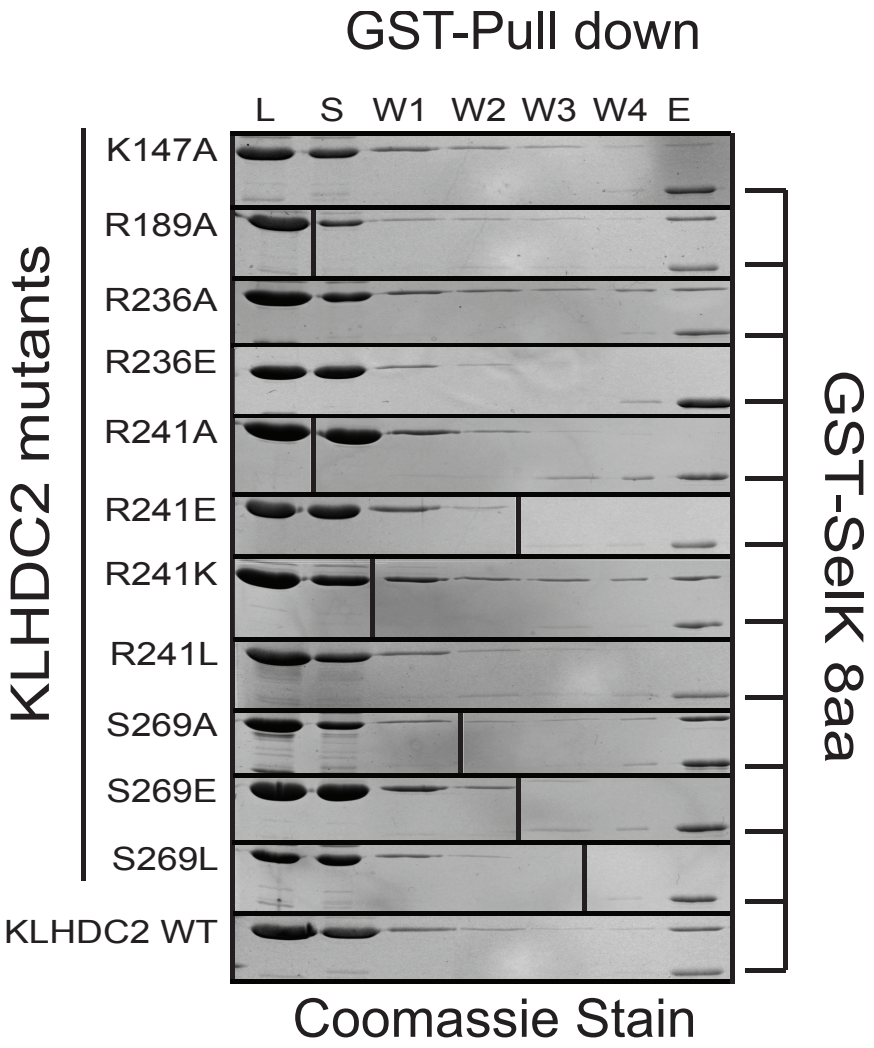



Figure S3. Binding of KLHDC2 mutants with 8 aa SelK degron fused to GST. Loading control (L), Supernatant (S, i.e. unbound), four washes (W1–W4), and elution (E) fractions were analyzed by SDS-PAGE with Coomassie stain. Vertical lines indicate discontinuity of the lanes in SDS-PAGE gels due to removal of molecular weight marker lanes.

**Table 1. Data Collection and Refinement Statistics**

| | KLHDC2-SelK | KLHDC2-SelS | KLHDC2-USP1 |
|---|---|---|---|
| PDB: | 6DO3 | 6DO4 | 6DO5 |
| **Data collection statistics** | | | |
| Space group | P 1 21 1 | P 1 21 1 | P 1 21 1 |
| Wavelength (Å) | 1 | 1 | 1 |
| Cell dimensions: a, b, c (Å) | 44.8, 87.8, 88.6 | 44.6, 88.6, 88.8 | 44.5, 87.1, 88.3 |
| Cell dimensions: a, b, g (°) | 90.0, 104.5, 90.0 | 90.0, 104.8, 90.0 | 90.0, 104.6, 90.0 |
| Resolution (Å)* | 50.0 - 2.2 (2.19 - 2.15) | 50.0 - 2.2 (2.24 - 2.20) | 50.0 - 2.2 (2.24 - 2.20) |
| Rmeas | 0.18 (0.68) | 0.16 (0.70) | 0.13 (0.60) |
| Mean I/σ(I) | 15.2 (2.2) | 11.2 (1.6) | 15.3 (1.6) |
| $CC_{1/2}$ (%) | (77.2) | (66.5) | (71.4) |
| Completeness (%) | 99.9 (99.2) | 99.3 (93.0) | 98.0 (80.3) |
| Redundancy | 6.9 (4.3) | 4.5 (3.3) | 5.9 (3.8) |
| **Refinement statistics** | | | |
| Resolution (Å) | 43.4-2.2 | 39.4-2.2 | 43.1-2.5 |
| Number of reflections | 35101 | 33734 | 22666 |
| R-work/R-free | 0.17/0.22 | 0.17/ 0.22 | 0.20/0.25 |
| Number of atoms | | | |
| Protein | 5136 | 5133 | 5095 |
| Ligand | 0 | 0 | 0 |
| Water | 208 | 230 | 128 |
| Average B-factor | 22.1 | 27.4 | 32.6 |
| RMSD bond length | 0.008 | 0.007 | 0.013 |
| RMSD bond angle | 1.160 | 0.894 | 1.303 |
| Ramachandran plot (%) | | | |
| Favored | 96.3 | 97.1 | 96.3 |
| Allowed | 3.7 | 2.9 | 3.5 |

*Values in parentheses are for the highest-resolution shell.

# Chapter 4. SMALL MOLECULE MEDIATED PROTEIN DEGRADATION

The following work has previously been published and was adapted from: Rusnac D.V., Zheng N. (2018) Overview of Protein Degradation in Plant Hormone Signaling. In: Hejátko J., Hakoshima T. (eds) Plant Structural Biology: Hormonal Regulations. Springer Nature
License Number 4720900051720

## 4.1 SYNOPSIS OF TARGETED PROTEIN DEGRADATION

Protein ubiquitination and degradation are tightly regulated processes that rely on the specific recognition of a target protein by its cognate E3 ligase. Despite the high specificity of the ubiquitin-proteasome system that has naturally evolved in the cell (Chapter 2.3), it is not uncommon for pathogenic viruses to hijack this system and alter its selectivity. Myriad viral proteins have been documented to achieve this by bridging noncognate cellular substrates to host E3 ligases, thereby, promoting their polyubiquitination and degradation to benefit viral replication and survival. Moreover, as the recognition of certain substrates by E3 ligases can be mediated by naturally occurring plant hormones or synthetic small molecules, the idea of compound-induced targeted protein degradation has arisen in the field. This new strategy involves reprogramming a ubiquitin ligase with a specific small molecule to recognize and ubiquitinate a neo-substrate, which is not the natural substrate of the E3. Currently, two distinct modalities have been proposed to achieve this goal, namely, Molecular Glue and PROteolysis TArgeting Chimeric molecule (PROTAC). In this chapter, I will describe the rationale and progress behind these two approaches along with their similarities and differences, as well as my efforts in developing KLHDC2 as a potential new E3 ligase platform that can be exploited by PROTACs for targeted protein degradation.

## 4.2 MOLECULAR GLUE

Besides PTMs, cellular signals such as hormones and secondary metabolites can directly participate in degron recognition by E3 ligases. In the green kingdom of life, two plant hormones, auxin and jasmonate (JA), have been shown to serve as molecular glue bridging the CRL1 F-box proteins TIR1 and COI1 to their specific substrates, respectively (Shabek and Zheng, 2014). These are the first known cases where E3s function in hormone perception and, as a result, regulate gene expression. The transcriptional regulation is achieved in the auxin and JA signaling pathways through the ubiquitination and degradation of AUX/IAA and JAZ transcription repressor proteins, respectively.

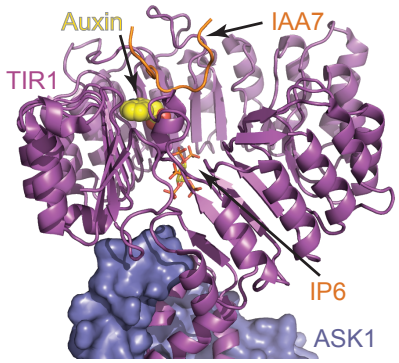



Figure 52. Structural representation of IAA7-Auxin-IP6-TIR1-ASK1 complex. Auxin-facilitated IAA7 degron recognition by TIR1. Inositol hexakisphosphate (IP6) serves as a co-factor of the ubiquitin ligase-based receptor. IAA7 and IP6 are in orange, Auxin is in yellow, TIR1 is in magenta and ASK1 is shown in blue. (PDB:2P1Q).

The crystal structure of the F-box protein, TIR1, has been determined in complex with ASK1, one of the plant homologues for SKP1 (Figure 52) (Tan et al., 2007). The complex adopts an overall mushroom-shaped structure, where the cap is composed of the TIR1 LRR domain and the stem contains the TIR1 F-box motif and ASK1. TIR1 folds into a twisted horseshoe-shaped solenoid and provides its top surface pocket for auxin-mediated degron binding. The crystal structure of TIR1 in complex with auxin and an AUX/IAA degron reveals that the hormone fills up a gap at the protein-protein interaction interface without inducing any detectable conformational changes of the F-box protein. Its unique mechanism of action helps raise the concept of Molecular Glue. By serendipity, an inositol hexakisphosphate (IP6) was discovered in the middle of the TIR1 LRR domain underneath the hormone-binding pocket. Its strategic location

and its conserved binding site strongly argue for a cofactor role in stabilizing hormone binding. Remarkably, JA-isoleucine, the active form of JA, acts through the same mechanism as auxin, mediating the interaction between the F-box protein COI1 and its substrate JAZ proteins (Sheard et al., 2010). Surprisingly, instead of IP6, COI1 uses a specific inositol pentakisphosphate as its cofactor, which is essential for its hormone sensing function. The TIR1 and COI1 F-box proteins, therefore, are each regulated by two naturally occurring small molecules, a molecular glue hormonal signal and an inositol polyphosphate cofactor, which might function as a proxy signal for phosphate abundance (Wild et al., 2016).

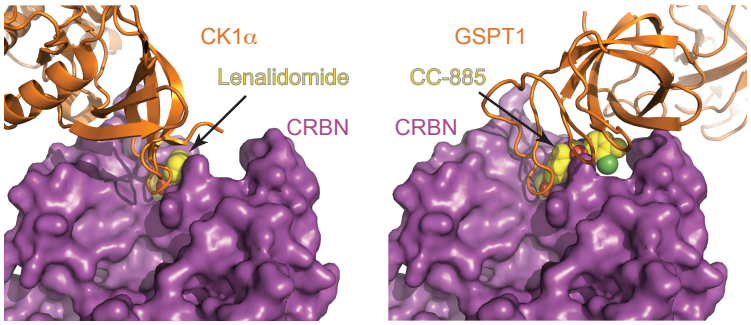



Figure 53. Structural representation of CK1α-lenalidomide-CRBN and GSPT1-CC-885-CRBN complexes. The recognition of CK1α by CRBN is promoted by lenalidomide (left) (PDB:5FQD). Structure of CRBN in complex with CC-885 and substrate protein GSPT1 (right). CRBN is displayed in magenta, the molecular glue in yellow, and the neo-substrate in orange (PDB:5HXB).

