UNIVERSITY OF CALIFORNIA SAN DIEGO

Regulation of Ubiquitylation and Ubiquitin-like Post-Translational Modifications

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in

Chemistry

by

Ryan James Lumpkin

Committee in charge:

Professor Elizabeth A. Komives, Chair
Professor Rommie Amaro
Professor Gouri Ghosh
Professor Randy Hampton
Professor Susan Taylor

2019

The dissertation of Ryan James Lumpkin is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

To my parents who have helped me get this far, my friends who supported me along the way, my PI who believed in me, and my cat who was always waiting for me.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Å	Angstrom
ARIH2	E3 ubiquitin-protein ligase ARIH2
ASB9	Ankyrin-repeat and SOCS Box protein 9
CKB	Creatine Kinase B
CUL5	Cullin 5
Da	Dalton
EloB/C	Elongin B and Elongin C
EDTA	Ethylenediaminetetraacetic Acid
FPLC	Fast Protein Liquid Chromatography
HDX	Hydrogen-Deuterium Exchange
MS	Mass Spectrometry
NAE1	NEDD8-activating enzyme E1 regulatory subunit
NEDD8	Neural precursor cell expressed developmentally down-regulated protein 8
PDB	Protein Data Bank
RBX2	Ring Box Protein 2
SUMO	Small Ubiquitin-like Modifier
UBA3	NEDD8-activating enzyme E1 catalytic subunit
UBE1	Ubiquitin Activating Enzyme 1
UBE2D2	Ubiquitin Conjugating Enzyme D2
UBE2F	Ubiquitin Conjugating Enzyme F
UBE2L3	Ubiquitin Conjugating Enzyme L3
UBE2R1	Ubiquitin Conjugating Enzyme R1

UB	Ubiquitin
wAlp	Wild-type Alpha-lytic Protease

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapter II, in full, is a reprint that the dissertation author was the principal researcher and author of. The material appears in *Nature Communications*. (**Lumpkin, R. J.,** Gu, H., Zhu, Y., Leonard, M., Ahmad, A. S., Clauser, K. R., Meyer, J. G., Bennett, E. J., and Komives E. A. (2017). Site-specific identification and quantitation of endogenous SUMO modifications under native conditions. *Nat. Commun.*, **8**, 1171.)

Chapter III, in full, is a reprint that the dissertation author was the principal researcher and author of. The material appears in *Molecular and Cellular Proteomics*. (**Lumpkin, R. J.,** and Komives, E. A. (2019). DECA, a comprehensive, automatic post-processing program for HDX-MS data. *Mol Cell Proteomics, in press)*

Chapter IV, is material that the dissertation author was the principal researcher and author of. The material has been submitted for publication. (**Lumpkin, R. J.,** Baker, R. W., Leschziner, A., and Komives, E. A. Structure and dynamics of the ASB9 CUL-RING E3 Ligase)

Chapter V, is material that the dissertation author was the principal researcher and author of. The material has been submitted for publication. (**Lumpkin, R. J.,** Ahmad, A. S., Blake, R., Condon, C., and Komives, E. A. Triggering CUL5-Mediated Ubiquitinylation.)

VITA

2014 Bachelor of Science, Chemistry, University of Denver, Denver, CO

2016 Master of Science, Chemistry, University of California San Diego

2019 Doctorate of Philosophy, Chemistry, University of California San Diego

PUBLICATIONS

Lumpkin, R. J., Ahmad, A. S., Blake, R., Condon, C., and Komives, E. A. Triggering CUL5-Mediated Ubiquitinylation. *In press*.

Lumpkin, R. J., Baker, R. W., Leschziner, A., and Komives, E. A. Structure and dynamics of the ASB9 CUL-RING E3 Ligase. *In press*.

Lumpkin, R. J., and Komives, E. A. (2019). DECA, a comprehensive, automatic post-processing program for HDX-MS data. *Mol Cell Proteomics, in press*.

Lumpkin, R. J., Gu, H., Zhu, Y., Leonard, M., Ahmad, A. S., Clauser, K. R., Meyer, J. G., Bennett, E. J., and Komives E. A. (2017). Site-specific identification and quantitation of endogenous SUMO modifications under native conditions. *Nat. Commun.*, **8**, 1171.

Capodagli, G. C., Deaton, M. K., Baker, E.A., **Lumpkin R. J.**, and Pegan S. D. Diversity of ubiquitin and ISG15 specificity amongst nairoviruses' viral ovarian tumor domain proteases. (2013) *J. Virol*. 87: 3815-27.

FIELDS OF STUDY

Major Field: Biochemistry

Studies in Biochemistry and Biophysics

Professor Elizabeth A. Komives

HONORS AND AWARDS

2018 Distinguished Graduate Student Fellowship (UC San Diego)

2015-2017 Molecular Biophysics Training Grant, trainee

2012-2013 Lynn H. Miller Departmental Scholarship (University of Denver)

ABSTRACT OF THE DISSERTATION

Regulation of Ubiquitylation and Ubiquitin-like Post-Translational Modifications

by

Ryan James Lumpkin

Doctor of Philosophy in Chemistry

University of California San Diego, 2019

Professor Elizabeth A. Komives, Chair

Ubiquitin (Ub) is an 8,500 Da, 76 amino acid protein that is among a family of similar Ubiquitin-like proteins. E3 Ligases facilitate the attachment of Ub onto lysine residues of substrate proteins. Ub can also be attached to another Ub molecule, forming polyubiquitin chains which signal for proteasomal degradation and the immune response. Small ubiquitin-like modifier (SUMO) is involved in cell cycle control and the DNA damage response. Another Ubiquitin-like protein, Neural precursor cell expressed developmentally down-regulated protein 8 (NEDD8), can modify some E3 Ligases to regulate their activation.

Chapter II presents a novel method for the side-by-side identification and quantification of

Ubiquitin and SUMO modifications. Trypsin digestion of a Ubiquitin-modified protein leaves a diglycyl-lysine at the site of modification, which can be detected by mass spectrometry. Unlike Ubiquitin, detection of SUMOylated proteins is limited by the lack of naturally occurring protease sites in the C-terminal tail of SUMO proteins. This chapter describes detection of endogenous SUMOylation using α-lytic protease, WaLP.

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) probes the solvent accessibility and conformational dynamics of proteins by measuring deuterium uptake on backbone amides over time. Chapter III describes an open-source software package designed to automate the correction, analysis and visualization of HDX-MS data.

Chapter IV investigates the structure, dynamics and regulation of an ASB9-CUL-RING E3 Ubiquitin Ligase. The structure of the substrate-receptor complex was determined for the first time by Cryo-EM, identifying changes in both the receptor and the substrate that were not predicted by previous models.

Chapter V further investigates the dynamics, regulation and activity of the ASB9-CUL-RING E3 Ubiquitin Ligase. *In vitro* assays, western blots and mass spectrometry data describe the locations and intensity of ubiquitin modifications. These data show that ARIH2, another E3 ligase, is necessary for initial substrate modification, and that the addition of other E2 enzymes can modulate the poly-ubiquitin chains formed. HDX-MS studies show that modification of the ASB9-CUL-RING ligase causes conformational restructuring that facilitates the binding and activation of ARIH2.

Chapter I:

Introduction

A. Proteome Complexity and Post-Translational Modifications

The human proteome contains approximately 20,000 distinct proteins, and there are an additional 22,000 isoforms of those proteins. This is hardly the limit of the complexity of the human proteome however, as proteins can be modified by various molecules and whole proteins. A particular modified form of a protein is described as a proteome, and estimates have placed the totally number of human proteoforms at approximately 6 million (1).

Post-translational modifications (PTMs) can drastically alter protein behavior and interactions. For example, histone proteins, which form DNA-bound nucleosomes, are modified with acetylation, methylation, and phosphorylation, among other proteins. This has been extensively studied, as these modifications perturb chromatin structure and interaction with effector molecules, thereby regulating transcription and other DNA processes (2).

PTMs can be split into two major groups based on size: chemical modifications and whole protein modifications. Phosphorylation, acetylation, and methylation are among the most common PTMs observed and they are all small molecules. The protein modification PTM family consists of modifications by Ubiquitin and Ubiquitin like proteins.

B. Ubiquitylation

The attachment of the 8.5kDa ubiquitin protein to a substrate, or ubiquitylation, can have a number of different functions, including cell cycle regulation, innate immunity, the DNA damage response, transcriptional regulation, and protein degradation. The hallmark of ubiquitylation is the formation of polyubiquitin chains. Ubiquitin can be linked to the N-terminal methionine or any of the 7 lysine residues on another molecule of ubiquitin, and the different chain types yield different cellular signals.

Ubiquitin was first discovered to have a role in protein degradation in 1982 (3). This process was found to be dependent upon three additional enzymes. The E1 activating enzyme is necessary for the initial conjugation of Ubiquitin to an E2 conjugating enzyme, which then incorporates into an E3 ligase. The E3 ligase is responsible for the catalyzing the final ubiquitin transfer to the substrate protein (4). In 1989, Chau et al. published research revealing that short-lived proteins with ubiquitylated with ubiquitin chains that were linked together at lysine 48 (5). K48 Polyubiquitylation is the most extensively studied form of ubiquitylation and is responsible for facilitating proteasomal degradation of proteins.

C. Ubiquitin-like proteins

Interferon-Stimulated Gene 15 (ISG15) is an 18 kDa Ubiquitin-like protein that is involved in the innate immune response. Over twice the size of Ubiquitin, it is a structural mimic of K63 diUbiquitin, a poly-ubiquitin chain type that has been implicated in cellular signaling and the immune response (6).

Small Ubiquitin-like Modifier (SUMO) is an 11 kDa Ubiquitin-like protein that is responsible for transcriptional regulation and the DNA Damage Response (7).

Neural-precursor cell-expressed developmentally-downregulated protein 8 (NEDD8) is a 9 kDa Ubiquitin like protein. It plays a role in the activation of one type of E3 Ubiquitin ligases.

D. E3 Ubiquitin Ligases

By 2000, two major families of E3 ubiquitin ligases had been discovered: the Homologous to E6-AP Carboxy-terminus (HECT) family and the Really Interesting New Gene (RING) family (Jackson 2000). These two families are best classified by the mechanism of ubiquitin transfer in each. HECT E3 ligases accept an activated ubiquitin from an E2 conjugating enzyme at a catalytic cysteine residue, from which the ligase transfers it to a bound substrate protein. RING E3 ligases

hold a Ubiquitin-conjugated E2 enzyme and a substrate and facilitate the Ubiquitin transfer from the E2 to the substrate protein. Recent estimates suggest that there are approximately 30 HECT E3 Ligases and 600 RING E3 Ligases (8).

A notable subfamily of RING Ligases is the Cullin (CUL)-RING family of E3 Ligases. These ligases are multi-subunit E3 ligases with the ability to incorporate a variety of substrate receptors. This unique functionality allows the ligases to reuse several of the components in the complex while maintaining specificity toward a wide set of proteins (9).

The CUL-RING E3 Ubiquitin Ligase is composed of one of seven CUL proteins (CUL1, CUL2, CUL3, CUL4a, CUL4b, CUL5, and CUL7) and one of two RING proteins (RBX1, RBX2). Adapter proteins interact with the N-terminal domain of the CUL proteins and mediate its interaction with a substrate receptor. Ubiquitin-bound E2 enzymes associate with the RING component of the ligase at the C-terminal domain of the CUL, and the ligase facilitates transfer of the ubiquitin to a substrate on the substrate receptor (9).

The CUL1-RBX1-Skp1-Fbox (SCF) complex is one of the first identified CUL-RING E3 ligases, and the structure of the complex was determined in 2002. Structurally, the CUL1 is distinct in having a long N-terminal stalk and a globular C-terminal domain. The N-terminal domain consists of three Cullin repeats, each composed of a bundle of helices. RBX1 interacts with the CUL1 at its C-terminal domain, forming a β-sheet with CUL1. CUL1 additionally acts as a scaffold, and it positions the substrate receptor and E2 binding components far apart (10). The N-terminal half of CUL5 was determined in complex with SOCS2-ELOB-ELOC in 2013 (11). The structure of pVHL-ELOB-ELOC-CUL2-RBX1 was determined in 2017 (12). These structures share the same general architecture describe for CUL1, suggesting that all CUL-RING ligases adopt these structural characteristics.

Neddylation of CUL proteins has been shown to facilitate the activation of CUL-RING ligases through the conformational rearrangement of the C-terminal domain of the CUL protein and by regulating the interactions of other proteins with the CUL. The COP9 Signalosome regulates this process by de-neddylating CUL proteins, which additionally decreases their affinity for E2 ubiquitin-conjugating enzymes. Furthermore, the CAND1 protein competes with substrate receptor to bind and sequester un-neddylated CUL proteins (9). These mechanisms keep equilibrium in the CUL-RING ubiquitylation pathway and prevent runaway ubiquitylation while maintaining a steady level of E3 Ligase components. Structures of the C-terminal domain of CUL5 before and after Neddylation were determined in 2008, demonstrating that the modification of CUL5 causes significant conformation rearrangement of the domain as well as the bound RING protein (13).

HIV and HPV have both been shown to hijack native ubiquitylation pathways through CUL-RING ligases, creating interest in the ligases as targets for drug discovery. Similarly, CUL-RING ligases are significant the design of proteolysis-targeting chimeric molecules (PROTACs) to facilitate targeted protein degradation (14).

E. The ASB Family of Protein Receptors

The Ankyrin Repeat and Suppressor of Cytokine Signaling (SOCS)-box (ASB) protein family consists of 18 proteins (ASB1-18) that act as substrate receptors for the CUL5-RBX2 ligase. ASB substrate receptors integrate into the ligase at the N-terminal domain of CUL5 through the adaptor proteins Elongin B and Elongin C. Each ASB has an ankyrin-repeat domain that is responsible for protein-protein binding and a SOCS-box domain that facilitates its interaction with Elongin B, Elongin C, and CUL5. The ASB9-CUL-RING ligase has been shown to ubiquitylate and promote degradation of Creatine Kinase (15).

F. References

1. R. Aebersold, J. N. Agar, I. J. Amster, M. S. Baker, C. R. Bertozzi, E. S. Boja, C. E. Costello, B. F. Cravatt, C. Fenselau, B. A. Garcia, Y. Ge, J. Gunawardena, R. C. Hendrickson, P. J. Hergenrother, C. G. Huber, A. R. Ivanov, O. N. Jensen, M. C. Jewett, N. L. Kelleher, L. L. Kiessling, N. J. Krogan, M. R. Larsen, J. A. Loo, R. R. Ogorzalek Loo, E. Lundberg, M. J. MacCoss, P. Mallick, V. K. Mootha, M. Mrksich, T. W. Muir, S. M. Patrie, J. J. Pesavento, S. J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schluter, S. Sechi, S. A. Slavoff, L. M. Smith, M. P. Snyder, P. M. Thomas, M. Uhlen, J. E. Van Eyk, M. Vidal, D. R. Walt, F. M. White, E. R. Williams, T. Wohlschlager, V. H. Wysocki, N. A. Yates, N. L. Young, B. Zhang, How many human proteoforms are there? *Nat Chem Biol* **14**, 206-214 (2018).

2. A. J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395 (2011).

3. A. Hershko, A. Ciechanover, Mechanisms of intracellular protein breakdown. *Annu Rev Biochem* **51**, 335-364 (1982).

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

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

6. L. A. Durfee, N. Lyon, K. Seo, J. M. Huibregtse, The ISG15 conjugation system broadly targets newly synthesized proteins: implications for the antiviral function of ISG15. *Mol Cell* **38**, 722-732 (2010).

7. S. Muller, C. Hoege, G. Pyrowolakis, S. Jentsch, SUMO, ubiquitin's mysterious cousin. *Nat Rev Mol Cell Biol* **2**, 202-210 (2001).

8. N. Zheng, N. Shabek, Ubiquitin Ligases: Structure, Function, and Regulation. *Annu Rev Biochem* **86**, 129-157 (2017).

9. M. D. Petroski, R. J. Deshaies, Function and regulation of cullin-RING ubiquitin ligases. *Nat Rev Mol Cell Biol* **6**, 9-20 (2005).

10. N. Zheng, B. A. Schulman, L. Song, J. J. Miller, P. D. Jeffrey, P. Wang, C. Chu, D. M. Koepp, S. J. Elledge, M. Pagano, R. C. Conaway, J. W. Conaway, J. W. Harper, N. P. Pavletich, Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**, 703-709 (2002).

11. Y. K. Kim, M. J. Kwak, B. Ku, H. Y. Suh, K. Joo, J. Lee, J. U. Jung, B. H. Oh, Structural basis of intersubunit recognition in elongin BC-cullin 5-SOCS box ubiquitin-protein ligase complexes. *Acta Crystallogr D Biol Crystallogr* **69**, 1587-1597 (2013).

12. T. A. F. Cardote, A. Ciulli, Structure-Guided Design of Peptides as Tools to Probe the Protein-Protein Interaction between Cullin-2 and Elongin BC Substrate Adaptor in Cullin RING E3 Ubiquitin Ligases. *ChemMedChem* **12**, 1491-1496 (2017).

13. D. M. Duda, L. A. Borg, D. C. Scott, H. W. Hunt, M. Hammel, B. A. Schulman, Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell* **134**, 995-1006 (2008).

14. E. Bulatov, A. Ciulli, Targeting Cullin-RING E3 ubiquitin ligases for drug discovery: structure, assembly and small-molecule modulation. *Biochem J* **467**, 365-386 (2015).

15. F. Okumura, A. Joo-Okumura, K. Nakatsukasa, T. Kamura, The role of cullin 5-containing ubiquitin ligases. *Cell Div* **11**, 1 (2016).

Chapter II:

Site-specific identification and quantitation of endogenous SUMO modifications under native conditions

A. Introduction

The family of small ubiquitin-like modifier (SUMO) proteins in humans includes four distinct genes with three types of members: SUMO1, SUMO2/3 (which differ by only three residues), and SUMO4. SUMO proteins regulate the function of various proteins by reversible covalent isopeptide bond attachment between the C terminus of SUMO and a free ε-amine group typically on lysine residues within target proteins (1), similar to ubiquitin (Ub). Ub conjugation mainly targets proteins for degradation by the proteasome, but has also been implicated in DNA repair, receptor signaling, and cell communication (2). The function of SUMO conjugation on target proteins is similarly diverse with SUMOylation catalyzing alteration of protein activity for targets involved in gene expression, DNA repair, nuclear import, heat shock, cell motility, and lipid metabolism (1,3). SUMO targets are generally low-abundance proteins, and the amount of the modification at steady state is also low (4). Given its importance in numerous cellular functions, several groups have developed proteomic methods for analysis of SUMOylated proteins.

Three general approaches have been previously utilized to isolate and identify the SUMOylated proteome (5). Ectopic expression of epitope-tagged SUMO followed by standard isolation techniques and mass spectrometry has been widely used in a variety of organisms (6-10). These approaches have yielded various maps of SUMO-interacting proteins but few sites of SUMOylation are identified by this approach and the extent to which exogenous expression of modified SUMOs alters substrate targeting is unknown. Immuno-affinity approaches utilizing antibodies that recognize endogenous SUMO2/3 or SUMO1 have been used to identify SUMO-interacting proteins under endogenous conditions (4). However, as with the epitope-tagging approach, few actual sites of SUMO modification were identified using this method. Therefore,

methods that allow proteome-level identification of endogenous SUMOylation sites are needed.

A robust proteomic method has been developed to measure thousands of endogenous ubiquitylation sites (11). The method takes advantage of the C-terminal sequence of Ub (RGG) (Fig. 2.1a). When cleaved with trypsin, ubiquitylated substrate proteins will generate peptides containing a Ub-remnant diglycyl-lysine (KGG) that can be enriched using specific antibodies and identified by tandem mass spectrometry (Fig. 2.1b) (12-14). Instead of the trypsin-friendly arginine residue preceding the C-terminal diGlycine sequence observed in the processed Ub sequence, mature human SUMO paralogs have a threonine preceding the C-terminal diGlycine sequence and no other tryptic cleavage sites near the C terminus (Fig. 2.1a). To overcome this problem, various schemes that introduce mutations within the C terminus of SUMO to render it more amenable to trypsin-based cleavage and identification by mass spectrometry have been developed for global profiling of SUMO attachment sites (15,16). Several groups have reported global profiling approaches in which mutant SUMOs with various affinity tags and protease recognition sites were introduced into cells. For example, Hendriks et al. introduced a lysine-deficient SUMO-3 with a C-terminal trypsin cleavage site, His_{10}–SUMO-3 K0 Q87R (17,18). The SUMOylated proteins were enriched by immobilized metal affinity chromatography (IMAC) and digested with Lys-C. Peptides modified with SUMO were then purified again with IMAC and finally digested with trypsin to generate a five amino acid C-terminal SUMO-remnant modification. This group has compiled all available data resulting in 7327 SUMOylation sites in 3617 proteins (17). Other groups have used similar engineering approaches to express mutant SUMOs and identify up to 1000 unique SUMOylation sites upon induction of cell stress (15,16,18-20). While these methods have proved effective in mapping SUMOylation sites, they all require exogenous expression of mutant versions of SUMO, which preclude analysis of SUMO-

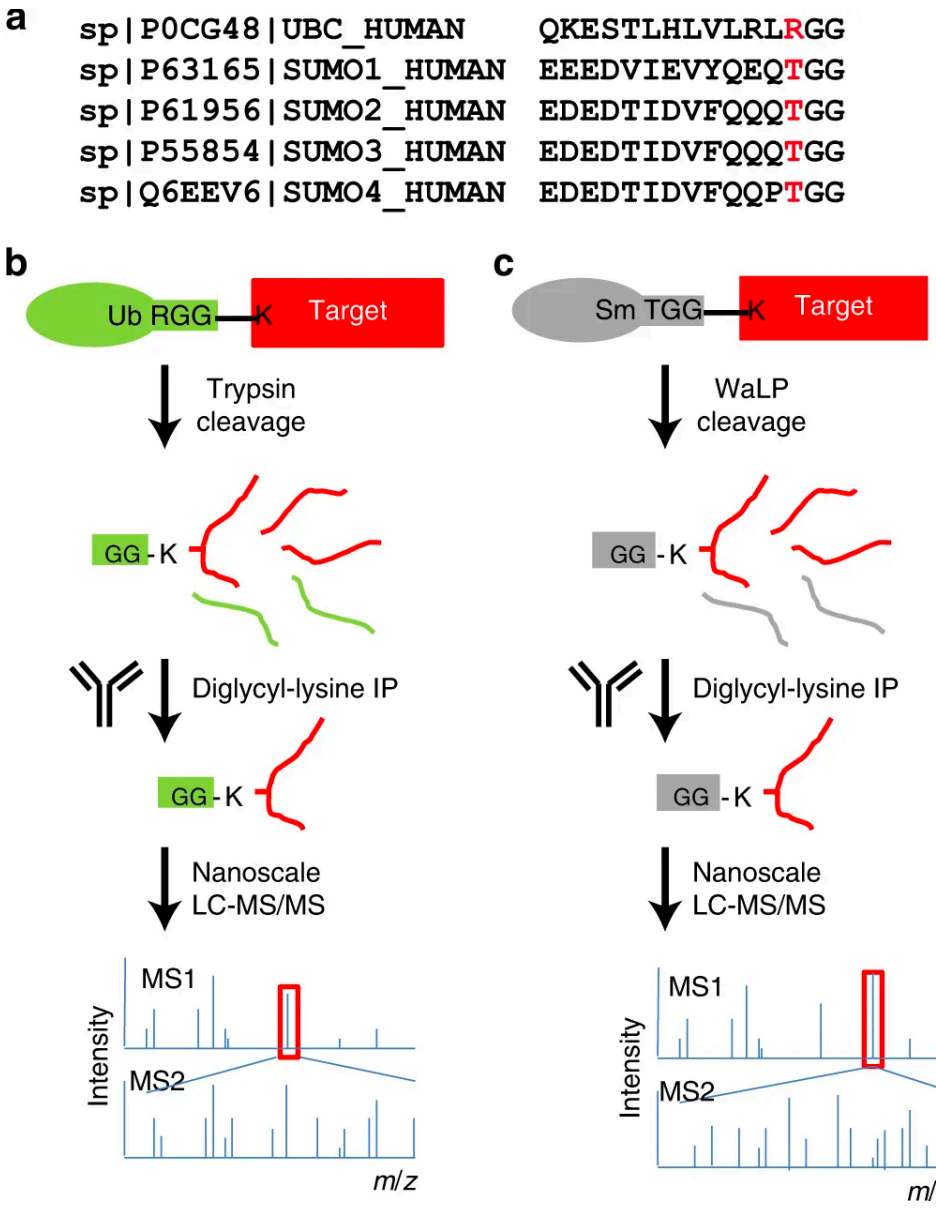

Figure 2.1: A strategy for mapping endogenous SUMO-modification sites.
a. C-terminal sequence alignment of processed human SUMO1-4 and ubiquitin. The proximal amino acid to the diGlycine C-terminal residues is indicated in red. b. Schematic depicting the Ub site mapping strategy. Proteins modified by ubiquitin are digested with trypsin leaving a diGlycine attached to the ε-amine of the lysine where ubiquitin was attached. An antibody specific for KGG-peptides is used to enrich peptides from ubiquitylated sites that are then identified by mass spectrometry. c. The same as b but WaLP digestion is used to generate the KGG-peptides from SUMO attachment sites.

modification sites in native settings or from human tissues. Currently, no method exists for identifying endogenous SUMO sites on a global proteome scale without introduction of mutant SUMO.

We recently described the application of wild-type α-lytic protease (WaLP) to proteome digestion for shotgun proteomics (21). Although relatively relaxed specificity was observed, WaLP prefers to cleave after threonine residues and rarely cleaved after arginine (21). In addition, WaLP generates peptides of the same average length as trypsin despite its more relaxed substrate specificity (21). We show here that WaLP cleaves at the C-terminal TGG sequence (all SUMO paralogs) leaving a SUMO-remnant KGG at the position of SUMO attachment in target proteins. The resulting KGG-containing peptides can then be identified using methods already developed for profiling the Ub-modified proteome as described above (Fig. 2.1c). The method allows identification of SUMO attachment sites under completely native conditions using the Ub-profiling workflow by simply substituting WaLP for trypsin. The same sample can be subjected to analysis of both the Ub- and SUMO-modified proteomes simply by digesting the sample with either trypsin or WaLP, respectively. We demonstrate the effectiveness of this parallel identification of ubiquitylation and SUMOylation sites in cells treated with proteasome inhibitors. We provide the description a unique method of identifying proteins containing individual lysine residues that are modified by both SUMO and Ub from the same sample. This method can be simply applied to any sample, including human tissues samples, to identify endogenous SUMOylation sites.

B. Materials and Methods

1. WaLP

WaLp was expressed from Lysobacter enzymogenesis type 495 (ATCC) using

Bachovichin's media supplemented with MEM vitamins and 60 g l^{-1}sucrose. *L. enz*. Was grown at 30 °C with shaking at 100 rpm for 3 days. WaLP was purified from the culture supernatant as described previously (33). Briefly, the protease was captured from the supernatant by batch binding on SP-sepharose, which is washed extensively and then eluted with high pH glycine buffer. After buffer exchange to pH 7.2, the enzyme was loaded onto an FPLC monoS 10/10 column and eluted with a gradient of 10 mM $NaHPO_4$, pH 7.2 to the same buffer containing 250 mM sodium acetate over 1 h.

2. SENP1 and SENP2 deSUMOylation

To verify the in vitro deSUMOylation and deubiquitylation assay, untreated HCT116 cells were harvested and lysed with denaturing lysis buffer (8 M urea, 150 mM NaCl, 50 mM Tris pH 7.8, 1 mM sodium orthovanadate, 1 mM NaF, 1 mM sodium 2-glycerophosphate, protease inhibitor tablet (Roche), 5 mM N-Ethylmaleimide (NEM, made fresh in methanol). Lysates were sonicated and centrifuged at 20,000×*g* for 10 min at 4 °C to remove insoluble material. Lysates were then diluted to 1 M urea using cold 50 mM Tris pH 7.8. Dithiothreitol (DTT) was then added to the lysates at a final concentration of 15 mM. Lysates were then untreated or treated with a mixture of SENP1/2 (Life Sensors) or Usp2cc (Enzo Life Sciences) for 4 h at 25 °C. For SENP1/2 treatment, we initially used 2 U mg^{-1}, 4 U mg^{-1}, and 6 U mg^{-1} concentrations for the validation experiments and 1 μg mg^{-1} (total protein) for Usp2cc. Reactions were quenched by addition of SDS sample buffer and samples were then processed for SDS-PAGE and immunoblotting. The antibodies used for immunoblotting were against ubiquitin (MAB1510, EMD Millipore, 1:1000 dilution), SUMO1 (4930, Cell Signaling Technologies, 1:1000 dilution), and SUMO2/3 (4971, Cell Signaling Technologies, 1:1000 dilution).

Hela cells were cultured in MEM media with 10% fetal bovine serum and 1%

penicillin/streptomycin at 37 °C with 5% CO_2. Heat shock was performed at 43 °C for 1 h, then Hela cells were washed with cold PBS, and harvested with 8 M urea lysis buffer (50 mM Tris pH 8.0, 8 M urea, 1 mM vanadate, 2.5 mM sodium pyrophosphate, 1 mM beta-glycerol-phosphate). Extracts were sonicated and centrifuged at 20,000×*g* for 15 min to remove insoluble material. Protein concentrations were measured by Bradford Assay.

The cell lysate was diluted fourfold by addition of 50 mM Tris buffer pH 8.0 to a final urea concentration of 2 M, and a DTT stock solution was added to a final concentration of 4.5 mM. Equal amounts of lysates were reduced with 4.5 mM DTT at 56 °C for 30 min and alkylated by adding iodoacetamide to 9 mM for 15 min in the dark. Additional DTT was added to lysate to concentration of 4.5 mM. In vitro deSUMOylation reaction was performed by adding specific SUMO protease 1 (SENP1) and SUMO protease 2 (SENP2) (Life Sensors, Malvern, PA) at 10 units per milligram of lysate protein and incubating with lysate for overnight at room temperature. In vitro de-ubiquitination reaction was performed by adding USP2cc (Enzo Life Sciences) at an enzyme-to-substrate ratio of 1:500 and incubating at RT for 4 h. Untreated control lysate samples were also prepared. Efficiency of SUMOylation cleavage by SUMO proteases and ubiquitin cleavage by USP2cc was verified by western blots using primary antibodies for SUMO1 (4930), SUMO2/3 (4971), and ubiquitin (3933) from Cell Signaling Technology. The lysates were digested by WaLP or trypsin, an enzyme-to-substrate ratio of 1:100 at 37 °C overnight with slow rotation. Digestion was stopped by adding 20% TFA solution to final TFA concentration of 1%. Peptides were then subject to C18 cleaning by Sep-Pak cartridges (Waters Corp.) and lyophilized.

3. Enrichment of KGG-containing peptides

For each 5–7 mg peptide sample, 10 µl of UbiScan beads (20 µl of slurry, Cell Signaling Technology) was used to immunopurify the KGG-containing peptides according to the CST

protocols. First, the digests were resuspended in 0.35 ml 2× IAP buffer (100 mM MOPS, 20 mM Na_2HPO_4, 100 mM NaCl, pH 7.5), the pH was adjusted to 7.5, and cleared by centrifugation at 10,000×*g* for 10 min at 4 °C. Then the digests were pre-cleared by rocking with protein A resin for 1 h at 4 °C. Next, the KGG-peptides were immunopurified from the digests by incubation with the UbiScan beads for 2 h with rocking at 4 °C. After isolating the beads by centrifugation at 1000×*g* for 1 min, the beads were washed two times with 1× IAP buffer, then four times with HPLC-grade water. The peptides were eluted in two steps. The beads were incubated at room temperature with 55 µl of 0.15% TFA for 10 min, centrifuged at 3500×*g* for 1 min, and the supernatant carefully saved. Then the beads were incubated for an additional 10 min with followed by 45 µl of 0.15% TFA for 10 min, centrifuged at 3500×*g* for 1 min, and the supernatant was combined with the first elution. The samples were analyzed by nLC-MS/MS on an orbitrap Fusion or LUMOS for WaLP-digested samples, and on a Q-Exactive for Ub-digested samples.

4. MS analysis of in vitro deSUMOylation experiment

Immunoprecipitated peptides were resuspended in 0.125% formic acid and analyzed by an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (Thermo) coupled to an EASY-nLC 1200 (Thermo). Each sample was split and analytical replicate injections were run to increase the number of identifications and provide metrics for analytical reproducibility of the method. Standard peptide mix (MassPREP™ Protein Digestion Standard Mix 1, Waters) was spiked in each sample vial in a total quantity of 100 fmol (33 fmol per injection) prior to LC-MS/MS analysis. The sample was loaded onto an EASY-Spray™ analytical column (PepMap™, 75 µm × 50 cm, C18, 2 µm, 100 Å, Thermo), which was connected to an EASY-Spray™ ionization source (Thermo). The column was heated to 45 °C for all runs. Mobile phase solvent A was composed of 0.1% formic acid and water. Mobile phase solvent B was composed

of 0.1% formic acid, 93.5% acetonitrile, and water. Peptides were separated using a gradient from 5% B to 32% B over 90 min and continued to 53% B over 5 min at a constant flow rate of 300 nl min^{-1}. Full MS scans were obtained with a range of *m/z* 300–1500 at a mass resolution of 120,000 (*m/z* 200), with an AGC target value of 4.0E5 and maximum injection time of 50 ms. To select peptides for MSMS analysis, ions with charge states from 2 to 7 were all included, dynamic exclusion was set to 60 s with mass tolerance 10 ppm, and intensity threshold was set at 2.0E3. Data-dependent mode was established at top speed of 3 s. Most intense precursor ions were selected and isolated with a window of 2 *m/z* and fragmented by collision-induced dissociation with a normalized collision energy of 35 and activation *Q* of 0.25. MS/MS spectra were acquired in the ion trap at enhanced scan rate with an AGC target value of 3.0E3 and maximum injection time at 350 ms. Real-time recalibration of mass error was performed using lock mass (34) with a singly charged polysiloxane ion *m/z* = 371.101237.

MS/MS spectra were evaluated using SEQUEST and the Core platform from Harvard University (35-37). Files were searched against the Swissprot homo sapiens FASTA database updated on September 2015. A mass accuracy of ±5 ppm was used for precursor ions and 0.02 Da for product ions. Enzyme specificity was limited to trypsin, with at least one tryptic (K- or R-containing) terminus required per peptide and up to four mis-cleavages allowed. No enzyme specificity was restricted for WaLP-digested samples. Cysteine carboxamidomethylation was specified as a static modification, oxidation of methionine residues was allowed, and digly remnant on lysine residue (+114.0429) was allowed for each enrichment sample set. Reverse decoy databases were included for all searches to estimate false discovery rates, and filtered using a 1% FDR in the linear discriminant module of core. Results were further narrowed by mass accuracy based on clustering of forward and reverse assignments in Xcorr vs. mass error plots. All

quantitative results were generated using Progenesis V4.1 (Waters Cooperation) and Skyline Version 3.1 to extract the integrated peak area of the corresponding peptide assignments.

Quantitative comparison of the KGG-containing peptides before and after SENP1/2 treatment was performed using Progenesis V4.1 (Waters Cooperation) and Skyline Version 3.1 to extract the integrated peak area of the corresponding peptide assignments according to previously published protocols (19, 20). Extracted ion chromatograms for peptide ions that changed in abundance between samples were manually reviewed to ensure accurate quantitation in Skyline. Statistical analysis of the quantitative data was done using two-tailed *t*-test between SUMO proteases treated and untreated groups. The maximum negative log-*p* value from three comparison pairs was used to indicate significance for abundance changes of a certain peptide between two groups. Bar graphs of the quantitative data were generated and clustered in Spotfire Decision Site (TIBCO Software AB) version 9.1.2.

5. SILAC quantitation of Ub and SUMO

HCT116 cells were grown on DMEM SILAC media (Thermo Scientific/Pierce) supplemented with 10% FBS, penicillin, streptomycin, arginine (85 mg l^{-1}), and either $^{13}C_6$ $^{15}N_2$ lysine (Cambridge Isotope Labs) or unlabeled lysine at 50 mg l^{-1}. Cells were expanded up to twenty 15 cm plates for heavy medium and twenty 15 cm plates in light medium to yielded 40 mg of total protein. The cells were treated with 10 μM MG132 (Sigma) dissolved in DMSO for 4 h or DMSO-only as the negative control. Cells were then washed with PBS, scraped into ice-cold PBS, and counted with a TC20 cell counter (Bio-Rad). Equal quantities of unlabeled and labeled cells from each condition were combined and stored at −80 until lysis.

Frozen cell pellets were thawed quickly and resuspended in 4 mls of denaturing lysis buffer containing 50 mM Tris, pH 8.2, 8 M urea, 75 mM NaCl, 1 mM Na_3VO_4, 1 mM β-

glycerophosphate, 1 mM NaF, 2 mM NEM, 1 mM PMSF, and Roche complete mini protease inhibitor. Cells were then sonicated on ice using 15W power output for three cycles of 30 s with 30 s rests in between. Insoluble material was precipitated by centrifugation at 20,000×*g* for 15 min at 4 °C and protein in the supernatant was quantified by BCA assay. Typically, 5–10 mg protein was digested for Ub identification and 10–15 mg protein was digested for SUMO identification. For Ub identification, lys-C was added to a final concentration of 10 ng μl^{-1} and incubated for 2 h at 37 °C. Then the digest was diluted to 2 M urea by addition of 50 mM Tris-HCl, pH 8.2 and trypsin (Sigma) was added to a final ratio of 1:100 and incubated at 37 °C overnight. For SUMO identification, the lysate was diluted to 2 M urea by addition of 50 mM Tris-HCl, pH 8.2 and WaLP was added to a final ratio of 1:100 and incubated at 37 °C overnight. Each reaction was stopped by acidification with TFA to 1% (v/v) and clarified by centrifugation at 20,000×*g* for 10 min at 4 °C. Peptide solutions were then desalted using tC18 Sep-Pak (either 200 or 500 mg, Waters) as previously described (38) and lyophilized.

For the SILAC experiments, initial purification of the peptides was performed using basic pH reverse phase chromatography on a 100 mm × 10 mm ID bridged-ethylene hybrid (BEH) C18 column with 5 µm particles (Waters) in 10 mM ammonium formate pH 10. Using a flow rate of 3 ml min^{-1} and a gradient of 0–100% 10 mM ammonium formate in 90% ACN over 1 h, thirty-two fractions were collected and every fourth fraction was pooled to obtain four fractions from each digest. After lyophilization and resuspension in 0.5% TFA, the peptides were desalted again with Sep-Pak tC18 cartridges (50 or 200 mg size), lyophilized, and stored at −80 °C.

