CRYO-ELECTRON MICROSCOPY SINGLE PARTICLE STUDIES OF HUMAN CANCER TARGETS: UBIQUITIN-SPECIFIC PROTEASE 7 (USP7), USP28, AND KEAP1-CULLIN3-RBX1 E3 LIGASE MACHINERY

by

Corey A Moore

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Biological Sciences

West Lafayette, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Andrew Mesecar, Chair

Department of Biological Sciences, Chair of Biochemistry Department

Dr. William Cramer

Department of Biological Science

Dr. Wen Jiang

Department of Biological Sciences, Department of Computer Science

Dr. David Thompson

Department of Chemistry

Approved by:

Dr. Andrew Mesecar

This dissertation is dedicated to all those who have believed in me. The family and the friends that I have lost, and the dreams that have faded…I'll never forget them.

ACKNOWLEDGMENTS

I would like to acknowledge the many people who have supported me and stayed by my side throughout the struggle that graduate school was. First and foremost, I would like to acknowledge my mother, Joyce Moore. She raised me in a very unconventional way, but it was in this unconventional manner that hardened me; that made me the person I am today. Although I would like anything to re-live my past, I must acknowledge the bright spots and the dark ones, for they made me the man I am today.

I would like to acknowledge and thank Brenda González. Brenda was immensely helpful to my studies, and never stopped believing in me. I would like to acknowledge and thank Frank Vago for his assistance in cryo-EM data collection and analysis. I would like to acknowledge and thank Lyman Monroe for his thoughtful discussion and help with computational studies. I would like to acknowledge and thank Dr. Alix McCloskey, Dr. Vishak Raman, Dr. Henry Leggett, Jim Zhang, Dr. Jozlyn Clasman, Dr. Dan Cholger, Adam and May Hamdani, and many others for their support. All of you were always willing to listen to me and always helped me to feel better, even on my worst days.

I would like to acknowledge the current and past members of the Mesecar Lab for their support. I would like to acknowledge and thank Prof. Bill Cramer, Prof. Cynthia Stauffacher, Prof. Wen Jiang, Prof. Nick Noinaj, Prof. Carol Post, Prof. Michael Gribskov, Prof. Diasuke Kihara, Prof. David Thompson, Prof. Andy Mesecar, Prof. Ignacio Camarillo, Prof. Morris Levy and his better half Maria (Mechas) Levy, Nic Steussy, Thomas Klose, Valorie Bowman, Steve Wilson, Russell Logsdon for their kindness, support, guidance, and assistance in my research and in my education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

BME:	beta-mercaptoethanol
BSA:	bovine serum albumin
CD:	catalytic domain
CHAPS:	3-[(3-chloamidopropyl)dimethylammonio]-1-propanesulfonate
DSF:	differential scanning fluorimetry
DUB:	deubiquitinase; deubiquitinating enzyme
GO:	graphene oxide
HAUSP:	Herpesvirus-associated ubiquitin-specific protease (old name for USP7)
HUBL:	HAUSP ubiquitin-like domain (UBL)
ITC:	isothermal titration calorimetry
SEC:	size exclusion chromatography
TEVp:	Tobacco Etch Virus protease
TRAF:	tumor necrosis factor receptor-associated factor
Ub:	ubiquitin
Ub-AMC:	ubiquitin-7-amino-4-methylcoumarin
Ub-MESNa:	ubiquitin sodium 2-mercaptoethansulfonate
Ub-PA:	ubiquitin-propargylamine
Ub-Rho110:	ubiquitin-rhodamine 110
USP:	ubiquitin-specific protease
VPP:	Volta phase plate

ABSTRACT

The following work describes the methodology and materials used to study three human protein complexes involved in the etiology and progression of cancer. The first, ubiquitin-specific protease 7 (USP7) is an isopeptidase that employs a unique auto-regulatory mechanism. The second is another ubiquitin-specific protease, USP28, which forms higher order states in solution. Lastly, the third case was a protein complex that utilizes an oxidation-sensitive dimeric protein, Keap1, and two components of an E3 ligase – Cul3-Rbx1. Each of these studies involved overcoming unique challenges for cryo-EM sample optimization. Not all yielded the quality of data that would result in high-resolution (< 6 Å) densities. Despite this, new information was discovered about each system.

USP7 has a unique mechanism of intramolecular regulation that stems from a hypothesized tethered-rheostat, whereby the c-terminal distal domains activate the catalytic domain via a hypothetical wide degree of conformational movement. My cryo-EM work, done in collaboration with the Wen Jiang lab, is the first comprehensive structural data that provides structural evidence for the movement of the tethered-rheostat. The particle set showed a great degree of conformational heterogeneity, even after a strategy was employed with a chemically-modified ubiquitin substrate to ameliorate these issues. The data showed that during the ubiquitin-bound state, after the release of a hypothetical substrate, but prior to the release of mono-ubiquitin, the HUBL4-5 domains do not remain engaged with the catalytic domain. This information suggests a change to existing models of catalysis.

Additionally, the structural model built from the cryo-EM density has revealed an interfacial region between domains that were previously not thought to interact. This interfacial region between the TRAF domain and HUBL1-3 represents a candidate location of binding for a mixed, non-competitive inhibitor of USP7 previously identified in the lab. Enzyme kinetics, DSF, and Glide molecular docking experiments all yielded data that corroborate this idea.

Structural studies on USP28 have been difficult as the multi-domain enzyme adopts oligomers in solution and is generally not amenable to crystallographic analysis. Prior to the work described herein, the only structural data were a solution NMR structure describing a few alpha-helical motifs in the N-terminus. During my graduate studies, two articles were published of the USP28 catalytic domain crystallographic structure. Both corroborated the existence of a dimer.

The USP28 catalytic domain migrates during analytical gel filtration assays with the apparent molecular weight of a tetramer. Furthermore, glutaraldehyde crosslinking experiments show the catalytic domain appears to adopt a tetrameric state, like the USP25 tetramer. The USP25 tetramer was published alongside the USP28 catalytic domain dimer, concluding that a USP28 tetrameric state was not observed. Upon cryo-EM data collection and single particle analysis, it was observed that the compositional heterogeneity of the dataset was too great for any meaningful reconstruction. Although, the dataset appeared to how the presence of the *E. coli* GroEL chaperone complex. Co-expression experiments confirmed that the GroEL chaperone complex migrates with USP28 throughout the purification and may be useful for purifying USPs for structural studies.

Currently, our lab has a single-angle X-ray scattering (SAXS) model of the Keap1-Cul3 E3 ligase complex. But, the field does not fully agree on the molecular stoichiometry or the overall structure-function of this oxidation sensor – E3 ligase complex. It is hypothesized that Keap1 forms a dimer through its BTB domain, and a single Cul3 molecule then binds this dimer. The oxidation state of Keap1 cysteines appears to be critical to the interaction, but the field remains uncertain about which residues are responsible for the interaction with the Cul3-Rbx1 E3 ligase. To better understand this interaction and to obtain structural information to corroborate the SAXS model, recombinant Keap1 and Cul3-Rbx1 were purified and their interaction was tested by ITC, gel filtration assay, and a new technique called *mass photometry*.

It was found that the Keap1 Cys151 residue is not the oxidation sensor critical to the interaction, contrary to what some in the field anticipated. Additionally, it was found that under oxidative conditions, WTKeap1 could not form a complex with Cul3-Rbx1. The complex was successfully purified and was measured by SDS-PAGE, gel filtration assay, and mass photometry, and then used for cryo-EM single particle analysis. Full data collection and analysis has not yet been completed. It is anticipated that like the data from mass photometry, analytical SEC, and cryo-EM single particle analysis will show the complex appears to show a 1:1 Keap1-Cul3 stoichiometry, as opposed to the anticipated 2:1 ratio.

CHAPTER 1. CYRO-EM SINGLE PARTICLE ANALYSIS OF USP7 REVEALS MECHANISTIC INFORMATION

1.1 Introduction

Human ubiquitin-specific protease 7 (USP7) is a particularly potent oncoprotein that has a demonstrated role in many cellular functions (Table 1.1) (1-30). USP7 has been described as a contributing factor in the progression of numerous diseases including prostate cancer, multiple myeloma (34), colon cancer (20-21), and due to its role, has generated interest as a drug target (16, 20-21, 28, 34). A major caveat to the development of cancer therapeutics in the USP family is that structure-activity-relationship studies of inhibitor optimization have relied mostly on the X-ray structures of the catalytic domain, rarely accounting for the other USP7 domains, TRAF and HUBL1-5 and how they recognize substrates and regulate the catalytic activity of USP7 (31-35). Although this method of inquiry has resulted in several therapeutic leads on inhibitors of USP7, these inhibitors are generally targeted to the active site, or to an adjacent site in the catalytic domain, and consequently it is anticipated this strategy may lead to off-target effects in closely related enzymes (31-35).

The domains surrounding the catalytic domain are thought to provide each USP with an individual ubiquitin-cleavage fingerprint, defining substrate specificity and regulation (1, 18, 36). USP7 is a unique member of the USP family due to its 6 domains, some of which lack significant homology with related enzymes (1, 18, 35). The domain architecture (illustrated in Fig. 1.1, below) includes a catalytic domain flanked by a TRAF domain, and five HUBL domains (termed 1-5) that are hypothesized to function as a tethered-rheostat activator of the catalytic domain (1, 3, 18, 36-39). The USP7 catalytic domain is characterized by a tripartite finger, palm, thumb fold organization, like that of coronavirus proteases, sharing a similar Cys-His-Asp catalytic triad (40-41).

The idea of a rheostat in biological functions comes from the term *rheostasis*, which refers to biochemical and physiological processes that can be regulated through a graduated and quantitative manner to adapt to physiological response, like a rheostat switch in a thermostat (42-43). Rheostats have also been used to describe regulation in several biological phenomena, including the immune responses, natural killer cell responsiveness, ubiquitination, and has been used to describe the role of USP7 in regulating endosomal protein recycling (44-47).

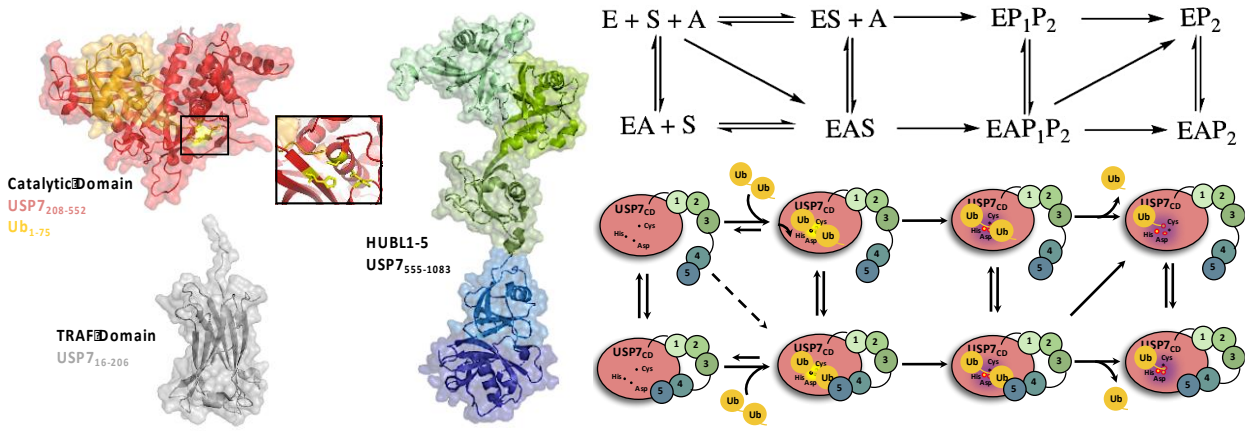


Figure 1.1. USP7 domain organization and catalytic mechanism scheme. (Left) USP7 catalytic domain with ubiquitin aldehyde bound (PDB: 1NBF; catalytic residues highlighted in inlet), TRAF domain (PDB: 2F1W), and HUBL1-5 (PDB: 2YLM). (Right) Enzymatic schematic and diagram illustrated with the applicable states with use of ubiquitin-aldehyde; HUBL-engaged states (bottom) and HUBL-disengaged states (top). The dashed line represents the "induced fit" model of engagement, where substrate binding induces HUBL engagement.

Biochemical and structural studies have uncovered a role for the TRAF domain in protein-protein interactions with substrates such as tumor suppressor p53, ubiquitin ligase MDM2, and viral DNA-binding protein EBNA1 (interacting proteins summarized in Table 1.1, below) (2-8). The HUBL1-5 rheostat has been found to extensively control the activity of the USP7 catalytic domain, truncations of which result in an 80 –fold reduction in K_m and k_{cat} (48-51).

Kinetic evaluation of USP7 within the context of the domains has revealed information suggesting a tethered-rheostat mechanism of intramolecular activation (48-51). TRAF domain deletion in USP7 results in no significant change to the K_m and k_{cat}, suggesting that the major function of the TRAF domain is suggested to provide substrate specificity by recognition specific peptides (3, 51). Correspondingly, it is thought that the TRAF domain may not be responsible for ubiquitin recognition or binding. However, truncation of HUBL4-5 reduces catalytic efficiency like a HULB1-5 truncation, suggesting that HUBL1-3 may not be necessary for activation (48-51).

Table 1.1. All currently known human USP7 deubiquitination targets and interacting peptides.

USP7 Deubiquitination Targets & Interacting Proteins		
Target Protein	**Associated Cellular Processes**	**Reference(s)**
MDM2	E3 ligase targeting p53 for ubiquitination	(2-4)
p53	Apoptosis and senescence	(2-8)
TSPYL5	Proliferation & tumor suppression	(9)
H2B	Chromatin and DNA regulation	(10)
Chk1	Cell cycle arrest & DNA repair	(11)
Claspin	Cell cycle arrest & DNA repair	(12)
UVSSA	Nucleotide excision repair	(13)
SCML2	Transcription repression	(14)
DNMT1	DNA methylation	(15-17)
BRCA1-A	DNA damage response & tumor suppression	(18)
RNF168	DNA damage response E3 ligase	(19)
FOXO4	Transcription factor; growth/differentiation	(20)
PTEN	Cell cycle regulator & tumor suppressor	(21)
MAGE-L2	Endocytic vesicle recycling and transport	(22)
GMPS	Nucleotide metabolism	(10, 23)
ICP0	Viral protein/Herpesvirus E3 ligase	(24)
EBNA-1	Viral protein/DNA replication	(25)
c-Myc	Transcription factor; cell growth/metabolism	(25)
N-Myc	Transcription factor; cell growth/metabolism	(26)
UHRF1	DNA methylation & chromatin modification	(27)
AR	Androgen receptor; hormone signaling	(28)
TRIM27	Apoptosis	(29)
RIP1	Apoptosis	(29)
HDMX/HDM2	DNA damage response	(30)

This difference in catalytic efficiency attributed to the loss of HUBL4-5 remains to be fully understood, but thorough biochemical analysis by Faesen *et al*. (49) defined the requirement of HUBL4-5 in USP7 catalysis as the c-terminal 19 residues. They hypothesized an interaction between this c-terminal peptide of HUBL domain 5 ($USP7_{1084-1102}$) and a switching loop of the catalytic domain ($USP7_{285-291}$) was responsible for this activation of USP7 (49). Crystallographic structures solved by Rougé *et al*. (50) built upon this model by revealing a density for the c-terminal peptide of HUBL5 bound within the newly defined activation cleft of the catalytic domain, resulting in movement of this switching loop and access of the catalytic residues. This study provided the first structural evidence for *in trans* activation of USP7 by the HUBL5 peptide; however, it is important to note the USP7 construct used was an artificial one. HUBL4-5 was tethered directly to the catalytic domain by a flexible glycine-serine linker, in lieu of HUBL1-3

(50). Kim *et al.* (51) have recently suggested a model of *in cis* activation by the HUBL1-5 rheostat supported by NMR data, but a structural density of this intramolecular binding has yet to be observed.

These two modes of activation: *in trans* and *in cis*, refers to the type of activation the HUBL1-5 rheostat imposes on the catalytic domain. *In trans* activation herein refers to catalytic domain activation via the HUBL domains of a second USP7 molecule, whereas *in cis* activation refers to activation by the HUBL rheostat within the same USP7 molecule.

To elicit this degree of activation by the HUBL1-5 rheostat, several groups have described the large degree of conformational flexibility that is supposedly maintained by the rheostat (48-52). Kim *et al.* (48) have determined the importance of the residues in the α-helical linker connecting the catalytic domain and HUBL1-3 for rheostat function, indicating amino acids that are required for conformational flexibility. Likewise, Pfoh, et al. have described the existence of potential hinge regions between HUBL2-3 and HUBL3-4 that are hypothesized to facilitate this movement of the rheostat to potentially support in cis and in trans activation of USP7 (52).

HUBL1-3 were originally described to serve primarily as a binding platform for protein-protein interactions with DNMT1, UHRF1, and ICP0, as well as the USP7 allosteric activator guanosine 5'- monophosphate synthase (GMPS) (10, 15-17, 23-24, 27). Although, our kinetic evaluation has uncovered a second role for HUBL1-3 as part of the rheostat which regulates the level of activation that HUBL4-5 impose upon USP7 (N. Hjortland, dissertation). Our data suggests that this rheostat function is negatively affected – at least in part – by the presence of the TRAF domain, likely indicating an interfacial region where an interaction may occur. (Nicole Hjortland, dissertation).

To obtain structural evidence of the HUBL tethered-rheostat and its hypothesized motion, a cryo-EM single particle analysis was performed with USP7. Additionally, the goal of the study involved using full-length USP7 in a substrate-bound state to determine if the HUBL5 peptide is engaged with the catalytic domain in the EP2 state (referring to FIG. 1.1). To this end, movies were collected of USP7 by cryo-TEM using a Volta Phase Plate (VPP) permitting visualization of USP7 particles in a variety of conformational states. The reconstruction process resulted in a sub-nanometer electron density from the single particle analysis.

1.2 Methods & Materials

1.2.1 Synthesis of Ubiquitin-Propagylamine (Ub-PA) probe

Codon-optimized human ubiquitin (amino acids 1-75) was expressed in a pTYB2 plasmid with a carboxy-terminal intein tag fused with a chitin binding domain (CBD) in BL21(DE3) RIPL *E. coli* (Rosetta) under 25 μg/mL chloramphenicol and 50 μg/mL carbenicillin antibiotic concentrations. Bacterial cells were cultured in auto-induction LB: Luria Broth supplemented with final concentrations of 0.3% (v/v) glycerol, 0.1% (w/v) glucose, and 0.2% (w/v) lactose at 37°C until an OD of 0.6, and then cooled to 18°C for 12 hours.

Cells were harvested via centrifugation at 5,000 x g for 20 min and resuspended in lysis buffer: 50 mM MES, pH 6.0, 350 mM NaOAc, 50 μg/mL lysozyme (Thermo Fisher), 25 μg/mL bovine DNase (Sigma), 1 tablet EDTA-free protease inhibitor cocktail (Roche). Cells were lysed via sonication (Branson Sonifier) at an amplitude of 65%, 1 min per gram cells, 6 s pulse on, 6 s pulse off. Lysate was clarified by centrifugation at 17,000 x g and 0.45 μm filtration. Clarified sample was applied to chitin beads (Boston Biochem) and allowed to react at 4°C for 2 hours, gently agitated. Sample-reacted chitin beads were applied to a fritted funnel and flowthrough was filtered, followed by 5 equivalent volume washes with 50 mM MES, pH 6.0, 350 mM NaOAc. Ubiquitin$_{1\text{-}75}$-MesNa was eluted from the chitin resin with the addition of 150 mM 2-mercaptoethanesulfonate (MesNa; Sigma), at 4°C, gently agitated for 12 hours.

Chitin resin was filtered by a fritted funnel and the ubiquitin$_{1\text{-}75}$-MesNa elutant was collected, filtered to remove excess resin, and supplemented with propargylamine HCL (Sigma) to a final concentration of 250 mM. This reaction was brought to pH 10, ambient temperature, and agitated gently for 12 hours. The reaction was concentrated using a 3,500 Da MWCO centrifugation concentrator (Millipore) applied to a Sephacryl S75 16/60 size exclusion column (GE Healthcare) via FPLC to remove contaminating proteins and buffer exchange into storage buffer: 25 mM Tris, pH 7.5, 150 mM NaCl, 5% (v/v) glycerol.

1.2.2 Purification of full-length USP7~Ub-PA adduct for cryo-EM imaging

Full-length human USP7 (amino acids 1-1102) was expressed in a BD BaculoGold (BD Bioscience) baculovirus expression system in SF9 insect cells (Expression Systems) with an engineered amino-terminal PreScission protease-cleavable hexa-histidine tag (10x His–GFP).

Baculovirus preparation and insect cell expression were performed as described by the manufacturer at 27°C, with an MOI of 5, and a 2-day infection period. SF9 cells were harvested via centrifugation at 5,000 x g for 20 min and resuspended in lysis buffer: 25 mM HEPES, pH 7.7, 500 mM NaCl, 5 mM beta-mercaptoethanol, 5% (v/v) glycerol, 10 mM $MgCl_2$, 50 μg/mL bovine DNase (Sigma), and 1 tablet of EDTA-free protease inhibitor cocktail (Roche). Resuspended cells were lysed by sonication (Branson sonifier) at 40% amplitude, 1 min per 2 g cells, 6 s pulse on, 6 s pulse off. Lysate was clarified by centrifugation at 17,000 x g for 20 min and 0.45 μm filtration and then applied via FPLC to a 5 mL NiNTA resin column (GE Healthcare).

The NiNTA column was washed with a 90% lysis buffer, 10% elution buffer mixture (elution buffer: 25 mM HEPES pH 7.7, 500 mM NaCl, 450 mM imidazole, 5 mM beta-mercaptoethanol, 5 % (v/v) glycerol) until baseline UV was reached. USP7 elution followed with an elution gradient of lysis buffer to elution buffer 10-100%; USP7 eluted at ~100-125 mM imidazole. Peak fractions were collected and analyzed by SDS-PAGE for purity and by ubiquitin-rhodamine110 (Boston Biochem).

Fractions with > 50% purity were pooled and placed in dialysis buffer: 25 mM HEPES, pH 7.7, 200 mM NaCl, 5% (v/v) glycerol, 5 mM beta-mercaptoethanol, with a 1:5 mass ratio of USP7 to PreScission protease, for 12 hr at 4°C, 10,000 MWCO dialysis tubing pore size. After dialysis, samples were again applied to a 5 mL NiNTA HiTrap column (GE Healthcare), this time collecting the flowthrough sample that does did not bind the resin. Samples were analyzed for purity and activity by SDS-PAGE and Ub-Rho110 cleavage assay, respectively.

Fractions with > 90% purity were pooled and concentrated using a 100 kDa MWCO centrifugation concentration (Millipore) and then FPLC-applied to a sephacryl S200 26/60 size exclusion column (GE Healthcare) in SEC buffer: 25 mM HEPES, pH 7.7, 150 mM NaCl, and 4 mM DTT. Peak fractions were analyzed by SDS-PAGE for purity and by a Ub-Rho110 assay for activity. USP7 fractions with > 1,000 μM/min/mg specific activity and > 95% purity on SDS-PAGE were used for cryo-EM specimen preparation and Ub-PA reaction.

FLUSP7 and Ub-PA and were mixed with a 20-molar excess to react at 4°C, overnight. Reacted USP7 was subjected to a second round of size exclusion in the same buffer to remove excess ubiquitin. Peak fractions were again analyzed by SDS-PAGE and Ub-Rho110 cleavage assay. Fractions with > 95% purity as evidence by SDS-PAGE and < 1.0 μM/min/mg specific

activity were used for cryo-EM specimen preparation. All steps after cell pellet harvest were performed at 4°C.

1.2.3 Isopeptidase activity assay for full-length USP7 and Ub-PA adduct

Isopeptidase activity of USP7 in Fig. 1.5 was measured using fluorogenic mono-ubiquitin-rhodamine110 (Ub-Rho110, Boston Biochem) and a plate reader with 485 nm excitation and 535 nm emission (BioTek). 1 nM [E] was and 500 nm [Ub-Rho110] final assay concentrations were diluted in Assay Buffer: 50 mM HEPES pH 7.5, 5 mM DTT, 0.1% (w/v) Triton X-100, 0.1 mg/mL BSA (Sigma). The assay was initiated with the addition of 25 μL of 2x enzyme to 25 μL of 2x substrate for a final volume of 50 μL in a 96-well black, opaque half-well assay plate (Corning). USP7 activity assays in Fig. 1.4 was measured with 5 mM [E] pre-incubated for 1 hour at room temperature with concentrations of mono-Ub-PA prior to initializing the reaction.

1.2.4 Cryo-EM specimen preparation, data acquisition, and movie processing

Three microliters of purified USP7~Ub-PA was applied to a Quantifoil R0.6/1 400 mesh gold grid (Electron Microscopy Sciences) without glow discharging. Bare gold grids were coated with 5 μL pyrene (1.0 mg/mL) after chloroform washing, and then coated with graphene oxide monolayer (Sigma). Grids were plunge-frozen in liquid ethane with a CP3 Vitrobot (Gatan) under the following conditions: 25°C temperature, > 90% relative humidity, and 9 s blotting time. Plunge-frozen grids were imaged by a FEI Titan Krios (300 kV, Thermo Fisher Scientific) fit with a Gatan Quantum-LS energy filter (20-eV zero-loss filtering) and a Gatan K2 Summit detector operating in super-resolution (SR) counting mode, and a Volta Phase Plate (VPP). Data acquisition was performed as movies of 64 frames over 8000 ms exposure acquired at a nominal magnification of 130,000x. During data collection, a total dose of 80 e^{-} Å^{-2} and a pixel size of 0.545 Å (SR mode) were used. Acquired movies were then processed during imaging with MotionCor2 motion correction and dose weighting and CTFFIND4 contrast transfer function (CTF) estimation (54-55). 800 total aligned movies were used for further single-particle analysis. Images were 2x2 binned, and images with CTF correction indicating a resolution > 7 Å or > 2 Å average drifts per frame were excluded from analysis. For data acquisition parameters and a summary of the single-particle analysis pipeline, refer to Appendix E.

1.2.5 Image Processing

Particle selection was performed with the Relion3.1 and ctffind4 programs (55-56). During initial processing, 1500 particles were reference-free manually selected from 60 images with Relion3.1 localization. The selected particles were used for generating auto-picking templates using reference-free 2D classified in Relion 3.1. Five 2D classes from this particle set were low-pass–filtered to 20 Å and used as templates for Relion3.1 auto-picking particle selection with the full dataset. The larger particle set was reference-free 2D classified in Relion 3.1, resulting in 30 higher-resolution averages appearing to cover a broader range of views. These were selected as templates for a second round of particle selection with Relion3.1. The final particle selection resulted in a total of 525,578 particles, which were then reprocessed with cryoSPARCv2 (57). After several rounds of 2D classification, a subset of 2D class averages was used to generate an initial model for 3D processing.

A final 2D class averages set of 269,921 particles was subjected to three further rounds of 3D classification with the major, HUBL1-5 extended conformation the input map (Fig. 1.8 & 1.9). The resulting 3D classes were then used to create an initial map of USP7 at approximately a resolution of 12 Å (0.143 FSC). From this classification, a dataset including 185,833 particles was then processed by Relion 3.1 using Refine3D. Two identical 3D class half-maps displaying sub-nanometer resolution of ~8.3 Å (0.143 FSC) and comprising a total of 185,833 particles were then combined into one dataset. This data was refined in post-processing with one of the two 8.3-Å 3D classes as an input model (15 Å low-pass–filtered) used to generate a mask in Relion3.1 mask creation. After masking and subsequent modulation transfer function correction, the reconstructions improved to a resolution of approximately 8.2 Å, as measured by 0.143 FSC, in which the refinement of the two halves of the dataset were separately refined and combined only during final map building (Fig. 1.9). For more information on working with the VPP dataset at low defocus and a representative micrograph, refer to Appendix F.

1.2.6 Multi-Body Refinement

Multi-body refinement was performed with two hypothesized bodies – a catalytic domain, TRAF domain, and HUBL1-2 body, and a HUBL3-5 body. First, map segmentation was performed by generating two masks in Chimera: one circular mask enclosing the HUBL3-5

domains and a second encircling the remainder of the density; coordinates were saved with normalized pixel values to [0, 1] (62). Relion "mask create" was used to ensure identical pixel size and box size as the reference map. The resulting masks were combined into in a single *.star* file.

Particle subtraction was performed by Relion/3.1 using the *optimizer.star* file from a '3D multi-body' jobtype as an input (61). The 'Particle subtraction' jobtype can take this file and run a subtraction utility. The resulting particle images stack for each body were then separated based on rotational and translational orientation, and signal was subtracted from the signal from the other bodies present. These new particle images resulted in a more focused refinement.

Multi-body refinement was performed using an angular sampling rate of 0.9° and a 0.5 pixels (0.55 Å) translational sampling rate. These parameters were selected for angular sampling due to computational cost and an appropriate sampling rate to describe the hypothetical flexibility (61).

1.3 Results

1.3.1 Reference-free 2D averages and 3D projections illustrate mobile HUBL domains

Current dogma suggests two mechanisms by which the HUBL domains can function as an activator of USP7's catalytic domain: 1) *in cis* activation where the HUBL domains bind intramolecularly into the *activation cleft* of the catalytic domain to increase catalytic efficiency, and 2) *in trans* activation where the HUBL domains bind the *activation cleft* intermolecularly – interacting with a second USP7 molecule. There is contention in the literature concerning what stage during the USP7 catalytic mechanism the H5 peptide binds the activation cleft, and at what stage the H5 peptide releases from the catalytic domain.

Rougé, *et al*. have resolved the H5 peptide electron density bound into the catalytic domain in a ubiquitin-bound state (50). And recently, Kim *et al.* have suggested a model whereby the H5 peptide is engaged after catalytic triad rearrangement by Ub-binding and remains engaged until after Ub-bond hydrolysis and release (51). Thus, it was expected that use of a chemically-modified mono-ubiquitin probe, ubiquitin-propargylamine (Ub-PA) might trap the full-length construct of USP7 in a similar H5-engaged conformation. It was thought that doing so should increase particle homogeneity, reducing the expected conformational mobility of the HUBL rheostat.

According to our single particle set, it is apparent that after Ub-PA conjugation, the HUBL domains do not appear to be engaged in a single, well-resolved conformation. As the averages and

3D projections show (Fig. 1.2), is appears the H5 domain lies distal from the catalytic domain, and the rheostat appears to move in the averages, adopting a variety of positions. The hypothesized position of the HUBL domains (Fig. 1.2, denoted by red arrows) show that the 4 distal HUBL domains, H3-5, move in a swinging motion relative to the catalytic domain covering nearly a ~60 Å distance. It would appear the HUBL domains are not bound *in cis* in this mono-ubiquitin-bound state, which we call the EP_2 state (Fig. 1.1). This adduct is thought to mimic the state of USP7 after cleavage of the isopeptide bond and release of a second molecule, with the first ubiquitin molecule cradled in the ubiquitin cleft, and the catalytic cysteine in a thioester adduct.

To investigate further, and to determine the identity of the mobile density, we employed molecular dynamics flexible fitting (MDFF) to determine the position of HUBL domains and measure how well the density correlates with existing crystallographic structures of USP7 (60). As Fig. 1.2 & Fig. 1.3 illustrates, the cryo-EM–derived electron density illustrates well-resolved domains for what was determined to be H3-5 in the mobile portion in the particle stacks (Fig. 1.2).

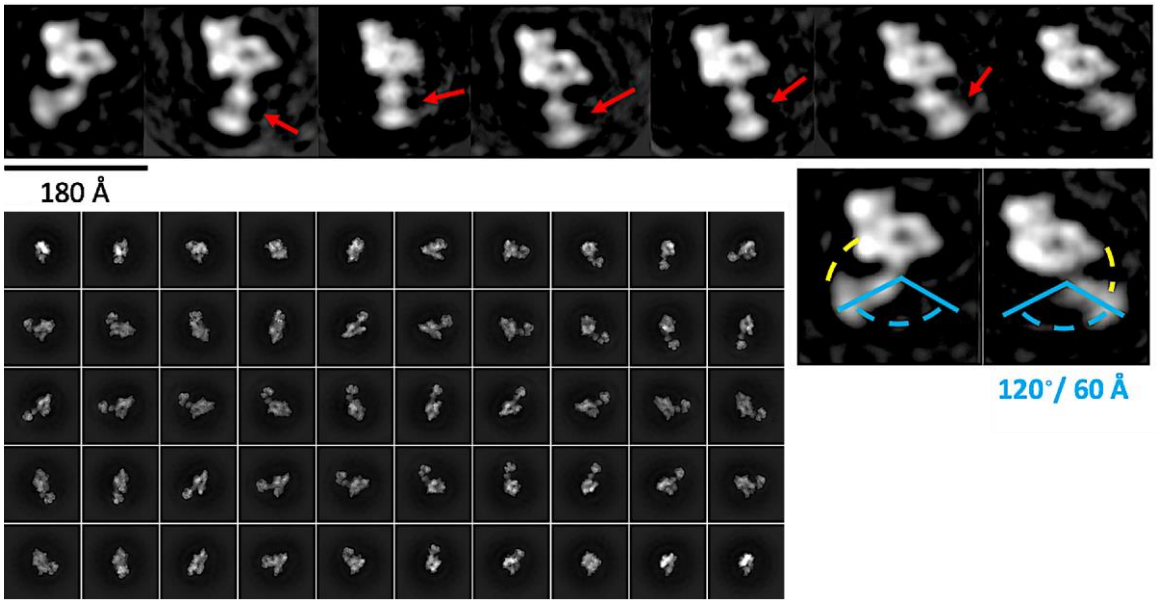


Figure 1.2. Reference-free 2D averages and 3D projections illustrate ubiquitin-bound HUBL5-unengaged state. (Top) Reference-free 2D averages show conformational heterogeneity whereby the HUBL domains (red arrow) appear to swing back-and-forth. (Right) The top leftmost and rightmost averages placed for side-by-side comparison. Distances measured show that the HUBL domains swing ~90 Å total distance and move through a ~60° angle of motion. The HUBL domain does not appear to move close enough to engage the catalytic domain. (Bottom) 50 3D projections of the density in Fig. 1 created in cryoSPARCv2. Projections show another representation of the HUBL5 domain distal from the catalytic domain.

1.3.2 Sub-nanometer resolution cryo-EM refined model of USP7

The resulting refined 8.2 Å cryo-EM density was fit with previously published crystallographic structures via MDFF (Fig. 1.3). The crystallographic structures were originally placed as a contiguous polypeptide (residues 1-1102), filled in by sequence alignment, into the volume without bias and allowed to move without imposed restraint to fit the density. The cross-correlation (cc) fit of the contiguous polypeptide chain, calculated by whole-structure RMSD, is depicted during energy minimization (Fig. 1.3). As the constraint for fitting was relaxed, the cc increases to 0.84, describing a relatively high confidence of agreement between the polypeptide and the cryo-EM volume.

The cryo-EM density is illustrated in Fig. 2 (bottom) by a series of 3D projections generated in cryoSPARCv2 that show the HUBL5 peptide ($USP7_{1084-1102}$) do not appear to be engaged *in cis* within USP7's catalytic domain. Additionally, during the reconstruction process, it was observed that in the 2D averages, there is a density that appears to adopt numerous positions (Fig. 1.2; top). As the red arrows indicate, the HUBL domains can appear to move up to 90 Å distance. MDFF molecular dynamics supports the idea of HUBL domain movement in this manner with simulations up to 40 nanoseconds illustrating a similar type of motion.

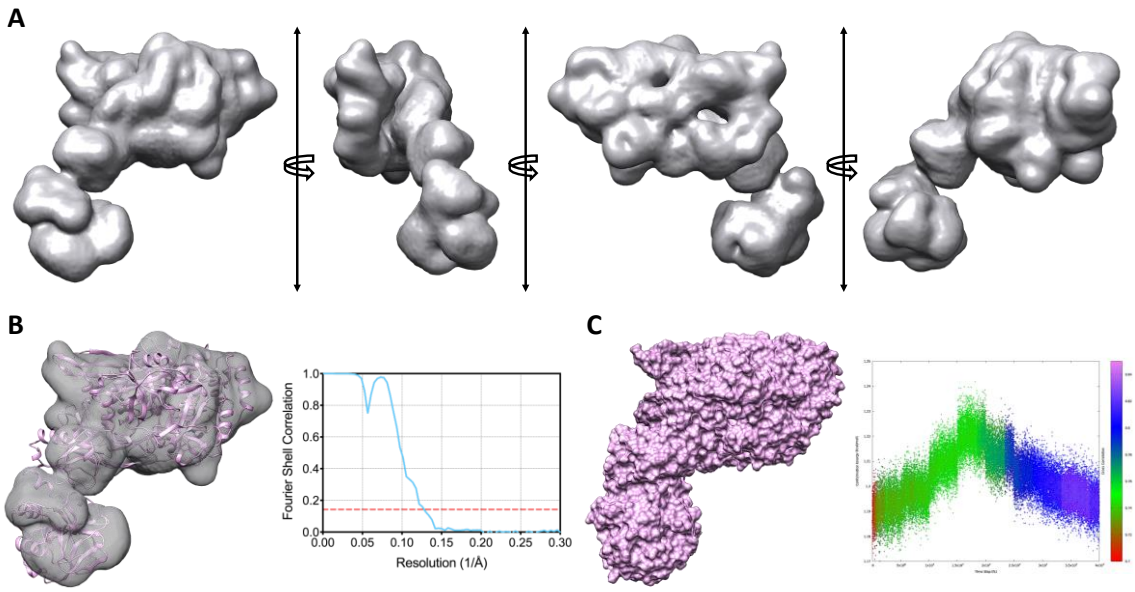


Figure 1.3. 8.2 Å cryo-EM density of USP7 fit with crystallographic structures by MDFF. (A) Cryo-EM single particle electron density of USP7 as viewed in Chimera. Each image is related by 90° rotation, showing all 4 sides. (B) Electron density (grey; surface) shown with ribbon diagram of final USP7 polypeptide modeled in by MDFF (pink). FSC at 0.143 from post-processing in Relion3.1 shows a final resolution of ~8.2 Å. The local minima in the FSC at ~20 Å resolution may be due to tight masking. (C) Surface representation of the higher-resolution USP7 model created from the density fitting. The plot to the right is 40,000 sampled time points (1 picosecond) over 40 nanoseconds, showing the cross-correlation (cc) of fit between the experimental cryo-EM volume and the crystallographic densities. Final cc is ~0.84, indicating a good agreement of fit.