Previously, I briefly mentioned CRBN as the target of the anticancer drug, thalidomide. Recent studies have shown that thalidomide and its derivatives, lenalidomide and pomalidomide (collectively known as immunomodulatory drugs, IMiDs), also work as molecular glues, rewiring CRBN to bind and degrade several clinically relevant substrates, including IKZF1, IKZF3, CK1α, and SALL4 (Krönke et al., 2015, Krönke et al., 2014, Lu et al., 2014, Matyskiela et al., 2018, Donovan et al., 2018). These substrates do not normally interact with CRBN in the absence of the small molecules. Therefore, they have been referred to as neo-substrates. The structure of DDB-CRBN-lenalidomine-CK1α complex has elucidated the detailed mechanism of interaction among all the components (Figure 53). The small molecule drug targets a tryptophan cage surface pocket of

CRBN and is stabilized by a combination of hydrogen bonds and hydrophobic interactions. Upon binding to the E3, lenalidomide offers a novel binding surface for capturing the neo-substrate. CK1α uses a β-hairpin loop to dock into the newly formed hydrophobic pocket, contacting both the drug and the E3 substrate receptor.

CC-885 is a lenalidomide derivative that reprograms CRBN to ubiquitinate yet another neo-substrate, GSPT1 (Matyskiela et al., 2016). This compound shares part of its structure with lenalidomide and interacts with CRBN in an almost identical manner (Figure 53). The additional moiety of CC-885 that differs from lenalidomide participates in the formation of extra hydrogen bonds with the E3 and presents a new hydrophobic interface for docking GSPT1. Despite adopting a completely different fold and packing against CRBN from a different direction, GSPT1 interacts with the compound-reshaped E3 pocket using a β-hairpin motif that is superimposable to the one found in CK1α. Based on structural and mutational analyses of the CRBN-substrate complexes, a key glycine residue within the β-hairpin shared among all neo-substrates has emerged. Although the target of thalidomide and its derivatives was elucidated *post-hoc*, their mechanism of action closely resembles that of the naturally occurring plant hormones, auxin and JA. Together, these E3-reshaping molecular-glue small molecules inspire the discovery of novel compounds with therapeutic potentials through targeted protein degradation.

## 4.3 PROTEOLYSIS TARGETING CHIMERIC MOLECULES

With the purpose of reprograming E3 ligases to ubiquitinate neo-substrates, Molecular Glues have been discovered by serendipity, whereas PROTACs have been rationally designed and developed. Proteolysis targeting chimeric molecules are bifunctional compounds, which contain two warheads, one with high affinity for the E3 ligase and the other with high affinity for the target

protein. These moieties with significant affinity for their partners are connected by a linker, which allows the neo-substrate to be brought into the close proximity of the ubiquitin ligase, ubiquitinated, and degraded by the proteasome. The proof of principle for PROTACs was first established by Craig Crews and Raymond Deshaies in 2001. They showed that MetAP2, which is not a substrate of SCF, can be ubiquitinated and degraded by $SCF^{ß\text{-}TRCP}$ in the presence of PROTAC-1. The compound is composed of ovalicin and IκBβ phosphopeptide, which bind tightly to MetAP2 and β–TRCP, respectively (Sakamoto et al., 2001). Due to the chemical properties of the IκBβ phosphopeptide, PROTAC-1 is not cell permeable and its activity was only validated in lysate-based assays. Without a drug-like ubiquitin ligase-targeting compound, PROTACs did not gain much traction in the early days.

Thanks to the identification of Cereblon as the target of IMiDs in 2010s, the PROTAC approach for targeted protein degradation has flourished. In 2015, James Bradner's group and Craig Crews' group reported the development of cell permeable IMiD-based PROTACs that can bind and degrade BRD4, a bromodomain-containing transcriptional co-activator critical for cell growth. In the past few years, over 100 PROTACs have been documented, many of which harness the activities of Cereblon and another CRL substrate receptor VHL (Paiva and Crews, 2019). Multiple successful examples of PROTACs have been reported, which enable compound-induced downregulation of numerous disease-relevant targets, such as BCR-ABL, BRD4, BRD9, ERRa, FKBP12, and RIPK. Due to the complexity of the ubiquitin-proteasome system, PROTAC-induced substrate-E3 spatial proximity does not always guarantee efficient substrate elimination. To visualize the mode of action of PROTACs and to improve their efficacy, structure studies of several PROTAC molecules targeting BRD4, a BET family member with two bromodomains, $BRD4^{BD1}$ and $BRD4^{BD2}$, have been determined.

The first glimpse of a PROTAC in action was derived from the crystal structure of the E3 ligase VHL in complex with $BRD4^{BD2}$ and the PROTAC molecule MZ1. With a three-unit PEG-linker, the bifunctional molecule MZ1 uses JQ1, a pan-BET inhibitor and VH032, a hydroxyproline mimetic to induce complex formation between VHL and BRD4. Due to the high affinity of JQ1 for the BET family of proteins, MZ1 was predicted to rewire VHL and promote the degradation of multiple BET proteins. Unexpectedly, the PROTAC molecule was shown to selectively destroy BRD4, but not other BET family members (Zengerle et al., 2015). Upon solving the structure of the VHL-MZ1-$BRD4^{BD2}$ tertiary complex, it was clear that besides the predicted interactions made by each warhead moiety of MZ1 to the respective partners, the PROTAC compound folds on itself and induces new protein-protein and ligand-protein interfaces (Figure 54). The E3-neo-substrate interface is characterized by both electrostatic and hydrophobic interactions and involves residues that are not conserved among the BET family members. This feature provides the structural explanation for the unexpected selectivity of MZ1 for $BRD4^{BD2}$ over other BET proteins. In the same study, various biophysical methods were used to establish positive cooperative complex formation, where the affinity of the proteins in the tertiary complex surpassed the sum of individual affinities of the MZ1 warheads for their substrates. Therefore, the compound-induced interface revealed in this structure not only explains substrate selectivity, but also provides the molecular basis of the unexpected positive cooperativity of MZ1 (Gadd et al., 2017).

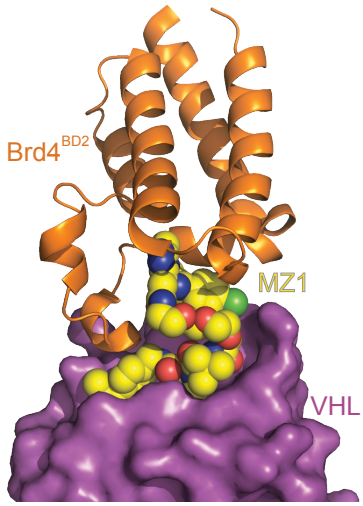



Figure 54. Structural representation of $BRD4^{BD2}$-MZ1-CRBN complex. The recognition of $BRD4^{BD2}$ by VHL is promoted by MZ1. VHL is displayed in magenta surface representation. $BRD4^{BD2}$ is shown in orange cartoon representation, while the model of MZ1 is presented as spheres in yellow and colored by element (PDB:5T35).

Structural studies on IMiDs-based PROTACs offered additional insights into the plasticity and functional versality of this promising class of protein degraders. Thanks to the high affinity of phthalimide moiety in IMiDs towards CRBN, the pharmacophore has been used widely to develop PROTACs for the degradation of various substrates. Similar to MZ1, phthalimide containing dBET23 sports JQ1 to recruit members of the BET protein family to CRBN for ubiquitination. The two warheads are connected through an 8-carbon linker (Nowak et al., 2018). As expected, the structure of DDB1ΔB-CRBN-dBET23-BRD4$^{BD1}$ shows that each warhead moiety of the PROTAC anchors to the binding pocket of its respective target protein (Figure 55). Also, as seen for MZ1, dBET23 induces hydrophobic and electrostatic protein-protein interactions between the neo-substrate and the E3. Interestingly, the surface used by CRBN to dock BRD4$^{BD1}$ in the presence of dBET23 differs from the ones involved in CK1α or GSPT1 recruitment mediated by Molecular Glue compounds, lenalidomide and CC-885, respectively. Moreover, in the presence of phthalimide-JQ1 PROTACs containing linkers of various length or with different linkage positions, CRBN- BRD4$^{BD1}$ were observed to adopt distinct binding conformations (Nowak et al., 2018). Structural superposition of the CRBN-BRD4$^{BD1}$ complexes induced by dBET23 and dBET57 revealed remarkable plasticity of binding between the same neo-substrates and the same E3 in the presence of variable PROTAC molecules. (Figure 56). Unlike what was seen for VHL-MZ1-BRD4$^{BD2}$, there was little to no

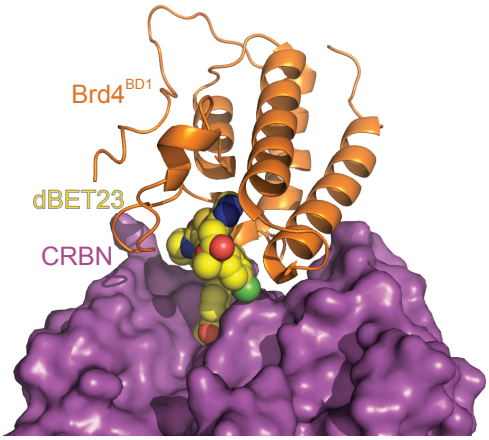



Figure 55. Structural representation of BRD4$^{BD1}$-dBET23-CRBN complex. The recognition of BRD4$^{BD1}$ by CRBN is promoted by dBET23. CRBN is displayed in magenta surface representation. BRD4$^{BD1}$ is shown in orange cartoon representation, while the model of dBET23 is presented as spheres in yellow and colored by element (PDB:6BN7).

cooperativity for the CRBN-BRD4$^{BD1}$ PROTACs, suggesting that the induced interface observed in the crystal structures might not contribute significantly to the tertiary complex formation.

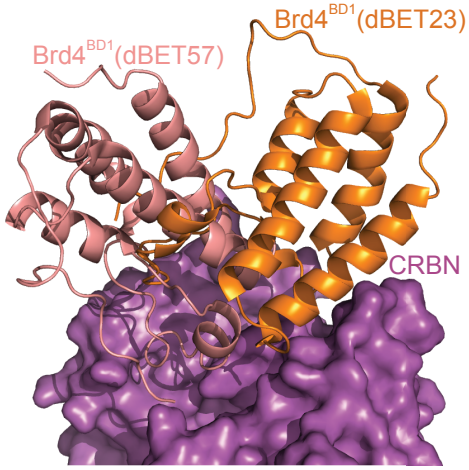



Figure 56. The structures of BRD4$^{BD1}$-dBET23-CRBN and BRD4$^{BD1}$-dBET57-CRBN. They were superimposed using CRBN-CTD. CRBN is displayed in magenta surface representation. BRD4$^{BD1}$(dBET23) is shown in orange cartoon representation, while BRD4$^{BD1}$(dBET57) is shown in light pink cartoon representation (PDB:6BN7 and 6BNB).

Despite the rapid progress made in the development of PROTACs, there are numerous examples of failed bifunctional molecules that are able to bridge neo-substrates and E3 without inducing substrate degradation. It is conceivable that, despite tertiary complex formation, the positioning of the substrate relative to the ubiquitin-conjugated and E3-anchored E2 might not be favorable for ubiquitin transfer. This could be due to the lack of available lysine receptors on the substrate or the unsurmountable gap separating the substrate from the E2. Therefore, the affinity of PROTACs towards the E3 and substrates are not the only determinants for productive ubiquitination.