6. Preparation of tissue samples

Murine tissue samples such as brain, heart, muscle, and liver were obtained from mature BALB/c mice (Cell Signaling Technology). Tissue was homogenized in 8 M urea lysis buffer

(50 mM Tris pH 8.0, 8 M urea, 1 mM vanadate, 2.5 mM sodium pyrophosphate, 1 mM beta-glycerol-phosphate). Lysate was reduced by 4.5 mM DTT for 30 min at 55 °C. Reduced lysate was alkylated with 10 mM iodoacetamide for 15 min at 25 °C in the dark. Sample was diluted fourfold with 50 mM Tris, pH 8, and digested overnight with WaLP at weight ratio of 1:100 at 37 °C overnight with slow rotation. Digested peptide lysate was acidified with 20% TFA to a final concentration of 1%, and peptides were desalted over 360-mg Sep-Pak Classic C18 columns (Waters, Milford, MA). Peptides were eluted with 40% acetonitrile in 0.1% TFA and lyophilized.

7. nLC-MS/MS for SILAC experiments

For the SILAC samples, WaLP digest data were collected on an Orbitrap Fusion mass spectrometer (Thermo Scientific) equipped with a Proxeon Easy nLC 1000. Samples were resuspended in 8 µl of 5% formic acid/5% acetonitrile and were loaded onto a 100 µm inner diameter fused-silica micro capillary with a needle tip pulled to an internal diameter less than 5 µm. The column was packed in-house to a length of 35 cm with a C18 reverse phase resin (GP118 resin 1.8 µm, 120 Å, Sepax Technologies). The peptides were separated using a 120 min linear gradient from 3 to 25% buffer B (100% ACN + 0.125% formic acid) equilibrated with buffer A (3% ACN + 0.125% formic acid) at a flow rate of 600 nL min^{-1}. Precursor spectra were collected with a target resolution of 120,000 in the Orbitrap using a scan range of 300–2000 *m/z*. The top 10 precursors with intensity greater than 5000 were fragmented sequentially with CID and ETD in the ion trap with the rapid scan rate, resulting in two separate spectra for each selected precursor ion.

For SILAC samples after WaLP digestion, two searches for each file were performed, one specifying fixed light lysine and one specifying fixed heavy lysine. All searches allowed variable oxidation of methionine, variable protein N-terminal methionine loss and acetylation at alanine or

serine, variable peptide N-terminal pyro-glutamate from Q, variable KGG, and fixed modification of cysteine (sample dependent). Following the conversion of the raw data to the open format .mzXML using Proteowizard (39), peptides were identified by database search (2015 Uniprot reviewed human proteome) with MS-GF+ trained for peptides from WaLP digestion as described previously (21). Current releases of MS-GF+ (2017-01-13) are available at https://omics.pnl.gov/software/ms-gf with development and support at https://github.com/sangtaekim/msgfplus. CID/ETD pairs were analyzed together when available from the Fusion data. The .mzid output from fixed heavy lysine database searches was processed using R scripts (40) to combine the mass of heavy lysine and diGlycine into one modification. R scripts used in this work are available at https://github.com/komiveslab/SUMO. All .mzid files were then converted to pepXML for compatibility with TPP (41) using idconvert.exe (Proteowizard 39). PeptideProphet was used to refine heavy and light identifications separately (42). iProphet was used to combine files corresponding to HPRP fractions from a single condition, and to combine the results from separate heavy and light database searches (43). PTMprophet was used to generate localization scores for KGG (41,44). SUMO-remnant-containing peptides were then filtered to 1% FDR by probability score, and ptm sites with localization scores below 0.9 were removed.

The trypsin-digested SILAC samples were analyzed on a Q-Exactive as previously described (14) and data were analyzed with the Core platform from Harvard University (36,37).

To quantify changes in SUMOylation or ubiquitylation upon MG132 treatment, the ratio of the peak areas for the heavy vs. the light were calculated and when protein modifications were identified by multiple peptides, the weighted average of the ratio was computed. The log(2) of each ratio was plotted and sites with log(2) ratios outside 1 s.d. from the median were flagged as

"changing." The proteins containing these sites were then analyzed for GO term enrichment with metascape (http://metascape.org/gp/index.html#/main/step1) using all human proteins as background.

8. Data availability

All data, spectra, and result files are accessible via FTP at: Public FTP link: massive.ucsd.edu/MSV000081018—Massive ID: MSV000081018. Proteome exchange ID: PXD006398. The data available on the web interface reflects the MS-GF+ output, before the 1% FDR filter was applied in TPP. The data that support the findings of this study are also available from the corresponding author on request.

C. Results

1. SUMOylation site profiling by WaLP digestion

Previous studies demonstrated the utility of WaLP digestion in shotgun proteomics platforms (21). Our observation that threonine was among the preferred amino acids in the P1 position for WaLP digestion (21) led us to hypothesize that digestion of SUMOylated proteins with WaLP, which should cleave after the threonine in the SUMO C-terminal sequence, TGG, would generate a SUMO-remnant diglycyl-lysine (KGG) at SUMOylation sites (Fig. 2.1a). The same workflow used for Ub-remnant profiling could then be used to globally profile SUMO attachment sites (Fig. 2.1b, c). Although WaLP can simply be substituted for trypsin during sample preparation, identification of non-tryptic peptides produced from WaLP digestion is challenging because search engines score on the basis of b and y ion series that are expected from tryptic peptides with a C-terminal positive charge. We previously demonstrated that identification of peptides arising from WaLP digestion benefits from electron transfer dissociation (ETD) (22), whereas higher-energy collisional dissociation generates internal ions that can complicate peptide

spectral matching (21). WaLP cleaves after at least four different amino acids requiring the use of "no enzyme" specificity in database searches, which is challenging for many publicly available search algorithms. To facilitate mapping of SUMOylation sites using WaLP digestion, we used the MS-GF+ search engine whose scoring function can be trained using identifications from an initial search (23,24).

2. Identification of novel SUMOylation sites using WaLP

As an initial test of the method, we digested lysates generated from human cell lines with WaLP and enriched for KGG-containing peptides using established protocols for Ub-modified peptide enrichment (14). Subsequent analysis by mass spectrometry and database searching using MS-GF+ resulted in the identification of 2051 unique KGG-containing peptides, which were confidently localized (PTMprophet score >0.9) to 1209 unique sites. Comparison of the SUMOylation sites identified using our WaLP digestion approach with the SUMOylation sites amalgamated by Hendriks et al. (17) revealed an overlap of only 30% (Fig. 2.2a). When compared to previously reported SUMOylation sites from Hendriks et al. (7710 sites) (17), Phosphosite Plus (780 sites) (25), and Uniprot (1863 sites) (26), 826 were novel. Peptides containing SUMO remnants from WaLP digestion may correspond to sequences that are not covered by tryptic digestion due to the abundance or lack of nearby tryptic cleavage sites (21). We found several examples of such cases. For example, we identified a novel SUMOylation site in DNA-PK (PRKDC, P78527), which would have been in a tryptic peptide of length 57 (Fig. 2.2b), and another in Sp100 (P23497), which would have been a tryptic peptide of only six amino acids (Fig. 2.2c). Multiple SUMOylation sites were found in over 60% of the proteins identified, with 5% having more than six SUMOylation sites. The SUMOylation sites identified in this study correspond to the expected motifs as described by others. We found that 31% of the SUMO

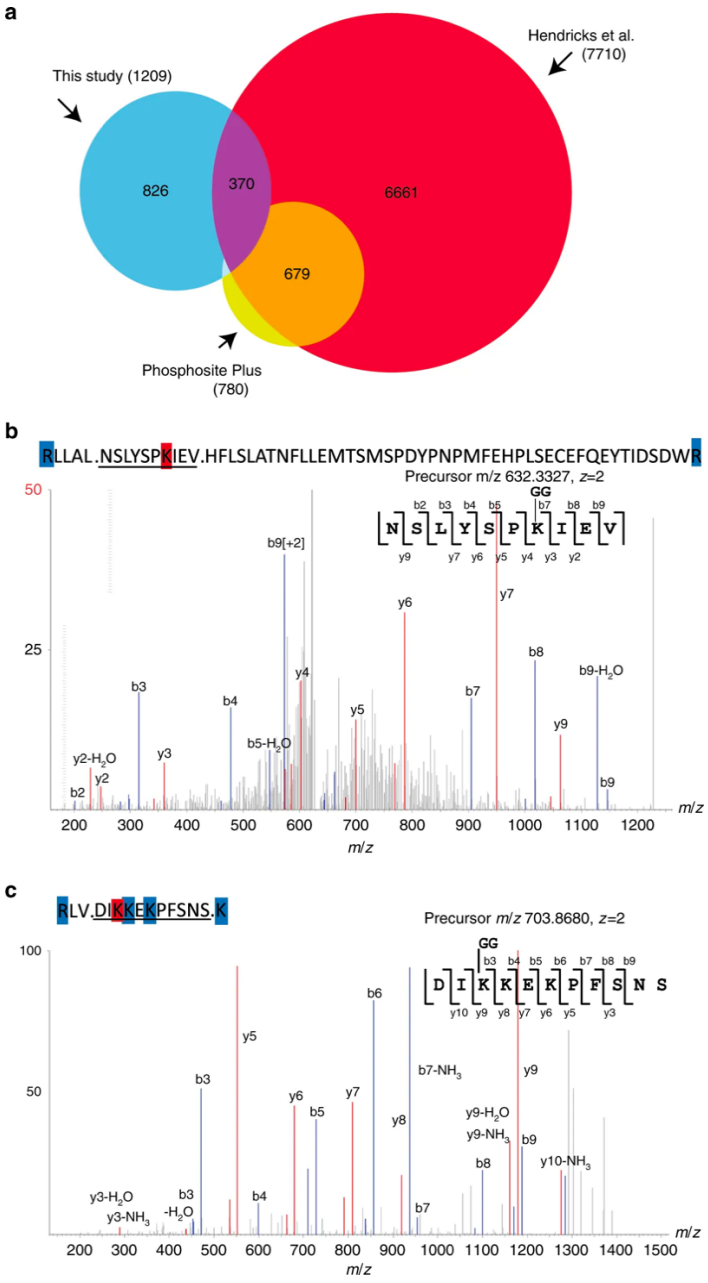

Figure 2.2: Summary of SUMOylation site identifications from this study and comparison with previous studies. a. Venn diagram showing overlap between the 1209 SUMO sites identified in this study from both Hela and HCT116 cell lines (blue) and all SUMO modifications from either Hendriks et al. (red) or Phosphosite Plus (yellow). b. Identified spectrum of a novel SUMO modification site identified in DNA-PK that would have resulted in a 57-amino-acid tryptic peptide. The sequence above the spectrum shows the tryptic cleavage sites in blue, the modification site is in red, and the matched sequence from the WaLP digest is underlined. c. Identified spectrum of a previously uncharacterized SUMO modification site in Sp100 that would result in only a six- residue tryptic peptide.

attachments occurred at the "forward" sequence motif (ΨKX[E|D]) (Fig. 2.3a), 9% corresponded to the "inverted" motif ([E|D]XK) (Fig. 2.3b), and 60% did not correspond to either consensus but these sites were somewhat enriched in acidic residues (Fig. 2.3c). Gene ontology term enrichment was performed to functionally annotate the SUMO-modified proteins. As reported previously, SUMOylation sites were primarily found in nuclear proteins and SUMOylated proteins were enriched for proteins involved in chromatin biology, RNA metabolism, and transcription (Fig. 2.4).

It is possible that the same lysine within an individual protein can be either modified by Ub or SUMO and that SUMOylation may antagonize ubiquitylation (27-29). In fact, 30% of the sites identified in this study were either previously reported to be ubiquitylated in the Phosphosite Plus database or were found in our experiments to be ubiquitylated. This overlap is similar to that previously reported by Hendriks et al. (18).

3. Validation of SUMOylation sites using in vitro deSUMOylation

Our previous studies on WaLP digestion specificity indicated that WaLP rarely cut after arginine residues (Fig. 2.1b) leads to the possibility of using WaLP to generate diGlycine-remnant peptides from SUMO-modified proteins and not Ub-modified proteins. However, when using WaLP to isolate and identify SUMOylated peptides in this study, we occasionally identified peptides whose MS/MS spectra matched best to peptides that resulted from cleavage after arginine either on the N or C terminus. These peptides were rare, with percentages varying from 3–8%, which we attribute at least partly to our inability to completely remove the trypsin used to detach cells during scale-up. Trypsin was avoided in cell harvesting, scraping was used instead. Even still, the fact that the P1 Arg was usually present in a sequence that had multiple positively charged residues in a row led us to question whether WaLP digestion could occasionally result in the

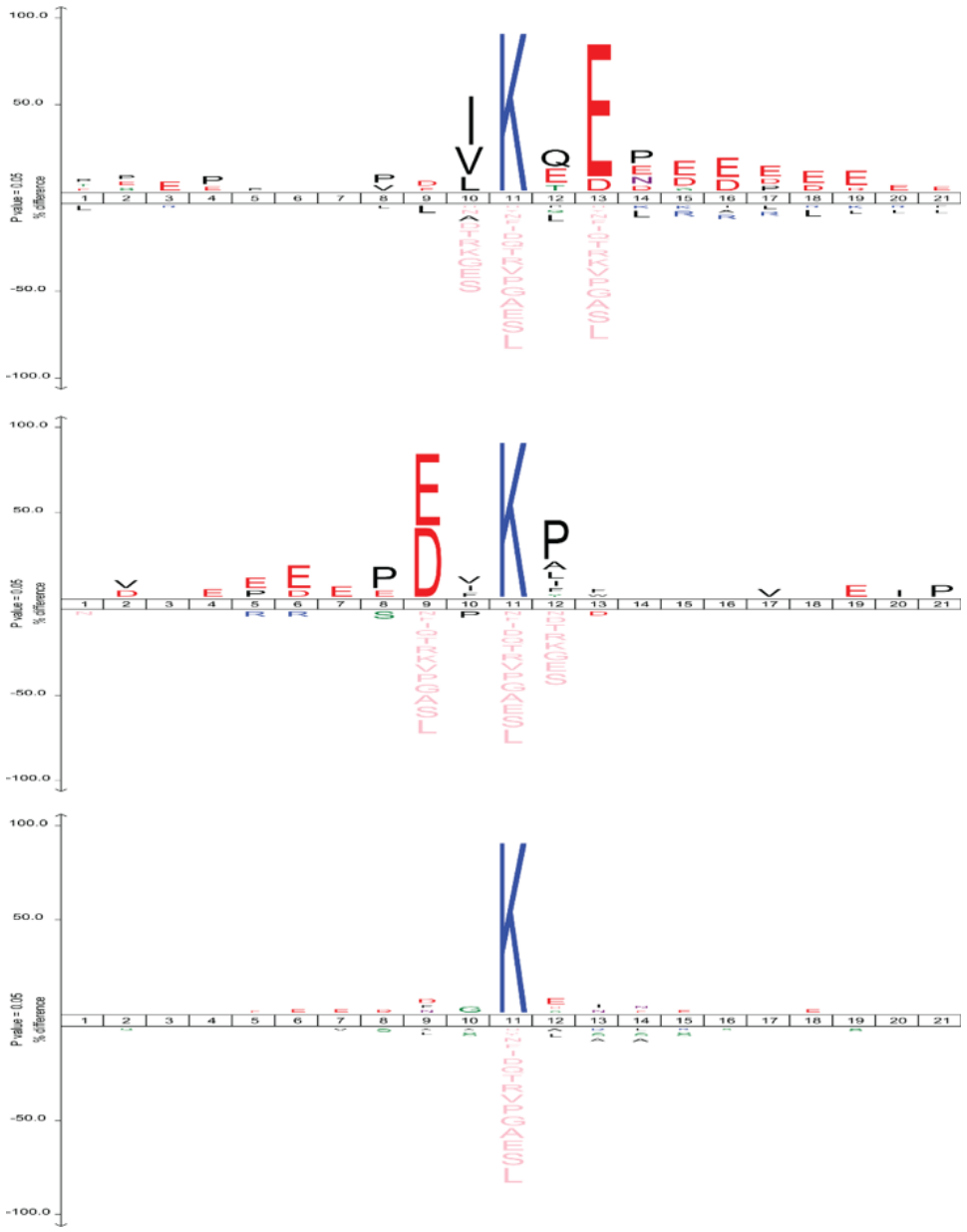

Figure 2.3: "IceLogo" motif analysis of the KGG sites discovered using WaLP digestion from Hela and HCT116 cells. SUMO attachment at the "forward" sequence motif (ΨKX[E|D], where Ψ indicates a hydrophobic amino acid) (a) occurred in 31% of identified sites, at the "inverted" motif ([E|D]XK) (b) in 9% identified sites, and at other sites (c) in 60% of the identified sites.

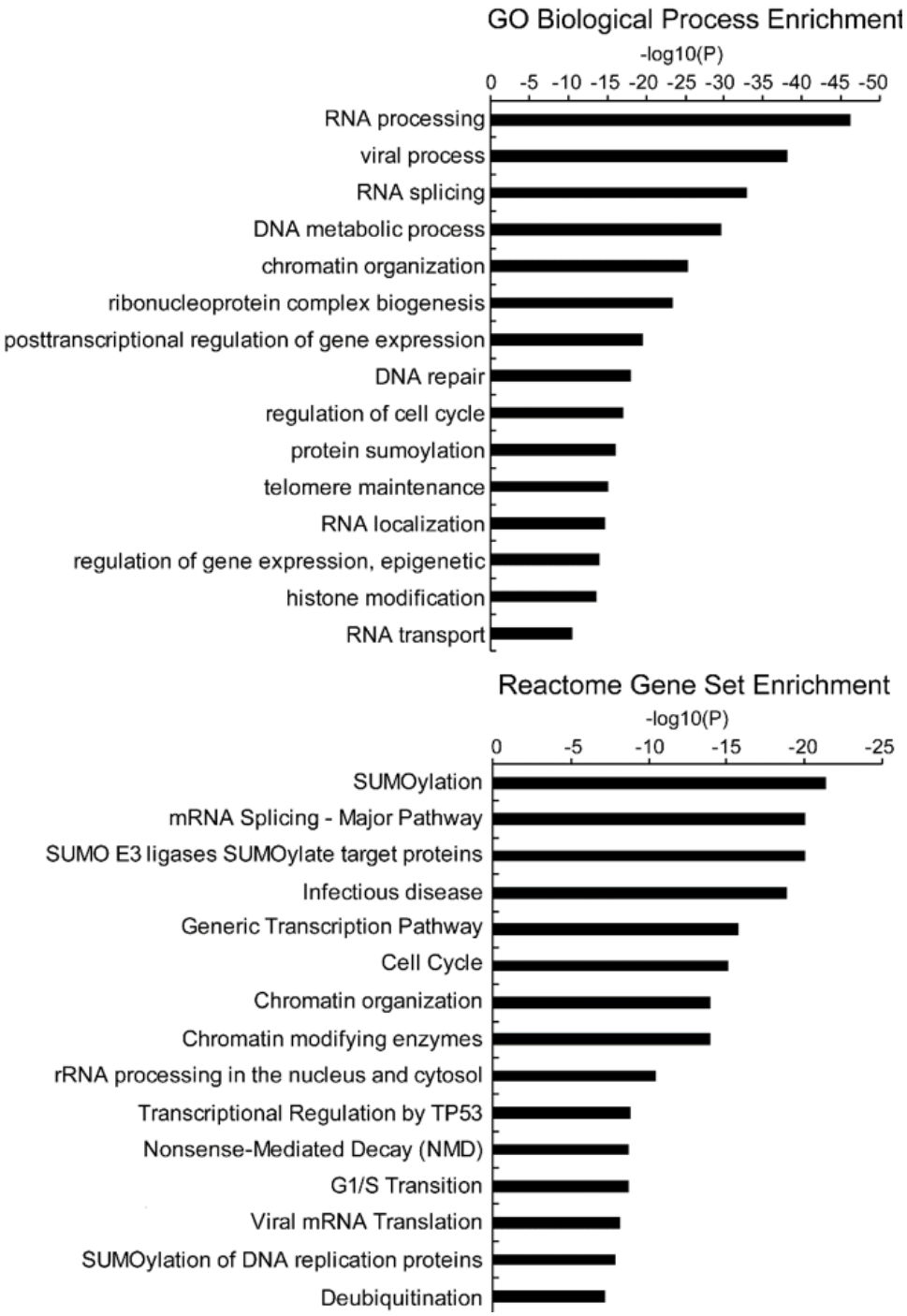

Figure 2.4: Gene ontology enrichment analysis of all proteins identified with SUMOylation sites from Hela and HCT116 cells. Depicted is the -log10 enrichment of selected enriched GO-biological processes (top) or GO-cellular compartments (bottom) using the predicted human proteome as background.

generation of a KGG-peptide from a ubiquitylated protein. To evaluate this possibility and to validate our approach, native cell lysates were either untreated or treated with recombinant deSUMOylating enzymes, SENP1 and SENP2, prior to digestion with either trypsin or WaLP. Western blot analysis revealed that in vitro treatment with SENP1/2 resulted in a robust reduction of SUMO1- and SUMO2/3-modified proteins while leaving ubiquitylated proteins unaltered (Fig. 2.5a). Digestion of SENP1/2-treated lysates with trypsin or WaLP and subsequent analysis by mass spectrometry revealed a dramatic reduction in the abundance of KGG-modified peptides from WaLP-digested samples compared to untreated samples (Fig. 2.5b). Importantly, the abundance of KGG-peptides resulting from trypsin cleavage, which would arise from ubiquitylated substrates, was unaltered by SENP1/2 treatment (Fig. 2.5b). The observation that 88% of sites identified after WaLP digestion were decreased at least twofold upon SENP1/2 treatment validated our approach. Conversely, treatment of lysates with a promiscuous deubiquitylating enzyme, Usp2cc, would result in a specific decrease in the amount of a KGG-peptide from a ubiquitylated site without altering SUMOylated peptides. Consistent with our prediction, less than 2% of KGG-peptides generated from WaLP-digested cells decreased upon Usp2cc treatment (a similar percentage to the FDR), whereas 97% of the KGG-containing peptides generated from trypsin digestion decreased after Usp2cc treatment (Fig. 2.5c). Taken together, we conclude that WaLP digestion and subsequent immuno-affinity enrichment of KGG-modified peptides specifically identifies endogenous SUMOylated peptides.

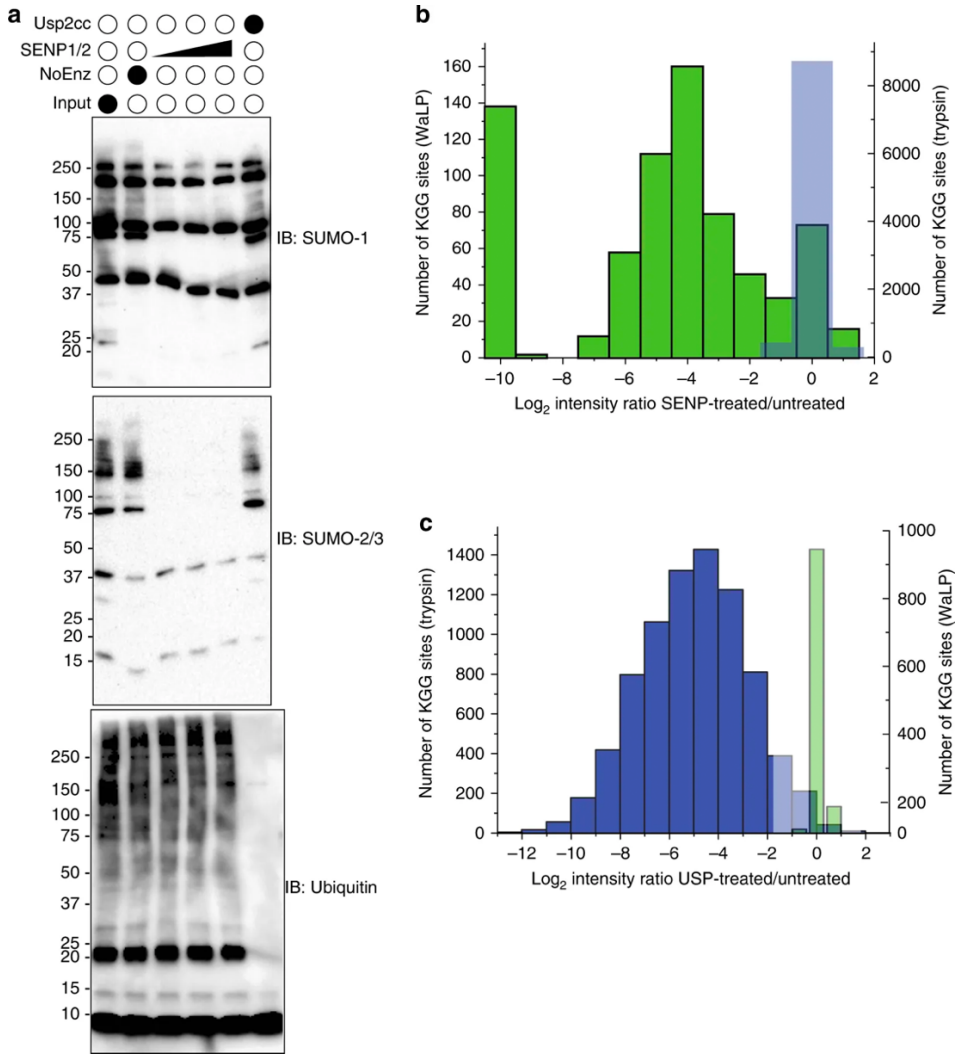

Figure 2.5: Identification and quantitation of SUMO and Ub modification sites upon in vitro deSUMOylation in Hela cell extracts. a. Extracts were either untreated or treated with varying amounts of SENP1 and SENP2 enzymes or Usp2cc and then analyzed by SDS-PAGE and blotted for SUMO1 or SUMO2/3 or ubiquitin. b. Bar graph showing results from quantitation of KGG-peptides after treatment with SENP1/2, illustrating the reduction of KGG sites observed after WaLP digestion. Green bars indicate SUMO modification sites identified and quantified upon WaLP digestion (left, y axis) and blue bars indicate Ub modification sites identified and quantified upon trypsin digestion (right, y axis). c. Bar graph showing results from quantitation of KGG-peptides after treatment with Usp2cc, illustrating the retention of KGG sites observed after WaLP digestion. Colors are the same as in b

4. Parallel Ub and SUMO site identification

A unique feature of our method is the ability to identify ubiquitylation and SUMOylation sites in parallel from the same cell or tissue lysate. The lysate can be simply split in half, with one half digested with WaLP and the other half digested with trypsin (Fig. 2.1c, d). The subsequent immuno-affinity enrichment steps are identical, although mass spectrometry and data processing were optimized for tryptic peptides for Ub and non-tryptic peptides for SUMO. As is already well established, SUMOylation is a lower abundance modification than ubiquitylation, and indeed, we identified 6472 Ub sites in the same samples in which 1209 SUMO sites were identified. Qualitative comparisons of the data revealed all possible types of dual modification scenarios. We found 2713 proteins that were ubiquitylated, 768 proteins that were SUMOylated, and 407 proteins that had both SUMO and Ub. Some had a large number of ubiquitylated lysines and few SUMOylated lysines while others had larger numbers of SUMO-modified lysines and few ubiquitylated sites. Those for which we observed both modifications mostly carried those modifications on different residues. We observed 243 lysines that were found to be Ub modified in the trypsin-digested samples and SUMO modified in the WaLP-digested samples. These results indicate that our method can identify proteins that are modified by both SUMO and Ub on the same lysine residues in the same sample.

5. Parallel quantification of Ub and SUMO sites

Previous studies indicated that SUMO modification could be stimulated by cell stresses such as heat shock and proteasome inhibition (30,31). Using SILAC-based quantitative proteomics, and our method allows for the parallel capture, identification, and quantification of SUMO- and Ub-modified proteins in response to various cell perturbations. As it is well established that proteasome inhibition results in global alteration of the Ub-modified proteome,

we evaluated the dynamic response of both the Ub- and SUMO-modified proteome to proteasome inhibition. Metabolically labeled cells were treated with MG132 and mixed with unlabeled cells prior to cell lysis. Again, the cell lysates were split and half was digested with trypsin for Ub analysis and the other half was digested with WaLP for SUMO analysis. We identified and quantified 330 unique SUMOylated sites and 2621 unique ubiquitylation sites from biological replicate samples. The results for both experiments were normally distributed (Fig. 2.6) and the results from the Ub analysis were comparable to previously published data in which proteasome inhibitors were utilized (14,32). For the SUMO analysis, 30 sites increased in abundance by at least twofold and 72 sites decreased in abundance by at least twofold upon MG132 treatment (18). Changes in modification abundance are not likely due to changes in protein levels as protein levels were not found to change significantly during a 4-h MG132 treatment (14). We observed 103 proteins for which both SUMOylation and ubiquitylation could be quantified. These proteins had a total of 591 modified lysines. Of the total 591 modified lysines, 53 were found to be both SUMOylated and ubiquitylated.

The proteins with the most abundant sites that were both SUMOylated and ubiquitylated were SUMO and Ub themselves, so we analyzed these to see if any sites were reciprocally regulated. It is important to note that we observed multiple modified peptides with more than one SUMO modification on SUMO proteins only in the WaLP digest. This indicates the presence of SUMO molecules that are simultaneously SUMOylated at different lysine residues and may have a branched-chain architecture. Interestingly, SUMO-modified ubiquitin was only observed on single lysine residues, suggesting that multiply SUMO-modified ubiquitin is a rare or nonexistent event. Conversely, peptides with more than one Ub modification on ubiquitin were observed only in the trypsin digest. These results are consistent with the presence of poly-SUMO (detected by

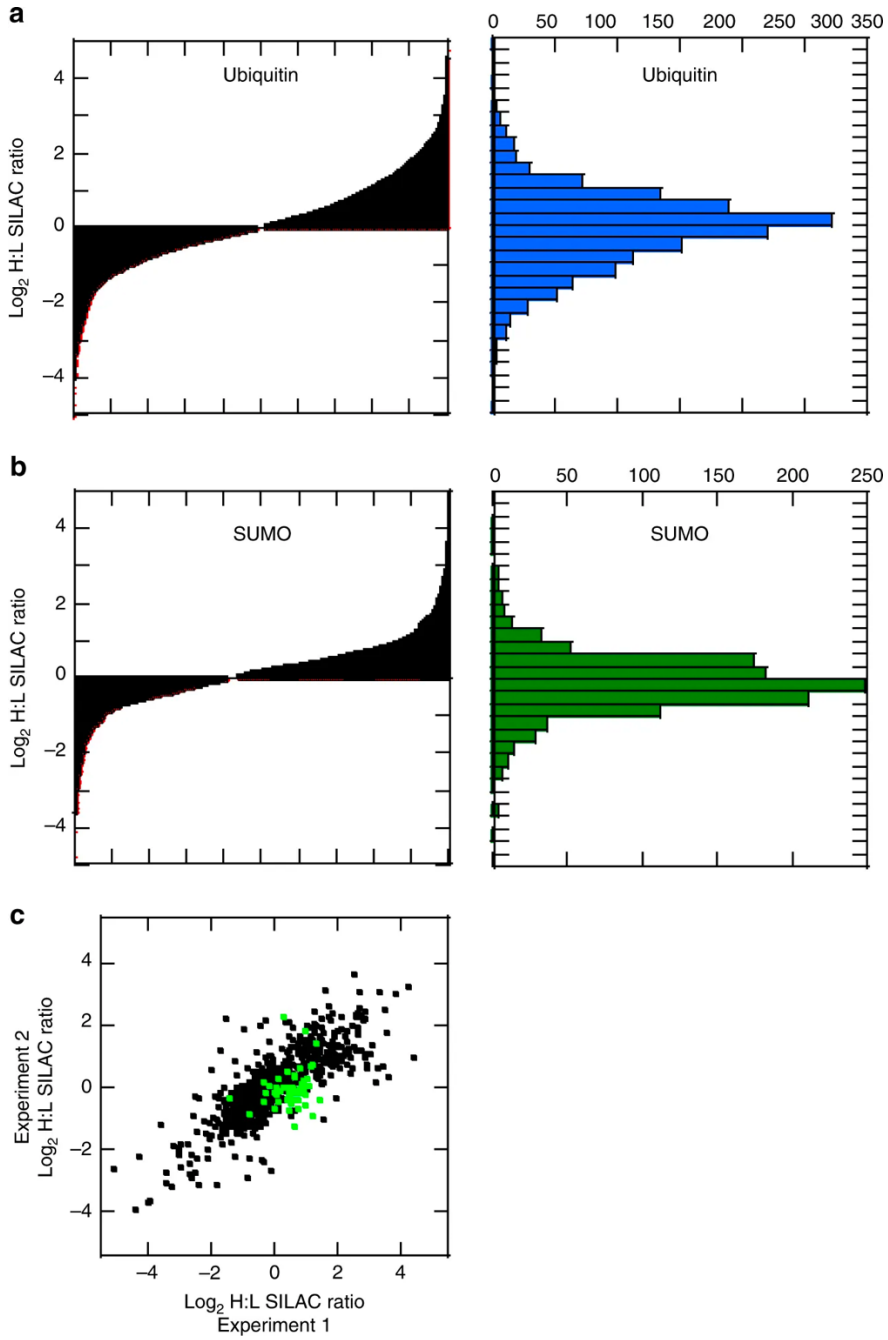

Figure 2.6: Parallel identification and quantitation of the SUMO and Ub-modified proteome upon MG132 treatment of HCT116 cells. Plot of normalized log(2) SILAC ratios (heavy/light) for all quantified unique Ub modification sites a or SUMO modification sites b upon MG132 treatment. The left graphs depict the ordered distribution of the SILAC ratios and right graphs depict the histogram of SILAC ratios. c Comparison of the SILAC ratios from biological replicates for ubiquitylation sites (black symbols) or SUMOylation sites (green symbols)

WaLP) and poly-Ub (detected by trypsin) but only mono-modifications of the opposite protein. We quantified ubiquitylation and SUMOylation of three lysine residues in SUMO2/3 upon MG132 treatment (Fig. 2.7a). Neither ubiquitylation nor SUMOylation of lysines 33 or 45 in SUMO2/3 were dramatically altered upon proteasome inhibition. However, ubiquitylation of lysine 11 on SUMO2/3 increased, whereas SUMOylation on this site decreased upon MG132 treatment (Fig. 2.7a). Cross-modification of SUMO2/3 by ubiquitin has been previously observed, and consistent with these results, ubiquitylation of SUMO2/3 on lysine 11 increases upon proteasome inhibition (31).

For ubiquitin, we were able to quantify Ub and SUMO modification of lysines 27, 29, 33, 48, and 63 (Fig. 2.7b), Both SUMOylation and ubiquitylation of ubiquitin responded similarly to MG132 treatment (Fig. 2.7b). Both SUMO and Ub modification of lysines 48, 29, and 33 increased in abundance upon MG132 treatment. Conversely, both SUMO and Ub modifications on lysine 63 of Ub decreased in abundance in the MG132-treated cells. Similarly, we observed cross-modification of another ubiquitin-like (Ubl) protein, Nedd8. Lysines 11 and 48 on Nedd8 were observed to be modified by both Ub and SUMO and the extent of modification was largely unperturbed by proteasome inhibition (Fig. 2.7c). Trypsin digestion followed by KGG immuno-affinity enrichment cannot distinguish between ubiquitylation or neddylation due to the presence of an arginine preceding the C-terminal diGlycine in both Ub and Nedd8. As such, it is possible that Nedd8 may be poly-neddylated at these sites rather than ubiquitylated. Consistent with this observation, DCN1 was observed to be diGlycine-modified at a lysine residue previously characterized to be a neddylation site and this same lysine residue was SUMOylated (Fig. 2.7c). Interestingly, the abundance of both SUMOylated and neddylated DCN1 decreased upon proteasome inhibition. These observations suggest that cross-modification of Ubl proteins is

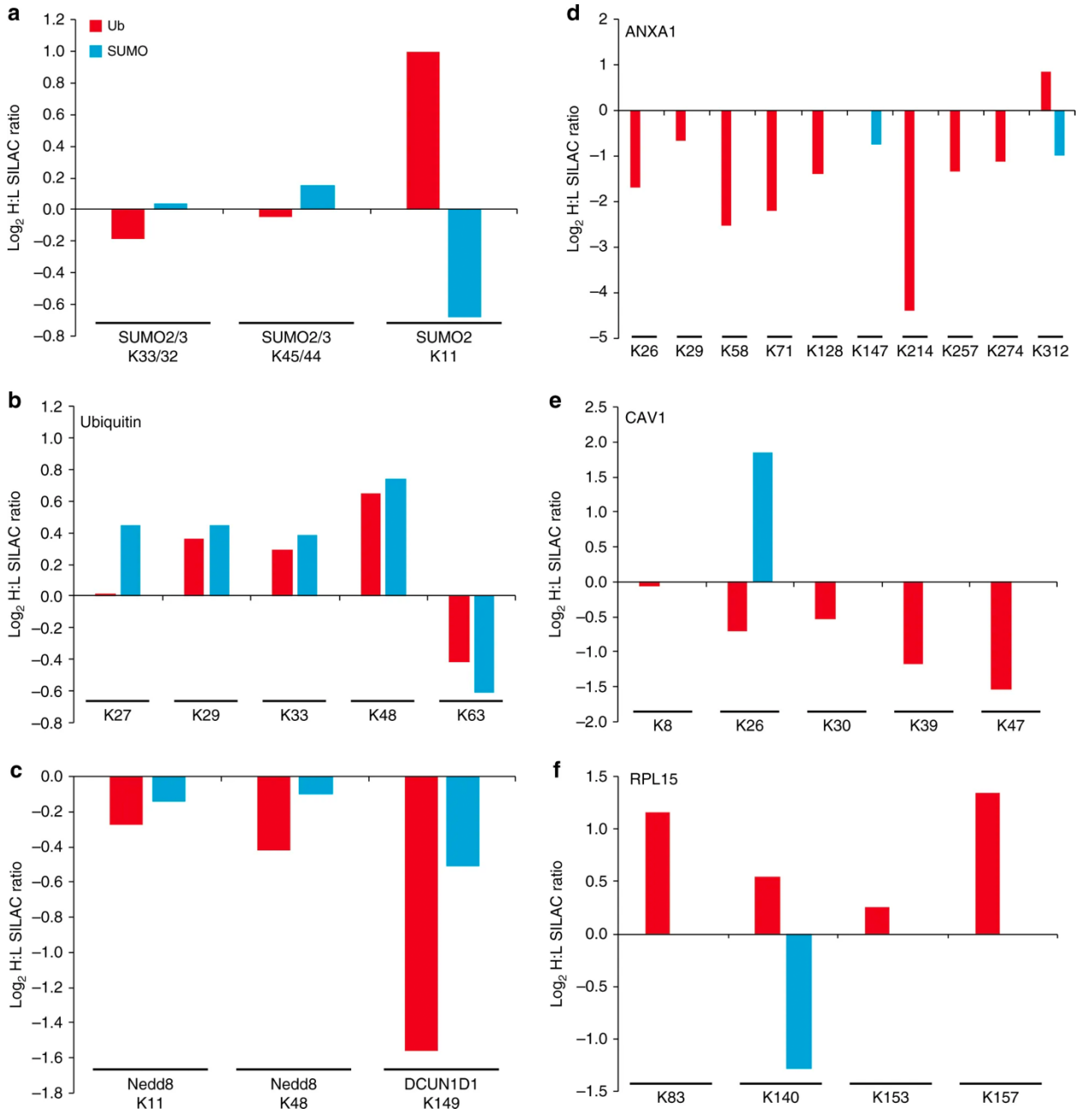

Figure 2.7: Log2 SILAC ratios (heavy-MG132-treated:light-untreated HCT116 cells) for quantified ubiquitin (red) and SUMO (blue)-modified peptides observed on SUMO-2/3. (a), ubiquitin (b) Nedd8 and DCN1 (c), annexin-1 (d), caveolin-1 (e), or RPL15 (f). The position of the modified lysine is indicated.

prevalent and may possibly be antagonistic modifications.