1.3.3 3D Multi-body refinement reveals tethered-rheostat motion

3D multi-body refinement is a useful analysis performed to refine a density with evidence of mobile components, or conformational heterogeneity. In the case of USP7, 2D averages and 3D classification show two clearly, well-defined lobes to the density: the HUBL domains, and the catalytic/TRAF domain. Furthermore, it appears these pieces move relative to one another throughout the particle stack, as visualized by the variety of conformational position evident in the 2D averages (Fig.1.2). This would suggest the USP7 map is a suitable candidate density for multi-body analysis.

One limitation to multi-body refinement relates to the individual body sizes used in the refinement. Relion suggests that for sufficient signal, each body should comprise a molecular weight of at least 100 kDa (61). Refinement of macromolecular complexes with smaller bodies becomes progressively difficult (61). There are few published examples of multi-body refinement

procedures to refine densities of significantly smaller bodies, such as the size of the distal HUBL domains: ~ 10 kDa each (~30 kDa body total) (61). However, with use of a phase plate, it is possible that signal increases for smaller bodies can help to overcome these size limits (61).

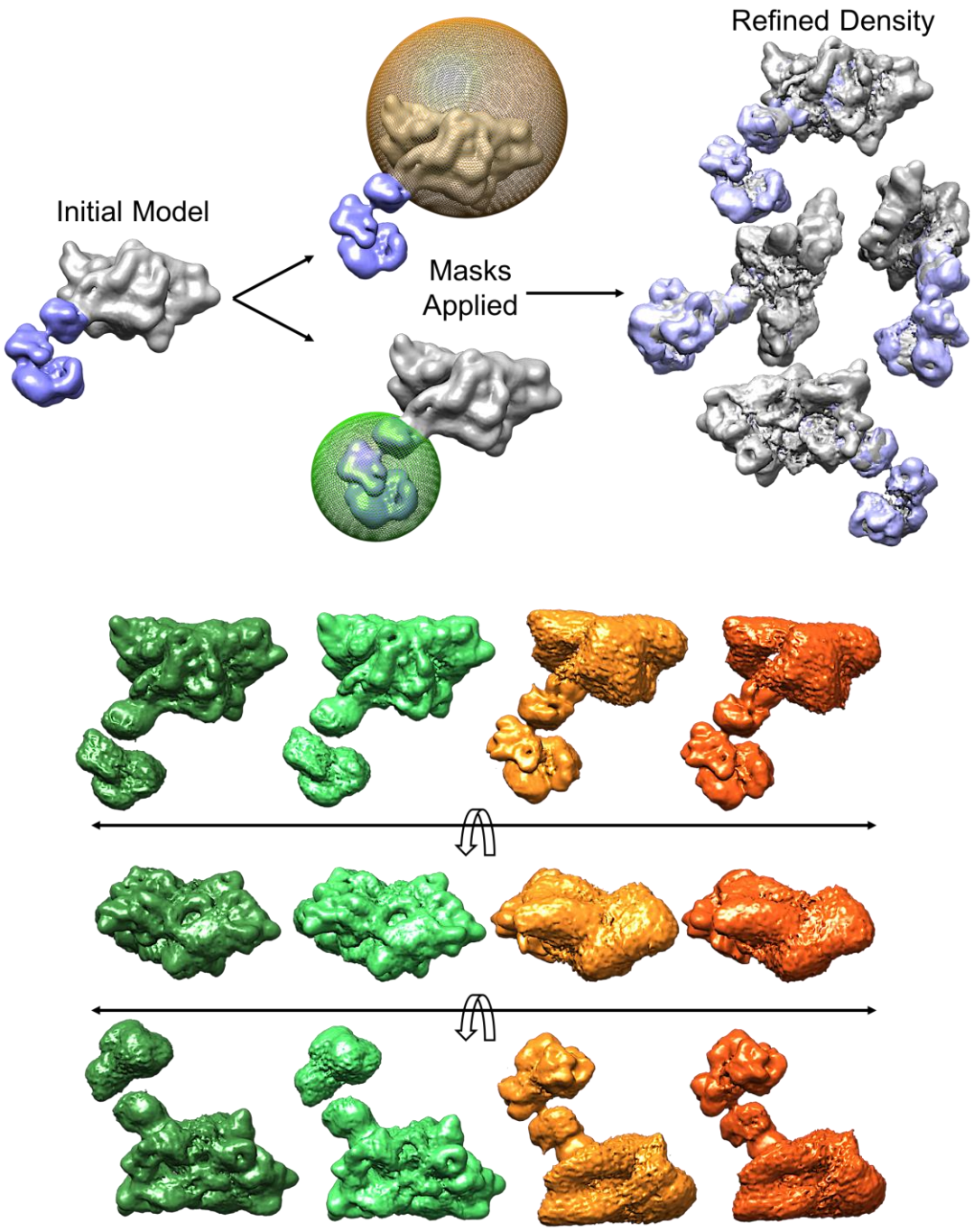


Figure 1.4. USP7 multi-body refinement to increase density resolution. (Top, Left) Initial model density colored based on segmentation; grey = catalytic domain portion; purple is mobile HUBL domain portion. (Top, Middle) Spherical masks at 1.5x the diameter of the body chosen to be masked were created in chimera; catalytic mask (masking out mobile HUBL portion) in orange mesh; HUBL mask (masking out catalytic domain) in green mesh. (Top, Right) Final multi-body refined densities. (Bottom) Half maps generated from each mask colored based on the mask applied. Fuzzy densities represent areas that were masked out and not refined. Green half maps represent refinement of the catalytic domain; orange half maps represent refinement of the HUBL domains.

Upon convergence, the solvent-corrected resolution estimates for the two bodies were ~6.1 Å for the HUBL domains and the catalytic domain (Fig. 1.10, top right).

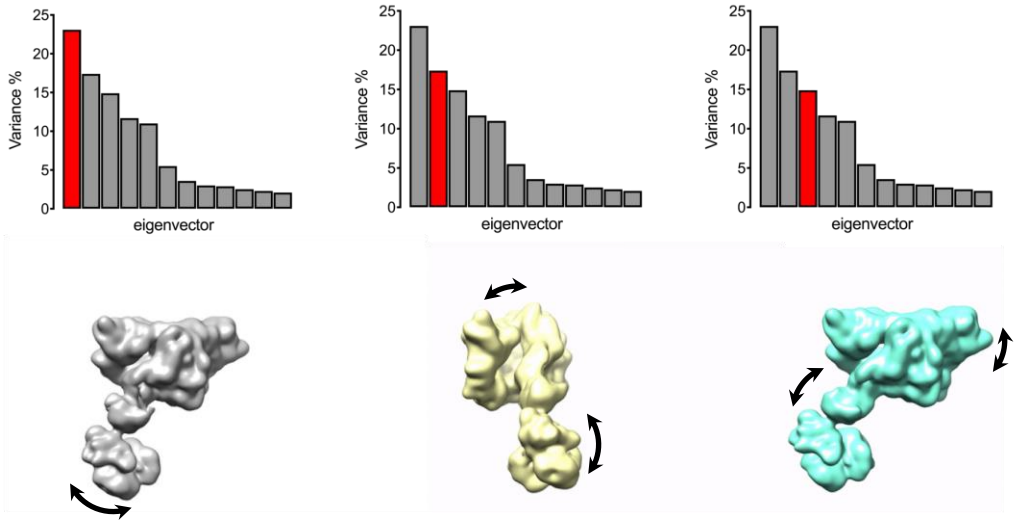


Figure 1.5. USP7 multi-body refinement illustrates HUBL movement. (Top) The contribution of all Eigenvectors to the variance. The first three Eigenvectors describe the highest movement and are shown as annotated movies below illustrating the directions of movement. Movies of the reconstructed body densities repositioned along the first three Eigenvectors reveal that a first Eigenvector (left; grey) corresponds to a swinging-like motion (denoted by black arrow). Motion along a second Eigenvector (middle; yellow) is reminiscent of a rolling motion. Movement along a third Eigenvector (right; blue) is comparable to a ratchet-like motion.

Principal component analysis (PCA) in Relion revealed that approximately 55% of the measured variance in the translations and rotations of the two bodies (catalytic-TRAF-HUBL1-2 and HUBL1-3) may be explained by the three major Eigenvectors (Fig. 1.5). Movies of the reconstructed densities repositioned along the above Eigenvectors reveal a first Eigenvector corresponding to a swinging-like motion of the HUBL domains with respect to the catalytic domain body. Motion along a second Eigenvector, albeit similar, is reminiscent of a rolling-like motion between the HUBL domains and the catalytic domain. Movement along a third Eigenvector is comparable to a ratchet-like motion. The movement in the HUBL domains appears to stem from motion along the HUBL3-4 and HUBL2-3 hinge regions, although more data analysis may be needed to confirm this. It is important to note that there was a normal distribution in the histogram of amplitudes along these Eigenvectors, indicating that structural heterogeneity along these Eigenvectors may be relatively continuous (61).

1.3.4 Cryo-EM-derived model validation

Table 1.2. Cryo-EM-derived model validation statistics for USP7~Ub-PA.

	USP7~Ub-PA
MolProbity score	2.4
Clashscore	4.2
Poor rotamers (%)	5
Ramachandran plot	
Favored (%)	86
Allowed (%)	9.0
Disallowed (%)	5.0

Table 1.2 illustrates the model validation statistics using MolProbity for the USP7~Ub-PA model obtained from the 8.2 Å cryo-EM map (Fig. 1.9) prior to multi-body refinement (63). The USP7~Ub-PA .pdb file was created from the crystallographic coordinates of $USP7_{CD}$~Ub-aldehyde (PDB: 1NBF), the TRAF domain (PDB: 2F1W), and the HUBL domains 1-5 (PDB: 4Z96), with all missing residues built-in, in one contiguous polypeptide. No restraints were imposed, and the coordinates were flexibly docked, using MDFF, into the cryo-EM map over a 40-ns simulation period using computing resources from Gilbreth (Kihara Lab). The resulting structural model (all-atoms) was then analyzed in MolProbity for structural validation summarized in Table 1.2.

According to several biochemical and structural studies published by Fasen *et al*. (49) the alpha-helical connector between the catalytic domain and HULB1 is thought to maintain its 7 alpha-helices in ubiquitin-bound states (48-50). Both the number of turns, and chemical properties of the residues in the linker of the rheostat, are thought to contribute significantly to the USP7 DUB function and are hypothesized to be maintained throughout all catalytic states of the enzyme.

After modeling a 7-turn helical linker, such as the crystallographic structure published by Kim, *et al.,* (PDB: 5FWI), it is anticipated that the MolProbity score and Clashscore might improve, indicating an improved model. The model described in Table 1.6 was used for subsequent structural analysis in the next section. The % poor rotamers, and % disallowed Ramachandran outliers are typical of models built from cryo-EM maps.

1.3.5 Compact HULB1-3 packing occludes activation cleft

Our cryo-EM density shows the HUBL domains occupy an unengaged state when USP7 catalytic domain is ubiquitin-bound by our mono-Ub-PA probe. It is thought that this more packed H1-3 conformation might sterically hinder the activation cleft, thus occluding H5 peptide binding – both *in cis* or *in trans* – explaining why no H5 engaged particles were observed in our dataset. Comparison of the Kim, *et al*. structure of catalytic domain and H1-3 (Fig. 1.11) shows the HUBL1-3 domains are packed more tightly to the catalytic domain (48). Comparison of Pfoh, *et al* structure of H1-3 (Fig. 1.6) by 'chimera matchmaker' shows good agreement at an RMSD of ~3 Å with our model, and the ICP0-bound peptide laying externally (52, 64). It is hypothesized that intermolecular regulators, like GMPS and ICP0, might influence USP7 activity by binding on this outside face and pulling the H1-3 domains away from the activation cleft.

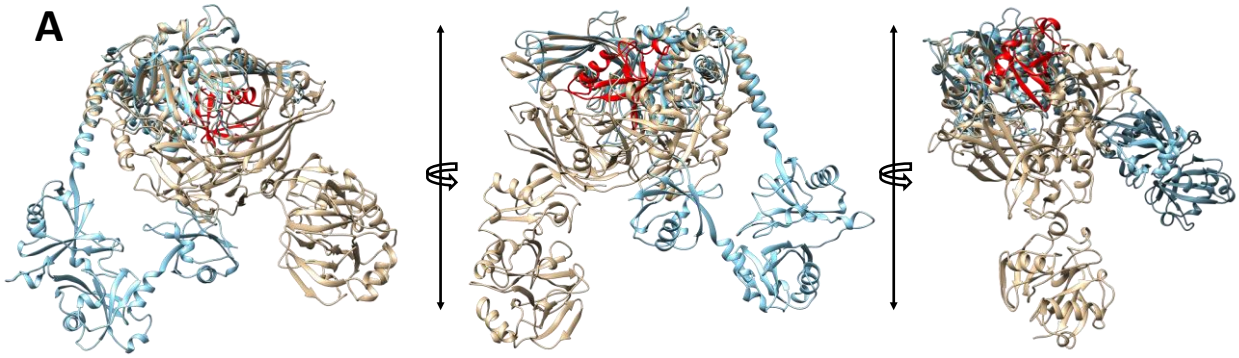


Figure 1.6. Comparison of USP7$_{CD}$-H1-3 and full-length model shows HUBL1-3 occlude activation cleft. (A) Comparison of the cryo-EM-derived structure (beige) with crystallographic structure published by Kim, *et al*. *J. Struc. Biol.* 2016 (blue; PDB: 5FWI) using chimera matchmaker (64). Alignment of the catalytic domains shows the H1-3 domain move inward an average of ~25 Å (average RMSD for each of the three HUBL domains). The catalytic domains were aligned with an average RMSD of ~0.5 Å. Extensive mutational analysis performed by Kim *et al*. suggested the number of turns in the α-helical linker was important for catalytic activity, and the same number of α-helical turns is maintained in our cryo-EM derived structure. Each turn from left-to-right represents a ~90° rotation about the axis.

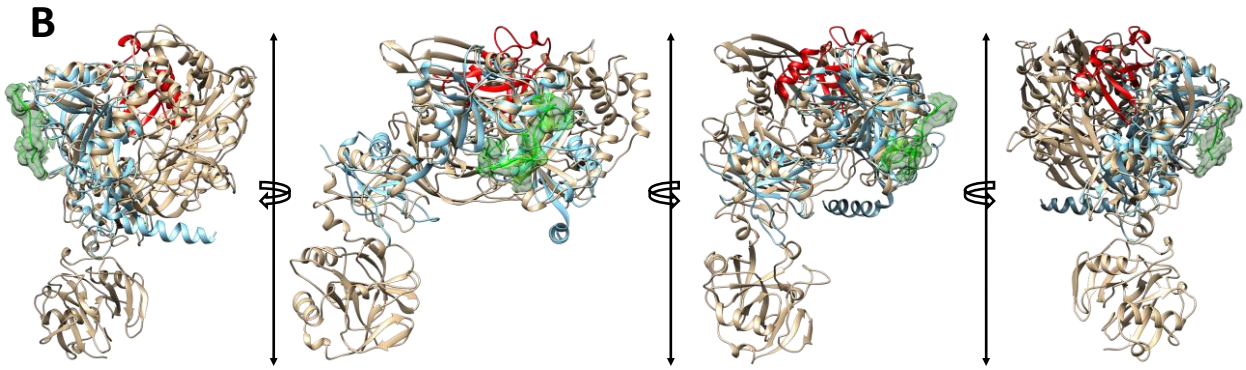


Figure 1.7. HUBL1-3 intramolecular binding pocket outward facing. Comparison of the cryo-EM-derived structure (beige) with crystallographic structure published by Pfoh *et al*. *PLoS Pathog*. 2015 (blue; PDB: 4WPI) using chimera's matchmaker (64). The H1-3 domains were aligned with an average RMSD of 5.0 Å. The ICP0 peptide is in green (stick and surface representation). This outward facing peptide binding pocket would suggest how intermolecular regulators like GMPS can increase catalytic efficiency by pulling H1-3 away from its position occluding the H5 peptide activation cleft. Each turn from left-to-right represents a ~90° rotation about the axis**.**

1.3.6 A flexible hinge between H2-H3 and H3-H4 in the USP7 rheostat

Several research groups have reported flexibility in USP7 HUBL domains (48-52). Structural, biochemical, and mutational analyses have identified two other regions in addition to the alpha-helical connector that are points of rheostat motion: HUBL2-3 hinge, and HUBL3-4 hinge. Below is a comparison of all currently published USP7 structures maintaining HUBL domains that illustrates some flexibility in these points.

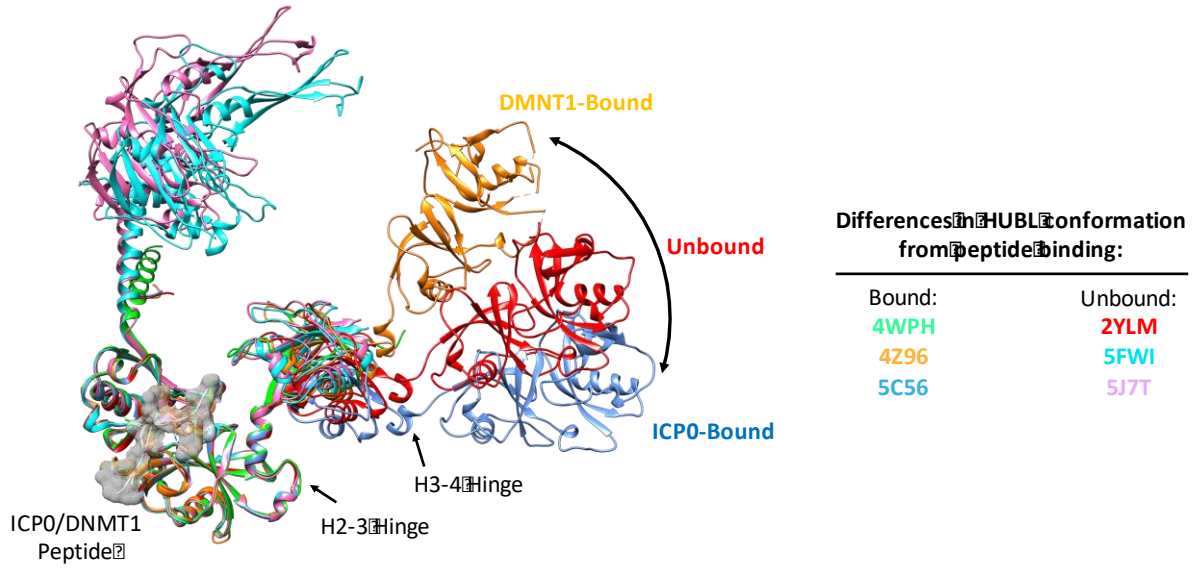


Figure 1.8. Structural analysis illustrates HUBL domain movement from intramolecular binding. (Right) Summary of the crystallographic coordinate PDB file numbers used to generate the comparison with color coding matching the structure to the left. All structures aligned are shown with the lowest permissible RMSD in Chimera matchmaker (64). (Left) The purple and cyan structures are USP7$_{CD}$-H1-3 illustrating the position of the alpha-helical connecter and the ICP0/DMNT1 binding pocket (ligands in surface representation). The HUBL1-3 domains are aligned between these two structures and the three crystallographic structures of HUBL1-5: DNMT1-bound (orange), apoHUBL1-5 (red), and ICP0-bound (blue). The position of HUBL1-5 appears to move irrespective of intramolecular binding, moving either inward or outward. In no position does the HUBL5 peptide appear to be within reach of the catalytic domain for *in cis* activation. In green is a crystallographic structure of HUBL1-3 to which all were aligned.

1.3.7 HUBL2-3 flexible hinge and ‘compact’ & ‘extended’ rheostat conformations

Evidently, this flexibility is also present in the cryo-EM analysis. In addition to flexibility at the H2-3 and H3-4 hinge points, Pfoh *et al*. (52) reported crystallographic structures for a ‘compact’ (PDB: 4WPH) and ‘extended’ (PDB: 2YLM) conformations for HUBL domains 1-3. They reported that, when superimposed, the major difference between the two structures was that the H3 domain had undergone a 35-degree motion about residue H792 (52). Further structural analysis suggested that the loop connecting H2 and H3 is a hinge that facilitated movement at this point in the HUBL rheostat. Structural alignment using chimera matchmaker of existing USP7 structures summarized in Fig. 1.8 show the degree of motion described in the HUBL rheostat about the H2-H3 hinge region (64). The cryo-EM–derived model compared with the same compact and

extended conformations of H1-3 solved by Pfoh, *et al*. also suggest movement about the H2-3 hinge (52).

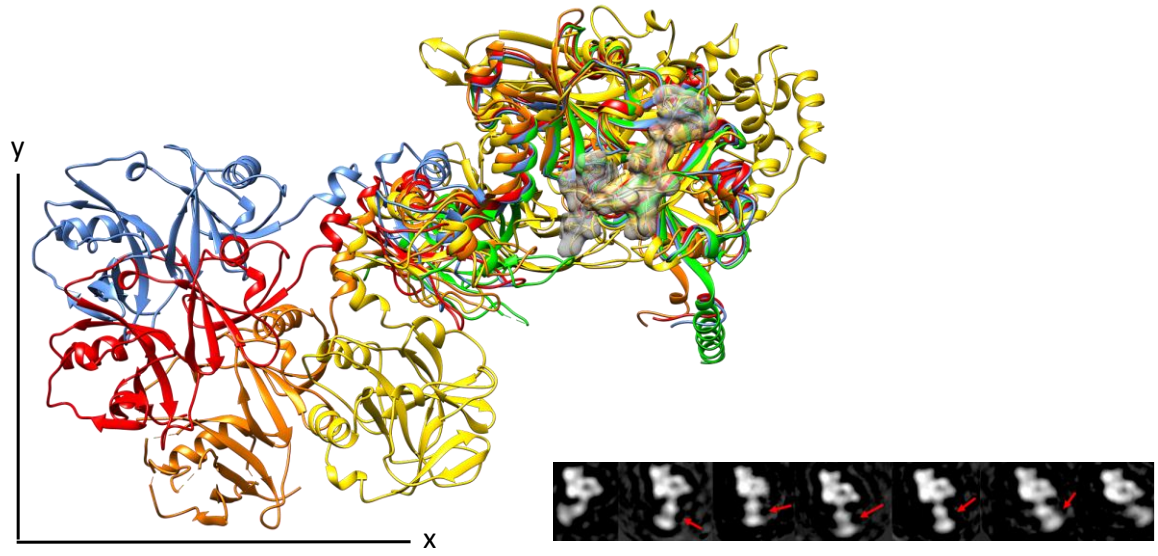


Figure 1.9. HUBL1-5 movement in USP7 illustrated by structural comparison to cryo-EM-derived model. (Left) Comparison of previous HUBL1-5 structures with the cryo-EM-derived model (gold), coloring is identical to Fig. 1.13. Alignment was performed by Chimera matchmaker, with an RMSD of ~3.0 Å in the HUBL1-2 region (64). The maps diverge at HUBL3-5. There is a significant degree of flexibility observed in H2-3 hinge and H3-4 hinge and where the alpha-helical linker is.

1.3.8 Alternate TRAF conformation illustrates flexible TRAF-hinge movement

Fig. 1.15 illustrates the comparison between the crystal structure published by Hu, *et al.* and the cryo-EM-derived model (3). Due to the lack of ordered electron density and high temperature factors in the residues linking the TRAF domain and catalytic domain, Hu, *et al*. suggested that these residues (199-206) form a "flexible linker" region, facilitating TRAF domain movement (3). They note that the substrate-binding residues of the TRAF domain and the ubiquitin-binding cleft of the catalytic domain are located on the same face of the molecule, which appears to move closer together in our cryo-EM-derived model.

The TRAF domain is observed in a more packed conformation, where the substrate-binding face of the TRAF domain twists towards the underside of USP7 catalytic domain. Using Chimera matchmaker to align the catalytic domains of the Hu, *et al*. structure and the catalytic domain of our cryo-EM-derived model to 1.0 Å RMSD, the TRAF domain appears to be positioned approx.

60 Å distance displaced, packing underneath the ubiquitin-binding cleft, and in proximity to H1-3 (3, 64).

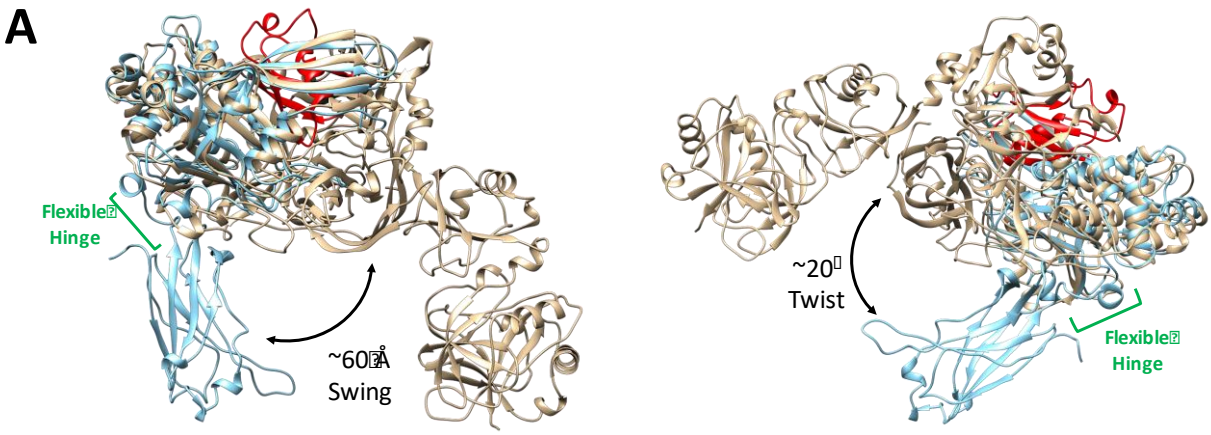


Figure 1.10. Comparison of cryo-EM derived model with crystallographic structure shows alternate TRAF domain conformation. Comparison of our cryo-EM-derived structure (beige) with crystallographic structure published by Hu, *et al*., (blue; PDB: 2F1Z) using chimera's matchmaker (3, 64). Alignment of the catalytic domains shows the TRAF domain moves ~60 Å and twists ~10˚. The catalytic domains were aligned with an average RMSD of ~1.0 Å. In the Hu, *et al*., crystallographic structure, the TRAF domain appears extended downward, away from the catalytic domain (3). Hu, *et al*., suggested the lack of contiguous electron density connecting the TRAF domain and the HUBL domains (denoted above in green) was due to a "flexible hinge" region that might suggest movement of the TRAF domain (3).

Structure alignment between the substrate-bound crystallographic structures of TRAF with our cryo-EM-derived model suggest that the p53/MDM2/EBNA1 TRAF peptide-binding groove twists ~10˚ from the front-face, towards the catalytic core of USP7 (Fig. 1.11 B). It is hypothesized that this movement might be conducive to a more favorable conformation of the ubiquitin chain of the substrate being pulled toward the active site (Fig. 1.10; Fig. 1.11). This close placement of the TRAF domain and the packed conformation of H1-3 supports a previous study on the influence of the TRAF domain on the H1-3 rheostat (66).

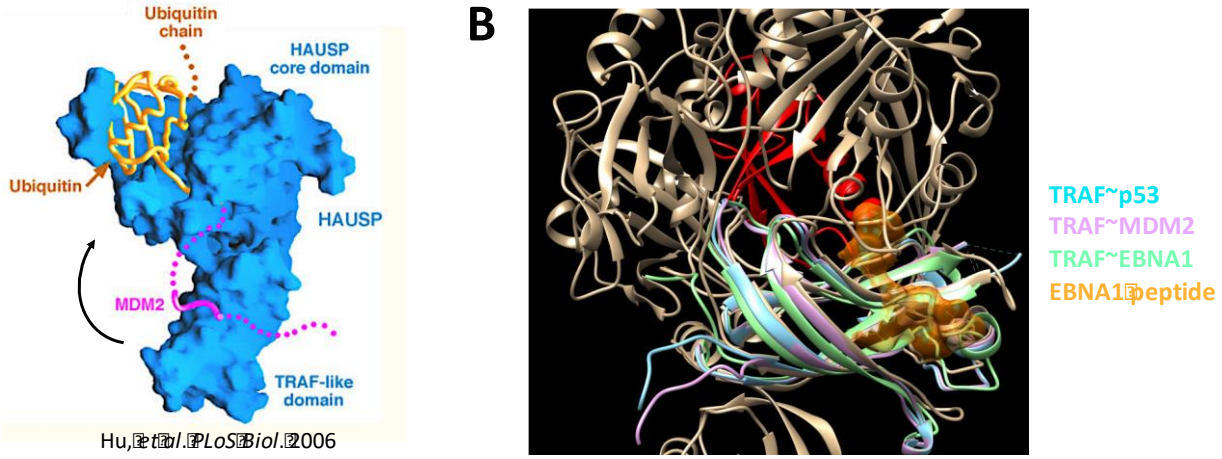


Figure 1.11. TRAF substrate-binding groove moves inwards towards the catalytic domain and twists towards ubiquitin-binding cleft. (Left) Illustrates the crystallographic structure published by Hu, *et al*. (blue in Fig. 1.9) which models the location of mono-Ub (yellow) and a ubiquitinated substrate, like MDM2 or p53 (pink) (3). (B) Structural comparison of the substrate-bound TRAF domains in comparison to the cryo-EM-derived model of USP7 TRAF domain. p53-bound (blue; PDB: 2F1X), MDM2-bound (purple; PDB: 2F1Y), EBNA1-bound (green; PDB: 1YY6), and EBNA1 peptide in orange. It is important to note that all peptides occupy the same binding groove, contacting the same residues. The TRAF domain appears to be rotated, orienting the peptide towards the interior. Ostensibly, this pulls the ubiquitinated peptide towards the active site. The region where EBNA1 would occupy appears occluded by the location of the H1-3 domains.

An analysis of existing structural data, and the new cryo-EM-derived USP7 structure suggest an interfacial region between TRAF and H1-3 domains. This data corroborates kinetics that showed the presence of the TRAF domain influences HUBL1-5 activation (66). The TRAF domain had a marked decrease in HUBL activation when H1-5 was present, but not H4-5. This suggests a mechanism whereby the TRAF domain negatively influences the activation of the H1-5 rheostat through an interaction with H1-3. The cryo-EM derived model appears to support this idea.

1.4 Discussion

Fig. 1.12 summarizes the intramolecular conformational movement of USP7 observed from structural analysis. Hu, *et al.*, suggested that due to high temperature factors of the TRAF domain linker electron density, that the domain was likely connected to the catalytic domain via a flexible linker consisting of up to ~10-15 residues (3). Our cryo-EM-derived model appears to show a different conformation of the TRAF domain that would be supported by such a flexible linker. This represents the 'TRAF domain' movement ostensibly pulling ubiquitinated substrates, towards the catalytic domain. This movement where the TRAF domain would bind a substrate, and the catalytic domain would cradle a single ubiquitin molecule would mimic the "base-cleavage" state described by Kategaya, *et al*. (65).

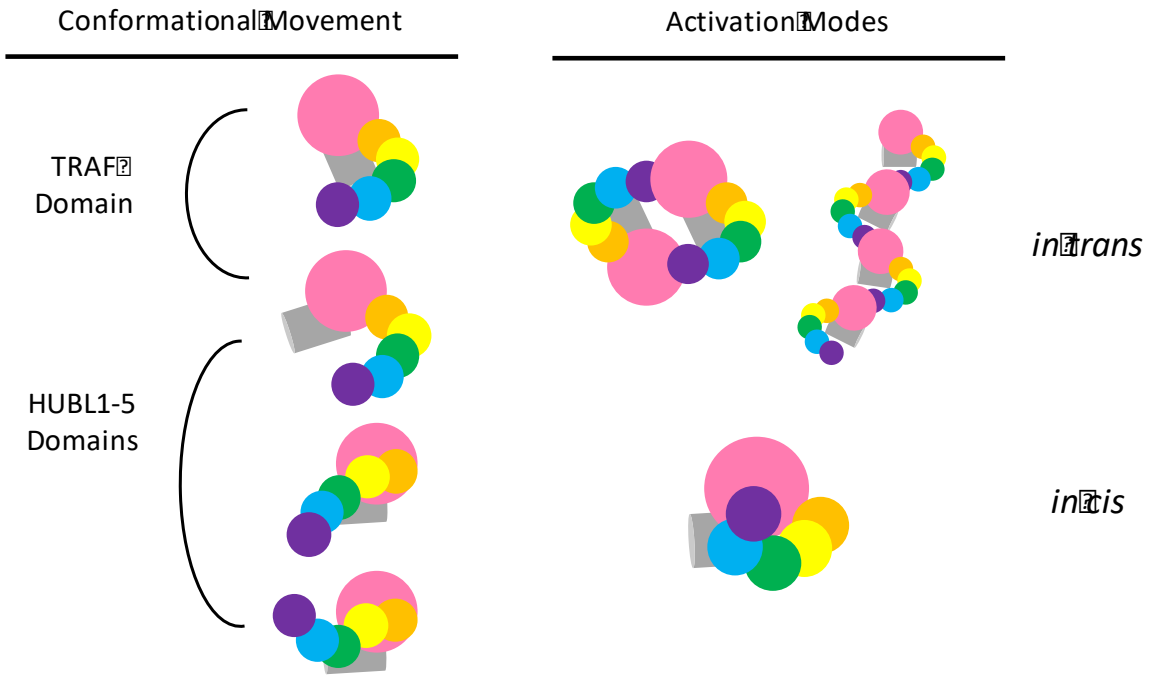


Figure 1.12. Schematic describing all hypothetical structural states of USP7. (Left) Conformational movement observed in cryo-EM particle dataset and crystallographic structures. Domain coloring is as follows: grey = TRAF; pink = catalytic domain; orange = H1; yellow = H2; green = H3; blue = H4; purple = H5. (Right) Hypothetical models of USP7 activation modes: *in trans* activation where a USP7 HUBL5 peptide activates a second USP7 molecule, and *in cis* activation where a USP7 HUBL5 peptide activates its own catalytic domain.

Current dogma suggests two mechanisms by which the HUBL domains can function as an activator of USP7's catalytic domain: 1) *in cis* activation and 2) *in trans* activation. The binding of the H5 peptide can be theoretically achieved in either activation mode (Fig. 1.18, right). When mono-Ub was applied to generate a thiol-ester adduct with the catalytic cysteine, we did not observe any *in cis* engaged particles. Further, the distance for the 19-residue H5 activation peptide to reach the catalytic domain *in cis* was too great in our model.

Our data suggest an update to the USP7 mechanism. Fig. 1.13 summarizes the states observed in our USP7 particle set. Contrary to the model published by Kim *et* al. (2019), our particle set did not show any H5 peptide-bound *in cis* states (51). Although it is important to note that this model depicts two distinct states: a closed HUBL conformation, where the tethered-rheostat draws nearer the catalytic domain but does not engage chemically. And, a second state where the HUBL domains are drawn close to the catalytic domain, and the HUBL5 peptide is engaged, making hydrogen-bonding and electrostatic interactions (51). It was hypothesized that covalent adducts of ubiquitin-bound USP7 could enrich for this state, perhaps decreasing particle heterogeneity. However, this was not the case. The EP_2 state (blue circle) was the predominant state in the particle set responsible for the result in the single particle analysis. Reference-free 2D averaging also illustrated significant movement by the HUBl3-5 domains, suggesting that the H5 peptide is not bound *in cis*, nor *in trans* during the EP_2 stage of catalysis.