## 4.4 Insights Into Molecular Glue and PROTAC

There are many benefits to developing molecular glues and PROTACs for degrading clinically relevant targets by hijacking the ubiquitin proteasome system. There are many unmet medical needs, which call for downregulation of a plethora of disease relevant proteins that are considered undruggable or unligandable. The undruggable targets often lack an active site that is important for its function but cannot be easily blocked by conventional drugs. The unligandable targets, on the other hand, do not even have any binding pocket that can be exploited by small molecules. PROTACs could be used for the degradation of undruggable targets, whereas, Molecular Glues

could be employed to destroy any target, including unligandable ones. Besides being able to downregulate targets that are intractable to conventional drugs, molecules capable of inducing targeted protein degradation have additional advantages over traditional therapeutics. Most, if not all, existing therapeutic compounds work at equilibrium conditions, where target occupancy is an important factor dictating efficacy. By contrast, protein degraders, such as Molecular Glue and PROTACs are not sequestered by the targets and operate sub-stoichiometrically. By inducing irreversible protein destruction, these molecules act catalytically, instead of stoichiometrically, in mediating the turn-over of the substrates. This property allows protein degraders to overcome the problems encountered by conventional drugs, such as target overexpression, competition with abundant natural ligands, and mutations in the target protein that evade the effect of inhibitors (Paiva and Crews, 2019, Pettersson and Crews, 2019).

Even though both Molecular Glues and PROTACs are able to reprogram E3 ligases to eliminate various targets, their different modes of action render them suitable for degrading different types of medically relevant substrates. Upon binding to a ubiquitin ligase, a Molecular Glue molecule reshapes the surface and alters the properties of the E3, thus allowing the docking of a target protein that has no detectable affinity for either the Molecular Glue or the E3. Importantly, the target protein does not have to possess a binding pocket for the protein degrader. The repertoire of the Molecular Glue target, therefore, encompasses essentially all proteins, particularly those that are unligandable. Moreover, Molecular Glues, which are not bifunctional molecules, could be very small in size and have a higher tendency to be drug-like in comparison to PROTAC molecules. While the concept of Molecular Glues sounds ideal, there is a major challenge for designing or identifying such protein degraders. Plant hormones, such as Auxin and JA, are naturally occurring small molecules. IMiDs were developed serendipitously. Without

detectable affinity toward the neo-substrate, it is not clear how one could go about identifying a Molecular Glue compound that can foster the interaction between a desirable target and an E3 enzyme. Therefore, the proof of concept for rationally developed Molecular Glue molecules for targeted protein degradation is yet to come.

PROTACs, on the other hand, are bifunctional molecules. They are required to have a moiety that has moderate to high affinity for the E3 ligase, a warhead that has moderate to high affinity for the target, and a linker between them. In order to fulfill the affinity need for PROTACs to bind, it is essential for the neo-substrate to feature a surface pocket. Therefore, unlike Molecular Glues, PROTACs are only suitable for degrading ligandable target proteins. Moreover, because of their bifunctional nature, PROTACs tend to be large in size, often exceeding a molecular weight of 500Da and violating Lipinski's Rule of Fives. This intrinsic caveat of all PROTAC compounds could lead to poor bioavailability problems and decrease their chance of having drug like properties and therapeutic values. Another challenge that PROTACs face is the hook effect associated with bifunctional molecules. Due to the strong affinity of their warheads towards the binding partners, the PROTAC molecules saturate the E3 ligases and the neo-substrates at high doses, sequestering them in separate binary complexes and preventing the ternary complex formation. This feature complicates the proper dosing of PROTACs to achieve therapeutic relief. Regardless of the aforementioned limitations, PROTACs can be rationally designed and multiple compounds have been proven to work exceptionally well. As a matter of fact, a PROTAC molecule degrading androgen receptor has already entered clinical trials.

## 4.5 KLHDC2 - A NEW PLATFORM FOR TARGETED PROTEIN DEGRADATION

The structural studies described above illustrate the importance of the PROTAC-induced neo-substrate-E3 interface in dictating substrate specificity. Due to the large number and diversity of

medically relevant targets, there is a need for developing more E3 ligase platforms for PROTACs. Based on my structural and biochemical analyses, KLHDC2 emerges as a promising new platform for PROTAC-mediated targeted protein degradation. KLHDC2 is not only expressed in different tissues, but also relatively abundant inside the cell, which makes the E3 suitable for targeting a wide variety of disease-related proteins. To date, this ubiquitin ligase is only known to participate in protein-quality control, a cellular process that is mediated by functionally redundant proteins. Therefore, rewiring KLHDC2 to downregulate neo-substrates is not expected to have detrimental effect on the cell. Moreover, this BC-box protein has been shown to be capable of degrading various neo-substrates fused to the C-end diglycine degron, suggesting that the E3 ligase can accommodate target proteins of different sizes and shapes, ubiquitinate and target them for proteasomal degradation.

Besides its tissue distribution, cellular function, and expression levels, KLHDC2 is an attractive platform for PROTACs because of its structural and biophysical characteristics. As revealed in my studies, this BC-box protein harbors a deep binding pocket, which is used for substrate degron recognition (Rusnac et al., 2018). The strong affinity between this E3 and the remarkably simple unmodified diglycine degron highlights its potential to house a degron-mimicking small molecule. The binding pocket of KLHDC2 has a well-defined shape with variable depth in different regions (~12Å in width and ~16Å in length). There might be enough room for the pocket to accommodate compounds of various shapes and sizes without sacrificing binding affinity. The pocket is also decorated by a balanced number of hydrophobic, polar, and charged amino acids, which can provide van der Waals packing surface and serve as hydrogen bond donors and acceptors for binding drug-like compounds that follow Lipinski's Rule of Fives. In order for an organic compound to function as the warhead of a PROTAC with drug-like

properties, it needs to have at least moderate affinity to the E3 and maintain a minimal molecular mass. In fact, due to the nature of PROTACs, which are bifunctional molecules with two warhead moieties and a linker, it would be highly desirable to keep each of those fragments as small as possible, preferably not exceeding 500 Da by a large margin. I have shown that KLHDC2 was able to bind the five amino acid C-end degron with a 431 Da molecular weight at 34 nM affinity. Therefore, the BC-box protein has intrinsic capability of binding small molecules tightly. Even more strikingly, I have demonstrated that the simple 132 Da diglycine motif is able to interact with the E3 ligase with an affinity of ~360 μM. This observation strongly suggests that the substrate binding interface of KLHDC2 can be exploited for developing PROTAC warheads with high affinity and minimal size.

To date, several approaches have been used in drug discovery and development to identify small molecule hit compounds. Among these, high-throughput screening (HTS) has been widely used in both industry and academia to search for hit compounds from large small-molecule libraries in a relatively quick manner. Despite the power of this method for drug discovery, there are multiple limiting factors that makes it prohibitive for practical reasons. Some of the major drawbacks are the requirement of a large amount of target proteins, a well-established and robust screening assay, and access to an HTS facility, not to mention sufficient funding. On average, the cost of high-throughput screening is estimated to be $1 per tested compound. Therefore, in order to screen against a small-sized diversity library with 100,000 compounds, it would cost around $100,000, which is more than a third of a modular NIH R01 budget or three times the annual salary of a graduate student.

Virtual drug screen is an alternative strategy in hit generation and has been gaining traction in recent years, thanks to the increasing power of computational hardware. It is not only cost-

effective, but also time effectual and labor efficient (Kar and Roy, 2013). A virtual screen entails computational docking of potential ligands with variable conformations into the binding pockets of protein targets with known structures. A scoring function is used to evaluate the probability of a ligand to interact with the target protein with high affinity. Currently, there are four general classes of scoring functions: force field, empirical, and knowledge-based, which are classical approaches, and a novel method, known as machine-learning (Ain et al., 2015). The force field scoring function relies on the predicted intermolecular electrostatic and van der Waals interactions between the atoms of the docked compound and target protein, as well as the desolvation energies of both components. While the empirical scoring function involves counting the number of interactions between the protein and the small molecule, the knowledge-based scoring function depends on potentials of mean force based on statistical observations. In contrast to these classical scoring functions, which are contingent on the predetermined functional form between binding affinity and structural features of the complex, machine-learning scoring function takes advantage of structural data from the PDB to predict functional form.

In collaboration with Accutar Biotech, we embarked on an ongoing endeavor in which artificial intelligence (AI)-aided virtual screening is used to search for small molecules that can dock to the deep degron-binging pocket of KLHDC2. After assessing the ability of a library of seven million compounds to fit the pocket of the BC-box protein, a data-driven atom-based scoring function which was learned from 100,000 protein crystal structures was used to select the top hits. Various filters were subsequently applied to single out drug-like compounds from the hit list. The resulting top hits all contain five or six membered ring structures with peripheral moieties, some of which can serve as hydrogen bond donors or acceptors. Carboxyl and/or carbonyl groups are present in the majority of these compounds, many of which also contain nitrogen, hydroxyl groups,

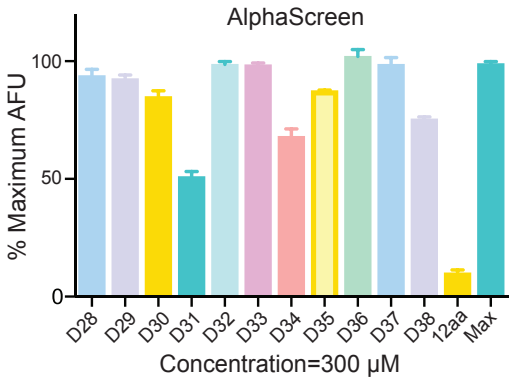



Figure 57. Activity of top virtual-screen hits tested at 300 µM. AlphaScreen competition assay for assessing the ability of artificial intelligence-aided virtual screen hit compounds to bind KLHDC2. The small molecules, D28 through D38, were tested at a final concentration of 300 µM. Free 12 aa SelK peptide in the presence of DMSO was used as a positive control. DMSO was added as a negative control. AFU: arbitrary fluorescence units. Data are measured in triplicates.

and halogens. With molecular weights ranging between 137-388 Da, the vast majority of compounds are soluble at concentrations of 100 mM in either DMSO or water. These drug-like properties make these hit compounds attractive PROTAC warheads.

In order to validate the hit compounds and to assess whether they directly target the degron binding pocket of KLHDC2, I examined the ability of the top 40 small molecules to disrupt the interaction between the 12aa SelK degron peptide and KLHDC2 using the AlphaScreen assay described in Chapter 3. First, I tested the effects of DMSO on the AlphaScreen readout and established that this organic solvent used to dissolve some of the hit compounds has detectable but marginal effects on the assay. When DMSO was kept below 5%, its effect on the AlphaScreen assay was negligible. Second, the KLHDC2 binding activity of each compound was measured at the final concentration of 300 µM, 1 mM, 5 mM or 50 mM with the 12 aa SelK tag-free peptide used as a positive control (Figure 57). The select top molecules showing activity in disrupting the interaction between KLHDC2 and the 12 aa biotinylated SelK peptide were then titrated at final concentrations ranging between 0.1 nM and 25 mM.