Separate from the highly abundant Ubl proteins, we observed a few examples of proteins that were modified by both SUMO and Ub on the same lysine residue with differential responses to proteasome inhibition. For example, we quantified two SUMOylation and nine ubiquitylation sites on annexin A1 (Fig. 2.7d). We observed SUMOylation at Lys 147, which has not been previously reported, and at Lys 312, a known SUMOylation and ubiquitylation site (18). This C-terminal lysine appears to experience reciprocal regulation by MG132 as the abundance of the SUMO-modified form was significantly decreased, whereas the ubiquitylation was significantly increased (Fig. 2.7d). Further, lysine 26 in caveolin-1 and lysine 140 in the large ribosomal protein 15 were modified by both Ub and SUMO and these modifications displayed divergent abundance alterations upon proteasome inhibition (Fig. 2.7e, f). These results indicate that reciprocal modification by SUMO and Ub on the same lysine within a target protein may be a relatively rare event, but that it clearly does occur at some sites in the proteome and that these modification events may impart differential functional outputs for the substrate proteins. These results also point to a unique strength of our method in that it allows for parallel mapping of regulated ubiquitylation and SUMOylation sites from a single sample.

6. Identification of SUMOylation in tissue samples

One advantage of our method is that it allows for the identification of endogenous SUMOylation sites from native tissues without exogenous SUMO protein expression. To establish this application of our method, we processed murine tissue from brain, heart, muscle, and liver to establish this application of our method, we processed murine SUMOylated proteins in vivo. We identified 144 unique SUMOylation sites across the four tissues. Overall, muscle and liver had the most similar SUMOylated proteins, sharing ~70% of the total sites observed in those tissues,

whereas brain had the most unique SUMOylated proteins that were not found in any other tissue type. We identified the well-characterized SUMOylated protein RANGAP1 in all tissues validating our approach. Interestingly, we again identified ubiquitin as a SUMO-modified protein with lysines 48 and 63 serving as the SUMOylation sites in all tissues. This result suggests that ubiquitin is SUMOylated at critical lysine residues in vivo and validates that our approach can be successfully applied toward the identification of endogenous SUMOylation sites in tissues.

D. Discussion

We set out to develop a method for global profiling of native SUMOylation events by taking advantage of the propensity of WaLP for cleavage after threonine. By simply substituting WaLP for trypsin, it was possible to immunopurify and identify a large number of KGG-containing peptides corresponding to SUMO remnants. A large number of the identified sites corresponded to as-yet-unreported SUMOylation events. Several reasons could explain the large number of new sites identified. First, as we reported previously, the orthogonal specificity of WaLP allows cleavage of proteins at sites that may not be accessible to trypsin (21). Second, although previous studies attempted to achieve minimal expression of their mutant SUMO construct, it is possible that slight overexpression of SUMO or the presence of mutant sequences could cause unnatural SUMO attachment. Third, our method does not differentiate between SUMO1–4, whereas Hendriks et al. examined only SUMOylation sites utilizing SUMO-3 attachment (18). One caveat is that WaLP also cleaves after Leu and to some extent Ile, so the method does not distinguish between SUMOylated, Fat10ylated, or Fub1ylated proteins. However, the observation that 88% of WaLP diGly proteins were reduced upon SENP1/2 treatment argues that the vast majority of observed KGG sites arise from SUMOylation. Additionally, analysis of the observed sites recapitulates previous reports of the expected motifs of SUMOylation and GO term enrichment

(19).

A powerful advantage of our method is that it allows for simultaneous determination of ubiquitylation and SUMOylation in the same sample. The same population of cells or tissue can be subjected to analysis of both Ub attachment and SUMO attachment simply by splitting the sample in two and digesting half with trypsin and the other half with WaLP. The samples can then be processed in parallel to immunopurify peptides for the presence of the KGG modification, and sequence by mass spectrometry. For the mass spectrometry, it is best to use optimized ionization approaches and data analysis tailored to the non-tryptic WaLP peptides. Although the commercial antibody used in this study was developed for enrichment of KGG-peptides from Ub modification, it appears to also efficiently immunopurify KGG-peptides from SUMO modification. Finally, the method allows for identification of SUMOylated proteins under native conditions including from tissue samples.

E. Acknowledgements

Chapter II, in full, is a reprint that the dissertation author was the principal researcher and author of. The material appears in Nature Communications. (Lumpkin, R. J., Gu, H., Zhu, Y., Leonard, M., Ahmad, A. S., Clauser, K. R., Meyer, J. G., Bennett, E. J., and Komives E. A. (2017). Site-specific identification and quantitation of endogenous SUMO modifications under native conditions. Nat. Commun., 8, 1171.)

F. References

1. T. Makhnevych, Y. Sydorskyy, X. Xin, T. Srikumar, F. J. Vizeacoumar, S. M. Jeram, Z. Li, S. Bahr, B. J. Andrews, C. Boone, B. Raught, Global map of SUMO function revealed by protein-protein interaction and genetic networks. Mol. Cell 33, 124–135 (2009).

2. A. Williamson, A. Werner, M. Rape, The Colossus of ubiquitylation: decrypting a cellular code. Mol Cell 49, 591-600 (2013).

3. A. Flotho, F. Melchior, Sumoylation: A Regulatory Protein Modification in Health and Disease. Annu Rev Biochem 82, 357-385 (2013).

4. J. Becker, S. V. Barysch, S. Karaca, C. Dittner, H. H. Hsiao, M. Berriel Diaz, S. Herzig, M. F. Urlaub H, Detecting endogenous SUMO targets in mammalian cells and tissues. Nat Struct Mol Biol 10.1038/nsmb.2526, 525-531 (2013).

5. K. Eifler, A. C. Vertegaal, Mapping the SUMOylated landscape. FEBS J 282, 3669-3680 (2015).

6. F. Golebiowski, I. Matic, M. H. Tatham, C. Cole, Y. Yin, A. Nakamura, J. Cox, G. J. Barton, M. Mann, R. T. Hay, System-wide changes to SUMO modifications in response to heat shock. Sci Signal 2 (2009).

7. A. C. Vertegaal, S. C. Ogg, E. Jaffray, M. S. Rodriguez, R. T. Hay, J. S. Andersen, M. Mann, A. I. Lamond, A proteomic study of SUMO-2 target proteins. J Biol Chem 279, 33791-33798 (2004).

8. J. A. Wohlschlegel, E. S. Johnson, S. I. Reed, J. R. Yates, Global analysis of protein sumoylation in Saccharomyces cerevisiae. J Biol Chem 279, 45662-45668 (2004).

9. J. T. Hannich, A. Lewis, M. B. Kroetz, S. J. Li, H. Heide, A. Emili, M. Hochstrasser, Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae. J Biol Chem 280, 4102-4110 (2005).

10. C. Denison, A. D. Rudner, S. A. Gerber, C. E. Bakalarski, D. Moazed, S. P. Gygi, A proteomic strategy for gaining insights into protein sumoylation in yeast. Mol Cell Proteomics 4, 246-254 (2005).

11. A. C. Carrano, E. J. Bennett, Using the ubiquitin-modified proteome to monitor protein homeostasis function. Mol Cell Proteomics 12, 3521-3531 (2013).

12. G. Xu, J. S. Paige, S. R. Jaffrey, Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. Nat Biotechnol 28, 868–873 (2010).

13. S. A. Wagner, P. Beli, B. T. Weinert, M. L. Nielsen, J. Cox, M. Mann, C. Choudhary, A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. Mol Cell Proteomics 10, M111 013284 (2011).

14. W. Kim, E. J. Bennett, E. L. Huttlin, A. Guo, J. Li, A. Possemato, M. E. Sowa, R. Rad, J. Rush, M. J. Comb, J. W. Harper, S. P. Gygi, Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell 44, 325-340 (2011).

15. M. Knuesel, H. T. Cheung, M. Hamady, K. K. Barthel, X. Liu, A method of mapping protein sumoylation sites by mass spectrometry using a modified small ubiquitin-like

modifier 1 (SUMO-1) and a computational program. Mol Cell Proteomics 4, 1626–1636 (2005).

16. H. A. Blomster, S. Y. Imanishi, J. Siimes, J. Kastu, N. A. Morrice, J. E. Eriksson, L. Sistonen, In vivo identification of sumoylation sites by a signature tag and cysteine-targeted affinity purification. J Biol Chem 285, 19324–19329 (2010).

17. I. A. Hendriks, R. C. D'Souza, B. Yang, M. Verlaan-de Vries, M. Mann, A. C. Vertegaal, Uncovering global SUMOylation signaling networks in a site-specific manner. Nat Struct Mol Biol 21, 927-936 (2014).

18. I. A. Hendriks, A. C. Vertegaal, A comprehensive compilation of SUMO proteomics. Nat Rev Mol Cell Biol 17, 581-595 (2016).

19. F. Lamoliatte, D. Caron, C. Durette, L. Mahrouche, M. A. Maroui, O. Caron-Lizotte, E. Bonneil, M. K. Chelbi-Alix, P. Thibault, Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. Nat Commun 5 (2014).

20. T. Tammsalu, I. Matic, E. G. Jaffray, A. F. M. Ibrahim, M. H. Tatham, R. T. Hay, Proteome-Wide Identification of SUMO2 Modification Sites. Science Signaling 7, rs2 1-10 (2014).

21. J. G. Meyer, S. Kim, D. A. Maltby, M. Ghassemian, N. Bandeira, E. A. Komives, Expanding proteome coverage with orthogonal-specificity α-lytic proteases. Mol Cell Proteomics 13, 823-835 (2014).

22. J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, D. F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Nat Acad Sci USA 101, 9528 -9533 (2004).

23. A. Guthals, C. Boucher, N. Bandeira, The Generating Function Approach for Peptide Identification in Spectral Networks. J Comput Biol in press (2014).

24. S. Kim, P. A. Pevzner, MS-GF+ makes progress towards a universal database search tool for proteomics. Nat Commun 5, 1-10 (2014).

25. P. V. Hornbeck, J. M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, V. Latham, M. Sullivan, PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res 40(Database issue), D261-270 (2012).

26. E. Meulmeester, M. Kunze, H. H. Hsiao, H. Urlaub, F. Melchior, Mechanism and consequences for paralog-specific sumoylation of ubiquitin-specific protease 25. Mol Cell 30 (2008).

27. A. Denuc, A. Bosch-Comas, R. Gonzàlez-Duarte, G. Marfany, The UBA-UIM domains of the USP25 regulate the enzyme ubiquitination state and modulate substrate recognition. PLoS One 4, e5571 (2009).

28. J. M. Desterro, M. S. Rodriguez, R. T. Hay, SUMO-1 modification of IkappaBalpha inhibits NF-kappaB activation. Mol Cell 2, 233-239 (1998).

29. K. Uzunova, K. Göttsche, M. Miteva, S. R. Weisshaar, C. Glanemann, M. Schnellhardt, M. Niessen, H. Scheel, K. Hofmann, E. S. Johnson, G. J. Praefcke, R. J. Dohmen, Ubiquitin-dependent proteolytic control of SUMO conjugates. J Biol Chem 282, 34167-34175 (2007).

30. J. Schimmel, K. M. Larsen, I. Matic, M. van Hagen, J. Cox, M. Mann, J. S. Andersen, A. C. Vertegaal, The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. Mol Cell Proteomics 7, 2107-2122 (2008).

31. J. M. Gendron, K. Webb, B. Yang, L. Rising, N. Zuzow, E. J. Bennett, Using the ubiquitin-modified proteome to monitor distinct and spatially restricted protein homeostasis dysfunction. Mol Cell Proteomics, mcp.M116.058420 (2016).

32. J. E. Mace, D. A. Agard, Kinetic and structural characterization of mutations of glycine 216 in alpha-lytic protease: a new target for engineering substrate specificity. J Mol Biol 254, 720-736 (1995).

33. J. V. Olsen, L. M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. M. Horning, M., Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 4, 2010-2021 (2005).

34. J. K. Eng, A. L. McCormack, J. R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5, 976-989 (1994).

35. E. L. Huttlin, M. P. Jedrychowski, J. E. Elias, T. Goswami, R. Rad, S. A. Beausoleil, J. Villen, W. Haas, M. E. Sowa, S. P. Gygi, A tissue-specific atlas of mouse protein phosphorylation and expression. Cell 143, 1174-1189 (2010).

36. J. Villen, S. A. Beausoleil, S. A. Gerber, S. P. Gygi, Large-scale phosphorylation analysis of mouse liver. . Proc Nat Acad Sci USA 104, 1488-1493 (2007).

37. J. Villén, S. P. Gygi, The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc 3 (2008).

38. D. Kessner, M. Chambers, R. Burke, D. Agus, P. Mallick, ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24, 2534–2536 (2008).

39. R. D. C. Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/ (2011).

40. D. D. E. Shteynberg, L. Mendoza, J. Slagel, H. Lam, A. Nesvizhskii, R. Moritz, PTMProphet: TPP Software for Validation of Modified Site Locations on Post-Translationally Modified Peptides. 60th American Society for Mass Spectrometry (ASMS) Annual Conference, Vancouver, Canada (2012).

41. A. Keller, A. I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74, 5383–5392 (2002).

42. D. Shteynberg, E. W. Deutsch, H. Lam, J. K. Eng, Z. Sun, N. Tasman, L. Mendoza, R. L. Moritz, R. Aebersold, A. I. Nesvizhskii, iProphet: Multi-level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. Mol Cell Proteomics 10, M111.007690 (2011).

43. D. Shteynberg, A. I. Nesvizhskii, R. L. Moritz, E. W. Deutsch, Combining results of multiple search engines in proteomics. Mol Cell Proteomics 12, 2383-2293. (2013).

Chapter III:

DECA, a comprehensive, automatic post-processing program for HDX-MS data

A. Introduction

Hydrogen deuterium exchange mass spectrometry (HDX-MS) probes protein structure and dynamics by measuring amide proton exchange. HDX reports on solvent-accessible surface area, protein-protein interfaces, and allosteric changes, and the data can be used to constrain docking or homology modeling (1–4). Our group recently showed that HDX-MS experiments in which proteins are incubated in a deuterated solvent for seconds to minutes probe microsecond to millisecond motions in samples (5).

Two caveats limit HDX-MS analysis. First, HDX-MS resolution is limited to the length of observable peptides. Approaches to achieve single amino acid resolution with HDX-MS include using several proteases to increase peptide overlap (6), ETD fragmentation to better localize deuterons on each peptide (3, 7), and deconvolution of isotopic envelopes to extract information from the peptide data itself (8, 9). Second, HDX-MS suffers from deuterium back exchange during sample handling and chromatography. Several methods exist for the correction of back exchange, a necessary step for downstream data processing to obtain reproducible numbers of amides exchanged. Correction for back exchange has previously been performed using several different methods including internal standards or using fully deuterated control samples to determine the extent of uptake on each peptide (8, 10, 11).

To decongest complex mixtures of biomolecules, it is helpful to integrate ion mobility (IMS) and m/z data simultaneously (12, 13). IMS-integrated HDX-MS uniquely excels in the study of large, complex protein samples (14, 15). Waters pioneered IMS as a third dimension of resolution for HDX-MS experiments with the SYNAPT system and the Protein Lynx Global Server/*DynamX* software workflow. IMS provides a third independent piece of information with which to identify each peptide and markedly improves the accuracy of ion clustering and

assignment of peptides from HDX-MS data (14).

As HDX-MS became more high-throughput, various efforts to automate aspects of the data analysis became available. *HXExpress*, a Microsoft Excel utility, was the first freely-available semi-automatic data analysis platform (16). This utility was further developed for deconvoluting overlapped peptide mass envelopes (17). Several automatic analysis platforms were developed beginning with *The Deuterator* and *HD Desktop*, which were further developed into *HDX Workbench*, a comprehensive fully automatic program that incorporates analysis of ETD fragments, isotopic fitting, overlapping peptide segmentation, statistical analysis, and data visualization (8). During this time, Schriemer's group developed *Hydra* (18), Mayer's group developed *Hexicon* (19), and Englander's group developed *ExMS* (9), all of which provide a similar list of functionalities. Realizing the need for more rigorous back-exchange correction, Z. Zhang developed *MassAnalyzer*, a fully-automated software that generated data in the form of protection factors (10). Two commercially available software programs were also developed, *HD Examiner* by Sierra Analytics and *DynamX* by Waters. These two software programs aptly demonstrate the two cultures that have grown out of the HDX-MS community; those researchers who desire a fully-automated platform such as that provided by *HD Examiner* and *HDX Workbench versus* those who wish to have some automation but with user-controlled examination of the raw data such as is provided by *DynamX*. Currently, only *HD Examiner* and *DynamX* allow import of IMS data. Several software programs also emerged that provided further data analysis once the initial centroid data was obtained. *HDX Analyzer* provides rigorous statistical analysis (20). *MEMHDX* (21) and *Deuteros* (22) allow downstream analysis of *HDX Workbench* and *DynamX* output data, providing further statistical analyses and data visualization. However, neither address back exchange analysis, which is missing from *DynamX*.

Here we present the Deuterium Exchange Correction and Analysis (DECA) software package, which was designed as a simple, rapid backend to *DynamX*, but can also be used for other data in .csv format. DECA provides several options for back exchange correction, resolution-increases by Overlapping Peptide Segmentation (OPS), a rigorous evaluation of the statistical significance of observed differences, and visualization tools. To our knowledge, DECA is the only freely-available software that provides all these functionalities in a single platform. In addition, DECA is able to extract detailed ion mobility, retention time, and deuterium uptake information from *DynamX* project files in order to measure the summary statistics of the ions assigned to each peptide, check the quality of the data through outlier analysis, and more accurately determine the statistical significance of differences between protein states than is possible with exported summary data. For data from other platforms which require back exchange correction and/or automatic analysis of overlapping peptides, DECA can be a helpful backend as well.

Here, we demonstrate the effectiveness of DECA on a representative HDX data set from *DynamX* and explain how it can be used for uptake data output files from other platforms. DECA is an initial attempt to provide software that will allow the HDX-MS community to comply with the recently published recommendations for performing, interpreting, and reporting HDX-MS experimental data. DECA is open-source with the intent that users will contribute additional functionalities and improvements.

B. Methods

1. Software Design.

DECA was written entirely in Python, and it implements modules from external libraries for visualization and analysis. The graphical interface implements the built-in Tkinter framework, and it was initially designed in PAGE, a Python GUI Generator. DECA performs statistical

analysis of imported data by implementing functions from the Scipy and Statsmodels libraries (23). Curve fitting in deuterium uptake plots uses a linear regression function from the Scipy library. Uptake plots, coverage maps and spectra are generated through functions from the Matplotlib library (24). The retention time prediction feature implements a function from the Pyteomics library (25). The PyInstaller script is used to package all DECA code and python dependences into executable binaries for OSX and Windows (26). The packages used are PSF, BSD, GPL or Apache licensed (All free software licenses, with some share-alike copyleft restrictions). The DECA source-code is compatible with Python 2.7 and Python 3.7, which are freely available and pre-installed on UNIX operating systems. This source code is executable through the python environment upon the installation of the external libraries mentioned above. The source code and binaries are available at github.com/komiveslab/DECA.

2. Data Format.

DECA primarily accepts tabular peptide lists of deuterium uptake data from CSV files. It was designed to import *DynamX* state data, *HDXWorkbench* data, and a third generic alternative. Sample data sets of each type are available in the supplementary information. Between the three data styles, DECA can import most tabular HDX-MS data with only minor alterations to the data set's column headers. For deeper statistical analysis and visualization of the raw spectra, a ".dnx" *DynamX* project file can be imported into DECA.

3. Data Analysis.

Each data set imported is evaluated to identify high relative standard deviations or errors in peptide mass values, retention times, and ion mobility values. The statistical significance is determined for every time point between each protein state in the data set. If a *DynamX* project file is imported, DECA interrogates every ion assigned to each peptide in order to flag outlier ion

assignments and report on the spread of m/z, retention time, and mobility values for each ion cluster.

4. Sample Dataset.

HDXMS experiments were performed and published previously (27). The data set includes peptides of a single protein, RelA, in two states: the RelA homodimer and the RelA-p50 heterodimer. Deuterium exchange was measured in triplicate at 30 s, 1 min, 2 min, and 5 min. Independent biological replicates of the triplicate experiment were performed to verify the results. Peptides were identified using the Protein Lynx Global Server and analyzed in *DynamX* 3.0 taking advantage of the ion mobility data. The peptides were identified from triplicate MS^E analyses and data were analyzed using PLGS 3.0 (Waters Corporation). Peptide masses were identified using a minimum number of 250 ion counts for low energy peptides and 50 ion counts for their fragment ions. The peptides identified in PLGS were then analyzed in DynamX 3.0 (Waters Corporation, Milford, MA) implementing a score cut-off of 7, the peptide must be present in at least 2 files, have at least 0.2 products per amino acid, a maximum MH+ error of 5 ppm, and less than a 5% error in the retention time. The relative deuterium uptake for each peptide was calculated by comparing the centroids of the mass envelopes of the deuterated samples *versus* the undeuterated controls following previously published methods (28) and corrected for back-exchange as previously described (29). The experiments were performed in triplicate, and independent replicates of the triplicate experiment were performed to verify the results.

5. Fitting of deuterium uptake plots.

A curve-fitting algorithm was implemented based on a least-squares minimization fit to the double- exponential function $y = a * (1 - \mathrm{e}^{-\mathrm{bt}}) + \mathrm{c} * (1 - \mathrm{e}^{-0.01t})$ to the HDX data. Constants a and b reflect the asymptote and rate of uptake by fast exchange, whereas constant c scales the

contribution of uptake by slow exchange. If the timepoints do not cover the curvature of the first exponential, the minimization may overfit the data. In such cases, the first exponential is minimized so $y_{t=\frac{1}{4}*n} = \frac{3}{4} * y_{t=n}$, where n is the lowest non-zero time point, to solve for the value of b that creates a curve that passes through the point at 3⁄4 of the uptake in 1⁄4 of the exposure of time point n. This curvature adjustment may be modified in the global settings. Constants a and c are subsequently minimized with the new b value to create a smooth curve to the first non-zero time point. Notably, in such scenarios where the data does not cover both sides of the inflection of the uptake curve, the curve fit does not reflect the rate of exchange of the peptide. DECA makes no attempt to extract the rates of exchange of individual amides from the peptide uptake curve. Such estimates are usually unreliable.

C. Results

1. Graphical User Interface.

The DECA interface enables rapid review and visualization of HDXMS data (Fig. 3.1). From the main window, the user can import a data file, which will display the contents of the file in a spreadsheet and enable data correction and visualization functions. Deuterium uptake plots are displayed adjacent to the data table, whereas additional figures will be displayed in separate windows. Diagnostic messages and errors are printed to a console window that opens with the main interface.

2. File Merging.

HDX experiments from a single study may be analyzed separately because of gaps in time between the experiments, changes to the instrumental setup, or in order to reduce the computational demand. Because data sets may contain redundant or complimentary data, DECA

may be used to intelligently merge data sets together by combining data or renaming proteins/states. Additionally, File Merge may be implemented either before or after performing back exchange correction on separate files in order to compare data sets with different levels of back exchange.

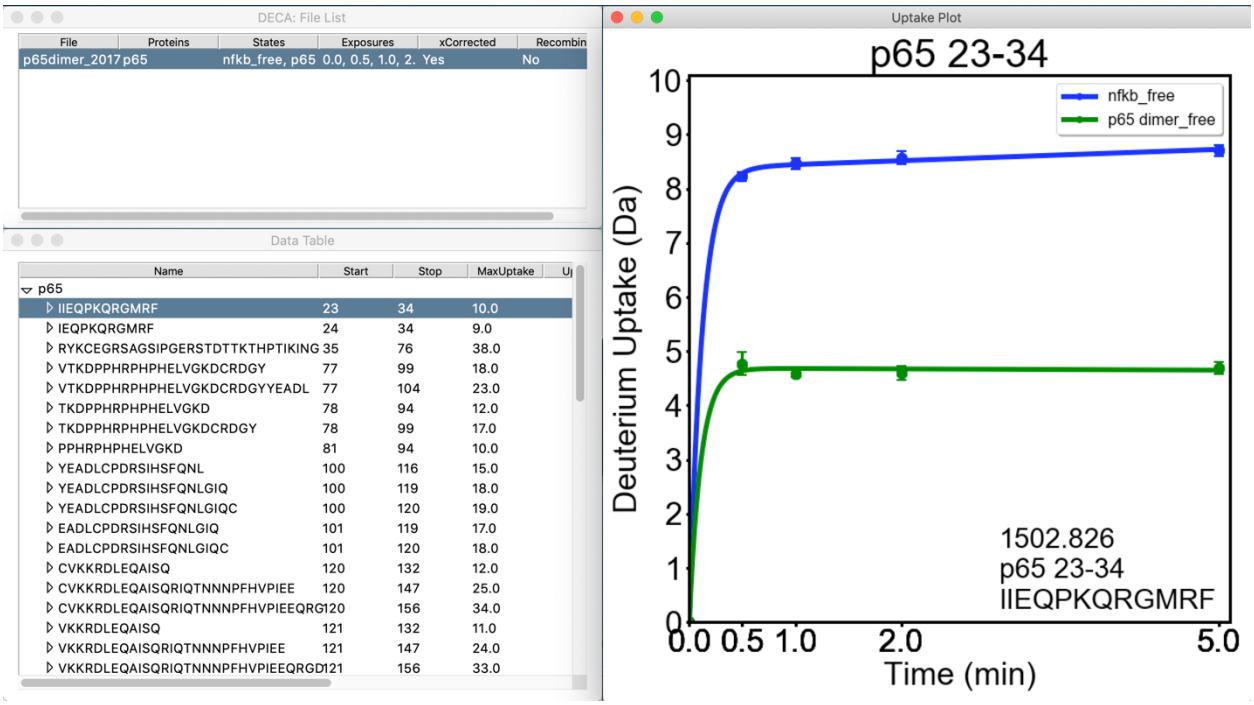

Figure 3.1: Example of the DECA Interface. DECA is a Python program built to run on Windows and MacOS. The GUI simplifies import and analysis of HDX-MS data in multiple formats. Typically, data can be rapidly screened using the Main window (top left) lists open projects, the data table containing the peptide list and their uptake data for the selected project, and the uptake plot is shown for each selected peptide.

3. Back Exchange Correction.

DECA is designed around solving the need to perform back exchange correction on hydrogen-deuterium exchange data before further processing or visualization. Two distinct forms of back exchange may influence deuterium uptake. Back exchange can occur from the quench to the point of sample injection into the mass spectrometer, and this can vary from peptide to peptide because of differences in retention time. The most accurate back exchange correction comes from

the use of a fully deuterated control sample in order to calculate the level of back exchange for each peptide.

$$\text{Correction Factor} = Uptake_{t\infty}/MaxUptake$$

$$\text{Correction} = \frac{Uptake}{CorrectionFactor}$$

The second form of back exchange is a systematic time point-dependent difference resulting from different liquid handling procedures for shorter and longer times of deuterium exposures. When a LEAP robot is used for sample preparation, for timepoints shorter than 2 min the mixing syringes skip a step resulting in a slightly lower back exchange.

$$Correction\ Factor_{t>t_{maxD}} = \frac{(Uptake_{\max D} - Uptake_t)}{Uptake_{\max D}}$$

$$Correction\ Factor_{t\leq t_{maxD}} = 1$$

$$\text{Correction} = \frac{Uptake}{Correction\ Factor_t}$$

Back Exchange Correction is implemented in DECA with two settings: (1) a global correction factor or a fully-deuterated exposure correction factor per peptide, and (2) the Long Exposure Adjustment Patch, a set of exposure correction factors which are applied universally to every peptide in the data set. These two settings may be used separately or in tandem. Back exchange corrected data can be saved and reimported as desired.

Fig. 3.2 demonstrates the necessity for both forms of back exchange correction. A perceptible difference of nearly a half deuteron from the first non-zero time point to the final time point demonstrates clear time point-dependent back ex- change. Following global back exchange correction, LEAP correction values should be obtained from the peptide with the greatest difference between the maximum observed uptake, in this case 3.5 Da at 30s, and the uptake at the final time point (or fully deuterated control). This should correspond to the peptide with the highest

fractional uptake (typically a spiked-in control peptide or a peptide corresponding to a completely disordered region of the protein). The correction factors for timepoints longer than the time point with the highest exchange will be set to adjust the uptake to the highest exchange value. The correction factors for each time point shorter than that with the highest exchange will be set to keep their ratio to the maximum the same. These correction factors are then applied to all other peptides.

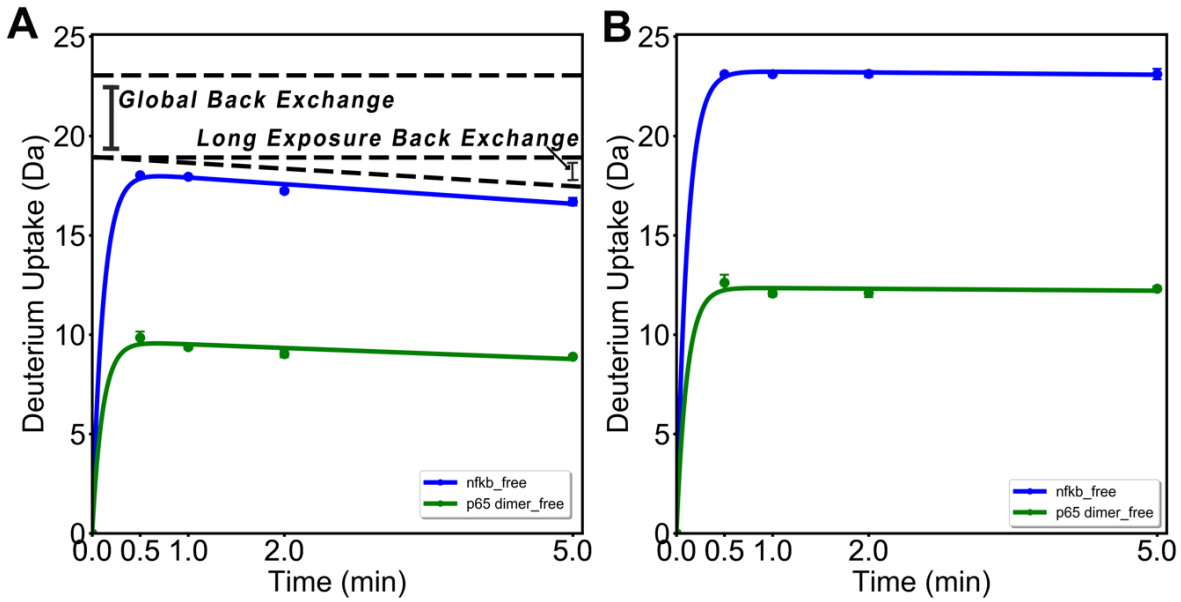

Figure 3.2: Global Back and Exposure Dependent Back Exchange Effects. A. HDX-MS data is affected by two forms of back exchange effects. This peptide should be fully exchanging in the state indicated by the blue line, yet at the longest exposure the uptake is only at ~65% of the maximum uptake. This deficiency is the global back exchange observed. The greater back exchange at higher exposures is the result of systematic error introduced by the LEAP sample handling robot. B. After applying both the back exchange correction and the correction for systematic error introduced by the LEAP robot, the plot on the corrected plot is obtained.

4. Generation of coverage maps from back-exchange corrected data.

At this point, the heat map and uptake plots representing the complete data set can be rapidly generated in DECA. We recommend plotting the coverage map of all the final peptides for which data is analyzed and presented in the manuscript as also recommended recently by the HDXMS community (30). DECA provides additional functions allowing visualization of deuterium uptake, fractional uptake, standard deviation of uptake, or relative standard deviation of uptake, for all real peptides in the data set according to sequence position. The coverage maps reveal the sequence coverage as well as the redundancy or average number of peptides that cover each amino acid. Several colormaps are available to choose from, and the range of the color assignment can be manipulated to highlight features in the data set.

5. Generation of publication-quality deuterium uptake plots from back-exchange corrected data.

A key feature of DECA is the generation of deuterium uptake plots after back exchange correction (Fig. 3.2). Plots generated by DECA allow accurate comparison of deuterium uptake for different regions of a protein sequence, which requires prior correction of back exchange. The deuterium uptake plots generated by DECA show the deuterium-uptake for each peptide, with time on the x axis and deuterium uptake in Da on the y axis with the maximum corresponding to the maximum possible uptake. Publication quality deuterium uptake plots generated in DECA have various features such as automatic plotting of error bars, plot annotation with the peptide sequence and residue numbers, and options for data symbol types and colors. The data are most readily visualized by scrolling through the peptides viewing each uptake plot.

In the data set used as an example here, deuterium uptake into the RelA homodimer was compared with uptake into the RelA-p50 heterodimer. We wanted to compare the uptake into both

RelA and p50 as well. These data were collected months apart and required file merging to present both the p50 uptake and the RelA uptake in a manner where they could be directly compared. This process, which took several weeks to be done manually took only one hour when done with DECA.

6. Comparison of overlapping peptides for increased resolution of deuterium uptake data.

DECA contains a feature called Overlapping Peptide Segmentation (OPS) which computationally increases the sequence resolution of the data. HDX-MS is limited by the size of peptides observable on a mass spectrometer, which usually is in the range of 10–30 amino acids. As a result, HDX data is spread over a large sequence range such that the uptake values may not always localize the exchange events effectively. Because of the use of nonspecific proteases for HDX-MS proteolysis, however, overlapping peptides are often produced. OPS exploits overlapping peptides to assign better-resolved uptake values to the non-overlapping regions in a manner similar to that previously described (31) (Fig. 3.3, Table 3.1). This propagates error, however, and may lead to a mischaracterization of the data, so DECA implements this OPS only once per data set. In other words, the function may not be repeatedly applied to continue generating overlaps from overlap peptides until there are no new peptides being generated. HDX-MS analysis programs such as HDXWorkbench, HDsite and HRHDXMS offer similar features.

7. Visualizing Highest Resolution Deuterium Exchange Data.

Heat maps, butterfly plots and pymol scripts for coloring 3D structures are generated in DECA using the highest resolution data at each residue by assigning the smallest peptide covering each amino acid to that position, including the OPS analysis, if performed (Fig. 3.4). The visualized data is subsequently taken from the assigned peptide at each position. Heat maps display data in the same format as coverage maps but with only a single value per residue instead of showing

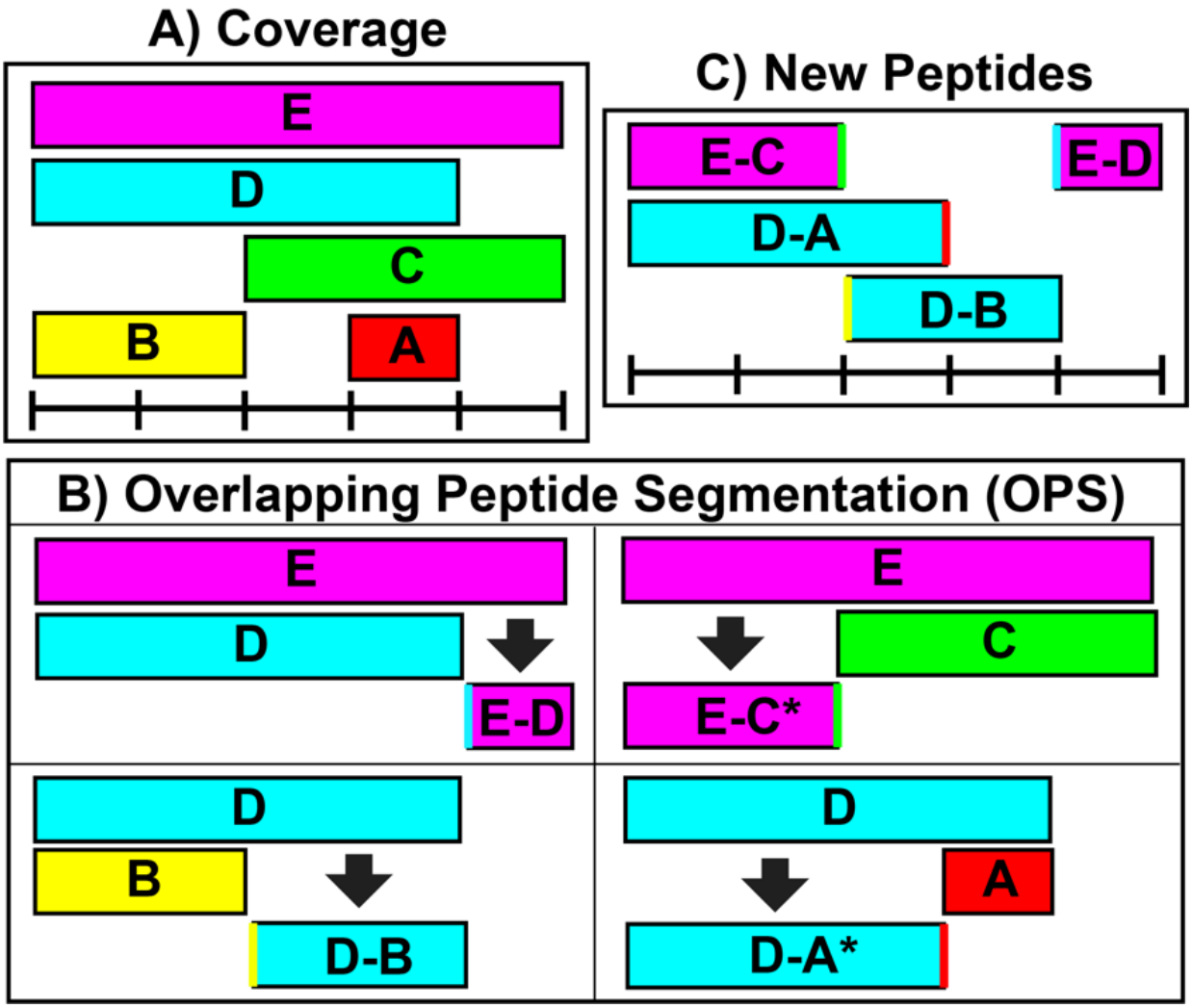

Figure 3.3: Overlapping peptide segmentation (OPS). Because of the partial specificity of "nonspecific" proteases like pepsin, the coverage resolution may be limited to longer peptide lengths. The resolution can be mathematically increased through Overlapping Peptide Seg- mentation (OPS). OPS takes advantage of peptide overlaps such as those in (A), where two peptides share exactly one terminus in com- mon. B, shows how OPS uses such pairs to yield new peptides (C).