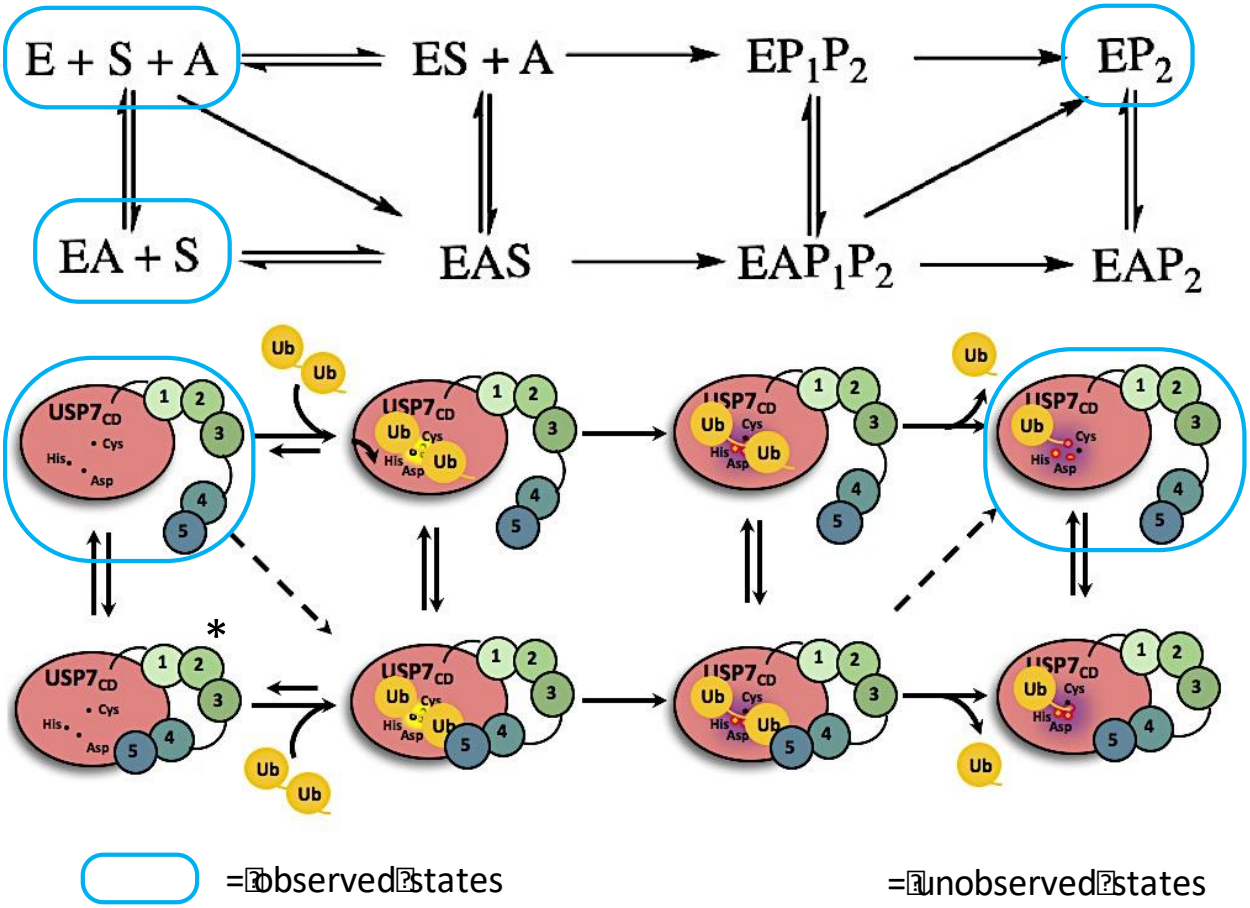


Figure 1.13. Schematic describing all hypothetical activation states of USP7 during catalysis. In this model of USP7 catalysis E = USP7 catalytic domain; S = ubiquitin isopeptide substrate (linkage between two ubiquitin molecules, or ubiquitin and target protein); A = activator tethered-rheostat (HUBL1-5); P_1 = initial product (the target protein or first ubiquitin to be released); P_2 = second product – i.e. the substrate bound into the catalytic domain cleft (ubiquitin cleaved from target protein). Since USP7~Ub-PA mimics the EP_2 state where USP7 is occupied with a mono-ubiquitin molecule after cleavage of the isopeptide bond and release of a ubiquitinated protein. In blue circles are the states that have been observed in previous structures or herein. Blacked out is a state that was expected in our particle set that was not observed. The asterisk denotes a state that exists, but not *in cis* like shown, but rather *in trans* with an artificial construct (50). It is important to note that *in trans* states can mimic the EA + S, EAS, and EAP_1P_2 states, wherein the presence of the activator is donated via a second USP7 molecule.

Reference-free 2D averaging and initial models suggest that in the ubiquitin-bound state (EP2, termed herein), the USP7 c-terminal HUBL5 peptide does not appear to be engaged to the activation cleft – neither *in cis* or *in trans*. According to the averaging and 3D projections obtained from the density, the HUBL domains appear to occupy a range of conformational states whereby

they remain flexible, moving via the hypothesized HUBL2-3 and HUBL3-4 hinges. Next, we successfully managed to use MDFF to flexible dock the atoms from crystallographic structures into the low-res volume to determine the position of domains with a high degree of agreement. This model was validated by MolProbity to measure the accuracy of conformation.

Comparison of the position of the TRAF domain in our cryo-EM map-derived model to the crystallographic structures published by Hu, *et al.*, showed the TRAF domain appears to adopt an alternative conformation, where the TRAF domain packs up closer to the HUBL1-3 domains. This supports the idea that there is a flexible linker between the catalytic domain and TRAF that allows for this movement, which was hypothesized by Hu, *et al*., from the high temperature factor of the density linking the TRAF and catalytic domains. This more packed conformation of TRAF positions the domain near HUBL1-3 appears to create a potential interfacial region in the mono-ubiquitin bound state (EP2 state; Fig. 1.1).

Molecular dynamics drug docking of our identified allosteric inhibitor and the pattern of inhibition that is elicited suggest that this interface is real. We validated the idea of an interaction between TRAF and HUBL1-3 by evaluating the effect on catalysis under steady-state kinetics in trans with HUBL1-5 versus HUBL1-3 with TRAF-USP7CD. Together these results suggest an alteration to the model of USP7 activation by which the TRAF domain negatively impacts the HUBL tethered-rheostat through an interaction with HUBL1-3. There is also an equilibrium in the mono-ubiquitin product-bound state, EP2 in our model, where the HUBL5 c-terminal peptide is mostly not engaged.

1.5 (references

1. Fraile, J. M., Quesada, V., Rodriguez, D., Freije, J. M. P., and Lopez-Otin, C. (2012) Deubiquitinases in cancer: new functions and therapeutic options. *Oncogene*. 31, 2373–2388.
2. Brooks, C.L., Gu, W. (2006) P53 ubiquitination: Mdm2 and beyond. *Mol. Cell*. 21, 307–315.
3. Hu, M., Gu, L., Li, M., Jeffrey, P. D., Gu, W., and Shi, Y. (2006) Structural Basis of Competitive Recognition of p53 and MDM2 by HAUSP/USP7: Implications for the Regulation of the p53–MDM2 Pathway. PLoS Biol. 4, e27.
4. Li, M., Brooks, C. L., Kon, N., and Gu, W. (2004) A Dynamic Role of HAUSP in the p53-Mdm2 Pathway. Mol. Cell. 13, 879–886.
5. Kon, N., Kobayashi, Y., Li, M., Brooks, C. L., Ludwig, T., and Gu, W. (2010) Inactivation of HAUSP in vivo modulates p53 function. Oncogene. 29, 1270–1279.
6. Cummins, J.M., Rago, C., Kohli, M., Kinzler, K.W., Lengauer, C., Vogelstein, B. (2004) Tumor suppression: disruption of HAUSP gene stabilizes p53. *Nature* 428, 1– 486.
7. Cummins, J. M., and Vogelstein, B. (2004) HAUSP is required for p53 destabilization. *Cell Cycle*. 3, 689–692.
8. Li, M., Chen, D., Shiloh, A., Luo, J., Nikolaev, A.Y., Qin, J., Gu, W. (2002) Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. *Nature* 416, 648–653.
9. Epping, M.T., Meijer, L.A.T., Krijgsman, O., Bos, J.L., Pandolfi, P.P., Bernards, R. (2011) TSPYL5 suppresses p53 levels and function by physical interaction with USP7. *Nat. Cell Biol.* 13, 102–108.
10. van der Knaap, J.A., Kumar, B.R.P., Moshkin, Y.M., Langenberg, K., Krijgsveld, J., Heck, A.J.R., Karch, F., Verrijzer, C.P. (2005) GMP synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. *Mol. Cell* 17, 695–707.
11. Alonso-de Vega, I., Mart.n, Y., Smits, V.A.J. (2014) USP7 controls Chk1 protein stability by direct deubiquitination. *Cell Cycle*. 13, 3921–3926.
12. Faustrup, H., Bekker-Jensen, S., Bartek, J., Lukas, J., Mailand, N. (2009) USP7 counteracts SCFbetaTrCP- but not APCCdh1-mediated proteolysis of Claspin. *J. Cell Biol*. 184, 13–19.
13. Schwertman, P., Vermeulen, W., Marteijn, J.A. (2013) UVSSA and USP7, a new couple in transcription-coupled DNA repair. *Chromosoma*. 122(4): 275-84.

14. Lecona, E., Narendra, V., Reinberg, D. (2015) USP7 cooperates with SCML2 to regulate the activity of PRC1. *Mol. Cell Biol*. 35(7): 1157-68.
15. Du, Z., Song, J., Wang, Y., Zhao, Y., Guda, K., Yang, S., Kao, H.-Y., Xu, Y., Willis, J., Markowitz, S.D., Sedwick, D., Ewing, R.M., Wang, Z. (2010) DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation driven ubiquitination. *Sci. Signal*. 3(146): ra80.
16. Cheng, J., Yang, H., Fang, J., Ma, L., Gong, R., Wang, P., Li, Z., and Xu, Y. (2015) Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. Nat. Commun. 6, 7023.
17. Saridakis, V., Sheng, Y., Sarkari, F., Holowaty, M. N., Shire, K., Nguyen, T., Zhang, R. G., Liao, J., Lee, W., Edwards, A. M., Arrowsmith, C. H., and Frappier, L. (2005) Structure of the p53 Binding Domain of HAUSP/USP7 Bound to Epstein-Barr Nuclear Antigen 1: Implications for EBV-Mediated Immortalization. Mol. Cell. 18, 25–36.
18. Sowa ME, Bennet EJ, Gygi SP, Harper JW. (2009) Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell*. 138(2): 389-403.
19. Zhu, Q., Sharma, N., He, J., Wani, G., Wani, A.A. (2015) USP7 deubiquitinase promotes ubiquitin-dependent DNA damage signaling by stabilizing RNF168. *Cell Cyc*le. 14(9): 1413-1425.
20. van der Horst, A., de Vries-Smits, A.M.M., Brenkman, A.B., van Triest, M.H., van den Broek, N., Colland, F., Maurice, M.M., Burgering, B.M.T. (2006) FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP. *Nat. Cell Biol*. 8: 1064–1073.
21. Song, M.S., Salmena, L., Carracedo, A., Egia, A., Lo-Coco, F., Teruya-Feldstein, J., Pandolfi, P.P. (2008) The deubiquitinylation and localization of PTEN are regulated by a HAUSP-PML network. *Nature*. 455: 813–817.
22. Hao, Y.-H., Fountain, M.D., Fon Tacer, K., Xia, F., Bi, W., Kang, S.-H.L., Patel, A., Rosenfeld, J.A., Le Caignec, C., Isidor, B., Krantz, I.D., Noon, S.E., Pfotenhauer, J.P., Morgan, T.M., Moran, R., Pedersen, R.C., Saenz, M.S., Schaaf, C.P., Potts, P.R. (2015) USP7 acts as a molecular rheostat to promote WASH-dependent endosomal protein recycling and is mutated in a human neurodevelopmental disorder. *Mol. Cell*. 59: 956–969.

23. Reddy, B. A., van der Knaap, J. A., Bot, A. G. M., Mohd-Sarip, A., Dekkers, D. H. W., Timmermans, M. A., Martens, J. W. M., Demmers, J. A. A., and Verrijzer, C. P. (2014) Nucleotide Biosynthetic Enzyme GMP Synthase Is a TRIM21-Controlled Relay of p53 Stabilization. *Mol. Cell*. **53**, 458–470
24. Holowaty, M.N., Frappier, L. (2004) HAUSP/USP7 as an Epstein-Barr virus target. *Biochem. Soc. Trans*. 32, 731–732.
25. Bhattacharya S, Ghosh MK. (2015) HAUSP regulates c-MYC expression via de-ubiquitination of TRRAP. *Cell. Onco*. 38: 265-77.
26. Tavana O, Li D, Dai C, Lopez G, Banarjee D, Kon N, Chen C, Califano A, Yamashiro DJ, Sun H, Gu W. (2016) HAUSP deubiquitinates and stabilizes N-Myc in neuroblastoma. *Nat. Med*. 22(10): 1180-1186.
27. Zhang ZM, Rothbart SB, Allison DF, Cai Q, Harrison JS, Li L, Wang Y, Strahl BD, Wang GG, Song J. (2015) An Allosteric Interaction Links USP7 to Deubiquitiantation and Chromatin Targeting of UHRF1. *Cell Rep*. 12(9): 1400-6.
28. Chen ST, Okada M, Nakato R, Izumi K, Bando M, Shirahige K. (2015) The Deubiquitinating Enzyme USP7 Regulates Androgen Receptor Activity by Modulating Its Binding to Chromatin. *J. Biol. Chem*. 290(35): 21713-23.
29. Zaman MM, Nomura T, Takagi T, Okumura T, Jin W, Shinagawa T, Tanaka Y, Ishii S. (2013) Ubiquitination/deubiquitination by the TRIM27-USP7 complex regulates tumor necrosis factor alpha-induced apoptosis. *Mol. Cell Biol*. 33(24): 4971-84.
30. Meulmeester, E., Maurice, M. M., Boutell, C., Teunisse, A. F. A. S., Ovaa, H., Abraham, T. E., Dirks, R. W., and Jochemsen, A. G. (2005) Loss of HAUSP-Mediated Deubiquitination Contributes to DNA Damage-Induced Destabilization of Hdmx and Hdm2. Mol. Cell. 18, 565–576.
31. Colland, F. (2010) The therapeutic potential of deubiquitinating enzyme inhibitors. Biochem. Soc. Trans. 38, 137–43.
32. Di Lello, P., Pastor, R., Murray, J. M., Blake, R. A., Cohen, F., Crawford, T. D., Drobnick, J., Drummond, J., Kategaya, L., Kleinheinz, T., Maurer, T., Rougé, L., Zhao, X., Wertz, I., Ndubaky, C., Tsui, V. (2017) Discovery of Small-Molecule Inhibitors of Ubiquitin Specific Protease 7 (USP7) Using Integrated NMR and in Silico Techniques. *J. Med. Chem*. 60(24): 10056-10070.

33. Gavory G., O'Dowd, C., R., Helm, M. D., Flasz J., Arkoudis E., Dossang A., Hughes C., Cassidy E., McClelland K., Odrzywol E., Page N., Barker O., Miel H., Harrison T. (2018). Discovery and characterization of highly potent and selective allosteric USP7 inhibitors. *Nat Chem Biol*. 14(2): 118-125.

34. Chauhan, D., Tian, Z., Nicholson, B., Kumar, K. G. S., Zhou, B., Carrasco, R., McDermott, J. L., Leach, C. A., Fulcinniti, M., Kodrasov, M. P., Weinstock, J., Kingsbury, W. D., Hideshima, T., Shah, P. K., Minvielle, S., Altun, M., Kessler, B. M., Orlowski, R., Richardson, P., Munshi, N., and Anderson, K. C. (2012) A Small Molecule Inhibitor of Ubiquitin-Specific Protease-7 Induces Apoptosis in Multiple Myeloma Cells and Overcomes Bortezomib Resistance. Cancer Cell. 22, 345–358.

35. Colland, F., Formstecher, E., Jacq, X., Reverdy, C., Planquette, C., Conrath, S., Trouplin, V., Bianchi, J., Aushev, V. N., Camonis, J., Calabrese, A., Borg-Capra, C., Sippl, W., Collura, V., Boissy, G., Rain, J.-C., Guedat, P., Delansorne, R., and Daviet, L. (2009) Small-molecule inhibitor of USP7/HAUSP ubiquitin protease stabilizes and activates p53 in cells. *Mol. Cancer Ther*. 8, 2286–2295.

36. Faesen, A. C., Luna-Vargas, M. P. A., Geurink, P. P., Clerici, M., Merkx, R., van Dijk, W. J., Hameed, D. S., El Oualid, F., Ovaa, H., and Sixma, T. K. (2011) The Differential Modulation of USP Activity by Internal Regulatory Domains, Interactors and Eight Ubiquitin Chain Types. Chem. Biol. 18, 1550–1561.

37. Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J.-W., Gu, W., Cohen, R. E., and Shi, Y. (2002) Crystal Structure of a UBP-Family Deubiquitinating Enzyme in Isolation and in Complex with Ubiquitin Aldehyde. *Cell*. **111**, 1041–1054.

38. Molland, K., Zhou, Q., and Mesecar, A. D. (2014) A 2.2 A resolution structure of the USP7 catalytic domain in a new space group elaborates upon structural rearrangements resulting from ubiquitin binding. *Acta Crystallogr. Sect. F*. **70**, 283–287.

39. Faesen, A. C., Luna-Vargas, M. P. A., and Sixma, T. K. (2012) The role of UBL domains in ubiquitin-specific proteases. *Biochem. Soc. Trans.* **40**, 539–545.

40. Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C., and Mesecar, A. D. (2006) Severe acute respiratory syndrome coronavirus papain-like protease: Structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5717–5722.

41. Clasman, J. R., Báez-Santos, Y. M., Mettelman, R. C., O'Brien, A., Baker, S. C., and Mesecar, A. D. (2017) X-ray Structure and Enzymatic Activity Profile of a Core Papain-like Protease of MERS Coronavirus with utility for structure-based drug design. *Sci. Rep.* **7**, 40292.
42. Mrosovsky, N. (1990) *Rheostasis : the physiology of change*, Oxford University Press
43. Bardwell, L. (2008) Signal transduction: turning a switch into a rheostat. *Curr. Biol.* **18**, R910-2.
44. O'Leary, C. E., Lewis, E. L., and Oliver, P. M. (2015) Ubiquitylation as a Rheostat for TCR Signaling: From Targeted Approaches Toward Global Profiling. *Front. Immunol.* **6**, 618.
45. Okazaki, T., Chikuma, S., Iwai, Y., Fagarasan, S., and Honjo, T. (2013) A rheostat for immune responses: the unique properties of PD-1 and their advantages for clinical application. *Nat. Immunol.* **14**, 1212–1218.
46. Brodin, P., Kärre, K., and Höglund, P. (2009) NK cell education: not an on-off switch but a tunable rheostat. *Trends Immunol.* **30**, 143–149.
47. Hao, Y.-H., Fountain, M.D., Fon Tacer, K., Xia, F., Bi, W., Kang, S.-H.L., Patel, A., Rosenfeld, J.A., Le Caignec, C., Isidor, B., Krantz, I.D., Noon, S.E., Pfotenhauer, J.P., Morgan, T.M., Moran, R., Pedersen, R.C., Saenz, M.S., Schaaf, C.P., Potts, P.R. USP7 acts as a molecular rheostat to promote WASH-dependent endosomal protein recycling and is mutated in a human neurodevelopmental disorder. Mol. Cell. 59, 956–969. 2015.
48. Kim, R. Q., van Dijk, W. J., and Sixma, T. K. (2016) Structure of USP7 catalytic domain and three Ubl-domains reveals a connector α-helix with regulatory role. *J. Struct. Biol.* **195**, 11–18.
49. Faesen, A. C., Dirac, A. M. G., Shanmugham, A., Ovaa, H., Perrakis, A., and Sixma, T. K. (2011) Mechanism of USP7/HAUSP Activation by Its C-Terminal Ubiquitin-like Domain and Allosteric Regulation by GMP-Synthetase. Mol. Cell. 44, 147–159.
50. Rougé, L., Bainbridge, T. W., Kwok, M., Tong, R., Di Lello, P., Wertz, I. E., Maurer, T., Ernst, J. A., and Murray, J. (2016) Molecular Understanding of USP7 Substrate Recognition and C-Terminal Activation. *Structure*. 24, 1335–1345.

51. Kim, R.Q., Geurink, P.P., Mulder, M.P.C., Fish, A., Ekkebus R., El Oualid, F., van Djik, W.J., van Dalen, D., Ovaa, H., van Igen, H., Sixma, T.K. (2019). Kinetic analysis of multistep USP7 mechanism shows critical role for target protein in activity. Nat. Commun. 10, doi: 10.1038/s41467-018-08231-5.

52. Pfoh, R., Lacdao, I. K., Georges, A. A., Capar, A., Zheng, H., Frappier, L., and Saridakis, V. (2015) Crystal Structure of USP7 Ubiquitin-like Domains with an ICP0 Peptide Reveals a Novel Mechanism Used by Viral and Cellular Proteins to Target USP7. PLoS Pathog. 11, 1–23.

53. Chari, A., Haselbach, D., Kirves, J. M., Ohmer, J., Paknia, E., Fischer N., Ganichkin O., Möller, V., Frye, J. J., Petzold, G., Jarvis, M., Tietzel, M., Grimm, C., Peters, J. M., Schulman, B. A., Tittmann, K., Markl, J., Fischer, U., Stark H. (2015) ProteoPlex – stability optimization of macromolecular complexes by spare-matrix screening of chemical space. *Nat. Meth.* 12(9): 859-865.

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

55. Rohou, A., Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struc. Biol*. 192(2): 216-221.

56. Zivanov, J., Nakane, T., Forsberg B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., Scheres S. H. J. (2018) New tools for automated high-resolution cryo-EM strucutre determination in RELION-3. *eLife*. 7:e42166.

57. Punjani, A., Rubenstein, J. L., Fleet, D. J., Brubaker, M. A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Meth*. 14(3): 290-96.

58. Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I. Ludtke, S.J. (2007) EMAN2: an extensible image processing suite for electron microscopy. *J. Struc. Biol*. 157(1): 38-46.

59. Li, L, Klose, T., Sun, C., Liu, Y., Jiang, W. (2018). 2.5 Å Resolution Cryo-EM Structure of Human Apo-ferritin Using an Optimized Workflow for Volta Phase Plate. Microsc. Microanal. 24 doi: 10.1017/S1431927618004993.

60. Trabuco, L. G., Villa, E., Mitra, K., Frank, J., Schulten K. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*. 16: 673-683.

61. Nakane, T., Kimanius, D., Lindahl, E., Scheres, S. H. W. (2018). Characterization of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife*. 7:e36861.

62. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., Ferrin, T. E. (2004) UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25(13): 1605-12.

63. Chen, V. B., Arendall III, B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Cryst.* D66: 12-21.

64. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., Ferrin, T. E. (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics*. 7(339). https://doi.org/10.1186/1471-2105-7-339.

65. Kategaya, L., Di Lello, P.D., Rougé L., *et al*. (2017) USP7 small-molecular inhibitors interfere with ubiquitin binding. *Nature*. 550: 534-538. https://doi.org/10.1038/nature24006.

66. Hortland, N. (2016) Defining the regulatory determinants in substrate catalysis by biochemical, biophysical, and kinetic studies for the development of specific small-molecule inhibitors of ubiquitin specific proteases 7 and 17. *Purdue University Open Acess Dissertations*. 1461. https://docs.lib.purdue.edu/open_access_dissertations/1461.

CHAPTER 2. NON-COMPETITIVE PYRAZOLE INHIBITORS OF USP7 DESTABILIZE THE TRAF AND H1-3 COMPONENTS OF THE USP7 REGULATORY REHOSTAT

2.1 Introduction

Homeostasis in the cell is largely controlled by a dynamic process of ubiquitination and deubiquitination that results in the carefully controlled half-life of proteins (1-4). In humans, there are 6 families of deubiquitinating enzymes (DUBs), ubiquitin carboxy-terminal hydrolases (UCHs), ovarian-tumor proteases (OTUs), Machado-Joseph disease protein domain proteases (MJDs), JAMM/MPN domain-associated metalloproteases, monocyte chemotactic protein-induced proteins (MCPIPs), and ubiquitin-specific proteases (USPs) – isopeptidases that catalyze the removal of ubiquitin from proteins – preventing degradation (1-4). The largest family of human DUBs is the USP family, named for its structural and mechanistic similarity (1-4). The defining structural characteristics of the USP family are a conserved fingers-palm-thumb papain-like catalytic domain fold and the canonical Cys-His-Asp/Asn catalytic triad (1-8).

The domains surrounding the catalytic domain, e.g. TRAF and HUBL1-5 (Fig. 2.1), are thought to provide each USP with an individual ubiquitin-cleavage fingerprint, defining substrate specificity and regulation (1-4, 9-10). USP7 is a unique member of the USP family due to its 6 ancillary domains which lack any significant homology with related enzymes (11). The domain architecture includes a catalytic domain flanked by a tumor necrosis factor receptor-associated factor (TRAF) substrate recognition domain, and five HAUSP ubiquitin-like (HUBL) domains, termed 1-5, that are hypothesized to function as a tethered-rheostat in regulation of USP7 catalytic activity (11). Biochemical and structural studies have uncovered a role for the TRAF domain in protein-protein interactions with DUB substrate targets such as tumor suppressor p53, ubiquitin ligase MDM2, and viral DNA-binding protein EBNA1 (Table 1.1). HUBL1-5 domains have been found to extensively control the activity of USP7, truncations of which result in up to an 80-fold reduction in K_m and k_{cat} (11-14).

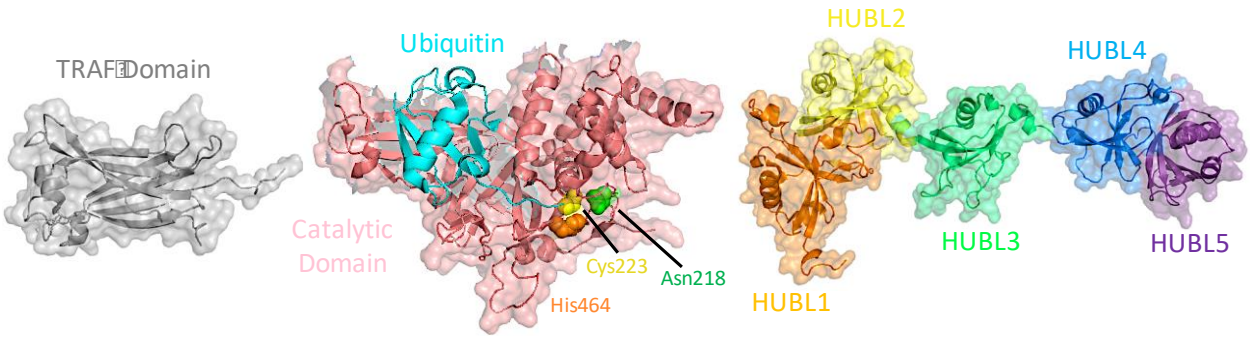


Figure 2.1. USP7 catalytic domain (CD) and surrounding domains. Grey = TRAF (PDB: 2F1W); pink = CD (1NEB); orange = HUBL1, yellow = HUBL2, green = HUBL3, blue = HUBL4, purple = HUBL5 (PDB: 2YLM). Mono ubiquitin is shown as a ribbon diagram colored in cyan bond to CD. The catalytic triad residues within the CD are colored as follows, green = Asn218, yellow = Cys223, orange = His464.

USP7 is a particularly potent oncoprotein that has a demonstrated role in many cellular functions (see Table 1.1; Chapter 1). A major caveat to the development of specific inhibitors as cancer therapeutics in the USP family is that studies have focused mainly on the conserved catalytic domain, rarely accounting for the unique surrounding or accessory domains and how they regulate each USP (15-21). There are 58 USPs in a human cell and each shares a similar structure in terms of their catalytic domain, catalytic triad residues, and residues that bind the c-terminus of ubiquitin (1). To design a selective USP7 inhibitor, these other USPs must be avoided. Competitive inhibitors, however, have proven effective in targeting USP7 with some degree of specificity among USPs (15-21). These studies have resulted in several therapeutic leads as inhibitors of USP7 generally targeted to the active site, or to an adjacent site in the catalytic domain (Fig. 2.6) (15-21). Consequently, it is hypothesized that optimization of lead competitive inhibitors may result in off-target effects through inhibition of closely related enzymes, or toxicity effects in healthy cells that depend on USP7 DUB activity.

Previously in our lab, Dr. Nicole Hjortland identified a series of small-molecule inhibitors that exhibited a robust structure-activity-relationship (SAR) along with selectivity for USP7 over other USPs (38). This series of pyrazole-derived non-competitive inhibitors of USP7 were first discovered through a high-throughput compound screen followed by hit optimization and ultimately resulted in the identification of compound **4** as a potent inhibitor **(**Fig. 2.2). A Lineweaver-Burk plot of the steady-state kinetic data revealed that compound 4 has no statistically significant effect on the K_m values for ubiquitin-rhodamine as a substrate with increasing inhibitor

concentrations (Fig. 2.2, left panel) (38). However, compound 4 decreases the turnover rate, i.e. k_{cat}, of USP7 (Fig. 2.2, right panel) (38). Compound **4** was further validated as a lead compound by measuring its inhibitory effect on di-ubiquitin peptide cleavage via SDS-PAGE gel analysis, which corroborated the steady-state inhibition results (38).

Compound **4** and its analogs also revealed a pattern of USP7 inhibition contingent upon the presence of the TRAF and HUBL1-3 domains (38). These results led to the hypothesis that these inhibitors do not bind within the active site, or within the ubiquitin-binding cleft. Instead, it was hypothesized that this class of compounds might bind at an interfacial region somewhere between the TRAF and HUBL1-3 domains, or to a pocket within one of the HUBL1-3 domains. Another possible explanation for these observations would be that the non-competitive inhibition may result from constraining the USP7 rheostat (HUBL1-3 domains) in an unfavorable conformation for catalysis, thereby lowering the turnover rate for the enzyme.

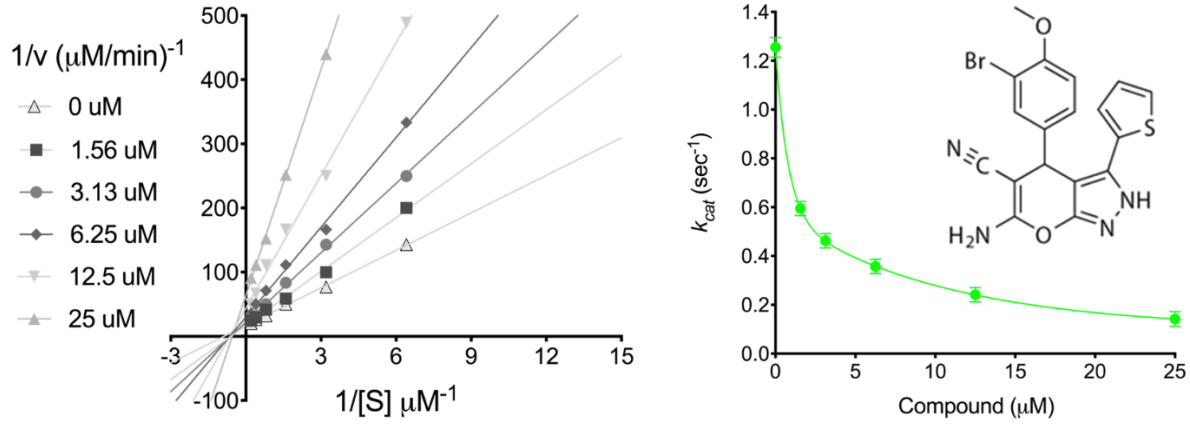


Figure 2.2. Inhibition of USP7 by compound 4 via steady-state kinetic studies and analysis. (Left) Lineweaver-Burk plot of steady-state kinetic rates versus substrate (Ub-Rhodamine 110) at fixed, variable compound **4** concentrations. Compound 4 (inlet, right) exhibits non-competitive inhibition as judged by the intersection on the x-axis (-1/Km). (Right) Plot of the individual $k_{cat,max}$ values with increasing inhibitor concentration showing that the compound decreases the turnover rate as expected for non-competitive inhibitors. These data were adapted from the dissertation of N. Hjortland (38).

The goal of this study described here in Chapter 2 was to determine which of USP7's domains are responsible for compound binding using differential scanning fluorimetry (DSF), computational docking, and molecular dynamics simulation on the cryo-EM derived structural model of USP7 determined in Chapter 1. The ability of compound **4** to bind to either the E or ES complexes of USP7 was investigated by DSF. The results suggest that a candidate drug-binding

location between HUBL1-3 and TRAF interfaces may exist, which may explain the non-competitive inhibitory mechanism of this series of compounds.

2.2 Methods & Materials

2.2.1 DSF methods for measuring USP7 stability with pyrazole compound treatment

Fluorescence signal (Ex 468 nm/507 nm Em) as a function of temperature (range: 25°C – 99°C) was collected on a QS5 (Applied Biosciences) qPCR instrument to measure protein melting transitions. Fluorescence data were measured at one minute intervals at a ramp rate of 1°C/min. The reaction volume for each well was 20 μL and contained a final USP7 concentration of 1.0 mg/mL and a 1x final concentration of GloMelt™ (Biotium). Concentrations of inhibitor were varied from 0.1 – 100 μM. Each concentration of inhibitor was diluted a final assay concentration of DMSO of ~2% (v/v) in the final assay volume. This ensured the lowest permissible DMSO to allow inhibitor solubility, while minimizing solvent effects on the protein. The buffer used for the inhibitor, protein, and dye dilution was identical to the SEC buffer described in Chapter 1. USP7FL, $USP7_{CD}$, TRAF, HUBL1-5, TRAF-$USP7_{CD}$, $USP7_{CD}$-H4-5, and $USP7_{CD}$-H1-5 and compounds were mixed together and allowed to incubate for 5 min, and the reaction was initiated by the addition of the dye. Plates were centrifuged at 1,000 x g for 30 s after sealing, prior to recording fluorescence, to eliminate air bubbles. All protein constructs were expressed and purified as described in Dr. Nicole Hjortland's dissertation (38).

The raw fluorescence data were exported from the 'Thermal Protein Analysis' software (Applied Biosciences) and the raw data was analyzed with a program to normalize the raw fluorescence values to more accurately compare melting temperatures between samples (Eq. 2.1).

$$Normalized\ Fluorescence\ (AU) = \frac{value - min}{max - min} \qquad \text{Eq. 2.1}$$

The fluorescence values were normalized between 0-1 using Eq. 2.1, as described before in Appendix A. With all fluorescence amplitudes normalized between 0-1, the temperature at 0.5 normalized fluorescence value (half maximal fluorescence) was identified as the nominal melting temperature (T_m) as suggested by Chari *et al*. (39) ProteoPlex protein stability assay.

2.2.2 Molecular dynamics for inhibitor binding candidate location

Docking of compound **4** to USP7 was conducted using Maestro in the Schrödinger software suite (Schrödinger LLC, New York, NY) (23-25). Compound **4** was prepared for docking using the Maestro LigPrep plugin (23-25). Compound **4** was docked to a reconstructed USP7 structure based on electron microscopy data using Glide (grid-based ligand docking with energetics) docking (23-25). For docking, the OPLS_2005 force field was used, and ionization states for compound **4** were determined under pH 8.0 ±2.0 using Epik (23-25). A cubic docking grid was generated to fit the entire USP7 molecule so that the binding location was unbiased. Docking was performed using Glide's XP (extra precision) setting with the ligand treated flexibly (23-25). From the docking run, the top 4 scoring compound **4** poses were retained, with the highest shown in Fig. 2.6.

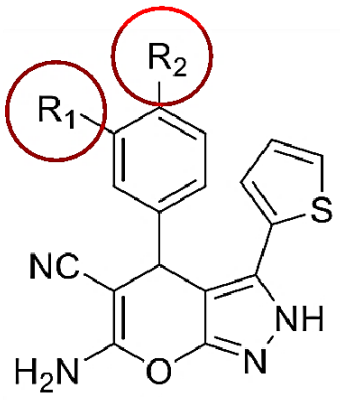
2.3 Results

2.3.1 Differential scanning fluorimetry (DSF) helps to identify the inhibitor binding domains

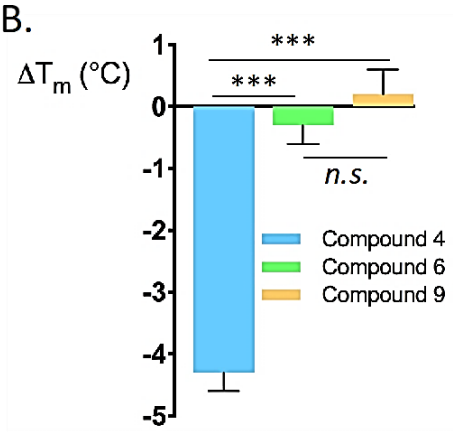
DSF has been used to identify crystallization conditions and hit compounds for over 2 decades (26-27). A DSF assay for measuring ligand binding involves measuring the fluorescence of a fluorescent dye (GloMelt™ in this case) binding to a protein as it unfolds due to heating (26). Ligand binding may alter the melting temperature of a protein, as observed during the transition state – or melting curve of the protein – by altering the unfolding temperature of a protein (26). Differential scanning fluorimetry has been successfully adapted for high-throughput ligand screening, as has been performed with poly-ADP ribosylating enzyme tankyrase, with the tuberculosis (TB) drug target, pantothenate synthetase (PS), and with an antibody such as the anti-cocaine monoclonal antibody h2E2 (28-30). In cases where larger compound screens are used, DSF has proven a useful strategy for validating a first-round of hits from HTS, such as was performed with tuberculosis (TB) PLP-dependent transaminase (BioA) (31). DSF has been proven as a useful method for aiding in the determination of the mechanism-of-action of inhibitors of multiple systems such as with hydroxyprostaglandin dehydrogenase, firefly luciferase, and glutathione S-transferase (32-34). It was hypothesized that a difference in melting temperature (T_m) as measured by DSF would be useful in determining which truncated forms of USP7 are affected by inhibitor binding.

The melting temperature for USP7 and a series of its truncated constructs was determined both in the presence and absence of compound **4** over the concentration of 0.1-100 µM. At compound concentrations of 100 µM, compound **4** had a destabilizing effect on $USP7_{FL}$ with a ΔT_m of -4.3°C ± 0.5°C compared to control compounds **6** and **9** (ΔT_m = -0.3°C ± 0.5°C and +0.2 ± 0.3°C) that do not inhibit USP7 (Fig. 2.3). The melting temperature effects are reported as a deviation in the melting temperature of USP7 in the presence of compound versus in the absence of compound (n = 3; reported as *mean* ± *S.D.*) with statistical significance calculated using a paired t-test. These three compounds were used as controls from the SAR kinetics to validate DSF as a method for determining the stabilizing or destabilizing effects of compounds on USP7. There was a statistically significant difference ($p < 0.001$) in the melting temperature of $USP7_{FL}$ between compound **4**, known to bind and inhibit USP7, and compounds **6** and **9** that do not bind and inhibit USP7 (Fig. 2.3) (38).

