# Insight into the biochemical interaction of the USP7 C-terminal domain

**Ira Kay Logatoc Lacdao**

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

**MASTER OF SCIENCE**

GRADUATE PROGRAM IN BIOLOGY
YORK UNIVERISTY
TORONTO, ONTARIO

December, 2013

## ABSTRACT

Ubiquitination is a post-translational modification, involving the covalent attachment of ubiquitin, important for the regulation of cellular pathways. Deubiquitinases reverse this process through the hydrolysis of ubiquitin moieties from substrates.

The deubiquitinase, Ubiquitin Specific Protease 7 (USP7), regulates many important nuclear processes. The USP7 N-terminus (USP7-NTD) is a TRAF domain which functions predominantly in substrate-binding while the USP7 C-terminus (USP7-CTD) was predicted to have 4-5 ubiquitin like domains (Ubl1-5). The aims of this thesis are biochemical and structural analyses of the C-terminus of human USP7. Various Ubl domains of human USP7-CTD were cloned, expressed, purified and crystallized. Subsequently, the biochemical interaction of an ICP0 peptide with these Ubl domains was investigated using *in vitro* GST-pulldown experiments and fluorescence polarization assays where the ICP0 peptide showed tightest interaction with Ubl123. Co-crystals of Ubl123 and ICP0 peptide need further refinement prior to structure determination.

## ACKNOWLEDGMENTS

I would to express my gratitude to Dr. Saridakis for accommodating and giving me this opportunity to work in her lab. When I started my research, I never really knew what to expect going into protein crystallography. A few of the things I learned from Dr, Saridakis was to work hard, persevere and a dash of luck does not hurt either. I thank her for her guidance and support throughout this journey. Indeed, it has been a long drawn out journey and so I definitely thank Dr. Saridakis for her patience and tolerance.

I would like to thank the members of my defense committee for their acceptance and their time of being in the committee. As my advisor, I would like to thank Dr. Sheng for her advice and her unique point of views on my project. I also would like to thank Dr. Wu for being an influential person during my undergraduate years and I was able to carry the lessons I learned from her throughout my graduate years.

What makes every journey interesting and unique are the connections and bonds we make along the way. For my colleagues, thank you for all your support and encouragement. Our endless discussions not only about our research but of life's lessons as well, helped me grow along the way.

Lastly, I would like to thank my family for their support throughout the years.

# TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

AA Amino acids

| | | |
|---|---|---|
| A | Ala | Alanine |
| C | Cys | Cysteine |
| D | Asp | Aspartic acid / Aspartate |
| E | Glu | Glutamic acid / Glutamate |
| F | Phe | Phenylalanine |
| G | Gly | Glycine |
| H | His | Histadine |
| I | Ile | Isoleucine |
| K | Lys | Lysine |
| L | Leu | Leucine |
| M | Met | Methionine |
| N | Asn | Asparagine |
| P | Pro | Proline |
| Q | Gln | Glutamine |
| R | Arg | Arginine |
| S | Ser | Serine |
| T | Thr | Threonine |
| V | Val | Valine |
| W | Trp | Tryptophan |
| Y | Tyr | Tyrosine |

| | |
|---|---|
| CT | Full-length USP7 C-terminal |
| CHFR | Checkpoint with forkhead and ring finger domains |
| Daxx | Death domain associated protein |
| E1 | Ubiquitin activating enzyme |
| E2 | Ubiquitin conjugating enzyme |
| E3 | Ubiquitin ligase |
| EBNA1 | Epstein-Barr nuclear antigen 1 |
| EBV | Epstein-Barr virus |
| FOXO4 | Forkhead box O 4 |
| GMPS | Guanosine monophosphate synthase |
| HAUSP | Herpesvirus-associated ubiquitin-specific protease |
| Hdm2 | Human double minute 2 |
| Hdmx | Human double minute X |
| HSV | Herpes simplex virus |
| HSV-1 | Herpes simplex virus type 1 |
| ICP0 | Infected cell protein 0 |
| IE | Immediate early |
| IKKγ | Inhibitor of nuclear factor kappa-B kinase subunit gamma |
| IP | Immunoprecipitation |
| kDa | killodalton |
| KSHV | Kaposi's sarcoma–associated herpesvirus |
| LANA | Latency-associated nuclear antigen |
| MCM-BP | Minichromosome maintenance complex binding protein |
| Mdm2 | Murine double minute 2 |
| MdmX | Murine double minute X |

| | |
|---|---|
| ND10 | Nuclear domain 10 |
| NLS | Nuclear localization signal |
| PML | Promyelocytic leukemia |
| PTEN | Phosphatase and tensin homologue |
| RING | Really Interesting New Gene |
| Ub | Ubiquitin |
| UbE2E1 | Ubiquitin-conjugating enzyme E2E1 |
| Ubl | Ubiquitin-like |
| UBP | USP yeast homologue |
| USP | Ubiquitin specific protease |
| USP7 | Ubiquitin specific protease 7 |
| USP7-CAT | USP7 catalytic domain |
| USP7-NTD | USP7 N-terminal domain |
| USP7-CTD | USP7 C-terminal domain |
| vIRF4 | viral interferon (IFN) regulatory factor 4 |
| TRAF | Tumor necrosis factor-receptor associated factor |

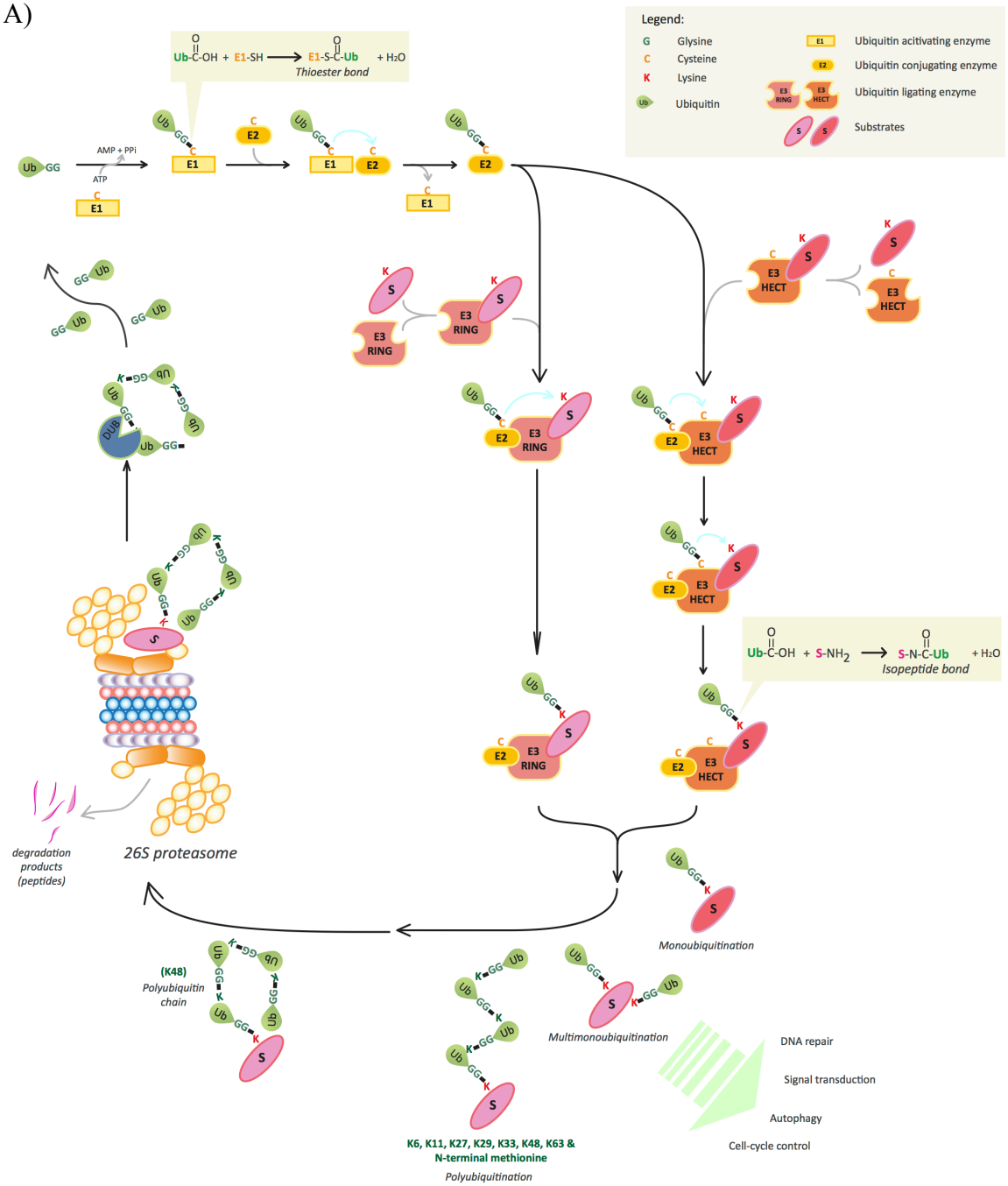
# CHAPTER 1: INTRODUCTION

Proteins undergo post-translational modification including phosphorylation, glycosylation and ubiquitination – each serving a different purpose in the cell. Ubiquitination is the covalent attachment, of a small protein, ubiquitin, to target proteins through a series of enzymatic reactions. The conjugation of ubiquitin allows the regulation of various cellular processes: proteasome-dependent protein degradation, endocytosis, membrane protein trafficking, signal transduction, immune response, nuclear transport and transcription [1-2].

## 1.1 Ubiquitination

Ubiquitin (Ub) is a small protein consisting of 76 amino acids and is about 8.5 kDa. It is highly evolutionarily conserved in eukaryotes. When ubiquitin is expressed in the cell, it usually exists as either a fusion protein with ribosomal subunits or as a polyubiquitin chain processed by proteases to generate ubiquitin monomers [3-4]. Ubiquitination is a multistep process involving an enzymatic cascade of which ubiquitin is covalently bound to the target protein (Figure 1-1A) [2]. First, ubiquitin activating enzyme (E1), in an ATP-dependent manner, activates and conjugates a Ub monomer to its active site cysteine forming a thiol-ester bond with the carboxy-terminal glycine of Ub [1,5-6]. Then through trans-thiolation, Ub from E1 is transferred into the active site cysteine of ubiquitin conjugating enzyme (E2) [5,7]. Finally, Ub-charged E2 transfers the ubiquitin to the substrate protein via a ubiquitin ligase (E3) [1]. Ubiquitin is covalently bound to the

A)


Legend:
G Glysine
C Cysteine
K Lysine
Ub Ubiquitin
E1 Ubiquitin acitivating enzyme
E2 Ubiquitin conjugating enzyme
E3 RING E3 HECT Ubiquitin ligating enzyme
S Substrates
Ub-C-OH + E1-SH → E1-S-C-Ub + H2O
Thioester bond
AMP + PPi
ATP
E3 RING
E3 HECT
DUB
Ub-C-OH + S-NH2 → S-N-C-Ub + H2O
Isopeptide bond
26S proteasome
degradation products (peptides)
Monoubiquitination
(K48) Polyubiquitin chain
Multimonoubiquitination
K6, K11, K27, K29, K33, K48, K63 & N-terminal methionine
Polyubiquitination
DNA repair
Signal transduction
Autophagy
Cell-cycle control

B)

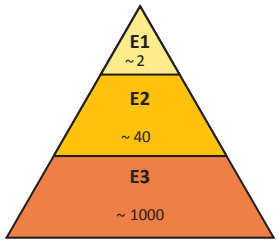

E1
~ 2
E2
~ 40
E3
~ 1000

Figure 1-1: Ubiquitination and the Ubiquitin proteasome pathway (UPS). A) Proteins get ubiquitinated through an enzymatic cascade. In an ATP-dependent manner, ubiquitin activating enzymes (E1) activates and conjugates ubiquitin to its active site cysteine forming a thiol-ester bond. Ubiquitin is then transferred from E1 to ubiquitin conjugating enzymes (E2) active site cysteine. From E2s, depending on the type of ubiquitin ligase (E3) either a HECT domain E3 or RING domain E3, ubiquitin can be directly or indirectly conjugated to the substrate, respectively. Ubiquitin's C-terminal glycine is covalently bound to lysine residues on the substrate. Proteins can be monoubiquitinated, multimonoubiquitinated and polyubiquitinated. Polyubiquitin chains are formed by subsequent addition of ubiquitin to a lysine on substrate-anchored ubiquitin. Ubiquitin has 7 lysines for polyubiquitin chain formation: K6, K11, K27, K29, K33 K48, K63. Linear ubiquitin chains can also be formed between the C-terminal glycine and N-terminal methionine (M1) of ubiquitin. The degree and type of ubiquitin conjugation dictates the fate and function of the substrate. For example, K48-polyubiquitin targets proteins for proteasomal degradation. Ubiquitin chains are spared from degradation and ubiquitins are recycled again into free ubiquitins by deubiquitinating enzymes (DUBs). B) In the cell, there are 2 E1s responsible for ubiquitin activation and the number of E3s (~1000) outnumbers E2s (~40). The diverse number of E3s provides the specificity needed to target diverse substrates for ubiquitination. Adapted from Kerscher, Felberbaum and Hochstrasser (2006); Hicke, Schubert and Hill (2005); Ye and Rape (2009).

substrate through an isopeptide bond linkage between the C-terminal glycine residue of ubiquitin and the epsilon-$NH_2$ (amino group) of a lysine residue in the substrate [1-2]. Linear ubiquitin chains can also be formed between the C-terminal glycine residue of ubiquitin and N-terminal methionine (M1) of ubiquitin [8]. Substrates can be monoubiquitinated, multimonoubiquitinated, or polyubiquitinated.

The number of human E1, E2 and E3 is best presented by a pyramid scheme (Figure 1-1B). E1, the ubiquitin activating enzyme, is at the top of the ubiquitination pathway [9]. The human genome encodes 2 E1s that are responsible for ubiquitin activation in the entire pathway [9]. On the second layer of the pyramid, activated ubiquitin is transferred to one of several E2 ubiquitin conjugating enzymes. There are almost ~ 40 E2s in the human genome [8,10]. Lastly, the bottom layer of the pyramid represents multiple E3s that the E2s cater to. E3 ubiquitin ligases are a diverse set of enzymes thus giving specificity as to which target protein is ubiquitinated [11-12]. There are 3 main classes of E3s: HECT, RING and U-box. HECT (homologous to E6-AP carboxyl terminus) E3s are able to accept ubiquitin from E2 to form a thiol-ester bond at their catalytic cysteines prior to ubiquitin transfer [13]. The HECT domain is about 350 residues and is highly conserved [13]. RING (Really Interesting New Gene) E3s bind to E2s and mediate the transfer of ubiquitin directly from E2 to the substrate [14]. The RING domain has a consensus sequence $CX2CX_{[9-39]}CX_{[1-3]}HX_{[2-3]}C/HX2CX_{[4-48]}CX2C$ (C: Cysteine, H: Histidine, X: any amino acid) [14]. These cysteines and histidines are able to coordinate two zinc ions [15]. U-box E3s function like RING E3s and have a RING-like fold [16]. However, they lack the cysteine and histidine Zn-coordination sites [16]. Between the E2s and E3s,

ubiquitination is able to generate specificity via E3's substrate recognition and the potential E2-E3 complex combinations that mediate ubiquitination [6].

Ubiquitination regulates cellular processes. The degree and type of ubiquitin conjugate formed dictates the fate and function of the target protein [6]. Monoubiquitination, which is the attachment of one ubiquitin to one lysine in the substrate, is involved in the regulation of transcription, signaling and translocation of the substrate [17-19]. Sometimes, ubiquitin monomers are added to multiple lysines in the substrate or multimonoubiquitination that can lead to endosome formation and degradation of the substrate in the lysosome [20]. Aside from ubiquitin's monomeric form, conjugated ubiquitins can also exist as polymeric chains. In order to form ubiquitin chains, the subsequent ubiquitin is attached to a lysine in a previously conjugated ubiquitin. There are seven lysines (K6, K11, K27, K29, K33 K48, K63) in ubiquitin that can possibly be involved in polyubiquitin chain formation (Figure 1-2B) [18]. Different polyubiquitin chain linkage leads to different consequences for the modified substrate [21]. Substrates tagged with a K48 linked ubiquitin chain are targeted for degradation by the proteasome [21]. The K48 linked ubiquitin chain is commonly referred to as the proteolytic function of ubiquitin while monoubiquitination, multimonoubiquitination and the other types of polyubiquitin linkages are referred to as the non-proteolytic function of ubiquitin. The K29 linked chain leads to degradation via the lysosome [22] while the K63 linked chain is involved in cellular processes such as DNA damage repair pathway and signal transduction [23]. It is also possible for mixed linkages but their existence and significance is currently being studied [24-25].

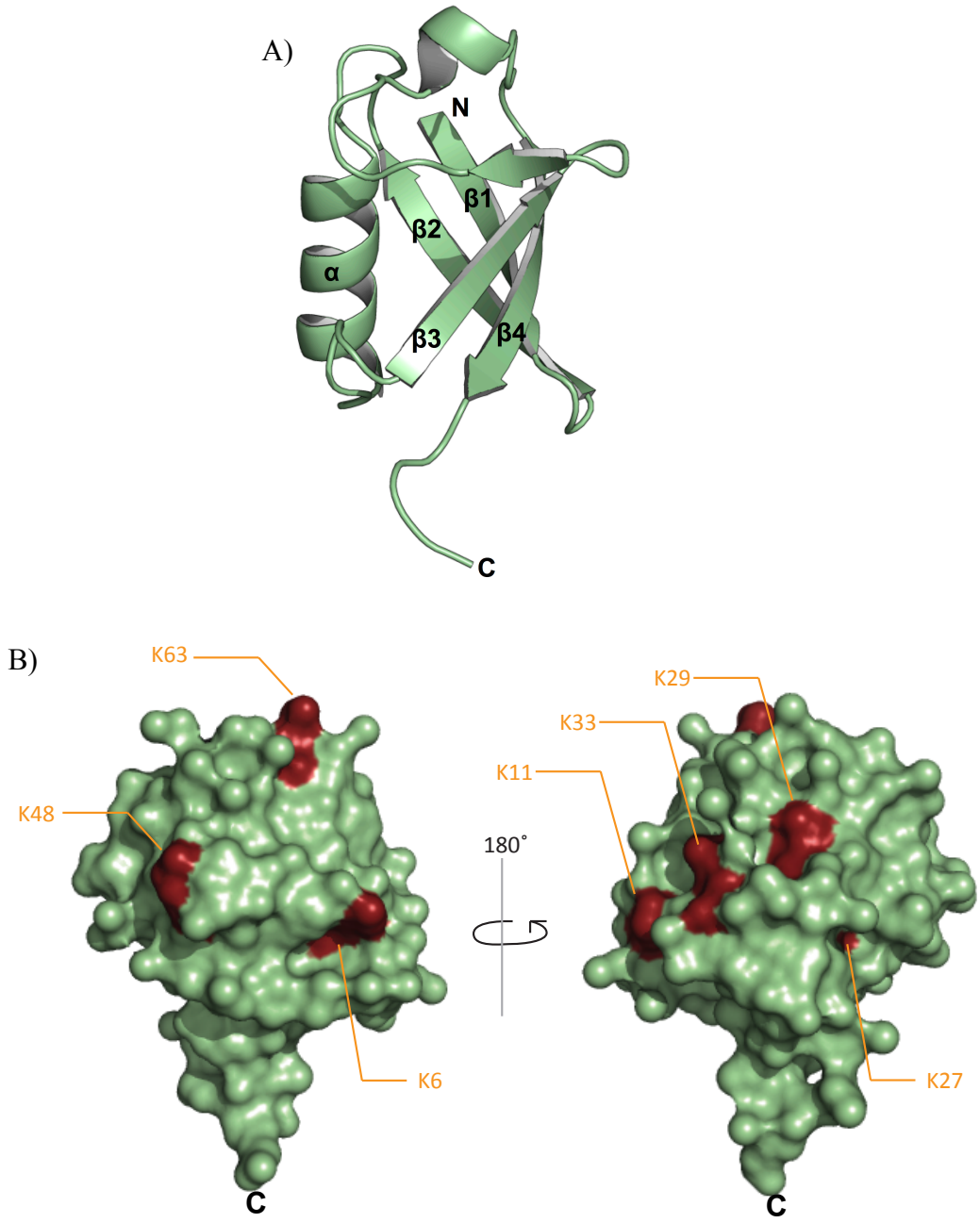



Figure 1-2: Molecular structure of ubiquitin. A) Ubiquitin has a characteristic ββαββ fold shown as ribbon diagram. B) Ubiquitin, shown as surface, has 7 lysine residues (K6, K11, K27, K29, K33, K48, K63) where potential polyubiquitin chains could be formed. In polyubiquitin chain formation, the C-terminal tail of subsequent ubiquitin is conjugated to the lysine of the previous ubiquitin. Rendered with Pymol. PDB: 1UBQ (Ubiquitin).

In any cellular processes, a particular signal can be attenuated or amplified by changing the protein concentration in the cell, which can be done by controlling the transcription of the gene, translation of mRNA and degradation of the protein itself. One of the ways to degrade proteins is through ubiquitin-mediated proteasomal degradation, also known as the ubiquitin-proteasome system (UPS). The UPS is important for the degradation of cytoplasmic and nuclear proteins [26]. As mentioned earlier, K48-tagged proteins are shuttled to the proteasome for destruction. The ability of the proteasome to recognize these ubiquitinated proteins is crucial in allowing for both selectivity and specificity [10].

Ubiquitin belongs to the ubiquitin superfamily of proteins. Even with varying sequence homology, the same 3-D structure, a unique fold called the ββαββ fold or β-grasp fold, is conserved amongst all family members [27]. Other small proteins that possess this fold are called ubiquitin-like proteins (Ubls) [27] (Figure 1-3). Like ubiquitin, Ubls are conjugated in a similar fashion with different Ubls having their own E1-E2-E3 enzymatic cascade [9]. In addition, Ubls are also involved in various cellular processes, some of which include cell cycle progression and immune response [28]. Some of the well-studied Ubls are neuronal- precursor-cell-expressed developmentally downregulated protein-8 (NEDD8) and small ubiquitin-related modifier (SUMO)-family (SUMO-1,-2,-3) [17,28].

## 1.2 Deubiquitination

Deubiquitinating enzymes (DUBs) are proteases, which process ubiquitin precursors, remove monoubiquitin and edit polyubiquitin chains through their

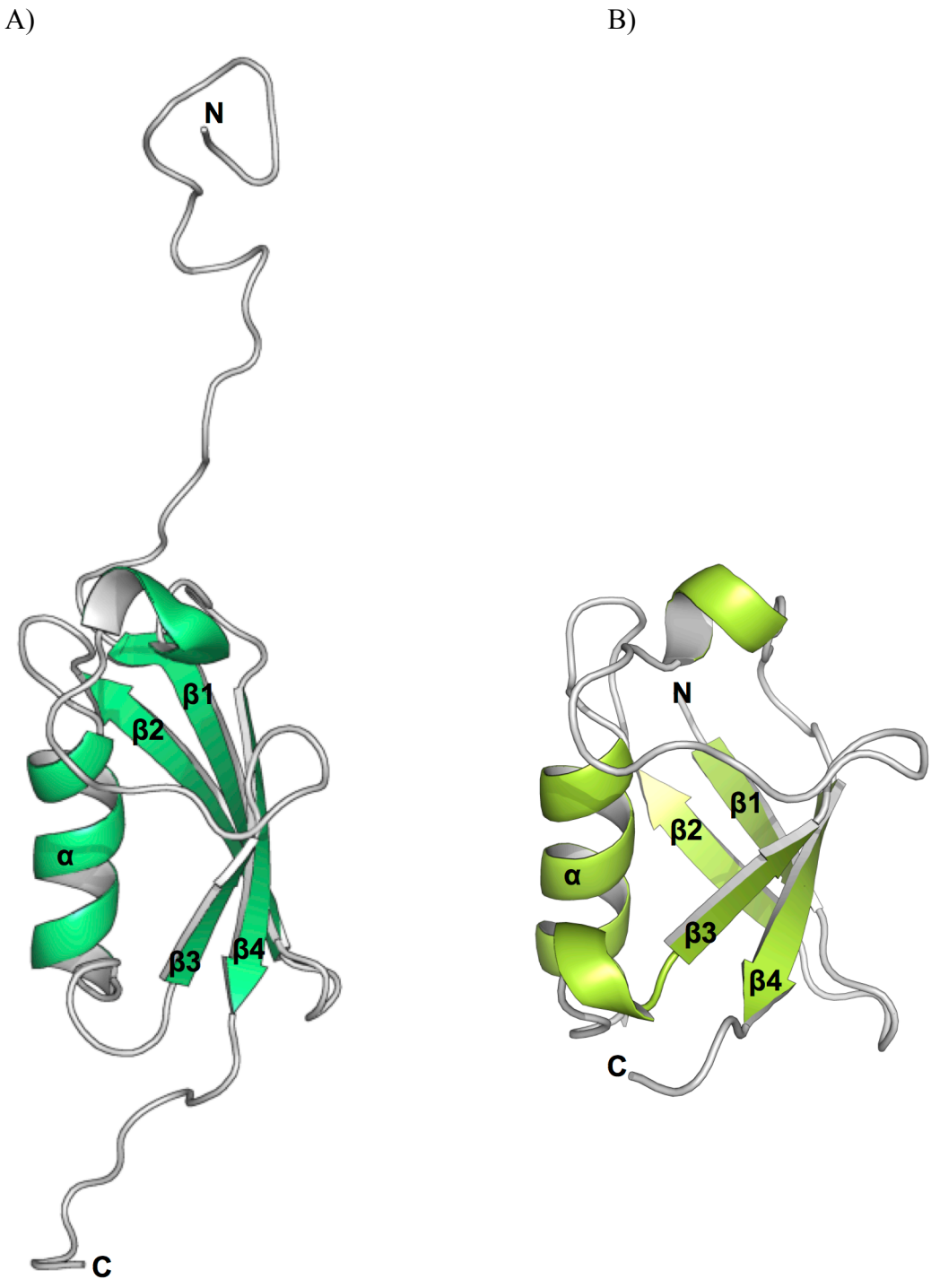


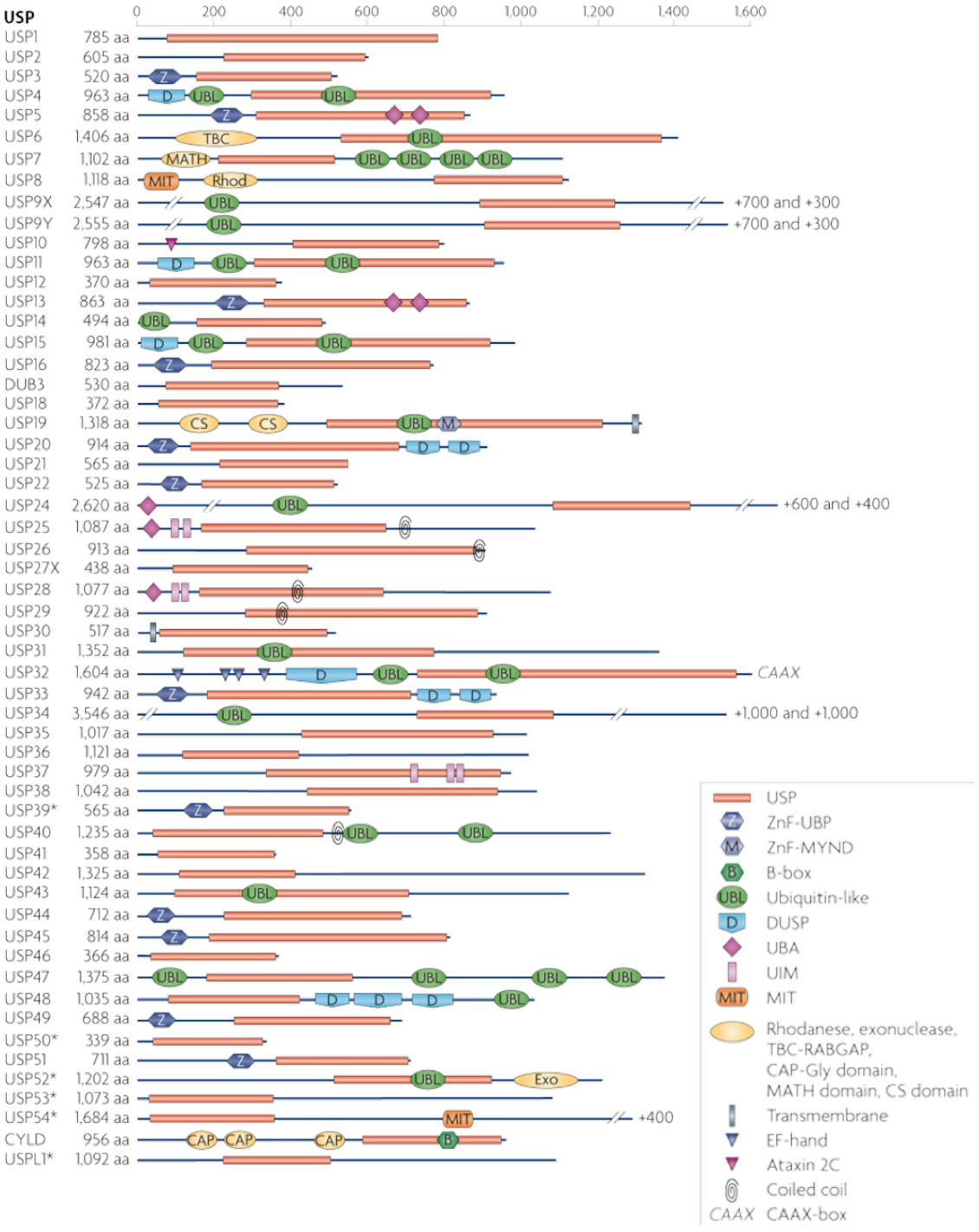

Figure 1-3: Ubiquitin-like proteins. A) SUMO-1 and B) NEDD8 have the same ββαββ fold characteristic to ubiqutin. PDB: 1NDD (NEDD8), 1A5R (SUMO-1). Both are shown as ribbon diagrams and rendered with Pymol.

isopeptidase activity [29]. There are approximately 100 DUB genes in the human genome [29], where 90 are cysteine proteases and 12 are metalloproteases [30]. There are 4 structural classes/families of 'papain-like' cysteine proteases: ubiquitin specific protease (USP or UBP in yeast), ubiquitin C-terminal hydrolases (UCH), ovarian tumor (OTU), and Josephin domain (MJD) [29-30]. The last structural class/family is a "zinc-dependent metalloprotease", JAB1/MPN/Mov34 (JAMM). These 5 classes are ubiquitin specific DUBs (Figure 1-4). Some DUBs prefer to process ubiquitin-like proteins [31]. These cysteine proteases belong to the Adenain family [31].