Table 3.1: Overlapping Peptide Segmentation (OPS). [1] The last two table entries are result peptides from the OPS analysis. [2] When the new peptide contains the N-terminus, the resulting value may be affected by the need to disregard the N-terminal amino acid which is known to exchange rapidly.

	Range	Uptake Range	Sequence	Uptake
Pep 1	1-16	2-16	GAPVILMSTFYWNQDE	8.78
Pep 2	1-11	2-16	GAPVILMSTFY	6.32
Pep 3	4-16	5-16	VILMSTFYWNQDE	7.89
Pep1-Pep2[1]	12-16	12-16	WNQDE	8.78-6.32 = 2.46
Pep1-Pep3	1-3	2-4[2]	GAP	8.78-7.89 = 0.89

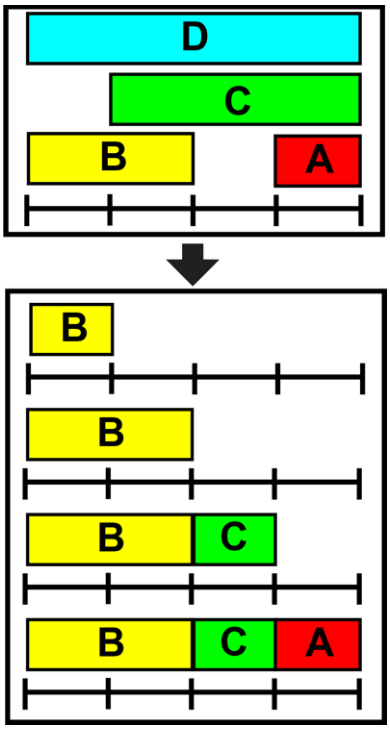

Figure 3.4: Peptide-to-Residue Assignment. In order to generate one-dimensional, residue-resolved data needed for heat maps, butterfly plots, and structure-coloring scripts, residues are assigned the data from the most representative (smallest) peptide covering that region of the protein.

multiple peptides covering each position (Fig. 3.5A). Butterfly plots consist of two line plots for the comparison of two states (Fig. 3.5B). OPS can be used to isolate sites of deuterium uptake and visualize them in these types of plots.

The PyMOL Script function creates a script that can be imported into PyMOL to easily assign data values to the b factor of each residue of a protein structure. This script simultaneously assigns a color gradient and range to visualize the differences by replacing the b factor data column (Fig. 3.5C). Importantly, this feature implemented in DECA considers the back-exchange-corrected uptake amounts which is important for the colors to be comparable across the protein molecule.

8. Statistical Significance.

DECA produces a report on the consistency and accuracy of ion assignments upon import

of a DynamX project file. DECA flags outliers in each ion cluster and calculates the standard deviation of the m/z, retention time, and mobility values for every charge state and every replicate. In the example data sets, assigned ions were within 5% of the mean retention time and mean mobility per peptide. 93–98% of all assigned ions had a search error below 10 ppm. DECA makes such details accessible through the GUI in DECA or exported as a spreadsheet. The raw spectra can be visualized along m/z, retention time, and mobility dimensions in 2D or 3D plots to evaluate peak overlaps and the complexity of the data set.

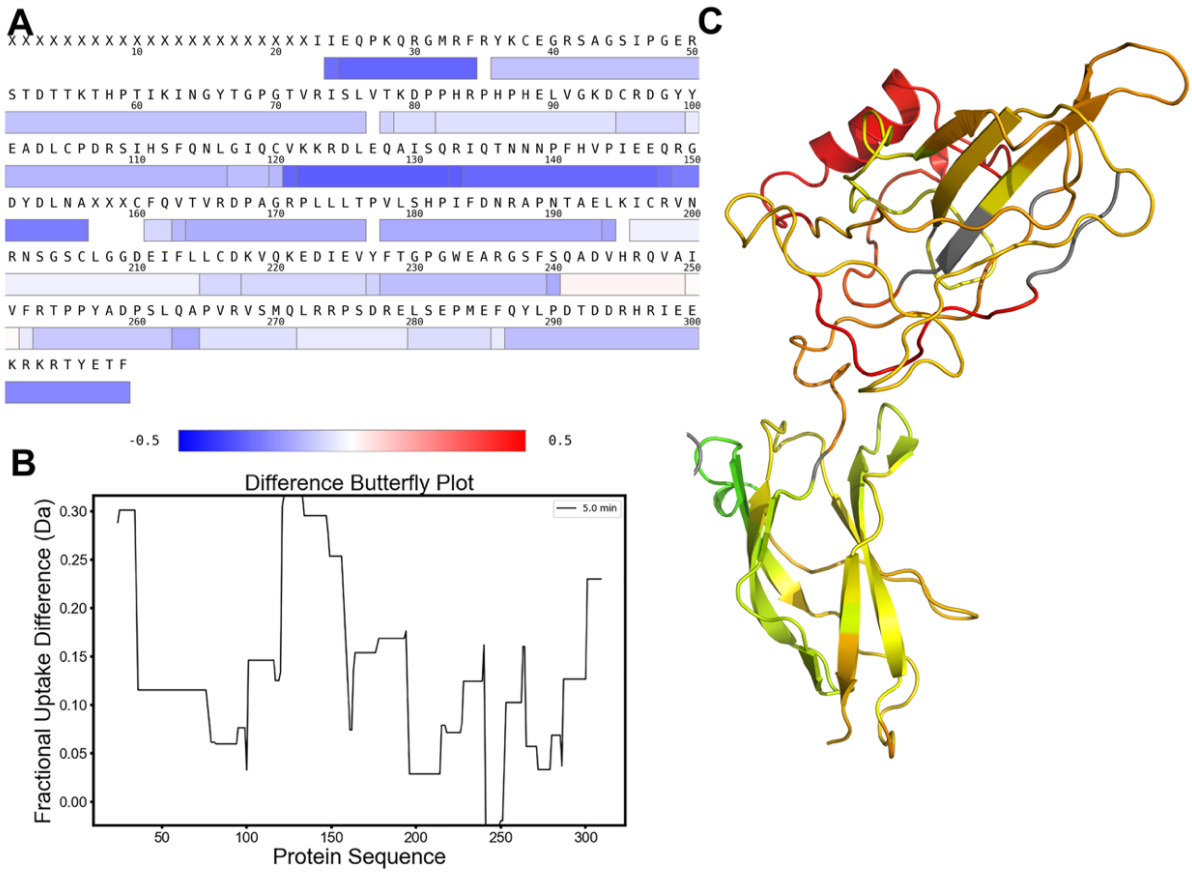

Figure 3.5: Visualizing and Comparing Uptake Data. Several options for the visualization of HDX-MS data are provided in DECA. The full coverage map including uptake information may be generated, or, through Peptide-to-Residue assignment, the data can be displayed in one of the three formats shown above. Heat Maps (A), PyMOL coloring (B), and Butterfly Plots (C) can visualize fractional uptake differences between protein states.

DECA can perform a series of tests to evaluate the statistical significance of differences observed between protein states, for example with or without ligand. For peptides from more than two states, DECA performs a Levene's test to ensure that the variance between the states is low enough to have confident results. This test is followed by a one-way ANOVA analysis to identify the presence of a statistical significance and a Tukey test on every combination of two states. For peptides with only two states, a t test is performed. These statistics are presented in the GUI and can be saved to a spreadsheet. Confidence intervals are plotted for each protein state with coloring matching the lines in the uptake plots. (Fig. 3.6).

9. Data export.

DECA is designed to quickly generate consistent, publication-quality uptake plots that can be edited as desired by saving the plots in ".svg", ".pdf", or ".png" or ".tiff" formats. These plots can be exported individually or all at once.

Data merged or processed by back-exchange correction or peptide recombination may be saved into a comma-separated value (CSV) formatted spreadsheet mimicking an import format. Exported spreadsheets additionally contain information about any back-exchange correction performed on the data set which complies with community recommendations (28).

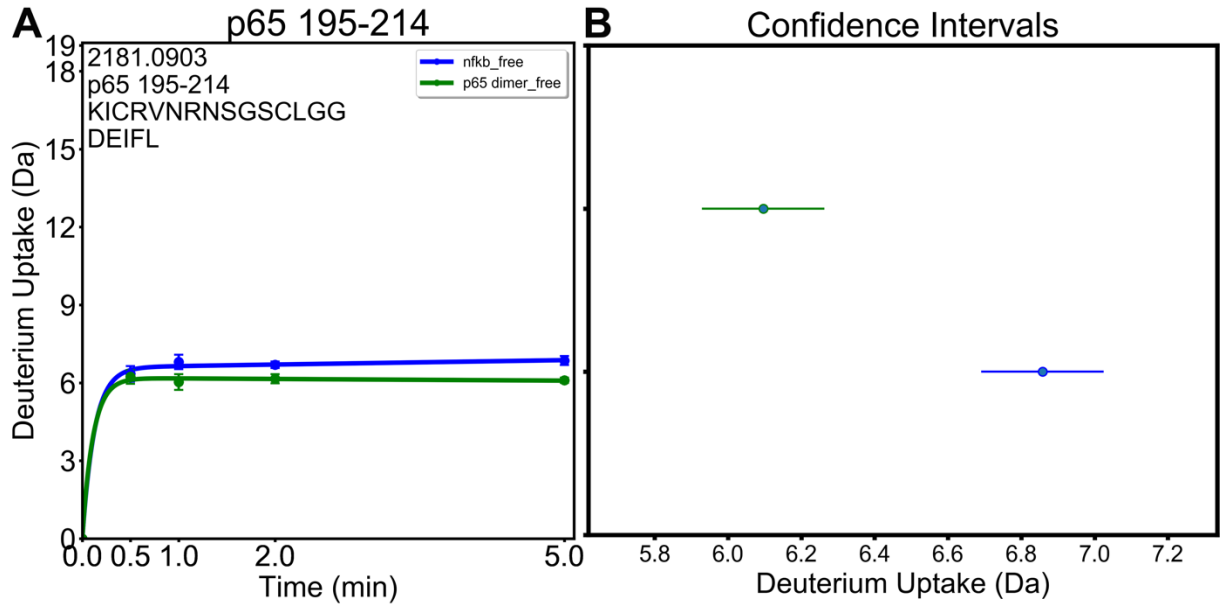

Figure 3.6: Analysis of Variance Determines Significance. Analysis of Variance tests are implemented in DECA to identify statistically significant differences between deuterium uptake in each peptide from different protein states. A, Uptake plot for a peptide from two different states with a deuterium uptake difference of 0.76 Da at 5 min. A t test was used to determine that the difference is significant with a p value of 2.7×10^{-3}. B, Confidence intervals can be calculated and plotted in DECA which illustrate the confident difference between the two states.

D. Discussion

HDX-MS is a rapidly growing technique, yet back-exchange, noisy data, instrumental drift, and poor peptide resolution limit the information content of this type of data. Significant additions to the HDX-MS workflow over the last decade include back-exchange correction, analysis of overlapping peptides, and extraction of data from isotopic envelopes. Waters' DynamX uniquely enables the study of high complexity data sets that otherwise would be limited by spectral overlap through the implementation of ion mobility. Because IMS enables extension to much larger data sets, an automatic downstream data analysis tool is required to not only ascertain the statistical significance underlying large data sets but also to prepare back-exchange corrected uptake plots and PyMOL scripts in a seamless and rapid manner.

Here we present DECA, a feature-rich data analysis back-end, that provides many of these functionalities in an open-source, cross-platform package. Although ion mobility enables the

separation of otherwise-overlapping spectra, automatic processing of large data sets in DynamX can result in occasional incorrect assignments, and manual correction is time-consuming and can sometimes leave the incorrect assignments undetected. DECA performs statistical evaluation on DynamX assignments and implements well established back exchange correction as well as the Long Exposure Adjustment Patch, which corrects for systematic time point-dependent differences we have observed when the LEAP robot is used for sample preparation. DECA enables overlapping peptide analysis that has been previously implemented to take advantage of high redundancy and sequence coverage to generate virtual peptides with higher resolution. DECA can subsequently export the analyzed, filtered, and corrected data to a spreadsheet, or it can produce publication-ready visuals, including 2D and 3D spectra, deuterium uptake plots, coverage maps, heat maps, butterfly plots, and pymol scripts from the back-exchange corrected data.

DECA is developed entirely in Python and compiled into executable binaries compatible with macOS or Windows. DECA can also be run directly from the source code on any computer with Python installed, enabling developers to modify and improve the software. Both the source code and the executable files are available at https://github.com/komiveslab/DECA.

F. Data Availability

The sample data set raw files, DynamX project, and peptides list are available at the MassIVE repository (massive.ucsd.edu) under data set ID: MSV000084200.

G. Acknowledgements

Chapter III, in full, is a reprint that the dissertation author was the principal researcher and author of. The material appears in *Molecular and Cellular Proteomics*. (**Lumpkin, R. J.,** and Komives, E. A. (2019). DECA, a comprehensive, automatic post-processing program for HDX-MS data. *Mol Cell Proteomics, in press)*

H. References

1. Mandell, J. G., Falick, A. M., and Komives, E. A. (1998) Identification of protein-protein interfaces by decreased amide proton solvent accessibility. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 14705–14710

2. Truhlar, S. M., Croy, C. H., Torpey, J. W., Koeppe, J. R., and Komives, E. A. (2006) Solvent accessibility of protein surfaces by amide H/2H exchange MALDI-TOF mass spectrometry. *J. Am. Soc. Mass Spectrom.* **17,** 1490 –1497

3. Rey, M., Sarpe, V., Burns, K. M., Buse, J., Baker, C. A., van Dijk, M., Wordeman, L., Bonvin, A. M., and Schriemer, D. C. (2014) Mass spec studio for integrative structural biology. *Structure* **22,** 1538–1548

4. Ramsey, K. M., Narang, D., and Komives, E. A. (2018) Prediction of the presence of a seventh ankyrin repeat in IkappaBepsilon from homology modeling combined with hydrogen-deuterium exchange mass spectrometry (HDX-MS). *Protein Sci* **27,** 1624–1635

5. Markwick, P., Peacock, R., and Komives, E. (2019) Accurate prediction of amide exchange in the fast limit reveals thrombin allostery. *Biophys. J.* **116,** 49–56

6. Kan, Z. Y., Walters, B. T., Mayne, L., and Englander, S. W. (2013) Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 16438 –16443

7. Zehl, M., Rand, K. D., Jensen, O. N., and Jorgensen, T. J. (2008) Electron transfer dissociation facilitates the measurement of deuterium incorporation into selectively labeled peptides with single residue resolution. *J. Am. Chem. Soc.* **130,** 17453–17459

8. Pascal, B. D., Willis, S., Lauer, J. L., Landgraf, R. R., West, G. M., Marciano, D., Novick, S., Goswami, D., Chalmers, M. J., and Griffin, P. R. (2012) HDX workbench: software for the analysis of H/D exchange MS data. *J. Am. Soc. Mass Spectrom.* **23,** 1512–1521

9. Kan, Z. Y., Mayne, L., Chetty, P. S., and Englander, S. W. (2011) ExMS: data analysis for HX-MS experiments. *J. Am. Soc. Mass Spectrom.* **22,** 1906–1915

10. Zhang, Z., Zhang, A., and Xiao, G. (2012) Improved protein hydrogen/deuterium exchange mass spectrometry platform with fully automated data processing. *Anal. Chem.* **84,** 4942–4949

11. Guttman, M., and Lee, K. K. (2016) Isotope labeling of biomolecules: structural analysis of viruses by HDX-MS. *Methods Enzymol.* **566,** 405–426

12. Henderson, S. C., Valentine, S. J., Counterman, A. E., and Clemmer, D. E. (1999) ESI/ion trap/ion mobility/time-of-flight mass spectrometry for rapid and sensitive analysis of biomolecular mixtures. *Anal. Chem.* **71,** 291–301

13. Kanu, A. B., Dwivedi, P., Tam, M., Matz, L., and Hill, H. H., Jr. (2008) Ion mobility-mass spectrometry. *J. Mass Spectrom.* **43,** 1–22

14. Iacob, R. E., Murphy, J. P., 3rd, and Engen, J. R. (2008) Ion mobility adds an additional dimension to mass spectrometric analysis of solution-phase hydrogen/deuterium exchange. *Rapid Commun. Mass Spectrom.* **22,** 2898–2904

15. Rand, K. D., Pringle, S. D., Murphy, J. P., Fadgen, K. E., Brown, J., and Engen, J. R. (2009) Gas-phase hydrogen/deuterium exchange in a travelling wave ion guide for the examination of protein conformations. *Anal. Chem.* **81,** 10019–10028

16. Weis, D. D., Engen, J. R., and Kass, I. J. (2006) Semi-automated data processing of hydrogen exchange mass spectra using HX-Express. *J. Am. Soc. Mass Spectrom.* **17,** 1700–1703

17. Guttman, M., Weis, D. D., Engen, J. R., and Lee, K. K. (2013) Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra. *J. Am. Soc. Mass Spectrom.* **24,** 1906–1912

18. Slysz, G. W., Baker, C. A., Bozsa, B. M., Dang, A., Percy, A. J., Bennett, M., and Schriemer, D. C. (2009) Hydra: software for tailored processing of H/D exchange data from MS or tandem MS analyses. *BMC Bioinformatics* **10,** 162

19. Lou, X., Kirchner, M., Renard, B. Y., Kothe, U., Boppel, S., Graf, C., Lee, C. T., Steen, J. A., Steen, H., Mayer, M. P., and Hamprecht, F. A. (2010) Deuteration distribution estimation with improved sequence coverage for HX/MS experiments. *Bioinformatics* **26,** 1535–1541

20. Liu, S., Liu, L., Uzuner, U., Zhou, X., Gu, M., Shi, W., Zhang, Y., Dai, S. Y., and Yuan, J. S. (2011) HDX-analyzer: a novel package for statistical analysis of protein structure dynamics. *BMC Bioinformatics.* **12,** S43

21. Hourdel, V., Volant, S., O'Brien, D. P., Chenal, A., Chamot-Rooke, J., Dillies, M. A., and Brier, S. (2016) MEMHDX: an interactive tool to expedite the statistical validation and visualization of large HDX-MS data sets. *Bioinformatics* **32,** 3413–3419

22. Lau, A. M. C., Ahdash, Z., Martens, C., and Politis, A. (2019) Deuteros: software for rapid analysis and visualization of data from differential hydrogen deuterium exchange-mass spectrometry. *Bioinformatics* **35,** 3171–3173

23. Jones, E., Oliphant, T., and Peterson, P. (2001) SciPy: open source scientific tools for python. https://www.scipy.org/

24. Hunter, J. D. (2007) Matplotlib: A 2D graphics environment. *Comput. Sci. Engineering* **9,** 90–95

25. Goloborodko, A. A., Levitsky, L. I., Ivanov, M. V., and Gorshkov, M. V. (2013) Pyteomics–a Python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass Spectrom.* **24,** 301–304

26. Cortesi, D. (2018) PyInstaller. http://www.pyinstaller.org/

27. Narang, D., Chen, W., Ricci, C. G., and Komives, E. A. (2018) RelA-Containing NFkappaB Dimers Have Strikingly Different DNA-Binding Cavities in the Absence of DNA. *J. Mol. Biol.* **430,** 1510–1520

28. Wales, T. E., Fadgen, K. E., Gerhardt, G. C., and Engen, J. R. (2008) High-Speed and High-Resolution UPLC Separation at Zero Degrees Celsius. *Anal. Chem.* **80,** 6815–6820

29. Ramsey KM, Dembinski HE, Chen W, Ricci CG, Komives, EA (2017) DNA and IKB Both Induce Long-Range Conformational Changes in NFKB. *J. Mol. Biol.* **429,** 999–1008

30. Masson, G. R., Burke, J. E., Ahn, N. G., Anand, G. S., Borchers, C., Brier, S., Bou-Assaf, G. M., Engen, J. R., Englander, S. W., Faber, J., Garlish, R., Griffin, P. R., Gross, M. L., Guttman, M., Hamuro, Y., Heck, A. J. R., Houde, D., Iacob, R. E., Jørgensen, T. J. D., Kaltashov, I. A., Klinman, J. P., Konermann, L., Man, P., Mayne, L., Pascal, B. D., Reichmann, D., Skehel, M., Snijder, J., Strutzenberg, T. S., Underbakke, E. S., Wagner, C., Wales, T. E., Walters, B. T., Weis, D. D., Wilson, D. J., Wintrode, P. L., Zhang, Z., Zheng, J., Schriemer, D. C., and Rand, K. D. (2019) Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16,** 595–602

31. Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W., and Woods, V. L. J. (2005) Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 15477–15482

Chapter IV:

Structure and dynamics of the ASB9 CUL-RING E3 Ligase

A. Introduction

Ubiquitylation is a post-translational protein modification in which the 8.5 kDa protein Ubiquitin (Ub) is covalently attached to a substrate protein following activation by the Ubiquitin-activating enzyme (E1) and transfer to a Ubiquitin-conjugating enzyme (E2). Ubiquitin Ligases (E3) facilitate the highly-specific covalent attachment of activated Ubiquitin (Ub) to substrate proteins through an isopeptide bond on an exposed lysine residue. Ubiquitylation leads to two general classes of modification, mono- and poly-ubiquitylation, wherein a molecule of ubiquitin is linked to one of the seven lysines in another molecule of ubiquitin. K48 polyubiquitin chains are the most studied form of polyubiquitylation and are responsible for targeting substrate proteins for proteasomal degradation (1).

The multi-subunit Cullin (CUL)-RING ligase (CRL) is the largest family of E3 ligases, and it is responsible for up to 20% of degradation through the proteasome (2). CRLs share a common structure, composed of a substrate receptor(s), a CUL, and a RING-box (RBX) protein. The RBX subunit is responsible for recruiting Ubiquitin-charged E2 enzymes to the ligase (1). Human proteins CUL2 and CUL5 bind substrate receptors through the adaptor proteins Elongin B (ELOB) and Elongin C (ELOC). The vHL-box and SOCS-box motifs in CUL2 and CUL5 cause them to recruit vHL-like or SOCS-box-containing substrate receptors, respectively. Several protein families contain SOCS-box domains, including the canonical Suppressor of Cytokine Signaling (SOCS), Single Stranded Binding (SSB), WD40 SOCS-box (WSB), and Ankyrin Repeat and SOCS-box (ASB) families (3).

The eighteen ASB proteins comprise the largest family of SOCS-box domain-containing E3 ligase substrate receptors. Each ASB protein contains an ankyrin repeat domain and a SOCS-box domain with CUL5-box and BC box motifs, and they specifically associate with CUL5

through the Elongin B and C (EB/C) adapter proteins (4-6). A few ASB proteins have known substrates that are marked for degradation: ASB2-CRL ubiquitylates Filamins A and B, ASB3 and 4 -CRLs ubiquitylate tumor necrosis factor receptor II, and ASB11-CRL ubiquitylates Notch ligand DeltaA (7).

Creatine Kinase brain-type (CKB) and mitochondrial type (MtCK) have been shown to be degraded in vivo in an ASB9-dependent manner (8, 9). ASB9 binds a dimer of CKB with sub-nanomolar affinity (10) with residues 19-24 being especially important for high affinity binding. (10, 11). Structures of ASB9 lacking the SOCS box domain, (12) and including the SOCS-box bound to ELOB and ELOC have been determined (7). In addition, structures of SOCS2-ELOB/C bound to the N-terminal half of CUL5 (PDB 4JGH) (13) and the C-terminal half of CUL5 bound to RBX1 (3DPL) (14) have been determined. While some details have been shown in a fragmented manner, a full mechanistic understanding of the ASB9 CRL assembly, substrate recognition, and dynamics remain unknown.

We present here a full model of the substrate-bound ASB9-ELOB/C-CUL5-RBX2 E3-ligase complex based on cryo-electron microscopy (cryo-EM) structures of CKB-ASB9-ELOB/C and of CUL5. By combining cryo-EM, HDX, and homology modeling, we are able to define the general architecture of the complex and propose conformational dynamics that underlie substrate recognition and ubiquitin loading. HDX-MS data confirms binding sites between CKB-ASB9 and ELOB/ELOC-CUL5. Most importantly, the HDX-MS reveals a pathway of allostery that connects substrate binding to conformational changes near important modules in CUL5 that control ligase activity, including the site of CUL5 neddylation and RBX2-E2 binding. This allosteric pathway suggests that substrate binding itself serves a mechanistic role in the activity of CUL-RING ligases.

B. Materials and Methods

1. Expression vectors.

Human ASB9-1 in pNIC-CTHF was obtained from the Structural Genomics Consortium and subcloned into pHis8 (KanR) with an N-terminal 8xHis tag and Human CKB was subcloned into pET11a (AmpR) as previously described (10). Human ELOB (full-length) and ELOC (17-112) were obtained in pACYC (CamR) (Structural Genomics Consortium). Human CUL5 and mouse RBX2 were obtained in pRSFDuet (KanR) with an N-terminal His tag, TEV cleavage site, and GB1 tag on CUL5 (gift from Nevan Krogan). 6xHis-GB1-TEV-CUL5 was subcloned into pET28a (KanR) and RBX2 was subcloned into pET11a (AmpR).

2. Protein Expression.

ASB9 was co-expressed with ELOB/ELOC and/or CKB using sequential transformation with the vectors described above into BL21(DE3) E. coli cells for ASB9-CKB, ASB9-ELOB-ELOC, and ASB9-CKB-ELOB-ELOC co-expression. Vectors containing ELOB/ELOC, CKB, or both were transformed into competent BL21(DE3) cells after which those cells were made competent again. ELOB/ELOC in pACYC was selected by Chloramphenicol (CAM) resistance, and CKB in pET11a was selected by Ampicillin (AMP) resistance. The vector for expression of ASB9 was transformed into ELOB/ELOC + CKB – containing BL21(DE3) cells and plated on a Kanamycin (KAN)-CAM-AMP LB agar plate. The pET28a vector for expression of CUL5 and the pET11a vector for expression of RBX2 were co-transformed into ELOB/ELOC - containing BL21(DE3) cells and plated on a KAN-CAM-AMP LB agar plate for co-expression of CUL5/RBX2/ELOB/ELOC.

All proteins were expressed as follows. A 5 mL M9-ZN (1.5x M9 salts, NZ-Amine media, 0.8% Dextrose, 1 mM MgSO4, 0.2 mM CaCl2) overnight culture was inoculated with a single

colony from the plate. A 20 mL M9-ZN starter culture was inoculated with 2 mL of the overnight culture and grown for 3 hours at 37°C. The 1 L M9-ZN growth culture was inoculated with the entire 20 mL starter culture and grown until OD600 = 0.8. After placing the cultures on ice for 15 min, protein expression was induced by addition of IPTG to a final concentration of 0.5 mM, and the cultures were transferred to an 18°C incubator for 16-18 hours. Because RBX2 contains a Zn binding domain, the RBX2-containing cultures were brought to 200 μM Zn by addition of a 1 M solution of ZnCl2 just prior to induction.

3. Protein Purification.

Cells from 1 L of culture were pelleted by centrifugation at 5000 rpm for 10 minutes, then re-suspended in 40 mL Resuspension Buffer (50 mM Tris-HCL pH 8.0, 100 mM NaCl, 10 mM Imidazole pH 8.0, 2 mM β-mercaptoethanol, 5% Glycerol) with Sigma Protease Inhibitor Cocktail (P2714) and 5 mM PMSF. Cells were lysed on ice by sonication with ten 30 seconds pulses with 45 seconds cool-down between each pulse. The lysate was clarified by centrifugation at 13000 rpm for 45 minutes. The clarified lysate was incubated with 2 mL Ni-NTA (in Resuspension Buffer) for 2 hours at 4°C with rocking. Ni-NTA beads were pelleted by centrifugation at 700 x g for 5 minutes. The supernatant was discarded, and the beads were washed with 10 mL Wash Buffer (50 mM Tris-HCL pH 8.0, 100 mM NaCl, 25 mM Imidazole pH 8.0, 2 mM β-mercaptoethanol, 5% Glycerol) for 30 minutes at 4°C. Ni-NTA beads were again pelleted by centrifugation at 700 x g for 5 minutes. The supernatant was discarded, and the beads were washed with 10 mL Elution Buffer (50 mM Tris-HCL pH 8.0, 100 mM NaCl, 250 mM Imidazole pH 8.0, 2 mM β-mercaptoethanol, 5% Glycerol) for 30 minutes at 4°C. Ni-NTA beads were again pelleted by centrifugation at 700 x g for 5 minutes. The supernatant was transferred to a 12-14 kDa Dialysis bag and dialyzed overnight in Dialysis Buffer (20 mM Tris-HCL pH 8.0, 100 mM NaCl, 5%

Glycerol, 1 mM DTT). Samples were concentrated to 2 mL and purified using size-exclusion chromatography over a Superdex S200 16 x 600 column in Dialysis Buffer. Peak fractions were combined and concentrated to 5 μM for analysis by HDX-MS or 20 μM for structural determination by Cryo-EM.

Sample compositions were identified and characterized according to the presence of the desired proteins as assessed by size exclusion chromatography, SDS-PAGE (10-15% Acrylamide gels), and nanospray LC MS/MS on a Lumos mass spectrometer after trypsin digestion.

To prepare CKB-ASB9-ELOB-ELOC or CUL5-RBX2 for structure determination, all four proteins were co-expressed and purified as described above. Following size-exclusion by Superdex S200 16 x 600, the samples were concentrated to 500 μL and re-purified on a S200 Increase 10/300 column. Collected fractions were concentrated to 20μM.

CKB-ASB9, ASB9-ELOB-ELOC, CKB-ASB9-ELOB-ELOC, ASB9-ELOB-ELOC-CUL5-RBX2, CKB-ASB9-ELOB-ELOC-CUL5-RBX2, and ELOB-ELOC-CUL5-RBX2, complexes were individually prepared for analysis by HDXMS. CKB-ASB, ASB9-ELOB-ELOC, CKB-ASB9-ELOB-ELOC, and ELOB-ELOC-CUL5-RBX2 complexes were expressed and purified from multiple-expression systems as described above. ASB9-ELOB-ELOC-CUL5-RBX2 and CKB-ASB9-ELOB-ELOC-CUL5-RBX2 were formed by combining the lysates of ASB9-ELOB-ELOC or CKB-ASB9-ELOB-ELOC with ELOB-ELOC-CUL5-RBX2 and purifying the combined proteins as described above.

4. Hydrogen Deuterium Exchange.

HDXMS experiments were conducted using a Waters nanoACQUITY UPLC system equipped with H/DX technology and a LEAP H/D-X PAL liquid handling system as previously described (23). The H¬2O buffer, was composed of 20 mM Tris-HCL pH 8.0, 100 mM NaCl, 5%

Glycerol, 1 mM DTT, and 0.5 mM EDTA, matching the Size-Exclusion Buffer used in the final stage of purification for each protein sample. This Buffer was lyophilized and resuspended in D2O (D2O Buffer). Using the LEAP robot, a 4 μL portion of a 5 μM protein sample was incubated for 5 min at 25 °C and then mixed with 56 μL of H2O Buffer as a control or D2O Buffer for deuteration times of 15 s, 30 s, 45 s, 60 s, or 120 s. The reaction was quenched with 60 μL of Quench Buffer (3 M guanidine, 0.1% formic acid, pH 2.66) at 0°C. A portion of the quenched sample (50 μL) was injected into the sample loop and subsequently digested on an in-line pepsin column (Immobilized Pepsin, Pierce Inc.) at 15°C. The digested samples were transferred to a BEH C18 Vanguard pre-column at 0°C and subsequently separated by analytical chromatography (Acquity UPLC BEH C18, 1.7 μM, 1.0 × 50 mm, Waters Corporation) using a 7-85% acetonitrile in 0.1% formic acid over 7.5 min also at 0°C, and electrosprayed into the Waters SYNAPT G2Si quadrupole time-of-flight mass spectrometer. The mass spectrometer was set to collect data in the Mobility, ESI+ mode; mass acquisition range of 200–2,000 (m/z); scan time 0.4 s. Continuous lock mass correction was accomplished with infusion of leu-enkephalin (m/z = 556.277) every 30 s (mass accuracy of 1 ppm for calibration standard). For peptide identification, the mass spectrometer was set to collect data in MSE, ESI+ mode instead.

The peptides were identified from triplicate MSE analyses of 5 μM CKB-ASB9, ASB9-ELOB-ELOC, and ELOB-ELOC-CUL5-RBX2 samples with data analysis using PLGS 2.5 (Waters Corporation). Peptide masses were identified using a minimum number of 250 ion counts for low energy peptides and 50 ion counts for their fragment ions. The peptides identified in PLGS were then analyzed in DynamX 3.0 (Waters Corporation) using a cut-off score of 6.5 and the deuterium uptake was corrected for back-exchange (typically 25%) as previously described (24). The relative deuterium uptake for each peptide was calculated by comparing the centroids of the

mass envelopes of the deuterated samples vs. the undeuterated controls following previously published methods (25).

5. Cryo-EM sample preparation and data collection.

The CKB-ASB9-ELOB-ELOC complex was purified as described above. The final protein buffer was 20 mM Tris, pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 1 mM DTT. To increase the angular distribution of particles in our dataset, we supplemented protein samples with the detergent n-Octyl-β-D-Glucopyranoside (β-OG). All samples were prepared using UltraAuFoil R 1.2/1.3 300 mesh gold grids (Quantifoil GmbH) that were glow discharged for 30 seconds at 20 mAmp. Samples (4 μL) were applied to grids within 10 minutes of charging and blotting. Plunge freezing was performed using a Vitrobot Mark IV robot (Thermo Fisher). The robot was set to 100% humidity, 4°C, blot force 20, and blot time 4 s. Multiple datasets were collected on grids prepared with varying protein (5-20 μM) and detergent (0.01-0.1% w/v β-OG) concentrations.

Samples were imaged using a Talos Arctica TEM (Thermo Fisher) operating at 200 keV in nano probe mode at a magnification of 36,000x, with a final pixel size of 1.16 Å. Parallel illumination of the microscope was performed according to (26). Dose fractionated movies were collected using a K2 Summit Direct Electron Detector (Thermo Fisher) operated in counting mode, with a defocus range of -0.6 μm and -2.5 μm, an exposure rate of ~6.7 e-/pixel/s, 200 ms frames, and a total exposure of ~50 e-/Å2. New camera gain references were collected before each dataset and the hardware dark reference was updated daily. Holes containing thin ice were manually selected and movies were collected automatically using the Leginon software suite (27). Motion correction and gain correction were performed using MotionCor2 (28), CTF estimation was performed on non dose-weighted micrographs using CTFFIND4 (29), and particle picking was performed using DogPicker.py (30); all data processing was performed on-the-fly using the

Appion software suite (26). Particles were extracted from dose-weighted, aligned micrographs and analyzed using Relion-3 (31) for 2D classification and cryoSPARC (32) for 3D classification and refinement. Resolution values are according to the 0.143 gold standard Fourier shell correlation (GSFSC) method (33).

All datasets were collected using the same exposure rate, total exposure, frame rate, and magnification, and then merged and processed as a single dataset. In total, 1983 movies were collected and 1735 remained after removing micrographs with crystalline ice or ethane contamination. An initial dataset of 1,309,964 particles was subjected to multiple rounds of 2D classification in Relion-3, with particles re-centered and re-extracted after each round. This yielded a final dataset of 368,831 'clean' particles. An ab initio 3D model was determined in cryoSPARC v2, asking for 3 initial models. The best model had clear secondary structure and was composed of 285,155 particles. A final refinement was performed using Non-Uniform Refinement in cryoSPARC v2, yielding a final reconstruction at a resolution of ~4.1 Å. Local resolution estimation was performed in cryoSPARC v2, showing a range from 3.5 to 8 Å resolution, with the bulk of the model below 4.0 Å. A sharpened map was used for model building in Rosetta, see Model building and validation, below.

Our cryo-EM map showed clear density for a CKB dimer (3DRB.pdb) and the ASB9-ELOB-ELOC heterotrimer (3ZKJ.pdb). To make an initial model, both crystal structures were docked into the cryo-EM map using Chimera (34). The CKB dimer seemed to be in the 'open' conformation, corresponding to chain A of 3DRB. An initial model was made using chain A of PDB 3DRB to make an "open" CKB dimer which was fitted into the CKB density. Chains D,E,F of the ASB9-ELOB-ELOC structure (PDB 3ZKJ) fit into our cryo-EM map of this part of the complex without any major conformational changes, except the N-terminal region of ASB9 needed

to be modeled because residues 1-34 of ASB9 were not defined in the 3ZKJ crystal structure. The CKB-ASB9-ELOB-ELOC model was then improved using RosettaCM and Rosetta FastRelax integrated into a cloud-based cryo-EM pipeline (35-37). First, our starting model was used to generate ~1000 models using RosettaCM. Then the models were sorted by Rosetta energy score, and the top 10% were then scored using MolProbity. The ten models with the best MolProbity score were selected as our top models. To build the N-terminal loop of ASB9 for which clear density was observed, a small region of ASB9 and CKB were extracted and used as an initial model for RosettaCM. No model was given for the ASB9 N-terminus (residues 25-35), and Rosetta CM was allowed to build this region de novo. ~500 models were generated. This initial building yielded a large range of realistic and unrealistic structures, and 10 models with good fit to the density were selected as starting models for another round of RosettaCM. ~1000 models were generated in this analysis, and the top ten models as judged by Rosetta energy score and MolProbity score had good agreement (per-residue RMSD < 2 Å). The top ten models from the two analyses were used to make ten chimeric models that were used as inputs for Rosetta FastRelax, which refines side chain placement and per-residue B-factors. This refinement improved the MolProbity and Clash Score for each model. A final model consisting of 10 models was deposited in the PDB.