A.



B.



C.

Compound	R_1	R_2	IC_{50} (µM)	Max % Inhibition
4	Br	OCH_3	4.1 ± 0.5	98 ± 4
6	Br	H	> 100	-
9	H	OH	> 100	-

Figure 2.3. Differential scanning fluorimetry (DSF) results of the pyrazole-scaffold derived inhibitors binding to USP7. DSF was used as an orthogonal method to corroborate the results from enzyme kinetics. (A) Structure of pyrazole-scaffold derived inhibitors. SAR studies reveals the importance of the R1 and R2 substituents as defined in (B). (C) The effect of 100 µM of compounds 4, 6 and 9 on $USP7_{FL}$ thermal stability. Columns represents the *mean* ± S.D. of triplicate measurements; statistical significance was determined with a two-sample t-test (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$) between the melting temp of treated and untreated sample. Testing USP7 constructs for binding by DSF

2.3.2 The role of TRAF and HUBL domains in binding and inhibitor of USP7 by compounds as determined by DSF

To determine the roles of the TRAF and HUBL domains in binding the inhibitors of USP7. DSF was used as an orthogonal method to investigate whether the compounds are binding directly to specific domains of USP7. This method has an advantage over kinetic studies because it is potentially able to determine if a compound binds an isolated, non-catalytic domain. Kinetic studies must have the catalytic domain present to determine a functional readout and it has been shown that the catalytic domain alone is unaffected by compounds such as compound **4** (38). The binding of compound **4** to USP7 domain constructs was therefore evaluated using DSF.

The ability of compound **4** to bind to the isolated catalytic domain of USP7 was determined first since it was previously shown that it does not inhibit its ability to catalyze hydrolysis of the ubiquitin-rhodamine substrate. As shown in Figure 2.4B, $USP7_{CD}$ in the presence of 100 μM compound exhibited no difference in melting temperature as the ΔT_m was only -0.9 °C ± 0.3°C, which is not statistically significant ($p = 0.055$) from the control in the absence of compound. These data and the kinetics data support the hypothesis that these inhibitors do not bind directly to the catalytic domain or that other domains are required for binding to the catalytic domain.

To test whether the TRAF domain is involved in binding compound **4**, the TRAF deletion construct, $USP7_{CD}$-H1-5, was subjected to DSF analysis. The data shown in Figure 2.4B illustrates that compound **4** destabilizes the $USP7_{CD}$-H1-5 construct by -2.6 ± 0.3 °C ($p < 0.001$) indicating that the compound is still able to bind these domains and that the TRAF domain is not entirely necessary for binding of the inhibitor. The decrease in ΔT_m of ~1.7 °C between $USP7_{FL}$ and $USP7_{CD}$-H1-5 domains suggest that the TRAF domain contributes to the overall binding destabilization brought about by compound **4**, but that it is not entirely necessary.

The ability of compound **4** to bind to the TRAF-$USP7_{CD}$ or TRAF domain alone was determined to assess whether the TRAF domain could bind the inhibitor without the catalytic or HUBL domains. Figure 2.4B shows that neither the TRAF-$USP7_{CD}$ or TRAF domains are capable of binding compound **4**. TRAF-$USP7_{CD}$ had a difference in the melting temperature (-1.0 ± 0.3°C), and although it was determined to be somewhat statistically significant ($p = 0.047$), previous kinetics data with compound **4** show no inhibition against this construct. Therefore, it is concluded that there is no direct interaction of compound **4** with this construct.

Interestingly, only constructs with HUBL1-3 present exhibited any significant thermal shifts upon compound **4** binding. A construct first described by Rougé, *et al.*, was used to determine if the c-terminal activator domains HUBL4-5 were kept intact without the HUBL1-3 portion of the tethered-rheostat, and if any effect on thermal stability when treated with compound **4** could be detected (13). As shown in Fig. 2.4B, the HUBL1-3 truncation, USP7$_{CD}$H4-5, exhibited no statistically significant difference in melting temperature in the presence of compound **4** (-0.8 °C ± 0.4 °C; p = 0.07), especially when compared to USP7$_{CD}$-H1-5 (p = 0.002). Correspondingly, HUBL1-5 on its own showed a robust change in thermal satiability when treated with compound 4 (-3.0 °C ± 0.3 °C; p < 0.001). The DSF results lead us to conclude that HUBL1-3, and perhaps to a lesser degree the TRAF domain, are responsible for the interaction of USP7 with these compounds and subsequently inhibition of USP7 catalytic activity.

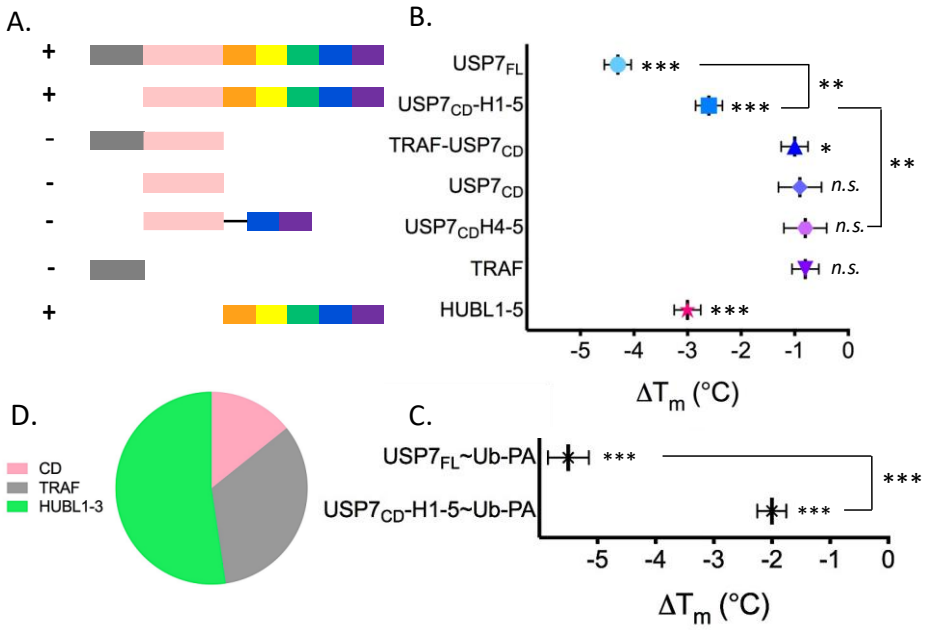


Figure 2.4. DSF results for APII-USP7 pyrazole-derived inhibitors of USP7. (A) Summary of inhibitor binding (+ denotes statistically significant change in thermal stability; - denotes no statistically significant change in thermal stability), schematic of USP7 constructs used for the experiment; grey = TRAF domain, pink = catalytic domain, orange = H1, yellow = H2, green = H3, blue = H4, purple = H5. (B) Thermal stability changes for the 100 μM inhibitor-treated construct vs the untreated construct. The untreated constructs were in buffer-DMSO composition identical to the inhibitor-treated samples. (C) Ub-PA conjugated samples. For all experiments in (B) and (C): points are *mean* ± *SD*; n = 3. Statistical significance was calculated with a paired t-test between treated and untreated melting temp: *** p <0.001, ** p <0.01, * p <0.05. (D) Relative contributions of each USP7 domain for compound 4 binding; grey = TRAF domain, pink = CD, green = HUBL1-3.

2.3.3 Evaluating compound binding to ES complex by DSF

The DSF binding studies described in the previous section were performed in the absence of substrate. The binding studies with constructs that contained the USP7 catalytic domain would represent binding to the free enzyme represented by E (vide infra Fig. 2.8). Since a pure non-competitive inhibitor can theoretically bind to the ES complex as well, the ability of compound **4** to bind to an ES complex was also evaluated. This was performed using a covalent vinyl sulfone adduct made through the addition of a ubiquitin-propargylamine (Ub-PA) probe. Constructs that could be saturated with the Ub-PA probe – $USP7_{FL}$ and $USP7_{CD}$-H1-5 – were tested for thermal shifts with compound **4** (Fig. 2.4C). There was a statistically significant decrease in thermal stability in both treated versus untreated ES complexes ($p < 0.001$). The magnitude of thermal shift was alike for the E and ES complexes: -4.3°C ± 0.5°C vs -5.5°C ± 0.5°C for full-length E and ES, respectively, and -2.6°C ± 0.4°C vs -2.0°C ± 0.3°C for $USP7_{CD}$-H1-5 E and ES, respectively (Fig. 2.4, B & C).

Fig. 2.3 D summarizes the DSF data as it relates to the observed contribution of each domain to thermal stability effects by compound **4**. Overall, these data corroborate the kinetics evaluation performed by Dr. Nicole Hjortland, suggesting that compound **4** makes significant contact with USP7 only when the HUBL1-3 domains or HUBL1-3 and TRAF domain are present (29). These data suggest that compound **4** may bind at an interfacial region connecting disparate portions between TRAF and HUBL1-3. Moreover, compound **4** can occupy a binding site that is available in both the free enzyme (E) and ES complex allowing it to inhibit USP7 through mixed, partial, non-competitive inhibition. According to the large destabilizing effect measured from thermal stability DSF, it can be inferred that compound **4** generates a large degree of instability in USP7 and may be affecting tethered-rheostat movement.

The DSF experiments were performed to observe differences in thermal stability of each of the enzymes domains – both apo and mono-ubiquitin-bound forms – to determine which were affected by the presence of the drug. Only when the TRAF domain and catalytic domain, HUBL domains 1-3 and catalytic domain, or HUBL domains alone were present did the drug bind (Fig. 2.4 A & B). The drug had no measureable effect on the catalytic domain alone, or the TRAF domain alone. Interestingly, although there was a small statistically significant decrease in thermal stability for the TRAF-CD construct ($p = 0.046$), the enzymatic activity of this construct did not appear to be affected by the presence of the inhibitor (29). It is thought that the drug binds the

domains to hold the enzyme in an unfavorable conformation that drives down the k_{cat} – acting as a noncompetitive inhibitor. It is hypothesized that in the absence of HUBL1-3, the TRAF domain – and thus the inhibitor – has little effect on driving the k_{cat} down any further in the TRAF- $USP7_{CD}$ construct.

According to the results summarized in Fig. 2.3 C, both apo and substrate-bound forms of USP7 were significantly destabilized by the inhibitor. This corroborates the idea that the inhibitor works as a mixed, non-competitive inhibitor. It is thought that all domains must be present for the drug to coordinate and have full inhibition on catalysis.

2.3.4 Glide molecular docking of inhibitor validates DSF and kinetics data

Glide drug docking software was used to dock compound **4**, without free-energy restraints imposed on the system, into the cryo-EM-derived model of USP7 (see section 1.6 and 1.7; Chapter 1) (23-25). Additionally, there was no voxel space restraint in that we did not confine inhibitor docking to any region of the model to reduce bias in finding candidate locations. From the kinetics and DSF experimentation, it was hypothesized that the inhibitor may bind somewhere in the region denoted by the red circle in Fig. 2.4 below.

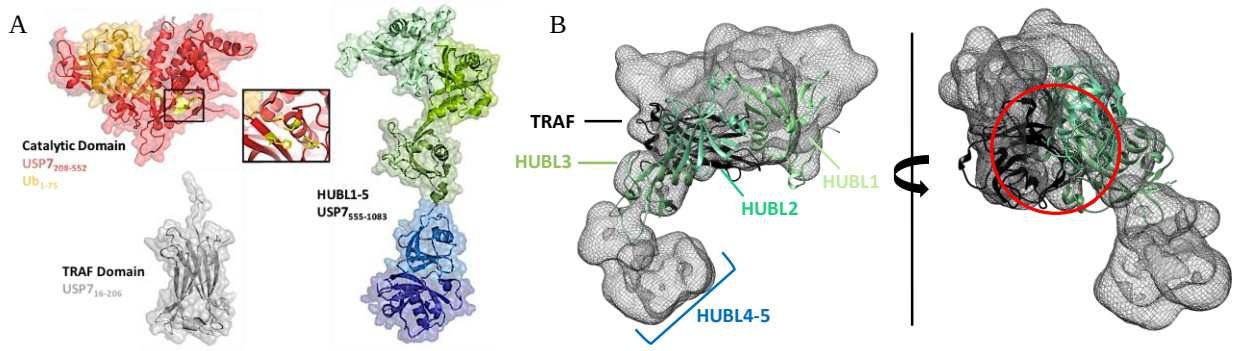


Figure 2.5. TRAF-HUBL1-3 interfacial region from cryo-Em derived model. (Left) USP7 catalytic domain with ubiquitin aldehyde bound (PDB: 1NBF; catalytic residues highlighted in inlet), TRAF domain (PDB: 2F1W), and HUBL1-5 (PDB: 2YLM). (Right) Cryo-EM-derived model of USP7 with TRAF-HUBL1-3 interfacial region highlighted by red circle.

It was found that compound **4** docked into the binding pocket illustrated in Fig. 2.6 with a free energy of -4.1 kJ/mol. All inhibitor orientations showed binding in this same binding pocket

depicted in Fig. 2.6 and the drug appeared to have no significant or meaningful interaction elsewhere. The candidate location residues highlighted for the interaction in Fig. 2.6 correspond to the TRAF domain (N10, T12, V13, M14, R16), the catalytic domain (N351, Y359), the HUBL2 domain (R634), and the HUBL3 domain (R730, D732).

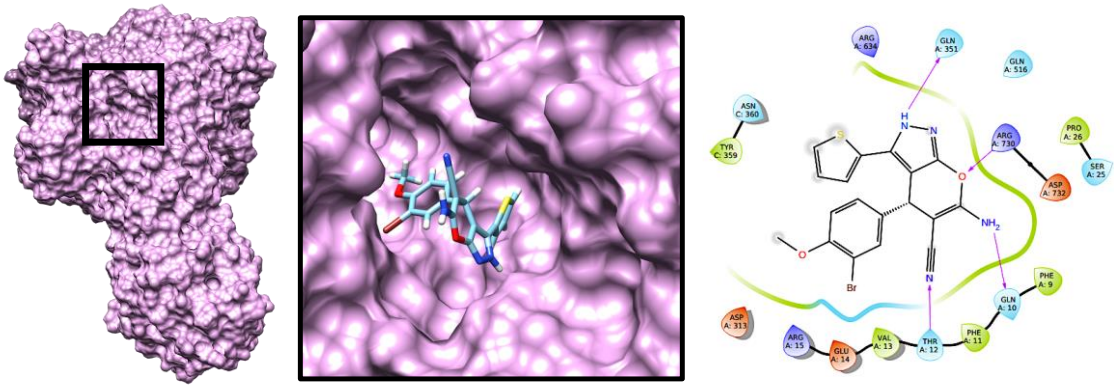


Figure 2.6. TRAF-HUBL1-3 interfacial region and corroborated by drug binding. (Left) Cryo-EM-derived model of USP7 with drug-binding pocket highlighted by black frame. (Middle) Close-up of binding pocket with APII-004 compound modeled in. (Right) Glide docking model of APII-004 with residues Gln10, Thr12, Val13, Glu14, Arg15 from the TRAF domain, Gln351 of the catalytic domain, and Arg730 from H3 domain.

The free energy of binding for compound **4** (-4.1 kJ/mol) may be related to the determined Ki, or rate of inhibition of compound 4, which is a measure of affinity for the compound of both the free enzyme (E) and enzyme-substrate (ES) complex, as depicted in Fig. 2.7 below. The Ki was experimentally determined using steady-state kinetics analysis to be 4.3 ± 0.2 (39). This experimentally determined Ki would suggest a theoretical ΔG from the equilibrium constant equation to be ΔG = -3.6 kJ/mol, which represents an approximately 13% difference between the theoretical ΔG and the ΔG measurement from drug docking experiments, showing good agreement. It is important to note that the compound **4** orientation depicted in Fig. 2.6 is one of several highly similar orientations providing binding energies on the order of -2 kJ/mol to -4 kJ/mol, in close agreement mathematically with the theoretical ΔG calculated from the Ki determined by N. Hjorltand (38).

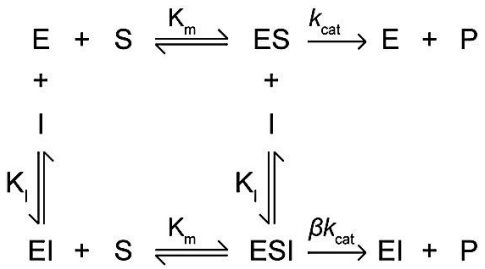


Figure 2.7. Theoretical inhibitor schematic of mixed, non-competitive inhibitor of USP7. A mixed, non-competitive inhibitor (I), by definition, has no effect on the K_m of an enzyme (E) for its substrate (S), and can bind both the E and ES complexes. Compound 4 (Fig. 2.7) does not alter USP7 K_m, but significantly drives down the turnover rate (k_{cat}), resulting in a reduced turnover rate, βk_{cat}.

This binding location represents a novel inhibitor binding site for USP7, as all previously published inhibitors of USP7 function as allosteric inhibitors binding the catalytic triad residues, or to an adjacent site in the catalytic domain (Fig. 2.7).

2.4 Discussion

USP7 is the etiological agent of several cancers and represents an attractive drug target due to its unique domain architecture and mechanism of inter- and intramolecular regulation. Several research groups have successfully identified small molecule inhibitors with demonstrated efficacy that are reportedly specific against USP7 summarized in Fig. 2.8, below (15-22, 35-37). Of these inhibitors few have accompanying *in vivo* data such as Chauhan *et al*. who have identified a compound series that inhibits cell growth and induces apoptosis in multiple myeloma cell lines by inhibiting USP7 (35). Colland *et al*. have developed a specific inhibitor of USP7 that induces apoptosis, and shows promise as part of a combinatorial therapy *in vivo* (36-27). However, all referenced compounds are targeted to the conserved catalytic domain of USP7 and may have significant off target effects when tested in the cell and in the clinic.

An alternative approach to targeting USP7 may be to develop non-competitive inhibitors that require the ancillary domains unique to USP7, perhaps providing selectivity and reduced toxicity in the cell. The [2, 3-c] pyrazole compounds described herein were identified from an inhibitory high-throughput screening trial as a compound that can inhibit mono-ubiquitin processing of $USP7_{FL}$, but not inhibit the catalytic domain. Inhibition was confirmed with a

chemically distinct, yet more physiologically-relevant, di-ubiquitin substrate by SDS-PAGE analysis.

DSF thermal shift experimentation showed that the pyrazole-derived compounds can be accommodated by both the E and ES complexes of USP7. The compound required HUBL1-3 and the TRAF domain to achieve full potency, suggesting the inhibitor is binding at an interfacial region between the domains (Fig. 2.7). The dimensions of the compound suggest these domains could reside ~5-15 Å distance from one another. It is hypothesized that binding across this interface would impose a strain on the HUBL1-3 rheostat, forcing an unfavorable conformation for ubiquitin processing, potentially explaining the decreased thermal stability observed after treatment.

The importance of H1-3 in the intramolecular activation of USP7 has long been overlooked and H1-3 were previously assumed to only serve as a scaffold for the binding of protein substrates DNMT1, UHRF1, and ICP0, or the allosteric activator GMPS (see Table 1.1 in Chapter 1). However, it appears H1-3 functions as part of the rheostat to regulate the level of influence the HUBL domains 4 and 5 have on USP7 catalysis (29). This model would suggest that interactions with H1-3, such as allosteric activator GMPS, adjusts the H1-3 rheostat to manipulate where it lies in relation to TRAF, and the ability of HUBL4-5 to activate the catalytic domain.

The use of an inhibitor that locks the HUBL rheostat in an unfavorable position may be an especially useful feature for targeting USP7 in cases such as GMPS overexpression. Whereas, competitive inhibitors of USP7 may still allow for *in trans* activation due to an unaffected rheostat; non-competitive inhibitors that impose unfavorable rheostat positioning may be more useful in abating *in trans* activation.

Overall, our results suggest the pyrazole compounds inhibit by disrupting the intramolecular regulation of USP7 imposed by the TRAF and HUBL1-3 domains. This mode of inhibition has proven successful as a biochemical tool in understanding USP7's mechanism of intramolecular activation. With further optimization, these compounds could be developed into promising small molecule inhibitors of USP7. As partial, non-competitive inhibitors, it is thought the compounds could greatly reduce the activity of USP7, but perhaps allow a small level of activity, lessening the effects of full inhibition of USP7 in the cell. Furthermore, inhibitors developed to target the unique multi-domain architecture of the USP7, as opposed to targeting the highly-conserved structure of the catalytic domain, could improve specificity and reduce toxicity.

In total, >30 derivatives were synthesized for defining the SAR between the [2,3-c] pyrazole scaffold and $USP7_{FL}$ (38). Changes to the substituents on the phenyl ring (R_1 & R_2; Fig. 2.1) had the largest impact on the IC_{50} and maximal inhibition (38). Dr. Hjortland tested a hypothesized interaction between TRAF and HUBL1-3 by evaluating the effect on catalysis under steady-state kinetics *in trans* with HUBL1-5 versus HUBL1-3 and TRAF-USP7CD (38). To more fully optimize our noncompetitive inhibitors structural evidence of the interactions would be needed to guide functional optimization of the inhibitor.

In summary, Fig. 2.8 below illustrates the domains necessary for interactions with USP7 inhibitors that have been identified to date. Most of the inhibitors have been shown to bind directly to the USP7 catalytic domain. The inhibitors most likely bind to an interface or pocket within the accessory domains, and thereby act as mixed, partial, non-competitive inhibitors.

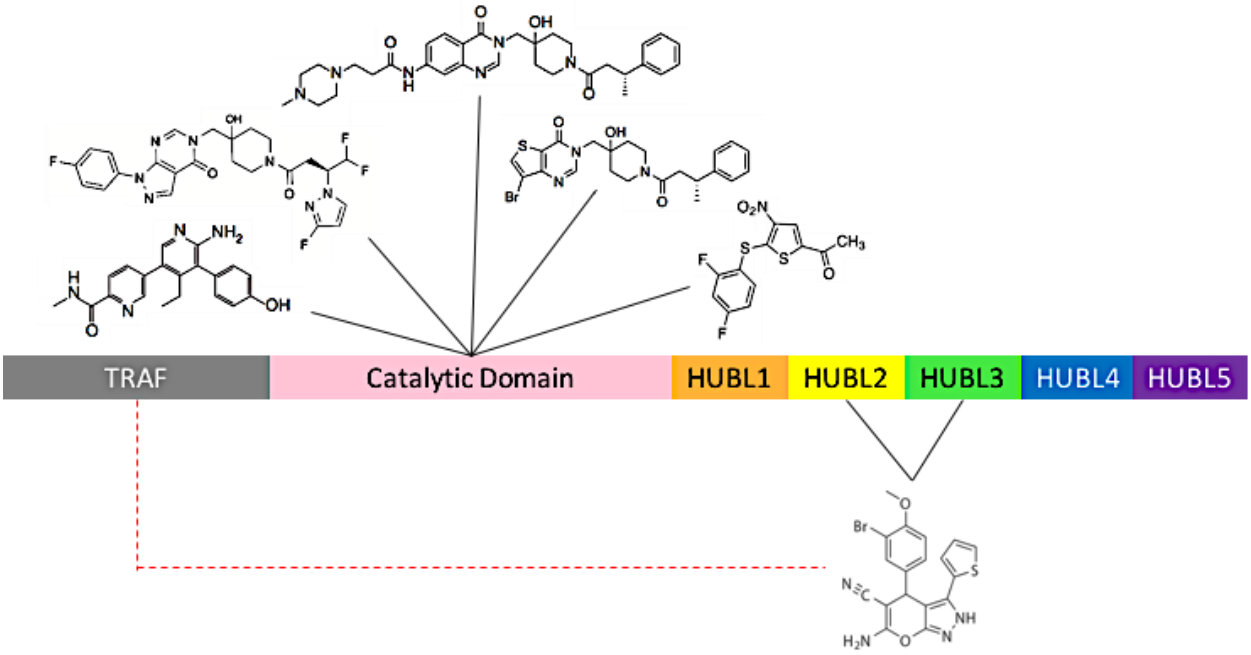


Figure 2.8. USP7 inhibitor map. (Top) Competitive inhibitors directed at the active site residues or an adjacent site in the catalytic domain that affects K_m (15-19). USP7 schematic is colored as before, grey = TRAF, pink = CD, orange = HUBL1, yellow = HUBL2, green = HUBL3, blue = HUBL4, purple = HUBL5. (Bottom) Mixed, partial, non-competitive inhibitor described herein; dashed red line refers to an observed importance for the TRAF domain in inhibition (see Fig. 2.3 A & B), but there is little evidence to suggest a direct interaction.

2.5 References

1. Fraile, J. M., Quesada, V., Rodriguez, D., Freije, J. M. P., and Lopez-Otin, C. (2012) Deubiquitinases in cancer: new functions and therapeutic options. *Oncogene*. **31**, 2373–2388.
2. Sowa, M. E., Bennet, E. J., Gygi, S. P., Harper, J. W. (2009) Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell*. 138(2): 389-403.
3. Eletr, A. M., Wilkinson, K. D. (2014) Regulation of Proteolysis by Human Deubiquitinating Enzymes. *Biochim. Biophys. Acta*. 1843(1): doi:10.1016/j.bbamcr.2013.06.027.
4. Young, M. J., Hsu, K. C., Lin, T. E. Chang, W. C., Hung, J. J. (2019) The role of ubiquitin-specific peptidases in cancer progression. *J. Biomed. Sci*. 26(42): https://doi.org/10.1186/s12929-019-0522-0.
5. Molland, K., Zhou, Q., and Mesecar, A. D. (2014) A 2.2 A resolution structure of the USP7 catalytic domain in a new space group elaborates upon structural rearrangements resulting from ubiquitin binding. *Acta Crystallogr. Sect. F*. **70**, 283–287.
6. Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C., and Mesecar, A. D. (2006) Severe acute respiratory syndrome coronavirus papain-like protease: Structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5717–5722.
7. Clasman, J. R., Báez-Santos, Y. M., Mettelman, R. C., O'Brien, A., Baker, S. C., and Mesecar, A. D. (2017) X-ray Structure and Enzymatic Activity Profile of a Core Papain-like Protease of MERS Coronavirus with utility for structure-based drug design. *Sci. Rep.* **7**, 40292.
8. Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J.-W., Gu, W., Cohen, R. E., and Shi, Y. (2002) Crystal Structure of a UBP-Family Deubiquitinating Enzyme in Isolation and in Complex with Ubiquitin Aldehyde. *Cell*. **111**, 1041–1054.
9. Faesen, A. C., Luna-Vargas, M. P. A., Geurink, P. P., Clerici, M., Merkx, R., van Dijk, W. J., Hameed, D. S., El Oualid, F., Ovaa, H., and Sixma, T. K. (2011) The Differential Modulation of USP Activity by Internal Regulatory Domains, Interactors and Eight Ubiquitin Chain Types. Chem. Biol. 18, 1550–1561.
10. Faesen, A. C., Luna-Vargas, M. P. A., and Sixma, T. K. (2012) The role of UBL domains in ubiquitin-specific proteases. *Biochem. Soc. Trans.* **40**, 539–545.

11. Kim, R. Q., van Dijk, W. J., and Sixma, T. K. (2016) Structure of USP7 catalytic domain and three Ubl-domains reveals a connector α-helix with regulatory role. *J. Struct. Biol.* **195**, 11–18.
12. Faesen, A. C., Dirac, A. M. G., Shanmugham, A., Ovaa, H., Perrakis, A., and Sixma, T. K. (2011) Mechanism of USP7/HAUSP Activation by Its C-Terminal Ubiquitin-like Domain and Allosteric Regulation by GMP-Synthetase. Mol. Cell. 44, 147–159.
13. Rougé, L., Bainbridge, T. W., Kwok, M., Tong, R., Di Lello, P., Wertz, I. E., Maurer, T., Ernst, J. A., and Murray, J. (2016) Molecular Understanding of USP7 Substrate Recognition and C-Terminal Activation. *Structure*. 24, 1335–1345.
14. Kim, R.Q., Geurink, P.P., Mulder, M.P.C., Fish, A., Ekkebus R., El Oualid, F., van Djik, W.J., van Dalen, D., Ovaa, H., van Igen, H., Sixma, T.K. (2019). Kinetic analysis of multistep USP7 mechanism shows critical role for target protein in activity. Nat. Commun. 10, doi: 10.1038/s41467-018-08231-5.
15. Kategaya, L., Di Lello, P.D., Rougé L., *et al*. (2017) USP7 small-molecular inhibitors interfere with ubiquitin binding. *Nature*. 550: 534-538. https://doi.org/10.1038/nature24006.
16. Turnbull, A. P., Ioannidis, S., Krajewski, W. W., *et al*. (2017) Molecular basis of USP7 inhibition by selective small-molecule inhibitors. *Nature*. 550: 481-486. https://doi.org/10.1038/nature24451.
17. Pozhidavaeva, A., Valles, G, Wang, F., Kanyo, J., Wright, D., Bezsonova, I. (2017) USP7-Specific Inhibitors Target and Modify the Enzyme's Active Site via Distinct Chemical Mechanisms. *Cell Chem. Biol*. 24: 1501-1512.
18. Gavory, G., O'Dowd, C. R., Helm, M. D., *et al*. (2017) Discovery and characterization of highly potent and selective allosteric USP7 inhibitors. *Nat. Chem. Biol.* 14: 118-125. https://doi.org/10.1038/nchembio.2528.
19. Lamberto, I., Liu, X., Seo, H. S., Schauer, N. J., Iacob, R. E., Hu W., Das, D., Mikhailova, T., Weisberg, E. L., Engen, J. R., Anderson, K. C., Chauhan, D., Dhe-Paganon, S., Buhrlage, A. J. (2017) Structure-Guided Development of a Potent and Selective Non-covalent Active-Site Inhibitor of USP7. *Cell Chem. Biol*. 24(12): P1490-1500.E11. DOI:https://doi.org/10.1016/j.chembiol.2017.09.003.

20. Schauer, N. J., Liu, X., Margin, R. S., *et al*. (2020) Selective USP7 inhibition elicits cancer cell killing through a p53-dependent mechansim. *Sci. Rep*. 10: 5324. https://doi.org/10.1038/s41598-020-62076-x.
21. Wertz, I. E., Murray, J. M. (2019) Structurally-defined deubiquitinase inhibitors provide opportunities to investigate disease mechanisms. *Drug Dicov. Today Technol*. 31. 109-123.
22. Hu, M., Gu, L., Li, M., Jeffrey, P. D., Gu, W., and Shi, Y. (2006) Structural Basis of Competitive Recognition of p53 and MDM2 by HAUSP/USP7: Implications for the Regulation of the p53–MDM2 Pathway. PLoS Biol. 4, e27.
23. Friesner, R.A., Murphy, R.B., Repasky, M.P., Frye, L.L., Greenwood, J.R., Halgren, T.A., Sanschagrin, P.C., Mainz, D.T. (2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* 49: 6177-6196.
24. J. L. (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem*. 47: 1750-1759.
25. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shaw, D. E., Shelley, M., Perry, J. K., Francis, P., Shenkin, P. S. (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* 47: 1739-1749.
26. Gao, K., Oerlemans, R., Groves, M. R. (2020) Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys Rev*. 12:85-104.
27. Pantoliano, M. W., Bone, R. F., Rhind, A. W., Salemme, F. R. (1997) Microplate thermal shift assay apparatus for ligand development and multi-variable protein chemistry optimization. U.S. Patent No. 6036920.
28. Dai, R., Geders, T. W., Liu, F. (2015) Fragment-based exploration of binding site flexibility if Mycobacterium Tuberculosis BioA. *J. Med. Chem*. 58:5208-5217.
29. Hung, A. W., Silvestre, H. L., Wen, S. (2009) Application of fragment growing and fragment linking to the discovery of inhibitors of mycobacterium tuberculosis panthothenate synthetase. *Agnew Chemie Int Ed.*
30. Larsson, E. A, Jansson, A., Ng, F. M. (2013) Fragment-baesd ligand design of novel potent inhibitors of tankyrases. *J. Med. Chem*. 56:4497-4508.

31. Kirley, T. L., Norman, A. B., Wetzel, H. N. (2020) A novel differential scanning fluorimetry analysis of a humanized anti-cocaine mAb and its ligand binding charactertics. *J. Immunol. Meth*. 476: 112676.

32. Niesen, F. H., Schultz, L., Jadhav, A., Bhatia, C., Guo, K., Maloney, D. J., Pilka, E. S., Wang, M., Oppermann, U., Heightman, T. D., Simeonov, A. (2010) High-Affinity Inhibitors of Human NAD^+-dependent 15-Hydroxyprostaglandin Dehydrogenase: Mechanism of Inhibition and Structure-Activity Relationships. *PLoS One*. 5(11): e13719.

33. Auld, D. S., Lovell, S., Thorne, N., Lea, W. A., Maloney, D. J., Shen, M., Rai, G., Battaile, K. P., Thomas, C. J., Simeonov, A., Hanzlik, R. P., Inglese, J., (2010) Molecular basis for the high-affinity binding and stabilizatoin of firefly luciferase by PTC124. *Proc. Natl. Acad. Sci. USA.* 107(11): 4878-4883.

34. Lea, W. A., Simeonov, A., (2012) Differential Scanning Fluorimetry Signatures as Indicators of Enzyme Inhibitor Mode of Action: Case Study of Glutathione S-Transferase. *PLoS One.* 7(4): e36219.

35. Chauhan, D., Tian, Z., Nicholson, B., Kumar, K. G. S., Zhou, B., Carrasco, R., McDermott, J. L., Leach, C. A., Fulcinniti, M., Kodrasov, M. P., Weinstock, J., Kingsbury, W. D., Hideshima, T., Shah, P. K., Minvielle, S., Altun, M., Kessler, B. M., Orlowski, R., Richardson, P., Munshi, N., and Anderson, K. C. (2012) A Small Molecule Inhibitor of Ubiquitin-Specific Protease-7 Induces Apoptosis in Multiple Myeloma Cells and Overcomes Bortezomib Resistance. *Cancer Cell*. **22**, 345–358.

36. Colland, F. (2010) The therapeutic potential of deubiquitinating enzyme inhibitors. *Biochem. Soc. Trans.* **38**, 137–43.

37. Colland, F., Formstecher, E., Jacq, X., Reverdy, C., Planquette, C., Conrath, S., Trouplin, V., Bianchi, J., Aushev, V. N., Camonis, J., Calabrese, A., Borg-Capra, C., Sippl, W., Collura, V., Boissy, G., Rain, J.-C., Guedat, P., Delansorne, R., and Daviet, L. (2009) Small-molecule inhibitor of USP7/HAUSP ubiquitin protease stabilizes and activates p53 in cells. *Mol. Cancer Ther.* **8**, 2286–2295.

38. Hortland, N. (2016) Defining the regulatory determinants in substrate catalysis by biochemical, biophysical, and kinetic studies for the development of specific small-molecule inhibitors of ubiquitin specific proteases 7 and 17. *Purdue University Open Acess Dissertations*. 1461. https://docs.lib.purdue.edu/open_access_dissertations/1461.

39. Chari, A., Haselbach, D., Kirves, J. M., Ohmer, J., Paknia, E., Fischer N., Ganichkin O., Möller, V., Frye, J. J., Petzold, G., Jarvis, M., Tietzel, M., Grimm, C., Peters, J. M., Schulman, B. A., Tittmann, K., Markl, J., Fischer, U., Stark H. (2015) ProteoPlex – stability optimization of macromolecular complexes by spare-matrix screening of chemical space. *Nat. Meth.* 12(9): 859-865.

CHAPTER 3. CRYO-EM SINGLE PARTICLE ANALYSIS OF KEAP1-CULLIN3-RBX1 E3 LIGASE

3.1 Introduction

Ubiquitination is a post-translational modification that functions to control protein levels and homeostasis in the cell. In humans, there are two major categories of ubiquitinating machineries: SCF E3 ligases, and RING E3 ligases (1). The prototypical RING E3 ligase functions as a multi-subunit enzymatic complex that sequesters free mono-ubiquitin and ligates it onto target proteins. The overall mechanism of an E3 ligase is denoted by three major catalytic events carried out in three major classes of enzymes: E1, E2, and E3 (Fig. 3.1). Thus far, it has been determined that the human genome encodes for 2 E1 enzymes, ~35 E2 enzymes, which can mix-and-match among other components to compose > 700 distinct E3 ligases (1).

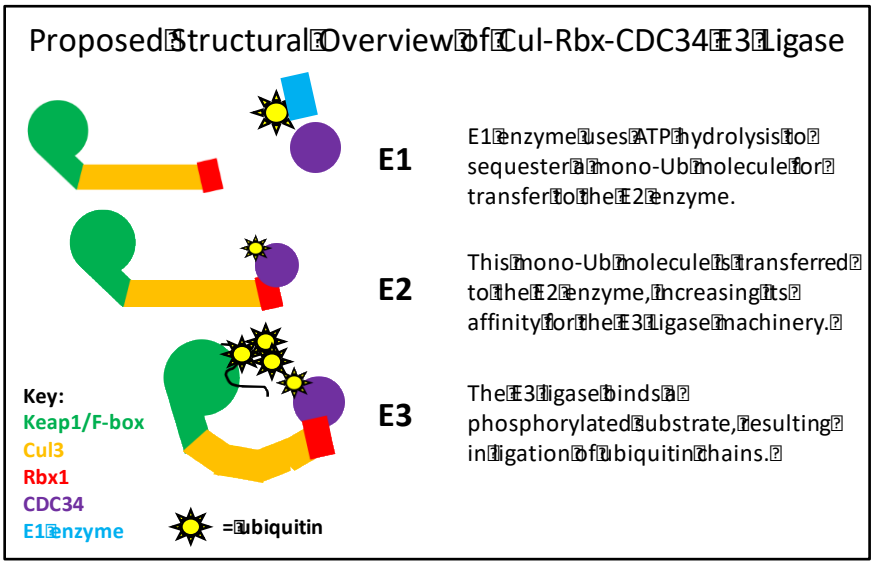


Figure 3.1. Proposed structural overview of Keap1-Cul3-Rbx1-E2 monomeric complex. Keap1 forms a dimer, but it is not entirely understood if the interaction with Cul3 would block dimerization. For clarity only a monomeric structure is illustrated. CDC34 is an E2 enzyme that would bind the Rbx1 protein.