In the ubiquitin-proteasome pathway, ubiquitin is expressed as a precursor protein and is processed by DUBs to expose the C-terminal glycine [4,32] (Figure 1-5). DUBS can be found associated with the proteasome to rescue improperly ubiquitinated substrates and also remove ubiquitin chains to prevent the degradation of ubiquitin. These unanchored polyubiquitin chains are then processed by DUBs replenishing the free ubiquitin pool in the cell [33-34]. DUBs are also able to reverse the modification on ubiquitinated substrates called chain editing by removal of monoubiquitin and polyubiquitin.

Deubiquitination is highly regulated with various roles in cell cycle control, protein degradation, gene expression, DNA damage repair, bacterial and viral infection [29]. There has been an emerging trend in the involvement of DUBs as regulators of signaling pathways, chromatin structure, endocytosis, and apoptosis [18]. DUBs are important for neuronal function, development and immunity [35]. DUBs have been linked to pathogens

A)


USP
0
200
400
600
800
1,000
1,200
1,400
1,600
USP1 785 aa
USP2 605 aa
USP3 520 aa
USP4 963 aa
USP5 858 aa
USP6 1,406 aa
USP7 1,102 aa
USP8 1,118 aa
USP9X 2,547 aa
USP9Y 2,555 aa
USP10 798 aa
USP11 963 aa
USP12 370 aa
USP13 863 aa
USP14 494 aa
USP15 981 aa
USP16 823 aa
DUB3 530 aa
USP18 372 aa
USP19 1,318 aa
USP20 914 aa
USP21 565 aa
USP22 525 aa
USP24 2,620 aa
USP25 1,087 aa
USP26 913 aa
USP27X 438 aa
USP28 1,077 aa
USP29 922 aa
USP30 517 aa
USP31 1,352 aa
USP32 1,604 aa
USP33 942 aa
USP34 3,546 aa
USP35 1,017 aa
USP36 1,121 aa
USP37 979 aa
USP38 1,042 aa
USP39* 565 aa
USP40 1,235 aa
USP41 358 aa
USP42 1,325 aa
USP43 1,124 aa
USP44 712 aa
USP45 814 aa
USP46 366 aa
USP47 1,375 aa
USP48 1,035 aa
USP49 688 aa
USP50* 339 aa
USP51 711 aa
USP52* 1,202 aa
USP53* 1,073 aa
USP54* 1,684 aa
CYLD 956 aa
USPL1* 1,092 aa
+700 and +300
+700 and +300
+600 and +400
CAAX
+1,000 and +1,000
+400
USP
ZnF-UBP
ZnF-MYND
B-box
Ubiquitin-like
DUSP
UBA
UIM
MIT
Rhodanese, exonuclease, TBC-RABGAP, CAP-Gly domain, MATH domain, CS domain
Transmembrane
EF-hand
Ataxin 2C
Coiled coil
CAAX-box

B)

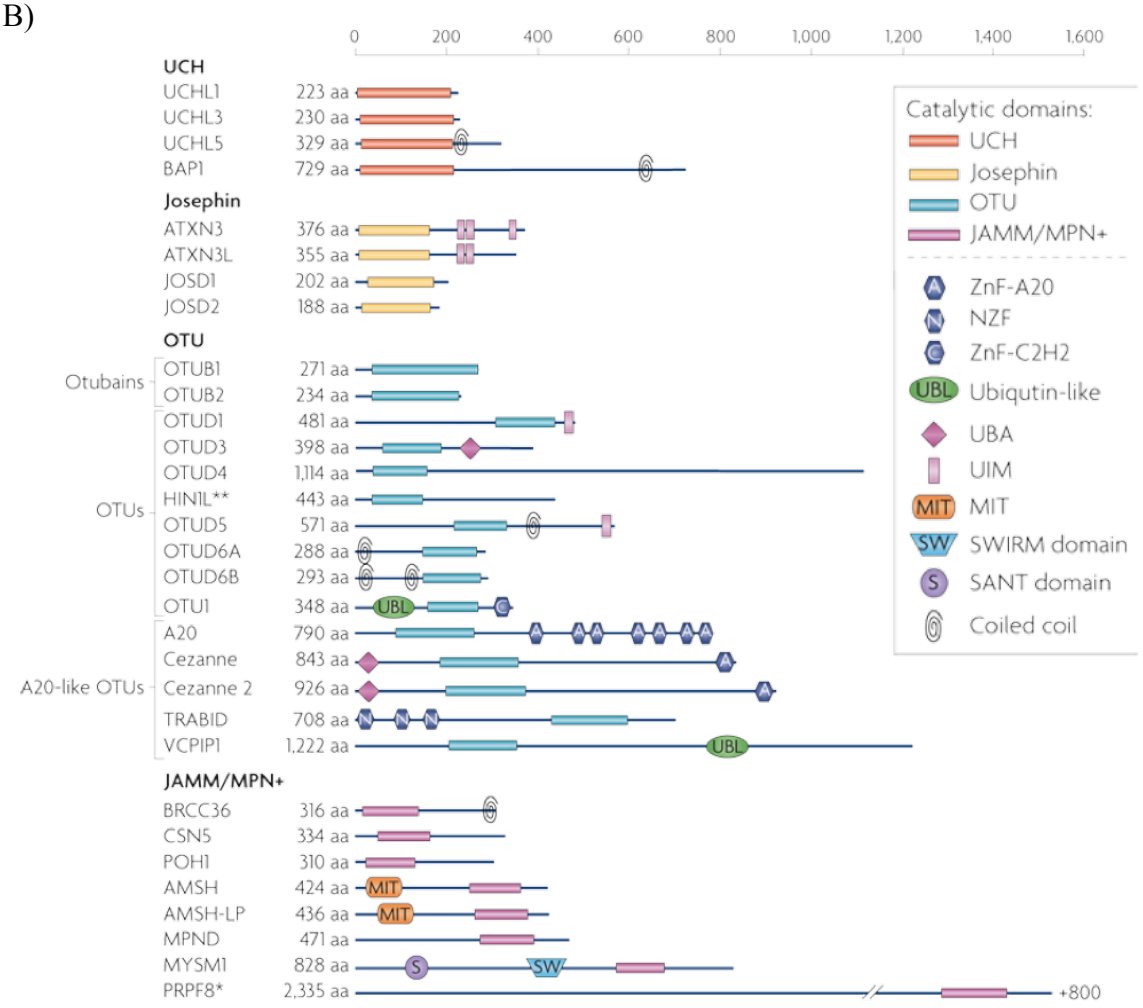



Figure 1-4: Families of deubiquitinating enzymes and common domains. A-B) For each DUB family (USP, UCH, OUT, Josephin and JAMM), the catalytic domain is conserved within the family and N-terminal/C-terminal domains have/contain a variety of domain folds. From Komander, Clague, and Urbé (2009).

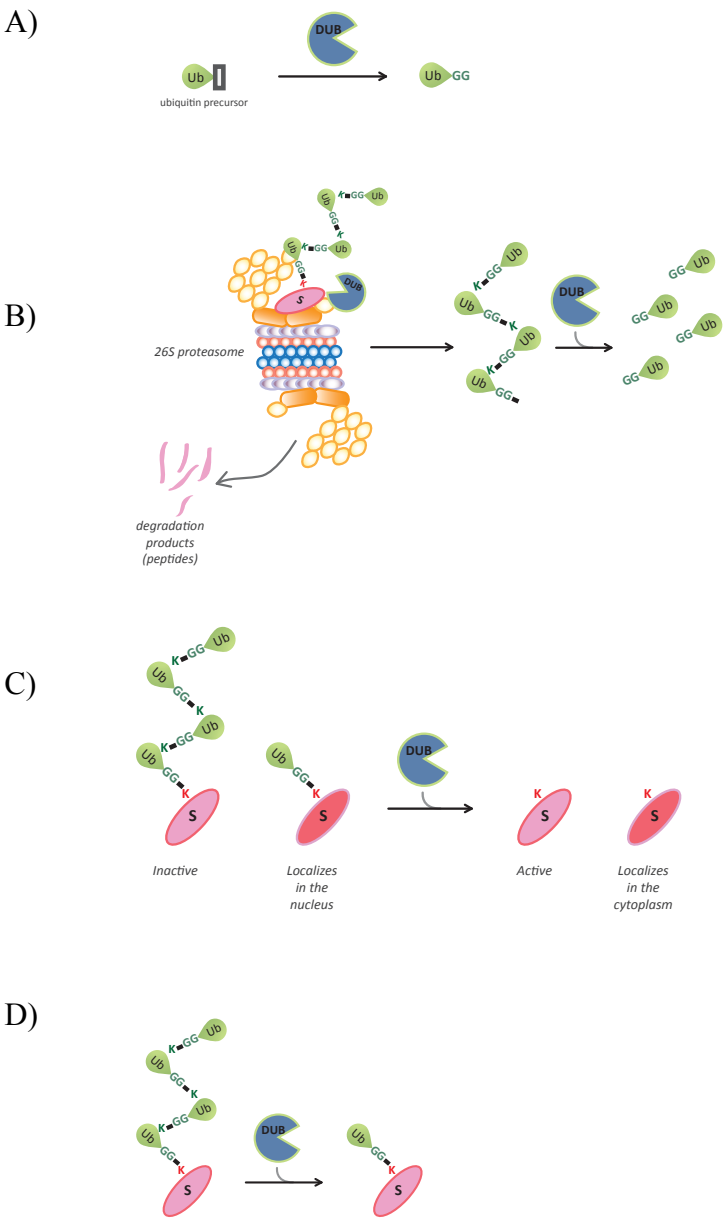



Figure 1-5: Deubiquitinating enzyme (DUB) activities. A) Ubiquitin is expressed as precursor proteins that need to be modified in order to be available to the cell. Ubiquitin precursors are processed by DUBs into free ubiquitin. B) DUBs remove ubiquitin chains from substrate destined for proteasome degradation. These ubiquitin chains are then converted to free ubiquitin. C) DUBs can regulate substrate activity and localization via ubiquitin hydrolysis. D) DUBs can also edit polyubiquitin chains, altering the fate of the substrate. Adapted from Wilkinson (1997).

where bacterial and viral proteins disrupt ubiquitination and deubiquitination in host cells during infection while mutations in some DUBs have been linked to cancers and neurologic diseases [2, 36]. Thus, DUBS are excellent candidates for the development of therapeutic drugs.

### 1.3 Ubiquitin specific protease 7 (USP7)

USP is the largest family of cysteine proteases. There are about 50 USPs in humans and about 16 UBPs (USP yeast homologue) [30]. The USP catalytic domain's fold is highly conserved (Figure1-4). Between different USPs, molecular weights vary from ~37-390 kDa. While the catalytic domain of USPs usually ranges from ~300-800 amino acids or ~33-88 kDa, the variation in molecular weight is a reflection of the diverse sequences N- or C-terminus of the catalytic domain (Figure 1-4) [30]. Structure prediction of these N-/C-terminal flanking sequences showed an abundance of ubiquitin-binding domains and a multitude of ubiquitin-like folds (Ubl) that probably play roles in protein-protein interaction, substrate specificity and enzymatic activity (Figure 1-4) [37,38].

USP7 is evolutionarily conserved in mammals with orthologs in *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*. USP7 is 1102 amino acids long with a molecular weight of 135 kDa. It has been found as a dimer in cells [39]. Some USP7 is closely associated with Nuclear Domain 10 (ND10) structures in the nucleus, where it localizes [40]. USP7 undergoes post-translation modification where S18 and S963 get phosphorylated and K869 gets ubiquitinated [41].

### 1.3.1 USP7 Interacting Proteins

Ubiquitin specific protease 7 (USP7), also known as herpesvirus-associated ubiquitin-specific protease (HAUSP), was first discovered as a cellular protein that bound to Herpes simplex virus type 1 (HSV-1) protein Infected Cell Protein 0 (ICP0) [40,42]. Since then, multiple proteins have been identified as substrates or binding partners including the well-known tumor suppressor, p53 [43], involving USP7 in cell survival pathways. Aside from USP7's ability to deubiquitinate p53, USP7 is also able to deubiquitinate and stabilize E3 ligase Human double minute 2 (Hdm2, human ortholog of Murine double minute 2 (Mdm2)), a negative regulator of p53 [38]. Similarly, MdmX, an Mdm2 homolog and p53 regulator, is stabilized via USP7 deubiquitination [44-45]. These interactions provide an overview of the complex role USP7 plays in the p53 pathway. Additional tumor suppressors now known as USP7 substrates include PTEN [46] and FOXO4 [47]. USP7 also plays a roles in cell cycle control particularly DNA damage response through interaction with Claspin [48] and Chfr [49]. Several other cellular proteins that interact with USP7 include GMPS [50-51], Ataxin-1 [52], histone H2B [50-51], and MARCH7 [53]. The variety of USP7 targets certainly gives us a peek into the diverse roles DUBs play in the cell making DUBs a prime target for viruses. Another herpes virus that targets USP7, aside from HSV-1, is the Epstein-Barr virus (EBV) whose viral protein Epstein-Barr Nuclear Antigen 1 (EBNA1) binds to the same domain of USP7 as p53 [54]. Structural domains within USP7 could be one of the contributing factors that dictate specificity for these targets. However, little

USP structural information is available particularly the substrate binding domains of USPs.

### 1.3.2 USP7 Domains

USP7 has 3 major domains (Figure 1-6): TRAF (Tumor necrosis factor-receptor associated factor) domain in the N-terminus (USP7-NTD), USP catalytic domain (USP7-CAT), and the C-terminus (USP7-CTD).

#### 1.3.2.1 USP7 N-terminal domain (USP7-NTD)

The USP7 catalytic domain is flanked between a N- and C-terminal domains (Figure 1-6). The N-terminal domain is from residues 1-207. Amino acids 1-62 are unstructured while the crystal structure of amino acids 63-205 shows two overlapping β-sheets called a Tumor necrosis factor-receptor associated factor (TRAF) domain (Figure 1-7A) [54]. Several co-crystal structures of USP7-NTD with peptides from substrate and interaction proteins have been determined: p53, Hdm2 (human ortholog of mouse Mdm2), HdmX (human ortholog of mouse MdmX), EBNA1, ubiquitin-conjugating enzyme E2E1 (UbE2E1) and Minichromosome maintenance complex binding protein (MCM-BP) (a few are shown in Figure 1-7B). The USP7-NTD TRAF domain preferentially interacts with peptides containing the P/AXXS (P: Proline, A: Alanine, X: any amino acid, S: Serine) motif for protein-protein interaction [52b] (Figure 1-7C). Interaction of the peptides' Serine with Aspartate (D164) in the binding pocket of USP7 was found to be critical for binding as

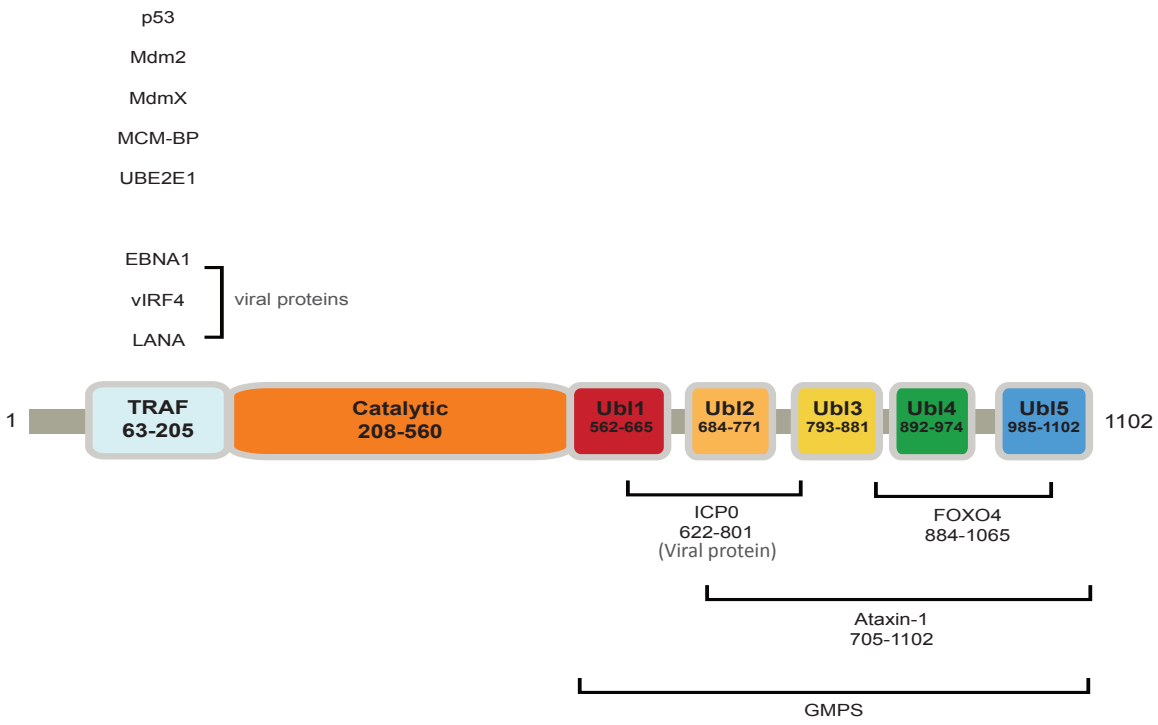



Figure 1-6: USP7 primary sequence map. USP7 TRAF domain in the N-terminus is from amino acids 63-205. Some USP7 N-terminal binding proteins and substrates include p53, MdmX, Mdm2, and EBNA1. The TRAF domain also interacts with some viral proteins such as EBNA1, vIRF4, and LANA. Amino acid residues 208-560 encodes for the USP7 catalytic domain. The USP7 C-terminal amino acids 560-1102 is predicted to have 5 ubiquitin-like (Ubl) domains each with a ββαββ fold: Ubl1 (residues 562-665), Ubl2 (residues 684-771), Ubl3 (residues 793-881), Ubl4 (residues 892-974), Ubl5 (985-1102). USP7 C-terminal binding proteins include GMPS, ATAXIN-1 (705-1102), ICP0 (622-801) and FOXO4 (884-1065).

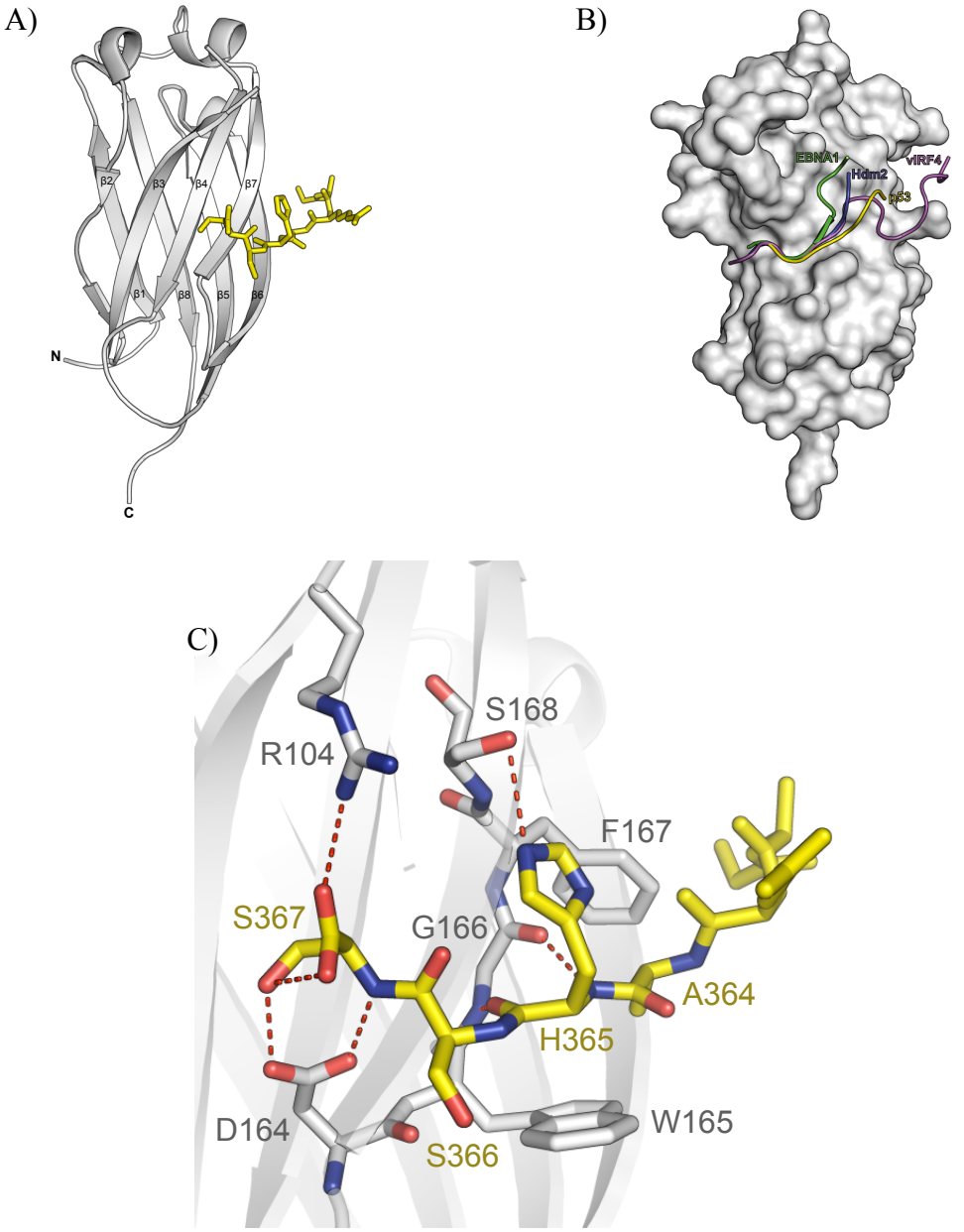



Figure 1-7: USP7 N-terminal Tumor necrosis factor-receptor associated factor (TRAF) domain. A) USP7 TRAF domain (63-205) β-sandwich (white ribbon diagram) with p53 361-367 peptide (yellow stick). B) Overlap of some of the known peptide containing the P/AXXS motif that interacts with the same binding pocket at the USP7 N-terminus. USP7 TRAF domain is shown in surface (white) and interacting peptides as ribbon diagrams. C) USP7 TRAF domain in white stick interacting with p53 peptide $^{364}$AHSS$^{367}$ in yellow stick. S366 is essential for binding. Hydrogen bonding was shown as a red dotted line. Oxygen was colored red and nitrogen was colored blue. PDB: 3MQS (USP7:Hdm2 395-401), 2FOJ (USP7:p53 361-367), 1YY6 (USP7:EBNA1), 2XXN (USP7:vIRF4). Rendered with Pymol. Adapted from Lee *et al*. (2011).

mutation of serine to alanine abolishes the interaction [55]. In the case of USP7-EBNA1 interaction, the EBNA1 peptide bound to the TRAF domain forms a β-strand and serves as an extension of one of the β-sheets [54]. Figure 1-7C highlights the interaction of p53 peptide's $^{364}$AHSS$^{367}$ with the USP7-NTD TRAF domain. The hydroxyl group of p53 S367 side chain and amino group of p53 S367 main chain forms a hydrogen bond with USP7 D164 side chain. Similarly, p53 S367 carbonyl group interacts with USP7 R104 side chain. In addition, the amine group of H365 side chain also forms a hydrogen bond with USP7 S168 side chain. Lastly, the side chain of p53 A364 forms a van der Waals interaction with USP7 W165 side chain.