Table 4.1: Cryo-EM data collection, refinement, and validation statistics

	Asb9-CKB-ELOB-ELOC complex
	PDB ID:
	EMDB ID:
Data Collection	
Microscope	Talos Arctica
Camera	K2 summit
Camera Mode	Counting
Magnification	36,000
Voltage (kV)	200
Total electron exposure (e-/Å^2)	50
Exposure rate (e-/pixel/sec)	6.7
Defocus Range (um)	0.6-2.5
Pixel Size (Å/pixel)	1.16
Micrographs collected (no.)	1983
Micrographs used (no.)	1735
Reconstruction	
3D Processing Package	cryoSPARC v2
Total Extracted picks (no.)	1,309,964
Refined particles (no.)	368,831
Final Particles (no.)	285,105
Symmetry	C1
Resolution (global) (Å)	4.1
FSC 0.143 (unmasked/masked)	4.9/4.1

Table 4.1: Cryo-EM data collection, refinement, and validation statistics (continued)

Local resolution range (Å)	3.5-8.0
Map sharpening *B*-factor	-232
Refinement	
Model refinement package	Rosetta
Number of models	10
Nonhydrogen atoms	9329 (per model)
Protein residues	1185 (per model)
B factors (Å^2)	
Protein residues	87.42
R.m.s. deviations	
Bond Lengths (Å)	0.0139
Bond angles (°)	1.35
Validation	
MolProbity score	1.17 (99th percentile)
Clashscore	1.62 (99th percentile)
Poor rotamers (%)	0
CaBLAM outliers (%)	0
Ramachandran plot	
Favored (%)	96.14
Allowed (%)	3.43
Disallowed (%)	0.43
EMRinger score	1.4
Map CC (*CCmask*)	0.752

6. Homology Modeling.

A model of the full ASB9-CRL was prepared by superimposing homologous domains of known structures to position each component in PyMOL. The CKB-ASB9-ELOB-ELOC structure was determined as described above. ELOB-ELOC in SOCS2-ELOB-ELOC-CUL5NTD(4JGH) were aligned with CKB-ASB9-ELOB-ELOC in PyMOL, and SOCS2-ELOB-ELOC from 4JGH were removed. To orient the CUL5CTD relative to CUL5CTD, CUL2 from CUL2-RBX1-ELOB-ELOC-VHL (5N4W) was aligned to residues 306-387 on CUL5NTD, and CUL5CTD from CUL5CTD-RBX1 (3DPL) was aligned to residues 383-424 on CUL2. After orienting the two halves of CUL5, CUL2-RBX1-ELOB-ELOC-VHL was removed from the model. Residues 208-238 of RNF4 in RNF4-UBE2D1-Ub (4AP4) were aligned to RBX1 to position UBE2D1-Ub, and RNF4 was removed. CDC34 (6NYO) was aligned with E2D1, and E2D1 was removed to yield a model containing CDC34.

Homology modeling with MODELLER 9.23 was performed for each of these superimposed complexes to refine loops, fill sequence gaps, correct mutations, and to model RBX2 by sequence homology to RBX1. Sequences extracted from the PDBs were aligned to the full-length Human sequences for each protein using the pairwise alignment tool EMBOSS Water from EMBL-EBI. An alignment file for MODELLER (38) was generated using the sequence alignments. MODELLER 9.23 was run using the default automodel class, to generate 10 models. The model with the lowest MOLPDF score was selected as the final model.

C. Results

1. The Structure of the CKB-ASB9-ELOB-ELOC Complex by Cryo-EM.

To study the assembly of the full CUL5 complex, we prepared a number of different sub-complexes of the E3 ligase complex containing CKB-ASB9-ELOB-ELOC-CUL5-RBX2 (Fig

4.1). A 4.1 Å structure of the substrate-receptor complex containing CKB, ASB9, ELOB and ELOC was determined through cryo-EM. A model was refined from this map using crystal structures of the CKB homodimer (PDB 3DRB) and the ASB9-ELOB-ELOC heterotrimer (PDB 3ZKJ) (Fig 4.2). CKB binds as a dimer to the N-terminal ankyrin repeat of ASB9 as previously predicted from HDX-MS data, kinetic data, and modeling (10, 11). The structure of the CKB homodimer (PDB 3DRB) has one monomer in an "open" conformation and the other in a "closed" conformation. The model that best fit the cryo-EM density was that with both CKB monomers in the "open" conformation.

ASB9 from ASB9-ELOB/C (3ZNG) aligned closely with the Cryo-EM density for the ASB9 subunit. The structure of ASB9 residues 25-34 was resolved to reveal a partial fold of a helix and a turn inside the active site of one monomer of the CKB dimer (Fig 4.2). Residues 19-34 of the N-terminus of ASB9 were present in the structure of the apo-ASB9-2 splice variant (3D9H) (12) and residues 23-34 were modeled for the docked structure of CKB and ASB9 (11), but neither of these predictions matched the Cryo-EM density. The partial helix positions D32 in the active site of one CKB monomer (Fig 4.2 inset), consistent with previous biophysical experiments demonstrating the critical role of this residue in the interaction (11).

2. Structure of full-length CUL5-RBX2.

To date, no full-length structure of CUL5 is available, however a structure of SOCS2-ELOB/C-CUL5NTD (4JGH) and of CUL5CTD-RBX1 (3DPL) are available. We determined a ~6 Å cryo-EM structure for the full-length CUL5-RBX2 complex. A severe preferred orientation in the cryo-EM grid limited the resolution of our map. However, both CUL5 crystal structures could be unambiguously docked into the cryo-EM density. Using these crystal structures as a starting point for modeling, we built a complete model of the CUL5-RBX2 complex (Fig 4.3).

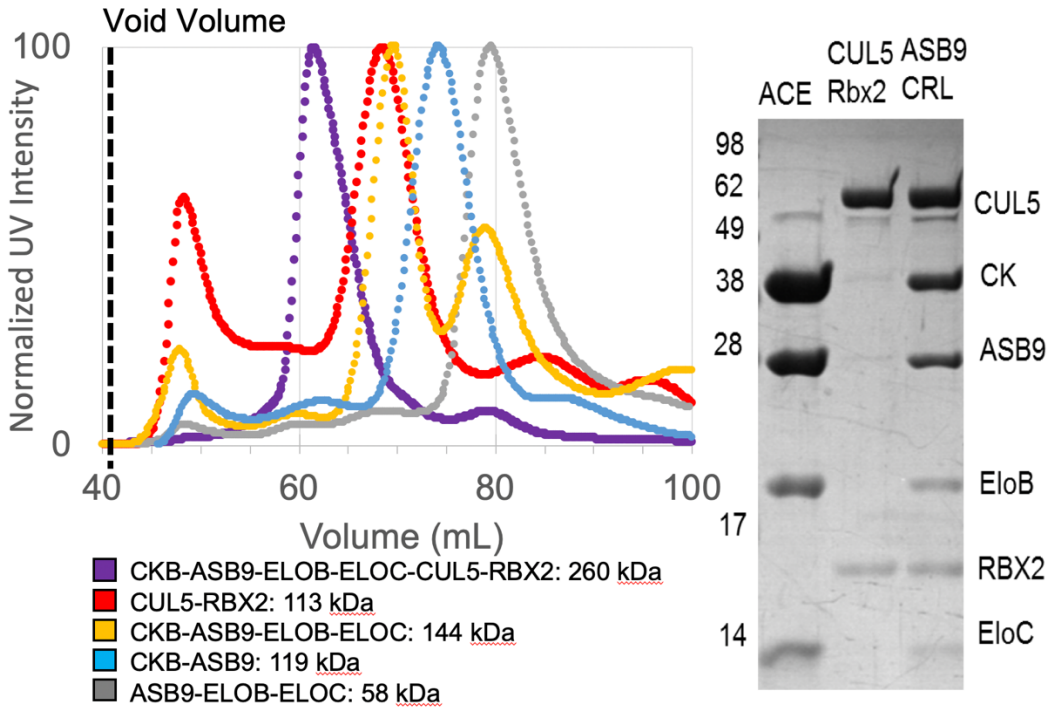

Figure 4.1: **Size exclusion chromatography** was used to purify the various complexes studied in this work. A) Elution profiles of the various complexes B) SDS PAGE showing the purity of the complexes and the presence of the subunits present in each.

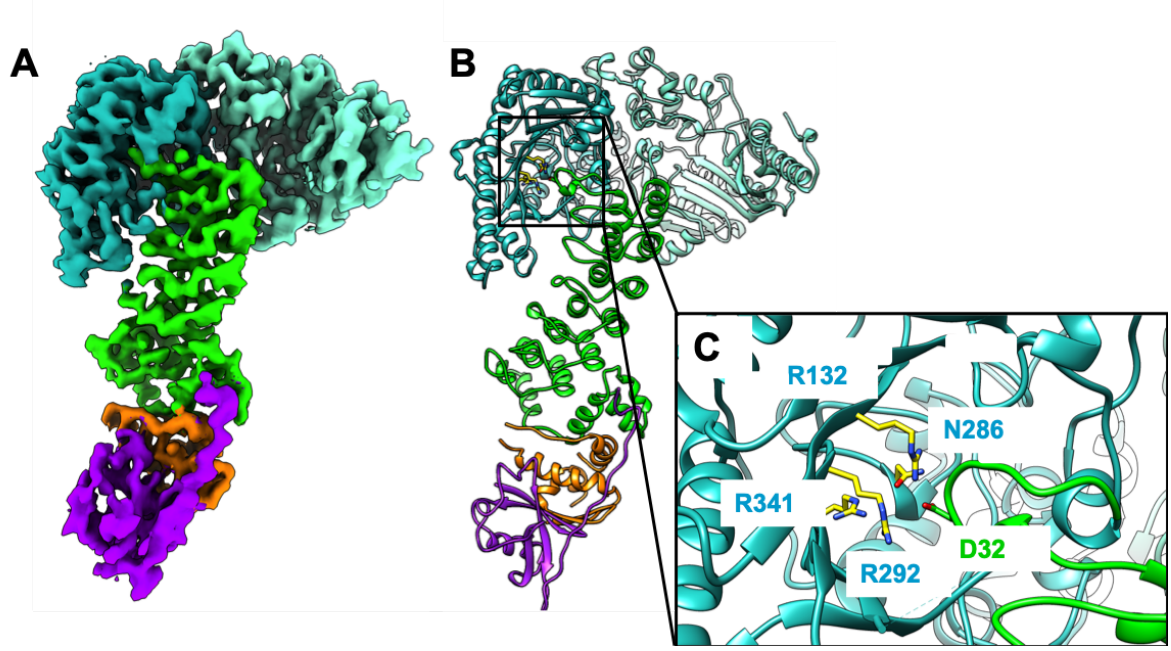

Figure 4.2: Structure of the CKB substrate bound to the ASB9 substrate receptor and adapter proteins ELOB/ELOC. A) The Cryo-EM density of CKB-ASB9-ELOB/C was determined to 4.1 Å. B) Model built by fitting the crystal structures for CKB dimer (PDB 3DRB) (light sea green, aquamarine) and for ASB9 (green) and ELOB/C (orange/purple), all from PDB 3ZKJ, and relaxed with RosettaCM. C) The ASB9 N-terminus inserts as one helical turn into the active site of one CKB monomer positioning the ASB9 aspartate D32, which was shown to be critical for high affinity binding, within 6 Å of three arginines (R132, R292, R341) and an asparagine (N286).

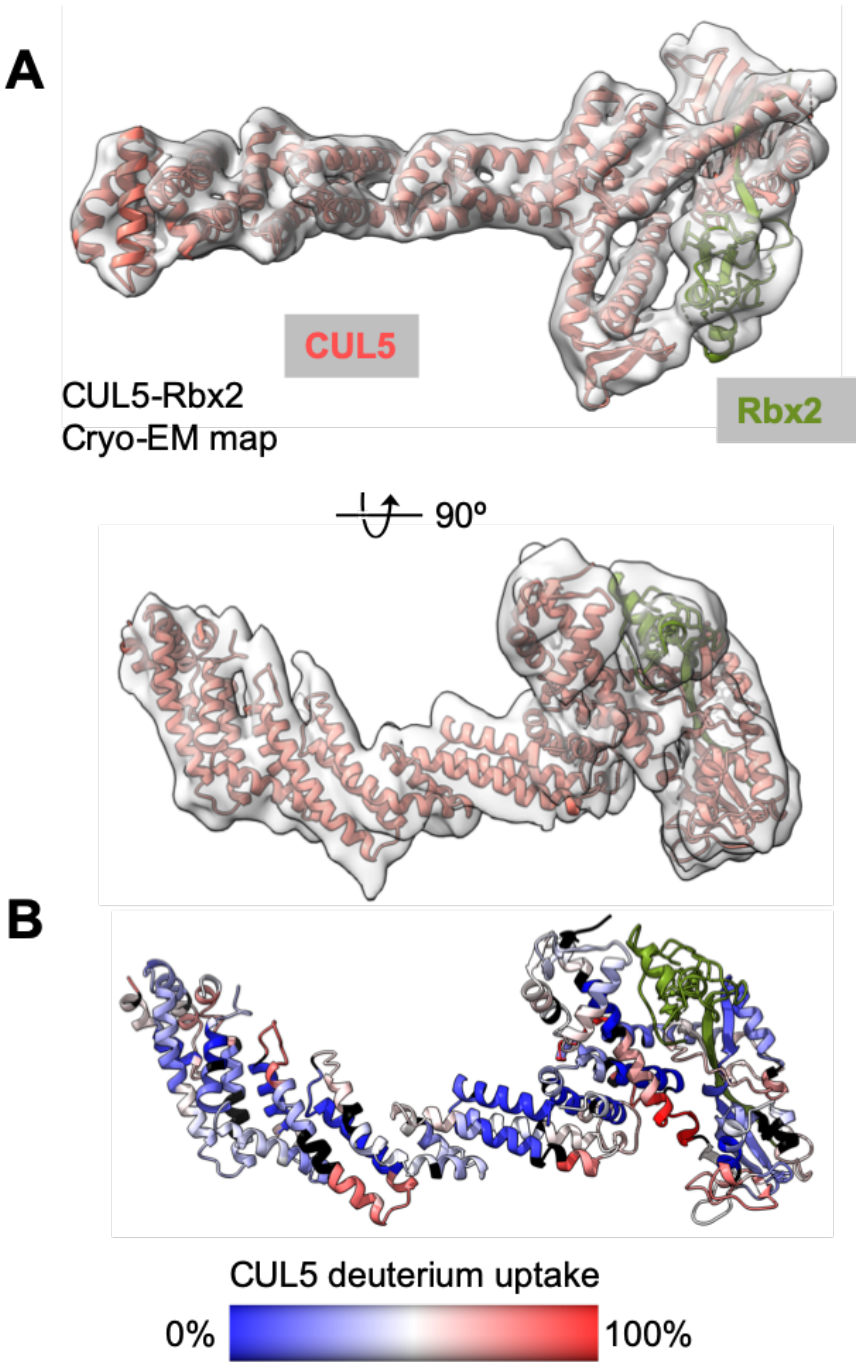

Figure 4.3: **Structure of full-length CUL5 from Cryo-EM map.** A) Models of the CUL5NTD structure (4JGH) and the CUL5CTD + RBX1 structure (3DPL) were fit into the Cryo-EM density for CUL5. B) Structure of CUL5-RBX2 showing the amide exchange results from the HDX-MS data. The exchange varied from 0% exchange (blue) to 100% exchange (red), Regions that were not covered in the HDX-MS data are colored black. RBX2 is colored olive. It is notable how much of the CUL5 shows very low amide exchange indicating a rigidly folded structure.

3. Full Model of the ASB9-CRL.

The structural model of the CKB-ASB9-ELOB/C was combined with the model of CUL5-RBX2 by superposition of ELOB/C. This process yielded a model of the ASB9CRL with the substrate, CKB bound (Fig 4.4). This model captures the Cryo-EM structure of the substrate bound to ASB9, the full structure of CUL5, and the full structure of CUL5 bound to RBX2.

4. HDX-MS mapping of the protein-protein interfaces in the CKB-bound ASB9-CRL.

We performed HDX-MS studies on various complexes to obtain more detailed information about the interfaces between the individual proteins (15). Differences in deuterium exchange of the N-terminal ankyrin repeat of ASB9 between ASB9-ELOB/C and CKB-ASB9-ELOB/C reflect the impact of association of CKB with ASB9. Previous HDX-MS experiments identified that residues 25-42 of ASB9 were protected by CKB but did not go into greater detail (10). In the current study, peptides covering ASB9 residues 1-8 were not detected, but residues 13-24 were found to be maximally exchanging in all states (Fig 4.5A, 4.6). Peptides corresponding to ASB9 residues 25-31, 25-42, 43-52, 50-59, 64-84, 85-99 and 100-107 (Fig 4.5A, 4.7A) all showed statistically significant decreases in exchange upon association with CKB, demonstrating that at least two ankyrin repeats of ASB9 engage in the CKB binding interaction. The decreased exchange was most notable in residues 50-59, the second helix of the first ankyrin repeat, with up to 64% less deuterium exchange in the presence of CKB.

Differences in deuterium uptake for CKB-ASB9 and CKB-ASB9-ELOB/C reflect the effects of ELOB/C interaction with the SOCS-box of ASB9. Peptides corresponding to ASB9 residues 176-214, 218-237, 238-248, 249-258, 249-261, 262-289, and 284-291 all showed decreased exchange in the presence of ELOB/C (Fig 4.5B, 4.7B). By comparing the

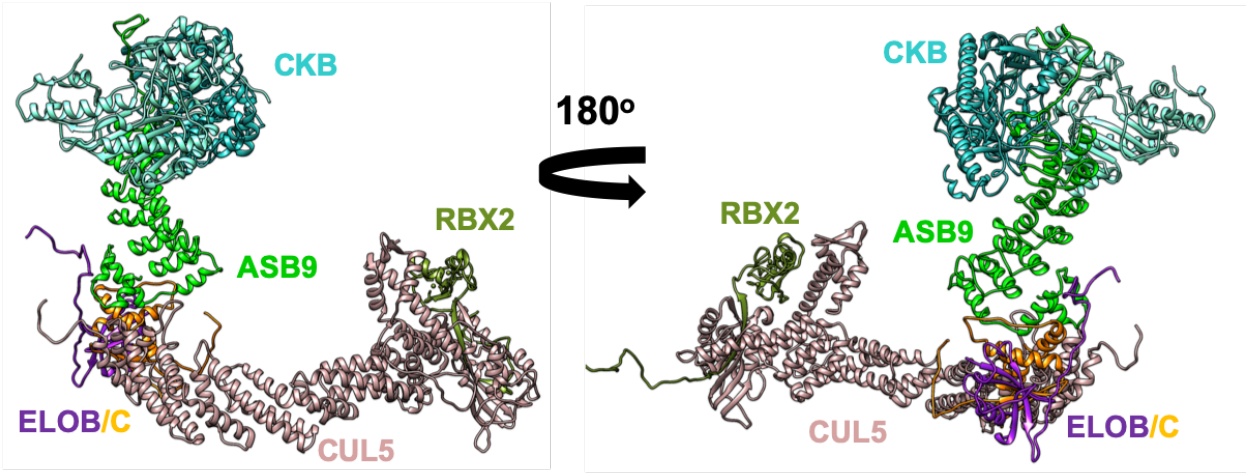

Figure 4.4: Model of the full ASB9-CRL. A complete model of the un-activated ASB9-CRL was assembled from the Cryo-EM model of CKB-ASB9-ELOB/C, with the Cryo-EM model of CUL5. The colors are as follows: CKB dimer (light sea green, aquamarine), ASB9 (green), ELOB/C (orange/purple), RBX2 (olive). A 180o rotation is also shown.

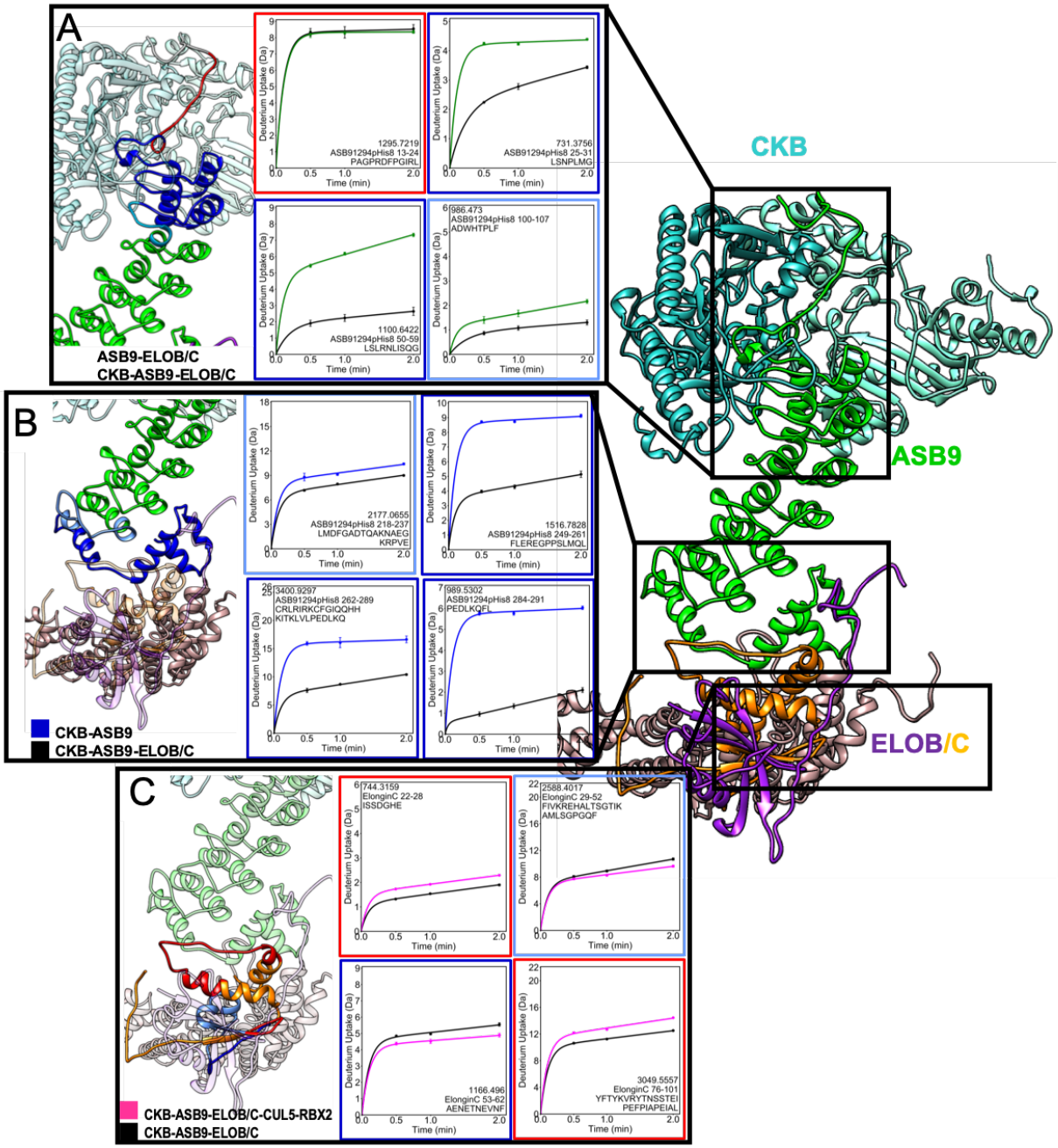

Figure 4.5: **HDX-MS and Cryo-EM data reveal protein interaction surfaces**. A) The HDX-MS data showed that ASB9 residues 13-24 exchanged completely in both free ASB9 and in CKB-bound ASB9 suggesting this region does not interact with CKB. ASB9 residues 25-99 markedly exchange upon interaction with CKB and protection extended through residue 107. B) The ASB9 SOCS box markedly decreased exchange upon binding to ELOB/C. C) ELOC showed both increased (residues 22-28 and 76-101) and decreased (residues 29-52, 53-62) exchange suggesting a hinging motion.

Figure 4.6: **Coverage map of ASB9 peptides used in the HDX-MS data analysis**

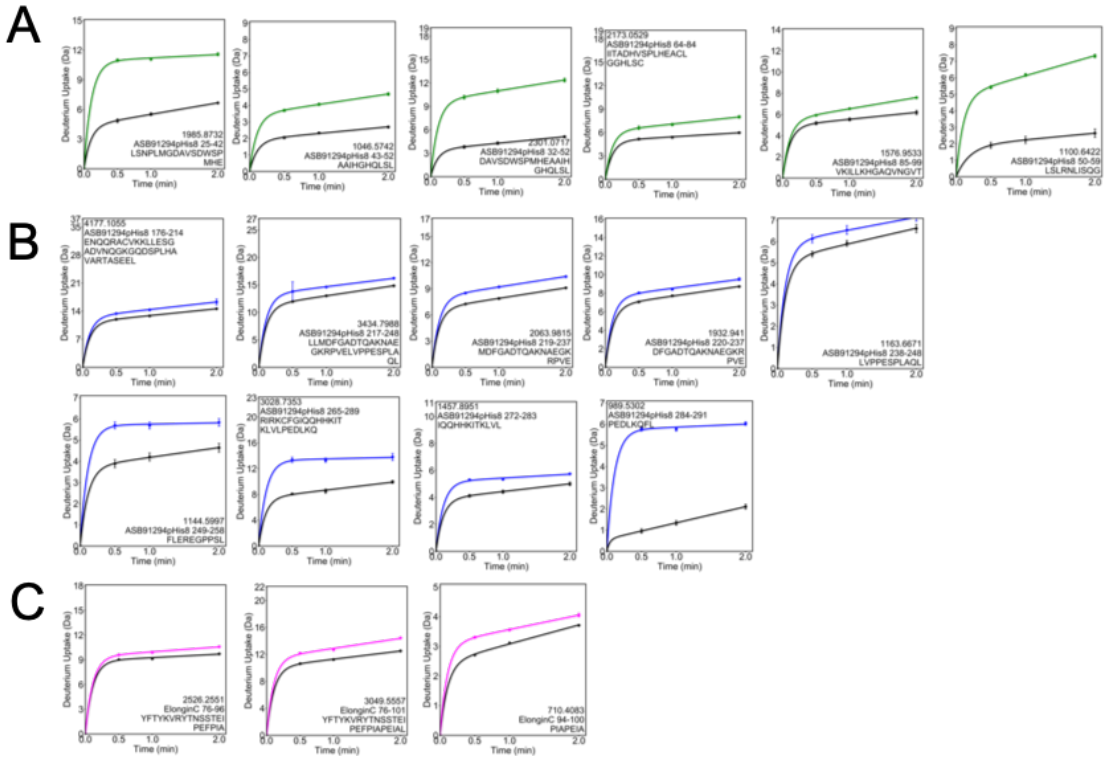

Figure 4.7: Additional Deuterium Uptake Plots for ASB9, CKB and ELOC.

deuterium uptake into residues 249-258 and residues 249-261, we determined that the majority of the difference in uptake was due to decreased exchange in residues 259-261 (16). These residues completely exchanged in ASB9 alone whereas only 0.5 Da exchanged in the ELOB-ELOC – bound state. This region corresponds to the two SOCS box subdomains; the BC-box motif, L-X-X-L-C-R, residues 258-263, and the CUL5-box motif is L-X-Φ-P-X-X-Φ-X-X-Ω-L, residues 281-291 in ASB9.

Differences in deuterium exchange between ASB9-ELOB/C and ASB9-ELOB/C-CUL5-RBX2 revealed the ELOC binding site on CUL5 (Fig 4.5C). Whereas ELOC residues 53-62 showed decreased exchange in the presence of CUL5, residues 22-28 immediately before, and residues 76-101, immediately after, showed increased exchange in the presence of CUL5 (Fig 4.5C, 4.7C and 4.8). These results suggested a rearrangement and/or hinging at ELOC takes place upon binding of CUL5.

5. HDX-MS reveals crosstalk between non-adjacent subunits of the E3 ligase.

By comparing variations of ASB9-CRL complexes, we discovered long-range changes in amide exchange indicative of allosteric crosstalk between non-adjacent subunits. We first compared ASB9-ELOB/C-CUL5-RBX2 with and without neddylation of CUL5. While we thought neddylation might induce changes in the substrate receptor, none were observed.

The amide exchange into CKB-ASB9-ELOB/C complex was compared to that of the CKB-ASB9 complex, to observe any differences in the substrate (CKB) and substrate receptor (ASB9) when ELOB/C was present. Indeed, CKB residues 177-193 and 193-203 showed decreased exchange in the presence of ELOB/C (Fig 4.9A, 4.10). This result indicates long-range communication through the substrate receptor, ASB9, to the substrate, CKB, when the ASB9 complex engages ELOB/C. We also analyzed how the ligase responded to a lack of substrate by

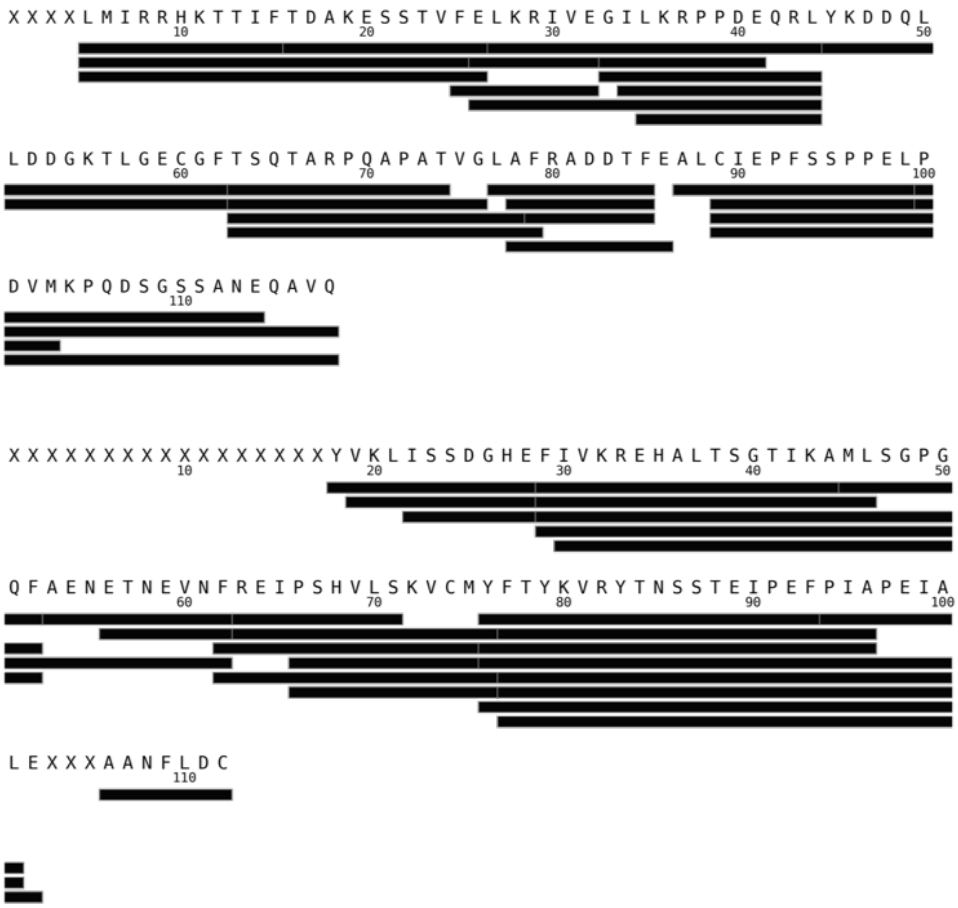

Figure 4.8: **Coverage map of ELOB and ELOC peptides used in the HDX-MS data analysis**.

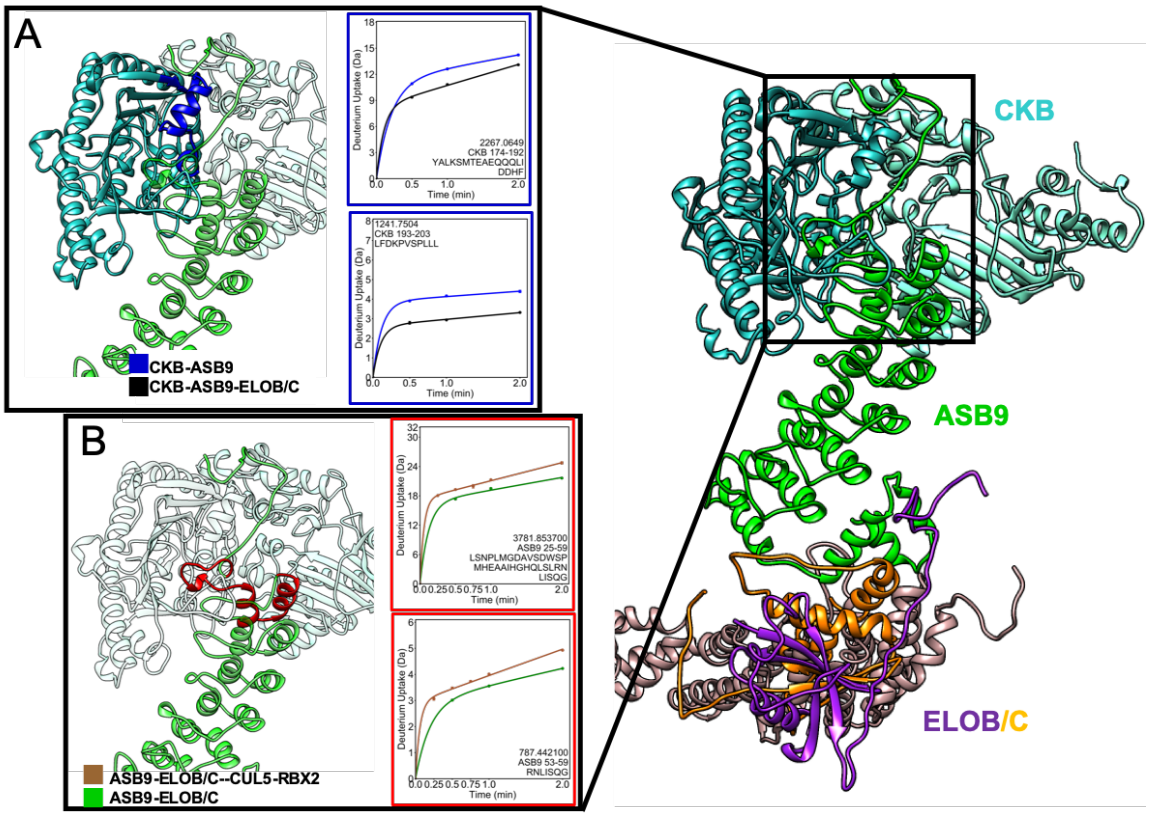

Figure 4.9: **Long-range allosteric crosstalk was observed** as A) decreased exchange in CKB residues 174-203 (bound at the N-terminus of ASB9) upon binding of ELOB/C to the C-terminal end of ASB9. B) When CUL5-RBX2 bound to ASB9-ELOB/C, the first ankyrin repeat of ASB9 showed increased exchange.

Figure 4.10: **Coverage map of CKB peptides used in the HDX-MS data analysis**

comparing deuterium exchange into the ASB9-ELOB/C complex to exchange into the ASB9-ELOB/C-CUL5-RBX2 complex. In the absence of CKB, ASB9 residues 25-42 and 53-59 showed higher deuterium exchange in the ASB9-ELOB/C-CUL5-RBX2 complex as compared to that in the ASB9-ELOB/C complex (Fig 4.9B). These results reveal even longer-range communication through the ASB9 substrate receptor and the ELOB/C adapter proteins so that when ELOB/C engage CUL5 in the absence of the substrate, the substrate binding site on ASB9 is more dynamic.

The most remarkable example of the observed long-range allostery was how CUL5 appeared to respond to the presence of substrate, CKB. When we compared the ASB9-ELOB/C-CUL5-RBX2 complex to the CKB-ASB9-ELOB/C-CUL5-RBX2 complex, decreased exchange in several regions of CUL5 was observed. This included CUL5 residues 7-21, 30-40, 105-114, 161-176, 214-229, 269-278, and 545-551. This result is most consistent with a twisting of the CUL5 molecule upon binding of CKB to ASB9 (Fig 4.11, 4.12).

D. Discussion

1. Structure of the substrate-bound ASB9-CRL based on Cryo-EM data.

We were able to obtain a 4.1 Å cryo-EM structure of the ASB9-CKB complex, the first such structure of an ASB substrate receptor bound to a bona fide ASB-CRL substrate. The cryo-EM density revealed an unexpected structuring of ASB9 residues 25-34, which pack into the active site of one subunit of the CKB dimer positioning Asp 32, a residue previously identified as critical for binding affinity, in the CKB active site (10, 11). In fact, ASB9 residues 25-34 form a partial helix in the presence of CKB whereas they exchange completely prior to binding. These data suggest that the structure between residues 25-31 may be responsible for the higher affinity binding with ASB9 1-252 as previously reported (10). It is interesting to speculate that the disordered N-terminal thirty-five residues of ASB9 may provide the promiscuous interactions necessary for

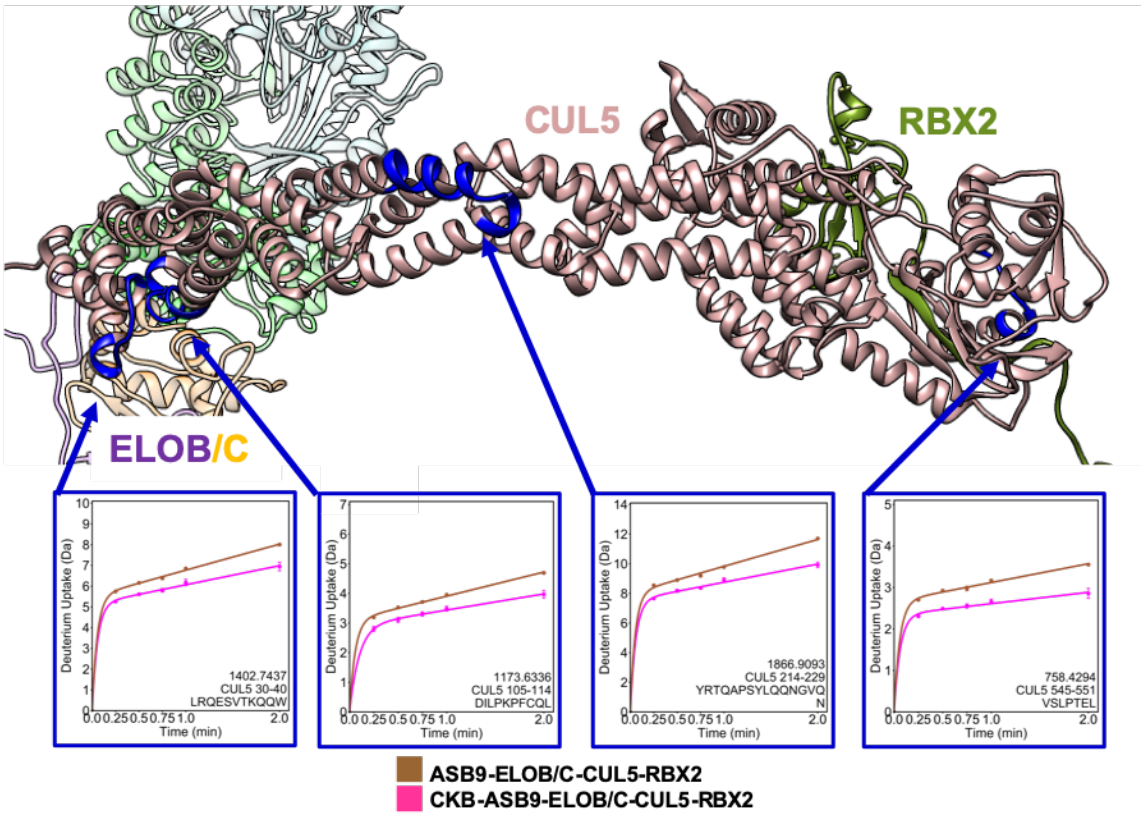

Figure 4.11: CKB binding induces changes in HDX in CUL5. CUL5 residues 30-40, 105-114, 214-229, and 545-551 showed decreased exchange upon CKB binding to the ASB9 CRL.