The canonical mechanism of an E3 ligase follows a cascade of thioester bond formation between the c-terminal aldehyde of ubiquitin and catalytic cysteines of E1/E2/E3 enzymes, resulting in ligation onto an amino group of a targeted substrate. First, an E1 enzyme is *activated* by a ubiquitin molecule onto the catalytic cysteine, requiring an ATP (1-2). This thioester-linked

ubiquitin is transferred to the E2 enzyme via a second thiol-esterification reaction, the precise details of which are not fully understood (1-2). Lastly, the E2-Ub conjugate transfers ubiquitin onto the target protein via an E3 enzyme in a two-step mechanism. The E2 active site cysteine transfers to the E3 to form a third thioester with ubiquitin c-terminal aldehyde, followed by ligation onto the target substrate (2). This ligation can be repeated to form ubiquitin chains of varying length and branching (1-2). The details of the thiol-esterification and ligation within an E3 ligase and how these complexes accommodate growing, branching patterns of ubiquitin remains poorly understood. This chapter will describe experimentation performed to generate, purify and optimize cryo-EM sample preparation for a RING E3 ligase with components described in Table 5.

Table 3.1. Keap1-Cullin3-Rbx1 E3 ligase cellular components for study in this chapter.

Keap1-Cul3-Rbx1 E3 Ligase Complex		
Protein	**Residues (Mw)**	**Function within E3 Ligase**
Cul3	768 (88.9 kDa)	Scaffold protein bridging Keap-Rbx1
Keap1	624 (69.7 kDa)	Substrate-binding protein
Rbx1	108 (12.2 kDa)	E2-binding protein (E1 enzyme)

Cullin-based RING E3 ligases (CRLs) feature a cullin-rbx1 protein backbone that acts as a scaffold between the E2 enzyme and the substrate-binding protein (Keap1). Keap1 functions to recognize substrates to be ubiquitinated through a degron consensus motif (7). To ligate ubiquitin in growing chains, it is hypothesized that a large degree of conformational flexibility in the cullin-rbx1 backbone, and oligomerization of Keap1 are proposed to occur (Fig. 3.1) (5, 7-8).

Several studies have shown Keap1 interaction with Nrf2, a potent regulator of the antioxidant response (AR), is thought to be the predominant mechanism for dealing with oxidative stress (3-4). Under quiescent conditions, Nrf2 concentration in the cytoplasm is maintained via sequestration by Keap1, which may facilitate the ubiquitination and proteasomal-mediated proteolysis of Nrf2 (3-4). Keap1 binding and degradation of Nrf2 in the cytosol is the primary mechanism for repressive effects of Keap1 on Nrf2-mediated ARE expression (3-5). Keap1 thus functions not only as a tumor suppressor but may also function as a metastasis suppressor via its interaction with Nrf2 (1). Since Nrf2 activation leads to a powerful antioxidant response, and Keap1 suppresses Nrf2 activation, this has made Keap1 an attractive drug target.

In the cell, Keap1 functions as a sensor for oxidants, electrophiles, and pro-inflammatory agents in the cytosol (3-5, 7-8). Deficiencies in Keap1-based protective mechanisms are associated with increased disease risk (8). Perhaps the most well-studied functions of Keap1 is its ability to shunt into the nucleus to bind antioxidant response elements (AREs) located in the promoter region of genes related to antioxidant stress-response (4). Keap1 performs this function by binding a transcription factor, Nrf2 (4). Luciferase reporter assays for ARE expression are a common activity assay to probe Keap1 functionality in the cell for testing complex formation with Cul3-Rbx1 E3 ligase (5, 7-11).

It is thought that under normal, reduced conditions in the cytosol, Keap1 can bind Nrf2 and associate with the Cul3 CRL, resulting in ubiquitin-mediated degradation and downregulation of Nrf2-signaling pathways (6-8). During oxidative stress in the cytosol, select Keap1 cysteine residues become oxidized, abrogating any interaction with the Cul3 CRL, preventing degradation of Nrf2 (6-8). Then, it is hypothesized that after exposure to electrophilic ARE inducers, Keap1-Nrf2 can translocate to the nucleus and alter gene expression in response to oxidation (9-11). Accordingly, the Keap1-Nrf2 axis of signaling remains an attractive therapeutic target for controlling oxidative stress and associated inflammation, but an understanding of the structure-function of the Keap1-CRL and Keap1-Nrf2 complexes are lacking.

Keap1 contains 27 cysteine residues which are proposed sensors of electrophilic ARE inducers, and leaves Keap1 sensitive to covalent modification by electrophiles (7-11). Currently, it is hypothesized that modification of highly reactive cysteines in Keap1, particularly Cys151, may decrease the binding of Keap1 with Cul3 (5, 9-10). Although, a crystallographic structure deposited on the PDB in 2017 (accession code: 5NLB) shows Cul3 bound to the Keap1 BTB domain with C151 outside the interaction interface, there is no biochemical data associated with this structure. Moreover, data from Eggler, *et al*., found that Cysteine mutation based on partial molar volume (PMV) was responsible for controlling Nrf2 levels, and potentially Keap1:Cul3 binding (11).

It is not fully understood which residues are critical to the interaction of Keap1 with Cul3, nor is it fully understood the binding kinetics involved in the interaction, and this information is important to understanding the regulation of the pathway. To this end, ITC was employed to determine a difference in binding between WT Keap1 and a C151S Keap1 mutant, and to determine the effect, if any, oxidizing WTKeap1 may have on complex formation for sample prep

considerations. Refeyn mass photometry, analytical size exclusion gel filtration assay, and cryo-EM single particle analysis was performed to identify the stoichiometry of the interaction between Keap1 and Cul3.

3.2 Methods & Materials

3.2.1 Purification of Keap1

Codon-optimized WT human Keap1 was expressed in BL21(DE3) pLys *E. coli* (Rosetta) with 30 μg/mL chloramphenicol and 100 μg/mL carbenicillin in LB. Cultures were incubated, shaken at 37°C until OD 600 nm = 0.6, cooled on ice for 20 min, supplemented with 0.5 mM IPTG final concentration, and then incubated, shaken at 18°C for 12 hrs. Cell pellet was harvested via centrifugation at 5,000 x g for 20 min. Cells were resuspended in 50 mM HEPES, pH 8.0, 500 mM NaCl, 1 mM TCEP, 25 μg/mL DNase (Sigma), 50 μg/mL Lysozyme (ThermoFisher), 1 tablet EDTA-free protease inhibitor cocktail (Roche) at 15 mL per gram cell pellet. Cells were lysed via sonication (Branson sonifier) at 65% amplitude, 6 sec pulse on, 6 sec pulse off, 1 min per gram cell pellet. Lysate was clarified by centrifugation at 17,000 x g for 20 min, and then 0.45 μm filtered. Clarified lysate was applied via FPLC to a 5 mL NiNTA HiTrap (GE Healthcare) column. The column was rinsed with 50 mM HEPES pH 8.0, 500 mM NaCl, 1 mM TCEP until baseline UV signal, at which point the column was rinsed with 10% isocratic gradient with 50 mM HEPES pH 8.0, 500 mM NaCl, 1 mM TCEP, 450 mM imidazole. Keap1 was eluted with a 10-100% gradient, eluting at ~120-150 mM imidazole. Keap1 was then pooled and dialyzed in 50 mM HEPES pH 8.0, 250 mM NaCl, 1 mM TCEP for 12 hrs, concentrated to 10 μM, centrifuged at 17,000 x g for 5 min to remove any particulates, and then used for ITC and gel filtration assays. All steps were carried out at 4°C.

3.2.2 Purification of Cul3-Rbx1

Codon-optimized WT human Cul3-Rbx1 were expressed in BL21(DE3) *E. coli* (Rosetta) via a pETDuet expression vector with 100 μg/mL carbenicillin in LB. Cultures were incubated, expression was induced, and the purification protocol was identical to as described above for Keap1.

3.2.3 Isothermal Titration Calorimetry

Samples for ITC were used in a Malvern PEAQ-ITC (Malvern Analytical) instrument. Proteins were purified and dialyzed in the same buffer: 50 mM HEPES pH 8.0, 250 mM NaCl, 1 mM TCEP. The ITC cell was 200 μL of Keap1 at 20 μM concentration, with 2 μL injections of 500 μM Cul3-Rbx1. Injections were every 120 s over for a total of 19 injections. Protein concentrations were determined using the Beer-Lambert law with the absorbance (A_{280nm}) read in a 0.05 cm cuvette (BioTek) and the average ExPASy ProtParam tool extinction coefficients for the primary sequences of Keap1 and Cul3 (12). The raw data was analyzed using the Malvern PEAQ-ITC software to obtain the enthalpy, stoichiometry of binding, and the entropy parameters.

3.2.4 Gel Filtration Interaction Assay

Keap1 and Cul3-Rbx1 reactions were injected individually or mixed in 1:1 molar stoichiometry (Keap1 homodimer to Cul3 monomer) and all reactions were incubated at ambient temperature for 1 hr. 250 μl volume of 28.6 μM Keap1 and 40 μM Cul3-Rbx1 final injections of ~1 mg total mass for each was injected to measure retention volumes of individual proteins. 500 uL of 1:1 mixture of Keap1 homodimer (21 μM Keap1 homodimer) and Cul3-Rbx1 (10.5 μM Cul3-Rbx1 monomer) was injected at a total mass amount of ~ 2 mg. All injections were via FPLC over a Sephacryl 16/60 S200 size exclusion column (GE Healthcare) in 50 mM HEPES, pH 8.0, 500 mM NaCl, 10% (v/v) glycerol, 1 mM TCEP. Flow rates were kept at 0.1-0.2 mL/min for all injections. 250 μl fraction volumes were collected to ensure high-purity samples from cryo-EM.

High-molecular weight standards kit (GE Healthcare) was used as standards for determining the void volume and the elution rate for molecular weight standards. All standards were resuspended in the same buffer at protein concentrations per manufacturer's protocol and incubated at the same temperature for the same period of time and were injected in injections per manufacturer's protocol. The standard curve was made using Equation 3.1 below:

$$K_{average}(K_{av}) = \frac{(V_e - V_o)}{(V_t - V_o)} \qquad (3.1)$$

The K_{av} is the average distribution coefficient for each protein, wherein V_e is the sample elution volume, V_o is the void volume determined by blue dextran, and V_t is the total resin bed volume for the Sephacryl 16/60 S200 size exclusion column (GE Healthcare). Once the K_{av} is

determined experimentally from elution volume, the molecular weight was calculated by using the following linear regression from the standards shown as Equation 3.2 below:

$$Experimental\ K_{average}(K_{av}) = -0.2883(M_w) + 1.770 \quad (3.2)$$

3.2.5 Preparation of Keap1-Cul3-Rbx1 complex for Cryo-TEM

Bare gold Quantifoil R 0.6/1.0, 400 mesh grids (EMS) were washed in chloroform, and then rinsed in buffer prior to sample application. Empty hole grids without substrate application were used for applying 5 µL of purified Keap1-Cul3-Rbx1 at 1.5 mg/mL of 0.5 mg/mL total protein concentration from the gel filtration assay. Grids were incubated with sample for 1 min and plunge-frozen in liquid nitrogen-cooled liquid ethane. Blotting and plunge-freezing was performed on a Gatan CryoPlunge3 at 4 sec and 6 sec blotting time and 95% humidity.

3.2.6 Refeyn mass photometry of Keap1-Cul3-Rbx1 complex

Samples from analytical size exclusion chromatography elution were collected, analyzed by SDS-PAGE for purity, and were diluted to 500 nM sample concentration. Samples were then diluted 1:10 by addition of 2 µL sample into 18 µL buffer for a final concentration of 50 nM, which corresponds to ~140 ng Keap1, 100 ng Cul3-Rbx, or approximately 240 ng of Keap1:Cul3 complex. Final concentration samples were imaged and analyzed by Refeyn™ mass photometry (13).

3.2.7 Cryo-TEM Data Acquisition

Five microliters of the purified Keap1-Cul3-Rbx1 and Keap1-Cul3-Rbx1-Nrf was applied to a Quantifoil R0.6/1 400 mesh grid (Electron Microscopy Sciences) without glow discharging at a range of concentrations from 0.5 – 1.5 mg/mL. Grids were used as bare gold, without substrate applied, to optimize thin ice over empty holes. The grids were plunge-frozen in liquid ethane with a CP3 Vitrobot (Gatan) under these conditions: temperature, 25°C; humidity > 90%; blotting time, 4 – 7 s. Frozen grids were imaged in a FEI Titan Krios (300 kV, Thermo Fisher Scientific) equipped with a Gatan Quantum-LS energy filter (20-eV zero-loss filtering) connected to a Gatan K3 Summit direct electron detector operating in super-resolution counting mode. Data acquisition

was performed as movies of 40 frames over 1300 ms exposure acquired at a nominal magnification of 165,000x. A total dose of 59.7 e^{-} Å^{-2} and a pixel size of 0.255 Å (SR mode) was used during data collection. The acquired movies were processed during the imaging session with motion correction and dose weighting by MotionCor2 (14) and contrast transfer function (CTF) estimation by CTFFIND4 (15). A total of 985 aligned movies for both data sets (with Nrf2 and without Nrf2) were collected. Images were analyzed by cryoSPARCv2 and CTFFIND4 programs to eliminate micrographs with resolutions > 7 Å and micrographs with defocus > ± 3 µm (15-16).

Table 3.2. Cryo-EM data acquisition and refinement parameters for Keap1-Cul3-Rbx1 and Keap1-Cul3-Rbx1-Nrf2.

Instrumentation	**Sample**
Titan Krios, K2 DED, VPP, Energy Filter	**Keap1-Cul3-Rbx1**
Magnification (x)	165,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	59.7
Defocus range (µm)	± 0.5 to ± 2.5
Pixel size (Å)	0.255 (SR); 0.51 2x2 bin
Exposure frames (#; time, ms)	40; 1300
Total aligned movies (#)	985
Movies used for Image Processing	222

3.2.8 Cryo-TEM Image Processing

Particles selection was performed with cryoSPARCv2 (16). During initial processing, 1500 particles were reference-free manually selected from 60 images with cryoSPARCv2 (16). The selected particles were used for generating templates for 'template picker' in cryoSPARCv2 using reference-free 2D classification in cryoSPARCv2 (16). Five 2D classes from this particle set were low-pass–filtered to 20 Å and used as templates for cryoSPARCv2 'template picker' particle selection with the full dataset. The larger particle set was reference-free 2D classified in cryoSPARCv2, resulting in 20 averages appearing to cover a broader range of views. These were selected as templates for a second round of particle selection with cryoSPARCv2. The final particle selection resulted in a total of 232,143 particles, which were then used for several successive rounds of 2D classification, resulting in the averages shown in Fig. 3.X representing 5,982 particles.

3.3 Results

3.3.1 ITC measurements of Keap1:Cul3 interaction

It has been hypothesized that oxidation of cysteines in the BTB domain of Keap1 is critical to the interaction with Cul3 (5, 7, 9-11). To test this hypothesis and to understand how the presence of reducing agent affects generating the Keap1:Cul3-Rbx1 complex for cryo-EM, ITC experiments were performed with either wild-type Keap1 under strong reducing conditions (1 mM TCEP; -0.33 V at pH 7), in the absence of reducing agent (Ox), or as a C151S Keap1 mutant.

ITC data (Fig. 3.2) suggests that Keap1 in the absence of reducing agent (green curve, top right) was unable to form a complex with Cul3; although it is important no experimentation was done to analyze the oxidation state of the Keap1 protein in this experiment. This finding corroborates previous studies which that found WTKeap1 when treated with oxidative reagents such as nitric oxide, Zn^{2+}, alkenals (9), or sulforaphane and isothiocyanate (10), that interaction between Keap1 and Cul3 was abrogated as evidenced by cellular ARE (antioxidant response element) luciferase reporter assays.

Interestingly, when WTKeap1 was purified in the presence of 1 mM TCEP as a reducing agent (Rd-Keap1; red curve), binding with Cul3-Rbx1 was observed (Fig. 3.2). When the WTKeap1 purified in the absence of reducing agent was resuspended in 1 mM TCEP (final sample concentration) and used for subsequent ITC runs, binding like WTKeap1 (Rd-Keap1; red curve) was observed (not shown).

Furthermore, the C151S Keap1 mutant was still able to form a complex with Cul3 in the presence of reducing agent, albeit at 3-fold diminished affinity (113 μM vs 407 μM) (Fig. 3.2). This binding behavior of the C151S Keap1 mutant corroborates findings by Eggler, *et al*. (11) with C151S Keap1 cellular studies. The C151S Keap1 mutant resulted in levels of ARE activation in the cell comparable to WTKeap1 (11). Interestingly, it was found that when Cys151 was mutated to Trp151 to mimic electrophile modification of Cys151, the level of ARE (antioxidant response element) activation in cells by Nrf2 was comparable to WTKeap1 treated with sulforaphane. These findings suggest that loss of Cys151 may not abrogate interaction with Cul3, but intact Cys151 may allow for perhaps more global changes in Keap1 that affect Cul3 binding through modification by larger electrophilic molecules such as sulforaphane and CDDO (11, 14, 15).

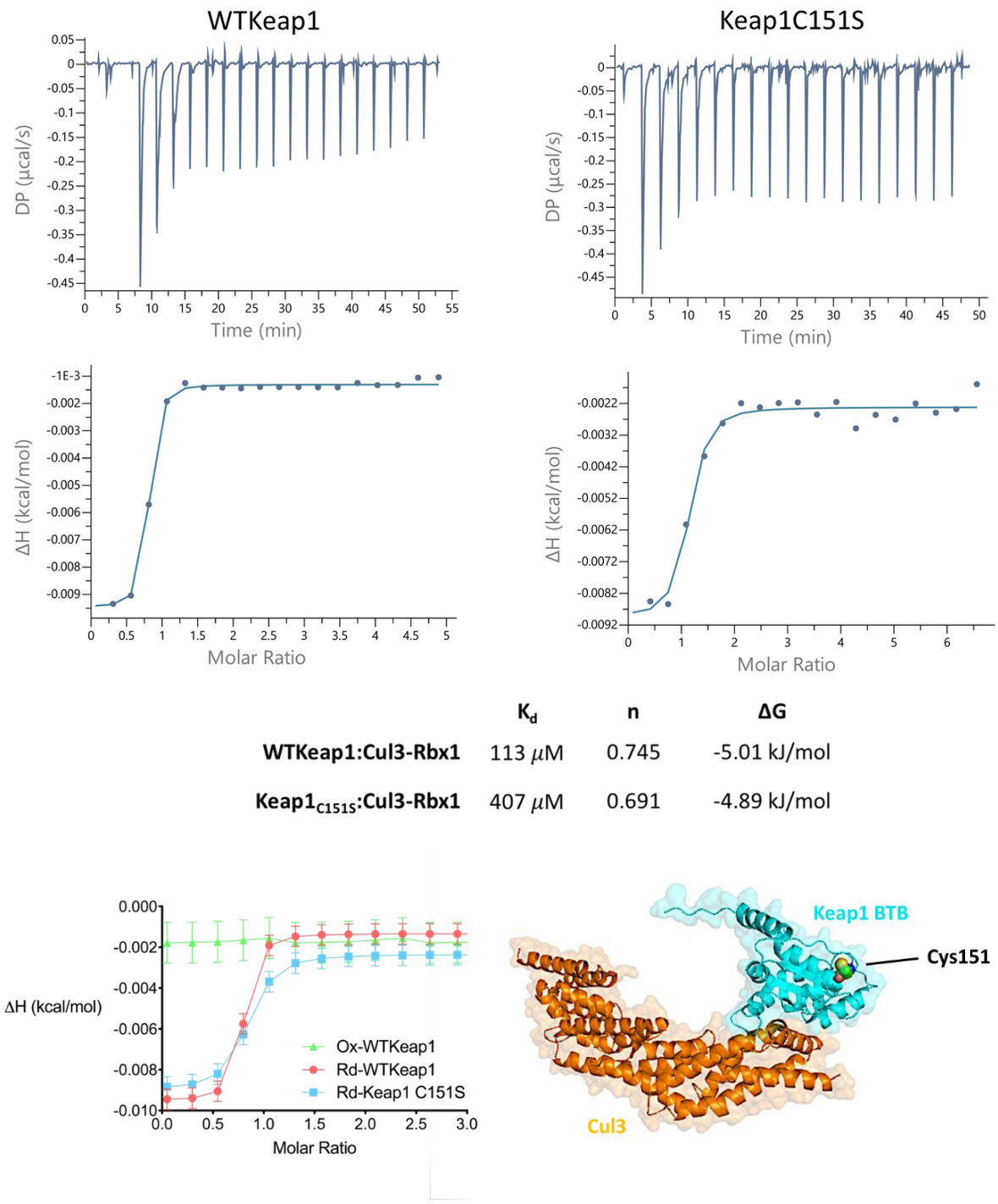


Figure 3.2. Cys151 of Keap1 is not critical to Cul3 interaction. ITC was performed with Keap1 at 20 µM concentration in the cell, with Cul3-Rbx1 at 500 µM in the syringe. It was determined that in the absence of reducing agent (Ox-WTKeap1) binding heats from an interaction was not detectable. Although, Keap1 in 1 mM TCEP (Rd-WTKeap1) had measurable binding with Cul3-Rbx1 at identical concentrations. Mutation of C151S led to increased K_D but did not prevent the interaction. Bottom right is a crystallographic structure (PDB: 5NLB) with cysteine 151 highlights in the BTB domain (cyan) but is not currently published.

The affinity (K_D) of Keap1 for Cul3-Rbx1 was calculated from fitting the enthalpy as a function of molar ratio, and was estimated at low affinities 1.13 x 10^{-4} M for WTKeap1 and 4.07 x 10^{-4} M for Cys151SKeap1 (Fig. 3.2), suggesting low affinities and weak binding interactions. Although there are no published K_D values for Keap1:Cul3-Rbx1 complex formation, these ITC-derived affinity values would suggest high micromolar protein concentrations for appreciable complex formation, which is not observed in the gel filtration assays (Fig. 3.5 below). There are, however, published ITC-derived K_D values for a closely related Keap1 protein, KLHL11, with Cul3 that showed K_D in the range of 20 nM – 650 nM, representing affinities that are 100-fold higher than Keap1:Cul3. However, it is important to note the KLHL11:Cul3 experiments were performed at 15℃ (versus 25℃ for Keap1:Cul3) and showed endothermic reactions as opposed to exothermic reactions for Keap1:Cul3. The researchers concluded the CUl3:KLHL11 interactions were hydrophobically-driven due to the observed largely entropically-driven interactions according to data analysis from observed heat of binding at differing temperatures (17).

Interestingly, the binding stoichiometry analysis from ITC binding curves for both WTKeap1 (n = 0.745) and the C151S Keap1 mutant (n = 0.691) are suggest somewhere between a 2:1 to 1:1 Keap1:Cul3-Rbx1 binding stoichiometry. Currently the two competing hypotheses regarding complex stoichiometry are a i) 2:2 Keap1 homodimer to 2 Cul3-Rbx1 molecules, which would be represented by an n of ~1 molar ratio, or ii) a 2:1 Keap1 homodimer to a single Cul3-Rbx1 molecule, represented by an n of ~0.5. It could be that the true stoichiometry is 2:2, but a falsely lower molar ratio may be indicative of a lower 'active' population, in that not all protein injected in the cell was able to react due to oxidative damage and/or loss of binding activity.

The Gibbs free energy parameters obtained from the best fit analysis of the ITC data in Fig. 3.3 would suggest that this interaction is primarily entropically-driven, with almost no enthalpy contribution. This finding would mean the interaction appears to have no contribution from enthalpy change (ΔH), but rather the free energy (ΔG) is due to changes in entropy (ΔS) (Fig. 2.6). The crystallographic structure depicted in Fig. 3.4 suggests that the interaction may be driven primarily by hydrophobic interactions, which may support the hypothesis of an entropically-driven interaction due to an entropy increase of the water molecules upon desolvation of the residue side chains (18). To elucidate the nature of the reaction more fully and to study the potential of an

entropically-driven hydrophobic effect, ITC performed at different temperatures may be suggested, as was done by Canning, *et al*. with KLHL11:Cul3 (17).

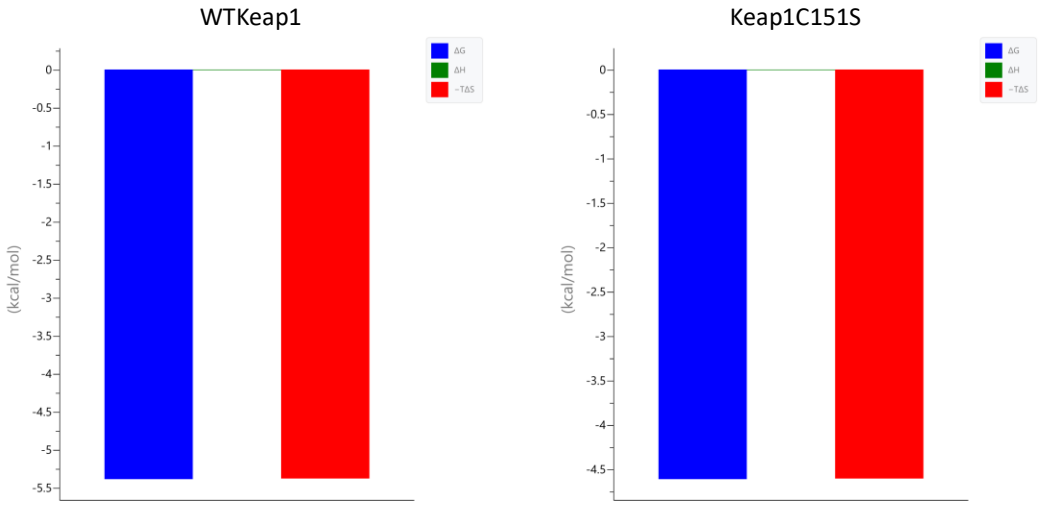


Figure 3.3. Keap1:Cul3 interaction is entropically-driven. According to best fit curves of ITC heats, the Keap1:Cul3 interaction appears to be entropically-driven for WTKeap1 and C151SKeap1. Values were extrapolated from best fits of raw heat measurements in Fig. 3.6 using Malvern PEAQ-ITC analysis software.

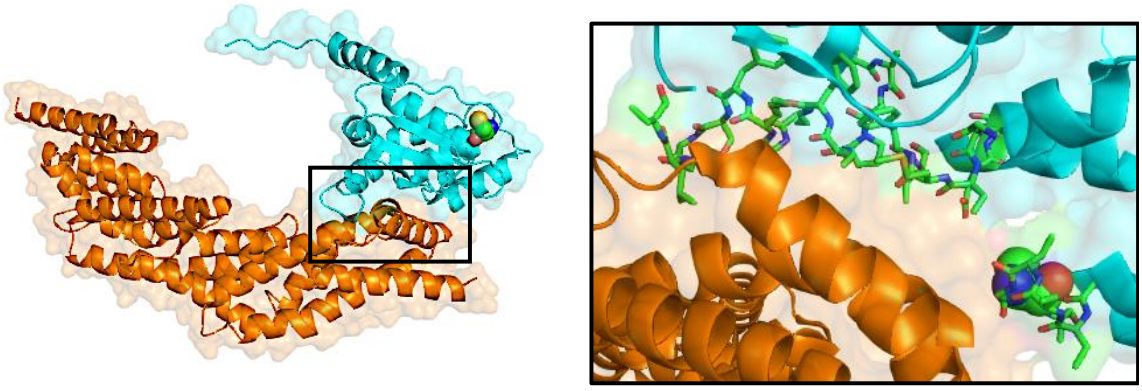


Figure 3.4. Residue critical to Keap1:Cul3 interaction. (Left) Crystallographic structure (PDB: 5NLB) with Keap1 BTB domain in cyan, and Cul3 in orange. (Right) Close-up of inlet showing BTB residues in green thought to be critical to interaction with Cul3, Cys196 is shown in sphere at bottom right.

According to a recently deposited crystallographic structure of the Keap1 BTB domain in complex with Cul3 (Fig. 3.8, PDB: 5NLB), it appears Cys196 is located at the interaction interface,

and not Cys151 (not yet published). In fact, structural analysis of the electron density shows all Keap1 side chains facing Cul3 appear to be a combination of aliphatic hydrophobic residues Pro104, Val105, Ala108, Met109, Gly113, Leu114, Gly157, Ala158, Val259, Met160, Ala190, Ile193, Gly194, Cys196, Val196, and aromatic hydrophobic residues Phe106, Phe110, and Tyr161 (PDB: 5NLB, 2017). Residues of the Cul3 interface, facing toward Keap1, include a combination of hydrophobic residues Phe54, Tyr58, Tyr62, Val55, Leu66, Ile122, Leu123, Met124, Tyr125, Val129 and polar residues Glu55, Arg59, Thr63, Arg128 (PDB: 5NLB, 2017). It is important to note that Cys196 is nearby Cul3 Thr63, likely indicating a hydrogen-bonding pair within the hydrophobic milieu. Oxidation of Cys196 is thus hypothesized to be a critical oxidation sensor of Keap1 that contributes to binding with Cul3, but the interaction appears to be entropically-driven likely due to the hydrophobic surfaces. However, given the 3.45 Å resolution and the lack of accompanying biochemical data with the deposited structure, it is not known the precise residues that are responsible for the interaction, although both sides of the interaction interface that would have displaced waters upon interaction are largely hydrophobic.

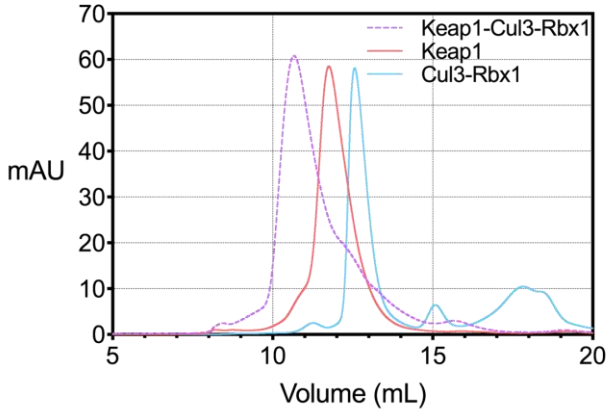
Crystallographic analysis of closely related Kelch-domain proteins such as KLHL11:CUl3, β-TrCP1:Cul3, and SOCS4:Cul3 interactions, as well as F-box intermediary protein Skp1:Cul3 interaction, all show largely hydrophobic-hydrophobic residue interactions between the BTB/BTB-like domains and Cul3 (17). The ITC data suggests that under oxidative conditions (lack of reducing agent) that Keap1 was unable to form an interaction with Cul3-Rbx1, but with TCEP added back in via dialysis the interaction was restored (not shown), indicating that some oxidation-sensitive residues might be critical to the interaction, and that oxidation at these residues is reversible. A hypothesis that may explain these findings is that oxidation of Keap1's 27 cysteines, such as Cys151, Cys195, and the like, by larger electrophilic modifications such as sulforaphane, CDDO, and isothiocyanate may sterically hinder, or otherwise dampen hydrophobic surface of Keap1 for binding Cul3.

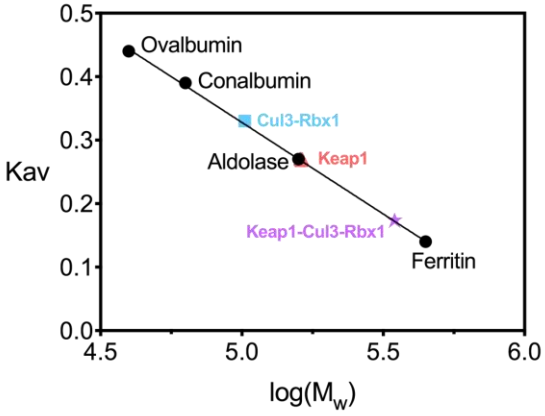
3.3.2 Gel filtration interaction assay suggests complex formation

Analytical size exclusion chromatography and analytical ultracentrifugation (AUC) experiments by Small, *et al*. Keap1:Cul3 suggest a 2:1 stoichiometry of a Keap1 homodimer to a single Cul3 molecule (10). While a later study by Iso, *et al*. showed AUC experimentation of the Keap1:Cul3 interaction that showed sedimentation that suggests a 2:2 stoichiometry of a Keap1 homodimer bound to a single Cul3 at each end (2 x Cullin3 molecules total) (19). Small-angle X-ray scattering (SAXS) modeling previously performed in our lab showed the possibility of a 1:1 Keap1:Cul3 or 2:1 Keap1 homodimer to Cul3 stoichiometry, with SAXS models showing complexes ~ 150 Å in length, but the SAXS data does not reliably differentiate between the two stoichiometries.

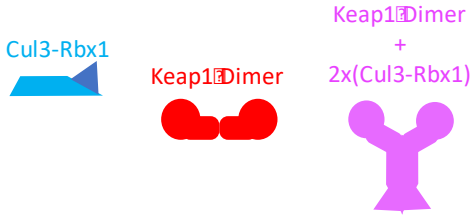
To investigate the stoichiometry of the complex and remove non-complex proteins prior to cryo-EM grid preparation, a gel filtration interaction assay and analytical size exclusion chromatography were performed. Fig. 3.5 shows that, when combined in a 2:1 WTKeap1:Cul3-Rbx1, or 1:1 Keap1 homodimer to Cul3-Rbx1, stoichiometry (purple curve), that the calculated molecular weight (M_w) from the standard curve corresponds to a molecular weight of ~348 kDa. This molecular weight suggests a 2:2 stoichiometry, wherein a Keap1 homodimer interacts with two Cul3-Rbx1 molecules and the shouldering of the purple curve may be attributed to excess Keap1 homodimer. This apparent molecular weight would support the Iso, *et al*. study.

Although, it is important to note that the complex may migrate larger than the actual molecular weight due to the apparent Stokes radii, wherein the hydrodynamic radius is larger than would be expected for the molecular weight due to the particular quaternary structure of the complex (20). The SAXS model built from data collected previously in our lab suggests a Keap1:Cul3-Rbx1 particle that would not adhere to strict spherical shape and would have a longer dimension ~ 15 nm. The data suggests each Keap1dimer may support two Cul3-Rbx1 molecules for a 1:1 molecular ratio, or a 200 kDa Cul3-Rbx1 to 142 kDa Keap1 mass ratio (1.4:1 mass ratio).





Sample	Estimated Mw	Calculated Mw	K_{av}
Cul3-Rbx1	100 kDa	103 kDa	0.33
Keap1	140 kDa	162 kDa	0.27
Keap1-Cul3-Rbx1	240 kDa	348 kDa	0.17



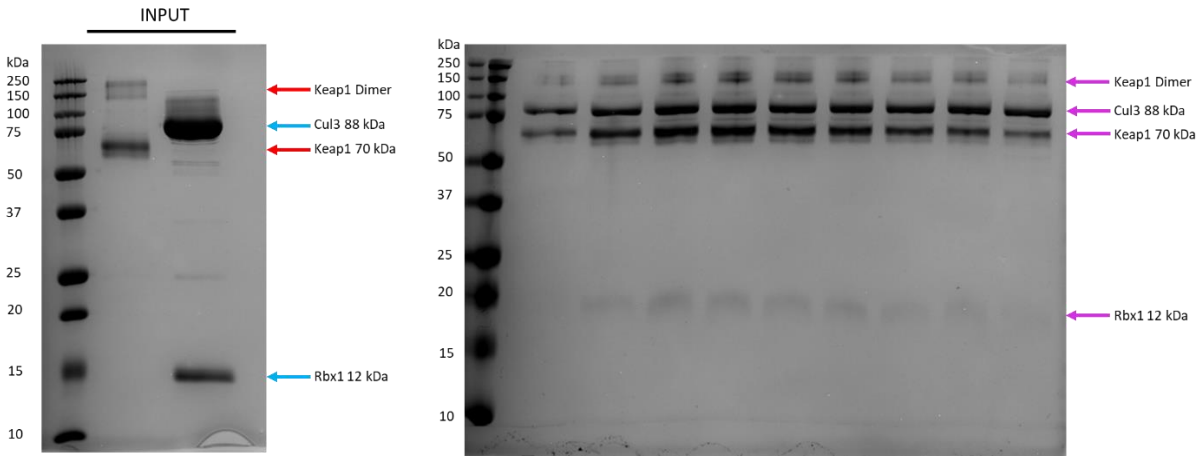


Figure 3.5. Gel filtration interaction assay and analytical size exclusion chromatography of Keap1-Cul3-Rbx1 interaction. (Top, Left) Elution profile of Superdex Sephacryl 16/20 24 size exclusion column (GE Healthcare). (Top, Right) Analytical size exclusion standards (black) and K_{av} of Cul3-Rbx1 (blue), Keap1 (red), and complex (purple). (Middle, Left) Table of estimated molecular weight (Mw) versus calculated molecular weights. Keap1 forms a stable dimer in solution; Cul3-Rbx1 are expected to be monomeric. (Middle, Right) Schematic illustration of anticipated stoichiometry from experimental data. (Bottom, Left) SDS-PAGE analysis of input samples of WTKeap1 and Cul3-Rbx1. (Bottom, Right) SDS-PAGE analysis of elution samples of Keap1-Cul3-Rbx1 injection; fractions were 0.2 mL; range of ~10 mL – 12.25 mL.