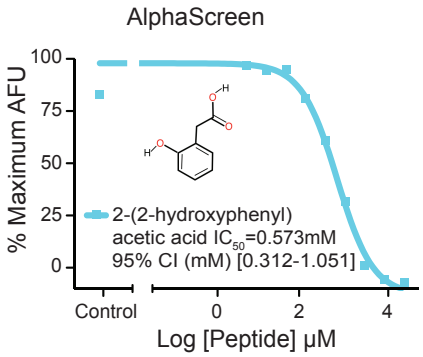



Figure 58. Activity of 2-(2-hydroxyphenyl) acetic acid. The dose response curve of the small molecule for disrupting KLHDC2- 12 aa SelK peptide binding is shown in cyan. The $IC_{50}$ value and the 95% confidence interval are listed. AFU: arbitrary fluorescence units. Data are measured in triplicates and represented as mean ± SEM.

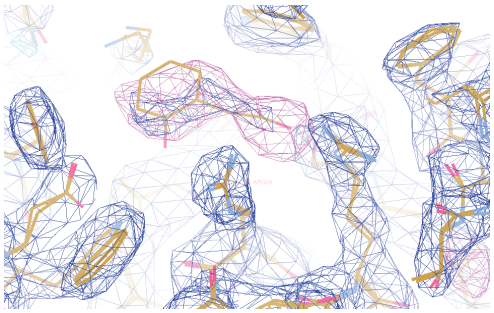


Figure 59. Positive Fo-Fc density for a hit compound. 2-(2-hydroxyphenyl) acetic acid is shown in sticks together with its positive Fo-Fc density (pink). KLHDC2 is shown in sticks together with its electron density (blue).

One of the top hits, a 152 Da small molecule, chemically known as 2-(2-hydroxyphenyl) acetic acid, was observed to have an $IC_{50}$ of 0.5 mM (Figure 58). This compound is a hydroxyl monocarboxylic acid and is derived from acetic acid with a methyl hydrogen substituted with a 2-hydroxyphenyl group. The presence of a carboxyl group in the small molecule led me to hypothesize that the negatively charged group would be interacting with the strictly conserved arginine residues of KLHDC2 in a similar manner as the common carboxyl group found in the C-end diglycine degrons does. Meanwhile, the obvious difference between this compound and the diglycine motif raises the immediate question as to how the cyclical ring would be accommodated in the deep binding pocket of KLHDC2.

In order to test the hypothesis and further understand the binding mode of the hit compound, the KLHDC2 crystals were soaked in the presence of 2-(2-hydroxyphenyl) acetic acid. After obtaining a compelling molecular replacement solution, a small yet strong positive density at the degron binding pocket was visualized in the Fo-Fc electron density map. This density can be explained by the 2-(2-hydroxyphenyl) acetic acid compound (Figure 59). As expected, the compound binds to KLHDC2 in an identical manner as the extreme C-terminal diglycine (Figure 60). The same

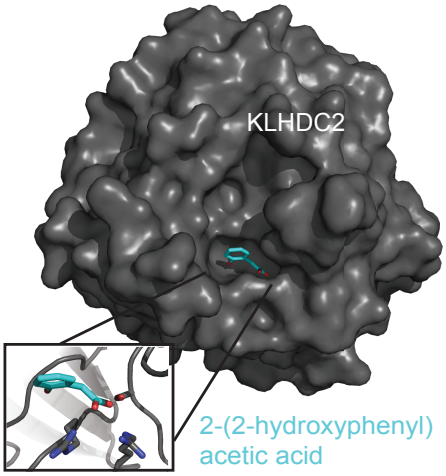



Figure 60. Structural mechanism of 2-(2-hydroxyphenyl) acetic acid binding to KLHDC2. 2-(2-hydroxyphenyl) acetic acid (shown in sticks representation in cyan) docks into the deep binding pocket of KLHDC2 (shown in surface representation in gray) in a nearly identical manner as the extreme C-terminal diglycine.

three KLHDC2 residues, Arg236, Arg241 and Ser269, which were previously revealed to recognize the carboxy group of the C-end degron, are making identical direct polar interactions with the carboxyl group of 2-(2-hydroxyphenyl) acetic acid. While the six-carbon ring is pi stacking against Tyr163, its hydroxyl group forms a hydrogen bond with Trp191. Overall, this compound takes advantage of the same interface residues on KLHDC2 as the diglycine motif does in order to anchor itself to the E3 ubiquitin ligase. Its affinity is on par with the simple diglycine peptide.

2-(2-hydroxyphenyl) acetic acid binds into the chamber C of the pocket, leaving the chamber N of KLHDC2 unoccupied and unmodified. Due to the limited number of bonds at the interface between the two, the affinity is noticeably low. It is conceivable that a larger molecule with more hydrogen bond donors and acceptors would allow additional interactions to be made with the remainder of the degron interface (chamber N), thereby, enhancing the affinity of the compound. After multiple iterative rounds of virtual drug screening followed by biophysical characterization with AlphaScreen, I was able to identify molecules with significantly augmented potencies in the low µM range (Figure 61). These hit compounds all share ring structures, carboxyl groups, and various hydrogen bond donors and acceptors. In order to probe if the binding mode of a leading compound, M01C, to the E3 ligase resembles 2-(2-hydroxyphenyl) acetic acid, the new compound was soaked into the KLHDC2 crystals. The resulting

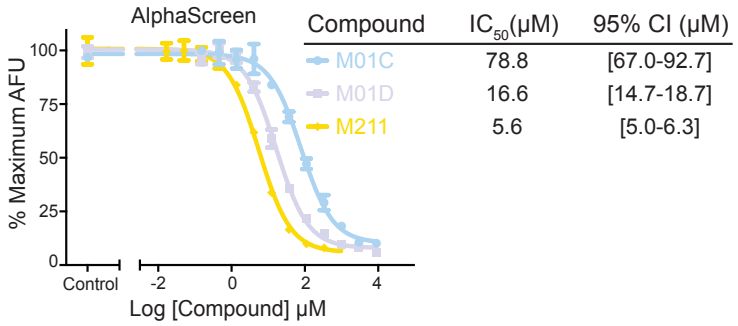



Figure 61. Activity of M01C, M01D, and M211. The dose response curves of the small molecules are shown in pastel blue, purple and yellow, respectively. The $IC_{50}$ value and the 95% confidence intervals are listed on the right. AFU: arbitrary fluorescence units. Data are measured in triplicates and represented as mean ± SEM.

complex structure unveiled that M01C docks to the kelch repeat domain of KLHDC2 in a similar manner as 2-(2-hydroxyphenyl) acetic acid. As predicted, M01C also took advantage of a larger portion of the binding pocket at chamber N and formed more hydrogen bonds with KLHDC2 (Figure 62). Due to proprietary reasons, it is premature to reveal the chemical structure of M01C in this chapter.

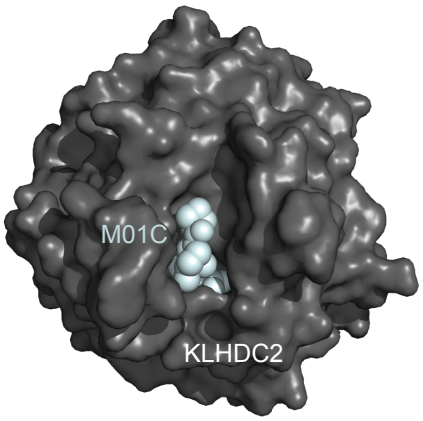



Figure 62. Structural basis of M01C-KLHDC2 interaction. M01C (shown in spheres representation in pastel blue) binds into the deep binding pocket of KLHDC2 (shown in surface representation in gray) in a nearly identical manner as the extreme C-terminal diglycine.

Now that I have hit compounds with higher affinity, the next step is to obtain proof of principle that KLHDC2 could serve as a PROTAC platform for targeted protein degradation. To do so, the small molecules will be used as a warhead to be conjugated to other established neo-substrate warheads via a linker. As an example, M01C could be linked to OVA, and the tertiary complex formation between KLHDC2, M01C-OVA and MetAP2 can be assessed using pull-down assays. The ubiquitination of MetAP2 will be determined by *in vitro* assays, while its degradation will be monitored through *in vivo* assays by western blotting against MetAP2. Once proof of concept is obtained, KLHDC2 can be used to degrade various target proteins involved in diseases such as cancer and Alzheimer's.

## 4.6 METHOD DETAILS

### 4.6.1 *Experimental Model and Subject Details*

For DNA extraction, *E. coli* DH5α was grown for 16 hr at 37 °C. For protein expression, *E. coli* BL21(DE3) was grown for 16 hr at 16-18 °C after 0.0015 mM IPTG induction. LB Broth Miller (Fisher BioReagents) was used for *E. coli* culture.

### 4.6.2 *Molecular Biology and Protein Purification*

The kelch repeat domain of human KLHDC2 (amino acid 22–362) was subcloned into the pET vector with an N-terminally fused six Histidine (HIS) tag and a glutathione-S-transferase (GST) tag followed by thrombin (Thr) and TEV cleavage sites, which were followed by SUMO, TEV and Thr cleavage site (HIS-GST-Thr-TEV-SUMO-TEV-Thr-GSGSGS-K2_shorter). The 24 liters of cell cultures were grown in LB broth at 37 °C to an $OD_{600}$ of 0.8 and induced with 0.0015 mM IPTG. After 16 hours, the cells were harvested, re-suspended and lysed in lysis buffer (20 mM Tris, pH 8.0, 200 mM NaCl, 5 mM DTT) in the presence of protease inhibitors (1 μg/ml Leupeptin, 1 μg/ml Pepstatin and 100 μM PMSF) using a microfluidizer. The GST-KLHDC2 protein was isolated from the soluble cell lysate by Pierce™ Glutathione Agarose (Thermo Scientific). For crystallization, the thrombin (1:50 mass ratio) cleaved KLHDC2 was further purified by Mono Q anion exchange chromatography (GE Healthcare) twice. The NaCl eluates were subjected to Superdex-200 size exclusion chromatography (GE Healthcare). The samples used for crystallization were concentrated by ultrafiltration to 15–20 mg $mL^{-1}$. All protein samples were flash frozen in liquid nitrogen for future use.

### 4.6.3 *Protein Crystallization*

The crystals of KLHDC2 were grown at 25 °C by the hanging-drop vapor diffusion method with 0.150 μL protein complex sample mixed with 0.075 μL volume of reservoir solution containing 0.03 M $MgCl_2*6H_2O$, 0.03 M $CaCl_2*2H_2O$, 0.05 M Tris (base), 0.05 M 2-(Bis(2-hydroxyethyl)amino)acetic acid (BICINE), 20% (v/v) Poly(ethylene glycol) methyl ether 500 (PEG500 MME), 10% (w/v) Poly(ethylene glycol) 20,000 (PEG 20,000), pH 8.5. Crystals of maximal sizes were obtained and harvested after a few days. Cryoprotection was provided by the

crystallization condition, which was supplemented with small molecules at the final concentration of 5 mM for soaking.

### 4.6.4 *Data Collection and Structure Determination*

After collecting a native dataset at Advanced Light Source Beamline 8.2.1, X-ray diffraction data were integrated and scaled with HKL2000 package (Otwinowski and Minor, 1997). The structure was solved by molecular replacement and phaser from the PHENIX suite of programs and the KLHDC2 structures described in chapter 3 as the search template (PDB: 6DO3, 6DO4, 6DO5). All structural models were built, refined and rebuilt using COOT (Emsley et al., 2010) and PHENIX (Adams et al., 2002). PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) was used to generate figures.