#### 1.3.2.2 USP7 Catalytic domain (USP7-CAT)

The USP7 catalytic domain was revealed to be residues 208-560 through limited proteolysis and X-ray crystallography [56]. It has the characteristic finger, palm, and thumb subdomains of USPs (Figure 1-8A). The catalytic residues C223, H464, D481 are found between the thumb and palm [56]. The co-crystal structure of the USP7 catalytic domain with ubiquitin-aldehyde revealed that interaction with ubiquitin induces a conformational change from an inactive to active state where the catalytic residues are assembled toward ubiquitin's C-terminal tail (Figure 1-8B) [56].

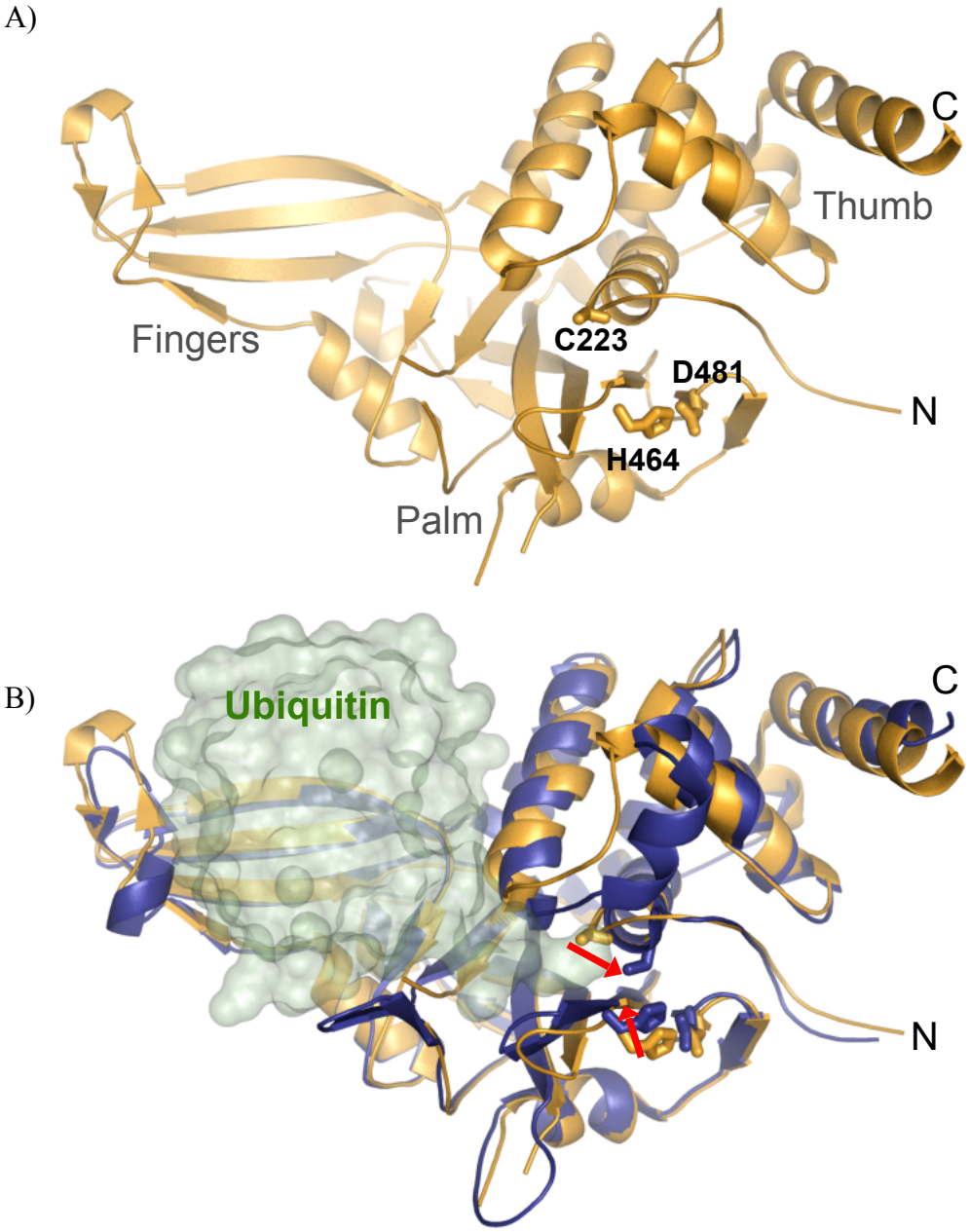



Figure 1-8: USP7 catalytic domain. A) The USP catalytic domain of USP7 (residues 208-560), shown as cartoon, has 3 sub-domains: finger, palm and thumb. The catalytic residues are C223, H464, and D481. B) When ubiquitin (pale green, surface) binds to the catalytic domain, the structure shifts (red arrow) from its inactive conformation (orange) to its active conformation (blue). The finger sub-domain, for the most part, holds ubiquitin, while ubiquitin's C-terminal tail sits between the thumb and the palm. Rendered with Pymol. PDB: 1NBF (USP7-CAT + Ub-Al, active), 1NB8 (USP7-CAT, inactive). Adapted from Hu *et al*. (2002).

#### 1.3.2.3 USP7 C-terminal domain (USP7-CTD)

As mentioned earlier, N/C-terminal domains flank the USP7 catalytic domain. The USP7 C-terminus (USP7-CTD) makes up almost 50% of the USP7 amino acid sequence, extending from amino acids 561 – 1102 (Figure 1-6). A recent structural bioinformatics analysis using consensus fold recognition showed that a number of USPs, including USP7, have ubiquitin-like (Ubl) domains (Figure 1-4A) [57]. Ubiquitin and ubiquitin-like proteins have a ββαββ fold or a β-grasp fold where the β-strands wrap around the major α-helix. USP7 was predicted to have 5 Ubl domains in its C-terminus [57] (Figure 1-6). An NMR structure of the first Ubl shows the ubiquitin-like fold (Figure 1-9). Previous studies using partial proteolysis, analyzed by SDS-PAGE and MADLI-TOF mass spectrometry, showed two major domains in the USP7 C-terminus that are resistant to proteolytic activity, amino acids 622-801 and 885-1061 [58]. After USP7 was first discovered as a cellular protein that interacts with the Herpes Simplex viral protein ICP0, later studies found that ICP0 interacts with the USP7 C-terminal region. Other USP7-CTD interacting proteins include GMPS and FOXO4.

### 1.3.3 The function of USP7 in viral interactions

The ubiqutination and deubiquitination of proteins is important for the regulation of many cellular processes. Viruses have been taking advantage of these pathways for viral entry to overcoming cellular defense mechanisms [2]. Some viruses also encode their own E3 ligases and DUBs [2].

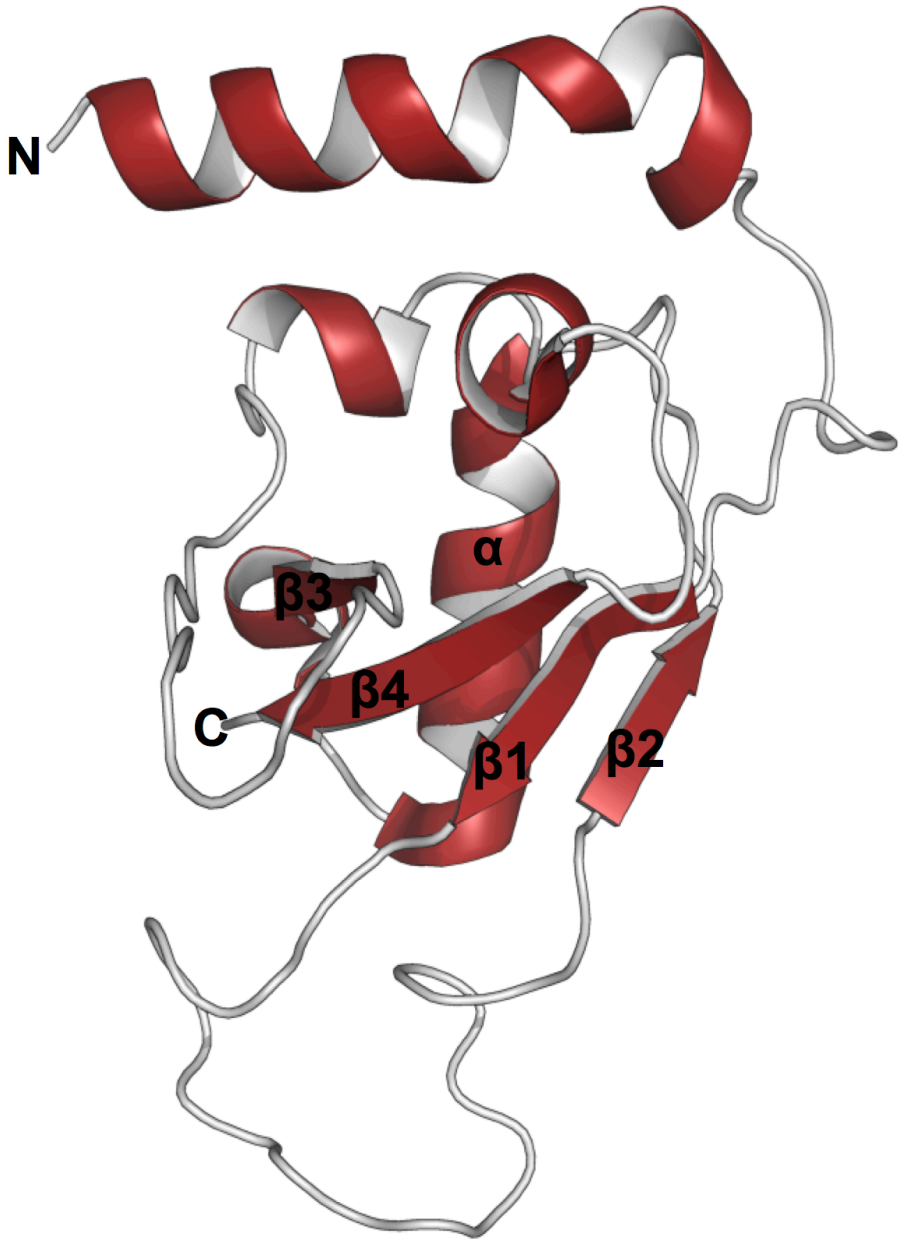



Figure 1-9: USP7 Ubl1 (residues 537-664) NMR structure. Ubl1 is in the C-terminal domain of USP7 and has the ubiquitin fold, ββαββ. Shown as ribbon diagram and rendered with Pymol. PDB: 2KVR (UBL1).

Herpesviridae virus family or herpesviruses are linear double-stranded DNA viruses whose genome is packaged in an enveloped capsid [59]. They are known for their lytic and latent viral life cycles [59]. During the initial infection of epithelial cells, the virus enters the cell by binding to cell surface receptors allowing the fusion of viral envelope with the plasma membrane and the contents of the tegument to be released into the cytoplasm [60-61]. The nucleocapsid is then transported towards the nucleus where the viral DNA is released [61]. Progeny viruses produced from successful transcription and translation of viral proteins and packing of replicated viral genome are released at the original site of infection [62]. From the epithelial cells, the virus is also able to enter sensory neurons that service these epithelial cells and establish a non-replicative or asymptomatic form in the cell nuclei called the latent cycle [63-64]. There is a lack of lytic gene expression and accumulation of latency associated transcripts during latency [63]. Latency is important for HSV life-long infection of the host and evasion of the immune system [63]. A viral strategy for establishment of viral lifelong infections is to encode viral proteins that antagonize different cellular signaling pathways, cell proliferation and apoptosis [65].

Several Herpesviruses, Kaposi's sarcoma–associated herpesvirus (KSHV), Epstein-Barr virus (EBV) and Herpes Simplex virus type 1 (HSV1), have shown to target the deubiquitinating enzyme USP7. With the multitude of roles USP7 plays in a wide variety of pathways it is not surprising that different viruses all target USP7. It is important to investigate how these viruses regulate USP7 activity and what role USP7 plays for their pathogenesis and life cycle.

#### 1.3.3.1 USP7 and EBNA1

Epstein-Barr virus (EBV) or human herpesvirus 4 (HHV-4) is a Herpesvirus that infects people worldwide and is known to predispose the host to different malignancies like Burkitt's lymphoma [2]. EBV viral protein Epstein-Barr nuclear antigen 1 (EBNA1) binds to USP7-NTD [54]. EBNA1, a latent gene product, binds to EBV viral genome and is responsible for its maintenance and replication for the life-long infection of the host [66]. The co-crystal structure of an EBNA1 peptide with USP7-NTD TRAF domain showed interaction at the same binding pocket as that of Mdm2 and p53 (Figure 1-7B). USP7 can also form a ternary complex with DNA-bound EBNA1 [51]. In comparison to p53 and Mdm2 peptides, the EBNA1 peptide bound with greater affinity to USP7-NTD [54-55]. EBNA1 is able to compete with p53 for the same binding pocket, destabilize p53 and interfere with p53-mediated apoptotic response to UV-induced DNA damage [54,58].

#### 1.3.3.2 USP7 and vIRF4

During virus infection, part of the host cell's immune response is to release interferons (IFNs), which is regulated by interferon (IFN) regulatory factors (IRFs) [67]. IRFs also regulate cell cycle arrest and apoptotic response during virus infection [67]. Kaposi's sarcoma-associated herpesvirus (KSHV) or human herpesvirus 8 (HHV-8), encodes viral homologs of IRFs (vIRFs), which are known to inhibit interferon response and p53's apoptotic activity [67-68]. vIRF4, a lytic gene product, binds to and inhibits USP7 deubiquitinating activity to regulate p53 [69]. vIRF4 residues 202-216 co-crystallized with

the USP7-TRAF domain [69]. The P/AXXS motif for this 15-mer peptide is at $^{211}$ASTS$^{214}$. The vIRF4 peptide binds to a more extensive surface of the TRAF domain, which could explain its higher binding affinity compared to other USP7 substrates [69]. Similar to EBV, KSHV utilizes pathways mediated by USP7 to progress virus lytic infection.

#### 1.3.3.3 USP7 and LANA

After the initial infection, herpesviruses are found in a repressed or latent state in the neurons where only a few genes are expressed. Periodically, the virus gets reactivated from its latent cycle to its lytic cycle at the original infection site. Latency-associated nuclear antigen 1 (LANA), a KSHV latent gene product, functions similar to EBNA1 from EBV [70]. LANA is also important for mediating genome maintenance and the replication of latent genome [71]. In addition, residues 971 to 986 of LANA also bind to USP7 TRAF domain [72]. Thus, KSHV has developed strategies to regulate USP7 activity at different stages of its life cycle; lytic cycle (vIRF4-USP7) and latent cycle (LANA-USP7).

#### 1.3.3.4 USP7 and ICP0

Herpes Simplex Virus type 1 (HSV-1) is a herpesvirus that commonly infects humans with symptoms of painful inflamed lesions around the mouth, also called cold sores [73]. Similar to other herpesviruses, it undergoes a lytic and latent cycle. Episodic reappearance of cold sores happens in response to stress and UV exposure where the lytic genes become reactivated from latent genomes and reestablish the lytic infection at these

cold sores or lytic infection sites [73]. The regulation of the switch between latent to lytic infection is poorly understood and had been an important research topic of HSV-1 biology.

As of 2013, there is no vaccine against HSV-1. Current antiviral treatments using pills and creams can target the virus during its lytic cycle that helps with healing and reducing the number of the cold sores and also their severity. One of the active compounds used in these drugs, acyclovir, helps to prevent lytic DNA replication leaving viruses that are in their latent cycle unaffected [74]. Thus, HSVs are life long infections. As details of the mechanisms behind HSV-1 life cycle and HSV-1 viral proteins' function becomes available, new opportunities for drug targets open up for future anti-viral treatments.

HSV-1, a member of Herpesviridae, has α, β, and γ gene classes [75]. The α genes or immediate early genes (IE) are the genes expressed within the first few hours of infection [76]. These IE gene products recruit cellular transcription factors to increase the transcription of IE genes and induce the activation of the β genes or early genes (E) [76]. β genes are responsible for DNA replication [77]. Early genes are also responsible for the activation of the γ genes or late genes (L) [77]. Late genes are the genes needed for successful viral packaging resulting in progeny virus release [78].

A HSV-1 immediate-early (IE) protein that has garnered a lot of interest is Infected Cell Protein 0 (ICP0) (also known as Vmw110). As an α gene product, ICP0 activates β genes and is known as a promiscuous gene transactivator as it is able to activate different viral and cellular promoters in the reporter assays [79]. ICP0 is also a part

of the tegument of viral particles and is tightly associated with the nucleocapsid [80]. The role of tegument ICP0 prior to IE gene expression is unknown [80]. However, the tegument could deliver proteins involved in virus infection initiation [80]. In early infection, ICP0 localizes to the nucleus [81] and then relocalizes to the cytoplasm during the course of infection [82].

ICP0 plays a role in the regulation of HSV-1 lytic and latent cycle [76,83]. It determines the balance between successful production and release of progeny virus during lytic replication and reactivation from latent infection [84,85]. This was known through replication phenotype characterization of ICP0-null mutants viruses [85]. These studies used mutant viruses where the ICP0 gene was removed from the HSV-1 genome and observed viral replication in infected cell lines *in vivo* [86]. ICP0 null mutant viruses had decreased viral growth [86-87]. ICP0 mutant viruses also required a high number of viral particles for an efficient productive infection [87]. However, successful infection highly decreased when the number of mutant viruses used was similar to that of wild type virus infection [88]. Thus, ICP0 is also required for productive low multiplicity of infection (m.o.i.) where there are a low number of virions added per cell during infection. Additionally, at low moi, ICP0 mutant viruses stayed in a quiescent state which can be reactivated with the expression of ICP0 [89-91]. When equivalent latent infections were established *in vivo* and in animal models, reactivation of ICP0 null-mutant was impaired [92-94]. Studying the mechanistic involvement of ICP0 during lytic and latent cycle could help with reoccurring HSV-1 infection.

ICP0 has several functional regions (Figure 1-10). It has a RING domain at the N-terminus, a nuclear localization signal (NLS), and a dimerization domain [79,95]. The RING domain has a zinc-binding RING finger motif and has E3 ligase activity [96,97]. ICP0 has a nuclear localization signal (nls) (residues 550-506) that enables it to translocate from cytosol to nucleus [98]. The cellular proteins, USP7 and ND10, also have binding sites in ICP0 [79,95]. Amino acids 618-638 are essential for ICP0 binding with USP7 [99]. The last 250 amino acids at the C-terminus of ICP0 consist the ND10 binding site and the dimerization domain [100].

ICP0 has a functional E3 ligase RING domain at N-terminal amino acids 116-156 [97]. The ICP0 RING-finger domain is an active E3 ligase *in vitro* where the purified ICP0 was reported to catalyze unanchored polyubiquitin chain formation in an E2-dependent manner with the help of either UbcH5a (UBE2D1) or UbcH6a (UBE2E1) [101]. *In vivo*, these E2s co-localize with ICP0 in the nucleus of transfected cells [101]. During infection, the RING-finger domain allows ICP0 to interact with the ubiquitin-proteasome pathway through the proteasome-dependent degradation of cellular proteins and with deubiquitinating enzymes. The E3 ligase activity of ICP0 is responsible for the disruption of nuclear domain 10 (ND10) through the proteasome-mediated degradation of sumoylated (SUMO-1) ND10 associated proteins, PML and Sp100 [102-103, 104]. HSV-1

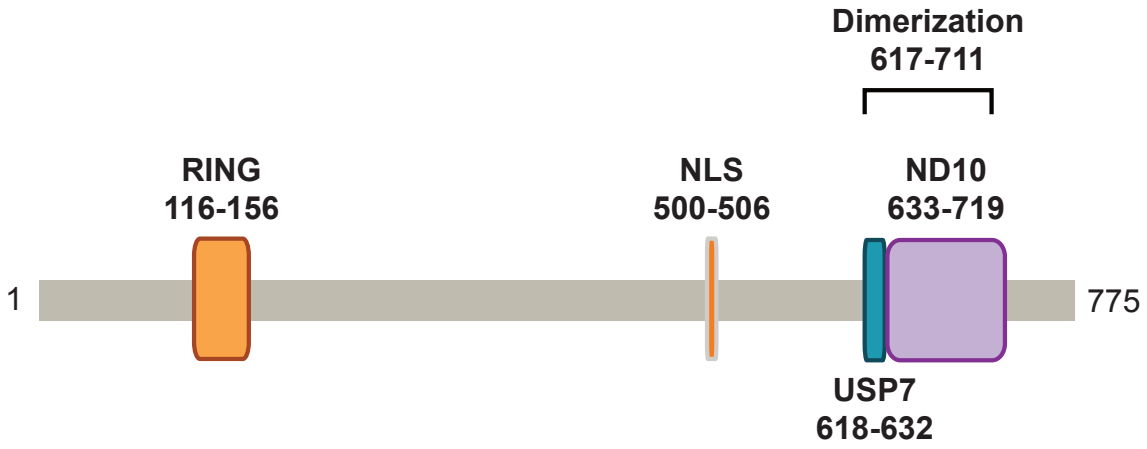



Figure 1-10: ICP0 primary sequence map. It has a functional RING-finger domain at residues 116-156. ICP0 localizes in the nucleus through its nuclear localization sequence, residues 500-506. ICP0 interacts with several cellular proteins. It interacts with deubiquitinating enzyme, USP7, at residues 618-632. Residues 633-719 allow ICP0 to localize near nuclear domain 10 (ND10) structures. ICP0 has a dimerization domain at residues 617-711. Adapted from Smith, Boutell, and Davido (2011).

disrupts ND10 within 2 hrs of infection induced by ICP0 [105]. In early HSV-1 infection, newly synthesized ICP0 localizes near ND10 structures in the nucleus via its ND10 binding site [105]. ND10s are small nuclear substructures seen as distinct foci of about 10 per nucleus [106]. Disruption of ND10 could be an important viral strategy because of ND10's involvement in antiviral response where some components of ND10, PML and Sp100, were induced by the interferon response [107-108]. In addition, viral replication of ICP0-null HSV-1 was restricted by both PML and Sp100 [109-110]. ND10 has been associated as a site of DNA transcription and replication for DNA viruses, including HSV [106].

Through co-immunoprecipitation with infected cell extracts and GST-pulldown assays, ICP0 was also found to bind to deubiquitinating enzyme, USP7, at amino acids 618-638. Amino acids K620 and K624 were found to be crucial for this interaction [99]. The ability of ICP0 to interact with USP7 contributes to its role in the activation of gene expression and stimulation of virus infection where deletion of the USP7 binding site and mutation of K620 reduced ICP0 activation of gene expression and virus growth, respectively [99]. ICP0 is unique from other herpesvirus viral protein mentioned earlier (EBNA1, vIRF4 and LANA) because it interacts at USP7-CTD instead of the USP7-NTD TRAF domain [58]. Similar to ICP0, some USP7 was also found to co-localize with PML at ND10 [40] and cause the destabilization of PML independent of USP7's catalytic activity [111]. At early infection, ICP0 can increase the proportion of ND10 that contains USP7 [40]. As an E3 ligase, ICP0 can catalyze ubiquitination of USP7 [112]. However, this effect of ICP0 to USP7 is secondary to USP7's ability to stabilize and regulate ICP0's

autoubiquitination for a mutation in ICP0's USP7-binding domain makes the mutant ICP0 less stable than wild type [113]. Also, ICP0 stability and expression level decreased during infection of USP7 RNAi-depleted cells [112-113]. ICP0 is also capable of translocating USP7 from the nucleus to the cytoplasm to inhibit Toll-like receptor signaling, possibly an immune evasion strategy by HSV-1 [114].

Lastly, ICP0's E3 activity is important for viral replication [85]. Viruses with a deleted RING domain or a mutated Zinc-finger have growth phenotypes similar to an ICP0 null-mutant [85]. These mutations also affect ICP0's ability to transactivate reporter gene expression and to initiate lytic infection and reactivate viral genome from latency [85].

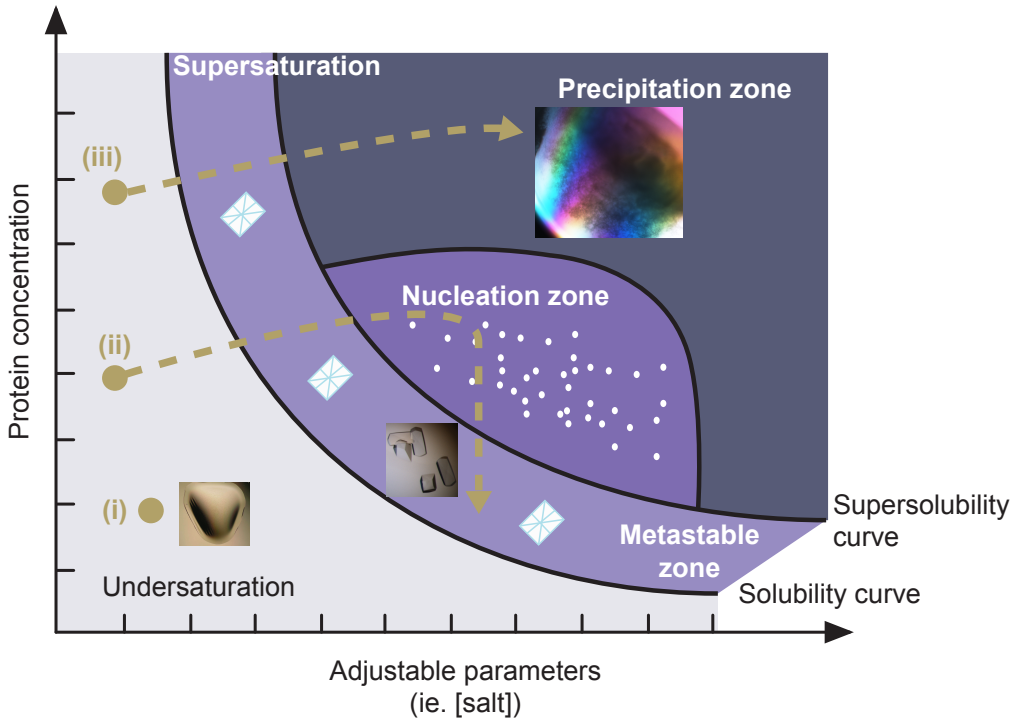
## 1.4 X-ray crystallography

X-ray crystallography is a method for the structure determination of proteins and nucleic acids at the atomic level. Crystal forms of proteins and nucleic acids, of significant size (~ 0.1 mm), are used to diffract high energy X-rays. The X-ray diffraction pattern is then interpreted using computer algorithms to obtain an electron density map, which will be used to map the location of the atoms in the protein. Ever since the first crystal structure of myoglobin was determined in 1958, the number of protein structures solved through x-ray crystallography has been increasing exponentially. Protein crystals are formed by slowly precipitating the protein out of solution in an organized manner. In a crystal, protein molecules arrange themselves in ordered repeats called the unit cell. One of the more common ways to set-up proteins for crystallization is through vapor diffusion method.