3.3.3 Mass photometry of Keap1-Cul3-Rbx1 shows complex not stable in nM range.

To better assay particle heterogeneity, complex stoichiometry, and measure particle integrity at concentrations used for cryo-EM grid preparation, the Refeyn™ mass photometry instrument was used. Mass photometry is a recently developed technology that can measure the hydrodynamic radius of particles from light scattering as particles settle onto a glass slide, imaged from underneath; the degree of scattering is proportional to molecular weight of macromolecules (13).

Samples eluted from analytical size exclusion summarized in Fig. 3.5 were diluted to ~ 500 nM concentrations and then diluted 1:10 into 20 µL sample volume for a final concentration of 50 nanomolar, which corresponds to ~140 ng Keap1, 100 ng Cul3-Rbx, and ~240-340 ng of Keap1:Cul3-Rbx1 complex. The protein masses are summarized below in the histogram plots of the particle distributions in Fig. 3.6, with the expected molecular weights shown underneath the sample name to the right. The particle homogeneity is calculated as the integration of the distribution of particles with a one standard deviation (1 sigma) statistical analysis applied to the normal distribution of particle size about the average molecular weight (13). The Keap1 sample was measured to be 91% homogeneous for particles of 142 kDa (101% of expected size). Cul3-Rbx1 was measured to be 87% homogenous for particles of 100 kDa (100% of expected size), and the complex was found to be composed of Keap1 and Cul3-Rbx1 particles that were not a part of a complex.

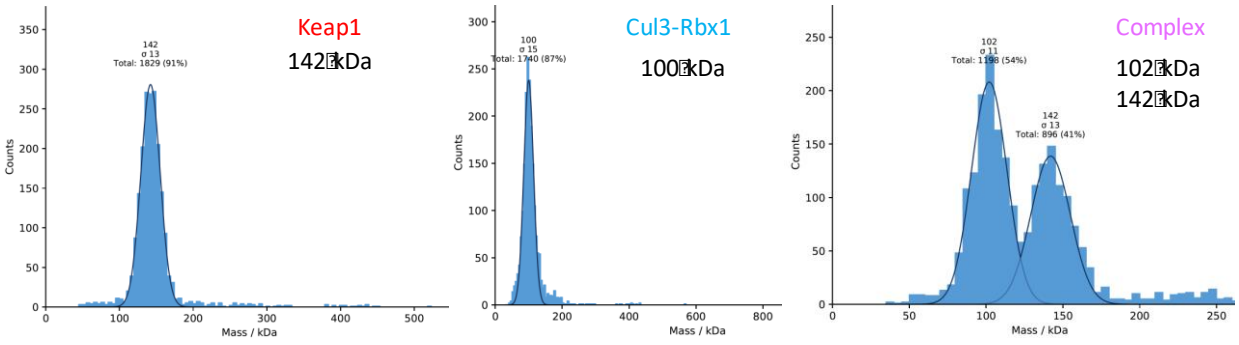


Figure 3.6. Refeyn™ instrument measurement of macromolecular homogeneity and particle size. (Left) WTKeap1 sample from Fig. 3.9 showing > 90% purity as a dimeric species of 142 kDa. (Middle) Cul3-Rbx1 sample from Fig. 3.9 showing 87% purity as a monomeric complex of 100 kDa. (Right) Keap1-Cul3-Rbx1 macromolecular complex from Fig. 3.9 as a nearly 1:1 ratio of Cul3-Rbx1 and Keap1. It is important to note that all samples are measured at < 1 µM final concentration, and this may have caused the complex to dissociate.

The Keap1:Cul3-Rbx1 complex sample was 95% composed of particles that were either Cul3-Rbx1, of Keap1. The sample represents a nearly 1.4:1 mass ratio based on counts between Cul3-Rb1 (54%) to Keap1 (41%), which matches well to the gel filtration interaction assay. 5% of total particles appeared between 200-250 kDa, although most particles were not observed in complex at nanomolar concentration. This data suggests that dilution to nanomolar concentrations, as was done for USP7 cryo-EM sample prep, is not conducive to maintaining the complex.

SAXS and AUC data generated by previous lab members led to the expectation that one Keap1 dimer binds one Cul3-Rbx1 molecule, resulting in an expected 2:1 stoichiometry. Although an intact Keap1-Cul3-Rbx1 complex could not be observed by mass photometry. The method cannot handle protein concentrations above nanomolar range without saturating the imaging of individual complexes. ITC experiments, while needing to be repeated, have estimated the K_D of the Keap1-Cul3 interaction to be in the micromolar range; and gel filtration assay reactions were performed at micromolar range, as well. Therefore, it is assumed the complex may have been intact and then dissociated through the serial dilution required to obtain sample concentration needed for mass photometry. Thus, for cryo-EM, the samples from Fig. 3.5 were kept at 5-10 µM, comparable to the concentration injected for the gel filtration experiment that resulted in complex formation.

3.3.4 Negative stain single particle analysis of Keap1 reveals dimeric structure at 14 Å

Negative stain images of Keap1 were found on the Mesecar network from 2013 (unpublished). The image analysis workflow involved importing raw micrographs into Relion3.0 for CTF estimation with CTFFIND4 followed by manual selection of 500 particles for auto-pick templates (15, 21). Auto-picking was performed in Relion3.0 with several rounds of 2D averaging providing reference-free 2D averaging results below in Fig. 3.7.

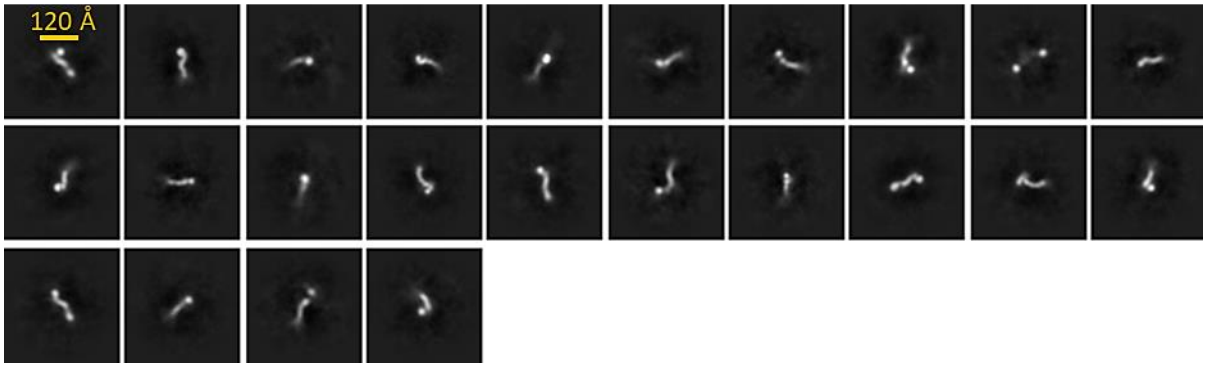


Figure 3.7. 2D Averaging illustrates conformational flexibility in the Keap1 dimer. Scale bar is 120 Å.

Reference-free 2D averaging performed in Relion3.0 from 10,000 particles from negative stain dataset. Each spherical shape at the distal end corresponds to a Kelch-like domain that is responsible for binding substrates.

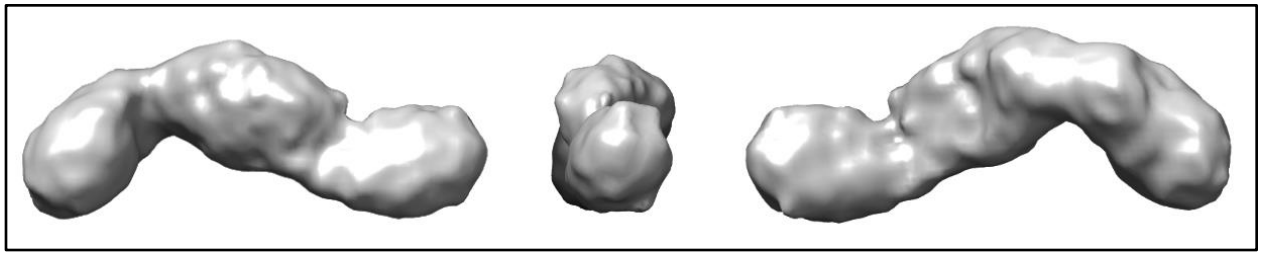


Figure 3.8. ~14 Å initial model of the Keap1 dimer.

The initial model of the WTKeap1 dimer was generated in cryoSPARCv2 from the subset of particles illustrated in Fig. 1. without symmetry restraints imposed (c1 symmetry). Compared to the published model (Fig. 3.9) from Dinkova-Kostova, *et al*. (2017), an idea of the locations of the BTB dimerization interface, oxidation-sensitive cysteines, and spherical Kelch domain relative to the conformational heterogeneity observed in the 2D averages can be observed.

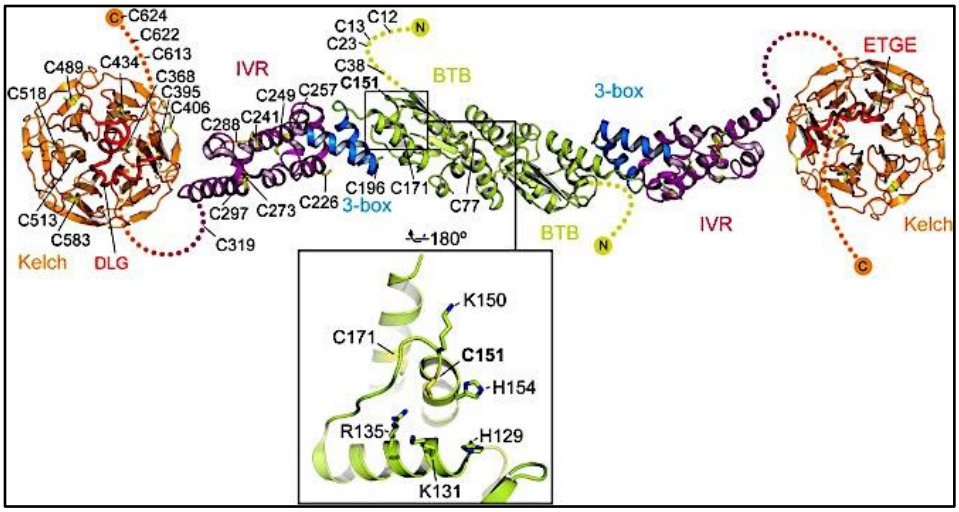


Figure 3.9. Keap1 dimer model from Dinkova-Kostova, *et al*. (2017). Reactive cysteines are indicated as potential oxidative sensors for function in the cytoplasm. C151 in the BTB domain is hypothesized to be a specific oxidation sensor critical to the interaction with Cul3. C196 Oxidation at this site is thought to abrogate the interaction with the E3 ligase machinery.

These negative stain results represent a promising lead for continuing TEM single particle work on the E3 ligase complex. Our lab had previously successfully purified Keap1 and Cul3/Rbx1 (6, 11), however a precise observation of the i) particle stoichiometry, ii) conformational heterogeneity, and iii) structural data to support our SAXS model is lacking. To analyze these aspects of the Keap1-Cul3-Rbx1 interaction and determine the overall quaternary structure, cryo-EM data collection and image analysis was performed.

3.3.5 Cryo-EM single particle analysis of Keap1-Cul3-Rbx1 shows particle heterogeneity

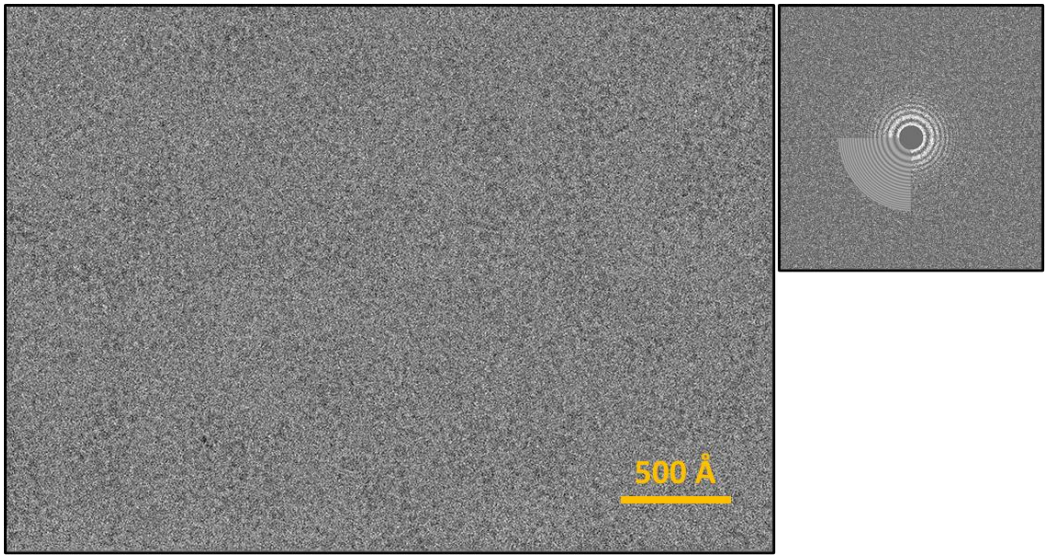


Figure 3.10. Keap1-Cul3-Rbx1 images acquired on Titan Krios at 165,000x nominal mag. (Top, Left) Representative image of Keap1-Cul3-Rbx1 on bare gold grids, scale bar is 500 Å. (Right) 2D power spectrum of the image to the left with CTF estimation done by CTFfind4.1 through cryoSPARCv2; estimated resolution was ~4.3 Å.

Figure 3.10 shows a representative micrograph of the Keap1-Cul3-Rbx1 dataset acquired as detailed in methods and materials section 3.2.8. Thin enough ice was achieved that the dataset was collected with the Volta Phase Plate (VPP). Image analysis was performed as detailed in methods and materials section 3.2.8 and resulted in a series of 2D averages that showed a significant degree of particle heterogeneity. Particles of various sizes ranging from ~200 Å - ~50 Å appeared in the earlier sets of 2D averages.

Overall, Fig. 3.11 shows 2D averages of size and feature like what was expected of the complete, intact Keap1-Cul3-Rbx1 particle. Box size variation experiments were performed based on dimensions obtained from SAXS data (~15 nm in longest dimension), analytical size exclusion gel filtration assays (~340 kDa apparent molecular weight), Keap1 homodimer negative stain initial model (~ 12 nm in longest dimension), and structural model of the Keap1:Cul3-Rbx1 E3 ligase from Canning, *et al*. (7, 17). Box sizes ranged from 1.5 – 2x of the longest anticipated dimension from the above experiments.

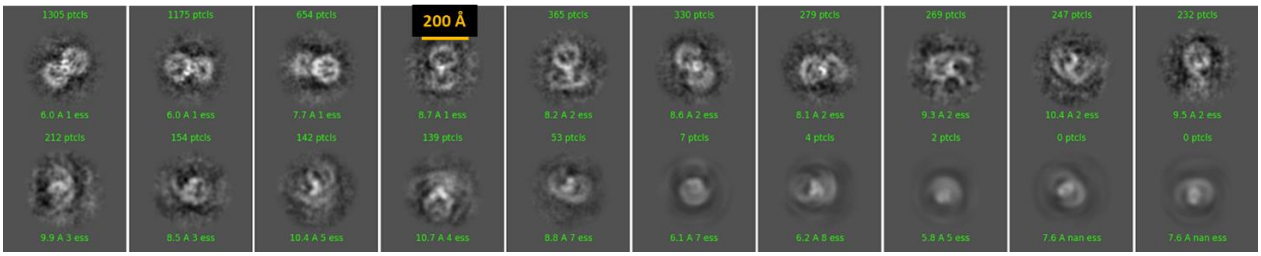


Figure 3.11. Keap1-Cul3-Rbx1 reference-free 2D averages. After 5 subsequent 2D averaging rounds, selecting for averages with particle size >100 Å in size, larger than Keap1 dimer initial model, 5,982 particles out of 232,143 particles (2.6%) were maintained throughout the averaging experiment that resemble the full complex based on the above experiments and previous findings in the literature.

Even among the 5,982 particles exhibited in the 2D averages below, there appears to be conformational differences. It is not clear if these averages represent the full complex, aggregation (multiple lobes), broken particles, or any other compositional heterogeneity issue. More sample optimization is necessary to more fully investigate the use of cryo-EM for studying the Keap1-Cul3-Rbx1 complex. Glutaraldehyde crosslinking experiments may be performed as a preliminary study to more accurately determine particle size and Keap1:Cul3-Rbx1 stoichiometry to build an initial model. In the event crosslinking, or otherwise fixing the complex is not performed, freezing trials should be performed to determine maximal concentrations for protein adsorption to avoid disassociation of particles, as very few larger complexes were observed in the particle set.

Initial modeling, with and without applying symmetry (cyclic c2 symmetry and dihedral d2 symmetry), in both Relion3.0 and cryoSPARCv2 of the particle stacks above that show the most feature and similarity resulted in featureless spheres of approximately the anticipated diameter of a full 2:2 and/or potentially a 2:1 Keap1:Cul3-Rbx1 dimer (16,21). It is still not clear which particle stoichiometry is correct from this data, and more particles would be necessary to improve views and initial model quality. Currently only ~2-3% of overall particles extracted from

micrographs appear to represent full complexes, and to proceed, strategies to better preserve the full complex should be tested.

3.4 Discussion

The integrity of the Keap1 oxidation sensors is critical to understanding the physiology and experimental considerations in purifying the Keap1-CUl3-Rbx1 E3 ligase complex. Currently there is no published K_D for the Keap1:Cul3-Rbx1 interaction by ITC which would inform appropriate concentration ranges for complex formation. Additionally, specific oxidation sensitive Keap1 residues been probed for their interaction with Cul3-Rbx1, but their contribution to the affinity have not been fully investigated. ITC experimentation was performed to determine the importance of reducing agent throughout the purification process and to determine if the Cys151 residue was critical to complex formation, as evidenced by findings in Eggler, *et al*. (11).

The C151SKeap1 mutant appeared to bind Cul3-Rbx1 at similar rates to wild-type, corroborating our lab's findings, and supporting the structure in PDB:5NLB of the Cul3-BTB domain, and runs contrary to other studies that would suggest Cys151 is the major oxidation sensor controlling E3 ligase binding, Nrf2 degradation, and ARE expression (5, 9-10). Although, no detectable binding was observed by ITC in the absence of reducing agent, the addition of reducing agent back into the sample appeared to restore binding, suggesting reversible oxidation damage was the cause of loss of binding. This introduces the concern that the air-water interface present during the grid freezing process may damage complexes and may be a hypothesis as to why thin ice may have damaged particles (22).

Table 3.3. Experimental considerations of four methods used to measure Keap1-Cul3-Rbx1 complex formation.

Experiment	Concentration	Complex	Experimental Comment
Refeyn Mass Photometry	50 nM	No complex	Works for higher affinity complexes
Cryo-EM Single Particle Analysis	5 – 10 µM	Very little	Complex dissociation observed
Gel Filtration Assay	10 – 40 µM	Complex	Complex formation most efficient
Isothermal Titration Calorimetry	20 – 500 µM	Complex	Weak affinity measured

Keap1-Cul3-Rbx1 appears to suggest a range of affinities based on the experiment used to characterize the complex (Table 3.2). ITC data analysis showed weak affinities in the 100-400 µM range, while gel filtration assays showed near complete complex formation as evidenced by

analytical size exclusion chromatography and SDS-PAGE at 10-40 μM range, suggesting a K_D in the low micromolar range, like that of KLHL11-Cul3 interaction (0.65 μM) and other BTB-domain proteins with Cul3 (Fig. 3.5) (17). Refeyn mass photometry appeared to show that conventional dilution of sample for cryo-EM sample preparation below the micromolar range, for instance as was performed with USP7~Ub-PA at ~135 kDa, resulted in loss of detectable Keap1:Cul3-Rbx1 complex. This would further support binding affinity in the micromolar range.

Overall, cryo-EM single particle analysis revealed that more sample optimization is necessary to collect a dataset amenable for sub-nanometer reconstruction. Of the >200,000 particles obtained from the best 222 micrographs, only 2-3% of the particle set appeared to resemble full complexes (Fig. 3.11). The remaining particles varied in compositional heterogeneity and likely represented a mixture of Keap1, Cul3-Rbx1, and potentially partially denatured and broken particles. Further studies may benefit from glutaraldehyde crosslinking experimentation to more accurately determine proper particle size, and thus optimal magnification and box size parameters for single particle pipeline without crosslinking reagent.

3.5 References

1. Shahd, F., Wells, O. S., Hill, A. D., D'Angiolella, V. D. (2019) Cullin Ring Ubiquitin Ligases (CRLs) in Cancer: Responses to Ionizing Radiation (IR) Treatment. *Front. Physiol*. 10:1144. doi: 10.3389/fphys.2019.01144.
2. Deol, K. K., Lorenz, S., Strieter, E. R. (2019) Enzymatic Logic of Ubiquitin Chain Assembly. *Front. Physiol*. 10:835. doi: 10.3389/fphys.2019.00835.
3. Dinkova-Kostova, A.T., Kostov, R.V., Canning, P. Keap1, the cysteine-based mammalian intracellular sensor for electrophiles and oxidants. (2017). *Arch. Biochem. & Biophys.* 617(1): 84-93.
4. Hashimoto, K. (2018) Essential Role of Keap1-Nrf2 Signaling in Mood Disorders: Overview and Future Perspective. *Front Pharmacol.* 9:1182. doi: 10.3389/fphar.2018.01182.
5. Cleasby, A., Yon, J., Day, P. J., Richardson, C., Tickly, I. J., Williams, P. A., Callahan J. F., Carr, R., Concha, N., Kerns, J. K., Qi, H., Sweitzer, T., Ward, P., Davies, T. G. (2014) Structure of the BTB Domain of Keap1 and Its Interaction with the Triterpenoid Antagonist CDDO. *PLoS ONE*. 9(6): e98896.

6. Small, E., Eggler, A., Mesecar, A.D. (2010) Development of an efficient *E. coli* expression and purification system for a catalytically active, human Cullin3-RINGBox1 protein complex and elucidation of its quaternary structure with Keap1. *Biochem Biophys Res Commun*. 400(4): 471-475.
7. Canning, P., Sorrell, F.J., Bullock, A.N. (2015) Structural basis of Keap1 interactions with Nrf2. *Free Radical Biol & Med*. 88: 101-107.
8. Deshmukh, P., Unni, S., Balasundaram, P. (2017) The Keap1-Nrf2 pathway: promising therapeutic target to counteract ROS-mediated damage in cancers and neurodegenerative diseases. *Biophys Rev*. 9(1): 41-56.
9. McMahon, M., Lamont, D.J., Beattie, K.A., Hayes, J.D. (2010) Keap1 perceives stress via three sensors for the endogenous signaling molecules nitric oxide, zinc, and alkenals. *PNAS*. 107(44): 18838-18843.
10 Zhang, D.D. and Hannink, M. (2003) Distinct Cysteine Residues in Keap1 Are Required for Keap1-Dependent Ubiquitination of Nrf2 and for Stabilization of Nrf2 by Chemopreventive Agents and Oxidative Stress. *Mol & Cell Biol*. 23(22): 8137-8151.
11. Eggler, A.L., Small, E., Mesecar, A.D. (2009) Cul3-mediated Nrf2 ubiquitination and ARE activation are dependent on the partial molar volume at position 151 of Keap1. *Biochem J*. 422(1): doi:10.1042/BJ20090471.
12. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server; (ln) John M. Walker (ed.): The Proteomics Protocols Handbook, *Humana Press*. pp: 571-607.
13. Young, G., Hundt, N., Cole, D., Fineberg, A., Andrecka, J., Tyler, A., Olerinyova, A., Ansari, A., Marklund, E. G., Collier, M. P., Chaundler, S. A., Tkachenko, O., Allen, J., Crispin, M., Billington, N., Takagi, Y., Sellers, J. R., Eichmann, C., Selenko, P., Frey, L., Riek, R., Galpin, M. R., Struwe, W. B., Benesch, J. L. P., Kakura, P. (2018) Quantitative mass imaging of single biological macromolecules. *Science*. 360(6387): 423-427.
14. Zheng, S. Q., Palovak, E., Armache, J. P., Verba, K. A., Cheng, Y., Agard, D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Meth*. 14: 331-332.
15. Rohou, A., Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struc. Biol*. 192(2): 216-221.

16. Punjani, A., Rubenstein, J. L., Fleet, D. J., Brubaker, M. A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Meth*. 14(3): 290-96.
17. Canning, P., Cooper, C. D. O., Krojer, T., Murray, J. W., Pike, A. C. W., Chaikuad, A., Keaters T., Thangaratnarajah, C., Hojzan V., Marsden, B. D., Gileadi, O., Knapp, S., von Delft, F., Bullock, A. N. (2013) Structral Basis for Cul3 Protein Aseembly with the BTB-Kelch Family of E3 Ubiquitin Ligases. *J. Biol*. Chem. 288(11): 7803-7814.
18. Ball, V., Maechling, C. (2009) Isothermal Microcalorimetry to Investigate Non Specific Interactions in Biophysical Chemistry. *Int. J. Mol. Sci.* 10(8): 3283-3315.
19. Iso, T., Suzuki, T., Baird, L., Yamamoto, M. (2016) Absolute Amounts and Status of teh Nrf2-Keap1-Cul3 Complex within Cells. *Mol Cell Biol.* 36(24): 3100-3112.
20. Armstrong, J. K., Wenby, R. B., Meliselman, H. J., Fisher, T. C. (2004) The Hydrodynamic Radii of Macromolecules and Their Effect on Red Blood Cell Aggregation. *Biophys J.* 87 (6): 4259-4270.
21. Zivanov, J., Nakane, T., Forsberg B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., Scheres S. H. J. (2018) New tools for automated high-resolution cryo-EM strucutre determination in RELION-3. *eLife*. 7:e42166.
22. D'Imprima, E., Floris, D., Joppe, M., Sanchez, R., Grininger, M., Kuhlbrandt, W. (2019) Protein denaturation at the air-watrer interface and how to precent it. *eLife*. eLife 2019; 8:e42747.

CHAPTER 4. CRYO-EM SINGLE PARTICLE ANALYSIS OF USP28

4.1 Introduction

USP28 is a human deubiquitinating enzyme (DUB) responsible for controlling the half-life of DNA-damage response proteins and numerous oncoproteins in the cell, including well-studied targets such as c-Myc, Notch, and LSD1 (Table 4.1) (1-14). Aberrant USP28 protein levels are linked to a variety of cancers, and USP28's DUB targets are implicated in disease and affect many signaling pathways in the cell (15-23). Mechanistically, USP28 shares the same catalytic triad and isopeptidase mechanism as related human DUB enzymes and viral proteases (24-29). Structurally, USP28 catalytic domain shares the characteristic papain-like fingers-thumb-palm domain of the catalytic domain, but contains a unique dimerization motif (24-28). Although the feasibility of targeting USP28 has not been thoroughly investigated, USP28 has emerged as a potential drug target for cancer. These studies have lacked in comprehensive understanding of USP28's structure-function.

USP28's role in regulating its DUB substrates are not well understood, particularly its relationship with an antagonistic enzyme complex – the SCF ligase. The prototypical SCF ligase consists of Skp1-Cul1-Fbw7, which together function as an E3 ligase, ligating ubiquitin onto proteins fated for proteasomal-mediated proteolysis (2-4, 6, 14). Biochemical and cellular studies have shown that the F-box and WD-repeat 7 (Fbw7) and UPS28 interact and engage in a diametrically-opposed, antagonistic regulatory mechanism (2-4, 6, 14).

Fbw7 is a substrate-binding component of an enzyme complex that ligates ubiquitin onto proteins for their destruction (5, 12). In the absence of its substrate, this SCF E3 ligase will ligate ubiquitin onto itself, resulting in degradation of Fbw7 (3). It is thought that this represents an auto-regulatory mechanism whereby the SCF ligase will tag Fbw7, it will release, and then be trafficked from the nucleus into the cytoplasm where it is degraded by the proteasome (3). USP28 is hypothesized to be the guardian of the SCF ligase, and potentially the primary DUB for Fbw7 (3, 5, 12). The details of this mechanism are not fully understood.

Outside of the catalytic domain and an NMR assignment for the first 130 residues, there is no useful structural information or mechanistic information on USP28 (28-30). USP28 exhibits dimeric and monomeric forms as determined by analytical ultracentrifugation and analytical size

exclusion chromatography (28-30). Although, there are no comprehensive published studies on the effect of dimerization on enzyme kinetic of USP28, steady-state kinetics experiments in Fig. 4.5 reflect first order reactions, suggesting that dimerization may not be necessary for catalysis. It is hypothesized that dimerization may be the result of hyper-physiological concentrations experienced during purification.

Due to the difficulty in obtaining crystallographic structural data outside of the catalytic domain, it is hypothesized that USP28 may exhibit conformational flexibility – like that of other multi-domain USPs. Therefore, it was suggested that cryo-EM be used to more adequately understand USP28 architecture and conformational movement. Herein is described the experiments that were performed to optimize the USP28 sample for cryo-EM, the kinetics of truncated forms, and an interesting result of GroEL chaperone contamination.

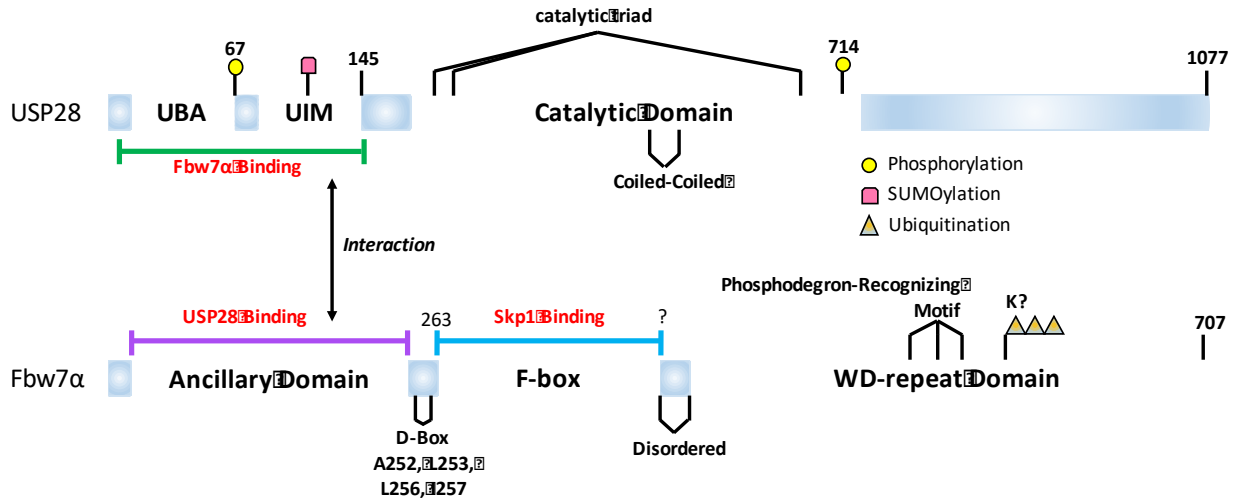


Figure 4.1 Interaction schematic of USP28 & Fbw7. (Top) USP28 structural organization; UBA = ubiquitin-associated domain; UIM = ubiquitin-interacting motif (8, 30-31). (Bottom) Fbw7 structural organization; ancillary domain is a putative term I used to described the domain that interacts with USP28 (8, 30-31). It was hypothesized that USP28's amino terminal domain might be responsible for recognition of ubiquitinated proteins, which is denoted as 'Fbw7 binding'.

Table 4.1. Human USP28 deubiquitination targets.

USP28 Deubiquitination Targets & Interacting Proteins			
Substrate	**Clinical Manifestation**	**Substrate Comments**	**Reference**
c-Myc	Tumorigenesis	USP28's deubiquitination is seemingly improved by binding through Fbw7α, however, it is not necessary.	(1-2)
Fbw7	Tumorigenesis	F-box and WD-repeat 7; substrate-binding component; E3 ligase component	(2-4, 6, 14)
c-Jun	Tumorigenesis	Jun proto-oncogene, AP-1 transcription factor subunit	(3-4)
NICDI	Tumorigenesis	NICDI is Notch Intracellular domain.	(3-4)
Cyclin E	Tumorigenesis	Binds Cdk2 during G_1 initiating DNA synthesis.	(5)
GSK-3β	Angiogenesis	Together regulate hypoxia-inducible transcription factor-1α (HIF-1α)	(6)
LSD1	Breast cancer	Lysine-specific demethylase 1 (LSD1)	(7)
Chk2	Lung sarcoma	Lung sarcoma cell line H460; Chk2-p53-PUMA pathway in DNA-damage-induced apoptosis. USP28 directly deubiquitinates p53.	(8-10)
53BP1/p53	Lung, sarcoma, prolonged mitosis		(8, 10)
Claspin	DNA-damage response		(12)
ZFN304	KRAS+ (CRCs)	Zinc finger DNA-binding protein: colorectal cancers (CRCs)	(13)
mTOR	Tumorigenesis	Mammalian target of rapamycin (mTOR)	(14)

4.2 Methods & Materials

4.2.1 Purification of USP28

Codon-optimized human USP28 was expressed on a pET11a expression vector in BL21 (DE3) *E. coli* (Novagen) from frozen glycerol stock on LB plate supplemented with 100 ug/mL carbenicillin. One well-isolated colony was selected to start an overnight culture; kept at 100 ug/mL carbenicillin. Once culture has reached mid-log phase, scale up to several liters of LB. Grow at 37°C until OD = 0.6-0.8, cool on ice water bath for 10 min, and induce to a final concentration of 0.5 mM IPTG. Let culture continue growing for 12-18 hrs at 18°C. All cultures shaken for adequate oxygenation. Harvest cells via centrifugation at 4,000 x g for 20 min, discard media. Resuspend pelleted cells in Buffer A at 5 mL per gram cell pellet, supplemented with lysozyme, DNase, 5 mM $MgCl_2$, and EDTA-free protease inhibitor cocktail (it does not affect

cysteine proteases). Cells were lysed via sonication on Branson sonifier: 70% amplitude, 1 min per 2 gram cells, 6 sec on, 6 sec off. Clarify lysate via centrifugation at 17,000 x g for 20 min followed by filtration by 0.45 μm syringe filter. Use sample pump to inject lysate over FPLC-linked NiNTA column (2 mL/min). Once lysate is injected (chased with buffer), continue rinsing for ~10 column volumes (CV) with Buffer A or until baseline UV is met. Wash again with 10% Buffer B for 10 CV. A lot of contaminating proteins will be removed in this step. It is done to improve resolution of recovered sample. To elute USP28, an elution gradient over 10 CV from 10-100% Buffer B is performed, collecting 5-10 mL fractions. Peak fractions are tested by ubiquitin-rhodamine110 (Boston Biochem) fluorogenic cleavage assay and SDS-PAGE analysis. Fractions are pooled and placed into TEVp dialysis buffer with TEV protease at a 1:10 protein ratio. Make sure dialysis volume is > 30-50 times the sample volume (for ~50 mL sample – use 2 L). Histidine-tag cleaved sample is re-injected by sample pump over the NiNTA column, with flowthrough collected in 5-10 mL fractions. Flowthrough is tested by Ub-Rho110 cleavage assay and SDS-PAGE analysis. At this point, USP28 can be reacted with ubiquitin-propargylamine or kept in apo form for SEC. Inject USP28 over S200 (300 mL) size exclusion column. The sample will migrate at an apparent molecular weight far above a monomer (122 kDa). Peak fractions are collected and analyzed by SDS-PAGE for purity and Ub-Rho110 for specific activity. Fractions that are 0.8 mg/mL+ and > 95% purity are used for cryo-EM sample prep.

4.2.2 ProteoPlex assay for USP28 sample optimization

USP28 purification conditions have been previously determined for studies, but no successful conditions have been optimized for structural studies by cryo-EM or kinetics. To optimize the final size exclusion buffer for apoUSP28 and USP28~Ub-PA, an adaptation of the ProteoPlex protein stability assay published by the Holger Stark lab was used (32). This method was performed as described in section 2.2. Results of the ProteoPlex assay for USP28 are in Appendix G.

4.2.3 Steady-state kinetics assay for USP28

Isopeptidase activity of USP28 was measured using fluorogenic mono-ubiquitin-rhodamine110 (Ub-Rho110, Boston Biochem) and a plate reader with 485 nm excitation and 535

nm emission (BioTek). 50 nM final USP28 concentration was added to initiate reaction with varying Ub-Rho110 final substrate concentrations (250 nM – 10 μM) in 1:1 volume ratio of 25 μL (2x reagent concentration) to a final reaction volume of 50 μL (1x reagent concentration) for adequate mixing. USP28 and Ub-Rho110 were diluted in Assay Buffer: 50 mM Tris pH 7.5, 100 mM NaCl, 5 mM DTT, 0.01% (w/v) CHAPS, and 0.1 mg/mL BSA (Sigma). The assay reactions were measured in a 96-well black, opaque half-well assay plate (Corning). Fluorescence was measured at 8 second intervals from 3 seconds post initiation to 20 minutes total reaction time.