### 4.6.5 *AlphaScreen Luminescence Proximity Assay*

AlphaScreen assays for determining and measuring protein-protein interactions were performed using EnSpire reader (PerkinElmer). GST-tagged KLHDC2 (WT) was attached to anti-GST AlphaScreen acceptor beads. Synthetic biotinylated 12 aa SelK degron peptide (Bio-Synthesis, Inc.) was immobilized to streptavidin-coated AlphaScreen donor beads. The donor and acceptor beads were brought into proximity by the interactions between the SelK peptide and KLHDC2. Excitation of the donor beads by a laser beam of 680 nm promotes the formation of singlet oxygen. When an acceptor bead is in close proximity, the singlet oxygen reacts with thioxene derivatives in the acceptor beads and causes the emission of 520-620 nm photons, which are detected as the binding signal. If the beads are not in close proximity to each other, the oxygen will return to its ground state and the acceptor beads will not emit light. Competition assays were performed in the presence of small molecules which were titrated at various concentrations.

The experiments were conducted with 0.12 nM of GST-KLHDC2 and 1.7 nM biotinylated 12 aa SelK peptide in the presence of 5 μg/ml donor and acceptor beads in a buffer of 25 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM TCEP, 0.1% Tween-20, and 0.05 mg/ml Bovine Serum Albumin. The concentrations of the compounds used in competition assays ranged from 0.1 nM to 25 mM. The experiments were done in triplicates. $IC_{50}$ values were determined using non-linear curve fitting of the dose response curves generated with Prism 4 (GraphPad).

# VITA

Domnița-Valeria Rusnac was raised in Chisinau, Moldova and is an alumna of Mircea Eliade High School, Class of 2008. Domnița attended undergraduate at the University of Montana where she received a Bachelor of Arts in Cell and Molecular Biology. She was awarded the President's Outstanding Senior Recognition Award, which recognized her as the outstanding graduating senior in her department on the basis of academic performance and devotion to excellence. At the University of Montana, Domnița worked with Dr. Klara Briknarova on investigating the interaction of Anastellin with fragments from the third type III domain of fibronectin and she determined the structure of the eleventh type III domain of fibronectin (11FN3) using X-ray crystallography.

Pursuing her interest in structure guided drug discovery, Domnița joined the Biological Physics, Structure, and Design graduate program at the University of Washington. Upon completing her rotations, Domnița joined the laboratory of Dr. Ning Zheng of the Pharmacology Department in June of 2015. Throughout her graduate work, Domnița elucidated the molecular mechanism that underlines the C-end diglycine degron recognition by $CRL2^{KLHDC2}$ and worked on assessing the ability of the E3 to serve as a new platform for targeted protein degradation. After completion of the University of Washington PhD program, Domnița will continue working for Dr. Ning Zheng to translate the latest findings of structures and functions of human ubiquitin ligases for therapeutic development.

**References**


ADAMS, P. D., GROSSE-KUNSTLEVE, R. W., HUNG, L. W., IOERGER, T. R., MCCOY, A. J., MORIARTY, N. W., READ, R. J., SACCHETTINI, J. C., SAUTER, N. K. & TERWILLIGER, T. C. 2002. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr,* 58**,** 1948-54.

AIN, Q. U., ALEKSANDROVA, A., ROESSLER, F. D. & BALLESTER, P. J. 2015. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdiscip Rev Comput Mol Sci,* 5**,** 405-424.

ALLEN, S. J., GILES, K., GILBERT, T. & BUSH, M. F. 2016. Ion mobility mass spectrometry of peptide, protein, and protein complex ions using a radio-frequency confining drift cell. *Analyst,* 141**,** 884-91.

AMBROGELLY, A., PALIOURA, S. & SÖLL, D. 2007. Natural expansion of the genetic code. *Nat Chem Biol,* 3**,** 29-35.

ANGERS, S., LI, T., YI, X., MACCOSS, M. J., MOON, R. T. & ZHENG, N. 2006. Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature,* 443**,** 590-3.

ARAI, T., KASPER, J. S., SKAAR, J. R., ALI, S. H., TAKAHASHI, C. & DECAPRIO, J. A. 2003. Targeted disruption of p185/Cul7 gene results in abnormal vascular morphogenesis. *Proc Natl Acad Sci U S A,* 100**,** 9855-60.

BANFIELD, M. J. 2015. Perturbation of host ubiquitin systems by plant pathogen/pest effector proteins. *Cell Microbiol,* 17**,** 18-25.

BHATTACHARYYA, S., YU, H., MIM, C. & MATOUSCHEK, A. 2014. Regulated protein turnover: snapshots of the proteasome in action. *Nat Rev Mol Cell Biol,* 15**,** 122-33.

BUSINO, L., BASSERMANN, F., MAIOLICA, A., LEE, C., NOLAN, P. M., GODINHO, S. I., DRAETTA, G. F. & PAGANO, M. 2007. SCFFbxl3 controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins. *Science,* 316**,** 900-4.

CALABRESE, M. F., SCOTT, D. C., DUDA, D. M., GRACE, C. R., KURINOV, I., KRIWACKI, R. W. & SCHULMAN, B. A. 2011. A RING E3-substrate complex poised for ubiquitin-like protein transfer: structural insights into cullin-RING ligases. *Nat Struct Mol Biol,* 18**,** 947-9.

CALLIS, J. 2014. The ubiquitination machinery of the ubiquitin system. *Arabidopsis Book,* 12**,** e0174.

CARDOTE, T. A. F., GADD, M. S. & CIULLI, A. 2017. Crystal Structure of the Cul2-Rbx1-EloBC-VHL Ubiquitin Ligase Complex. *Structure,* 25**,** 901-911.e3.

CAVADINI, S., FISCHER, E. S., BUNKER, R. D., POTENZA, A., LINGARAJU, G. M., GOLDIE, K. N., MOHAMED, W. I., FATY, M., PETZOLD, G., BECKWITH, R. E., TICHKULE, R. B., HASSIEPEN, U., ABDULRAHMAN, W., PANTELIC, R. S., MATSUMOTO, S., SUGASAWA, K., STAHLBERG, H. & THOMÄ, N. H. 2016. Cullin-RING ubiquitin E3 ligase regulation by the COP9 signalosome. *Nature,* 531**,** 598-603.

CHAMOVITZ, D. A., WEI, N., OSTERLUND, M. T., VON ARNIM, A. G., STAUB, J. M., MATSUI, M. & DENG, X. W. 1996. The COP9 complex, a novel multisubunit nuclear regulator involved in light control of a plant developmental switch. *Cell,* 86**,** 115-21.

CHEN, S. J., WU, X., WADAS, B., OH, J. H. & VARSHAVSKY, A. 2017. An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science,* 355.

CHOI, W. S., JEONG, B. C., JOO, Y. J., LEE, M. R., KIM, J., ECK, M. J. & SONG, H. K. 2010. Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. *Nat Struct Mol Biol,* 17**,** 1175-81.

COPE, G. A., SUH, G. S., ARAVIND, L., SCHWARZ, S. E., ZIPURSKY, S. L., KOONIN, E. V. & DESHAIES, R. J. 2002. Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of Nedd8 from Cul1. *Science,* 298**,** 608-11.

COUX, O., TANAKA, K. & GOLDBERG, A. L. 1996. Structure and functions of the 20S and 26S proteasomes. *Annu Rev Biochem,* 65**,** 801-47.

DA FONSECA, P. C., KONG, E. H., ZHANG, Z., SCHREIBER, A., WILLIAMS, M. A., MORRIS, E. P. & BARFORD, D. 2011. Structures of APC/C(Cdh1) with substrates identify Cdh1 and Apc10 as the D-box co-receptor. *Nature,* 470**,** 274-8.

DECORSIÈRE, A., MUELLER, H., VAN BREUGEL, P. C., ABDUL, F., GEROSSIER, L., BERAN, R. K., LIVINGSTON, C. M., NIU, C., FLETCHER, S. P., HANTZ, O. & STRUBIN, M. 2016. Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature,* 531**,** 386-9.

DESHAIES, R. J. & JOAZEIRO, C. A. 2009. RING domain E3 ubiquitin ligases. *Annu Rev Biochem,* 78**,** 399-434.

DICKINSON, M. E., FLENNIKEN, A. M., JI, X., TEBOUL, L., WONG, M. D., WHITE, J. K., MEEHAN, T. F., WENINGER, W. J., WESTERBERG, H., ADISSU, H., BAKER, C. N., BOWER, L., BROWN, J. M., CADDLE, L. B., CHIANI, F., CLARY, D., CLEAK, J., DALY, M. J., DENEGRE, J. M., DOE, B., DOLAN, M. E., EDIE, S. M., FUCHS, H., GAILUS-DURNER, V., GALLI, A., GAMBADORO, A., GALLEGOS, J., GUO, S., HORNER, N. R., HSU, C. W., JOHNSON, S. J., KALAGA, S., KEITH, L. C., LANOUE, L., LAWSON, T. N., LEK, M., MARK, M., MARSCHALL, S., MASON, J., MCELWEE, M. L., NEWBIGGING, S., NUTTER, L. M., PETERSON, K. A., RAMIREZ-SOLIS, R., ROWLAND, D. J., RYDER, E., SAMOCHA, K. E., SEAVITT, J. R., SELLOUM, M., SZOKE-KOVACS, Z., TAMURA, M., TRAINOR, A. G., TUDOSE, I., WAKANA, S., WARREN, J., WENDLING, O., WEST, D. B., WONG, L., YOSHIKI, A., MACARTHUR, D. G., TOCCHINI-VALENTINI, G. P., GAO, X., FLICEK, P., BRADLEY, A., SKARNES, W. C., JUSTICE, M. J., PARKINSON, H. E., MOORE, M., WELLS, S., BRAUN, R. E., SVENSON, K. L., DE ANGELIS, M. H., HERAULT, Y., MOHUN, T., MALLON, A. M., HENKELMAN, R. M., BROWN, S. D., ADAMS, D. J., LLOYD, K. C., MCKERLIE, C., BEAUDET, A. L., BUĆAN, M., MURRAY, S. A., CONSORTIUM, I. M. P., LABORATORY, J., INFRASTRUCTURE NATIONALE PHENOMIN, I. S. C. D. L. S. I., LABORATORIES, C. R., HARWELL, M., PHENOGENOMICS, T. C. F., INSTITUTE, W. T. S. & CENTER, R. B. 2016. High-throughput discovery of novel developmental phenotypes. *Nature,* 537**,** 508-514.

DONG, C., ZHANG, H., LI, L., TEMPEL, W., LOPPNAU, P. & MIN, J. 2018. Molecular basis of GID4-mediated recognition of degrons for the Pro/N-end rule pathway. *Nat Chem Biol,* 14**,** 466-473.

DONOVAN, K. A., AN, J., NOWAK, R. P., YUAN, J. C., FINK, E. C., BERRY, B. C., EBERT, B. L. & FISCHER, E. S. 2018. Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane Radial Ray syndrome. *Elife,* 7.

DOU, H., BUETOW, L., SIBBET, G. J., CAMERON, K. & HUANG, D. T. 2012. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat Struct Mol Biol,* 19**,** 876-83.

DRISCOLL, D. M. & COPELAND, P. R. 2003. Mechanism and regulation of selenoprotein synthesis. *Annu Rev Nutr,* 23**,** 17-40.

DUDA, D. M., BORG, L. A., SCOTT, D. C., HUNT, H. W., HAMMEL, M. & SCHULMAN, B. A. 2008. Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell,* 134**,** 995-1006.