Figure 4.12*:* Coverage map of CUL5 peptides used in the HDX-MS data analysis

interaction with other putative ASB9 substrates which so far have only been identified by affinity mass spectrometry (6). The list of putative ASB9 substrates includes many metabolic enzymes, which, like CKB, adopt dimeric structures.

We were also able to obtain a cryo-EM structure of full-length CUL5 bound to RBX2, which has previously been only structurally characterized using fragments of the N- and C-terminal domains of CUL5. HDX-MS data of CUL5 validated the new model showing low exchange throughout the core of the entire CUL5 with only the connecting loops between the helices showing higher exchange. The core of CUL5 exchanged very little throughout, indicating that the three repeated helical bundles form one folded structure without any obvious hinge points between the helical repeats. In fact, engagement of the CUL5 with CKB-ASB9-ELOB/C actually resulted in even further decreases in exchange throughout the CUL5 molecule again suggesting that in the full E3 ligase, the CUL5 behaves as a rigid rod with no hinge points.

2. Allosteric Changes in the ASB9-CRL.

Comparison of the HDX-MS data for the ASB9-ELOB/C vs. ASB9-ELOB/C-CUL5-RBX2 revealed that the presence of CUL5-RBX2 causes increased exchange in the first ankyrin repeat of ASB9. Increased exchange is generally considered indicative of increased dynamics. It is interesting to speculate that increased dynamics at the substrate binding site may promote CKB binding by a fly-casting mechanism (17). Assembly of the CRL without its substrate bound would be dangerous as CRL components are known to autoubiquitylate (18). A mechanism such as fly-casting (17) to promote substrate binding under these conditions might be desirable. Comparison of the HDX-MS data for the CKB-ASB9 vs. CKB-ASB9-ELOB/C revealed the converse allosteric changes. In this case, the regions of the CKB homodimer that interface with ASB9 showed decreased exchange in the presence of ELOB/C. Finally, comparison of the HDX-MS data for the

CKB-ASB9-ELOB/C-CUL5-RBX2 vs. the ASB9-ELOB/C-CUL5-RBX2 complex showed that CKB binding to ASB9 induced decreased exchange at specific surface loops in CUL5 spanning from residue 25 through residue 551. This may possibly be due to CKB binding inducing changes in ELOB/C which then cause twisting of the CUL5. These results imply that CKB may induce allosteric crosstalk through ASB9-ELOB/C and then through CUL5, a series of three other proteins.

The mechanism of allostery in the ASB9 CRL appears to involve two proteins, ASB9 and CUL5, which behave as rigid rods. Indeed, the amide exchange throughout the folded core of each of these proteins is very low and remained low regardless of which sub-complexes were studied. ASB9 is an ankyrin repeat protein which is stabilized by interactions between its ankyrin repeats. As was observed in IκBα (19), such a one-dimensional architecture is an ideal structure through which to transmit allosteric changes since it is minimally frustrated (20) through its long cylindrical core and should move as a rigid body. It is interesting that a similar phenomenon is observed in CUL5, which could be considered a quasi-repeat protein, as it has a similar cylindrical architecture of repeated helical bundles, and it also appears to move as a rigid body. Neither the HDX-MS data nor the cryo-EM data show any indication of bending or hinging of the CUL5. The only place where a hinge-like motion is suggested by the HDX-MS data is in ELOB/C, in which we observe decreased exchange in one region and increased exchange in the segments preceding and following this segment. (Fig 4.13)

Our results strongly suggest that the CUL5 CRLs do not actually bend to close the distance between the substrate and the RBX2-E2-Ub, as we and others have previously suggested (11, 21). Instead, it appears that the CRL is built of two rigid rod-like proteins (ASB9 and CUL5) with a single hinge at ELOB/C. Flexibly tethered subunits such as RBX2 are then attached to CUL5,

enabling Ub transfer by a ball-and-chain mechanism. This suggestion is bolstered by recent observations that the CUL5 CRLs require ARIH2, a Ring between Ring ligase, to install the first Ub on bound substrates (22).

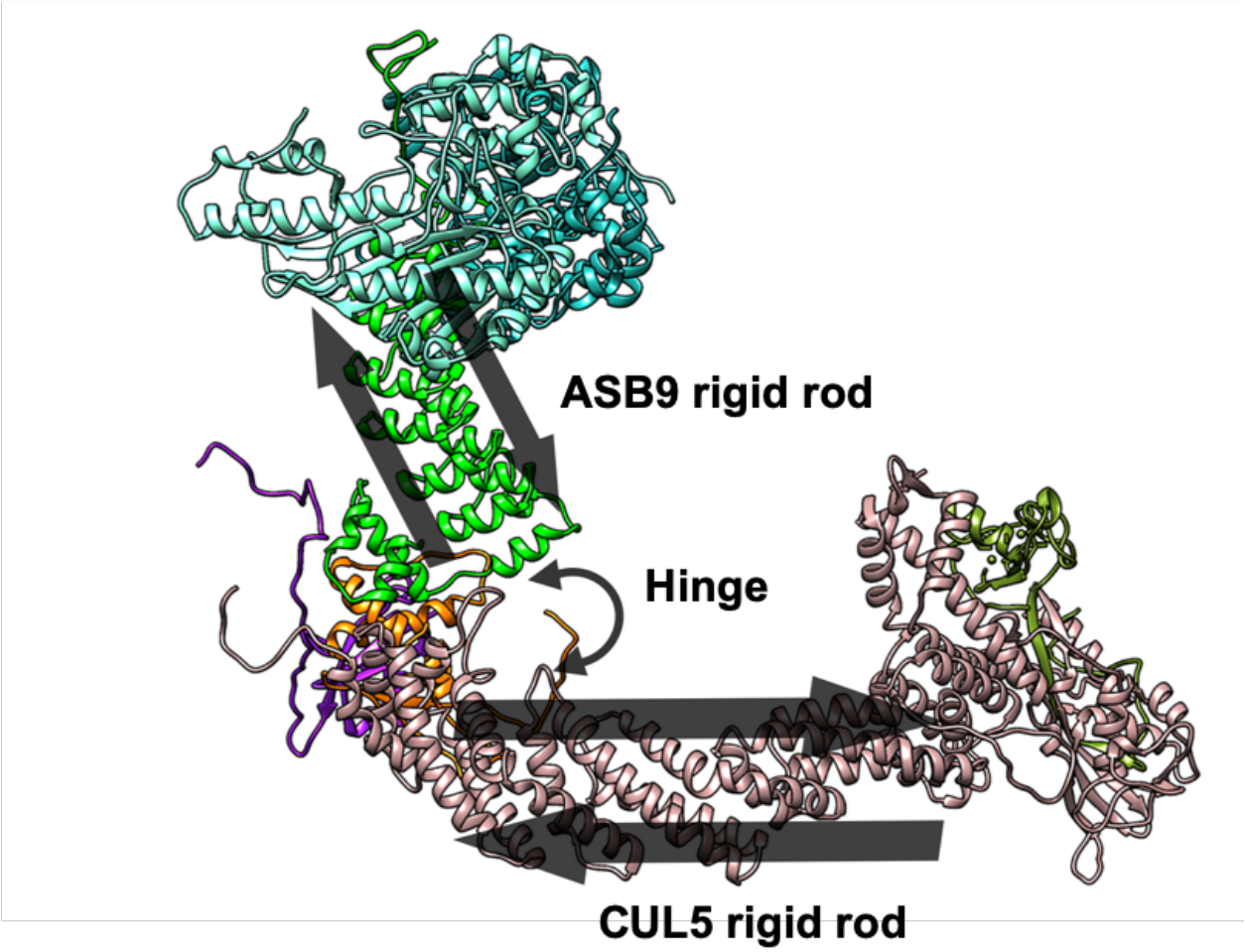

Figure 4.13: **The predicted allosteric path from substrate, CKB, through the CRL**. CKB binding communicates through the substrate receptor, ASB9 which behaves as a rigid rod, through ELOB/C which behaves as a hinge, and finally through CUL5, which also behaves as a rigid rod. Both ASB9 and CUL5 appear to move as rigid bodies in the allosteric network.

F. Acknowledgements

Chapter IV, is material that the dissertation author was the principal researcher and author of. The material has been submitted for publication. (**Lumpkin, R. J.,** Baker, R. W., Leschziner, A., and Komives, E. A. Structure and dynamics of the ASB9 CUL-RING E3 Ligase)

G. References

1. M. D. Petroski, R. J. Deshaies, Function and regulation of cullin-RING ubiquitin ligases. Nature reviews. Molecular cell biology 6, 9-20 (2005).

2. J. R. Lydeard, B. A. Schulman, J. W. Harper, Building and remodelling Cullin-RING E3 ubiquitin ligases. EMBO reports 14, 1050-1061 (2013).

3. T. Kamura, K. Maenaka, S. Kotoshiba, M. Matsumoto, D. Kohda, R. C. Conaway, J. W. Conaway, K. I. Nakayama, VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. Genes Dev 18, 3055-3065 (2004).

4. J. Kohroki, T. Nishiyama, T. Nakamura, Y. Masuho, ASB proteins interact with Cullin5 and Rbx2 to form E3 ubiquitin ligase complexes. FEBS Lett 579, 6796-6802 (2005).

5. F. Okumura, A. Joo-Okumura, K. Nakatsukasa, T. Kamura, The role of cullin 5-containing ubiquitin ligases. Cell division 11, 1 (2016).

6. C. A. Andresen, S. Smedegaard, K. B. Sylvestersen, C. Svensson, D. Iglesias-Gato, G. Cazzamali, T. K. Nielsen, M. L. Nielsen, A. Flores-Morales, Protein Interaction Screening for the Ankyrin Repeats and Suppressor of Cytokine Signaling (SOCS) Box (ASB) Family Identify Asb11 as a Novel Endoplasmic Reticulum Resident Ubiquitin Ligase. J Biol Chem 289, 2043-2054 (2014).

7. J. R. C. Muniz, K. D. Guo, N. J. Kershaw, V. Ayinampudi, F. von Delft, J. J. Babon, A. N. Bullock, Molecular Architecture of the Ankyrin SOCS Box Family of CuI5-Dependent E3 Ubiquitin Ligases. J Mol Biol 425, 3166-3177 (2013).

8. M. A. Debrincat, J. G. Zhang, T. A. Willson, J. Silke, L. M. Connolly, R. J. Simpson, W. S. Alexander, N. A. Nicola, B. T. Kile, D. J. Hilton, Ankyrin repeat and suppressors of cytokine signaling box protein asb-9 targets creatine kinase B for degradation. J Biol Chem 282, 4728-4737 (2007).

9. S. Kwon, D. Kim, J. W. Rhee, J. A. Park, D. W. Kim, D. S. Kim, Y. Lee, H. J. Kwon, ASB9 interacts with ubiquitous mitochondrial creatine kinase and inhibits mitochondrial function. BMC Biol 8, 23 (2010).

10. D. Balasubramaniam, J. Schiffer, J. Parnell, S. P. Mir, R. E. Amaro, E. A. Komives, How the ankyrin and SOCS box protein, ASB9, binds to creatine kinase. Biochemistry 54, 1673-1680 (2015).

11. J. M. Schiffer, R. D. Malmstrom, J. Parnell, C. Ramirez-Sarmiento, J. Reyes, R. E. Amaro, E. A. Komives, Model of the Ankyrin and SOCS Box Protein, ASB9, E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer. Structure 24, 1248-1256 (2016).

12. X. Fei, Y. Zhang, X. Gu, R. Qiu, Y. Mao, C. Ji, Crystallization and preliminary X-ray analysis of the splice variant of human ankyrin repeat and suppressor of cytokine signaling box protein 9 (hASB9-2). Protein Pept Lett 16, 333-335 (2009).

13. Y. Kim, M. Kwak, B. Ku, H. Suh, K. Joo, J. Lee, J. Jung, B. Oh, Structural basis of intersubunit recognition in elongin BC-cullin 5-SOCS box ubiquitin-protein ligase complexes. Acta Crystallogr D Biol Crystallogr 69, 1587-1597 (2013).

14. D. M. Duda, J. L. Olszewski, J. P. Schuermann, I. Kurinov, D. J. Miller, A. Nourse, A. F. Alpi, B. A. Schulman, Structure of HHARI, a RING-IBR-RING ubiquitin ligase: autoinhibition of an Ariadne-family E3 and insights into ligation mechanism. Structure 21, 1030-1041 (2013).

15. J. G. Mandell, A. Baerga-Ortiz, S. Akashi, K. Takio, E. A. Komives, Solvent accessibility of the thrombin-thrombomodulin interface. J Mol Biol 306, 575-589 (2001).

16. R. Lumpkin, E. A. Komives, DECA, a comprehensive, automatic post-processing program for HDX-MS data. Mol Cell Proteomics 10.1074/mcp.TIR119.001731 (2019).

17. P. J. Shoemaker BA, Wolynes PG., Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A 97, 8868-8873 (2000).

18. P. Zhou, P. M. Howley, Ubiquitination and degradation of the substrate recognition subunits of SCF ubiquitin-protein ligases. Mol Cell 2, 571-580 (1998).

19. H. E. Dembinski, K. Wismer, J. D. Vargas, G. W. Suryawanshi, N. Kern, G. Kroon, H. J. Dyson, A. Hoffmann, E. A. Komives, Functional importance of stripping in NFkappaB signaling revealed by a stripping-impaired IkappaBalpha mutant. Proc Natl Acad Sci U S A 114, 1916-1921 (2017).

20. D. U. Ferreiro, J. A. Hegler, E. A. Komives, P. G. Wolynes, Localizing Frustration in native proteins and protein assemblies Proc Natl Acad Sci U S A 104, 19819-19824 (2007).

21. N. R. Liu J, Flexible cullins in cullin-RING E3 ligases allosterically regulate ubiquitination. J Biol Chem 286, 40934-40942. (2011).

22. R. Huttenhain, J. Xu, L. A. Burton, D. E. Gordon, J. F. Hultquist, J. R. Johnson, L. Satkamp, J. Hiatt, D. Y. Rhee, K. Baek, D. C. Crosby, A. D. Frankel, A. Marson, J. W. Harper, A. F. Alpi, B. A. Schulman, J. D. Gross, N. J. Krogan, ARIH2 Is a Vif-Dependent Regulator of CUL5-Mediated APOBEC3G Degradation in HIV Infection. Cell Host Microbe 26, 86-99 e87 (2019).

23. D. Narang, W. Chen, C. G. Ricci, E. A. Komives, RelA-Containing NFkappaB Dimers Have Strikingly Different DNA-Binding Cavities in the Absence of DNA. J Mol Biol 430, 1510-1520 (2018).

24. K. M. Ramsey, H. E. Dembinski, W. Chen, C. G. Ricci, E. A. Komives, DNA and IkappaBalpha Both Induce Long-Range Conformational Changes in NFkappaB. J Mol Biol 429, 999-1008 (2017).

25. T. E. Wales, K. E. Fadgen, G. C. Gerhardt, J. R. Engen, High-speed and high-resolution UPLC separation at zero degrees Celsius. Anal Chem 80, 6815-6820 (2008).

26. M. A. Herzik, Jr., M. Wu, G. C. Lander, Achieving better-than-3-A resolution by single-particle cryo-EM at 200 keV. Nat Methods 14, 1075-1078 (2017).

27. C. Suloway, J. Pulokas, D. Fellmann, A. Cheng, F. Guerra, J. Quispe, S. Stagg, C. S. Potter, B. Carragher, Automated molecular microscopy: the new Leginon system. J Struct Biol 151, 41-60 (2005).

28. S. Q. Zheng, E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14, 331-332 (2017).

29. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol 192, 216-221 (2015).

30. N. R. Voss, C. K. Yoshioka, M. Radermacher, C. S. Potter, B. Carragher, DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. J Struct Biol 166, 205-213 (2009).

31. J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl, S. H. Scheres, New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7 (2018).

32. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290-296 (2017).

33. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180, 519-530 (2012).

34. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-1612 (2004).

35. M. A. Cianfrocco, I. Lahiri, F. DiMaio, A. E. Leschziner, cryoem-cloud-tools: A software platform to deploy and manage cryo-EM jobs in the cloud. J Struct Biol 203, 230-235 (2018).

36. R. Y. Wang, M. Kudryashev, X. Li, E. H. Egelman, M. Basler, Y. Cheng, D. Baker, F. DiMaio, De novo protein structure determination from near-atomic-resolution cryo-EM maps. Nat Methods 12, 335-338 (2015).

37. R. Y. Wang, Y. Song, B. A. Barad, Y. Cheng, J. S. Fraser, F. DiMaio, Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. Elife 5 (2016).

38. A. Sali, T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234, 779-815 (1993).

Chapter V:

Allosteric Regulation of CUL5-Mediated Ubiquitylation

A. Introduction

Ubiquitin (Ub) is a post-translational modification (PTM) that requires an Activating Enzyme (E1), transfer to a Conjugating Enzyme (E2), and incorporation into the target protein by a Ub Ligase (E3) (Fig 5.1A). Association of substrates with E3s leads to the covalent attachment of Ub onto exposed lysine residues. Ub may be further conjugated onto the N-terminal methionine or any of its seven lysine residues, leading to the formation of poly-Ub chains. Poly-ubiquitin linkage type is determined by the E2 enzyme, and the formation of polyubiquitin chains may involve more than one E2 enzyme (1). Both mono-ubiquitin and the various poly-ubiquitin chains lead to distinct signals. K48 chains signal for proteolysis, while K63 chains coordinate processes such as innate immunity, protein trafficking and endocytosis. Atypical chains, M1/K6/K11/K27/K29/K33, are involved in the DNA damage response, cell cycle control, Wnt/β-catenin signaling, TCR signaling, protein trafficking, autophagy, angiogenesis, and innate immunity (2).

Ub Ligases can be categorized into three major groups: HECT (Homologous to E6-AP carboxyl terminus) E3s, RING (Really Interesting New Gene) E3s, and RBR (Ring-Between-RING) E3s. RING E3s account for approximately 90% of known Ub Ligases, and the most common form of RING E3s is the Cullin-RING Ligase (CRL). These E3s are multi-subunit complexes consisting of a substrate receptor, adaptor proteins, a Cullin protein (CUL1, CUL2, CUL3, CUL4, CUL5, CUL7), and a RING protein (RBX1, RBX2). The RING protein associates with Ub-bound E2 enzymes to facilitate the direct transfer of Ub from the E2 to bound substrate proteins. While all CUL proteins can interact with RBX1, RBX2 binds specifically to CUL5 (3). CUL2 and CUL5 have been shown to bind Elongin B and Elongin C (ELOB/C), through which CUL2 binds vHL-box substrate receptors and CUL5 binds Suppressor of Cytokine Signaling

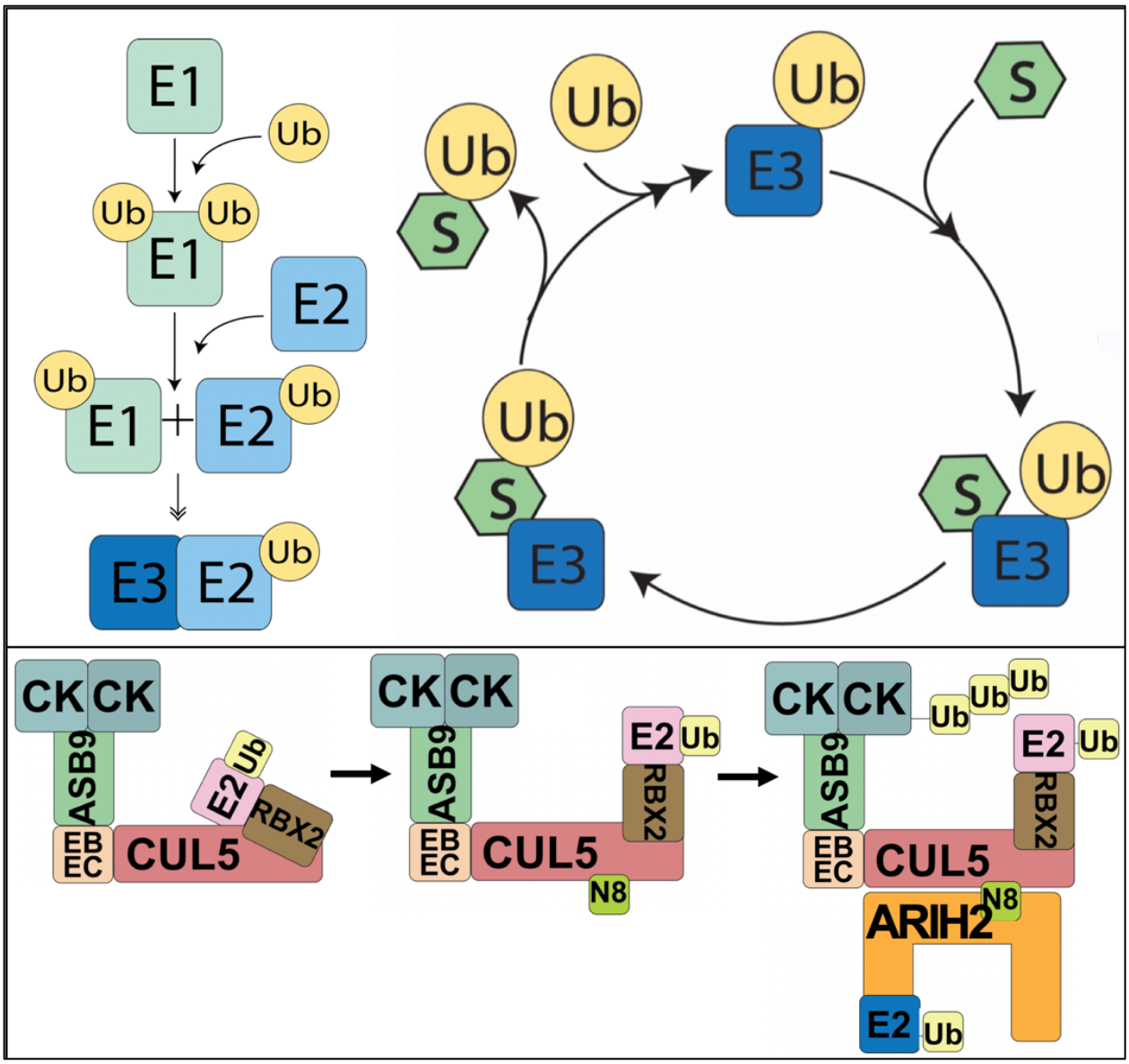

Figure 5.1: Schematics of the ubiquitylation cascade as mediated by the ASB9 CUL5 E3 ligase. A) Ubiquitin is activated through a three-step enzymatic cascade. The Ubiquitin-activating enzyme (E1) binds ATP and catalyzes adenylylation of Ubiquitin. The active-site cysteine on E1 attacks the Ub-AMP complex to form a thioester bond. Through a trans-thioesterification reaction Ubiquitin is transferred to the active-site cysteine on a Ubiquitin-conjugating enzyme (E2). Ubiquitin Ligases (E3) facilitate the highly-specific covalent attachment of activated Ubiquitin (Ub) to bound substrate proteins through an isopeptide bond on an exposed lysine residue. B) Schematic showing the states of the ASB9-CUL-RING Ligase explored in this work: un-activated, activated by Nedd8, and activated by Nedd8 with RING-Between-RING Ligase present.

(SOCS)-box substrate receptors. Conserved CUL and BC-box domains in the substrate receptors determine this specificity. Several protein families contain SOCS-box domains, including the canonical Suppressor of Cytokine Signaling (SOCS), Single Stranded Binding (SSB), WD40 SOCS-box (WSB), and Ankyrin Repeat and SOCS-box (ASB) (3). CUL5 was originally determined to have characteristics of an E3 ligase based on its similarity to CUL2 in binding ELOB and ELOC and ASB2 (4).

ARIH1 and ARIH2 are RING-Between-RING ligases (RBRL), a group of E3 ligases containing characteristics of both HECT and RING ligases (5). These E3s contain two RING domains and an In-between-RING domain, but unlike RING ligases, the E3s associate with a cysteine-reactive E2 enzyme E2L3 that transfers Ub to the catalytic cysteine on the E3 (6). These E3s have been shown to be auto-inhibited, and this is relieved upon binding to CRLs. ARIH1 has been shown to interact with CUL1-4 while ARIH2 specifically interacts with CUL5. The interaction between ARIH2 and CUL5 has been shown to be dependent upon the N-terminus of ARIH2, likely due to attraction between the ARIH2 acidic N-terminus and a basic patch on CUL5 (7). This interaction has a symbiotic effect, whereupon activation of the RBRL by the CRL enables the RBRL to ubiquitinate the substrate on the CRL. (8)

The largest subgroup of CRL E3 ligases is the ASB-CUL5 CRL class. The ASB family contains 18 proteins each of which has been shown to bind around 10 proteins (9), presumably substrates for ubiquitylation although that remains to be demonstrated. The ASB proteins contain an ankyrin repeat domain (ARD) consist of varying numbers of ankyrin repeats which are helix-turn-helix-loop motifs that stack upon one another forming a cooperatively folded unit (10). ARDs have a well-established role in facilitating protein-protein binding (11). ARD, presumably involved in substrate binding, and a SOCS-box domain that contains conserved BC-box and

CUL5-box motifs. ASB9 was found to pulldown both Brain Type Creatine Kinase (CKB) and Mitochondrial Creatine Kinase (MtCK) from HEK293 cells and induce ubiquitylation and degradation of Creatine Kinase in a SOCS-box dependent manner (12, 13). ASB9 was subsequently shown to bind a dimer of CKB at the first ankyrin repeat (14). Structural determination of the ASB9-ELOB/C complex (15) and modeling of CKB, CUL5 and RBX1 led to models of the full ligase that depicted CKB bound to ASB9 approximately 70Å away from the E2-Ub complex on RBX2, suggesting that large conformational changes in the ligase must be necessary to facilitate ubiquitylation of CKB (16). ARIH2 has been shown to co-precipitate with ASB9 in pulldown proteomics studies (17), and recently it has been shown to help ubiquitinate CKB in the presence of the ASB9-CRL. (18)

Here, we present biochemical, biophysical, and structural data on the ASB9-CRL we have assembled in vitro. We explore how and when this E3 ligase ubiquitylates one of its substrates, CKB. We demonstrate that the ASB9-CRL does not significantly ubiquitylate CKB in the absence of ARIH2, and that the addition of ARIH2 is sufficient to add multiple Ub moieties to CKB. We present data regarding the necessity of neddylation and demonstrate that neddylation is required for high affinity binding of ARIH2. We used mass spectrometry to identify four lysine residues on CKB that are modified by the ASB9-CRL-ARIH2-E2L3 complex, and we show that the addition of other E2s such as CDC34 or E2D2 modulate the polyubiquitin linkage activity of the ligase, but not the sites of ubiquitylation on CKB. Using hydrogen-deuterium exchange mass spectrometry (HDX-MS), we show that neddylation rearranges the C-terminal domain of CUL5 whereupon several regions become more solvent exposed, while a region of RBX2 corresponding to the NEDD8 E2 binding site decreases in exchange. The association of the ASB9-CRL with ARIH2 decreases deuterium exchange at some of the same CUL5 regions that increased exchange upon

neddylation, and the binding event activates ARIH2 through an allosterically triggered opening of the auto-inhibited Ariadne domain. These results are interpreted in light of models of the complete ligase including one in which ARIH2 is docked to CUL5 that is validated by the HDX-MS data.

B. Materials and Methods

1. Expression vectors.

Human ASB9-1 in pNIC-CTHF was obtained from the Structural Genomics Consortium and subcloned into pHis8 (KanR) with an N-terminal 8xHis tag. Human CKB was subcloned into pET11a (AmpR) as previously described (14). Human ELOB (full-length) and ELOC (17-112) were obtained in pACYC (CamR) (Structural Genomics Consortium). Human CUL5 and Mouse RBX2 were obtained in pRSFDuet (KanR) with an N-terminal His tag, TEV cleavage site, and GB1 tag on CUL5 (gift from Nevan Krogan). 6xHis-GB1-TEV-CUL5 was subcloned into pET28a (KanR) and RBX2 was subcloned into pET11a (AmpR).

Human UBA1 in pET21b (AmpR) with an N-terminal 6xHis tag was obtained from Addgene (Plasmid #34965). Human UBE2D2 in pET-SUMO (AmpR) with an N-terminal 6xHis tag and SUMO solubility tag was obtained from Addgene (Plasmid #60443) and was subcloned into pET28a with an N-terminal 6xHis tag and TEV cleavage site. Human CDC34 in pDEST17 with an N-terminal 6xHis tag was obtained from Addgene (Plasmid #18674). Human Ub in pcDNA3 (AmpR) with an N-terminal HA tag was obtained from Addgene (Plasmid #18712) and subcloned into pET28a (KanR) with an N-terminal 6xHis tag and TEV cleavage site. Human NAE1 and UBA3 were obtained in pGEX4T1 (AmpR) with a GST tag (gift from Brenda Schulman). Human UBE2F in pDEST17 (AmpR) with an N-terminal 6xHis tag was obtained from Addgene (Plasmid #15800) and was subcloned into pET28a (KanR) with an N-terminal 6xHis tag and TEV cleavage site. Human Nedd8 in pcDNA3 (AmpR) with an N-terminal HA tag was

obtained from Addgene (Plasmid #18711) and was subcloned into pET28a (KanR) with an N-terminal His tag. Codon-optimized human ARIH2 was synthesized with an N-terminal 6xHis tag, MBP solubility tag, and TEV cleavage site, and it was subcloned into pET11a (AmpR) (GENEWIZ, Inc.). Codon-optimized human UBE2L3 was synthesized and subcloned into pET28a (KanR) with an N-terminal 6xHis tag and TEV cleavage site. TEV Protease was obtained in pRK793 (AmpR) with an N-terminal MBP solubility tag, TEV cleavage site, and 6xHis tag. (Gift from César Ramirez)

2. Protein Expression.

ASB9 was co-expressed with ELOB/ELOC and/or CKB using sequential transformation with the vectors described above into BL21 E. coli cells for ASB9-CKB, ASB9-ELOB/C, and ASB9-CKB-ELOB/C protein complexes. Vectors containing ELOB/ELOC, CKB, or both were transformed into competent BL21 cells after which those cells were made competent again. ELOB/ELOC in pACYC was selected for by Chloramphenicol (CAM) resistance, and CKB in pET11a was selected for by Ampicillin (AMP) resistance. The vector for expression of ASB9 was transformed into ELOB/ELOC + CKB – containing BL21 cells and plated on a Kanamycin (KAN)-CAM-AMP LB agar plate. The pET28a vector for expression of CUL5 and the pET11a vector for expression of RBX2 were co-transformed into ELOB/ELOC - containing BL21 cells and plated on a KAN-CAM-AMP LB agar plate for co-expression of CUL5/RBX2/ELOB/ELOC. Vectors for expression of UBE1 and Ub were co-transformed into BL21 cells and plated on a KAN-AMP LB agar plate for co-expression of UBE1/Ub. Vectors for expression of UBE2D2, UBE2L3, UBE2F, Ub and Nedd8 were transformed individually into BL21 cells and plated on KAN LB agar plates. Vectors for expression of NAE1/UBA3, CDC34, and ARIH2 were transformed individually into BL21 cells and plated on AMP LB agar plates.

All proteins were expressed as follows. A 5 mL M9-ZN (1.5x M9 salts, NZ-Amine media, 0.8% Dextrose, 1 mM MgSO4, 0.2 mM CaCl2) overnight culture was inoculated with a single colony from the plate. A 20 mL M9-Zn starter culture was inoculated with 2 mL of the overnight culture and grown for 3 hours at 37°C. The 1 L M9-ZN growth culture was inoculated with the entire 20 mL starter culture and grown until OD600 = 0.8. After placing the cultures on ice for 15 min, protein expression was induced by addition of IPTG to a final concentration of 0.5 mM, and the cultures were transferred to an 18°C incubator for 16-18 hours. Because RBX2 and ARIH2 contain Zn binding domains, the cultures containing either of those proteins were brought to 200 μM Zn by addition of a 1M solution of ZnCl2 just prior to induction.

3. Protein Purification.

Cells from 1 L of culture were pelleted by centrifugation at 5000 rpm for 10 minutes, then re-suspended in 40 mL Resuspension Buffer (50 mM Tris-HCL pH 8.0, 100 mM NaCl, 10 mM Imidazole pH 8.0, 2 mM β-mercaptoethanol, 5% Glycerol) with Sigma Protease Inhibitor Cocktail (P2714) and 5 mM PMSF. Cells were lysed on ice by sonication with ten 30 seconds pulses with 45 seconds cooldown between each pulse. The lysate was clarified by centrifugation at 13000 rpm for 45 minutes. The clarified lysate was incubated with 2 mL Ni-NTA (in Resuspension Buffer) for 2 hours at 4°C with rocking. Ni-NTA beads were pelleted by centrifugation at 700xg for 5 minutes. The supernatant was discarded, and the beads were washed with 10 mL Wash Buffer (50 mM Tris-HCL pH 8.0, 100 mM NaCl, 25 mM Imidazole pH 8.0, 2 mM β-mercaptoethanol, 5% Glycerol) for 30 minutes at 4°C. Ni-NTA beads were again pelleted by centrifugation at 700xg for 5 minutes. The supernatant was discarded, and the beads were washed with 10 mL Elution Buffer (50 mM Tris-HCL pH 8.0, 100 mM NaCl, 250 mM Imidazole pH 8.0, 2 mM β-mercaptoethanol, 5% Glycerol) for 30 minutes at 4°C. Ni-NTA beads were again pelleted by centrifugation at 700xg

for 5 minutes. The supernatant was transferred to a 12-14 kDa Dialysis bag and dialyzed overnight in Dialysis Buffer (20 mM Tris-HCL pH 8.0, 100 mM NaCl, 5% Glycerol, 1 mM DTT). Samples were concentrated to 2 mL and purified using size-exclusion chromatography over a Superdex S200 16 x 600 column in Dialysis Buffer. Peak fractions were combined and concentrated to 5 μM for analysis by HDX-MS.

Neddylation reactions following purification of all necessary components resulted in only partial Nedd8ylation of CUL5. We overcame this limitation by Nedd8ylating CUL5 during its purification, following elution from Ni-NTA but prior to size-exclusion chromatography. Lysates for NAE1/UBA3, NEDD8, and the CUL5-containing complex were combined and co-purified. Following elution from Ni-NTA and dialysis, UBE2F was added to the sample. Full nedd8ylation occurred within 30 min at 4°C. The neddylated CUL5 complex was isolated from NAE1/UBA3, UBE2F, and excess NEDD8 in the subsequent size exclusion chromatography step.

GB1 was cleaved from CUL5 and MBP was cleaved from ARIH2 by adding TEV to the proteins at a 1:10 molar ratio and incubating at 4°C for 24 hours. TEV and the cleaved tags were removed by size-exclusion over a Superdex S200 Increase 10/300 column in 20 mM Tris-HClH pH 8.0, 100 mM NaCl, 5% Glycerol, 1 mM DTT.

Sample compositions were identified and characterized according to the presence of the desired proteins as assessed by size exclusion chromatography, SDS-PAGE (10-15% Acrylamide gels), and nanospray LC MS/MS on a Lumos mass spectrometer after trypsin digestion.

ELOB/C-CUL5-RBX2, ELOB/C-CUL5(NEDD8)-RBX2, ARIH2, and ELOB/C-CUL5(NEDD8)-RBX2-ARIH2 protein samples were individually prepared for analysis by HDX-MS. ELOB/C-CUL5-RBX2, ELOB/C-CUL5(NEDD8)-RBX2, and ARIH2 were expressed and purified as described above. ELOB/C-CUL5(NEDD8)-RBX2-ARIH2 was formed by combining

the ELOB/C-CUL5-RBX2 and ARIH2 following neddylation of CUL5. The combined complex was isolated by size-exclusion over a Superdex S200 Increase 10/300 column in 20 mM Tris-HCL pH 8.0, 100 mM NaCl, 5% Glycerol, 1 mM DTT.

4. Hydrogen Deuterium Exchange.

HDX-MS experiments were conducted using a Waters nanoACQUITY UPLC system equipped with H/DX technology and a LEAP H/D-X PAL liquid handling system as previously described (27). The H¬2O Buffer was composed of 20 mM Tris-HCL pH 8.0, 100 mM NaCl, 5% Glycerol, and 1 mM DTT, matching the Size-Exclusion Buffer used in the final stage of purification for each protein sample. This Buffer was lyophilized and resuspended in D2O (D2O Buffer). A 4 μL portion of a 5 μM protein sample was incubated for 5 min at 25 °C and then mixed with 56 μL of H2O Buffer as a control or D2O Buffer for deuteration times of 15 s, 30 s, 45 s, 60 s, or 120 s. Our lab has previously shown that measurement of amide exchange during this time regime best captures allosteric transitions (28).The reaction was quenched with 60 μL of Quench Buffer (3 M guanidine, 0.1% formic acid, pH 2.66) at 0°C. A portion of the quenched sample (50 μL) was injected into a sample loop and subsequently digested on an in-line pepsin column (Immobilized Pepsin, Pierce Inc.) at 15°C. BEH C18 Vanguard pre-column, separated by analytical chromatography (Acquity UPLC BEH C18, 1.7 μM, 1.0 × 50 mm, Waters Corporation) using a 7-85% acetonitrile in 0.1% formic acid over 7.5 min, and electrosprayed into the Waters SYNAPT G2Si quadrupole time-of-flight mass spectrometer. The mass spectrometer was set to collect data in the Mobility, ESI+ mode; mass acquisition range of 200–2,000 (m/z); scan time 0.4 s. Continuous lock mass correction was accomplished with infusion of leu-enkephalin (m/z = 556.277) every 30 s (mass accuracy of 1 ppm for calibration standard). For peptide identification, the mass spectrometer was set to collect data in MSE, ESI+ mode instead.

The peptides were identified from triplicate MSE analyses of 5 μM ELOB/C-CUL5(NEDD8)-RBX2 and ARIH2 samples by PLGS 2.5 (Waters Corporation). Peptide masses were identified using a minimum number of 250 ion counts for low energy peptides and 50 ion counts for their fragment ions. The peptides identified in PLGS were then analyzed in DynamX 3.0 (Waters Corporation) and the deuterium uptake was corrected for back-exchange as previously described (29). The relative deuterium uptake for each peptide was calculated by comparing the centroids of the mass envelopes of the deuterated samples vs. the undeuterated controls following previously published methods (30).