In the vapor diffusion method, the protein drop mixed with crystallization solution is allowed to reach equilibrium against a reservoir of concentrated crystallization solution [115]. Because of the sealed environment that the protein drop and reservoir solution are in, water diffuses out of the protein drop into the reservoir solution until there is equal concentration of precipitant in both [115]. During this process, the protein drop becomes supersaturated where protein molecules reach their solubility limit [115]. At this point, proteins can either form amorphous or crystalline precipitates. Protein solubility can be affected by temperature, precipitant concentration (salts, polymers, etc) and buffer pH [116-117]. Figure 1-11A shows the phase or solubility diagram of the protein crystallization solution. The y-axis is the protein concentration while the x-axis is the precipitant (ie. salt) concentration. The solubility curve signifies the protein solubility limit for a specific crystallization solution. Above this line, supersaturation, proteins will precipitate out of solution. Some proteins will easily form amorphous precipitates at the precipitation zone. Spontaneous crystal formation occurs in the nucleation zone, which decreases the protein concentration allowing for crystal growth in the metastable zone. Crystallization drops that are undersaturated will remain clear. Few big crystals are preferred over many small crystals. Crystals may appear overnight, months and even years.

Crystallization experiments using vapor diffusion could either be set-up as a sitting drop or a hanging drop (Figure 1-11B). Sitting drop vapor diffusion method is usually set-up in a 96-well format where the protein drop sits on the crystallization plate sealed with an optically clear tape. In this sealed environment, the difference in the precipitant concentration between the crystallization drop and reservoir solution drives

A)




B)

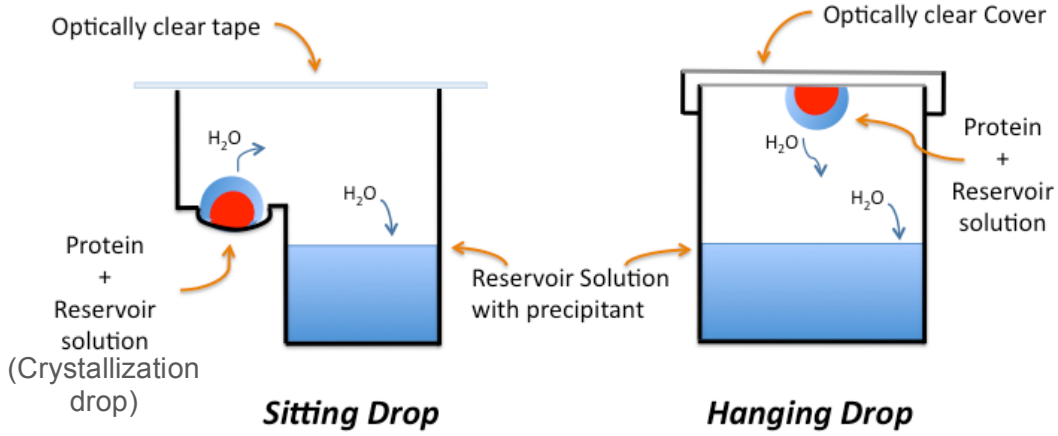



Figure 1-11: Protein crystallography. A) Protein crystallization phase diagram represents the different states a crystallization drop can go through during protein crystal trials. Protein concentration and adjustable paramaters like salt concentration affects the solubility of the protein in the crystallization buffer. The drop can stay clear (i) if it is undersaturated. If supersaturated, it can form protein crystals (ii) when in the metastable zone or amorphous precipitates (iii) when in the precipitation zone. B) Crystallization trials can be set-up using the sitting drop or hanging drop technique. The system is closed and the difference in precipitant concentration between the crystallization drop and reservoir solution drives water vapour to leave the crystallization drop until equilibrium is reached. This slowly concentrates the protein and allows the drop to reach supersaturation where protein crystals could appear.

water vapour to leave the crystallization drop and into the reservoir solution until equilibrium is reached. This process slowly concentrates the protein and allows the drop to reach supersaturation (in the phase diagram Figure 1-11A). Since it is impossible to predict which crystallization solution would induce protein crystal formation, sitting drop in 96-well format is ideal for screening multiple conditions at one time. As little as 1 μl of protein sample can be plated using into each well of the 96 well plate. Commercially available crystallization suites like Qiagen's JCSG-I, II, Hampton Research's PEG ion screen, are used in sparse matrix screening. Crystal trial plates can be incubated at room temperature or at 4ºC. Once a crystal condition hit is obtained from the sitting drop method, the crystallization condition is usually refined using hanging drop vapor diffusion. Hanging drop is set-up in a 15- or 24-well format where the protein drop hangs from the cover. Protein and precipitant concentration is refined to find the optimal condition for crystal growth in order to increase crystal size and crystal quality. This refinement stage also ensures that the crystallization condition is reproducible.

There are a variety of ways to optimize or refine protein crystallization conditions: seeding, additive screen, limited proteolysis. In seeding, crystals can be harvested and crushed using a seeding tool (ie. Seed Bead from Hampton Research). Using the same crystallization condition that the original crystal was obtained from, the protein drops are seeded with a few crystals. These seed crystals serve as nuclei where crystallization continues and usually produce fewer but larger crystals. An additive screen where different additives (chaotrophes, divalent cations, denaturing agents, etc.) are used to supplement the protein drop is also commercially available. This may help improve the

size and quality of the crystal. Sometimes, proteins have disordered loops which could be preventing it from forming crystals and this may be solved through limited proteolysis where diluted amounts of proteases are added to the protein to cleave the disordered loops [118].

Protein crystallographers encounter a wide variety of challenges from obtaining significant amount of soluble proteins, obtaining crystal hits from crystallization screen to acquiring a diffraction resolution sufficient for structure determination. Very pure homogenous proteins samples are required for crystallization as any impurities can affect the crystal packing thus affecting the quality of the crystal. If the native protein does not produce crystals on its own on the initial crystallization screen, another screen can be set-up where a ligand or interacting partner can be added to form a complex with the native protein and crystallize them together, a process called co-crystallization.

Once a decent sized protein crystal is obtained, it is then prepared for X-ray diffraction data collection. The crystal is scooped out of the drop using a circular synthetic fiber loop and flash frozen with a cryoprotectant into liquid nitrogen. Cryoprotectants (ie. glycerol and ethylene glycol) prevent ice formation, which can destroy the protein crystal during freezing [119]. Freezing protein crystals minimizes radiation damage during data collection [119]. X-rays directed towards the protein crystal are diffracted and create a diffraction pattern consisting of spots or reflections. A detector then measures the intensities and positions of each reflection. These reflections are then analyzed using computer algorithms to calculate the electron density map of the

molecule. Amino acids are modeled into the electron density map to generate the three-dimensional crystal structure.

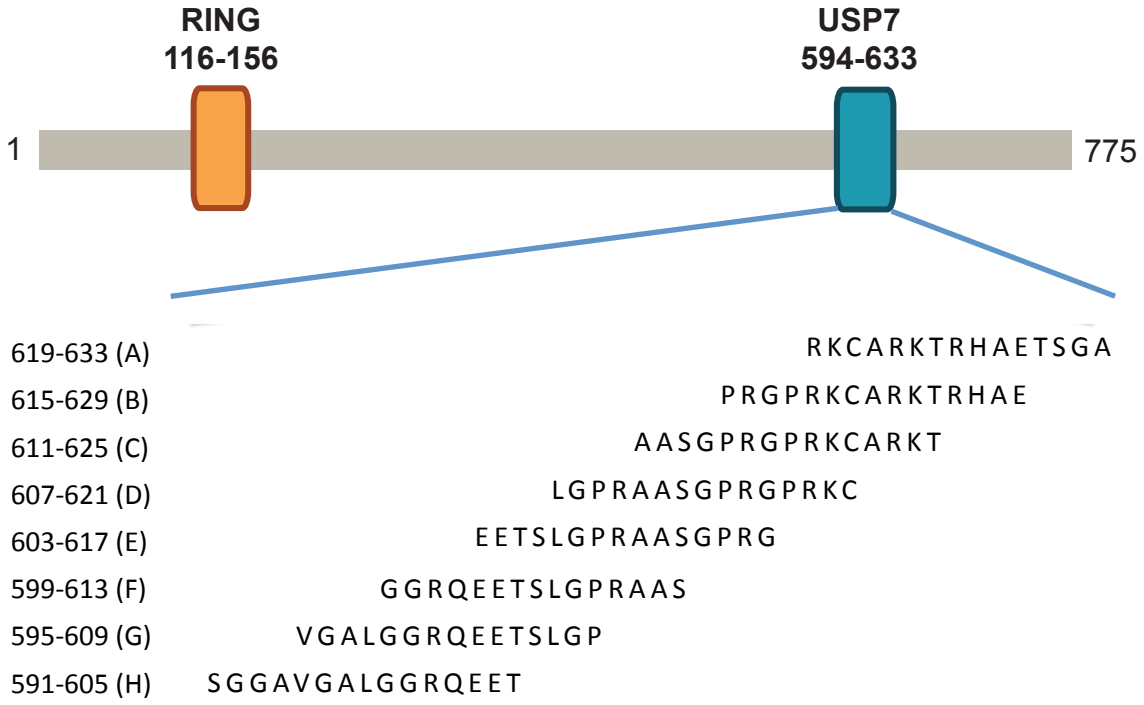
## 1.5 Research Objectives

Unlike the USP7 N-terminus and catalytic domains, little is known about the three dimensional structure of the USP7 C-terminus. Secondary structure predictions have shown 4-5 possible ubiquitin-like domains in the C-terminus. One of the objectives of this thesis is to determine the structure of USP7 C-terminal domain using X-ray crystallography and see how these Ubls fold and interact with each other. HSV-1's ICP0 interference of the UPS through its E3 ligase activity and interaction with USP7, a deubiquitinating enzyme, has been thought to be important for the survival of the virus. *In vitro* experiments through co-immunoprecipitation of ICP0-USP7 complex using infected cell extracts have shown strong and specific binding of ICP0 (residues 594-633) to USP7 where ICP0 amino acids K620 and K624 being crucial for this interaction [99]. K620 and K624 are part of three grouped positively charged doublets found in ICP0 residues 619-627 (Appendix A). In another study, Holowaty *et al*. (2003), partially digested USP7 with trypsin and chymotrypsin and found that the USP7 C-terminal have two protease resistant domains, residues 622-801 and 885-1061 through mass spectrometry. Partially digested USP7 was applied to a glutathione-sepharose column containing GST-tagged ICP0 C-terminal (residues 594-775) [58]. Of the two USP7 C-terminal protease resistant domains, residues 622-801 bound to the column [58]. A previous graduate student in the lab, Hong Zheng, cloned and purified His-tagged USP7 560-870. In order to determine the essential

residues in ICP0 important for USP7 interaction, GST-tagged fifteen amino acid (15-mer) long peptides that span the USP7 binding domain (residues 594-633) of ICP0 were cloned and purified (Figure 1-12A). Using GST-pulldown assays, ICP0(B) 615-629 peptide was shown to bind to USP7 560-870 with the strongest affinity (Figure 1-12B). This leads to another objective of this study, to identify which Ubl domain/domains is/are responsible for this interaction.

To study the biochemical basis of interaction between USP7 and ICP0, specifically to identify which amino acids from USP7 and ICP0 are important for their interaction, I have carried out a series of biochemical and biophysical experiments, including a GST-pulldown assay of different His-tagged USP7 C-terminal constructs with both wild type and mutated GST-tagged ICP0 peptides. From these assay, wild type GST-ICP0 peptide ($^{615}$PRGPRKCARKTRHAE$^{629}$) interacted with both USP7 Ubl 12 (residues 535-776) and USP7 Ubl 123 (residues 535-888). The ICP0 mutant peptides lost this interaction when either K620 or K624 is mutated. Using fluorescence polarization, the binding affinity between FITC-labeled ICP0 peptide and USP7 C-terminal constructs was characterized. Lastly, X-ray crystallography was used to obtain the co-crystal structure of USP7 C-terminus and ICP0 peptide. These results would help in development of treatments against HSV-1 infection.

A)




B)

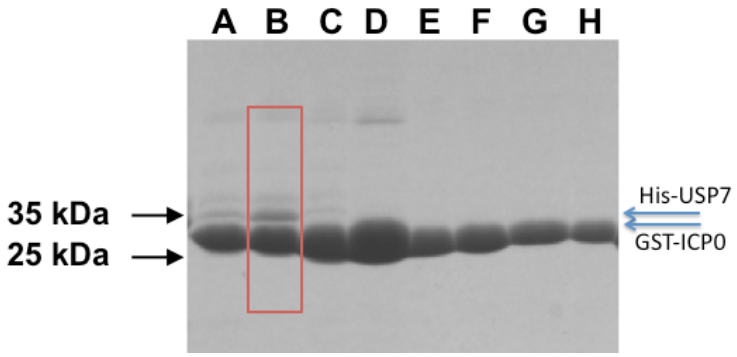



Figure 1-12: Interaction of USP7 560-870 with GST-tagged ICP0 peptides. A) Eight 15-amino acid long GST-tagged ICP0 peptides (named A-H) that span ICP0's USP7 binding site were cloned and purified. B) GST-pulldown assay of His-tagged USP7 560-870 with GST-ICP0 peptides A-H. Proteins that bound to the GST-Sepharose column and eluted with reduced glutathione were run in 15% SDS-PAGE gel stained with Coomassie blue. His-USP7 560-870 (~36 kDa). GST-ICP0 peptides (~28 kDa). GST-ICP0(B) peptide residues 616-629 strongly interacted with USP7 560-870. This was an experiment previously done by Hong Zheng.

# CHAPTER 2: MATERIALS AND METHODS

## 2.1 Cloning

### 2.1.1 Template and vector

USP7 1-1102 in pCANmycUSP7 was a gift from Dr. Lori Frappier (University of Toronto).

### 2.1.2 Primers

Primers synthesized by IDT technologies were resuspended in Milli Q water to a final concentration of 50 μM. The primers were used for Polymerase Chain Reaction (PCR) where the Tm used was calculated as 2°C for A/T and 4°C for G/C. Primers used are summarized in Table 2-1.

### 2.1.3 Polymerase Chain Reaction (PCR)

Each polymerase chain reaction mixture contained 0.05 U/μL *Pfu* polymerase, 1X *Pfu* buffer (20 mM Tris-HCl pH 8.8, 1 mM MgSO4, 10 mM KCl, 10 mM $(NH4)_2SO4$, 0.1 % Triton X-100, 0.1 mg/mL nuclease-free BSA), 0.3 mM of each dNTP (dATP, dCTP, dGTP, dTTP), 50 ng template DNA, 50 pmol of forward and reverse primers and Milli Q water to a total reaction volume of 50 μL. The cycling parameters started with 10 minutes at 95°C for the initial denaturation. Next, there was 30 cycles of 1 minute

subsequent denaturation, 1 minute at the chosen annealing temperature (usually <5-10°C of the calculated Tm) and 1-3 minutes at the extension temperature of 72°C (1 min/kb of PCR target). Lastly, the final extension was 20 minutes at 72°C. After PCR, samples were electrophoresed on an agarose gel.

Table 2-1: USP7 C-terminal primers.

| Construct (residues) | Ubl | Primer sequence |
|---|---|---|
| 535-776 | 12 | 533f<br>5′-ttg tat ttc cag ggc atg gat att cct cag ttg gtg gag-3'<br>775r<br>5′-caa gct tcg tca tca atc att ttc agg gtc atc ctt ctg-3' |
| 535-888 | 123 | 533f<br>5′-ttg tat ttc cag ggc atg gat att cct cag ttg gtg gag-3'<br>885r<br>5′-caa gct tcg tca tca aga gtc tgt gat ttt cat ctt aag ctg-3' |
| 535-890 | 123b | 533f<br>5′-ttg tat ttc cag ggc atg gat att cct cag ttg gtg gag-3'<br>890r<br>5′-caa gct tcg tca tca ctc aaa gtc tgt gat ttt cat ctt aag-3' |
| 680-888 | 23 | 680f<br>5′-ttg tat ttc cag ggc atg gat aaa gat cat gat gta atg-3'<br>885r<br>5′-caa gct tcg tca tca aga gtc tgt gat ttt cat ctt aag ctg-3' |
| 680-1102 | 2345 | 680f<br>5′-ttg tat ttc cag ggc atg gat aaa gat cat gat gta atg-3'<br>1102r<br>5′-caa gct tcg tca tca gtt atg gat ttt aaa ggc ctt-3' |
| 776-1102 | 345 | 775f<br>5′-ttg tat ttc cag ggc atg gat aac agt gaa tta ccc acc gc-3'<br>1102r<br>5′-caa gct tcg tca tca gtt atg gat ttt aaa ggc ctt-3' |
| 886-1102 | 45 | 885f<br>5′-ttg tat ttc cag ggc atg aca gac tct gag aac agg cga agt-3'<br>1102r<br>5′-caa gct tcg tca tca gtt atg gat ttt aaa ggc ctt-3' |
| 884-1102 | 45a | 884f<br>5′-ttg tat ttc cag ggc aaa atc aca gac ttt gag aac agg-3′<br>1102r<br>caa gct tcg tca tca gtt atg gat ttt aaa ggc ctt-3' |
| 680-776 | 2 | 680f<br>5′-ttg tat ttc cag ggc atg gat aaa gat cat gat gta atg-3'<br>775r<br>5′-caa gct tcg tca tca atc att ttc agg gtc atc ctt ctg-3' |
| 776-888 | 3 | 775f<br>5′-ttg tat ttc cag ggc atg gat aac agt gaa tta ccc acc gc-3'<br>885r<br>5′-caa gct tcg tca tca aga gtc tgt gat ttt cat ctt aag ctg-3' |

### 2.1.4 Agarose gel electrophoresis

1X DNA loading dye (10 mM Tris-HCl pH 7.6, 0.03 % bromophenol blue, 0.03 % xylenecyanol FF, 60 % glycerol, 60 mM EDTA) was added to 5 µL of each PCR reaction and loaded on 1% agarose gel (1 g Agarose, 1X TAE (40 mM Tris acetate, 1mM EDTA) in 100 mL $dH_2O$). Along side the samples, 5 µL of DNA ladder (Fermentas Gene Ruler 1 kb DNA ladder or Invitrogen 100 bp DNA ladder) was also loaded. The gel was electrophoresed with 0.5X TAE for 30 minutes at 100 V. The gel was soaked in ethidium bromide solution (0.5 µg/mL EtBr, 0.5X TAE in 100 mL $dH_2O$) to stain the DNA and viewed/photographed under ultraviolet light.

### 2.1.5 Agarose gel purification of PCR product

Amplified DNA in the 1 % agarose gel was purified using a DNA gel purification kit (GE Healthcare). The gel containing the PCR band was excised and mixed with the capture buffer in a microfuge tube. The gel was dissolved at 60° C and transferred to the purification column. After centrifugation of the column at 16,100 xg (Eppendorf Benchtop Centrifuge 5415D) for 1 minute, the flow through was discarded and the column was washed with 500 µL wash buffer. After centrifugation of the column at 16,100 xg for 1 minute, the flow through was discarded and the DNA was eluted with 40 µL $ddH_2O$. The PCR DNA product was concentrated using a speed vacuum system (Thermo) to around 100 ng/µL and stored at -20° C.

### 2.1.6 In-Fusion™ ligation reaction

100 ng/µL PCR DNA product was added to 100 ng/µl BseRI linearized p15TV-L vector, 1X In-Fusion™ (Clontech) reaction master mix and dd$H_2O$ to a final reaction volume of 2.5 µL. The mixture was at 37° C for 15 minutes and 55° C for 15 minutes. The ligation reaction was stored at -20° C before transformation.

### 2.1.7 Transformation into Electrocompetent DH5α

50 µl of electrocompetent *Escherichia coli* (*E. coli*) DH5α (F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^-$ $m_K^+$), λ–) was added to the In-Fusion reaction mixture and incubated on ice for 10 minutes. The cells were placed in a chilled 0.1 cm gap electrophoration cuvette and then in a BioRad Gene pulser set to 1700 V. After electropulsing, the cells were added to tube of 100 µl pre-warmed SOC medium (2 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 0.05 % (w/v) NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) and incubated at 37°C for 30 minutes. The cells were plated on LB-agar plates (Luria Broth (10 g tryptone, 5 g yeast extract, 10 g NaCl, d$H_2O$ up to a liter), 1.5 % agar (15 g/L)) containing 5 % (w/v) sucrose and 100 µg/mL ampicillin. The plates were then incubated overnight in a 37°C incubator.

### 2.1.8 Mini-preparation of plasmid DNA of individual colonies

Colonies that grew overnight on the sucrose-ampicillin plates were inoculated into 5 mL Luria Broth (LB) with 100 µg/mL ampicillin and incubated overnight at 37°C and

shaking at 200 rpm (Innova 40, New Brunswick Scientific). The cells were pelleted in a 2 mL microfuge tube at 16,100 xg (Eppendorf Benchtop Centrifuge 5415D) for 1 minute. QIAprep Spin Miniprep kit (Qiagen) was used to extract the plasmid. 250 µL buffer P1 was used to resuspend the cells and vortexed. Next, 250 µL buffer P2 was added and mixed by inverting the tubes. To precipitate genomic DNA and other bacterial proteins, 350 µL buffer N3 was added. The tubes were then centrifuged for 10 minutes at 16,100 xg. The cell lysate was transferred to a QIAprep column and centrifuged at 16,100 xg for 1 minute. The column was then washed with 500 µL buffer PE and centrifuged at 16,100 xg for 1 minute. To make sure that the column was dry, it was centrifuged for another minute at 16,100 xg. Plasmid was eluted from the column by the addition of 40 µL dd$H_2O$ and centrifugation for 1 minute at 13000 rpm. The plasmids were stored at -20° C.

### 2.1.9 Positive clone screening and sequencing

The plasmids from mini-prepped colonies were tested for positive clones through PCR using the same primers to make the clones. The clones were then verified by sequencing (York University Core Facility).

## 2.2 His-tagged protein expression

### 2.2.1 Transformation into BL21(DE3)mgk

Positive clones were transformed into chemically competent *E. coli* BL21(DE3)mgk ($F^-$ ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1

sam7 nin5])). This strain of BL21 has a plasmid that codes for rare tRNA codons (mgk), which are tRNA not commonly found in abundance in *E. coli*. This is especially useful when expressing eukaryotic proteins in *E. coli*. 1 µL of plasmid was added to a tube of 25 µL chemically competent BL21(DE3)mgk cells and incubated for 10 minutes on ice. The cells were quickly heat-shocked for 45 seconds on a 42° C water bath and then placed on ice for 2 minutes. 100 µL of warm LB was added to the cells and then incubated at 37° C for 30 minutes. The cells were then plated on pre-warmed LB-agar plates with 100 µg/mL ampicillin and 50 µg/mL kanamycin. The plates were incubated at 37° C overnight. 5 mL of LB, inoculated with colonies that grew on the plate, was incubated overnight at 37° C and shaking at 200 rpm. The overnight culture was used to prepare a glycerol stock with equal ratios of overnight culture and 30 % (w/v) glycerol-LB mixture. The glycerol stocks were stored at -80°C.

### 2.2.2 Protein expression induction

The glycerol stocks were used to inoculate 50 mL LB medium with 100 µg/mL ampicillin and 50 µg/mL kanamycin incubated overnight at 37° C and shaking at 200 rpm. The overnight cultured was transferred to 1 L Terrific broth (TB) (12 g/L tryptone, 24 g/L yeast extract, 9.4 g/L $K_2HPO_4$, 2.2 g/L $KH_2PO_4$) with 100 µg/mL ampicillin and 50 µg/mL kanamycin. The culture was allowed to grow at 37° C and shaking at 200 rpm (Innova 43R, New Brunswick Scientific) until an $OD_{600nm}$ of 1. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the culture, final concentration of 0.4 mM, to induce protein expression and incubated at 16° C and shaking at 200 rpm overnight. The

cells were collected by centrifugation for 30 minutes at 4° C at 7,500 xg (Beckman Coulter Avanti J-E centrifuge with a JA-9.1 rotor). The cell pellets were either stored at -80° C or used immediately for protein purification.

### 2.2.3 Selenomethionine-labeling

An overexpressing His-tagged clone in *E. coli* BL21(DE3)mgk was grown overnight in 50 mL LB medium with 100 µg/mL ampicillin and 50 µg/mL kanamycin incubated overnight at 37°C and shaking at 200 rpm. 1 L of minimal media (1X M9 (6 g $Na_2HPO_4$, 3 g $KH_2PO_4$, 1 g $NH_4Cl$, 0.5 g NaCl), 1 mM $MgSO_4$, 0.4 % (w/v) Glucose, 0.00005 % (w/v) Thiamine, 0.0038 mg/mL $FeSO_4$, 0.00005 % (w/v) Biotin, 100 µg/mL ampicillin and 50 µg/mL kanamycin in 1L water) was prepared. The overnight culture was centrifuged for 5 minutes at 1,100 xg (Sorvall Legend RT+). The pellet was washed twice with 25 mL of minimal media and then added to the rest of the minimal media. The culture was allowed to grow at 37° C and shaking at 200 rpm until an $OD_{600nm}$ of 1. Amino acid supplements (L-Lysine 100 mg/L, L-Phenylalanine 100 mg/L, L-Threonine 100 mg/L, L-Isoleucine 50 mg/L, L-Leucine 50 mg/L, L-Valine 50 mg/L, and L-Selenomethione 50 mg/L) were added to the media and incubated for another 15 minutes for inhibition of methionine biosynthesis. IPTG was added to the culture, final concentration of 0.4 mM, to induce protein expression and incubated at 16° C and shaking at 200 rpm overnight. The cells were collected by centrifugation for 30 minutes at 4° C at 7,500 xg (Beckman Coulter Avanti J-E centrifuge with a JA-9.1 rotor). The cell pellets were immediately used for protein purification.

### 2.2.4 His-tag Cleavage

TEV protease (60 mg TEV per 1000 mg protein), 2.5 mM $CaCl_2$, 1mM DTT and purified protein were mixed and dialyzed overnight in 20 mM Tris pH7.5 and 500 mM NaCl at 4°C.

## 2.3 GST-tagged protein expression

GST-tagged ICP0(B) peptide was cloned by Hong Zheng and Andraya Lewis. GST-tagged ICP0(B) mutant peptides were cloned by Tim Shigapov. Clones in *E. coli* BL21 (DE3) were grown overnight in 5 mL LB medium with 100 mg/mL ampicillin at 37° C and shaking at 200 rpm. The overnight cultured was transferred to 50 mL TB medium with 100 μg/mL ampicillin. The culture was allowed to grow at 37° C and shaking at 200 rpm until an $OD_{600nm}$ of 1. IPTG was added to the culture, final concentration of 0.4 mM, to induce protein expression and incubated at 16° C and shaking at 200 rpm overnight. The cells were collected by centrifugation for 20 minutes at 1,100 xg (Sorvall Legend RT+) and used for protein purification.