DUDA, D. M., OLSZEWSKI, J. L., TRON, A. E., HAMMEL, M., LAMBERT, L. J., WADDELL, M. B., MITTAG, T., DECAPRIO, J. A. & SCHULMAN, B. A. 2012. Structure of a glomulin-RBX1-CUL1 complex: inhibition of a RING E3 ligase through masking of its E2-binding surface. *Mol Cell,* 47**,** 371-82.

DUDA, D. M., SCOTT, D. C., CALABRESE, M. F., ZIMMERMAN, E. S., ZHENG, N. & SCHULMAN, B. A. 2011. Structural regulation of cullin-RING ubiquitin ligase complexes. *Curr Opin Struct Biol,* 21**,** 257-64.

EMANUELE, M. J., ELIA, A. E., XU, Q., THOMA, C. R., IZHAR, L., LENG, Y., GUO, A., CHEN, Y. N., RUSH, J., HSU, P. W., YEN, H. C. & ELLEDGE, S. J. 2011. Global identification of modular cullin-RING ligase substrates. *Cell,* 147**,** 459-74.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66**,** 486-501.

ENCHEV, R. I., SCOTT, D. C., DA FONSECA, P. C., SCHREIBER, A., MONDA, J. K., SCHULMAN, B. A., PETER, M. & MORRIS, E. P. 2012. Structural basis for a reciprocal regulation between SCF and CSN. *Cell Rep,* 2**,** 616-27.

FISCHER, E. S., SCRIMA, A., BÖHM, K., MATSUMOTO, S., LINGARAJU, G. M., FATY, M., YASUDA, T., CAVADINI, S., WAKASUGI, M., HANAOKA, F., IWAI, S., GUT, H., SUGASAWA, K. & THOMÄ, N. H. 2011. The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell,* 147**,** 1024-39.

FUKUTOMI, T., TAKAGI, K., MIZUSHIMA, T., OHUCHI, N. & YAMAMOTO, M. 2014. Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1. *Mol Cell Biol,* 34**,** 832-46.

GADD, M. S., TESTA, A., LUCAS, X., CHAN, K. H., CHEN, W., LAMONT, D. J., ZENGERLE, M. & CIULLI, A. 2017. Structural basis of PROTAC cooperative recognition for selective protein degradation. *Nat Chem Biol,* 13**,** 514-521.

GEYER, R., WEE, S., ANDERSON, S., YATES, J. & WOLF, D. A. 2003. BTB/POZ domain proteins are putative substrate adaptors for cullin 3 ubiquitin ligases. *Mol Cell,* 12**,** 783-90.

GODINHO, S. I., MAYWOOD, E. S., SHAW, L., TUCCI, V., BARNARD, A. R., BUSINO, L., PAGANO, M., KENDALL, R., QUWAILID, M. M., ROMERO, M. R., O'NEILL, J., CHESHAM, J. E., BROOKER, D., LALANNE, Z., HASTINGS, M. H. & NOLAN, P. M. 2007. The after-hours mutant reveals a role for Fbxl3 in determining mammalian circadian period. *Science,* 316**,** 897-900.

GOLDBERG, A. L. 2003. Protein degradation and protection against misfolded or damaged proteins. *Nature,* 426**,** 895-9.

GOLDENBERG, S. J., CASCIO, T. C., SHUMWAY, S. D., GARBUTT, K. C., LIU, J., XIONG, Y. & ZHENG, N. 2004. Structure of the Cand1-Cul1-Roc1 complex reveals

regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. *Cell,* 119**,** 517-28.
GORITSCHNIG, S., ZHANG, Y. & LI, X. 2007. The ubiquitin pathway is required for innate immunity in Arabidopsis. *Plant J,* 49**,** 540-51.
GUHAROY, M., BHOWMICK, P., SALLAM, M. & TOMPA, P. 2016. Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system. *Nat Commun,* 7**,** 10239.
GUO, Y., DONG, L., QIU, X., WANG, Y., ZHANG, B., LIU, H., YU, Y., ZANG, Y., YANG, M. & HUANG, Z. 2014. Structural basis for hijacking CBF-β and CUL5 E3 ligase complex by HIV-1 Vif. *Nature,* 505**,** 229-33.
HAO, B., OEHLMANN, S., SOWA, M. E., HARPER, J. W. & PAVLETICH, N. P. 2007. Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol Cell,* 26**,** 131-43.
HAO, B., ZHENG, N., SCHULMAN, B. A., WU, G., MILLER, J. J., PAGANO, M. & PAVLETICH, N. P. 2005. Structural basis of the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase. *Mol Cell,* 20**,** 9-19.
HERSHKO, A. & CIECHANOVER, A. 1998. The ubiquitin system. *Annu Rev Biochem,* 67**,** 425-79.
HORI, T., OSAKA, F., CHIBA, T., MIYAMOTO, C., OKABAYASHI, K., SHIMBARA, N., KATO, S. & TANAKA, K. 1999. Covalent modification of all members of human cullin family proteins by NEDD8. *Oncogene,* 18**,** 6829-34.
HORVATH, C. M. 2004. Weapons of STAT destruction. Interferon evasion by paramyxovirus V protein. *Eur J Biochem,* 271**,** 4621-8.
HUANG, L., KINNUCAN, E., WANG, G., BEAUDENON, S., HOWLEY, P. M., HUIBREGTSE, J. M. & PAVLETICH, N. P. 1999. Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science,* 286**,** 1321-6.
IVAN, M., KONDO, K., YANG, H., KIM, W., VALIANDO, J., OHH, M., SALIC, A., ASARA, J. M., LANE, W. S. & KAELIN , W. G. 2001. HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science,* 292**,** 464-8.
JIN, J., ARIAS, E. E., CHEN, J., HARPER, J. W. & WALTER, J. C. 2006. A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1. *Mol Cell,* 23**,** 709-21.
JIN, J., CARDOZO, T., LOVERING, R. C., ELLEDGE, S. J., PAGANO, M. & HARPER, J. W. 2004. Systematic analysis and nomenclature of mammalian F-box proteins. *Genes Dev,* 18**,** 2573-80.
JIN, J., LI, X., GYGI, S. P. & HARPER, J. W. 2007. Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging. *Nature,* 447**,** 1135-8.
KAELIN, W. G. 2005. Proline hydroxylation and gene expression. *Annu Rev Biochem,* 74**,** 115-28.
KAR, S. & ROY, K. 2013. How far can virtual screening take us in drug discovery? *Expert Opin Drug Discov,* 8**,** 245-61.
KIM, D. Y., KWON, E., HARTLEY, P. D., CROSBY, D. C., MANN, S., KROGAN, N. J. & GROSS, J. D. 2013. CBFβ stabilizes HIV Vif to counteract APOBEC3 at the expense of RUNX1 target gene expression. *Mol Cell,* 49**,** 632-44.

KISH-TRIER, E. & HILL, C. P. 2013. Structural biology of the proteasome. *Annu Rev Biophys,* 42**,** 29-49.
KOMANDER, D., CLAGUE, M. J. & URBÉ, S. 2009. Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol,* 10**,** 550-63.
KOMANDER, D. & RAPE, M. 2012. The ubiquitin code. *Annu Rev Biochem,* 81**,** 203-29.
KOREN, I., TIMMS, R. T., KULA, T., XU, Q., LI, M. Z. & ELLEDGE, S. J. 2018. The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell*.
KRAFT, C., VODERMAIER, H. C., MAURER-STROH, S., EISENHABER, F. & PETERS, J. M. 2005. The WD40 propeller domain of Cdh1 functions as a destruction box receptor for APC/C substrates. *Mol Cell,* 18**,** 543-53.
KRÖNKE, J., FINK, E. C., HOLLENBACH, P. W., MACBETH, K. J., HURST, S. N., UDESHI, N. D., CHAMBERLAIN, P. P., MANI, D. R., MAN, H. W., GANDHI, A. K., SVINKINA, T., SCHNEIDER, R. K., MCCONKEY, M., JÄRÅS, M., GRIFFITHS, E., WETZLER, M., BULLINGER, L., CATHERS, B. E., CARR, S. A., CHOPRA, R. & EBERT, B. L. 2015. Lenalidomide induces ubiquitination and degradation of CK1α in del(5q) MDS. *Nature,* 523**,** 183-8.
KRÖNKE, J., UDESHI, N. D., NARLA, A., GRAUMAN, P., HURST, S. N., MCCONKEY, M., SVINKINA, T., HECKL, D., COMER, E., LI, X., CIARLO, C., HARTMAN, E., MUNSHI, N., SCHENONE, M., SCHREIBER, S. L., CARR, S. A. & EBERT, B. L. 2014. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science,* 343**,** 301-5.
KURZ, T., OZLÜ, N., RUDOLF, F., O'ROURKE, S. M., LUKE, B., HOFMANN, K., HYMAN, A. A., BOWERMAN, B. & PETER, M. 2005. The conserved protein DCN-1/Dcn1p is required for cullin neddylation in C. elegans and S. cerevisiae. *Nature,* 435**,** 1257-61.
LANDER, G. C., ESTRIN, E., MATYSKIELA, M. E., BASHORE, C., NOGALES, E. & MARTIN, A. 2012. Complete subunit architecture of the proteasome regulatory particle. *Nature,* 482**,** 186-91.
LASKOWSKI, R. A. & SWINDELLS, M. B. 2011. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model,* 51**,** 2778-86.
LECHTENBERG, B. C., RAJPUT, A., SANISHVILI, R., DOBACZEWSKA, M. K., WARE, C. F., MACE, P. D. & RIEDL, S. J. 2016. Structure of a HOIP/E2~ubiquitin complex reveals RBR E3 ligase mechanism and regulation. *Nature,* 529**,** 546-50.
LI, T., CHEN, X., GARBUTT, K. C., ZHOU, P. & ZHENG, N. 2006. Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. *Cell,* 124**,** 105-17.
LI, T., ROBERT, E. I., VAN BREUGEL, P. C., STRUBIN, M. & ZHENG, N. 2010. A promiscuous alpha-helical motif anchors viral hijackers and substrate receptors to the CUL4-DDB1 ubiquitin ligase machinery. *Nat Struct Mol Biol,* 17**,** 105-11.
LIAKOPOULOS, D., DOENGES, G., MATUSCHEWSKI, K. & JENTSCH, S. 1998. A novel protein modification pathway related to the ubiquitin system. *EMBO J,* 17**,** 2208-14.
LIN, H. C., HO, S. C., CHEN, Y. Y., KHOO, K. H., HSU, P. H. & YEN, H. C. 2015. SELENOPROTEINS. CRL2 aids elimination of truncated selenoproteins produced by failed UGA/Sec decoding. *Science,* 349**,** 91-5.