5. Ubiquitylation Reactions.

Ubiquitylation reactions were carried out in a buffer containing 20mM Tris-HCl pH 7.5, 5 mM MgCl2, 0.5 mM DTT, and 2 mM ATP with 0.5 μM UBE1, 5 μM UBE2D2 or CDC34 and/or ARIH2+E3L3, 100-200 uM Ub, and 5 μM ASB9-CRL as previously described (31). Reactions were incubated at 37°C for 3 hours, quenched with reducing SDS-PAGE buffer, boiled at 90°C for 10 minutes prior to SDS PAGE.

Thioester transfer of Ubiquitin between UBE1 and UBE2D2 as well as the covalent ubiquitylation of CKB were measured by monitoring the molecular weights of the proteins as previously described (32). Samples were separated by 10% polyacrylamide SDS-PAGE in replicate for parallel detection by Coomassie Blue staining and by anti-Ub Western Blotting.

Samples were transferred from acrylamide gel to nitrocellulose by wet electrophoretic transfer in 1 L of cold transfer buffer (25mM Tris Base, 200 mM Glycine, 20% Methanol) at 100 V for 30 minutes. After incubating in Blocking Buffer (1% w/v Casein in 1xTBS) for 1 hour, the blot was incubated overnight with rocking at 4°C with rabbit Anti-NEDD8 Antibody (CST 2754) diluted 1:4500 in Blocking Buffer, or rabbit Anti-Creatine kinase B type Antibody (Abcam

ab38211) diluted 1:500 in Blocking Buffer. The blot was washed three times with TTBS (TBS with 0.05% Tween) and then incubated with secondary antibody Anti-Mouse IgG (H+L), HRP Conjugate (Promega W4021) or Anti-Rabbit IgG (H+L), HRP Conjugate (Promega W4011) for 1 hour with rocking at 23°C. The blot was washed twice with TTBS and finally with TBS to remove remaining surfactant. Blot was developed by incubation with Bio-Rad Clarity Western ECL Substrate (1705061) for 5 min at 23°C and was imaged using a Bio-Rad Chemi-Doc with Bio-Rad Image Lab. The Colorimetric Blot default protocol was used to image the molecular weight ladder and the Chemi High Resolution Blot default protocol was used to detect chemiluminescence with exposures from 1-10 s.

6. ImageJ Gel Analysis.

Coomassie stained gels and anti-CKB blots were imported into ImageJ (33). The ”Subtract Background” function was used to lower the background signal. Each lane of the images was selected and plotted as gel mobility vs. intensity. Areas were calculated for each peak in the lane.

7. NanoLC Mass Spectrometry for ubiquitin analysis.

Whole solution samples were electrophoresed on 10% Tris-SDS poly-acrylamide gels in SDS gel running buffer for 5 min at 180 V. The gel lanes were excised from the bottom of the stacking gel to the dye front. For mass spectrometry of individual gel bands, samples were separated on 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad 4561083) in SDS gel running buffer for 60 min at 180 V. The gels were stained for 10 min in 0.25% w/v Coomassie Blue/40% MeOH/10% Acetic Acid and de-stained for 10min in 20% MeOH/10% Acetic Acid after which gel bands were excised.

Each excised gel band was cut into small pieces and de-stained by first washing once with water and twice with 10 mM Ammonium Bicarbonate (AmmBic). The gel was incubated in 200

µL 50% Acetonitrile (ACN)/ 50 mM AmmBic for 1 hr at 37°C. The supernatant was removed, and the gel was washed twice with 200 µL water. 100µL ACN was added to dehydrate the gel and incubated for 5 min at 37°C, after which the supernatant was removed. This process was repeated two more times.

Proteins in the gel were acetylated by addition of a mixture of 5 µL acetic anhydride and 45 µL 100 mM AmmBic. If necessary, 1M AmmBic was added to bring the pH to between 7-8. The reaction was incubated for 1 hr at 37°C. The supernatant was removed, and the gel was washed three times with 200 µL water. ACN (100µL) was added to dehydrate the gel and incubated for 5 min at 37°C, after which the supernatant was removed. This process was repeated two more times. The gel was dried in a speed vac for 5 min at 60oC, then cooled to 23°C (34). Cysteines were reduced by addition of 250 µL 10 mM DTT/100 mM AmmBic and incubation for 45 min at 56°C. Cysteines were alkylated by addition of 250 µL 55 mM iodacetamide/100 mM AmmBic and incubated for 30 min at 23°C in the dark. The supernatant was removed, and the gel was incubated in 250 µL 100 mM AmmBic for 5 min at 23°C.ACN (250 µL) was added to the 100 mM AmmBic and the gel was incubated for 15 min at 23°C. The supernatant was removed, and the gel was dried in a speed vac for 5 min.

Proteins were digested by adding 50 µL of 12.5 ng/µL Trypsin in 2 mM CaCl2 / 50 mM AmmBic was added to the gel. The gel sample was incubated for 45 min at 4°C, then transferred to 37°C overnight after which time the supernatant containing the tryptic peptides was collected. To extract any remaining peptides, 20 µL of 1% Formic Acid / 2 % ACN was added to the gel and incubated for 20 min at 23°C and the supernatant was collected. Finally, 20 µL of ACN was added to the sample and incubated for 15 min at 23°C and the supernatant was collected. The Formic

Acid/ACN steps were repeated two more times. All of the collected supernatants were pooled, dried in a speed vac, and then stored at -20°C.

Trypsin-digested samples were analyzed by ultra-high pressure liquid chromatography (UPLC) coupled with tandem mass spectroscopy (LC-MS/MS) using nano-spray ionization on an Orbitrap fusion Lumos hybrid mass spectrometer (Thermo) interfaced with nano-scale reversed-phase UPLC (Thermo Dionex UltiMate™ 3000 RSLC nano System) using a 25 cm, 75-micron ID glass capillary packed with 1.7-µm C18 (130) BEHTM beads (Waters corporation). Peptides were eluted from the C18 column into the mass spectrometer using a linear gradient (5–80%) of ACN (Acetonitrile) at a flow rate of 375 nl/min for 1h. The buffers used to create the ACN gradient were: Buffer A (98% H2O, 2% ACN, 0.1% formic acid) and Buffer B (100% ACN, 0.1% formic acid). Mass spectrometer parameters are as follows; an MS1 survey scan using the orbitrap detector (mass range (m/z): 400-1500 (using quadrupole isolation), 120000 resolution setting, spray voltage of 2200 V, Ion transfer tube temperature of 275 C, AGC target of 400000, and maximum injection time of 50 ms) was followed by data dependent scans (top speed for most intense ions, with charge state set to only include +2-5 ions, and 5 second exclusion time, while selecting ions with minimal intensities of 50000 at in which the collision event was carried out in the high energy collision cell (HCD Collision Energy of 30%), and the fragment masses where analyzed in the ion trap mass analyzer (With ion trap scan rate of turbo, first mass m/z was 100, AGC Target 5000 and maximum injection time of 35ms). Protein identification and label free quantification was carried out using Peaks Studio 8.5 (Bioinformatics solutions Inc.)

Digestion of a ubiquitinylated peptide by trypsin leaves a ‘digly’ modification because trypsin cleaves after the Arg three from the C-terminus of Ub. tThe diglycyl remnant has a mass of 114.04 Da on the lysine at the site of the ubiquitylation.(19) To quantify the amount of Ub at

each lysine in CKB, the sum of the areas of all peptides with a digly at a particular lysine was divided by the sum of the areas for all peptides covering that lysine. This ratio could then be used to compare levels of ubiquitylation between samples.

8. Homology Modeling.

A model of a neddylated intermediate of ASB9-CRL with UBE2F-NEDD8 bound was prepared. Modelling started from the CKB-ASB9-ELOB/C-CUL5NTD model described in Lumpkin et al., submitted. To orient the CUL5CTD relative to CUL5NTD, CUL2 from CUL2-RBX1-ELOB/C-VHL (5N4W) was first aligned to residues 306-387 on CUL5NTD in the starting model using PyMOL. CUL1CTD from CUL1-RBX1-UBC12~NEDD8-DCN1 (4P5O) was aligned to residues 383-424 on CUL2 from CUL2-RBX1-ELOB/C-VHL (5N4W). Following orientation of the two halves of CUL5, CUL2-RBX1-ELOB/C-VHL was removed from the model. Prior to model refinement, sequences extracted from the PDBs were aligned to the full-length human sequences for each protein using the pairwise alignment tool EMBOSS Water from EMBL-EBI. The model was refined with the sequence alignment by MODELLER (35) to minimize loop energies, fill sequence gaps, correct solubility mutations, model CUL5CTD from CUL1CTD, and to model RBX2 from RBX1. MODELLER 9.23 was run using the default automodel class. The model with the lowest MOLPDF score of the 10 models generated was selected as the final model. A model of ASB9-CRL with neddylated CUL5 was prepared by superimposing homologous domains of known structures to position each component in PyMOL. The CKB-ASB9-ELOB/C-CUL5NTD structure described in Lumpkin et al., submitted. To orient the CUL5CTD relative to CUL5NTD, CUL2 from CUL2-RBX1-ELOB/C-VHL (5N4W) was first aligned to residues 306-387 on CUL5NTD in the starting model using PyMOL. CUL5CTD from CUL5CTD(NEDD8)-RBX1 (3DQV) was aligned to residues 383-424 on CUL2 from CUL2-RBX1-ELOB/C-VHL

(5N4W). Following orientation of the two halves of CUL5, CUL2-RBX1-ELOB/C-VHL was removed from the model. Residues 208-238 of RNF4 in RNF4-UBE2D1-Ub (4AP4) were aligned to RBX1 to position UBE2D1-Ub, and RNF4 was removed. CDC34 (6NYO) was aligned with UBE2D1, and UBE2D1 was removed to yield a model containing CDC34. Prior to model refinement, sequences extracted from the PDBs were aligned to the full-length Human sequences for each protein using the pairwise alignment tool EMBOSS Water from EMBL-EBI. The model by MODELLER (35), using the sequence alignments, to minimize loop energies, fill sequence gaps, correct solubility mutations, and to model RBX2 from RBX1. MODELLER 9.23 was run using the default automodel class. The model with the lowest MOLPDF score of the 10 models generated was selected as the final model.

ARIH2 was aligned to the sequence of ARIH1 extracted from the crystal structure of the ARIH1-E2L3-Ub complex (PDB 5UDH) using the pairwise alignment tool EMBOSS Water from EMBL-EBI. Zinc chelating residues C186, C189, C203, H205, C208, C211, C231, C236, C276, C281, C297, C299, C304, C307, H312, C317, C344, C347, C357, C362, C367, C372, C375, H382, and C389 in ARIH1 correspond to residues C139, C142, C156, H158, C161, C164, C183, C188, C228, C233, C249, C252, C257, C260, H265, C270, C297, C300, C310, C315, C318, C323, C326, H333, and C340 in ARIH2. The two protein sequences were aligned so that the Zinc-chelating residues were exactly aligned. analysis of the ARIH2 sequence predicted residues 66-83, 88-97, 102-111, 113-120, 162-174, 194-200, 204-220, 257-260, 271-277, 281-293, 349-400, 404-434, 440-491 to be helical. Residues 66-84, 88-98, 102-110, 113-120, 148-150, 162-172, 194-200, 204-221, 270-281, 361-399, 404-433, and 438-491 were modeled as helical from alignment with ARIH1. Of these, only two helical regions were predicted and not modeled, and one region was

modeled as helical that wasn't predicted to be. Substitution of ARIH2 for ARIH1 gave the structure of ARIH2-E2L3-Ub based on the structure of the complex (PDB 5UDH).

9. Docking the N-terminus of ARIH2 into ASB9-CRL.

The flexible N-terminus of ARIH2 was docked onto ASB9-CRL using GalaxyWEB PepDock (22), CABS-dock (23), and HPEPDOCK (21). These docking programs were selected for their ability to model long peptides (>20 aa) in large proteins (>300 aa) and to discriminately score various models. PepDock uses template structures of peptide-protein interactions to predict binding interfaces between the given peptide sequence and protein structure and scores the models based on protein and interaction similarities to the template. CABS-dock simulates and clusters trajectories of the given peptide sequence interacting with the submitted protein structure, returning the predominant clusters with RMSD scores, cluster density, and contact maps. HPEPDOCK performs blind peptide-protein docking of a given sequence to a submitted structure and returns models with calculated docking energies.

The peptide 'DSNEEDYDPNCEEEEEEEEDDPGDIEDYYV', corresponding to residues 11-40 of ARIH2, was docked onto CUL51-780-RBX2 from the post-neddylation homology model, using PepDock. 10 models were generated from four protein-peptide template structures.

The peptide 'DSNEEDYDPNCEEEEEEEEDDPGDIEDY', corresponding to residues 11-38 of ARIH2 (truncated due to peptide length limitations), was docked onto CUL5300-780 (truncated due to protein length limitations), from the post-neddylation homology model, using CABS-dock. Fifty Simulation cycles were performed, and 10 models were generated. The model with the highest cluster density and no steric clashes with CUL5 was selected as the best model.

The peptide 'MSVDMNSQGSDSNEEDYDPNCEEEEEEEEDDPGDIEDYYVGVASD

VEQQGADAFDPE', corresponding to residues 1-57 of ARIH2, was docked onto CUL5(NEDD8)-RBX2, from the post-neddylation homology model, using HPEPDOCK. Ten models were generated.

10. Building ARIH2 onto ASB9-CRL.

The best CABS-dock model, consisting of residues 11-38 of ARIH2 bound to CUL5, was selected to begin building ARIH2 onto the full ASB9-CRL homology model. Residues 34-60 of ARIH2 were modeled using PEP-FOLD3 (24). The model with the lowest sOPEP energy score was chosen. The docked peptide, the PEP-FOLD3 peptide, and the homology model for ARIH2 (as described above) were used to model ARIH2 onto the ASB9-CRL homology model (as described above). MODELLER 9.23 was run using the default automodel class. The model with the lowest MOLPDF score of the 10 models generated was selected as the final model.

C. Results

1. Assembly of an active ASB9-CRL from purified proteins in vitro.

Using a process of strategic co-expression and mixing of crude E. coli lysates, we were able to assemble the following partial and complete ASB9-CRL protein complexes; CKB-ASB9-ELOB/C-CUL5(+/-NEDD8)-RBX2, CUL5(NEDD8)-RBX2, ARIH2-CUL5(NEDD8)-RBX2, ARIH2 (Fig 5.1B, 5.2). Neddylation did not require the CKB-ASB9 components and proceeded efficiently if E. coli lysates containing CUL5, NAE1/UBA3, and NEDD8 were combined and co-purified on Ni-NTA. Following dialysis, UBE2F was added and full neddylation occurred within 30 min at 4°C.

To demonstrate ubiquitylation activity of the CKB-ASB9-ELOB/C-CUL5 (+/-NEDD8)-RBX2, we initially added E2D2 and Ub and tested for ubiquitylation of CKB. Taking advantage of the fact that trypsin leaves a diglycyl fragment on the lysine that bore the Ub (19), we were able

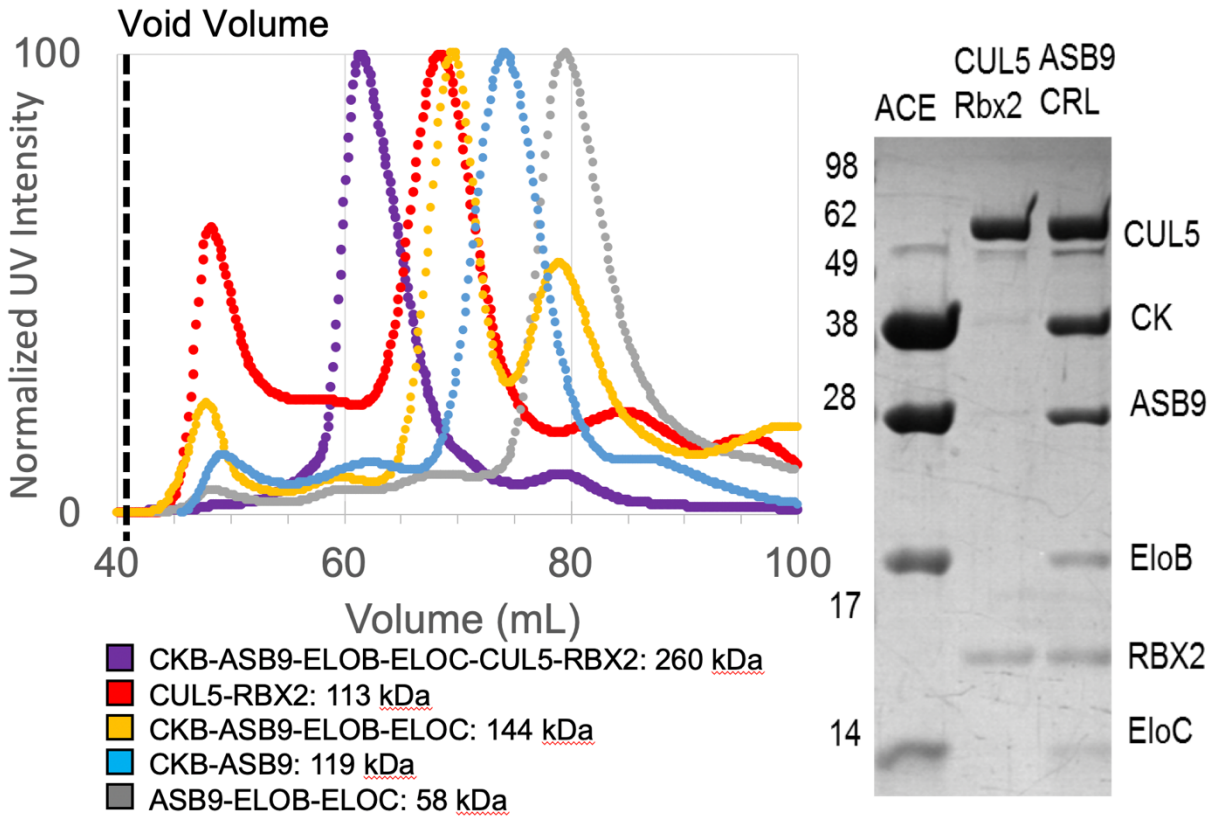

Figure 5.2: **Purification and Separation of the ASB9-CRL**. A) Elution profiles of the various complexes B) SDS PAGE showing the purity of the complexes and the presence of the subunits present in each.

to observe both CKB ubiquitylation and autoubiquitylation of the ligase. Because the reactions were performed in vitro with purified proteins, we were able to quantify the ratio of ubiquitylated to non-ubiquitylated lysines. In the non-neddylated CUL5 sample, we observed prevalent modification of CUL5 K724 (observed peptide: T715QEAIIQIMK(+114.04)MR726), the site of neddylation. Coomassie-stained gels and anti-Ub blotting showed that when CKB was absent, the band for ASB9 disappeared becoming polyubiquitylated and the band for CUL5 also had multiple Ub attachments. When CKB was present, modification of ASB9 was much reduced, and CUL5 migrated primarily as a monoubiquitylated species (Fig 5.3). To disentangle the role of NEDD8 vs Ub attachment at K724, we prepared the K724R mutant. This mutant was able to facilitate ubiquitylation by E2D2 nearly as well as the wild type.

2. ARIH2 promotes CKB ubiquitylation.

Although we were able to observe ubiquitylation of CKB *in vitro* with only E2D2 present, the reaction seemed very inefficient and the yields very low. We therefore searched for a missing component. Substitution of E2D2 with alternative E2, CDC34 did not yield measurable amounts of CKB-Ub. Consistent with a recent report, addition of ARIH2 dramatically improved the yields of CKB-Ub (1). CKB is a dimer, and we observed 50% of the CKB being rapidly ubiquitylated by anti-CKB western blotting which showed high molecular weight CKB bands only in the presence of ARIH2 (Fig 5.4A). Coomassie Blue-stained SDS-PAGE of the reactions showed that the ratio of CKB, which is a dimer, to ASB9 remained 2:1 until ARIH2 was added at which point one equivalent of CKB disappeared resulting in a ratio of CKB:ASB9 of 1:1, apparently due to ubiquitylation of CKB (Fig 5.4B , C). Analysis of MS/MS data collected on the highest molecular weight band revealed only the presence of CKB and Ub in different ratios with the highest ratio of Ub to CKB present when ARIH2 and either CDC34 or E2D2 were present (Figure 5.4D, Table

5.1). By measuring the ratio of ubiquitylated lysine-containing peptides to the same peptides without ubiquitylation, we could show that CKB was most highly modified on K45 and K381 but modification was also observed on K101 and K107 (Table 5.2). In addition, we could determine that ARIH2-CDC34 produced primarily K48-linked Ub chains whereas ARIH2-E2D2 produced a mixture of K48 and K63-linked chains (Table 5.3).

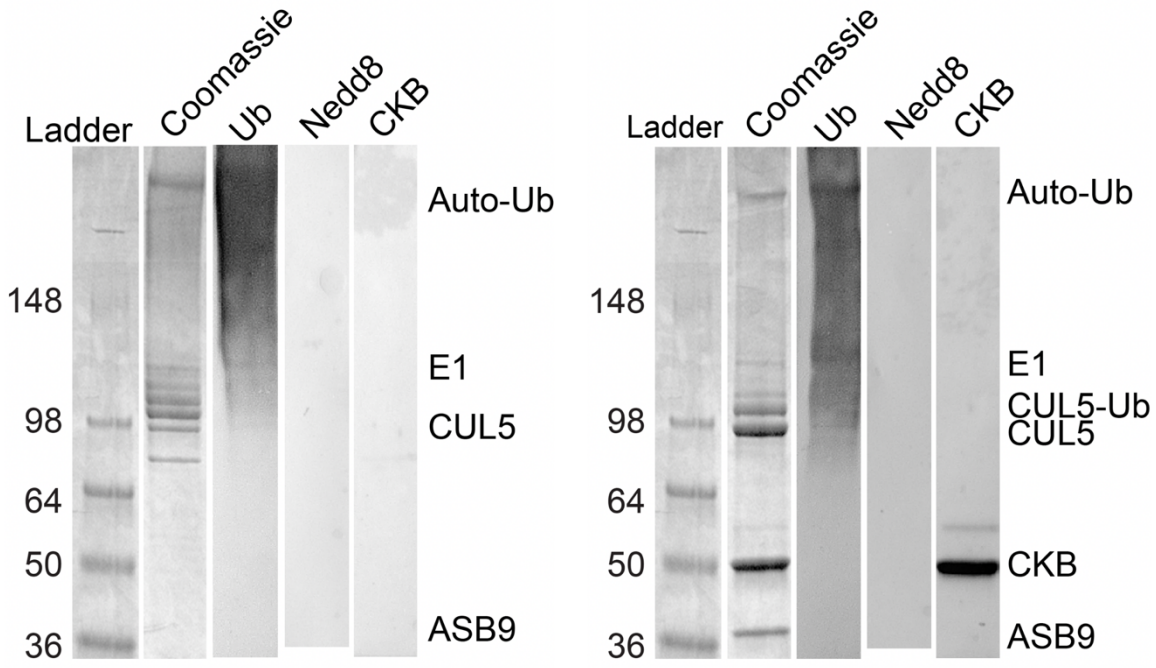

Figure 5.3: **Auto-ubiquitylation Assay**. Left, Ubiquitylation Assay in the absence of the substrate, CKB. Right, Ubiquitylation Assay in the presence of the substrate, CKB.

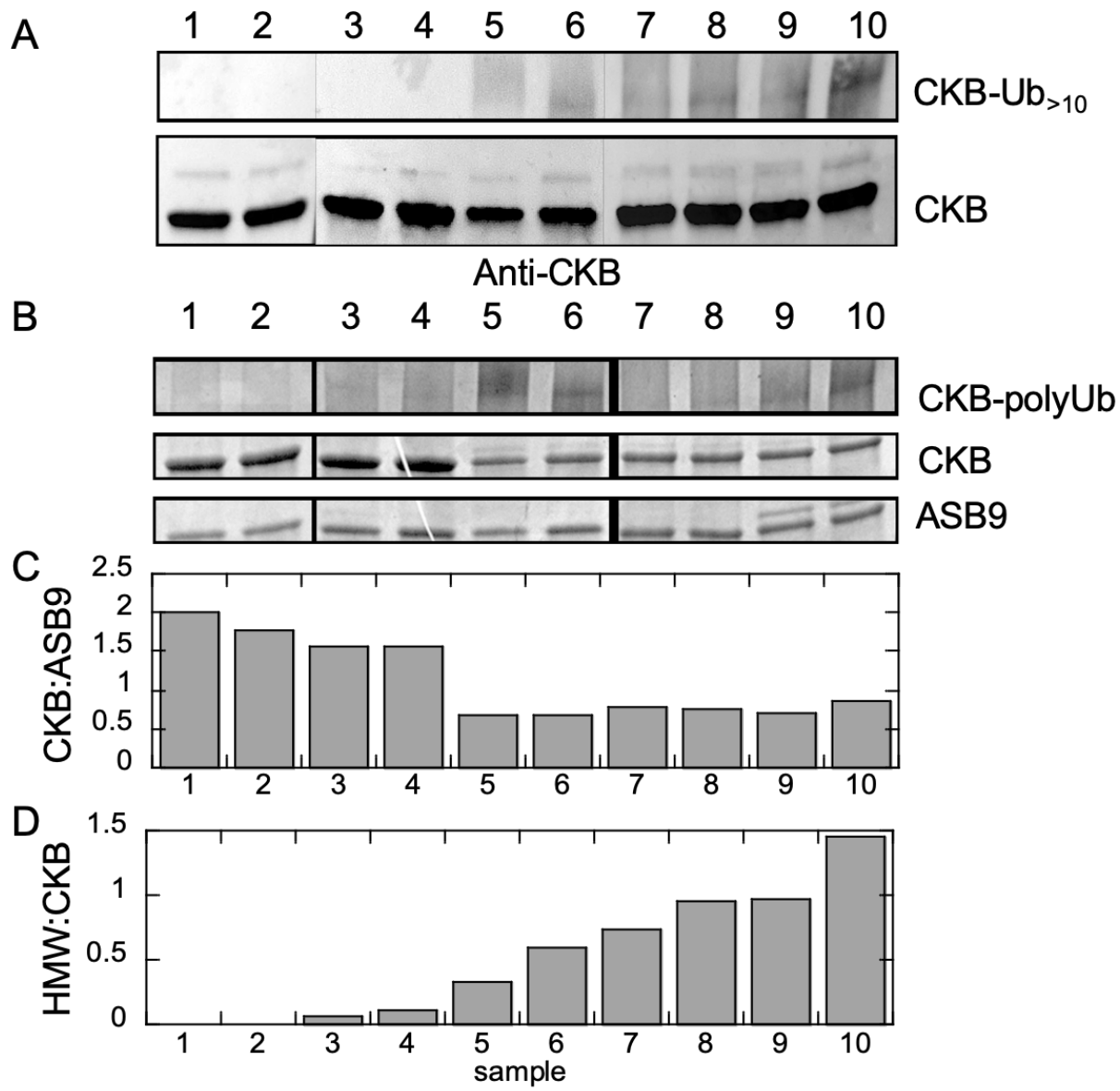

Figure 5.4: **In vitro ubiquitylation assays showing AriH2 induced ubiquitylation**
A) Anti-CKB blot for reactions containing various components reveal the most efficient combination of components for polyubiquitylation. Lanes 1, 3, 5, 7, 9 contain CKB-ASB9-ELOB/C-CUL5-RBX2. Lanes 2, 4, 6, 8, 10 contain CKB-ASB9-ELOB/C-CUL5(NEDD8)-RBX2. Lane-specific additional components are: E2D2 (lane 1, 2), CDC34 (lanes 3, 4), ARIH2-E2L3 (lanes 5, 6), ARIH2-E2L3-E2D2 (lanes 7, 8), and ARIH2-E2L3-CDC34 (lanes 9, 10). B) Coomassie blue-stained SDS-PAGE of the same ubiquitylation reactions as in A show that in absence of ARIH2 (lanes 1-4) the stoichiometry of CKB:ASB9 is 2:1 indicating very little consumption of CKB to produce CKB-polyUb. In contrast, in the presence of ARIH2 (lanes 5-10) the CKB:ASB9 stoichiometry is 1:1 indicating the consumption of 1 equivalent of CKB to produce CKB-polyUb. C) ImageJ analysis of the Coomassie blue-stained gel and the anti-CKB blot. The intensities of each band were calculated and the ratio of the CKB:ASB9 is plotted. D) Image J analysis was also used to quantify the intensity of the high molecular weight CBK-polyUb band compared to the CKB band.

Table 5.1: **Lysine Ubiquitylation counts from MS**

	K45	K101	K107	K381
ACECR-Ari	68 %	22 %	7 %	32 %
ACECR-Ari-E2D2	37 %	1 %	0 %	40 %
ACECR-Ari-CDC34	47 %	13 %	5 %	28 %

Table 5.2: **Poly-Ubiquitylation counts from MS**

	K33	K48	K63
ACECR-Ari	4 %	23 %	11 %
ACECR-Ari-E2D2	1 %	23 %	26 %
ACECR-Ari-CDC34	2 %	25 %	11 %

Table 5.3: **Individual Gel bands**

	CKB	Ub
ACECR-Ari	37%	61%
ACECR-Ari-E2D2	5%	94%
ACECR-Ari-CDC34	7%	93%

3. Models of the structures of the neddylated ASB9-CRL with ARIH2 bound.

To gain an understanding of how the ASB9-CRL might be transferring Ub to these sites on CKB, we first labeled the modified lysines on the 4.1 Å cryoEM structure of CKB-ASB9 (Lumpkin et al., submitted). It is interesting to note that there are many other surface lysines on CKB that we did not observe to be modified indicating that the ligase specifically modifies certain lysines. This is in line with our observation that only one subunit of CKB gets modified (see above). Using a combination of cryoEM data and homology modeling, we previously obtained a model of the CKB-ASB9-ELOB/C-CUL5-RBX2 complex (Lumpkin et al., submitted). Here, we further modeled the neddylated ASB9-CRL complex.

Because we observed a requirement for ARIH2 in high efficiency ubiquitylation of CKB, we decided to also attempt to build a model of ARIH2 and to dock it to CUL5. ARIH1 and ARIH2 have a sequence identity of 33% and a similarity of 54.9%. Both proteins have two RING domains and an In-Between-RING domain that each chelate two Zinc atoms. Care was taken to ensure that the Zn-coordinating residues were aligned between ARIH2 and ARIH1 to preserve this functionally relevant structure. A homology model was generated from the sequence alignment between ARIH1 (5UDH) and ARIH2 using MODELLER. NetSurfP 2.0 (20). Only two helical regions were predicted and not modeled, and one region was modeled as helical that wasn't predicted to be. The prediction did not identify β-sheets in the structure consistently, most likely due to the short sequence lengths. While sequence similarity was only 54.9%, the 74% of buried residues in the model were enriched in the residues were similar in the alignment between ARIH1 and ARIH2, supporting the validity of the three-dimensional model. (Fig. 5.5A)

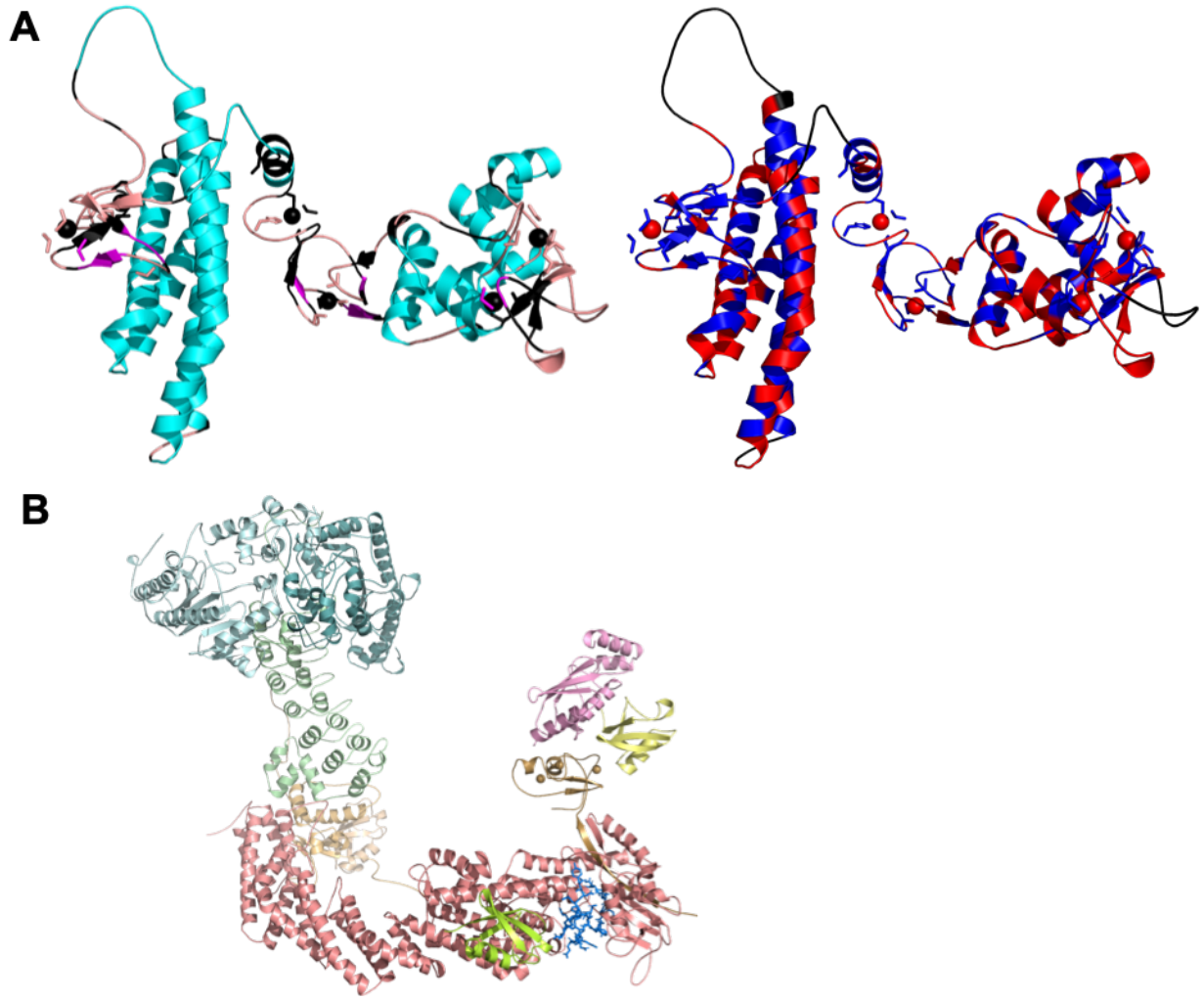

Figure 5.5: **Homology Modeling and Docking of ARIH2**. A) A homology model of ARIH2 was constructed from ARIH1 as described in the Results section. The models are colored according to secondary structure, with helices in cyan, sheets in yellow, and loops in salmon. Residues that are similar or identical between ARIH1 and ARIH2 in blue, and non-conserved residues in red. B) Model of the neddylated ASB9 CRL modified by attaching NEDD8 to the model from Lumpkin et al., submitted. Molecular docking of the ARIH2 disordered N-terminal residues 11-38 using CABS-dock resulted in a highly populated solution in which the ARIH2 segment docked into the positively-charged channel in CUL5 that is predicted to be more open in CUL5(NEDD8) (3DPL, 3DQV).

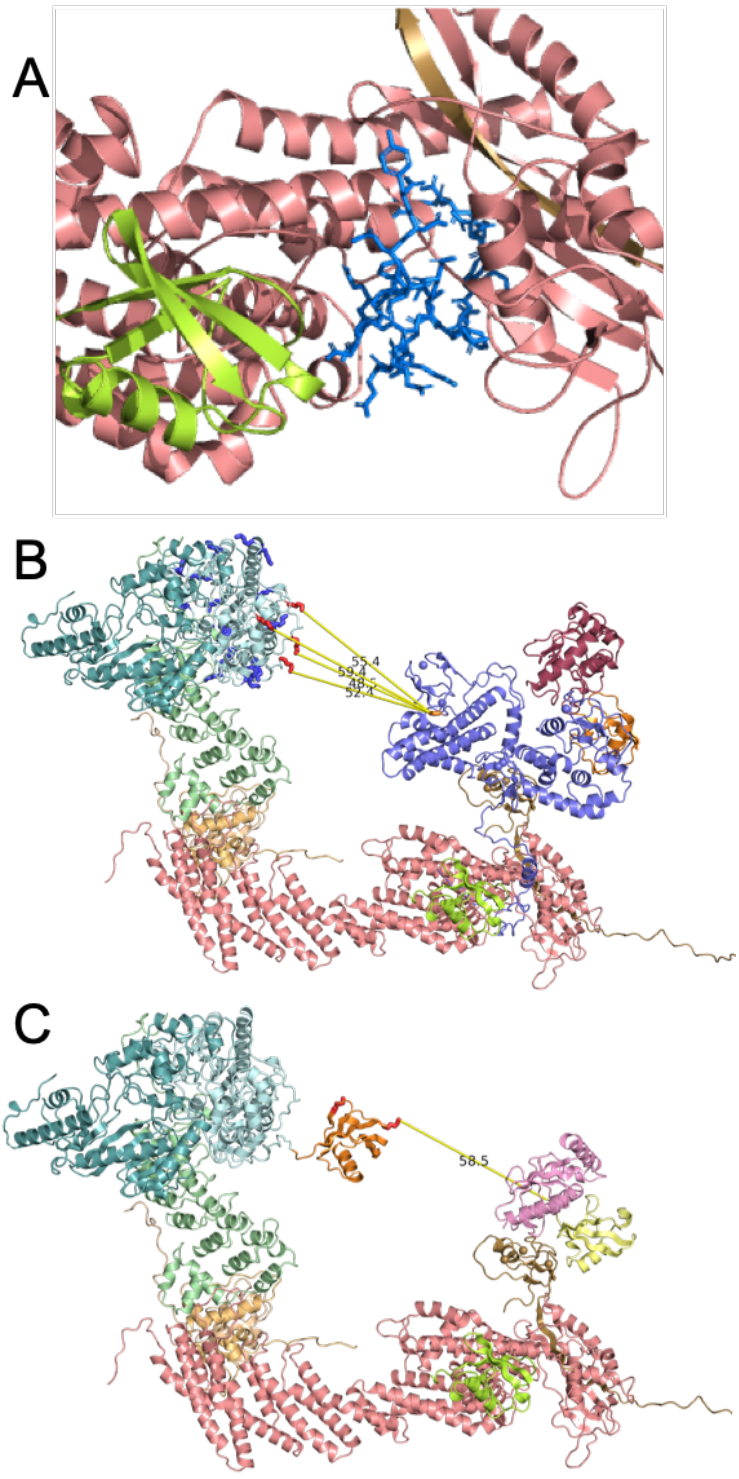

Figure 5.6: Homology model of AriH2 and docking to CUL5. A) Molecular docking of the ARIH2 disordered N-terminal residues 11-38 using CABS-dock resulted in a highly populated solution in which the ARIH2 segment docked into the positively-charged channel in CUL5 that is predicted to be more open in CUL5(NEDD8) (3DPL, 3DQV). B) Distances from the catalytic cysteine on ARIH2 to modified lysines on CKB. C) Distance from the catalytic cysteine on RBX2 to K48 on Ub.