## 2.4 Protein purification

### 2.4.1 Nickel Affinity Chromatography

Cell pellets were resuspended with Binding buffer (20 mM Tris pH 7.5, 500 mM NaCl, 5 mM imidazole) and 1X protease inhibitor (0.05 mM phenylmethylsulfonyl fluoride (PMSF), 0.1 mM benzamidine, cold absolute ethanol). The cells where then lysed using a Branson Sonifier at 30% amplitude, 10 sec ON, 15 sec OFF for 12 minutes. After sonication, another 1X of protease inhibitor was added. The cell lysates were centrifuged at 41,400 xg for 30 minutes at 4° C (Beckman Avanti J-E with a JA-25.5 rotor). During centrifugation, 5 mL of Ni-NTA metal-affinity chromatography resin was washed with dd$H_2O$ and equilibrated with 10 mL of binding buffer. The supernatant was then added to the column of equilibrated nickel resin. The supernatant and nickel resin were incubated for 30 minutes at 4° C. After incubation, any proteins in the cell lysate that did not interact with the nickel resin were allowed to flow through (FT). A sample of the FT was saved for SDS-PAGE gel analysis. The column was washed with 10 mL of Binding buffer (B). A sample of B was saved for SDS-gel analysis. The column was further washed with 100 mL of Wash buffer (W) (20 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole). A sample of W was saved for SDS-gel analysis. Proteins that bound to the nickel resin were eluted with Elution buffer (E) (20 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole). 1 mM EDTA and 0.33 mM DTT were added to the eluted fraction. 1X of protein dye (62.6 mM Tris-Cl pH 6.8, 2 % (w/v) SDS, 10 % (v/v) glycerol, 0.1 M

DTT, 0.01 % (w/v) bromophenol blue) was added to the purification fractions (FT, B, W, E) and electrophoresed on 15% SDS-PAGE.

### 2.4.2 Glutathione-S-Transferase (GST) affinity chromatography

Cell pellets were re-suspended with 1X phosphate buffered saline (PBS) (1.44 g $Na_2HPO_4$, 0.24 g $KH_2PO_4$, 8 g NaCl, and 0.2 g KCl in 1L $ddH_2O$ at final pH 7.4). The cells where then lysed using a Branson Sonifier at 50 % amplitude, 6 sec ON, 2 sec OFF for 3 minutes. The cell lysates were centrifuged at 41,400 xg for 30 minutes at 4° C (Beckman Coulter Avanti J-E with a JA-25.5 rotor). During centrifugation, 1-2 mL of Glutathione-Sepharose (GE Healthcare) beads or 10 mg GST-protein/1mL dry beads was equilibrated with 1X PBS. The supernatant was added to the column of equilibrated Glutathione-Sepharose beads and incubated for an hour at 4°C. After incubation, any proteins in the cell lysate that did not interact with the Glutathione-Sepharose beads were allowed to flow through (FT). The column was then washed three times with 1X PBS (W). Bound proteins were eluted with elution buffer (50 mM Tris pH 8.0, 500 mM NaCL, 30 mM reduced glutathione) (E).

### 2.4.3 Size-exclusion chromatography (Gel filtration)

Superdex 200 16/60 (GE Healthcare) or Sephacryl 200 16/60 (GE Healthcare) columns were equilibrated with the desired buffer. Filtered, purified, concentrated protein was injected into the AKTA Purifier (GE Healthcare) and chromatographed at a flow rate

of 0.5 mL/min. The eluted fractions from the desired peaks were collected and concentrated for further experiments.

### 2.5 GST-pulldown assay

For GST-pulldown assay, 100 µL of Glutathione-Sepharose beads were equilibrated with pulldown buffer (1X Protease inhibitor cocktail (0.1 mM Benzamidine, 0.05 mM PMSF), 5 % glycerol, 50 mM Tris pH 8.0, 100 mM NaCl and 5 mM DTT). 250 µg of GST-ICP0 peptide was added to the beads. The volume was brought up to 500 µL with pulldown buffer and incubated at 4° C rotating for 1.5 hrs. The beads:GST-ICP0 mixture was centrifuged at 100 xg (Eppendorf Benchtop Centrifuge 5415D) for 30 seconds. The supernatant was removed and 500 µg of various His-tagged USP7 proteins were added. The volume was brought up to 500 µL with pulldown buffer and incubated at 4° C rotating for 1.5 hrs. The mixture was transferred to a micro column and the flow through was collected. The column was washed twice with 500 µL of pulldown buffer and the wash fractions were collected. The proteins were eluted from the beads with 50 µL elution buffer (50 mM Tris pH 8.0, 500 mM NaCL, 30 mM reduced glutathione) and centrifuged at 600 xg for 30 seconds into a 1.5 mL microfuge tube. This was repeated 5 times to ensure that all of the proteins were eluted. GST was used as a negative control. Eluted proteins were analyzed by SDS-PAGE. A mixture of 250 µg of GST-ICP0 peptide and 500 µg of various His-tagged USP7 proteins in 500 µL pulldown buffer were also electrophoresed with the eluted proteins as a representation of the proteins added to the beads.

## 2.6 Fluorescence polarization assay

Purified USP7 proteins were titrated (0 to 250 μM) with N-terminally labeled FITC-ICP0(B) peptide (synthesized by CanPeptide, 40 nM) in 0.01% Triton X-100, 150 mM NaCl, 20mM Tris pH 7.5. Polarized emission of FITC-ICP0(B) was measured with BioTeK-Synergy2. The excitation wavelength was 485 nm while the emission wavelength was 528 nm. The polarization data were plotted and analyzed using GraphPad Prism version 5.0 to calculate $K_D$ values.

## 2.7 Protein crystallization

### 2.7.1 Protein crystallization trials

Initial protein crystallization trials were set-up using vapour diffusion sitting drop technique in 96-well plates (Douglas Instruments). The 96-well plates' buffer reservoir wells were filled with 100 μL of crystallization buffer from sparse matrix crystallization suites JCSGI-IV, JCSG+, PACT, PEGsII and ClassicsII (QIAGEN) using a 100 μL multichannel pipette. 1μL of concentrated purified protein was aliquoted using a 1μL multichannel pipette in the crystallization well and mixed with 1μl of crystallization buffer. The plates were sealed with optically clear tape and were incubated at either 4°C or room temperature. Plates were checked regularly.

Protein co-crystallization trials were set-up the same way. Interacting protein: peptide were mixed in a 1:2 ratio and allowed to interact for a few minutes before plating.

### 2.7.2 Crystal refinement

Once protein crystals appear in the crystallization trials, the crystallization conditions were refined using the vapour diffusion hanging drop technique in 15-well or 24-well format. Some of the factors optimized during refinement were protein concentration, pH, salt concentration, precipitant concentration, additives, affinity tag cleavage, and seeding.

## 2.8 Crystal diffraction

Protein crystals were harvested with a cryo loop and submerged into a cryo solution. The cryo solution was composed of crystallization buffer with 10-20% of cryoprotectant (ie. glycerol, 2-Methyl-2,4-pentanediol (MPD), alcohol, etc.). The cryo solution was optimized for each different protein crystal. The diffraction of frozen crystals were tested on a MicroMax007 rotating anode diffractometer (Rigaku) with a Saturn 944+ detector at -180°C. X-ray datasets were collected on most crystals that diffracted greater than 3.5Å.

# CHAPTER 4: RESULTS

## 4.1 Cloning

The C-terminus of USP7 was predicted to have 5 ubiquitin-like (Ubl) domains. Using PHYRE, a secondary structure prediction program (*www.sbg.bio.ic.ac.uk/~**phyre**/*), the secondary structure of the USP7 C-terminus showed the presence of the ubiquitin fold, ββαββ. Five possible Ubl domain boundaries were highlighted (Figure 4-1). USP7 C-terminal constructs were designed and cloned base on the possibility of having Ubls as individual domains (Table 4-1). These constructs were cloned into p15TV-L vector. The DNA fragments for the constructs were amplified using polymerase chain reaction (PCR) and the PCR fragments were inserted into the p15TV-L vector. Positive clones were identified using PCR and confirmed by sequencing. All of the positive clones were transformed into *E. coli* BL21(DE3)mgk cells for protein expression.

The GST-tagged ICP0(B) $^{615}$PRGPRKCARKTRHAE$^{629}$ mutant peptides were cloned by Tim Shigapov (a previous undergraduate student in the lab) into the pGEX-2TK vector.

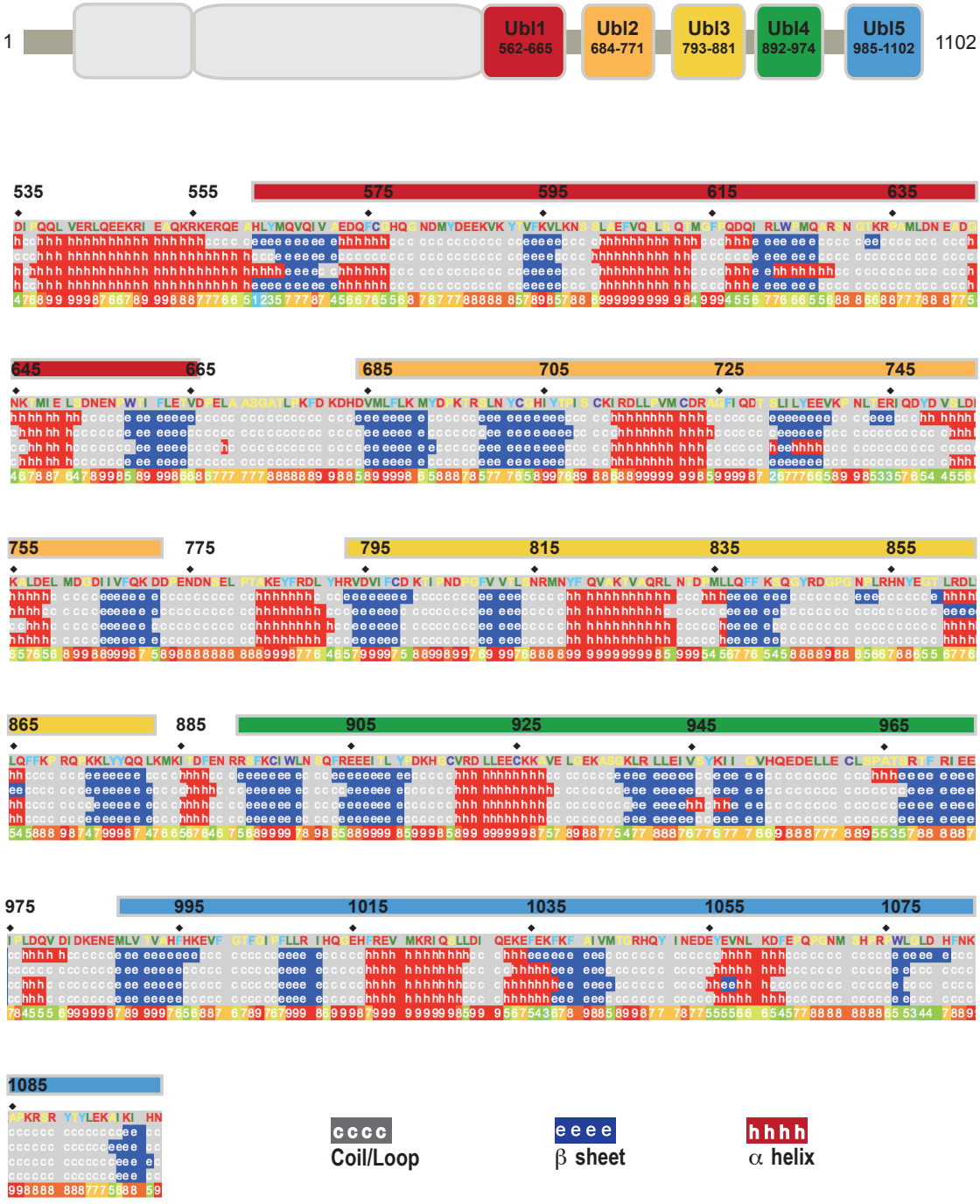



Figure 4-1: USP7 535-1102 or USP7 C-terminus secondary structure prediction. Possible ubiquitin-like (Ubl) domain boundaries are colored and labeled as Ubl1 to Ubl5. Ubl amino acid position: Ubl1 (562-665), Ubl2 (684-771), Ubl3 (793-881), Ubl4 (892-974), Ubl5 (988-1102).

Table 4-1: USP7 C-terminus constructs

| Construct * | Ubl |
|---|---|
| 535-776 | 12 |
| 535-888 | 123 |
| 535-889 | 123b |
| 680-888 | 23 |
| 680-1102 | 2345 |
| 776-1102 | 345 |
| 886-1102 | 45 |
| 884-1102 | 45a |
| 680-776 | 2 |
| 776-888 | 3 |
| 535-1102 ** | 1-5 |

* Number indicated represents the residue number of human USP7
** USP7 C-terminal (535-1102) was previously cloned by Hong Zheng.

## 4.2 Protein Expression and Purification

### 4.2.1 His-tagged USP7 C-terminal constructs

Ten USP7 C-terminal constructs successfully cloned in p15TV-L were transformed in *E. coli* BL21(DE3)mgk cells for protein expression. All of constructs were overexpressed. Nine out of the ten constructs were soluble and were successfully purified. A summary of the USP7 C-terminal constructs is shown in Table 4-2. USP7 535-1102 (Ubl1-5) or full-length USP7 C-terminal (CT) (prepared by Hong Zheng during her M.Sc.), was also expressed in BL21(DE3)mgk cells and produced soluble protein that was later purified.

Table 4-2: 6XHis-tagged USP7 C-terminal constructs

| Construct | Ubl | Vector | MW(kDa)* | Soluble | Expression | Property |
|---|---|---|---|---|---|---|
| 535-776 | 12 | p15TV-L | 30.5 | yes | high | degrades |
| 535-888 | 123 | p15TV-L | 43.6 | yes | high | degrades |
| 535-890 | 123b | p15TV-L | 43.9 | yes | high | degrades |
| 680-888 | 23 | p15TV-L | ~25.3 | no | low | - |
| 680-1102 | 2345 | p15TV-L | ~48.8 | yes | low | - |
| 776-1102 | 345 | p15TV-L | 41.2 | yes | high | aggregates |
| 886-1102 | 45 | p15TV-L | ~26.2 | yes | medium | aggregates |
| 884-1102 | 45a | p15TV-L | 28.4 | yes | medium | - |
| 680-776 | 2 | p15TV-L | ~13 | yes | low | - |
| 776-888 | 3 | p15TV-L | ~14.7 | yes | low | - |
| 535-1102 | 1-5 | p15TV-L | 69.2 | yes | high | degrades |

* MW includes the 6X-Histadine tag and TEV recognition site

*Expression and purification of USP7 535-776 (Ubl12)*

6XHis-tagged USP7 535-776 (Ubl12) was purified by Ni-NTA affinity chromatography and size exclusion chromatography. Ubl12 was highly overexpressed and soluble. Eluted Ubl12 showed some minor degradation with the main protein band around 30 kDa (Figure 4-2A). Size exclusion chromatography using Superdex200 HiPrep 16/60 in 20 mM Tris pH7.5 167 mM NaCl, shows peak broadening of the major peak at the peak tail or right side of the peak (Figure 4-2C). There were also two smaller peaks on the left side possibly from high molecular weight aggregated proteins. Fractions from the peak tail electrophoresed on SDS-PAGE gel, shows degradation products (Figure 4-2B). Thus, fractions closest to the middle of peak were collected to avoid most of these smaller molecular weight products. Purified Ubl12 was further used for GST-pulldown assay, fluorescence polarization assay and protein crystallization trials.

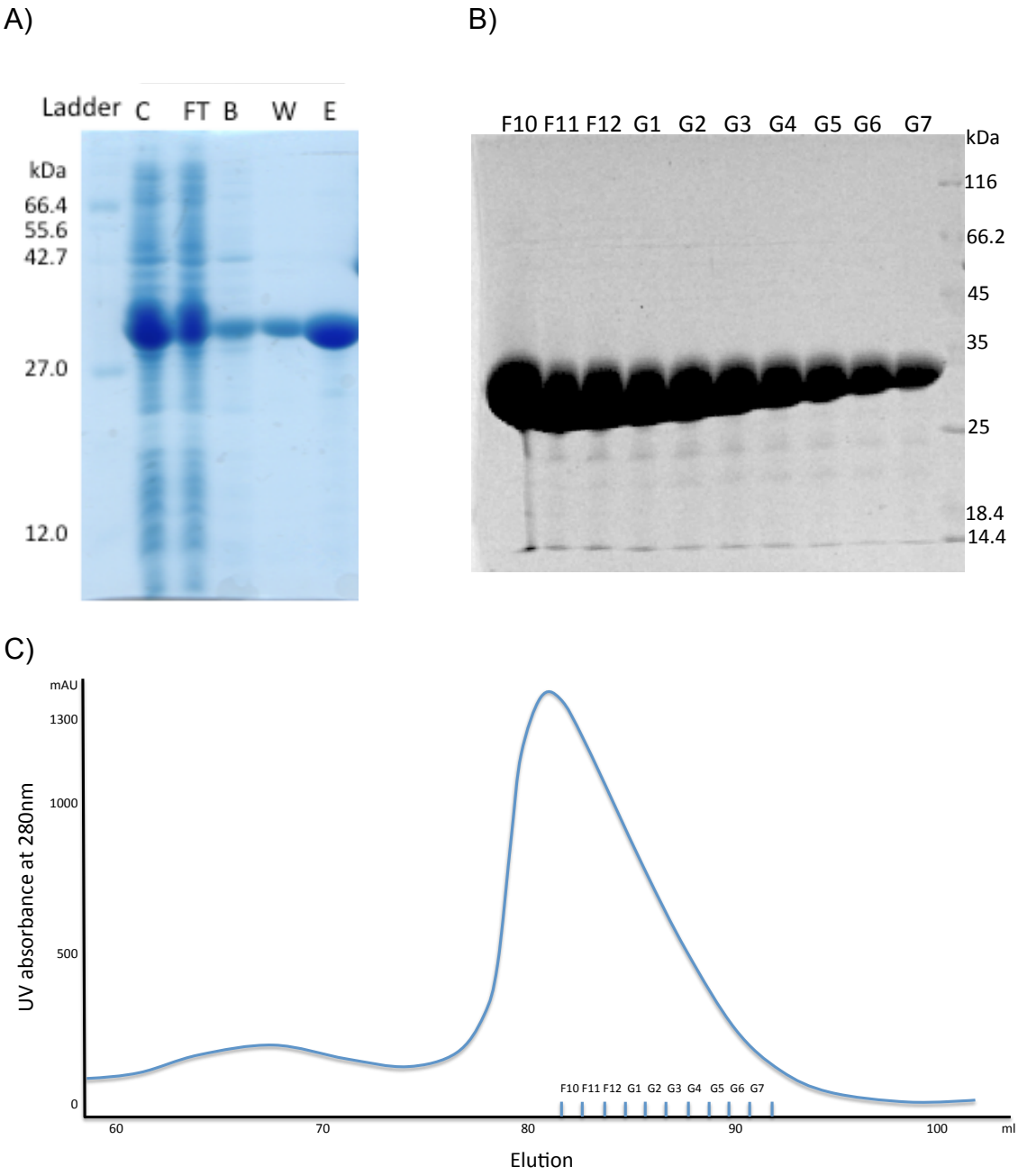



Figure 4-2: 6XHis-tagged USP7 535-776 (Ubl12) purification. Protein samples were run in a Coomassie stained SDS-PAGE gel. A) Ubl12 Ni-NTA affinity chromatography purification fractions: Crude (C), Flow through (FT), Binding (B), Wash (W), and Elute (E). Ubl12 eluted protein band ~30 kDa. B) Ubl12 size exclusion chromatography fraction samples. C) Ubl12 size exclusion chromatography profile.

*Expression and purification of USP7 535-888 (Ubl123)*

6XHis-tagged USP7 535-888 (Ubl123) was purified by Ni-NTA affinity chromatography and size exclusion chromatography. Ubl123 was highly overexpressed and soluble. Similar to Ubl12, degradation products were seen in eluted Ubl123 ~42 kDa (Figure 4-3A). Gel filtration using Sephacryl200 HiPrep 16/60 in 20 mM Tris pH7.5 167 mM NaCl, shows one major peak with peak broadening (Figure 4-3C). SDS-PAGE gel shows that these fractions were also highly degraded (Figure 4-3B). Again, gel filtration fractions that showed the least degradation were collected. Purified Ubl123 was further used for GST-pulldown assay, fluorescence polarization assay and protein crystallization trials.

The 6X-Histidine tag was also cleaved from Ubl123 using TEV protease. The lower molecular weight of Ubl123 indicated that the 6X-Histidine tag was cleaved (Figure 4-4A). His-cut Ubl123 was purified using affinity chromatography where Ubl123-Hiscut did not adhere to the column and was found in the binding and wash fractions. Gel filtration using Sephacryl100 in 20 mM Tris pH7.5, 500 mM NaCl, showed one major peak (Figure 4-4C).

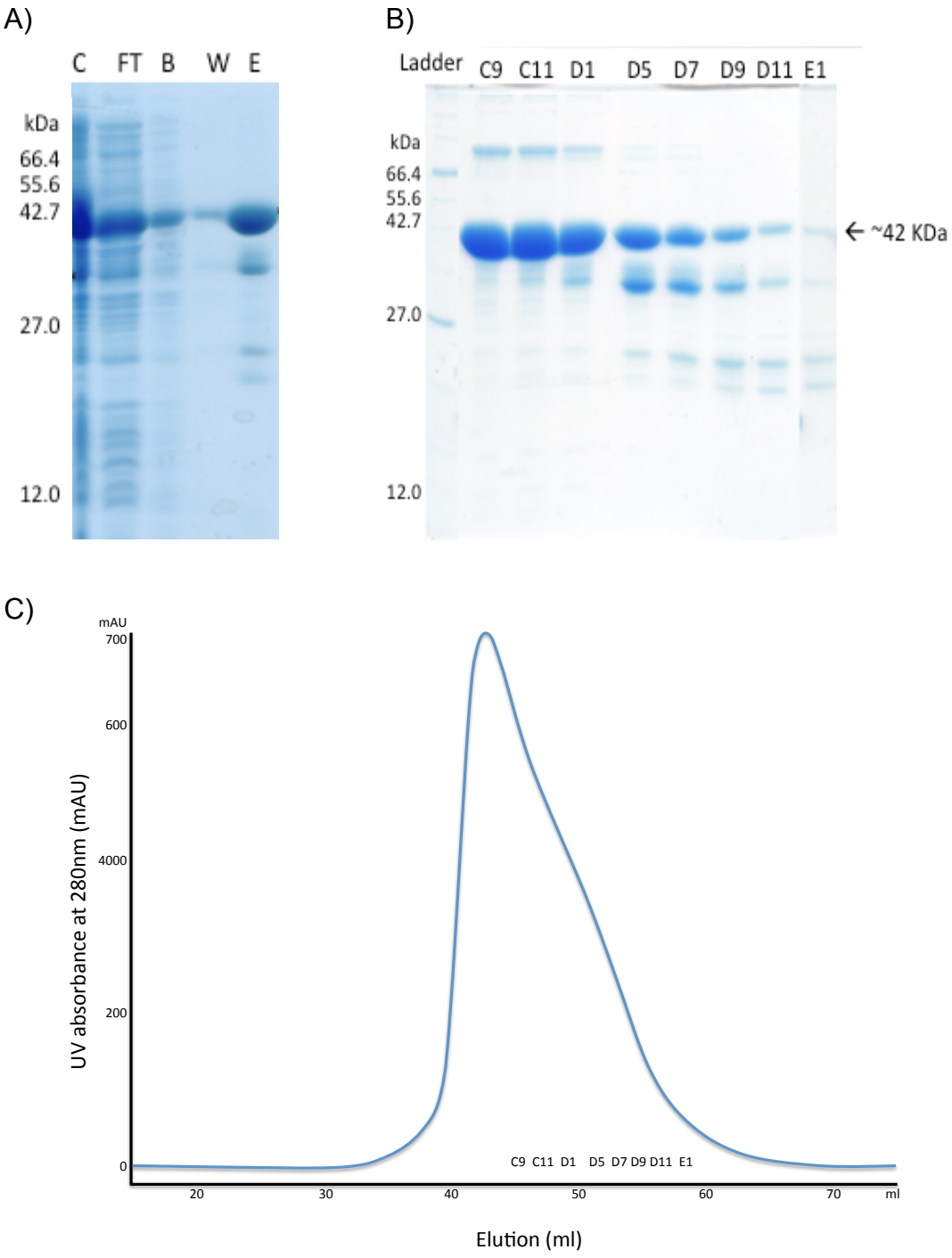



Figure 4-3: 6XHis-tagged USP7 535-888 (Ubl123) purification. Protein samples were run in a Coomassie stained SDS-PAGE gel. A) Ubl123 Ni-NTA affinity chromatography purification fractions: Crude (C), Flow through (FT), Binding (B), Wash (W), and Elute (E). Ubl123 eluted protein band ~42 kDa. B) Ubl123 size exclusion chromatography fraction samples. C) Ubl123 size exclusion chromatography profile.

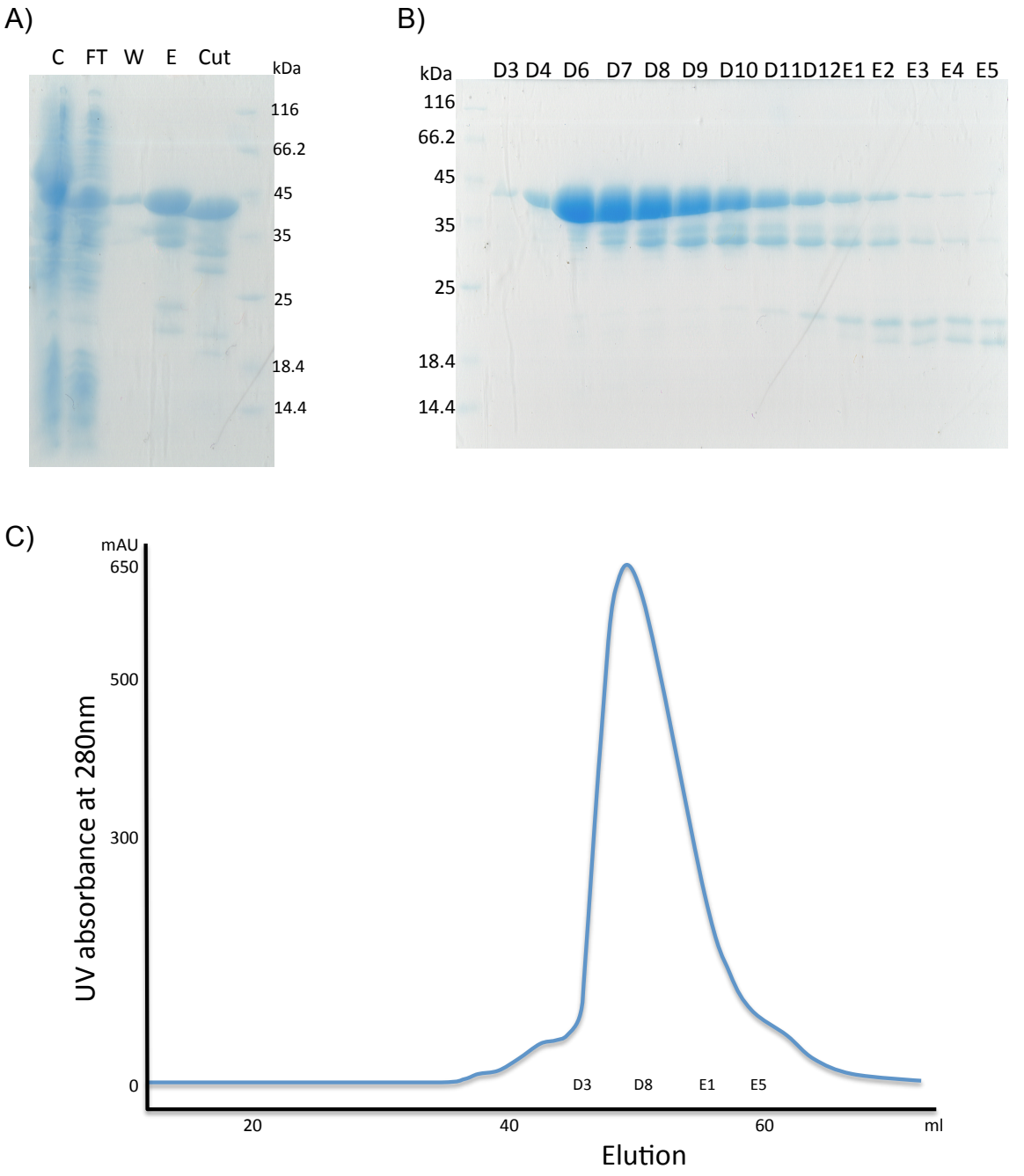



Figure 4-4: USP7 535-888 (Ubl123) 6XHis-tag cut purification. Protein samples were run in a Coomassie stained SDS-PAGE gel. A) Ubl123 Ni-NTA affinity chromatography purification fractions in comparison with overnight digestion of the 6X-His tag: Crude (C), Flow through (FT), Binding (B), Wash (W), Elute (E), and Ubl123 tag cleaved (Cut). Ubl123-Hiscut protein band lower than ~42 kDa. B) Ubl123-Hiscut gel filtration fraction samples. C) Ubl123-Hiscut size exclusion chromatography profile.