LIN, H. C., YEH, C. W., CHEN, Y. F., LEE, T. T., HSIEH, P. Y., RUSNAC, D. V., LIN, S. Y., ELLEDGE, S. J., ZHENG, N. & YEN, H. S. 2018. C-Terminal End-Directed Protein Elimination by CRL2 Ubiquitin Ligases. *Mol Cell,* 70**,** 602-613.e3.
LINGARAJU, G. M., BUNKER, R. D., CAVADINI, S., HESS, D., HASSIEPEN, U., RENATUS, M., FISCHER, E. S. & THOMÄ, N. H. 2014. Crystal structure of the human COP9 signalosome. *Nature,* 512**,** 161-5.
LIU, J., FURUKAWA, M., MATSUMOTO, T. & XIONG, Y. 2002. NEDD8 modification of CUL1 dissociates p120(CAND1), an inhibitor of CUL1-SKP1 binding and SCF ligases. *Mol Cell,* 10**,** 1511-8.
LIU, X., REITSMA, J. M., MAMROSH, J. L., ZHANG, Y., STRAUBE, R. & DESHAIES, R. J. 2018. Cand1-Mediated Adaptive Exchange Mechanism Enables Variation in F-Box Protein Expression. *Mol Cell,* 69**,** 773-786.e6.
LU, G., MIDDLETON, R. E., SUN, H., NANIONG, M., OTT, C. J., MITSIADES, C. S., WONG, K. K., BRADNER, J. E. & KAELIN, W. G. 2014. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science,* 343**,** 305-9.
LUCAS, X. & CIULLI, A. 2017. Recognition of substrate degrons by E3 ubiquitin ligases and modulation by small-molecule mimicry strategies. *Curr Opin Struct Biol,* 44**,** 101-110.
MAHROUR, N., REDWINE, W. B., FLORENS, L., SWANSON, S. K., MARTIN-BROWN, S., BRADFORD, W. D., STAEHLING-HAMPTON, K., WASHBURN, M. P., CONAWAY, R. C. & CONAWAY, J. W. 2008. Characterization of Cullin-box sequences that direct recruitment of Cul2-Rbx1 and Cul5-Rbx2 modules to Elongin BC-based ubiquitin ligases. *J Biol Chem,* 283**,** 8005-13.
MARÍN, I. 2010. Diversification and Specialization of Plant RBR Ubiquitin Ligases. *PLoS One,* 5**,** e11579.
MARÍN, I. 2013. Evolution of plant HECT ubiquitin ligases. *PLoS One,* 8**,** e68536.
MATTA-CAMACHO, E., KOZLOV, G., LI, F. F. & GEHRING, K. 2010. Structural basis of substrate recognition and specificity in the N-end rule pathway. *Nat Struct Mol Biol,* 17**,** 1182-7.
MATYSKIELA, M. E., COUTO, S., ZHENG, X., LU, G., HUI, J., STAMP, K., DREW, C., REN, Y., WANG, M., CARPENTER, A., LEE, C. W., CLAYTON, T., FANG, W., LU, C. C., RILEY, M., ABDUBEK, P., BLEASE, K., HARTKE, J., KUMAR, G., VESSEY, R., ROLFE, M., HAMANN, L. G. & CHAMBERLAIN, P. P. 2018. SALL4 mediates teratogenicity as a thalidomide-dependent cereblon substrate. *Nat Chem Biol,* 14**,** 981-987.
MATYSKIELA, M. E., LU, G., ITO, T., PAGARIGAN, B., LU, C. C., MILLER, K., FANG, W., WANG, N. Y., NGUYEN, D., HOUSTON, J., CARMEL, G., TRAN, T., RILEY, M., NOSAKA, L., LANDER, G. C., GAIDAROVA, S., XU, S., RUCHELMAN, A. L., HANDA, H., CARMICHAEL, J., DANIEL, T. O., CATHERS, B. E., LOPEZ-GIRONA, A. & CHAMBERLAIN, P. P. 2016. A novel cereblon modulator recruits GSPT1 to the CRL4(CRBN) ubiquitin ligase. *Nature,* 535**,** 252-7.
METZGER, M. B., HRISTOVA, V. A. & WEISSMAN, A. M. 2012. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J Cell Sci,* 125**,** 531-7.
MEYER, H. J. & RAPE, M. 2014. Enhanced protein degradation by branched ubiquitin chains. *Cell,* 157**,** 910-21.

MIN, J. H., YANG, H., IVAN, M., GERTLER, F., KAELIN, W. G. & PAVLETICH, N. P. 2002. Structure of an HIF-1alpha -pVHL complex: hydroxyproline recognition in signaling. *Science,* 296**,** 1886-9.
MOSADEGHI, R., REICHERMEIER, K. M., WINKLER, M., SCHREIBER, A., REITSMA, J. M., ZHANG, Y., STENGEL, F., CAO, J., KIM, M., SWEREDOSKI, M. J., HESS, S., LEITNER, A., AEBERSOLD, R., PETER, M., DESHAIES, R. J. & ENCHEV, R. I. 2016. Structural and kinetic analysis of the COP9-Signalosome activation and the cullin-RING ubiquitin ligase deneddylation cycle. *Elife,* 5.
MÉSZÁROS, B., KUMAR, M., GIBSON, T. J., UYAR, B. & DOSZTÁNYI, Z. 2017. Degrons in cancer. *Sci Signal,* 10.
NANGLE, S., XING, W. & ZHENG, N. 2013. Crystal structure of mammalian cryptochrome in complex with a small molecule competitor of its ubiquitin ligase. *Cell Res,* 23**,** 1417-9.
NGUYEN, H. C., YANG, H., FRIBOURGH, J. L., WOLFE, L. S. & XIONG, Y. 2015. Insights into Cullin-RING E3 ubiquitin ligase recruitment: structure of the VHL-EloBC-Cul2 complex. *Structure,* 23**,** 441-449.
NOWAK, R. P., DEANGELO, S. L., BUCKLEY, D., HE, Z., DONOVAN, K. A., AN, J., SAFAEE, N., JEDRYCHOWSKI, M. P., PONTHIER, C. M., ISHOEY, M., ZHANG, T., MANCIAS, J. D., GRAY, N. S., BRADNER, J. E. & FISCHER, E. S. 2018. Plasticity in binding confers selectivity in ligand-induced protein degradation. *Nat Chem Biol,* 14**,** 706-714.
ORLICKY, S., TANG, X., WILLEMS, A., TYERS, M. & SICHERI, F. 2003. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell,* 112**,** 243-56.
OSAKA, F., KAWASAKI, H., AIDA, N., SAEKI, M., CHIBA, T., KAWASHIMA, S., TANAKA, K. & KATO, S. 1998. A new NEDD8-ligating system for cullin-4A. *Genes Dev,* 12**,** 2263-8.
OTWINOWSKI, Z. & MINOR, W. (eds.) 1997. *Processing of X-ray Diffraction Data Collected in Oscillation Mode,* New York: Academic Press.
PADMANABHAN, B., TONG, K. I., OHTA, T., NAKAMURA, Y., SCHARLOCK, M., OHTSUJI, M., KANG, M. I., KOBAYASHI, A., YOKOYAMA, S. & YAMAMOTO, M. 2006. Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol Cell,* 21**,** 689-700.
PAIVA, S. L. & CREWS, C. M. 2019. Targeted protein degradation: elements of PROTAC design. *Curr Opin Chem Biol,* 50**,** 111-119.
PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem,* 25**,** 1605-12.
PETTERSSON, M. & CREWS, C. M. 2019. PROteolysis TArgeting Chimeras (PROTACs) - Past, present and future. *Drug Discov Today Technol,* 31**,** 15-27.
PETZOLD, G., FISCHER, E. S. & THOMÄ, N. H. 2016. Structural basis of lenalidomide-induced CK1α degradation by the CRL4(CRBN) ubiquitin ligase. *Nature,* 532**,** 127-30.
PICKART, C. M. 2001. Mechanisms underlying ubiquitination. *Annu Rev Biochem,* 70**,** 503-33.
PIERCE, N. W., KLEIGER, G., SHAN, S. O. & DESHAIES, R. J. 2009. Detection of sequential polyubiquitylation on a millisecond timescale. *Nature,* 462**,** 615-9.
PIERCE, N. W., LEE, J. E., LIU, X., SWEREDOSKI, M. J., GRAHAM, R. L., LARIMORE, E. A., ROME, M., ZHENG, N., CLURMAN, B. E., HESS, S., SHAN, S. O. & DESHAIES,

R. J. 2013. Cand1 promotes assembly of new SCF complexes through dynamic exchange of F box proteins. *Cell,* 153**,** 206-15.
PINTARD, L., WILLIS, J. H., WILLEMS, A., JOHNSON, J. L., SRAYKO, M., KURZ, T., GLASER, S., MAINS, P. E., TYERS, M., BOWERMAN, B. & PETER, M. 2003. The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. *Nature,* 425**,** 311-6.
PLECHANOVOVÁ, A., JAFFRAY, E. G., MCMAHON, S. A., JOHNSON, K. A., NAVRÁTILOVÁ, I., NAISMITH, J. H. & HAY, R. T. 2011. Mechanism of ubiquitylation by dimeric RING ligase RNF4. *Nat Struct Mol Biol,* 18**,** 1052-9.
PRUNEDA, J. N., LITTLEFIELD, P. J., SOSS, S. E., NORDQUIST, K. A., CHAZIN, W. J., BRZOVIC, P. S. & KLEVIT, R. E. 2012. Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol Cell,* 47**,** 933-42.
RAINA, K. & CREWS, C. M. 2017. Targeted protein knockdown using small molecule degraders. *Curr Opin Chem Biol,* 39**,** 46-53.
REITSMA, J. M., LIU, X., REICHERMEIER, K. M., MORADIAN, A., SWEREDOSKI, M. J., HESS, S. & DESHAIES, R. J. 2017. Composition and Regulation of the Cellular Repertoire of SCF Ubiquitin Ligases. *Cell,* 171**,** 1326-1339.e14.
ROTIN, D. & KUMAR, S. 2009. Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol,* 10**,** 398-409.
RUSNAC, D. V., LIN, H. C., CANZANI, D., TIEN, K. X., HINDS, T. R., TSUE, A. F., BUSH, M. F., YEN, H. S. & ZHENG, N. 2018. Recognition of the Diglycine C-End Degron by CRL2. *Mol Cell,* 72**,** 813-822.e4.
SAHA, A. & DESHAIES, R. J. 2008. Multimodal activation of the ubiquitin ligase SCF by Nedd8 conjugation. *Mol Cell,* 32**,** 21-31.
SAKAMOTO, K. M., KIM, K. B., KUMAGAI, A., MERCURIO, F., CREWS, C. M. & DESHAIES, R. J. 2001. Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc Natl Acad Sci U S A,* 98**,** 8554-9.
SCHMALEN, I., REISCHL, S., WALLACH, T., KLEMZ, R., GRUDZIECKI, A., PRABU, J. R., BENDA, C., KRAMER, A. & WOLF, E. 2014. Interaction of circadian clock proteins CRY1 and PER2 is modulated by zinc binding and disulfide bond formation. *Cell,* 157**,** 1203-15.
SCHULMAN, B. A., CARRANO, A. C., JEFFREY, P. D., BOWEN, Z., KINNUCAN, E. R., FINNIN, M. S., ELLEDGE, S. J., HARPER, J. W., PAGANO, M. & PAVLETICH, N. P. 2000. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature,* 408**,** 381-6.
SCOTT, D. C., MONDA, J. K., BENNETT, E. J., HARPER, J. W. & SCHULMAN, B. A. 2011. N-terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex. *Science,* 334**,** 674-8.
SCOTT, D. C., RHEE, D. Y., DUDA, D. M., KELSALL, I. R., OLSZEWSKI, J. L., PAULO, J. A., DE JONG, A., OVAA, H., ALPI, A. F., HARPER, J. W. & SCHULMAN, B. A. 2016. Two Distinct Types of E3 Ligases Work in Unison to Regulate Substrate Ubiquitylation. *Cell,* 166**,** 1198-1214.e24.
SCOTT, D. C., SVIDERSKIY, V. O., MONDA, J. K., LYDEARD, J. R., CHO, S. E., HARPER, J. W. & SCHULMAN, B. A. 2014. Structure of a RING E3 trapped in action reveals ligation mechanism for the ubiquitin-like protein NEDD8. *Cell,* 157**,** 1671-84.