4.2.4 Cryo-EM specimen preparation, data acquisition, and movie processing

Three microliters of the purified apo full-length USP28 was applied to a Quantifoil R0.6/1 400 mesh grid (Electron Microscopy Sciences) without a prior glow discharge. Grids were coated with five microliters pyrene (1.0 mg/mL) after chloroform washing, and then coated with graphene oxide monolayer (Sigma). The grids were plunge-frozen in liquid ethane with a CP3 Vitrobot (Gatan) under these conditions: temperature, 25°C; humidity > 90%; blotting time, 6-9 s. Frozen grids were imaged in a FEI Titan Krios (300 kV, Thermo Fisher Scientific) equipped with a Gatan Quantum-LS energy filter (20-eV zero-loss filtering) connected to a Gatan K2 Summit detector operating in super-resolution counting mode. Data acquisition was performed as movies of 64 frames over 8000 ms exposure acquired at a nominal magnification of 130,000x. A total dose of 59.30 e^{-} Å^{-2} and a pixel size of 0.545 Å (SR mode) was used during data collection. The acquired movies were processed by motion correction and dose weighting by MotionCor2, and contrast transfer function (CTF) estimation by CTFFIND4 (33-34). A total of 2,000 aligned movies from image acquisition were used for further single-particle processing. Images displaying a resolution less than 7 Å during CTF correction or average drifts higher than 2 Å per frame were excluded from the analysis.

Table 4.2. Cryo-EM data acquisition and refinement parameters for USP28.

Instrumentation	**Sample**	
Titan Krios, K2 DED, VPP, Energy Filter	**USP28**	**USP28~Ub-PA**
Magnification	130,000	130,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	59.3	59.3
Defocus range (μm)	0.5 to 2.5	0.5 to 2.5
Pixel size (Å)	0.545 (SR); 1.09 2x2 bin	0.545 (SR); 1.09 2x2 bin
Exposure frames (#; time)	40; 1300 ms	40; 1300 ms
Total aligned movies (#)	2,000	2,000

4.2.5 Image processing and single particle analysis

Particles were selected with the Relion3.1, cryoSPARCv2, and ctffind4 programs (34-36). During the initial processing, 1500 particles were manually selected from 100 preselected images with cryoSPARCv2 manual picker (36). The selected particles were used for generating auto-picking templates using reference-free 2D classified in cryoSPARCv2 (36). Five 2D classes from this particle set were low-pass–filtered to 10 Å and used as templates for cryoSPARCv2 template-picking particle selection on all aligned images (36). This larger dataset of particles was reference-free 2D classified in cryoSPARCv2, and 18 higher-resolution class averages covering a broader range of views were selected as templates for a subsequent round of Relion3.1 particle selection. Fig. 4.3 shows representative reference-free 2D averages of 208,137 particles demonstrating a large degree of conformational heterogeneity in the particle set.

4.3 Results

4.3.1 Kinetic evaluation of USP28 catalytic efficiency

Several USP enzymes, such as USP1, 2, 7, 8, 11, 12, 16, and 17 – approximately 14% of the human USP enzyme family – have been studied for their catalytic efficiency in cleaving commercially-available fluorogenic mono-ubiquitin substrates (Table 4.2). Of these, USP7 has demonstrated the importance of regulatory domains for intra- and intermolecular regulation of the catalytic domain (Chapter 1). It is not known how many of the USP family of enzymes share these characteristics with USP7, but this information is paramount to developing a reliable structure-activity-relationship for the development of cancer therapeutics in the USP family of enzymes.

Currently, there are no published enzyme kinetics evaluating the role of the N- and C-terminal domains of USP28 in catalysis. As Fig. 4.5 describes, structural and sequence homology

studies on USP28 within the context of the USP family suggests the presence of 3 putative domains: 1) the N-terminal domain, 2) the catalytic domain, and 3) a C-terminal domain (27-30). From the work of a prior student in our lab (Katie Molland), the constructs described in Fig. 4.5 below were generated. Purifications of all constructs are described in section 4.2.1. All kinetics assays were performed as described in section 4.2.1. As Fig. 4.5 and Table 4.2 shows, truncation of the N- and C-terminal resulted in catalytic efficiencies within 2-fold of the full-length, and were not considered statistically significant increases. This indicates that USP28 is not anticipated to carry an intramolecular regulation capacity, like USP7. Steady-state enzyme kinetics analysis have shown USP28 has a K_m like $USP8_{CD}$ or USP12, a turnover rate like USP7 or USP17, and a catalytic efficiency like USP11 (Table 4.2).

It is important to note that the steady-state kinetics analysis of USP28 appear to show first-rate order reactions, regardless of the presence of the N- and C-terminal domains. The results in 4.3.2 and the recent publications by Gersch *et al*. and Sauer, *et al*., appear to show USP28 adopts higher order states, particularly heterodimeric states, discussed in further detail below (28-30). Although, the rates of catalysis of USP28 with increasing [S] do not appear to show an exponential increase at $k[S]^2$ rates, it is important to note that USP28 concentrations for enzyme kinetics is far below micromolar range, and appreciable amounts of homodimer formation are likely not present.

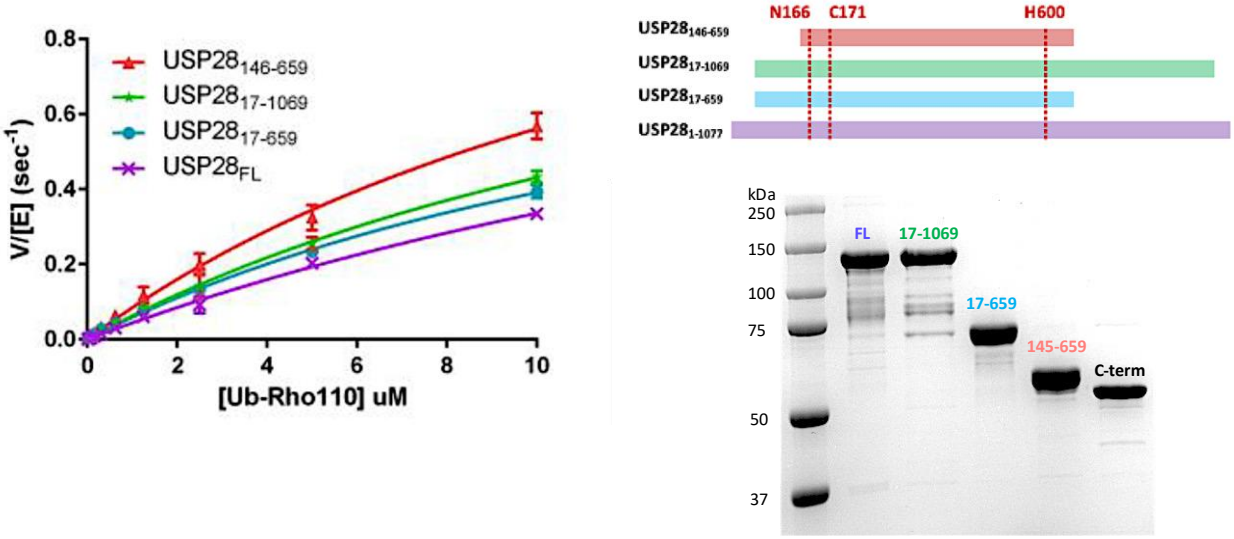


Figure 4.2 USP28 steady-state enzyme kinetics. (Left) USP28 rates of catalysis for constant [E] with increasing [S]. Saturation was not observed with up to 10 μM; replicates with higher [S] were not possible due to material constraints. (Right; Top) Schematic of constructs with catalytic residues indicated at top. (Right; Bottom) SDS-PAGE gel of purity of all enzymes used for kinetics color-matching the curves. C-term is residues 659-1077, and does not contain catalytic activity.

Table 4.3. All currently known USP family enzyme kinetics parameters. All constructs are with parameters as reported in indicated reference; * denotes Ubiquitin-AMC substrate, all others performed with Ub-Rhodamine110. Constructs are assumed full-length, unless denoted.; CD = catalytic domain.

Enzyme	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ((10^{-3}) $s^{-1}M^{-1}$)	Reference
$USP28_{1-1077}$	28.0 ± 5.5	1.28 ± 0.20	45.7 ± 11.4	
$USP28_{17-1069}$	18.7 ± 2.8	1.24 ± 0.13	66.1 ± 12.0	
$USP28_{17-659}$	17.0 ± 2.0	1.06 ± 0.08	62.2 ± 8.70	
$USP28_{146-659}$	16.8 ± 2.6	1.51 ± 0.16	89.6 ± 17.0	
$USP1_{21-785}$	9.71 ± 0.9	0.08 ±0.003	9	[38]
$USP2_{CD}$	2.4 ± 0.2	0.35 ± 0.03	146	[39]
USP7	2.89 ± 0.1	1.37 ± 0.01	482	[38]
USP7*	1.46 ± 0.07	1.24 ± 0.02	850 ± 40	[Hjortland; unpub.]
$USP7_{CD}$*	1.6 ± 0.2	0.040 ±0.002	25 ± 3	[Hjortland; unpub.]
$USP7_{CD}$-H4-5*	2.2 ± 0.2	1.33 ± 0.05	600 ± 70	[Hjortland; unpub.]
$USP8_{CD}$	17.39 ± 2.5	7.90 ± 0.45	464	[38]
USP11	0.77 ± 0.13	0.074 ± 0.003	96	[38]
USP12	12.01 ± 4.2	0.0023 ±0.003	0.2	[38]
USP16	1.42 ± 0.25	1.47 ± 0.02	1500	[40]
USP17*	1.76 ± 0.05	1.47 ± 0.02	880 ± 50	[40]

4.3.2 Glutaraldehyde crosslinking of USP28 catalytic domain reveals tetrameric state

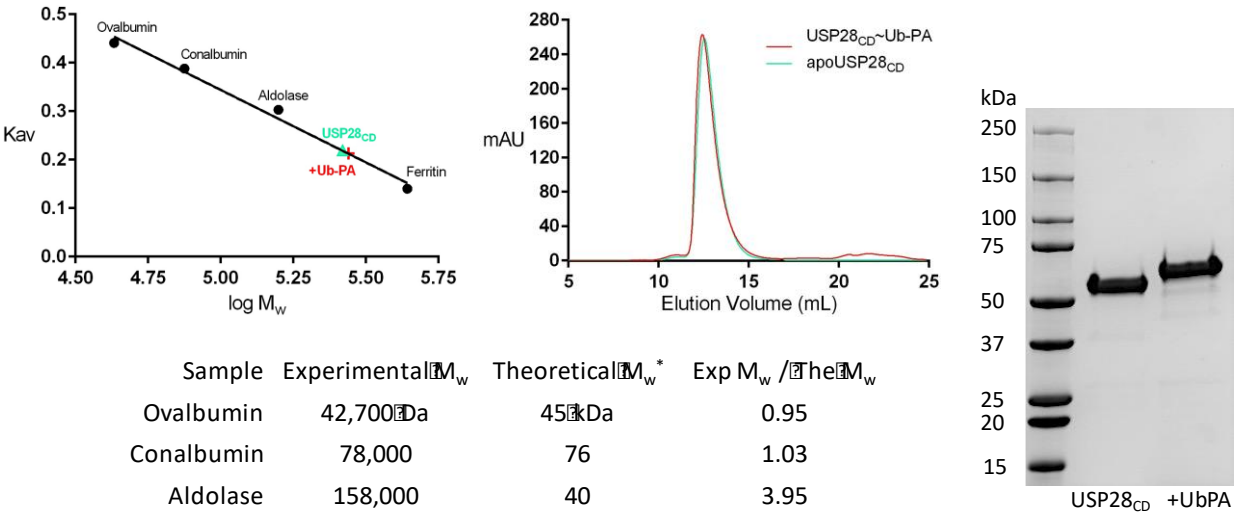
Two research reports were published in Molecular Cell in 2019 that described the USP28 catalytic domain as a dimeric species (28-29). The crystallographic structures of the two different constructs, albeit very similar, showed USP28 packed as a dimer (28-29). Both groups published analytical ultracentrifugation and size exclusion chromatography that corroborated the idea of a USP28 catalytic domain dimer. A previous study published 5 years prior in 2014 by Zhen, *et al*., hinted that the full-length USP28 enzyme eluted via size exclusion as a dimeric species, as well (30). These findings present that USP28 can adopt higher order species, much like USP25, the closet USP family relative of USP28.

USP25 has been shown by Gersch *et al*. and Sauer, *et al*., to adopt a tetrameric species, despite a high degree of structural and sequence similarity (28-29). Analytical size exclusion chromatography and chemical crosslinking experiments performed with our lab's construct of USP28 catalytic domain support the idea of a tetrameric species, like USP25.

Glutaraldehyde crosslinking (Ted Pella reagent) was performed using an 8% (v/v) stock aqueous solution, EM grade, 100.12 MW, Cat # 18421. Glutaraldehyde was diluted from stock

solution in USP28 sample at concentrations from 0.001 – 0.05 % (v/v) and incubated for 1 min. Crosslinking reactions were quenched by the addition of 0.5 M Tris buffer, pH 7.5 to a final concentration of 1 mM in the reaction. Proteins samples (10 μg) were added with SDS-PAGE buffer, and run on a 10% acrylamide SDS-PAGE gel for analyzing crosslinking. As an orthogonal method to measure molecular weight, and to remove crosslinking-induced aggregation, size exclusion was performed with 1 mg total protein samples were loaded onto a Sephacryl 16/60 24 mL SEC column (GE Healthcare) in SEC buffer. All steps were performed at 4°C.

As Fig. 4.6 demonstrates, analytical size exclusion trials performed with USP28 catalytic domain showed both apo and mono-ubiquitin adducts (USP28~Ub-PA) appear to migrate as tetrameric species. This was not observed with the full-length enzyme. To eliminate the potential of elution artifacts due to hydrodynamic radii, crosslinking was performed and then analyzed by SDS-PAGE to measure the size of any complexes formed. Fig. 4.6 shows that apo USP28, USP28~Ub-PA, and glut-crosslinked apo USP28, all elute similar on the Sephacryl 16/60 24 mL SEC column (GE Healthcare), with the addition of significant aggregation in the void volume from the crosslinking reagent. SDS-PAGE of glutaraldehyde-crosslinked samples shows a concentration-dependent decrease in the monomeric species, and the presence of a band between 150-250 kDa, indicating a higher order species than dimer. These results suggest that our lab's catalytic domain construct, which is ~40 residues longer than constructs used for crystallographic studies, adopts a tetrameric state, like USP25 (28-30).



Sample	Experimental M_w	Theoretical M_w*	Exp M_w / The M_w
Ovalbumin	42,700 Da	45 kDa	0.95
Conalbumin	78,000	76	1.03
Aldolase	158,000	40	3.95
apoUSP28$_{CD}$	**236,000**	**58**	**4.07**
USP28$_{CD}$~Ub-PA	**263,000**	**66.5**	**3.96**
Ferritin	440,000	19.8	22.2

Figure 4.3 USP28 catalytic domain analytical size exclusion. (Top, Left) Elution of GE Healthcare size exclusion molecular weight markers to produce a standard curve to compare USP28 catalytic domain. (Top, Middle) Elution profile of apo USP28$_{CD}$ and USP28$_{CD}$~Ub-PA. (Bottom) Summary of analytical SEC Elutions. (Right) Final SDS-PAGE of injected samples, gel-shift corresponds to ~8.5 kDa shift form Ub-PA conjugation.

4.3.3 GroEL chaperone protein complex associates with USP28

As the representative micrograph in Fig. 4.3 shows, there are larger, symmetrical particles, not reminiscent of USP28, present in the dataset. This prompted a study to attempt to reconstruct these particles and try to identify the contaminant. Fig. 4.3 shows 4 representative reference-free 2D averages demonstrating several views of the contaminant made from manual picking in Relion (35). The averages, and the subsequent initial models showed the particle appeared to have 7-fold (D7) symmetry. An initial model from < 200 particles, shown in Fig. 4.3, suggests the contaminant is likely the GroEL chaperone protein complex present in the competent *E. coli* strain. It is important to note that GroEL might serve as a useful benchmark standard for our K3.

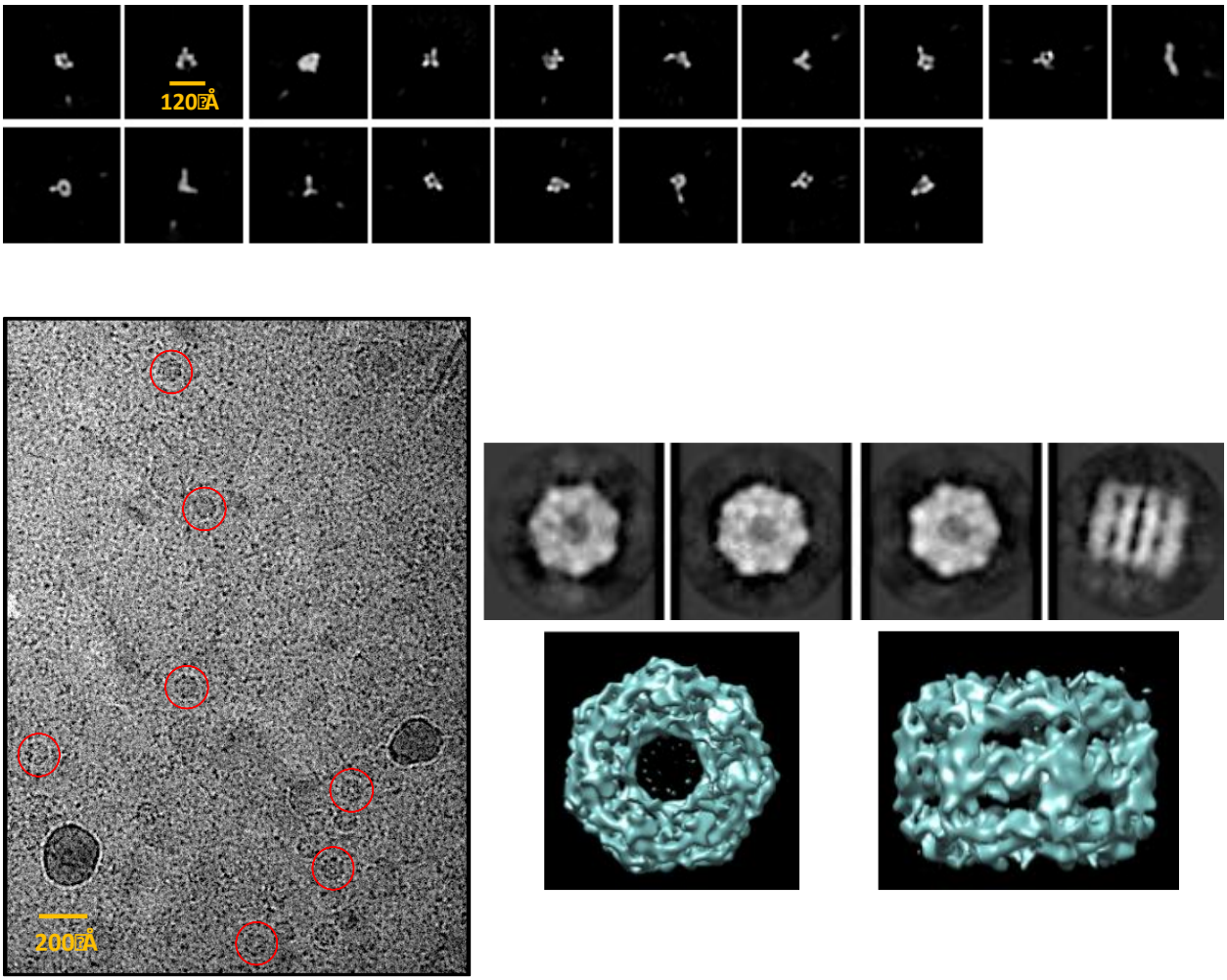


Figure 4.4 USP28 cryo-EM single particle analysis. (Top) Reference-free 2D averages for USP28 demonstrate high degree of conformational heterogeneity. (Bottom, Left) Representative micrograph with contaminating particles circled in red. Scale bar is 200 Å. (Bottom, Right) Reference-free 2D averaging and initial model of contaminating particles.

4.3.4 GroEL consensus sequence mapping in USP28

After the observation that contaminating GroEL was present in the USP28 sample, despite a robust purification strategy, it was thought that human USP28 might contain consensus sequences that solicit GroEL chaperone binding. It was hypothesized that if a consensus sequence could be identified, and then mutated, recombinant USP28 could be recovered without affecting catalytic efficiency, eliminating any GroEL contamination.

A research report published by Chen, *et al*., (Cell, 1999) described 31 unique physiologically-relevant peptides that were found to be among the tightest-binding to GroEL (37).

Several were co-crystallized with GroEL and indicated that they bind directly in the chaperone fold on the interior of the full complex (37).

To test for the presence of any of these consensus sequences in USP28, the sequences were first generated into FASTA files for each individual GroEL consensus sequence, and then individually blasted against the wild-type USP28 protein sequence using PRALINE local consensus sequence alignment (http://www.ibi.vu.nl/programs/pralinewww/). This resulted in an empirical 'score' and percent identity to indicate 'hot spots' in the USP28 sequence. As Fig. 4.4 shows, USP28's C-terminal domain appears to be enriched in sequences that closely resemble GroEL chaperone consensus sequences.

It is important to note that this was performed with the final hexa-histidine tag-cleaved protein sequence of USP28, and not the histidine tag + TEVp sequence that would be present in the cell. As can be seen in Fig. 4.4, many GroEL consensus sequences are histidine-rich. Therefore, it is possible that GroEL binding is due to the presence of the purification tag; however, the presence of a histidine tag has not been reported to substantially increase GroEL contamination in purifying recombinant USPs. Additionally, only the primary sequence was considered due to the chaperone's function in binding nascent polypeptides lacking appreciable folding. Consensus sequences could be mapped to regions with secondary structure that eliminates binding.

Sequence	Score (% Identity)	Residues (span)	Sequence	Score (% Identity)	Residues (USP28)
SWMTTPWGFLHP	124 (25%)	1004-1015 (13)	QEIYLTPRGPQQ	118 (17%)	828-840 (13)
FHYEIWIPPHRG	120 (25%)	559-619 (21)	IDRTQMWRQSDL	126 (18%)	882-893 (11)
SSPWWLVSFTST	106 (17%)	610-621 (12)	INRDHPLHAGQP	138 (33%)	1005-1017 (13)
SHSLIWRIPLLH	134 (33%)	826-848 (23)	HQTPQSLARWSL	110 (17%)	889-900 (12)
IYVPWYYAENLP	134 (33%)	826-838 (12)	HSLRAIQLITGM	126 (33%)	23-34 (12)
YNYSWNGVVFVP	100 (17%)	582-593 (12)	LPSHHHHRVPAA	108 (17%)	944-961 (18)
AQSTPLMKPQKS	122 (25%)	859-870 (12)	IPTYHHHHPSLR	110 (17%)	882-893 (12)
DQTTLQRFLGSH	106 (33%)	1022-1033 (12)	QMTHHHTHRPPI	122 (25%)	351-362 (12)
QTIKPPITVHPS	132 (42%)	356-368 (12)	DLHSHHHGHMNH	116 (8%)	881-892 (12)
QYNHILGYLPFQ	148 (46%)	824-834 (11)	SMHHHHRPASPT	96 (8%)	887-898 (12)
IMDPQNSKVTVA	138 (42%)	1064-1077 (14)	WIGDAKSSLHHA	146 (33%)	1010-1025 (16)
LPIQNAKRSMVS	118 (17%)	865-876 (12)	HNHPHTTSHVSM	110 (25%)	889-900 (12)
IMSPWDESFWNY	116 (17%)	884-895 (12)	HNSIIYHWHTLP	124 (25%)	601-617 (17)
ASESYVLFPGTR	108 (25%)	378-392 (15)	HFNHNHRGFHLI	110 (17%)	592-603 (12)
SNWHGPLSYQLM	136 (40%)	886-897 (10)	AASPHYSSSHSH	108 (17%)	880-891 (12)
ALPLQDTAATLS	124 (42%)	953-964 (12)			

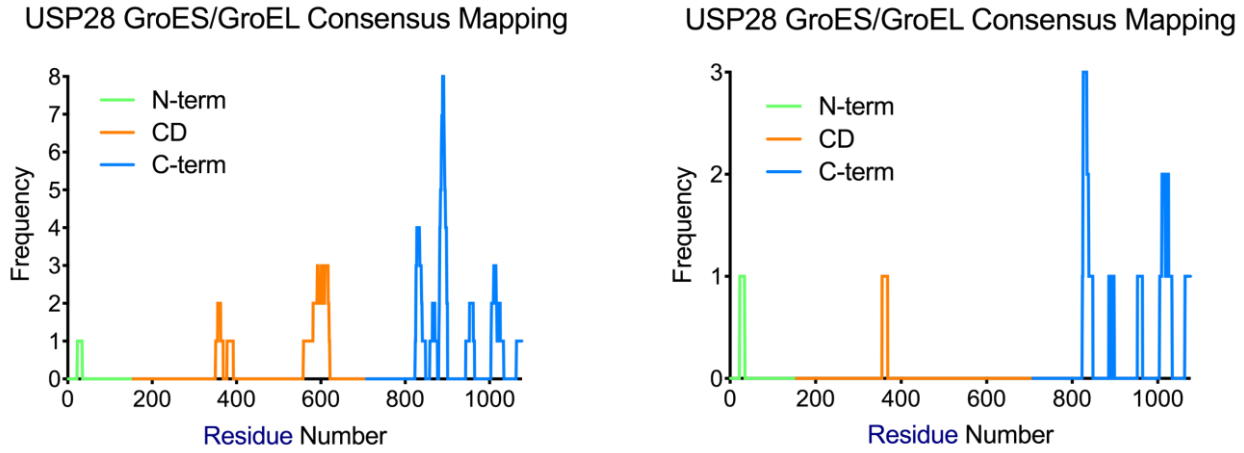


Figure 4.5. USP28 GroEL chaperone consensus sequence mapping. (Top) Published GroEL consensus sequences 'Sequence' mapped to USP28 primary sequence 'Residues'. Scores are from PRALINE local consensus sequence alignment, with percent identity indicated. Residues indicate the span of USP28 primary sequence that best matches the GroEL consensus. 'Span' indicates the presence of gap penalties (> 12 represents gaps in USP28 sequence; < 11 represents gap in GroEL consensus sequence). (Bottom; Left) Mapping of GroEL consensus sequences to USP28 primary for all GroEL consensus sequences. (Bottom; Right) Mapping of GroEL consensus sequences to USP28 for only consensus sequences that matched with ≥ 33% identity.

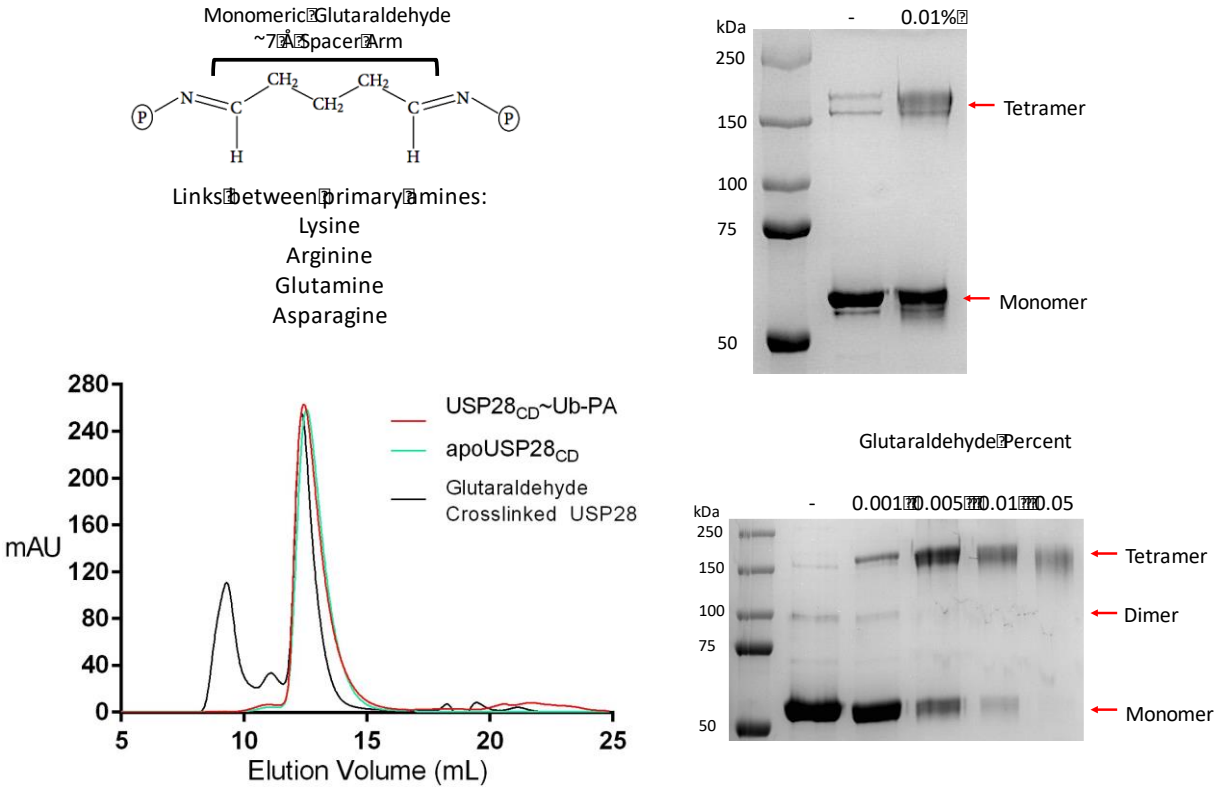


Figure 4.6 USP28 catalytic domain glutaraldehyde crosslinking. (Top) Glutaraldehyde linker arm schematic. (Top, Right) SDS-PAGE gel without reducing agent or boiling sample showing the before and after of glutaraldehyde addition at room temp. (Bottom, Left) Size exclusion chromatography (S200; GE Healthcare) of apo, Ub-PA-bound, and crosslinked apo USP28 catalytic domain. (Bottom, Right) Concentration dependent glutaraldehyde crosslinking of USP28.

4.4 Discussion

USP28 has been shown by Gersch *et al*. and Sauer, *et al*., to adopt a dimeric species, although our analytical SEC and crosslinking experiments appear to show that $USP28_{CD}$ may adopt a tetrameric structure in solution (28-30). Although, it is important to note that our construct has amino acids that were truncated for crystallization. Closer inspection of the space groups and packing indicated that a tetrameric structure was not mistaken as a dimer of homodimers (28-29).

Our steady-state kinetics experimentation of USP28 show that truncation of the N- and C-terminal domains has no statistically significant effect on K_M or k_{cat}, in that all were within 2-fold of wild-type. The reactions appear to proceed with first-rate kinetics, indicating that dimerization is likely not a factor, however, the assays are performed well below micromolar rage (50 nM), and thus higher order states are likely a small proportion of the equilibrium of states.

Since USP28 was hypothesized to adopt a D2 symmetry dimeric structure of ~250 kDa, it was thought that the enzyme would be amendable to study by cryo-EM. Data collection showed that purification of recombinant USP28 resulted in substantial GroEL contamination, despite migration on size exclusion showing a native, dimeric molecular weight. More data collection and analysis might be needed to further investigate strategies for studying USP28 by cryo-TEM.

4.5 References

1. Popov, N., Wanzel, M., Madiredjo, M., Zhang, D., Beijersbergen, R., Bernards, R., Moll, R., Elledge, S. J., Eilers, M. (2007) The ubiquitin-specific protease USP28 is required for MYC stability. *Nat. Cell Biol*. 9(7): 765-74.
2. Popov, N., Herold, S., Llamazares, M., Schülein, C., Eilers, M. (2007) Fbw7 and Usp28 Regulate Myc Protein Stability in Response to DNA Damage. *Cell Cycle*. 6(19): 2327-31.
3. Schülein-Volk, C., Wolf, E., Zhu, J., Xu, W., Taranets, L., Hellmann, A., Jänicke, L. A., Diefenbacher, M. E., Eilers, M., Popov, N. (2014) Dual regulation of Fbw7 function and oncogenic transformation by USP28. *Cell Rep*. 6(9): 1099-109.
4. Welcker, M., Clurman, B. E. (2008) FBW7 ubiquitin ligase: a tumor suppressor at the crossroads of cell division, growth and differentiation. *Nat. Rev. Cancer*. 8: 83-93.
5. Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W., Pavletich, N. P. (2007) Structure of a Fbw7-Skp1-Cyclin E Complex: Multisite-Phosphorylated Substrate Recognition by SCF Ubiquitin Ligases. *Mol. Cell*. 26(1): 131-43.
6. Flügel, D., Görlach, A., Kietzmann, T. (2012) GSK-3β regulates cell growth, migration, and angiogenesis via Fbw7 and USP28-dependent degradation of HIF-α. *Blood*. 119(5): 1292-300.
7. Wu, Y., Wang, Y., Yang, X. H., Kang, T., Zhao, Y., Wang, C., Evers, B. M., Zhou, B.P. (2013) The Deubiquitinase USP28 Stabilizes LSD1 and Confers Stem-Cell-like Traits to Breast Cancer Cells. *Cell Rep.* 5: 244-36.
8. Zhang, D., Zaugg, K., Mak, T. W., Elledge, S. J. (2006) A Role for the Deubiquitinating Enzyme USP28 in Control of the DNA-Damage Response. *Cell*. 126: 529-42.
9. Samara, N. L., Datta, A. B., Berndsen, C. E., Zhang, A., Yao, T., Cohen, R. E., Wolberger, C. (2010) Structural Insights into the Assembly and Function of the SAGA Deubiquitinating Module. *Science*. 328: 1025-29.

10. Cuella-Martin, R., Oliveira, C., Lockstone, H. E., Snellenberg, S., Grolmusova, N., Chapman, J. R. (2016) 53BP1 Integrates DNA Repair and p53-Dependent Cell Fate Decisions via Distinct Mechanisms. *Mol Cell*. 64(1):51-64. 2016.
12. Bassermann, F., Frescas, D., Guardavaccaro, D., Busino, L., Peschiarolli, A., Pagano, M. (2008) The Cdc14B-Cdh1-Plk1 Axis Controls the G2 DNA-Damage-Response Checkpoint. *Cell*.134: 256-67.
13. Serra, R. W., Fang, M. Park, S. M., Hutchinson, L., Green, M. R. (2014) A KRAS-directed transcriptional silencing pathway that mediates CpG methylator phenotype. *eLife*. 3:e02313.
14. Mao, J. H., Kim, J., Wu, D., Climent, J., Kang, H. C.,DelRosario, R., *et al*. (2008) FBXW7 targets mTOR for degradation and cooperates with PTEN in tumor suppression. *Science*. 321:1499-1502.
15. Ren, K., Li, Y., Lu, H., Li, Z., Han, X. (2017) miR-3940-5p Functions as a Tumor Suppressor in non-Small Cell Lung Cancer Cells by Targeting Cyclin D1 and Ubiquitin Specific Peptidase-28. *Transl Oncol*. 10(1):80-9.
16. Cao, C., Vasilatos, S. N., Bhargava, R., Fine, J. L., Oesterreich, S., Davidson, N. E., Huang, Y. (2017) Functional interaction of histone deacetylase 5 (HDAC5) and lysine-specific demethylase 1 (LSD1) promotes breast cancer progression. *Oncogene*. 36(1):133-45.
17. Shah, S. P., Morin, R. D., Khattra, J., Prentice, L., *et al*. (2009) Mutational evolution in a lobular breast tumor profiled at single nucleotide resolution. *Nature*. 461:809-13.
18. Li, Y. Y., Wu, C., Chen, S. M., Shah, S. S., Wangpaichitr, M., Feun, L. G., Suarez, M., Prince, J., Savaraj, N. (2016) BRAF inhibitor resistance enhances vulnerability to arginine deprivation in melanoma. *Oncotarget*. 7(14):17665-80.
19. Wang, Z., Song, Q., Xue, J., Zhao, Y., Qin, S. (2016) Ubiquitin-specific protease 28 is overexpressed in human glioblastomas and contributes to glioma tumorigenicity by regulating MYC expression. *Exp Biol Med (Maywood)*. 241(3):255-64.
20. Diefenbacher, M. E., Chakraborty, A., Blake, S. M., Mitter, R., Popov, N., Wilers, M., Behrens, A. (2015) Usp28 counteracts Fbw7 in intestinal homeostasis and cancer. *Cancer Res*. 75(7):1181-6.
21. Radeva, M., Hofmann, T., Altenberg, B., Mothes, H., Ritcher, K. K., Pool-Zobel, B., Greulich, K. O. (2008) The database dbEST correctly predicts gene expression in colon cancer patients. *Curr Pharm Biotechnol*. 9(6):510-5.

22. Han, H., Sun, D., Li, W., Shen, H., Zhu, Y., Li, C., Chen, Y., Lu, L., Li, W., Zhang, J., Tian, Y., Li, Y. (2013) A c-Myc-MicroRNA functional feedback loop affects hepatocarcinogenesis. *Hepatology*. 57(6);2378-89.
23. Guo, G., Xu, Y., Gong, M., Cao, Y., An, R. (2014) USP28 is a potential prognostic marker for bladder cancer. *Tumor Biol*. 35(5):4017-22.
24. Fraile, J. M., Quesada, V., Rodriguez, D., Freije, J. M. P., and Lopez-Otin, C. (2012) Deubiquitinases in cancer: new functions and therapeutic options. *Oncogene*. 31, 2373–2388.
25. Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C., and Mesecar, A. D. (2006) Severe acute respiratory syndrome coronavirus papain-like protease: Structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5717–5722.
26. Clasman, J. R., Báez-Santos, Y. M., Mettelman, R. C., O'Brien, A., Baker, S. C., and Mesecar, A. D. (2017) X-ray Structure and Enzymatic Activity Profile of a Core Papain-like Protease of MERS Coronavirus with utility for structure-based drug design. *Sci. Rep.* **7**, 40292.
27. Ndubaku, C., Tsui, V. (2015) Inhibiting the Dubiquitinating Enzymes (DUBs). *J Med Chem*. 58:1581-95.
28. Gersch M., Wagstaff, J. L., Toms, A. V., Graves, B., Freund, S. M. V., Komander, D. (2019) Distinct USP25 and USP28 Oligomerization States Regulate Deubiquitinating Activity. *Mol. Cell*. 74(3): 436-451.
29. Sauer, F., Klemm, T., Kollampally, R. B., Tessmer, I., Nair, R. K., Popov, N., Kisker, C. (2019) Differential Oligomerization of the Deubiquitinatses USP25 and USP28 Regulates Their Activities. 74(3): 421-435.f.
30. Zhen, Y., Knobel, P. A., Stracker, T. H., Reverter, D. (2014) Regulation of USP28 Deubiquitinating Activity by SUMO Conjugation. *J. Biol. Chem.* 289: 34838-50.
31. Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W., Pavletich, N. P. (2007) Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell*. 26(1): 131-43.