The next step in building the model of the ASB9-CRL-bound ARIH2 was to dock the N-terminal domain of ARIH2 to CUL5. Acidic residues 22-31 of ARIH2 were targeted for docking to CUL5 based on mutagenesis experiments that identified a basic patch on CUL5 that was necessary for ARIH2-CUL5 protein binding (7). Attempts to perform this docking with HPEPDOCK (21) yielded 10 structures with little in common with each other and with insignificant contact with CUL5. Docking with PepDock (22) produced more meaningful results, identifying a particular region of CUL5CTD to which ARIH2 residues 11-40 preferentially docked. However, the Estimated Accuracy score for all of the models was less than 0.4, indicating that less than 40% of the residue interactions in each model were likely to be correct. Furthermore, many of the docked structures seemed improbable because the ARIH2 sequence was threaded through CUL5. In addition, the ARIH2 sequences that were not threaded had contacts with CUL5 that did not match previous mutagenesis. Finally, we performed docking with CABS-Dock (23) which resulted in 1000 models that formed 10 clusters allowing us to identify the top ten models. Of these top ten models, eight fit into the same channel on CUL5, near the basic patch predicted to be involved in binding (7). (Fig 5.6A, 5.5B) This interaction was later validated by the HDX-MS data presented here.

The structure of residues 34-60 of ARIH2 was predicted using PEP-FOLD3 (24) with the default parameters. The model with the lowest sOPEP score predicted residues 34-50 to be helical. To build the full model of CKB-ASB9-ELOB/C-CUL5(NEDD8)-RBX2-ARIH2-E2L3-Ub, the sequences of the docked 11-38 residues from CABS-dock, the predicted residues 34-60 from PEP-FOLD3, and the homology model of ARIH2 from MODELLER were aligned to the full sequence of ARIH2. This alignment was used to model the full ARIH2 protein assembled with CKB-ASB9-ELOB/C-CUL5(NEDD8)-RBX2. The final model of the ARIH2-bound ASB9 CRL shows the

ARIH2 catalytic cysteine at 48 Å from K48 on CKB, 40Å from K101 on CKB, 52 Å from K107 on CKB, and 50 Å from K381 on CKB. (Fig 5.6B). Priming of CKB by ARIH2 with a single ubiquitin significantly decreases the distance to E2D2 on RBX2. (Fig 5.6C) Flexibility in the N-terminus of ARIH2 and RBX2 facilitate the ubiquitylating activity of the ASB9-CRL by increasing the reach of the catalytic domains. (Fig. 5.7)

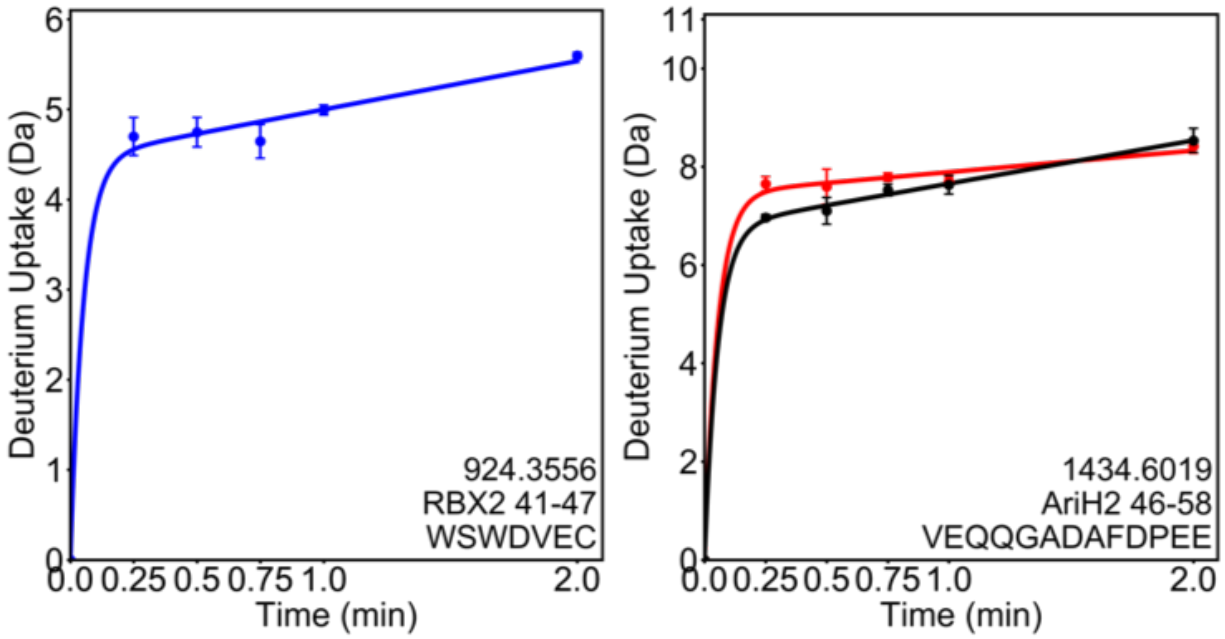

Figure 5.7: Flexible Regions of RBX2 and ARIH2. A disordered, highly flexible region is present in the N-terminal domains of both ARIH2 and RBX2, as revealed by HDX-MS.

4. ARIH2 Binds neddylated CUL5.

ARIH2 and CUL5 were evaluated for binding through size exclusion chromatography (Fig 5.8). ARIH2 was mixed with an equimolar amount of either ELOB/C-CUL5(NEDD8)-RBX2 or ELOB/C-CUL5- RBX2 and then the resulting mixture was separated by size exclusion. While the sample with neddylated CUL5 eluted as a mostly homogenous peak, the non-neddylated sample eluted in two clearly separate peaks. SDS-PAGE of the SEC fractions indicated that ARIH2 coeluted preferentially with neddylated CUL5.

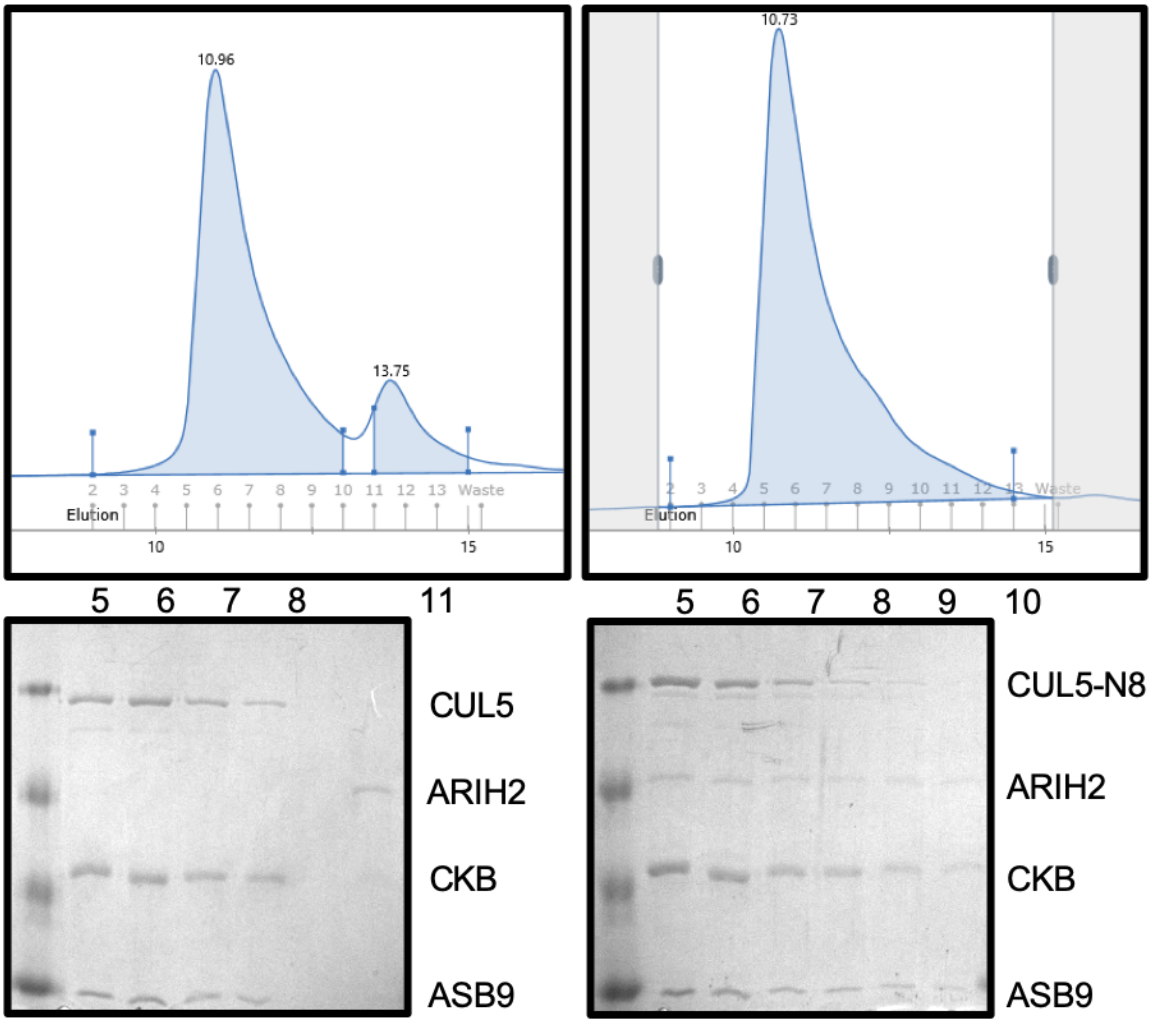

Figure 5.8: Size Exclusion Chromatography of the ARIH2-ASB9-CRL interaction. Complexation and subsequent separation of ARIH2 and the ASB9-CRL with and without Neddylation demonstrates that co-elution of the dual ligase complex together from size exclusion depends on the neddylation of CUL5.

5. Neddylation restructures CUL5 and RBX2.

Neddylation of CUL5 is known to play an important role in the activation and regulation of CUL-RING ligases (25). Crystal structures of the C-terminal domain of CUL5 before and after neddylation show that the modification reorients the C-terminal domain of CUL5 and increases the angle between CUL5 and RBX2 (3DPL, 3DQV). We performed HDX-MS experiments on CKB-ASB9-ELOB/C-CUL5-RBX2 and CKB-ASB9-ELOB/C-CUL5(NEDD8)-RBX2 to determine the protein-protein interfaces and any long-range conformational alterations that may occur upon neddylation. We found CUL5 residues 439-450, 489-496, 497-504, and 698-709 exchanged more when CUL5 was neddylated (Fig 5.9). We found RBX2 residues 52-58 exchanged less when CUL5 was neddylated. Interestingly, NEDD8 attachment to CUL5 did not cause any observable protection, at a putative interface and instead it causes increases in exchange on the C-terminal domain of CUL5 at residues 439-450, 489-504, and 698-709. This may indicate that the NEDD8 molecule is flexibly tethered to K724 on CUL5.

Additionally, neddylation of CUL5 causes a decrease in exchange on RBX2 at residues 52-58. This sequence is solvent exposed both before and after neddylation (3DPL, 3DQV). The structure in the neddylation transition state shows that this same region is part of the binding interface for the NEDD8 E2 (4P50). These results suggest that neddylation may induce long-range conformational alterations at the E2 binding site on RBX2.

6. ARIH2 Binding to the ASB9-CRL causes transiently lower exchange in CUL5 and significant increases in exchange on ARIH2.

HDX-MS experiments were used to probe conformational changes in ARIH2 and in CUL5 upon interaction between ARIH2 and the ASB9-CRL. We compared free ARIH2 and free ELOB/C-CUL5(NEDD8)-RBX2, to the complex between them (ELOB/C-CUL5(NEDD8)-

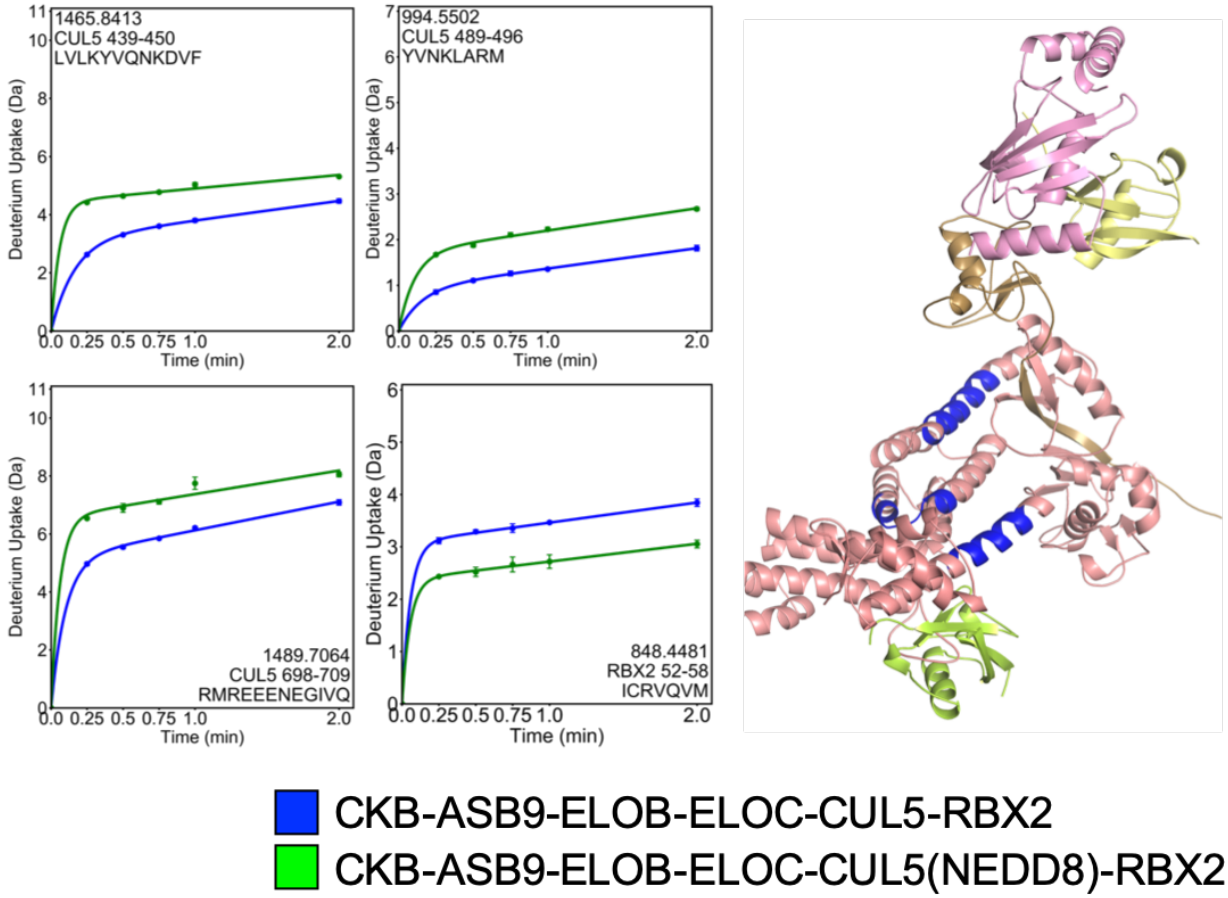

Figure 5.9: HDX-MS reveals how the ASB9-CRL responds to neddylation of CUL5. Crystal structures of CUL5-RBX1 before and after NEDD8ylaton (3DPL, 3DQV) show that NEDD8ylation results in a 180-degree rotation in the CUL5 stalk and an increased angle between CUL5 and RBX2. We compared ELOB/ELOC-CUL5-RBX2 +/- NEDD8, ASB9-ELOB/ELOC-CUL5-RBX2 +/- NEDD8, and CKB-ASB9-ELOB/ELOC-CUL5-RBX2 +/- NEDD8 using HDX-MS. NEDD8ylation in all of these complexes results in increased exchange across three sites on the C-terminal domain of CUL5 as well as decreased exchange on one region of RBX2 that binds to UBE2F-NEDD8 in the pre-Neddylated state. (I only want to compare two states in the paper, but I think it is interesting that the NEDD8ylation effect is not dependent on the substrate receptor complex. So maybe we can have some additional plots or info in the Supplementary).

RBX2-ARIH2). Unfortunately, the N-terminal 38 residues were not covered by HDX-MS, and these residues contain a highly acidic patch of residues that are believed to interact with a positively charged patch on the C-terminus of CUL5. We did not detect any other surface of ARIH2 that showed decreased exchange upon binding to the CRL (Fig 5.10). However, several regions of ARIH2 (residues 361-367, 369-381, 382-395, 404-413) showed major increases in deuterium exchange in the comparison of free ARIH2 to ELOB/C-CUL5(NEDD8)-RBX2. These regions surround the domain of ARIH2 that has been previously reported to be autoinhibited prior to CUL5 binding (7). These results strongly indicate that the release of ARIH2 autoinhibition upon interaction with CUL5 involves increased dynamics/exposure of residues 361-367, 369-381, 382-395, 404-413 in ARIH2.

Several regions on CUL5 showed decreased exchange upon ARIH2 binding. The basic patch on the $CUL5_{CTD}$ has previously been suggested to interact with the acidic tails of both CDC34 and ARIH2. Mutations at R417, R417, K418, K423, K424, K676, K679, K682, and R683 on CUL5 affected ARIH2 binding (7). Here we compared the interaction of ARIH2 with CUL5 by analyzing the ELOB/C-CUL5(NEDD8)-RBX2 and ELOB/C-CUL5(NEDD8)-RBX2-ARIH2 complexes by HDX. The regions 333-344, 366-383, 415-429, 439-450, 465-471, 477-488, 528-544, and 639-648 of CUL5 all show decreases in exchange in the presence of ARIH2, however this effect is transient, disappearing by the 2 min timepoint. (Fig 5.11) Notably, neddylation caused increased exchange in CUL5 at regions 439-450 and 489-504, at or between the same areas affected by ARIH2 binding.

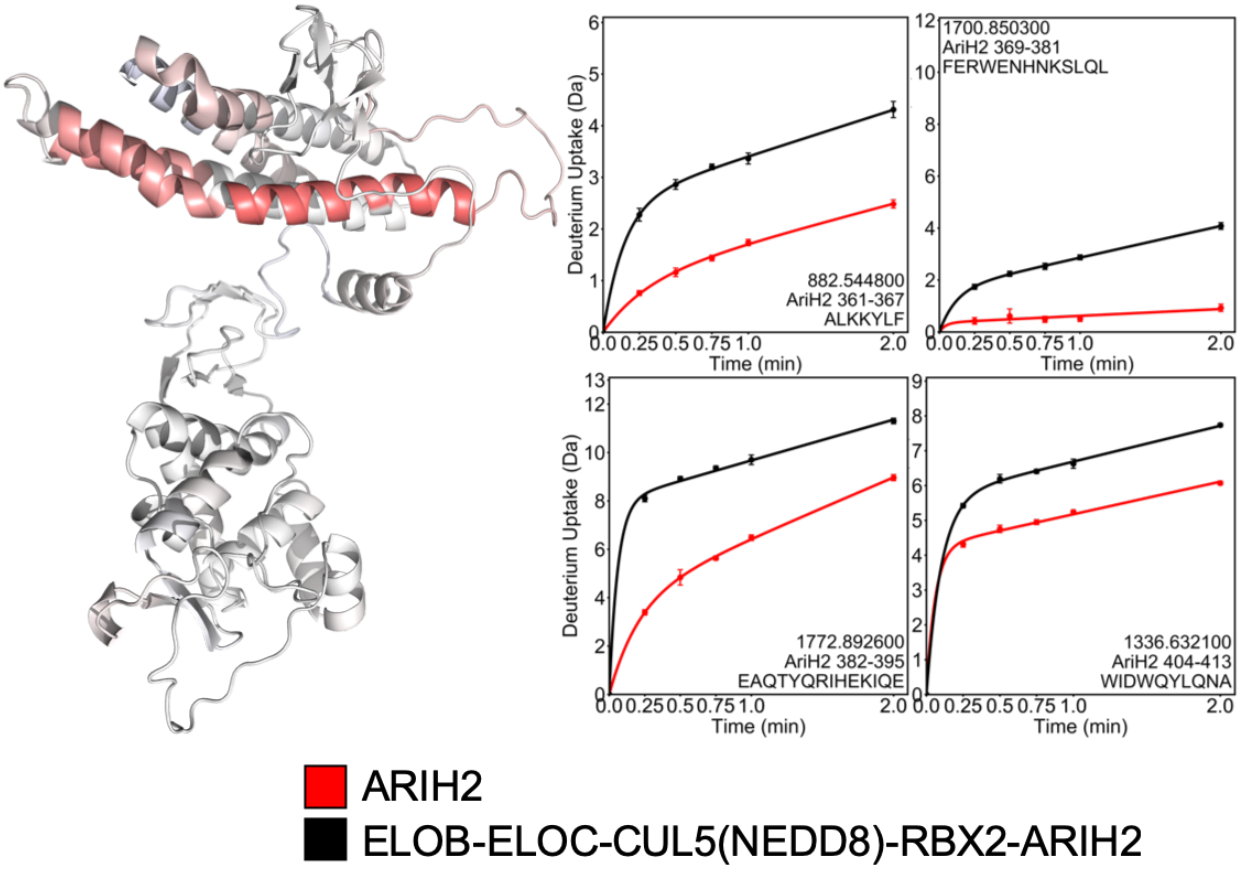

Figure 5.10*:* HDX-MS on AriH2. HDX-MS studies of ARIH2 in complex with the ASB9-CRL show a small transient protection of the N-terminal sequence of ARIH2. More notably, the association of ARIH2 with ASB9-CRL results in major increases in exchange in ARIH2 throughout the auto-inhibited domain at residues 338-416.

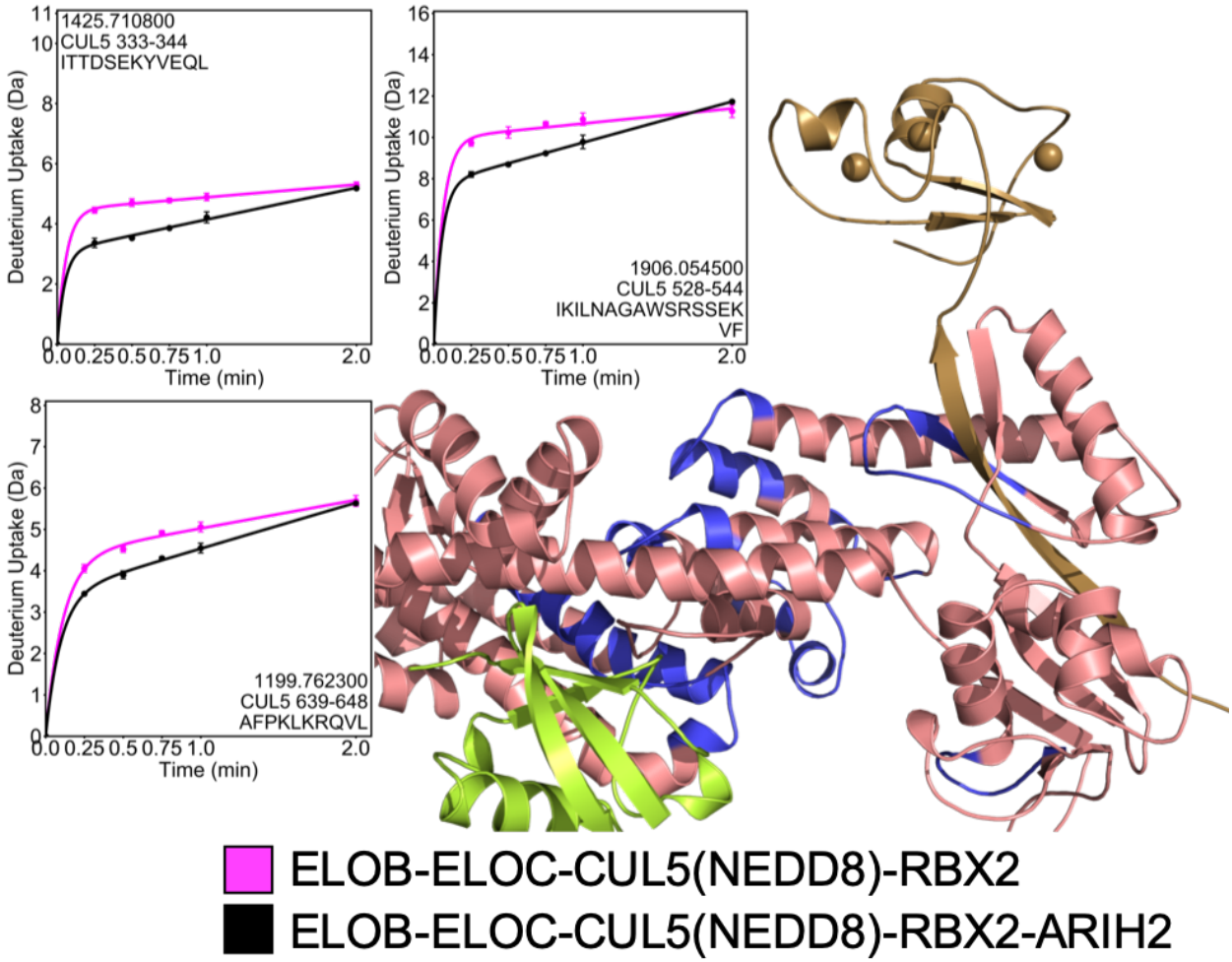

Figure 5.11: **HDX-MS on CUL5**. HDX-MS studies of ASB9-CRL in complex with the ARIH2 show peptides with decreased deuterium exchange as compared to the free ASB9-CRL throughout the C-terminus of CUL5, many of which cover the same regions as peptides with increased deuterium exchange upon NEDD8ylation.

D. Discussion

We were able to reconstruct the full ASB9-CRL from recombinant proteins in vitro and identify the smallest set of proteins that were sufficient for CKB ubiquitylation. While neddylation of CUL5 increased the amount of ubiquitylation by E2D2, neither E2D2 nor CDC34 significantly ubiquitylated CKB. The addition of ARIH2-E2L3 to the ASB9-CRL triggered ubiquitylation of approximately half of the CKB, detected by mass spectrometry and visualized on Coomassie-stained gels and anti-CKB blots. In the presence of the ASB9-CRL and ARIH2, CKB was found to be ubiquitylated at K45, K101, K107, and K381. Mass spectrometry analysis revealed that CDC34 predominantly makes K48 chains, while E2D2 and ARIH2-E2L3 yield multiple polyubiquitylation chain types, including K11, K33, K48, and K63. The inclusion of CDC34 with ARIH2 tunes the chain specificity, increasing the ratio of K48 to other chain types formed. While only one-half of the CKB dimer becomes modified, the addition of CDC34 increases the level of polyubiquitylation on CKB. Furthermore, the high molecular weight species indicative of polyubiquitylated CKB converge at a particular molecular weight, suggesting that the distance between the E2 on RBX2 and the mono-ubiquitylated CKB regulates polyubiquitin chain length in the ASB9-CRL.

ASB9-CRL and ARIH2 were modeled individually and docking experiments were performed to identify the binding interface between the two ligases. ARIH2 was modeled from ARIH1, which lacks structure for the N-terminal residues that correspond to 1-57 of ARIH2, which have been suggested to facilitate binding to CUL5 (7). Docking residues 11-38 of ARIH2 revealed that acidic residues in that domain interact with basic residues on the CUL5¬CTD. Using this information and prediction of the structure between 38-57, the model of ARIH2 was modeled onto the ASB9-CRL. While the catalytic cysteine of E2D2-Ub on RBX2 is positioned 60-90 Å from

modified lysines on CKB, the catalytic cysteine on ARIH2 is 40-50 Å away from the modified lysines. Additionally, residues 47-60 are predicted to be unstructured, giving ARIH2 significant flexibility to reach further in order to perform ubiquitylation.

Comparison of several states of the ASB9-CRL revealed allosteric mechanisms that activate the ligase. Neddylation triggers increased exchange in multiple regions of the CUL5CTD, and decreased exchange in one region of the RBX2, residues 52-58. Our model of the ASB9-CRL-E2F-Nedd8 shows that this region of RBX2 may interact with the E2F-NEDD8 complex to facilitate Neddylation. These HDX results suggest that Neddylation allosterically changes RBX2, which is in agreement with previous studies suggesting that the RBX2's specificity for E2 enzymes changes following neddylation (26).

Comparison of free ARIH2 and the ASB9-CRL-ARIH2 complex revealed that the interaction between ARIH2 and the CUL5CTD allosterically changes the Ariadne domain of ARIH2 that is initially autoinhibited. These changes reveal the catalytic cysteine, allowing ARIH2 to ubiquitylate substrates on the ASB9-CRL. No regions of ARIH2 decreased in exchange upon interaction with ASB9-CRL, suggesting that the binding interface is at the N-terminus of ARIH2 that lacked coverage in out HDX experiments.

Comparison of free ASB9-CRL and the ASB9-CRL-ARIH2 complex identified several regions of decreased exchange that reveal the binding interface of ARIH2 on ASB9-CRL. The location of the observed changes matched previous observations that mutations to basic residues in the CUL5CTD¬ altered binding to ARIH2. The decreased exchange in the CUL5CTD was transient for all regions with the change. While ARIH2 co-elutes from size exclusion in complex with the neddylated ASB9-CRL, this suggests that ARIH2 has a flexible interaction with the CUL5CTD, thus allowing the interacting residues to slowly exchange over time.

F. Acknowledgements

Chapter V, is material that the dissertation author was the principal researcher and author of. The material has been submitted for publication. (**Lumpkin, R. J.,** Ahmad, A. S., Blake, R., Condon, C., and Komives, E. A. Triggering CUL5-Mediated Ubiquitinylation.)

G. References

1. D. Komander, M. Rape, The ubiquitin code. Annu Rev Biochem 81, 203-229 (2012).

2. M. Akutsu, I. Dikic, A. Bremm, Ubiquitin chain diversity at a glance. J Cell Sci 129, 875-880 (2016).

3. T. Kamura, K. Maenaka, S. Kotoshiba, M. Matsumoto, D. Kohda, R. C. Conaway, J. W. Conaway, K. I. Nakayama, VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. Genes Dev 18, 3055-3065 (2004).

4. M. L. Heuze, F. C. Guibal, C. A. Banks, J. W. Conaway, R. C. Conaway, Y. E. Cayre, A. Benecke, P. G. Lutz, ASB2 is an Elongin BC-interacting protein that can assemble with Cullin 5 and Rbx1 to reconstitute an E3 ubiquitin ligase complex. J Biol Chem 280, 5468-5474 (2005).

5. K. K. Dove, R. E. Klevit, RING-Between-RING E3 Ligases: Emerging Themes amid the Variations. J Mol Biol 429, 3363-3375 (2017).

6. D. M. Duda, J. L. Olszewski, J. P. Schuermann, I. Kurinov, D. J. Miller, A. Nourse, A. F. Alpi, B. A. Schulman, Structure of HHARI, a RING-IBR-RING ubiquitin ligase: autoinhibition of an Ariadne-family E3 and insights into ligation mechanism. Structure 21, 1030-1041 (2013).

7. I. R. Kelsall, D. M. Duda, J. L. Olszewski, K. Hofmann, A. Knebel, F. Langevin, N. Wood, M. Wightman, B. A. Schulman, A. F. Alpi, TRIAD1 and HHARI bind to and are activated by distinct neddylated Cullin-RING ligase complexes. EMBO J 32, 2848-2860 (2013).

8. D. C. Scott, D. Y. Rhee, D. M. Duda, I. R. Kelsall, J. L. Olszewski, J. A. Paulo, A. de Jong, H. Ovaa, A. F. Alpi, J. W. Harper, B. A. Schulman, Two Distinct Types of E3 Ligases Work in Unison to Regulate Substrate Ubiquitylation. Cell 166, 1198-1214 e1124 (2016).

9. C. A. Andresen, S. Smedegaard, K. B. Sylvestersen, C. Svensson, D. Iglesias-Gato, G. Cazzamali, T. K. Nielsen, M. L. Nielsen, A. Flores-Morales, Protein Interaction Screening for the Ankyrin Repeats and Suppressor of Cytokine Signaling (SOCS) Box (ASB) Family

Identify Asb11 as a Novel Endoplasmic Reticulum Resident Ubiquitin Ligase. Journal of Biological Chemistry 289, 2043-2054 (2014).

10. J. Li, A. Mahajan, M. D. Tsai, Ankyrin repeat: a unique motif mediating protein-protein interactions. Biochemistry 45, 15168-15178 (2006).

11. L. K. Mosavi, D. L. Minor, Jr., Z. Y. Peng, Consensus-derived structural determinants of the ankyrin repeat motif. Proc Natl Acad Sci U S A 99, 16029-16034 (2002).

12. M. A. Debrincat, J. G. Zhang, T. A. Willson, J. Silke, L. M. Connolly, R. J. Simpson, W. S. Alexander, N. A. Nicola, B. T. Kile, D. J. Hilton, Ankyrin repeat and suppressors of cytokine signaling box protein asb-9 targets creatine kinase B for degradation. J Biol Chem 282, 4728-4737 (2007).

13. S. Kwon, D. Kim, J. W. Rhee, J. A. Park, D. W. Kim, D. S. Kim, Y. Lee, H. J. Kwon, ASB9 interacts with ubiquitous mitochondrial creatine kinase and inhibits mitochondrial function. BMC Biol 8, 23 (2010).

14. D. Balasubramaniam, J. Schiffer, J. Parnell, S. P. Mir, R. E. Amaro, E. A. Komives, How the ankyrin and SOCS box protein, ASB9, binds to creatine kinase. Biochemistry 54, 1673-1680 (2015).

15. J. R. C. Muniz, K. D. Guo, N. J. Kershaw, V. Ayinampudi, F. von Delft, J. J. Babon, A. N. Bullock, Molecular Architecture of the Ankyrin SOCS Box Family of CuI5-Dependent E3 Ubiquitin Ligases. Journal of Molecular Biology 425, 3166-3177 (2013).

16. J. M. Schiffer, R. D. Malmstrom, J. Parnell, C. Ramirez-Sarmiento, J. Reyes, R. E. Amaro, E. A. Komives, Model of the Ankyrin and SOCS Box Protein, ASB9, E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer. Structure 24, 1248-1256 (2016).

17. E. L. Huttlin, R. J. Bruckner, J. A. Paulo, J. R. Cannon, L. Ting, K. Baltier, G. Colby, F. Gebreab, M. P. Gygi, H. Parzen, J. Szpyt, S. Tam, G. Zarraga, L. Pontano-Vaites, S. Swarup, A. E. White, D. K. Schweppe, R. Rad, B. K. Erickson, R. A. Obar, K. G. Guruharsha, K. Li, S. Artavanis-Tsakonas, S. P. Gygi, J. W. Harper, Architecture of the human interactome defines protein communities and disease networks. Nature 545, 505-509 (2017).

18. R. Huttenhain, J. Xu, L. A. Burton, D. E. Gordon, J. F. Hultquist, J. R. Johnson, L. Satkamp, J. Hiatt, D. Y. Rhee, K. Baek, D. C. Crosby, A. D. Frankel, A. Marson, J. W. Harper, A. F. Alpi, B. A. Schulman, J. D. Gross, N. J. Krogan, ARIH2 Is a Vif-Dependent Regulator of CUL5-Mediated APOBEC3G Degradation in HIV Infection. Cell Host Microbe 26, 86-99 e87 (2019).

19. J. Peng, D. Schwartz, J. E. Elias, C. C. Thoreen, D. Cheng, G. Marsischky, J. Roelofs, D. Finley, S. P. Gygi, A proteomics approach to understanding protein ubiquitination. Nat Biotechnol 21, 921-926 (2003).

20. M. S. Klausen, M. C. Jespersen, H. Nielsen, K. K. Jensen, V. I. Jurtz, C. K. Sonderby, M. O. A. Sommer, O. Winther, M. Nielsen, B. Petersen, P. Marcatili, NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. Proteins 87, 520-527 (2019).

21. P. Zhou, B. Jin, H. Li, S. Y. Huang, HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. Nucleic Acids Res 46, W443-W450 (2018).

22. H. Lee, L. Heo, M. S. Lee, C. Seok, GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization. Nucleic Acids Res 43, W431-435 (2015).

23. M. Kurcinski, A. Badaczewska-Dawid, M. Kolinski, A. Kolinski, S. Kmiecik, Flexible docking of peptides to proteins using CABS-dock. Protein Sci 10.1002/pro.3771 (2019).

24. A. Lamiable, P. Thevenet, J. Rey, M. Vavrusa, P. Derreumaux, P. Tuffery, PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res 44, W449-454 (2016).

25. J. R. Lydeard, B. A. Schulman, J. W. Harper, Building and remodelling Cullin-RING E3 ubiquitin ligases. EMBO Rep 14, 1050-1061 (2013).

26. D. M. Duda, L. A. Borg, D. C. Scott, H. W. Hunt, M. Hammel, B. A. Schulman, Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. Cell 134, 995-1006 (2008).

27. D. Narang, W. Chen, C. G. Ricci, E. A. Komives, RelA-Containing NFkappaB Dimers Have Strikingly Different DNA-Binding Cavities in the Absence of DNA. J Mol Biol 430, 1510-1520 (2018).

28. P. R. L. Markwick, R. B. Peacock, E. A. Komives, Accurate Prediction of Amide Exchange in the Fast Limit Reveals Thrombin Allostery. Biophys J 116, 49-56 (2019).

29. K. M. Ramsey, H. E. Dembinski, W. Chen, C. G. Ricci, E. A. Komives, DNA and IkappaBalpha Both Induce Long-Range Conformational Changes in NFkappaB. J Mol Biol 429, 999-1008 (2017).

30. T. E. Wales, K. E. Fadgen, G. C. Gerhardt, J. R. Engen, High-speed and high-resolution UPLC separation at zero degrees Celsius. Anal Chem 80, 6815-6820 (2008).

31. Y. S. Choo, Z. Zhang, Detection of protein ubiquitination. J Vis Exp 10.3791/1293 (2009).

32. M. Scheffner, U. Nuber, J. M. Huibregtse, Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. Nature 373, 81-83 (1995).

33. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9, 671-675 (2012).

34. A. Villar-Garea, L. Israel, A. Imhof, Analysis of histone modifications by mass spectrometry. Curr Protoc Protein Sci Chapter 14, Unit 14 10 (2008).

35. A. Sali, T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234, 779-815 (1993).