SDS-PAGE gel shows that these fractions were also degraded (Figure 4-4B). Fractions with the least degradation were pooled and used for protein co-crystallization trials except that the buffer used was 20 mM Tris pH7.5 167 mM NaCl.

Selenomethionine-labeled Ubl123 was also expressed and purified by Ni-NTA affinity and size exclusion chromatography. Collected protein samples were used for protein co-crystallization trials.

*Expression and purification of USP7 535-890 (Ubl123b)*

6XHis-tagged USP7 535-890 (Ubl123b) was purified by Ni-NTA affinity chromatography and size exclusion chromatography. Ubl123b was a modified Ubl123 constructs and also behaved similarly to Ubl123 with observable degradation patterns after purification. Protein eluted at ~44 kDa (Figure 4-5A). Gel filtration using Superdex200 HiPrep 16/60 in 20 mM Tris pH7.5 500 mM NaCl, showed one major peak with peak broadening (Figure 4-5C). SDS-PAGE gel also showed degradation from sampled fractions (Figure 4-5B). Samples collected were used for fluorescence polarization assay.

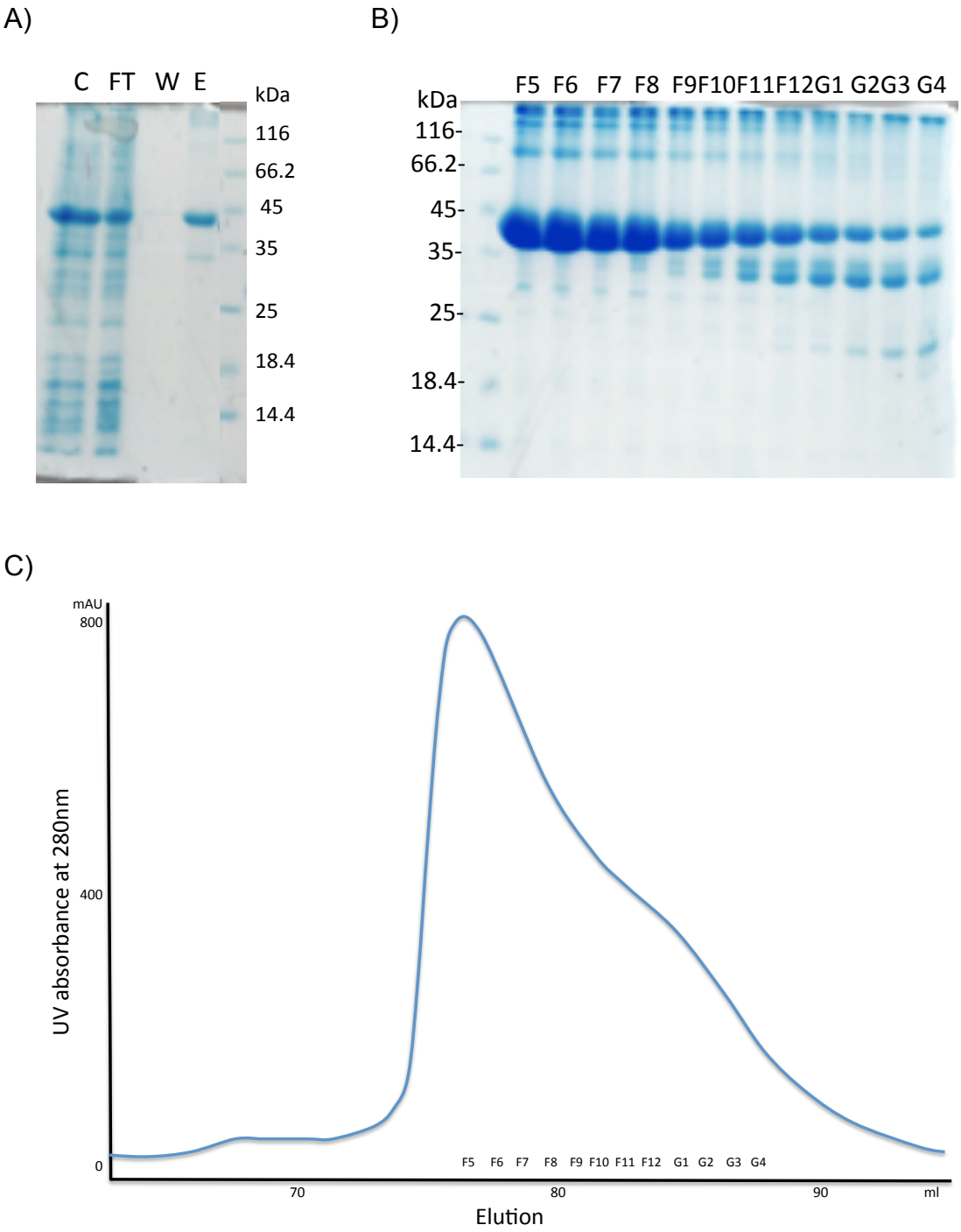



Figure 4-5: 6XHis-tagged USP7 535-890 (Ubl123b) purification. Protein samples were run in a Coomassie stained SDS-PAGE gel. A) Ubl123 Ni-NTA affinity chromatography purification fractions: Crude (C), Flow through (FT), Binding (B), Wash (W), and Elute (E). Ubl123b protein eluted at ~44 kDa. B) Ubl123b size exclusion chromatography fraction samples. C) Ubl123b size exclusion chromatography profile.

*Expression and purification of USP7 776-1102 (Ubl345)*

6XHis-tagged USP7 776-1102 (Ubl345) was purified by Ni-NTA affinity chromatography and size exclusion chromatography. Ubl345 was soluble and had medium level of protein expression. Ubl345 aggregated once purified. To minimize this, 5 mM β-mercaptoethanol (β-Me) (final concentration) was added to the purification buffers and gel filtration buffers. Protein eluted at ~41 kDa (Figure 4-6A). Size exclusion chromatography using Sephacryl S-100 in 20 mM Tris pH7.5, 500 mM NaCl, and β-Me, showed two peaks (Figure 4-6C). The left peak is the higher molecular weight peak for aggregated proteins while the right peak is the Ubl345 peak, shown in the fractions run on SDS-PAGE gel (Figure 4-6B). Collected samples were used for GST-pulldown assay, fluorescence polarization assay and protein crystallization trials. The buffer was also changed to 20mM Hepes pH 8.0 to help stabilize the protein at 4° C and this sample was used for protein crystallization. The 6X-Histadine was also cut off using TEV protease. This protein sample was used for protein crystallization. Similar to Ubl123, Ubl345 was labeled with Selenomethionine for protein crystallization.

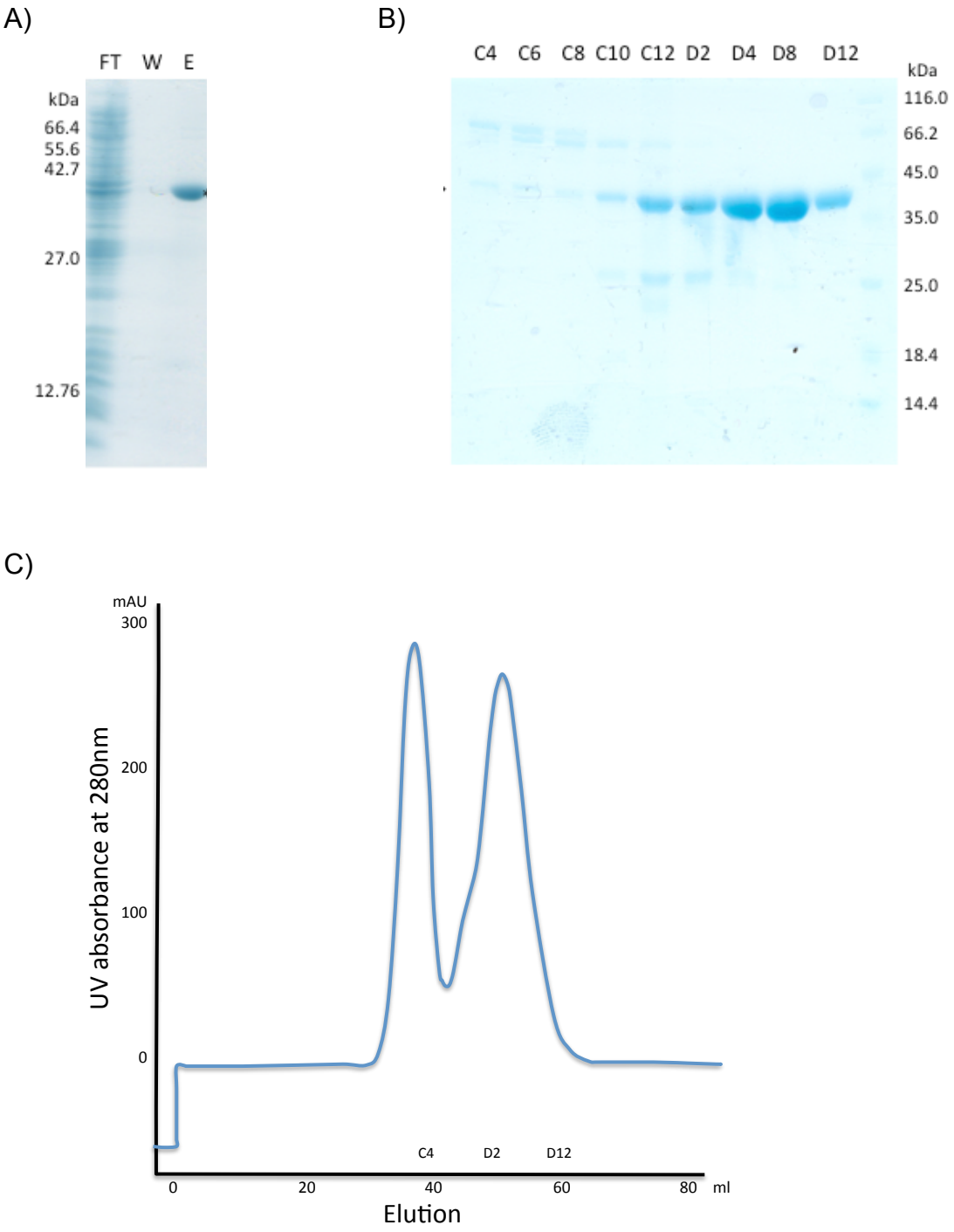
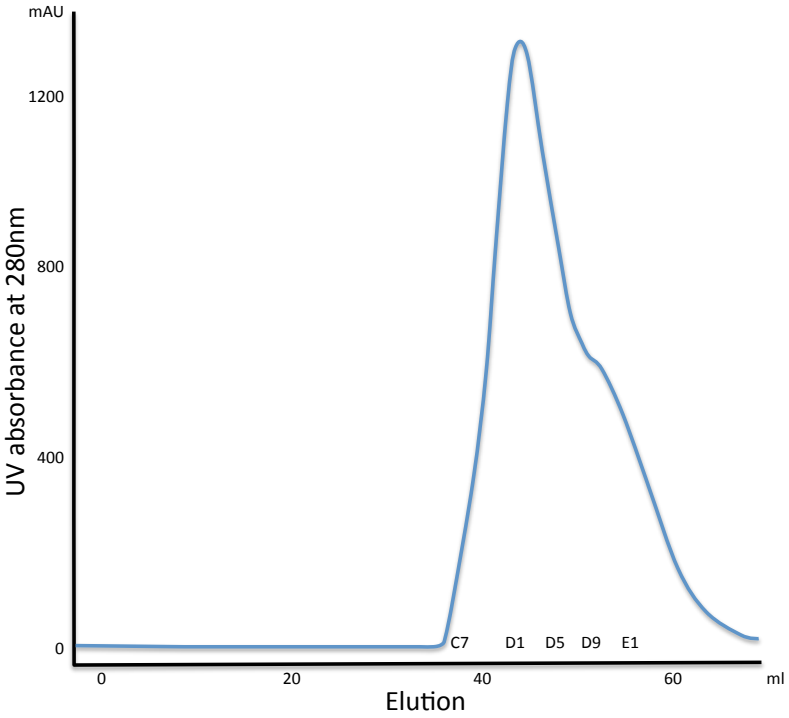



Figure 4-6: 6XHis-tagged USP7 776-1102 (Ubl345) purification. Protein samples were run in a Coomassie stained SDS-PAGE gel. A) Ubl345 Ni-NTA affinity chromatography purification fractions: Flow through (FT), Wash (W), and Elute (E). Ubl345 protein eluted at ~41 kDa. B) Ubl345 size exclusion chromatography fraction samples. C) Ubl345 size exclusion chromatography profile.

*Expression and purification of USP7 535-1102 (Ubl1-5)*

6XHis-tagged USP7 535-1102 (Ubl1-5) was purified by Ni-NTA affinity and size exclusion chromatography. Similar to Ubl123, Ubl1-5 tends to degrade. Gel filtration using Superdex100 in 20 mM Tris pH7.5 500 mM NaCl, show peak trailing which is a sign that the protein contains degradation products (Figure 4-7A). SDS-PAGE gel shows bands of smaller molecular weight than the Ubl1-5 band at ~69 kDa (Figure 4-7B). The least degraded samples were collected and used for GST-pulldown assay, fluorescence polarization assay and protein crystallization trials.

A)




B)

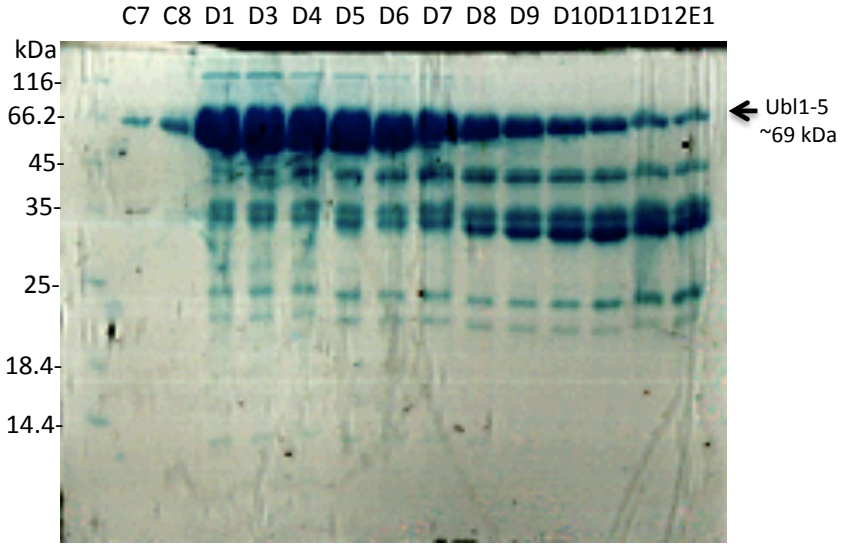



Figure 4-7: 6XHis-tagged USP7 535-1102 (Ubl1-5) purification. Protein samples were run in a Coomassie stained SDS-PAGE gel. A) Ubl1-5 size exclusion chromatography profile. B) Ubl1-5 size exclusion chromatography fraction samples. Ubl1-5 protein band at ~69 kDa.

### 4.2.2 GST-tagged ICP0 peptides

GST-pulldown assays done previously by Hong Zheng showed that GST-ICP0(B) $^{615}$PRGPRKCARKTRHAE$^{629}$ interacted with USP7 560-870 (Figure 1-12). The same ICP0 peptide construct was used to test interaction with the new USP7-Ubl constructs using GST-pulldown assay. ICP0 $^{615}$PRGPRKCARKTRHAE$^{629}$ peptide has 3 positively charged doublets $^{619}$RK$^{620}$, $^{623}$RK$^{624}$, and $^{626}$RH$^{627}$. From a previous study, mutating K620I and K624I from these doublets abolished USP7 binding *in vitro* (Appendix A)(Everett 1999). Using GST-pulldown assays, USP7-Ubl constructs and mutant GST-ICP0(B) mutants with mutations on either or both of the charged doublets, $^{619}$RK$^{620}$ and $^{623}$RK$^{624}$ were tested for interaction.

## 4.3 GST-pulldown assays

### 4.3.1 USP7 C-terminal constructs interaction with GST-ICP0(B) 615-629

Building on the new knowledge that the USP7 C-terminus has 5 Ubiquitin-like (Ubl) domains, the Ubl(s) which is/are important for ICP0 interaction were identified by repeating the GST-pulldown assay. The USP7-Ubl constructs (Ubl123, Ubl12, Ubl345, Ubl45a) were used in a, GST-pulldown assay using GST-ICP0(B) $^{615}$PRGPRKCARKTRHAE$^{629}$, from Figure 1-12, as bait (Figure 4-8). Full-length USP7 (FL) and full-length USP7 C-terminal (USP7-CTD) were used as positive controls while GST alone was used as a negative control. Both FL and USP7-CTD were seen in the eluate (E) indicating that they both interacted with GST-ICP0(B). USP7 N-terminus

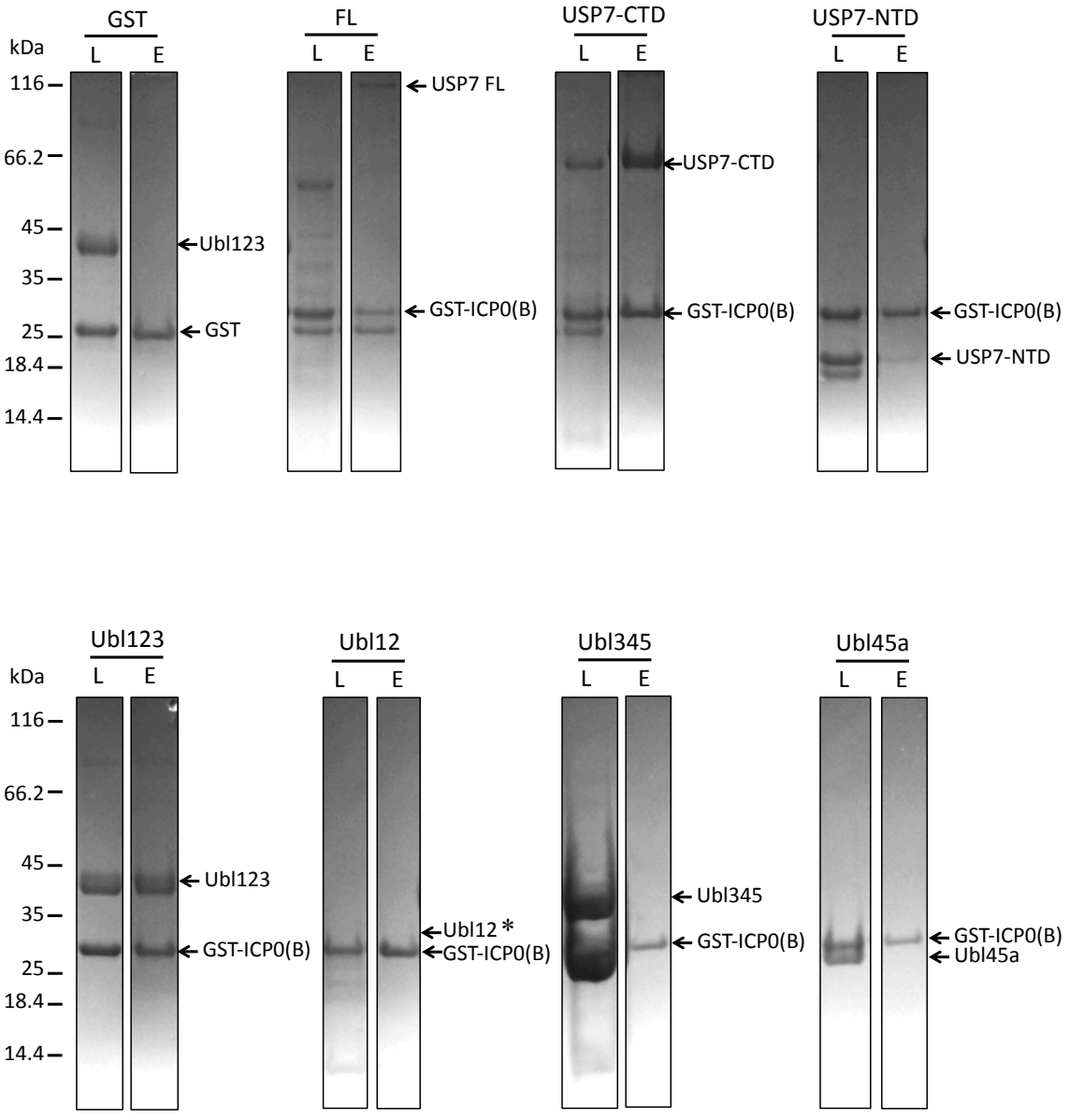



Figure 4-8: Interaction of USP7 constructs to GST-tagged ICP0 peptide (residues 615-629) in GST pull-down assays. Proteins that bound to GST-ICP0(B) charged glutathione-agarose beads were eluted with reduced glutathione and analyzed by Coomassie blue stained NuPAGE® Novex 4–12% Bis-Tris Midi Gel. The negative control was GST alone and the positive controls were full-length USP7 C-terminus (CT, aa. 535-1102) and full length (FL) His-tagged USP7. The column load (L) and eluate (E) are shown. GST ~26 kDa, GST-ICP0(B) ~27.6 kDa, Full length USP7 (FL) ~123 kDa, USP7 C-terminus (CT) ~69 kDa, Usp7 N-terminus (N) ~20 kDa, Ubl123 ~44 kDa, Ubl12 ~30 kDa, Ubl345 ~41 kDa, Ubl45a ~28.4 kDa. *Ubl12 and GST-ICP0(B) could not be separated because they run at the same molecular weight.

(USP7-NTD), Ubl12, Ubl 345, and Ubl45a were not found in the eluate indicating that these proteins were unable to interact with GST-ICP0(B). Ubl123 was observed in the eluate strongly suggesting that interaction with GST-ICP0(B) was contained within the first 3 Ubls. The interaction was specific between Ubl123 and GST-ICP0(B) as Ubl123 was not eluted when GST alone was used.

### 4.3.2 USP7 535-888 (Ubl123) interaction with GST-ICP0(B) 615-629 mutants

Using the results from the previous experiment (Figure 4-8), the identity of the ICP0 residues responsible for interaction with USP7 were identified. A similar study was done by Everett et al. (1999), using a GST-tagged ICP0 C-terminal construct 553-712 and single/double mutations on the three groups of positively charged doublets found between residues 619-627 (Appendix A). A more conservative mutagenesis approach was used whereby the charge was altered but the bulkiness or size of the residue was kept (Appendix A). Mutating the lysines at position 620 and 624 resulted in lack of binding by USP7 while mutating the arginines at position 619, 623, and 626 resulted in reduced USP7 binding (Everett 1999). Only two of the three doublets, residues $^{619}$RK$^{620}$ and $^{623}$RK$^{624}$, were mutated in ICP0(B) $^{615}$PRGP**RK**CA**RK**TRHAE$^{629}$ (Table 4-3). A more aggressive approach to the mutations was undertaken where the charged and bulky amino acids, arginine and lysine, were mutated to alanine which is small and uncharged. USP7 535-888 (Ubl123) which showed strong interaction with ICP0(B) 615-629 was used with the ICP0(B) 615-629 mutants in a GST-pulldown assay (Figure 4-9). In the pulldown assay, GST was used as a negative control and the eluates were shown for the mutant

peptides. There was a lack of Ubl123 binding with mutants K1, K2, K1K2, R1K1 and R1K1R1K2 while mutants R1, R2 and R1R2 showed Ubl123 binding equivalent to wild type (WT) ICP0(B) peptide (E). Table 4-3 shows a summary of the GST-pulldown assay and the location of each mutated amino acid.

Table 4-3: Summary of USP7 535-888 (Ubl123) binding to GST-tagged ICP0 mutant peptides (aa. 615-629). Amino acids were mutated to alanine.

| Mutation | ICP0 peptide sequence | USP7 535-888 binding |
|---|---|---|
| WT | 615 619 620 623 624 629<br>P R G P R K C A R K T R H A E | + |
| R1 | P R G P A K C A R K T R H A E | + |
| K1 | P R G P R A C A R K T R H A E | -- |
| R2 | P R G P R K C A A K T R H A E | + |
| K2 | P R G P R K C A R A T R H A E | -- |
| R1R2 | P R G P A K C A A K T R H A E | + |
| K1K2 | P R G P R A C A R A T R H A E | -- |
| R1K1 | P R G P A A C A R K T R H A E | -- |
| R1K1R2K2 | P R G P A A C A A A T R H A E | -- |

(+) binding (--) lack of binding

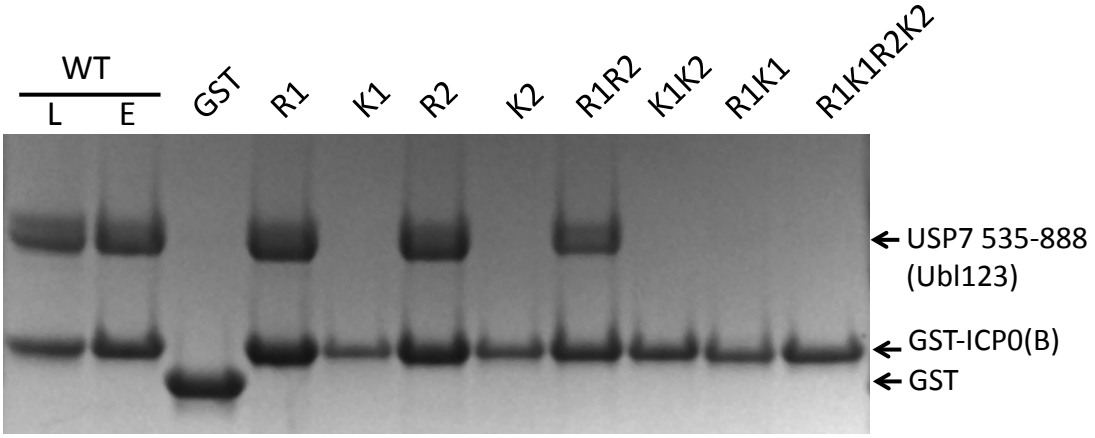



Figure 4-9: Interaction of USP7 535-888 (Ubl123) to GST-tagged ICP0(B) mutant peptide (residues 615-629) in GST pull-down assays. USP7 535-888 (His-Ubl123) that bound to GST-ICP0(B) mutant peptide charged Glutathione-Sepharose beads were eluted with reduced glutathione and analyzed by Coomassie stained NuPAGE® Novex 4–12% Bis-Tris Midi Gel. The negative control was GST alone and the positive control was wild-type (WT) ICP0 peptide (residues 615-629). The column load (L) and eluate (E) for the WT are shown while the rest are all eluates.