32. Chari, A., Haselbach, D., Kirves, J. M., Ohmer, J., Paknia, E., Fischer N., Ganichkin O., Möller, V., Frye, J. J., Petzold, G., Jarvis, M., Tietzel, M., Grimm, C., Peters, J. M., Schulman, B. A., Tittmann, K., Markl, J., Fischer, U., Stark H. (2015) ProteoPlex – stability optimization of macromolecular complexes by spare-matrix screening of chemical space. *Nat. Meth.* 12(9): 859-865.
33. Zheng, S. Q., Palovak, E., Armache, J. P., Verba, K. A., Cheng, Y., Agard, D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Meth*. 14: 331-332.
34. Rohou, A., Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struc. Biol*. 192(2): 216-221.
35. Zivanov, J., Nakane, T., Forsberg B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., Scheres S. H. J. (2018) New tools for automated high-resolution cryo-EM strucutre determination in RELION-3. *eLife*. 7:e42166.
36. Punjani, A., Rubenstein, J. L., Fleet, D. J., Brubaker, M. A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Meth*. 14(3): 290-96.
37. Chen, L., Siger, P. B. (1999) The Crystal Structure of a GroEL/Peptide Complex. *Cell*. 99(7): P757-768.
38. Faesen, A. C., Luna-Vargas, M. P. A., Geurink, P. P., Clerici, M., Merkx, R., van Dijk, DS Hameed W. J., El Oualid, F., Ovaa, H., Sixma, T. K. (2011) The differential modulation of USP activity by internal regulatory domains, interactions and eight ubiquitin chain types. *Chem Biol*. 18:1550-61.
39. Bozza, W. P., Liang, Q., Gong, P., Zhuang, Z. (2012) Transient kinetic analysis of USP2-catalyzed deubiquitination reveals a conformational rearrangement in the K48-Linked diubiquitin substrate. *Biochemistmisty*. 51:10075-86.
40. Hjortland, N. M., Mesecar, A. D. (2016) Steady-state kinetic studies reveal that the anti-cancer target Ubiquitin-Specific Protease 17 (USP17) is a highly efficient deubiquitinating enzyme. *Arch Biochem Biophys*. 612:35-45.

APPENDIX A. PROTEOPLEX ASSAY FOR CRYO-EM SAMPLE PREP

ProteoPlex for Condition Optimization of USP7

USP7 purification conditions have long been determined for structural studies by x-ray crystallography, but conditions have not yet been optimized for structural studies by cryo-EM. To optimize the final size exclusion buffer for apoUSP7 and USP7~Ub-PA, an adaptation of the ProteoPlex protein stability assay was used (53). This method results in a high-throughput pipeline for sparse-matrix screening of final purification conditions – to identify conditions that keep USP7 most thermally stable. Generally, the assay requires dilution of > 95% purity enzyme from the currently optimized purification, an instrument that can measure fluorescence with careful temperature control, and the ability to adapt 96 or 385-well plates.

Thermal Stability Assay

Fluorescence as a function of temperature was collected on a QS5 qPCR instrument (Applied Biosciences) with the SYBR™ Green filter setting to measure protein melting transitions. A final protein concentration of 1.0 mg/mL, with 1x final concentration of GloMelt™ (Biotium), was diluted into a final volume of 20 μL. The following screens were performed for determining: i) the optimal buffer and pH, ii) the optimal salt and ionic strength, and iii) the appropriate type of polyol and its concentration. The plates were set up in the following way:

Table A.1. ProteoPlex adapted spare-matrix buffer screening. All buffers were at a final concentration of 0.1 M – like that typical of crystallization screen buffer concentrations. SPG = succinic acid, sodium phosphate, glycine in 33.3 mM equimolar stoichiometry; St. represents the standard to which each row of thermal melts was compared, which corresponds to the current purification buffer at equal concentration (100 mM HEPES, pH 7.5).

SPG	MOPS	PIPES	MES	NaCitrate	BisTris	Na/KPO_4	Bicine	HEPES	Imidazole	Tris	St.
5.6	6.5	6.1	5.5	5.5	5.8	5.8	7.6	6.8	6.2	7.5	-
6.0	6.7	6.3	5.7	5.75	6.0	6.1	7.8	7.0	6.45	7.7	-
6.4	6.9	6.5	6.9	6.0	6.2	6.4	8.0	7.2	6.7	7.9	-
6.8	7.1	6.7	6.1	6.25	6.4	6.7	8.2	7.4	6.95	8.1	-
7.2	7.3	6.9	6.3	6.5	6.6	7.0	8.4	7.6	7.2	8.3	-
7.6	7.5	7.1	6.5	6.75	6.8	7.3	8.6	7.8	7.45	8.5	-
8.0	7.7	7.3	6.7	7.0	7.0	7.6	8.8	8.0	7.7	8.8	-
8.4	7.9	7.5	6.9	7.25	7.2	7.9	9.0	8.2	7.95	9.0	-

Table A.2. ProteoPlex adapted spare-matrix polyol screening. All buffers were at low final concentration in percent (v/v) or percent (w/v) depending on state at ambient temperature and pressure. It is typically suggested that polyols be removed from the final buffer for optimizing ice thickness when plunge-freezing. Although USP7 is purified with 5% (v/v) glycerol, screening was done for low concentrations of polyols for due diligence. St. represents 5% (v/v) glycerol. Suc = sucrose, Glu = glucose, Lac = lactose, Gal = galactose, Man = mannitol, Sor = sorbitol, Gly = glycerol.

Suc	Glu	Lac	Gal	Man	Sor	Gly	Pentaerythritol	MPD	PEG400	PEG3350	St.
1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	-
2%	2%	2%	2%	2%	2%	2%	2%	2%	2%	2%	-
3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	-
4%	4%	4%	4%	4%	4%	4%	4%	4%	4%	4%	-
5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	-
6%	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%	-
7%	7%	7%	7%	7%	7%	7%	7%	7%	7%	7%	-
8%	8%	8%	8%	8%	8%	8%	8%	8%	8%	8%	-

Table A.3. ProteoPlex adapted spare-matrix salt Screening. It was observed through numerous dialysis experiments that USP7 was not stable at salt concentrations below 100 mM for NaCl, resulting in aggregation and loss of protein. This corresponds to a conductivity of ~10 mS/cm; therefore, salts were screened at 100+ mM. Traditionally, ionic strength of buffers for freezing should be as close to pure water as possible to improve the ability to optimize thin ice without getting hexagonal ice contamination or salt crystal formation. St. = 150 mM NaCl.

$MgCl_2$	NaCl	$CaCl_2$	LiCl	KCl	$ZnCl_2$	$MnCl_2$	NH_4Cl	K_2SO_4	$(NH_4)_2SO_4$	$NH_4CH_3CO_2$	St.
100 mM	100	100	100	100	100	100	100	100	100	100	-
120 mM	120	120	120	120	120	120	120	120	120	120	-
140 mM	140	140	140	140	140	140	140	140	140	140	-
160 mM	160	160	160	160	160	160	160	160	160	160	-
180 mM	180	180	180	180	180	180	180	180	180	180	-
200 mM	200	200	200	200	200	200	200	200	200	200	-
220 mM	220	220	220	220	220	220	220	220	220	220	-
240 mM	240	240	240	240	240	240	240	240	240	240	-

Raw Data Analysis

The raw fluorescence data was exported from the 'Thermal Protein Analysis' software as a .csv file. From here, the data was treated with a program to normalize the raw fluorescence values to compare melting temperatures between samples. These fluorescence values were normalized between 0-1 using the equation below:

$$Normalized\ Fluorescence\ (AU) = \frac{value - min}{max - min}$$

The 'min' value corresponds to the minimal temperature prior to the melting transition. This signifies the beginning of the transition state. For instance, 30-40 degrees C just before the

increase in fluorescence from melting was observed. The ‘max’ corresponds to the maximal fluorescence amplitude after the determined ‘min’. Oftentimes there is a large amplitude of fluorescence from dye binding to protein surfaces and/or unfolded portions before the transition. It is assumed the transition state occurs after the local minima in the beginning of the melting curve, and the maximal signal from melting is reached after this point. Once normalized, all fluorescence amplitudes within the transition state will be between 0 and 1, with a curve resembling a Boltzmann distribution. This way, differences in the slopes and maxima can be more accurately compared between samples.

The optimal condition for USP7 was HEPES buffer pH 7.4-7.8, 180-220 mM NaCl, and nearly all percents of glycerol. The addition of polyols at the concentration range tested did not greatly affect the melting temperature. Thus, it was concluded that the final purification buffer would be: 25 mM HEPES pH 7.7, 200 mM NaCl, 4 mM DTT.

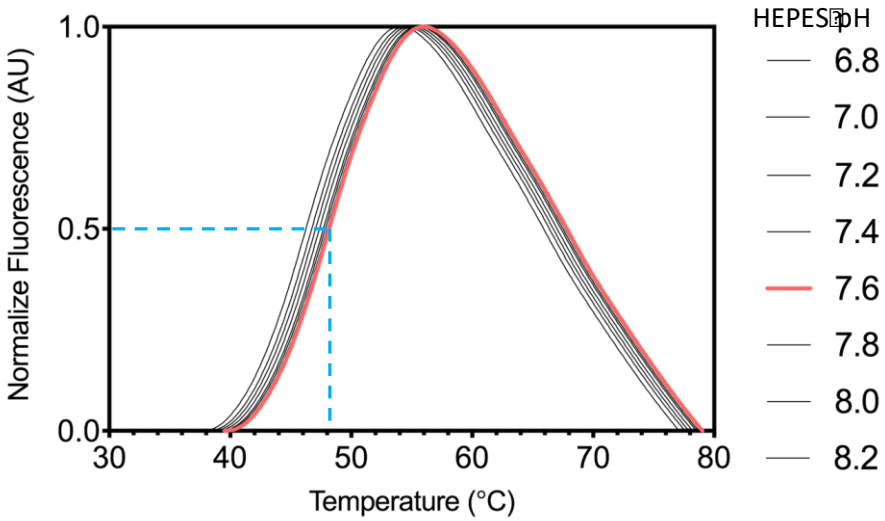


Figure A.1. ProteoPlex thermal stability assay example. Above is an example condition series of USP7 for the highest stability buffer, HEPES. All non-optimal conditions are in black, with the highest thermal stability condition, HEPES pH 7.6, highlighted in red. The range of melting temps was approx. 2-3˚C for all buffers tested. The blue dashed line represents the half-maximal fluorescence, which is observed as the melting temperature for comparison between conditions. Optimal condition for USP7 above is HEPES pH 7.6, corresponding to a T_m = 47.5˚C. Each curve represents the mean of three replicates.

APPENDIX B. PRESCISSION PROTEASE FOR USP7 TAG CLEAVAGE

Human Rhinovirus 3C Protease (PreScission Protease) for USP7 purification tag cleavage

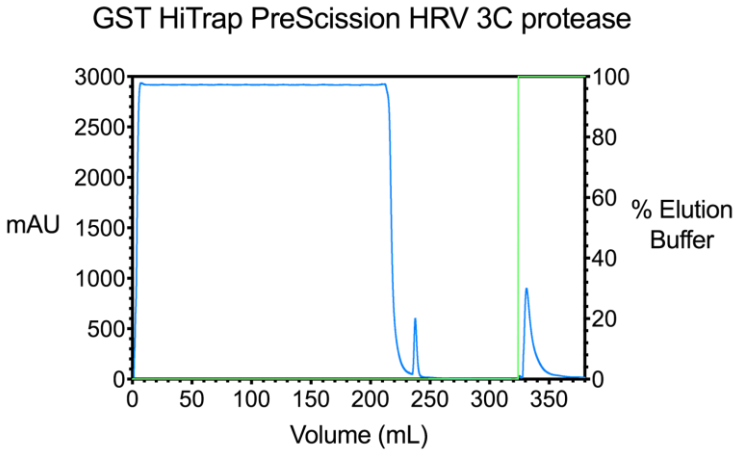
A bacterial cell pellet of PreScission protease from was grown and harvested 7 years prior to purification. PreScission protease is the Human Rhinovirus 3C protease that can be used for high-fidelity cleavage of purification tags with the recognition sequence: Leu-Phe-Gln/Gly-Pro (slash denotes isopeptide hydrolysis). This purification also served as an experiment to determine the feasibility of recovering active protease after 7-year storage at -80°C. Usual circumstances would dictate that a pellet of this age not be used for protein purification; however, we no longer possess stocks of purified PreScission protease, nor was the plasmid or original bacterial stock present. The codon-optimized GST-Human Rhinovirus 3C protease was expressed in BL21 (DE3) *E. coli* on a pGEX-6P expression plasmid under carbenicillin (50 μg/mL). The culture was grown at 37°C until an OD_{600nm} of ~0.6, at which point it was supplemented with IPTG to 0.5 mM final concentration and grown at 30°C for 15 hours. The cell pellet was harvested via centrifugation at 5,000 x g for 20 min.

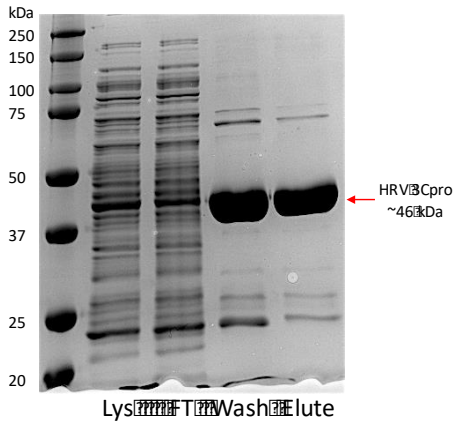
The pellet was thawed on ice and resuspended in lysis buffer: 25 mM HEPES pH 7.5, 250 mM NaCl, 5 mM BME, 5% (v/v) glycerol, 50 μg/mL DNase (Sigma), 25 μg/mL lysozyme (Thermo Fisher), and 5 mM $MgCl_2$ at a volume of 3 mL per gram cells. The cells were sonicated by a Branson sonifier 65% amplitude, 6 s pulse on, 6 s pulse off at 1 min per 2 g cell pellet. The lysate was clarified by ultracentrifugation at 80,000 x g for 30 min and then applied via FPLC to a 5 mL GST HiTrap column (GE Healthcare).

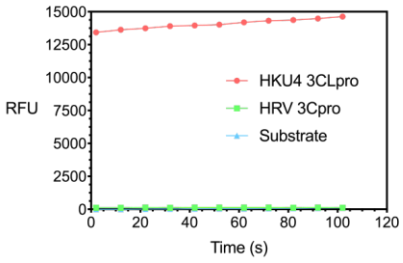
After sample was applied, the GST column was rinsed with wash buffer: 25 mM HEPES pH 7.5, 250 mM NaCl, 5 mM BME, 5% (v/v) glycerol until UV baseline was reached. Then, sample was eluted with elution buffer: 20 mM HEPES pH 8.0, 150 mM NaCl, 10 mM glutathione (reduced), 5 mM BME. Eluted sample was tested via SDS-PAGE assay for the presence of a 46 kDa protein and via a fluorogenic peptide cleavage assay using a peptide to test for the activity of other 3C viral proteases with the following peptide sequence: ESATLQ/SGLRKAK. Peptide cleavage was measure on a BioTek instrument (488 Ex; 520 Em), reading fluorescence every 10 s with HKU4 3CLprotease as a positive control. Protein concentration was 200 nM and peptide

substrate was 2 μM, [2 μM HRV PreScission 3C protease was added at the end to test for the presence of any isopeptide bond cleavage].

Samples showing > 90% purity were pooled and placed into dialysis buffer: 50 mM HEPES pH 8.0, 150 mM NaCl, 10% (v/v) glycerol, 5 mM EDTA, 5 mM DTT for 12 hours and snap frozen in liquid nitrogen to be stored at -80°C. All purification steps were carried out at 4°C over the course of 48 hrs.







	Amount	Total Recovered
Wash Peak	15 mg	
Elutant	30 mg	
Dialysis		28 mg (93%)
5 mL GST HiTrap	50 mg	45 mg (90%)

Figure B.1. PreScission HRV 3C protease GST purification. (Top left) Chromatograph of 5 mL GST HiTrap FPLC purification, (top right) 10% SDS-PAGE analysis, HRV 3C protease appears at ~46 kDa (see Dr. Hjortland's Dissertation for reference), (bottom right) activity assay using ESATLQ/SGLRKAK fluorescent peptide substrate – no cleavage was detected up to 2 μM HRV 3C concentration, (bottom left) purification summary – 28 mg protease were recovered from this run (~30 g cells; nearly 1 mg/g cell pellet recovered).

APPENDIX C. CRYO-EM SAMPLE PREP PIPELINE FOR USP7

Summary of USP7~Ub-PA and grid preparation strategy

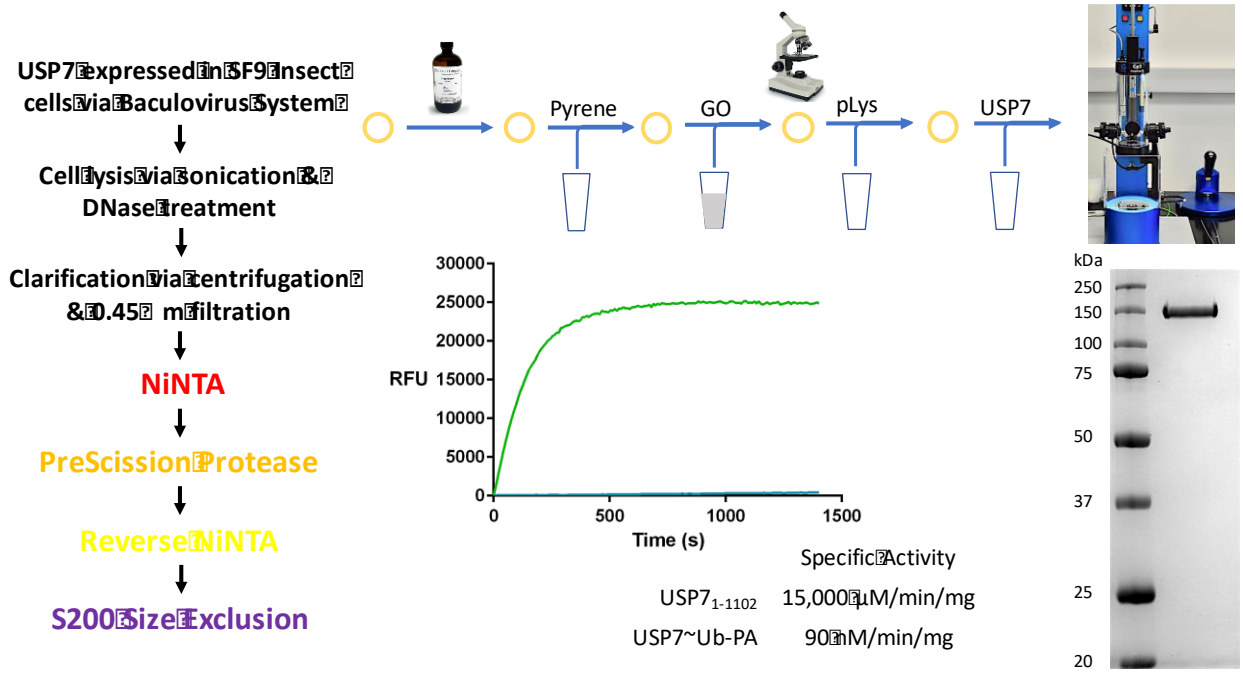


Figure C.1. Purification summary and grid preparation overview for USP7 cryo-TEM imaging. (Left) Schematic of steps involved in USP7 purification. (Middle) Mono-Ub-Rho110 (BostonBiochem) fluorescence cleavage assay showing apoUSP7 (green) and USP7~UB-PA (blue) with specific activity indicated below, showing > 99% inhibition. (Right) Final SDS-PAGE gel of reacted USP7~Ub-PA used for cryo-TEM. (Top) Schematic of grid preparation; first is chloroform rinsing of gold, Quantifoil R1.2/1.3, 400 mesh grids (EMS), followed by application of pyrene, then graphene oxide (GO; Sigma), and then poly-lysine (pLys) to coat the grid in a basic amino acid surface, then application of USP7, and lastly plunge-freezing in liquid nitrogen-cooled liquid ethane using the CP3 cryoplunge instrument (Gatan).

APPENDIX D. MONO-UBIQUITIN-PRORARGYLAMINE (UB-PA) AS A PROBE FOR USP7

Effect of mono-Ub-PA on USP7

During initial screening by negative stain TEM, a large degree of compositional heterogeneity was observed in USP7. It was hypothesized that use of a synthetic mono-ubiquitin probe might lock USP7 into an *in cis* conformation, like that of the proposed EP_2 state. The resulting covalent adduct with mono-Ub would then increase particle homogeneity, alleviating any pitfalls in single particle analysis due to heterogeneity.

To this end, one strategy takes advantage of an intein-mediated amide to thioester transthioesterification reaction, otherwise referred to as: N to S acyl migration. $Ubiquitin_{1-75}$ (lacking the final carboxy-terminal glycine) followed by an intein peptide sequence is collected on chitin beads via a chitin-binding domain (CBD) fused to the intein stretch. N to S acyl migration occurs spontaneously under purification conditions to form the thioester intermediate, which can be eluted from chitin beads with *Sodium 2-mercaptoethanesulfonate* (MesNa) to generate the Ub-MesNa adduct. Subsequent reaction of the MesNa thioester with different "warheads", or reactive functional groups, can yield desired ubiquitin probes. The strategy here uses propargylamine HCl to replace MesNa, where the terminal alkyne acts as the Michael acceptor that can form a covalent adduct with the active site cysteine. Several chemical modifications have been successful for USP adducts, for example, Ub-aldehyde used for USP7 catalytic domain crystallographic structure (37). Although, whether mono-Ub will serve to lock full-length USP7 into any specific conformation has yet to be investigated. It was hypothesized that use of this mono-Ub probe would decrease particle heterogeneity by conformationally locking UPS7 into a HUBL5-engaged state (EAP2, Fig. 1.1).

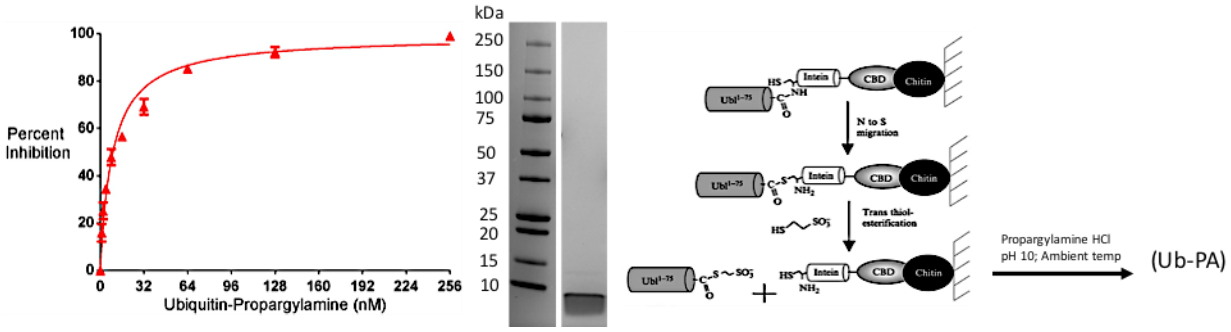


Figure D.1. Ubiquitin-propargylamine (Ub-PA) as chemically-modified mono-ubiquitin probe for USP7. (Left) IC_{50} determination of Ub-PA for USP7; IC_{50} is determined to be ~15 nM. Assay conditions were 5 nM USP7, 500 nM Ub-Rho. Saturation of USP7 is observed at 20-fold+ molar excess Ub-PA. (Middle) Final SDS-PAGE analysis of Ub-PA. After S75 size exclusion samples are generally > 95% purity, showing a band ~ 8.5 kDa. (Right) Schematic detailing chemistry occurring at the c-terminal end during the purification of Ub-PA.

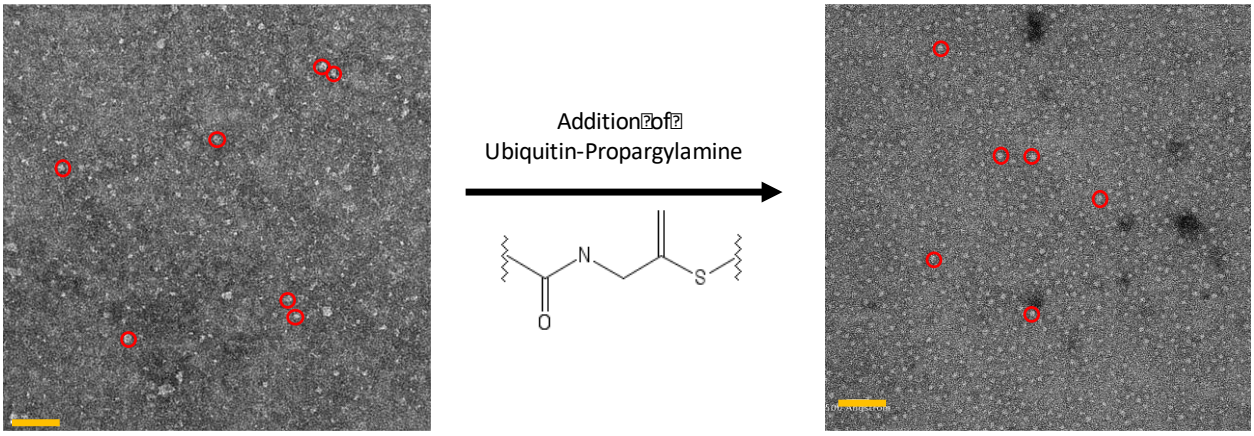


Figure D.2. Ubiquitin-propargylamine (Ub-PA) improved particle composition as seen by negative stain TEM. (Left) apoUSP7 visualized at 0.06 mg/mL and stained with uranyl-formate. (Right) Conjugation of ubiquitin-propargylamine resulted in USP7 (at 0.06 mg/mL) molecules more conducive to imaging and single particle analysis. The orange scale bar represents ~100 nm in each image. Red circles enclose representative particles for comparison.

APPENDIX E. CRYO-EM SINGLE PARTICLE ANALYSIS PIPELINE

Cryo-EM Single Particle Analysis Pipeline for USP7

Table E.1. Cryo-EM data acquisition and refinement parameters for USP7~Ub-PA.

Instrumentation	Sample
Titan Krios, K2 DED, VPP, Energy Filter	**USP7~Ub-PA**
Nominal Magnification	130,000
Voltage (KeV)	300
Electron exposure (e–/Å^2)	80
Defocus range (μm)	0.5 to 1.5
Pixel size (Å)	0.545 (SR); 1.09 2x2 bin
Exposure frames (#; time)	64; 8000 ms
Symmetry imposed	C1
Initial particle images (no.)	525,578
Final particle images (no.)	183,833
FSC threshold	0.143
Map resolution (Å)	8.2 Å

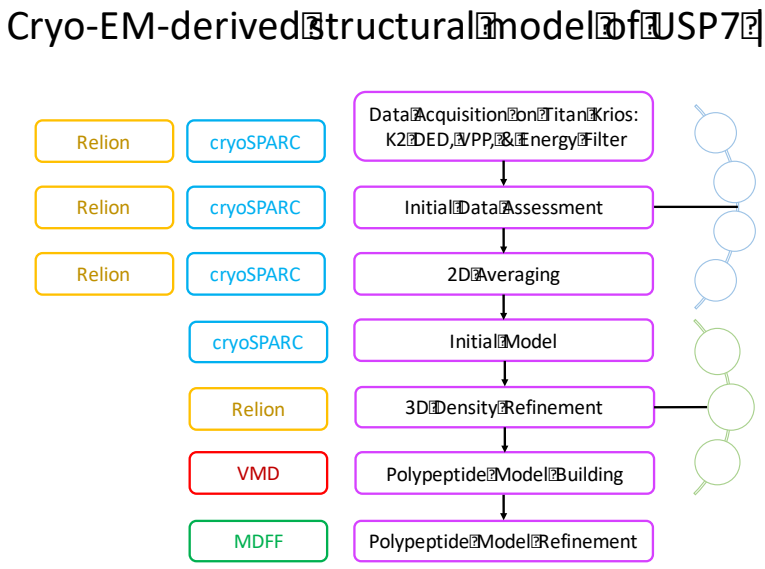


Figure E.1. Overview of Image Processing Pipeline of USP7. (Left) Software used for each step, color-coded: orange is Relion versions 2.1.0 – 3.1.0, blue is cryoSPARCv1/v2, red is VMD and green is MDFF. The usage of VMD and MDFF was performed in collaboration with Lyman Monroe from the Kihara lab. (Middle) Major steps of the image processing pipeline color intensity increasing toward final steps performed for model described herein, with sub-steps (right) in grey. Relion and cryoSPARC were both used for posterity in comparing outputs from different software. The final pipeline that resulted in the data in Fig 1.7 – 1.10 is discussed below.

APPENDIX F. CTF ESTIMATION WITH LOW-FOCUS VPP IMAGES

CTF Estimation with VPP

Typically, the power spectrum (Fourier transform) of an electron micrograph will show the effect of the contrast transfer function (CTF). This will appear as alternating light and dark rings emanating from the center outward (Thon rings), showing the relationship between contrast and spatial frequency in the image. Aberrations in the Thon rings can provide information on parameters such as astigmatism and drift in the image. Adequate estimation of these features by the CTF estimation script (ctffind4 in this case) can improve the quality of the single particle reconstruction.

Due to the size of USP7~Ub-PA (~137 kDa), and the relatively small defocus used for imaging (± 1.5 μm), there were few Thon rings observed in the 2D power spectrum. Even after manually viewing by e2ctfit.py in Eman2, CTF estimation by gctf/ctffind4 in Relion and cryoSPARC, most images exhibited very weak signal resulting in poor CTF fitting (55-58). It has been reported that use of the Volta Phase Plate can degrade the signal in the power spectra due to the large defocus errors or "bad spots" in the VPP during data collection (59). Correspondingly, this remains to be a step that could benefit from further experimentation in the future.

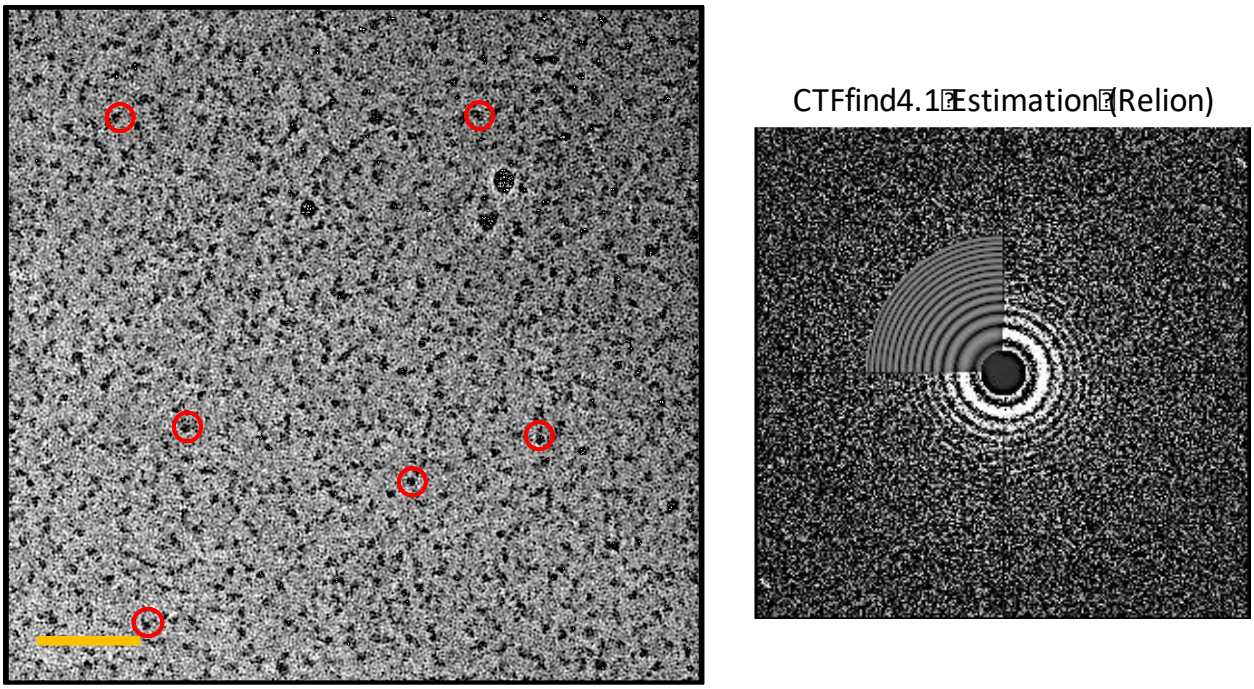


Figure F.1. USP7 images acquired with VPP on Titan Krios at 130,000x nominal mag. (Left) Representative image of USP7~Ub-PA on pyrene-graphene oxide (pGO) coated gold grids. Red circles are representative particles that would be manually selected for auto-picking templates. Orange scale bar represents ~100 nm distance. (Right) 2D power spectrum of the image to the left with CTF estimation done by CTFfind4.1 through Relion. This is an image of one of the higher signal power spectra that was maintained for reconstruction. 2D averaging illustrates significant domain movement in USP7 EP2 state.

APPENDIX G. PROTEOPLEX ASSAY FOR USP28 SAMPLE OPTIMIZATION

ProteoPlex assay results for USP28

The optimal condition for USP28 was Bicine buffer pH 7.8-8.8, 180-220 mM NaCl, and nearly all per cents of glycerol. The addition of polyols at the concentration range tested did not greatly affect the melting temperature. Thus, it was concluded that an acceptable final purification buffer for cryo-EM would be: 25 mM Bicine pH 8.4, 200 mM NaCl, 1 mM TCEP.

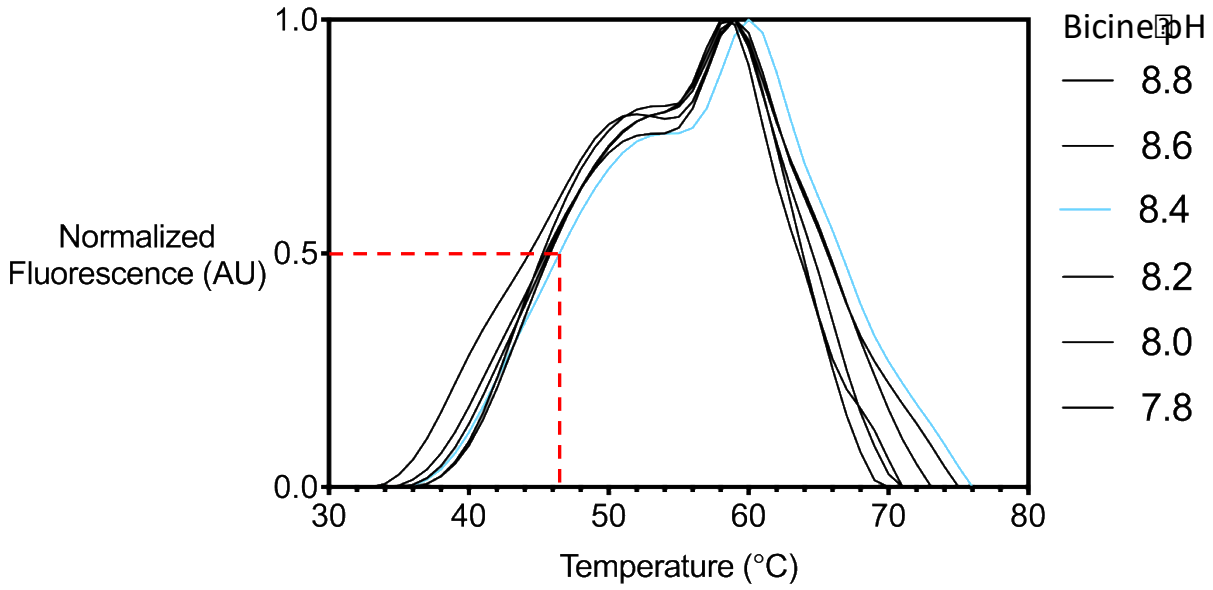


Figure G.1 ProteoPlex thermal stability assay for USP28. Above is an example condition series of USP28 for the highest stability buffer, Bicine. All curves are in black, with the highest melting temp, Bicine pH 8.4, highlighted in blue. The range of melting temps was approx. 3-5˚C for all buffers tested. The red dashed line represents the half-maximal fluorescence, which is observed as the melting temperature. Optimal condition for USP28 above is Bicine pH 8.4, corresponding to a T_m = 47.5˚C. Each curve represents the mean of three replicates. It is important to note the bimodal melting curve. It is thought the first is due to dimer dissociation, with full denaturation resulting in the higher amplitude peak.