SHABEK, N. & ZHENG, N. 2014. Plant ubiquitin ligases as signaling hubs. *Nat Struct Mol Biol,* 21**,** 293-6.
SHEARD, L. B., TAN, X., MAO, H., WITHERS, J., BEN-NISSAN, G., HINDS, T. R., KOBAYASHI, Y., HSU, F. F., SHARON, M., BROWSE, J., HE, S. Y., RIZO, J., HOWE, G. A. & ZHENG, N. 2010. Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature,* 468**,** 400-5.
SHIROGANE, T., JIN, J., ANG, X. L. & HARPER, J. W. 2005. SCFbeta-TRCP controls clock-dependent transcription via casein kinase 1-dependent degradation of the mammalian period-1 (Per1) protein. *J Biol Chem,* 280**,** 26863-72.
SIEPKA, S. M., YOO, S. H., PARK, J., SONG, W., KUMAR, V., HU, Y., LEE, C. & TAKAHASHI, J. S. 2007. Circadian mutant Overtime reveals F-box protein FBXL3 regulation of cryptochrome and period gene expression. *Cell,* 129**,** 1011-23.
SKOWYRA, D., CRAIG, K. L., TYERS, M., ELLEDGE, S. J. & HARPER, J. W. 1997. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell,* 91**,** 209-19.
SPRAGUE, E. R., REDD, M. J., JOHNSON, A. D. & WOLBERGER, C. 2000. Structure of the C-terminal domain of Tup1, a corepressor of transcription in yeast. *EMBO J,* 19**,** 3016-27.
SPRATT, D. E., WALDEN, H. & SHAW, G. S. 2014. RBR E3 ubiquitin ligases: new structures, new insights, new questions. *Biochem J,* 458**,** 421-37.
STEBBINS, C. E., KAELIN, W. G. & PAVLETICH, N. P. 1999. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science,* 284**,** 455-61.
STEWART, M. D., RITTERHOFF, T., KLEVIT, R. E. & BRZOVIC, P. S. 2016. E2 enzymes: more than just middle men. *Cell Res,* 26**,** 423-40.
STIEGLITZ, B., RANA, R. R., KOLIOPOULOS, M. G., MORRIS-DAVIES, A. C., SCHAEFFER, V., CHRISTODOULOU, E., HOWELL, S., BROWN, N. R., DIKIC, I. & RITTINGER, K. 2013. Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. *Nature,* 503**,** 422-6.
SUZUKI, T. & YAMAMOTO, M. 2015. Molecular basis of the Keap1-Nrf2 system. *Free Radic Biol Med,* 88**,** 93-100.
TAKAHASHI, J. S. 2017. Transcriptional architecture of the mammalian circadian clock. *Nat Rev Genet,* 18**,** 164-179.
TAN, X., CALDERON-VILLALOBOS, L. I., SHARON, M., ZHENG, C., ROBINSON, C. V., ESTELLE, M. & ZHENG, N. 2007. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature,* 446**,** 640-5.
TASAKI, T., MULDER, L. C., IWAMATSU, A., LEE, M. J., DAVYDOV, I. V., VARSHAVSKY, A., MUESING, M. & KWON, Y. T. 2005. A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol Cell Biol,* 25**,** 7120-36.
TASAKI, T., SRIRAM, S. M., PARK, K. S. & KWON, Y. T. 2012. The N-end rule pathway. *Annu Rev Biochem,* 81**,** 261-89.
TASAKI, T., ZAKRZEWSKA, A., DUDGEON, D. D., JIANG, Y., LAZO, J. S. & KWON, Y. T. 2009. The substrate recognition domains of the N-end rule pathway. *J Biol Chem,* 284**,** 1884-95.

THROWER, J. S., HOFFMAN, L., RECHSTEINER, M. & PICKART, C. M. 2000. Recognition of the polyubiquitin proteolytic signal. *EMBO J,* 19**,** 94-102.
TOMKO, R. J. & HOCHSTRASSER, M. 2013. Molecular architecture and assembly of the eukaryotic proteasome. *Annu Rev Biochem,* 82**,** 415-45.
TOYAMA, B. H. & HETZER, M. W. 2013. Protein homeostasis: live long, won't prosper. *Nat Rev Mol Cell Biol,* 14**,** 55-61.
TREMPE, J. F., SAUVÉ, V., GRENIER, K., SEIRAFI, M., TANG, M. Y., MÉNADE, M., AL-ABDUL-WAHID, S., KRETT, J., WONG, K., KOZLOV, G., NAGAR, B., FON, E. A. & GEHRING, K. 2013. Structure of parkin reveals mechanisms for ubiquitin ligase activation. *Science,* 340**,** 1451-5.
TRON, A. E., ARAI, T., DUDA, D. M., KUWABARA, H., OLSZEWSKI, J. L., FUJIWARA, Y., BAHAMON, B. N., SIGNORETTI, S., SCHULMAN, B. A. & DECAPRIO, J. A. 2012. The glomuvenous malformation protein Glomulin binds Rbx1 and regulates cullin RING ligase-mediated turnover of Fbw7. *Mol Cell,* 46**,** 67-78.
ULJON, S., XU, X., DURZYNSKA, I., STEIN, S., ADELMANT, G., MARTO, J. A., PEAR, W. S. & BLACKLOW, S. C. 2016. Structural Basis for Substrate Selectivity of the E3 Ligase COP1. *Structure,* 24**,** 687-696.
VAN DER VEEN, A. G. & PLOEGH, H. L. 2012. Ubiquitin-like proteins. *Annu Rev Biochem,* 81**,** 323-57.
VARSHAVSKY, A. 2011. The N-end rule pathway and regulation by proteolysis. *Protein Sci,* 20**,** 1298-345.
VIERSTRA, R. D. 2009. The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat Rev Mol Cell Biol,* 10**,** 385-97.
WAUER, T. & KOMANDER, D. 2013. Structure of the human Parkin ligase domain in an autoinhibited state. *EMBO J,* 32**,** 2099-112.
WEI, N., CHAMOVITZ, D. A. & DENG, X. W. 1994. Arabidopsis COP9 is a component of a novel signaling complex mediating light control of development. *Cell,* 78**,** 117-24.
WELCKER, M. & CLURMAN, B. E. 2008. FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. *Nat Rev Cancer,* 8**,** 83-93.
WENZEL, D. M., LISSOUNOV, A., BRZOVIC, P. S. & KLEVIT, R. E. 2011a. UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature,* 474**,** 105-8.
WENZEL, D. M., STOLL, K. E. & KLEVIT, R. E. 2011b. E2s: structurally economical and functionally replete. *Biochem J,* 433**,** 31-42.
WILD, R., GERASIMAITE, R., JUNG, J. Y., TRUFFAULT, V., PAVLOVIC, I., SCHMIDT, A., SAIARDI, A., JESSEN, H. J., POIRIER, Y., HOTHORN, M. & MAYER, A. 2016. Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains. *Science,* 352**,** 986-90.
WILKINSON, K. D. 2005. The discovery of ubiquitin-dependent proteolysis. *Proc Natl Acad Sci U S A,* 102**,** 15280-2.
WU, G., XU, G., SCHULMAN, B. A., JEFFREY, P. D., HARPER, J. W. & PAVLETICH, N. P. 2003. Structure of a beta-TrCP1-Skp1-beta-catenin complex: destruction motif binding and lysine specificity of the SCF(beta-TrCP1) ubiquitin ligase. *Mol Cell,* 11**,** 1445-56.
WU, S., ZHU, W., NHAN, T., TOTH, J. I., PETROSKI, M. D. & WOLF, D. A. 2013. CAND1 controls in vivo dynamics of the cullin 1-RING ubiquitin ligase repertoire. *Nat Commun,* 4**,** 1642.

XIA, Z., WEBSTER, A., DU, F., PIATKOV, K., GHISLAIN, M. & VARSHAVSKY, A. 2008. Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. *J Biol Chem,* 283**,** 24011-28.

XING, W., BUSINO, L., HINDS, T. R., MARIONNI, S. T., SAIFEE, N. H., BUSH, M. F., PAGANO, M. & ZHENG, N. 2013. SCF(FBXL3) ubiquitin ligase targets cryptochromes at their cofactor pocket. *Nature,* 496**,** 64-8.

XU, L., WEI, Y., REBOUL, J., VAGLIO, P., SHIN, T. H., VIDAL, M., ELLEDGE, S. J. & HARPER, J. W. 2003. BTB proteins are substrate-specific adaptors in an SCF-like modular ubiquitin ligase containing CUL-3. *Nature,* 425**,** 316-21.

YAMAMOTO, M., KENSLER, T. W. & MOTOHASHI, H. 2018. The KEAP1-NRF2 System: a Thiol-Based Sensor-Effector Apparatus for Maintaining Redox Homeostasis. *Physiol Rev,* 98**,** 1169-1203.

YAMOAH, K., OASHI, T., SARIKAS, A., GAZDOIU, S., OSMAN, R. & PAN, Z. Q. 2008. Autoinhibitory regulation of SCF-mediated ubiquitination by human cullin 1's C-terminal tail. *Proc Natl Acad Sci U S A,* 105**,** 12230-5.

YU, X., YU, Y., LIU, B., LUO, K., KONG, W., MAO, P. & YU, X. F. 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science,* 302**,** 1056-60.

ZEMLA, A., THOMAS, Y., KEDZIORA, S., KNEBEL, A., WOOD, N. T., RABUT, G. & KURZ, T. 2013. CSN- and CAND1-dependent remodelling of the budding yeast SCF complex. *Nat Commun,* 4**,** 1641.

ZENGERLE, M., CHAN, K. H. & CIULLI, A. 2015. Selective Small Molecule Induced Degradation of the BET Bromodomain Protein BRD4. *ACS Chem Biol,* 10**,** 1770-7.

ZHENG, J., YANG, X., HARRELL, J. M., RYZHIKOV, S., SHIM, E. H., LYKKE-ANDERSEN, K., WEI, N., SUN, H., KOBAYASHI, R. & ZHANG, H. 2002a. CAND1 binds to unneddylated CUL1 and regulates the formation of SCF ubiquitin E3 ligase complex. *Mol Cell,* 10**,** 1519-26.

ZHENG, N., SCHULMAN, B. A., SONG, L., MILLER, J. J., JEFFREY, P. D., WANG, P., CHU, C., KOEPP, D. M., ELLEDGE, S. J., PAGANO, M., CONAWAY, R. C., CONAWAY, J. W., HARPER, J. W. & PAVLETICH, N. P. 2002b. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature,* 416**,** 703-9.

ZHENG, N. & SHABEK, N. 2017. Ubiquitin Ligases: Structure, Function, and Regulation. *Annu Rev Biochem,* 86**,** 129-157.

ZHUANG, M., CALABRESE, M. F., LIU, J., WADDELL, M. B., NOURSE, A., HAMMEL, M., MILLER, D. J., WALDEN, H., DUDA, D. M., SEYEDIN, S. N., HOGGARD, T., HARPER, J. W., WHITE, K. P. & SCHULMAN, B. A. 2009. Structures of SPOP-substrate complexes: insights into molecular architectures of BTB-Cul3 ubiquitin ligases. *Mol Cell,* 36**,** 39-50.

ZIMMERMAN, E. S., SCHULMAN, B. A. & ZHENG, N. 2010. Structural assembly of cullin-RING ubiquitin ligase complexes. *Curr Opin Struct Biol,* 20**,** 714-21.