## 4.4 Fluorescence polarization assay of USP7 constructs with FITC-ICP0 617-627 peptide

The interaction between two proteins can be quantified and represented by its dissociation constant, $K_D$. The protein-protein interaction can be strong, low $K_D$ (~nM) or the interaction can be weak, high $K_D$ (>100 μM). A way to measure $K_D$ is through fluorescence polarization. In fluorescence polarization, a probe (ie. fluorescein isothiocyanate (FITC) tagged peptide) with a specific excitation and emission wavelength is excited with monochromatic light. When these peptide probes are excited in solution, the peptide tumbles quickly due to its size and the emitted light is more scattered. The detected polarization value of the peptide is low. When an interacting protein is added with the peptide, the excited probe tumbles less and the emitted light is more focused. The detected polarization value is higher. When the polarization value of the probe is plotted against an increasing concentration of interacting protein, the $K_D$ value can be calculated.

The GST-pulldown of USP7 constructs with GST-ICP0(B) $^{615}$PRGPRKCARKTRHAE$^{629}$ peptide showed qualitatively that FL USP7, Ubl1-5 and Ubl123 were able to bind to the GST-ICP0(B) peptide. Ubl123 interaction with FITC-ICP0 $^{617}$GPRKCARKTRH$^{627}$ was quantified using fluorescence polarization. Fluorescence polarization allows the quantification of protein binding to a fluorophore-labeled peptide through excitation-emission of the fluorophore and rotation of the peptide.

Fluorescein isothiocyanate (FITC) a fluorophore is attached to the ICP0 peptide $^{617}$GPRKCARKTRH$^{627}$, which is excited by polarized light. Polarization occurs when the excited FITC-peptide emits light in the same direction as the excitation light. However, since peptides are small molecules, they tumble a lot in solution. The emitted light is no longer parallel to the excitation light and is depolarized. Since the quantity measured for fluorescence polarization assay is polarized light, free peptides have low polarization value. When a peptide binds to protein making the peptide heavier, the peptide tumbles less and would have a higher polarization value. The concentration of the labeled peptide was kept constant at 40 nM and increasing protein was added (concentration of up to 200 μM) shown in the x-axis. Protein binding to the peptide is measured in terms of millipolarization (mP), shown in the y-axis. Measurement of $K_D$ of USP7 constructs to FITC-labeled ICP0 peptide was measured through binding saturation.

ICP0 binds to the C-terminal of USP7 and the binding has been narrowed down to Ubl123 through GST-pulldown assay. Ubl123, Ubl123b and Ubl1-5 have $K_D$ values ranging from 0.36 to 1.65 μM (Table 4-4). Full-length USP7's dissociation constant was 7.54 μM and, was higher than the USP7 C-terminal constructs. Addition of USP7 N-terminal with ICP0 peptide resulted to negative mP values indicating that ICP0 does not interact with USP7-NTD. Looking at the data points for Ubl345, there was no clear saturation even though the program tried to fit a saturation curve (Figure 4-21). The fluorescence polarization assay showed that Ubl45a was binding to ICP0 peptide with a dissociation constant of 13.98 μM possibly suggesting a second binding site with a much lower affinity.

Table 4-4: Summary of USP7 constructs' average dissociation constant ($K_D$) to FITC-ICP0 617-627

| Construct | $K_D$ (μM) | SEM |
|---|---|---|
| FL USP7 | 7.54 | ±0.96 |
| USP7-CTD (Ubl1-5) | 0.86 | ±0.05 |
| USP7-NTD | -- | -- |
| Ubl12 | 3.78 | ±0.26 |
| Ubl45a | 13.98 | ±1.21 |
| Ubl123 | 1.65 | ±0.25 |
| Ubl123b | 0.36 | ±0.04 |
| Ubl345 | -- | -- |

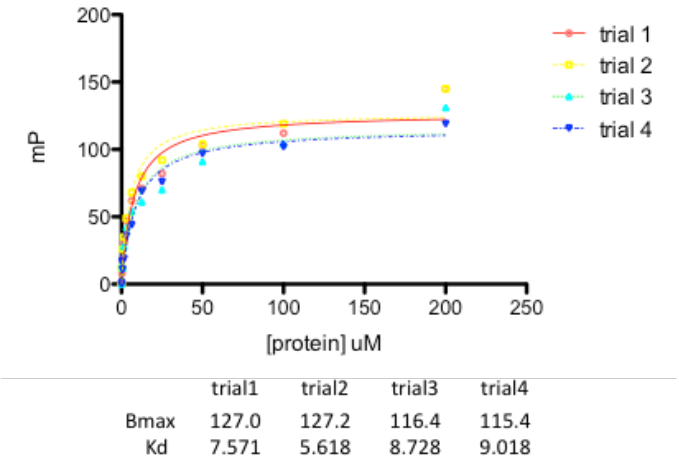
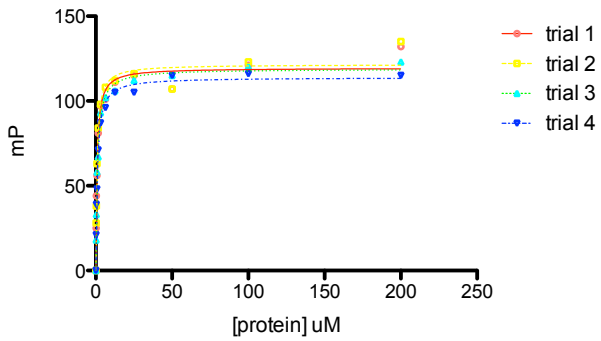
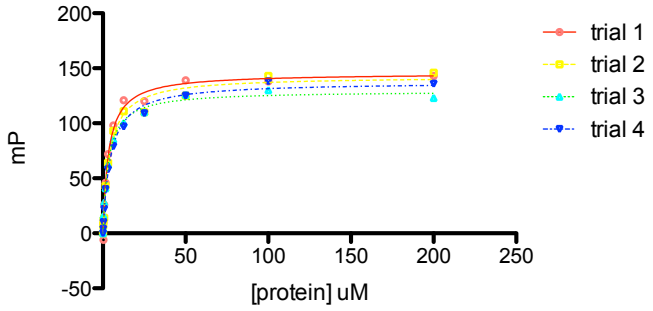



Figure 4-10: Binding affinity of full length USP7 (FL) to FITC-ICP0 peptide. Fluorescence polarization (mP) was plotted as a function of protein concentration using nonlinear regression. The concentration of the FITC-ICP0 peptide was 40 nM. The data were fitted to a one-site specific binding model using GraphPad Prism version 5.0. $K_D$ ranged from 5.6-9.0 μM.



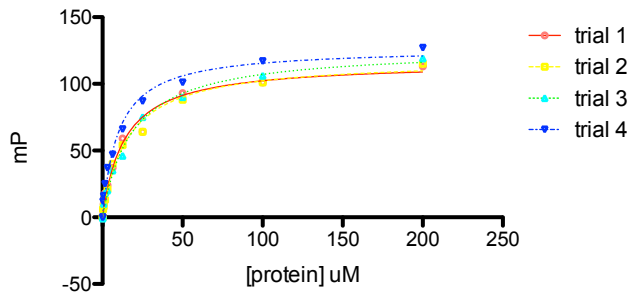

| | trial1 | trial2 | trial3 | trial4 |
|---|---|---|---|---|
| Bmax | 119.5 | 121.5 | 118.9 | 113.9 |
| Kd | 0.7812 | 0.7514 | 1.001 | 0.9463 |

Figure 4-11: Binding affinity of USP7 535-1102 (Ubl1-5) or full length USP7 C-terminal (CT) to FITC-ICP0 peptide. Fluorescence polarization (mP) was plotted as a function of protein concentration using nonlinear regression. The concentration of the FITC-ICP0 peptide was 40 nM. The data were fitted to a one-site specific binding model using GraphPad Prism version 5.0. $K_D$ ranged from 0.8-1.0 μM.



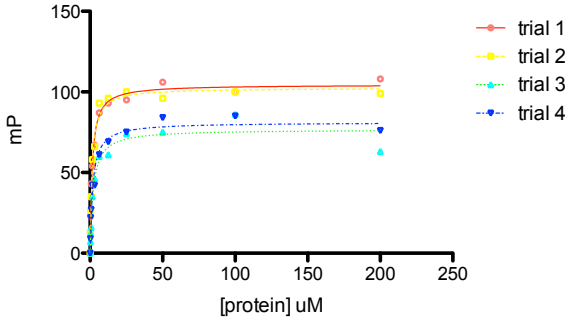

| | trial1 | trial2 | trial3 | trial4 |
|---|---|---|---|---|
| Bmax | 145.5 | 142.6 | 129.4 | 137.5 |
| Kd | 3.422 | 4.009 | 3.324 | 4.563 |

Figure 4-12: Binding affinity of USP7 535-776 (Ubl12) to FITC-ICP0 peptide. Fluorescence polarization (mP) was plotted as a function of protein concentration using nonlinear regression. The concentration of the FITC-ICP0 peptide was 40 nM. The data were fitted to a one-site specific binding model using GraphPad Prism version 5.0. $K_D$ ranged from 3.3-4.6 μM.



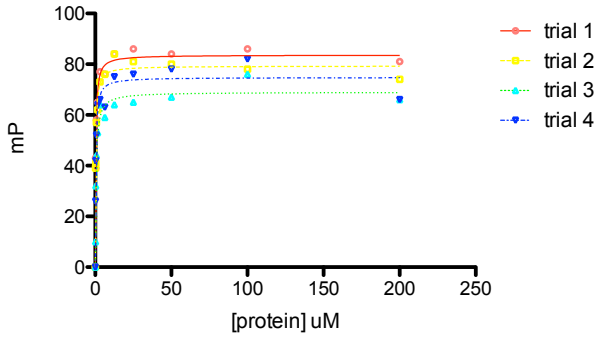

| | trial1 | trial2 | trial3 | trial4 |
|---|---|---|---|---|
| Bmax | 116.4 | 118.4 | 126.7 | 127.0 |
| Kd | 13.74 | 15.72 | 18.23 | 9.910 |

Figure 4-13: Binding affinity of USP7 884-1102 (Ubl45a) to FITC-ICP0 peptide. Fluorescence polarization (mP) was plotted as a function of protein concentration using nonlinear regression. The concentration of the FITC-ICP0 peptide was 40 nM. The data were fitted to a one-site specific binding model using GraphPad Prism version 5.0. $K_D$ ranged from 9.9-18.2 μM.



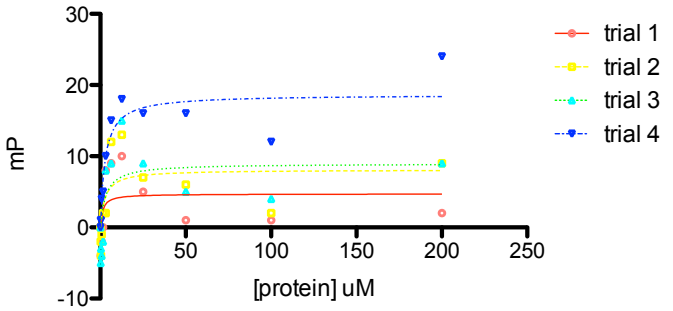

| | trial1 | trial2 | trial3 | trial4 |
|---|---|---|---|---|
| Bmax | 104.5 | 102.4 | 76.72 | 81.10 |
| Kd | 1.399 | 1.296 | 2.106 | 1.774 |

Figure 4-14: Binding affinity of USP7 535-888 (Ubl123) to FITC-ICP0 peptide. Fluorescence polarization (mP) was plotted as a function of protein concentration using nonlinear regression. The concentration of the FITC-ICP0 peptide was 40 nM. The data were fitted to a one-site specific binding model using GraphPad Prism version 5.0. $K_D$ ranged from 1.3-2.1 μM.




| | trial1 | trial2 | trial3 | trial4 |
|---|---|---|---|---|
| Bmax | 83.63 | 79.29 | 69.00 | 74.80 |
| Kd | 0.3182 | 0.2898 | 0.5259 | 0.3363 |

Figure 4-15: Binding affinity of USP7 535-890 (Ubl123b) to FITC-ICP0 peptide. Fluorescence polarization (mP) was plotted as a function of protein concentration using nonlinear regression. The concentration of the FITC-ICP0 peptide was 40 nM. The data were fitted to a one-site specific binding model using GraphPad Prism version 5.0. $K_D$ ranged from 0.3-0.5 μM.




| | trial1 | trial2 | trial3 | trial4 |
|---|---|---|---|---|
| Bmax | 4.708 | 8.093 | 8.958 | 18.65 |
| Kd | 1.382 | 2.695 | 3.235 | 2.733 |

Figure 4-16: Binding affinity of USP7 776-1102 (Ubl345) to FITC-ICP0 peptide. Fluorescence polarization (mP) was plotted as a function of protein concentration using nonlinear regression. The concentration of the FITC-ICP0 peptide was 40 nM. The data were fitted to a one-site specific binding model using GraphPad Prism version 5.0.

## 4.5 Protein crystallization

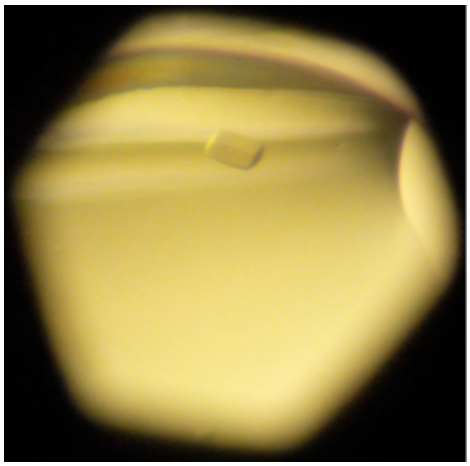
Table 4-5: USP7 constructs set-up for crystal trials.

| Constructs | Amino Acids | Expression | Soluble | Crystallization Screen | Crystals | Crystal Refinement | Diffracted |
|---|---|---|---|---|---|---|---|
| Ubl1-5 | 535-1102 | High | Yes | Yes | | | |
| Ubl123 | 535-888 | High | Yes | Yes | Yes | | Yes |
| Ubl12 | 535-776 | High | Yes | Yes | | | |
| Ubl2345 | 680-1102 | Low | Yes | | | | |
| Ubl345 | 776-1102 | High | Yes | Yes | Yes | Yes | Yes |
| Ubl45 | 886-1102 | High | Yes | Yes | | | |
| Ubl23 | 680-888 | Low | | | | | |
| Ubl2 | 680-776 | Low | Yes | | | | |
| Ubl3 | 776-888 | Low | Yes | | | | |

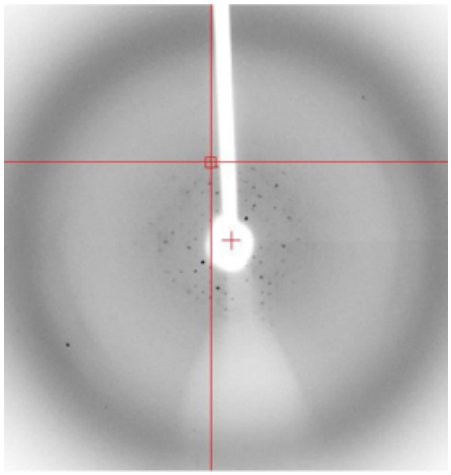
### 4.5.1 USP7 535-888 (Ubl123) protein crystallization

Ubl123 has a tendency to degrade after purification. Crystallization trials were set-up as soon as possible once the proteins were pooled after size exclusion chromatography. Protein concentrations of ~15 mg/ml and ~31 mg/ml were plated with QIAGEN crystallization suites JCSGI-IV and JSCG+ and incubated at 4° C. Crystals appeared 1-3 weeks after plating in JCSGI-G3: 1.0 M Lithium chloride, 0.1 M Citric acid pH 5.0, 10% (w/v) PEG6000 (final pH5) (Figure 4-17). Crystals from this condition diffracted at ~9 Å (Figure 4-22B). However, these crystals could not be refined.

A)



B)



C)

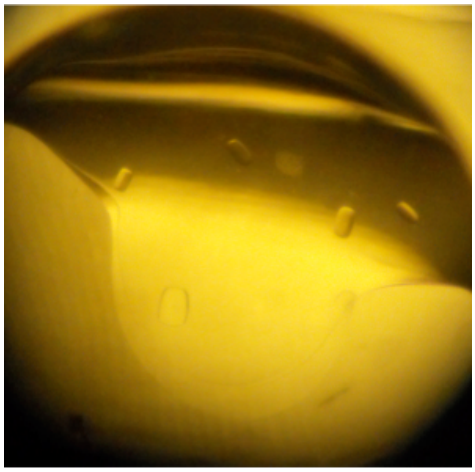


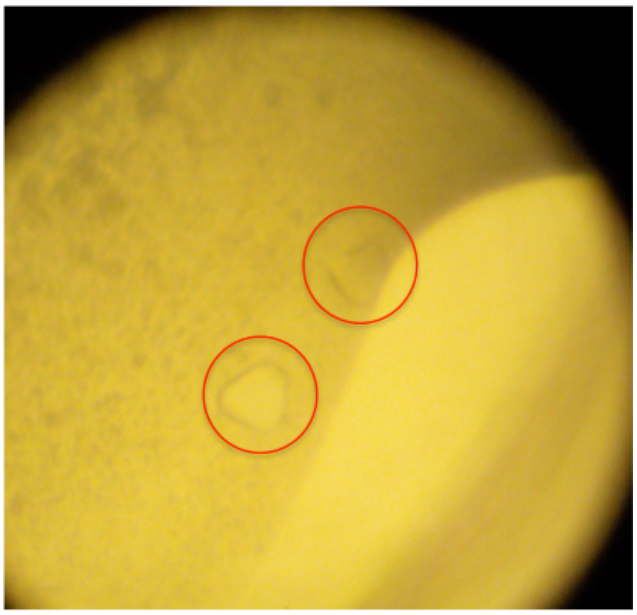
Figure 4-17: USP7 535-888 (Ubl123) protein crystals. A) Ubl123 crystal in initial crystallization hit condition: 1.0 M Lithium chloride, 0.1 M Citric acid pH 5.0, 10% (w/v) PEG6000 (final pH 5), ~15 mg/ml, 4° C. B) Ubl123 crystal diffracted at ~9 Å. C) Ubl123 crystal in initial crystallization hit condition: 1.0 M Lithium chloride, 0.1 M Citric acid pH 5.0, 10% (w/v) PEG6000 (final pH 5), ~31 mg/ml , 4° C.

### 4.5.2 USP7 776-1102 (Ubl345) protein crystallization

Ubl345 aggregates after purification. To minimize aggregation, β-mercaptoethanol was added to buffers and proteins were used for crystallization trials as soon as possible after size exclusion. Protein concentration of ~23 mg/ml was plated with QIAGEN crystallization suites JCSGI-IV and JSCG+ and incubated at 4° C. Crystals were discovered after three months in JCSGIII-F10: 0.2 M Potassium/Sodium tartrate, 0.1 M Sodium citrate pH 5.6, 2 M Ammonium sulfate (Figure 4-18A). Crystals from this condition diffracted at ~9 Å (Figure 4-18B). This crystallization condition was then further refined.

Crystallization condition refinement efforts for Ubl345 crystal included changing the salt concentration, crystal seeding, pH, protein:crystallization buffer ratio, additives screening, cutting the 6x-Histadine tag and Selenomethionine labeling. Increasing the salt concentration from 0.2 M to 0.27 M Potassium/Sodium tartrate resulted in a lot of tiny crystals or microcrystals (Figure 4-19A). Thus, the salt concentration was kept at 0.2 M Potassium/Sodium tartrate. Next, buffer pH ranging from 4 to 12 was tested and 0.1 M Sodium citrate pH 6 gave the biggest crystals. These same crystals were then crushed and used as seed crystals. The resulting crystals were fewer and bigger (Figure 4-19B). Since the crystals grow in 4°C, crystal refinement plates were set-up at 4°C to minimize the temperature fluctuation in the protein drop.

A)



B)

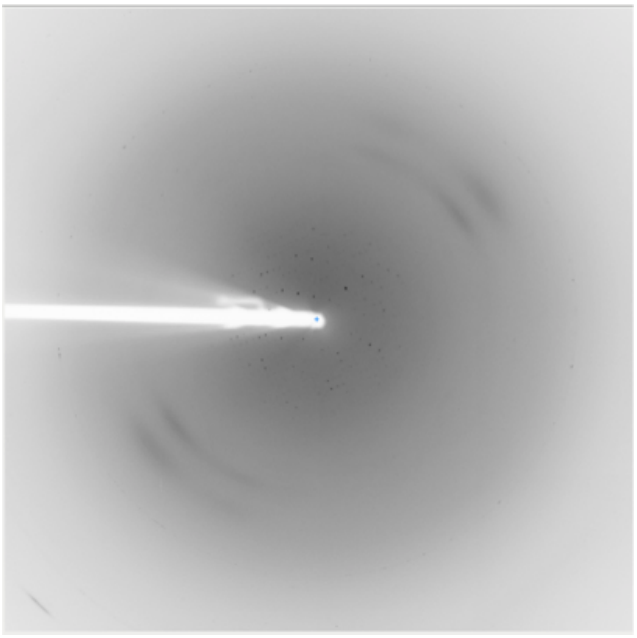


Figure 4-18: USP7 776-1102 (Ubl345) protein crystals. A) Ubl345 in initial crystallization hit condition: 0.2M Potassium/Sodium tartrate, 0.1M Sodium citrate pH 5.6, 2M Ammonium sulfate, ~23mg/ml protein, 4° C. B) Ubl345 crystal diffracted at ~7 Å.

In addition, protein to crystallization buffer ratio normally kept at 1:1 was changed to 0.75 µl of protein and 1 µl of crystallization buffer in order to dilute the protein concentration. This also resulted in fewer but bigger crystals (Figure 4-19C). Additives were added to the protein:crystallization buffer drop during one of the refinements. Crystals the formed in the drop where 0.01 M GSH (L-Glutathione reduced), GSSG (L-Glutathione oxidized), an additive, was added to a final concentration of ~ 0.001 M gave improved diffraction, ~3.5-4 Å, (Figure 4-19D) in comparison to the diffraction of Ubl345's original crystal, ~7 Å, (Figure 4-19B). Ubl345 has a 6X-Histadine tag at its N-terminus. The tag was removed and the protein plated for crystallization trials. When the histadine tag was cut, crystals appeared in JCSGIII-F10: 0.2 M Potassium/Sodium tartrate, 0.1 M Sodium citrate pH 5.6, 2 M Ammonium sulfate, ~23 mg/ml protein, 4° C (Figure 4-20A). This was also the same condition where His-Ubl345 crystal formed. When this condition was refined to 0.2 M Potassium/Sodium tartrate, 0.1 M Sodium citrate pH 5.8, 2 M Ammonium sulfate, the protein crystals diffracted at ~4 Å (Figure 4-20B). Selenomethionine labeled His-Ubl345 formed crystals in condition 0.2 M Potassium/Sodium tartrate, 0.1 M Sodium citrate pH 6, 2 M Ammonium sulfate, ~17 mg/ml, 4° C (Figure 4-20C).

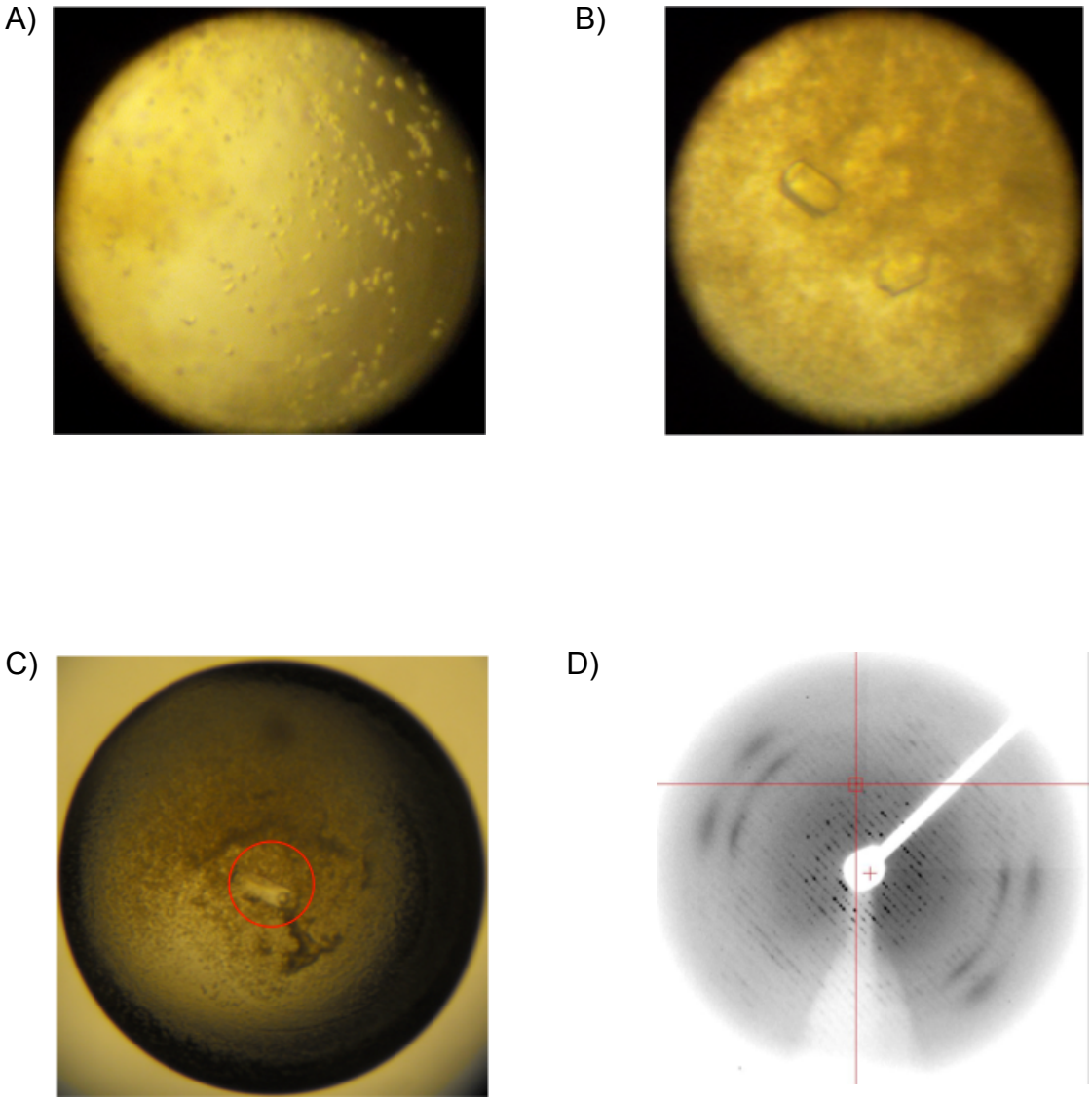



Figure 4-19: USP7 776-1102 (Ubl345) refinement crystals. A) Higher salt concentration, 0.27 M Potassium/Sodium tartrate, gave microcrystals. B) Changing the pH to 6 and seeding gave fewer and bigger crystals. C) Plating in 4°C and changing protein:crystallization buffer ratio resulted to bigger crystals. D) Adding GSH (L-Glutathione reduced), GSSG (L-Glutathione oxidized) improved crystal diffraction to ~3.5-4 Å.

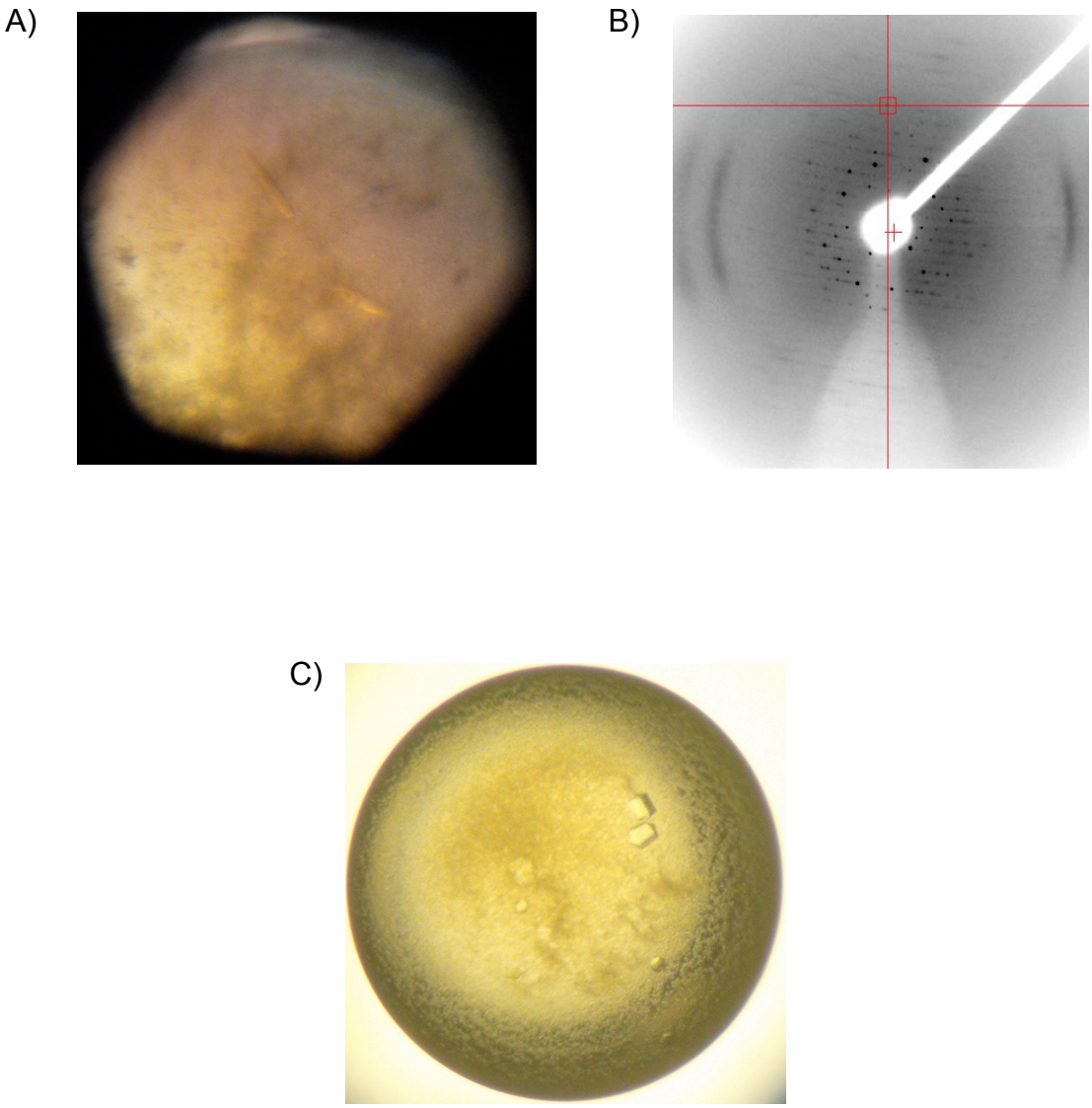



Figure 4-20: His-tagged cut Ubl345 and Selenomethionine labeled Ubl345 protein crystals. A) Crystals in crystallization hit condition JCSGIII-F10: 0.2 M Potassium/Sodium tartrate, 0.1 M Sodium citrate pH 5.6, 2 M Ammonium sulfate, ~23 mg/ml protein, 4° C. B) Refined His-cut Ubl345 crystal in 0.2 M Potassium/Sodium tartrate, 0.1 M Sodium citrate pH 5.8, 2 M Ammonium sulfate diffracted at ~4 Å. C) Selenomethionine labeled Ubl345 protein crystals in 0.2 M Potassium/Sodium tartrate, 0.1 M Sodium citrate pH 6, 2 M Ammonium sulfate, ~17 mg/ml, 4° C .

### 4.5.3 USP7 535-888 (Ubl123) with ICP0 617-627 peptide co-crystallization

The GST-pulldown assay and fluorescence polarization assay results, mentioned earlier, indicated that Ubl123 forms a tight interaction with GST-ICP0 615-629 peptide and FITC-ICP0 617-627, respectively. In order to determine the mechanism of their interaction, Ubl123 needed to be co-crystallized with ICP0 $^{617}$GPRKCARKTRH$^{627}$ peptide. Ubl123 has a tendency to degrade after purification. Crystallization trials were set-up as soon as possible once the proteins samples were pooled after size exclusion chromatography. Ubl123 was concentrated to ~40 mg/ml. ICP0 peptide powder was added to Ubl123 at a 1:2 molar ratio. Because the ICP0 peptide contained a cysteine, DTT to a final concentration of 5 mM was added to the protein:peptide solution. The protein:peptide mixture was plated with QIAGEN crystallization suites JCSGI-IV, JSCG+, PACT, PEGsII and ClassicsII and incubated at room temperature. Crystals appeared after three days in about seventy-four crystallization conditions (Appendix C). Figure 4-21 shows some of the crystals formed.

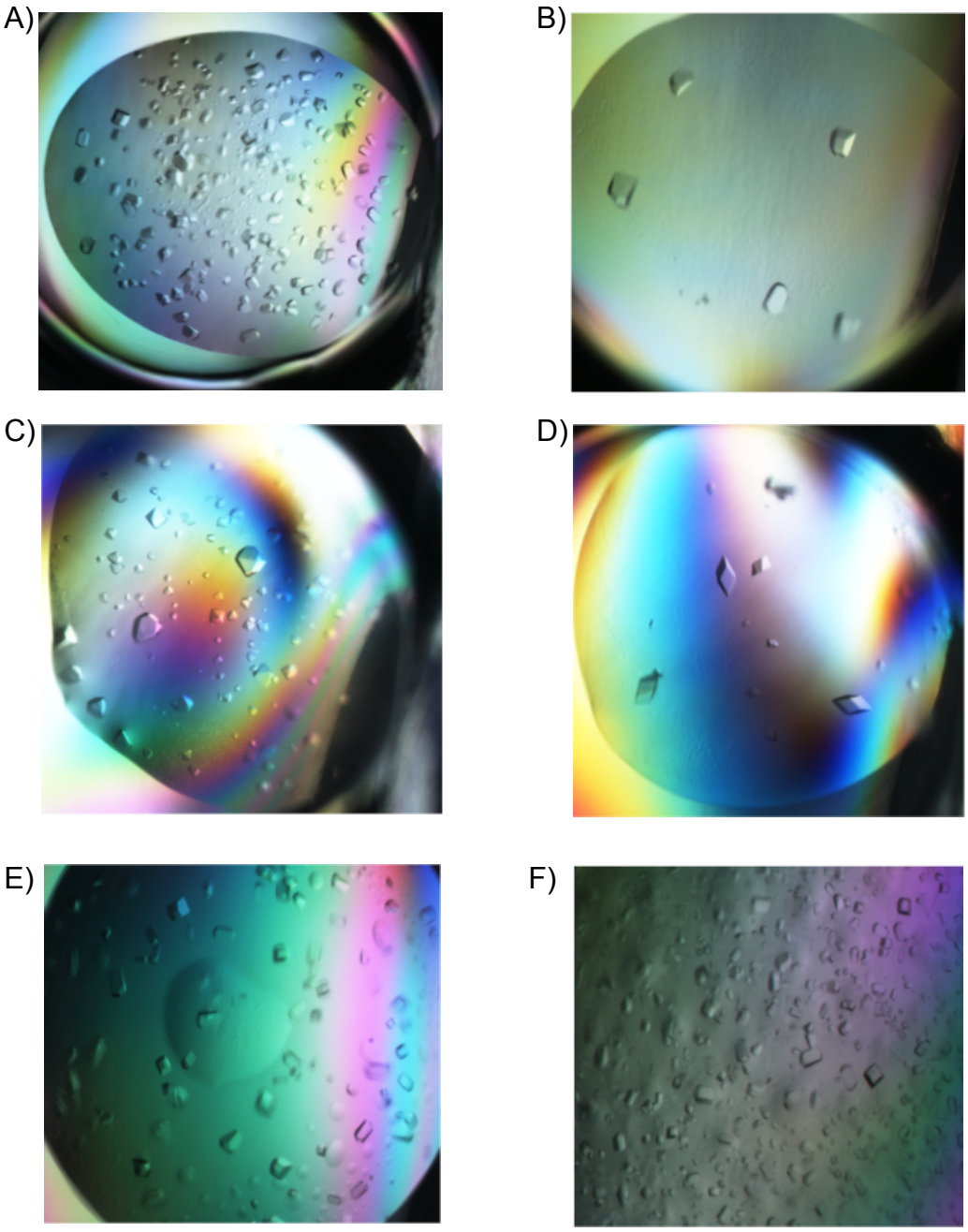



Figure 4-21: Ubl123-ICP0 co-crystals initial crystallization hit condition. Ubl123 and ICP0 were mixed at a 1:2 molar ratio using eight crystallization suites and incubated at room temperature. A) JCSGIII-D8: 0.1 M HEPES pH 6.5, 5% (w/v) PEG 6000. B) JCSGIII-B1: 0.2 M Magnesium chloride, 0.1 M Tris pH 8.5 20% (w/v) PEG 8000. C) ClassicsII-H4: 0.2 M Ammonium citrate pH 7.0, 20 % (w/v) PEG 3350. D) JCSGI-F8: 0.1 M Sodium citrate pH 5.5, 20 % (w/v) PEG 3000. E) JCSGII-C10: 0.1 M HEPES pH 6.5, 10% (w/v) PEG 6000. F) PEGsII-E4: 0.1 M HEPES pH 7.5, 10 % (w/v) PEG 4000, 5 % (v/v) Isopropanol.

Since there were so many conditions containing crystals, conditions were chosen that produced few large crystals. One of the conditions refined was PEGsII-H4: 4% (w/v) PEG 8000. Different molecular weight of PEGs (Polyethylene glycol) ranging from PEG 1000-20000 and different additives available in the lab (ie, isopropanol, ethanol, glycerol, etc.) were tried. One of the conditions forming crystals during refinement was in 4% (w/v) PEG 6000 (Figure 4-22A). Many hexagonal crystals formed from this refinement. However, the diffraction quality was poor. Another condition that was refined was JCSGI-F8 0.1 M Sodium citrate pH 5.5, 20% (w/v) PEG 3000. For this condition, the pH, percentage of PEG, used PEG with different molecular weight and added additives were refined. Crystals from this refinement were bigger and had more of a football shape to them (Figure 4-27B-F). The diffraction quality of the crystals was better. In particular, a crystal from condition 0.1 M Sodium citrate pH 6, 16% (w/v) PEG 3000, additive: 0.1M L-proline (amino acid) (Figure 4-22B) diffracted ~ 2.85 Å (Figure 4-23). Diffraction data was collected from the crystal.

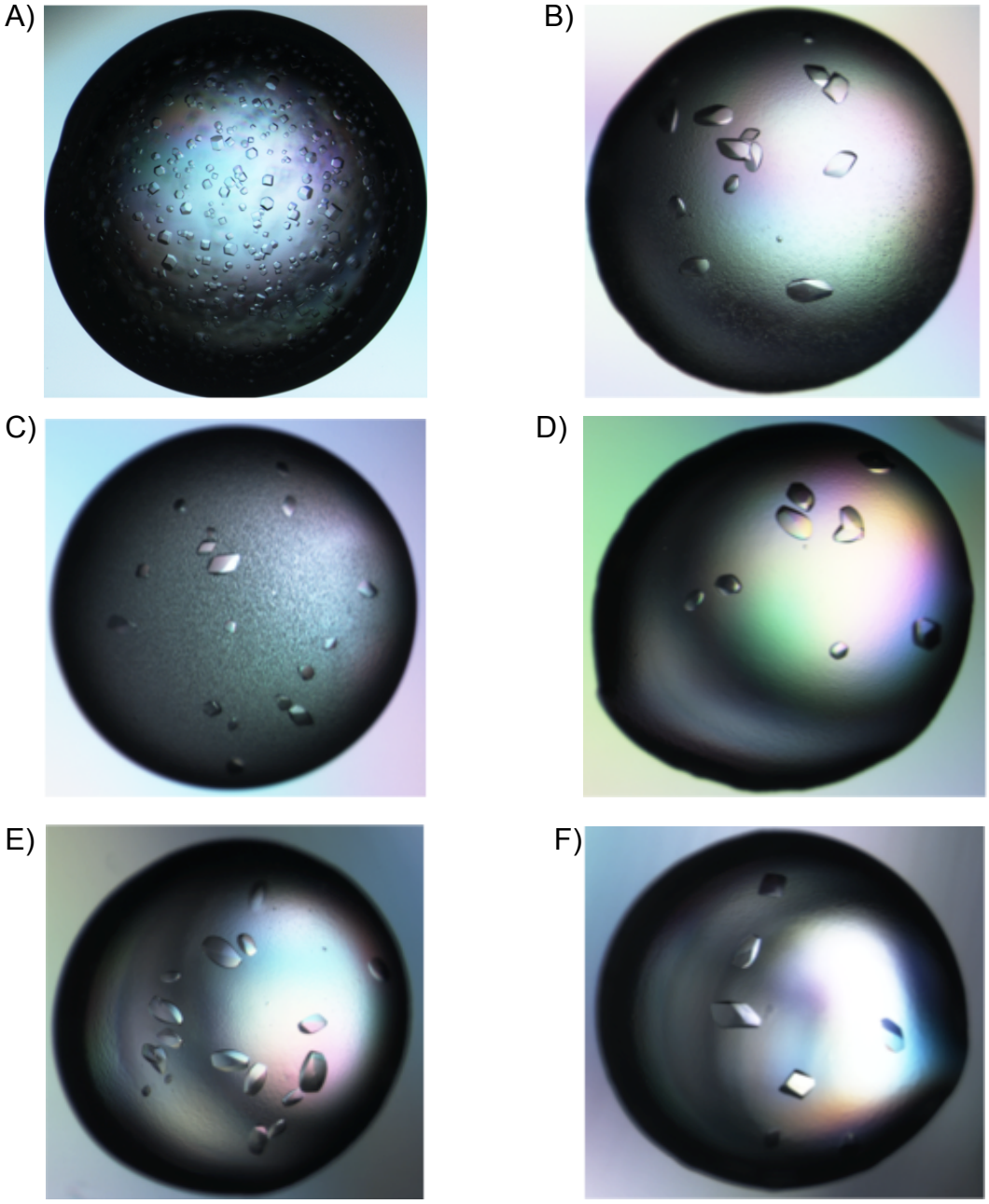



Figure 4-22: Ubl123-ICP0 617-627 co-crystals refinement. Ubl123 and ICP0 617-627 were mixed at a 1:2 molar ratio and incubated at room temperature. A) 4% (w/v) PEG 6000. B) 0.1 M Sodium citrate pH 6, 16% (w/v) PEG 3000, 5% (v/v) Propylene glycol. C) 0.1 M Sodium citrate pH 6, 16% (w/v) PEG 3000, 5% (v/v) Isopropanol. D) 0.1 M Sodium citrate pH 6, 16% (w/v) PEG 3000, additive: 0.1M L-proline (amino acid). E) 0.1 M Sodium citrate pH 6, 16% (w/v) PEG 3000, additive: 0.1 M sodium bromide (dissociating agent). F) 0.1 M Sodium citrate pH 6, 16% (w/v) PEG 3000, additive: 0.1 M trimethylamine HCl (chaotrope).

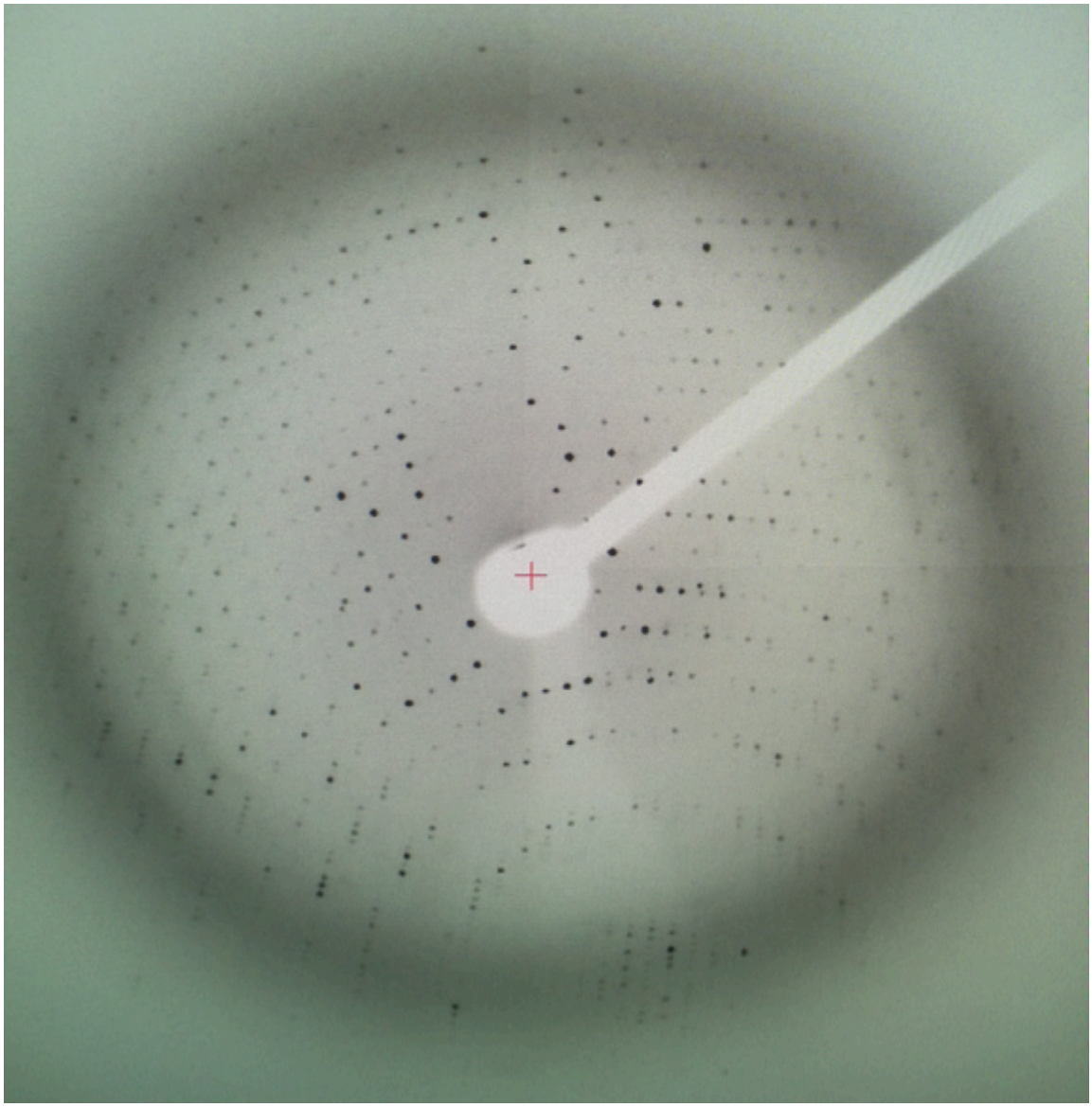


Figure 4-23: Ubl123-ICP0 617-627 co-crystals diffraction. The crystals were refined at 0.1 M Sodium citrate pH 6, 16% (w/v) PEG 3000, additive: 0.1M L-proline (amino acid), room temperature. It diffracted ~ 2.85 Å and diffraction data was collected.

Co-crystallization trials of selenomethionine-labeled Ubl123 and ICP0 617-627 peptide were also set-up. Selenomethionine easily gets oxidized therefore DTT was added to the protein:peptide mixture to a final concentration of 5 mM. The protein:peptide mixture was plated with QIAGEN crystallization suites JCSGI, JCSG+, PEGsII, PACT and ClassicsII. Crystals formed in several conditions (Appendix C) and some are shown in Figure 4-24. Two of the conditions refined were PEGsII-E1: 0.1 M Tris pH 8.5, 0.8 M LiCl, 8% (w/v) PEG 4000 and JCSGI-E3: 0.1 M Sodium/Potassium phosphate pH 6.4, 0.2 M NaCl, 10% (w/v) PEG 8000. During refinement, pH, salt concentration and PEG concentration were changed. Figure 4-25 shows some of the crystals formed during refinement. The crystals diffracted at ~5-7 Å.

Surprisingly, when excess Selenomethionine-labeled Ubl123:ICP0 617-627 mixture was left in a microfuge tube overnight at room temp, crystals formed inside the microfuge tube (Figure 1-26). It turns out that the protein:peptide mixture was able to crystallize in its own buffer, 167 mM NaCl, 20 mM Tris pH 7.5, 5 mM β-mercaptoethanol. These crystals were then refined using an additive screen, which then produced big crystals (Figure 1-27). Diffraction data were collected from these crystals, however, the structure could not be determined.

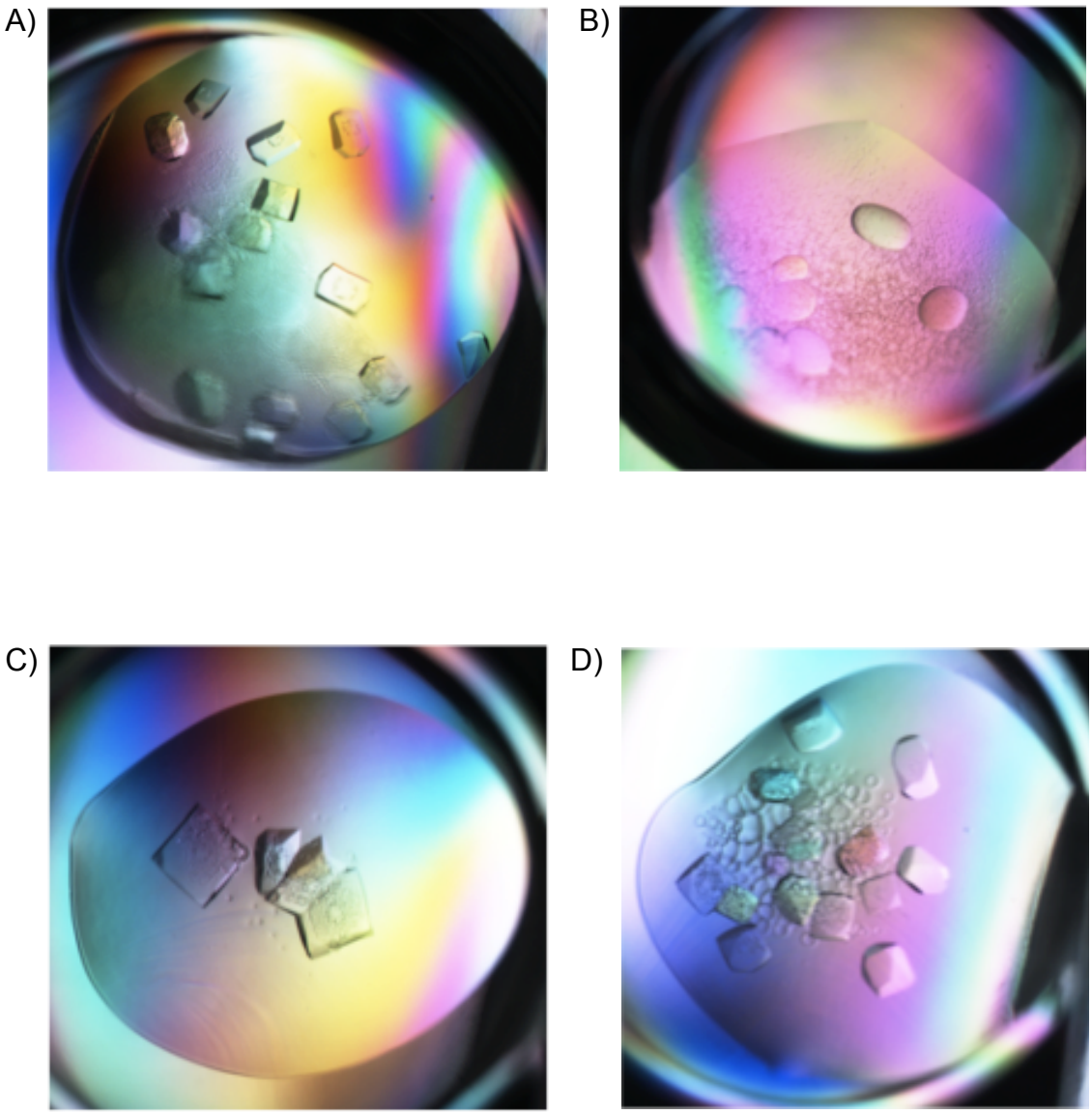



Figure 4-24: Selenomethionine-labeled Ubl123-ICP0 617-627 co-crystals initial crystallization hit condition. Sel-Ubl123 and ICP0 617-627 were mixed at a 1:2 molar ratio and incubated at room temperature. A) ClassicsII-F3 0.064 M Sodium citrate pH 7.0, 0.1 M HEPES pH 7.0, 10% (w/v) PEG 5000 MME. B) ClassicsII-C7 0.8 M Sodium/Potassium tartrate, 0.1 M TRIS pH 8.5, 0.5% (w/v) PEG 5000 MME. C) ClassicsII-C4 0.56 M Sodium citrate pH 7.0. D) ClassicsII- B10 0.8 M Succinic acid pH 7.0.

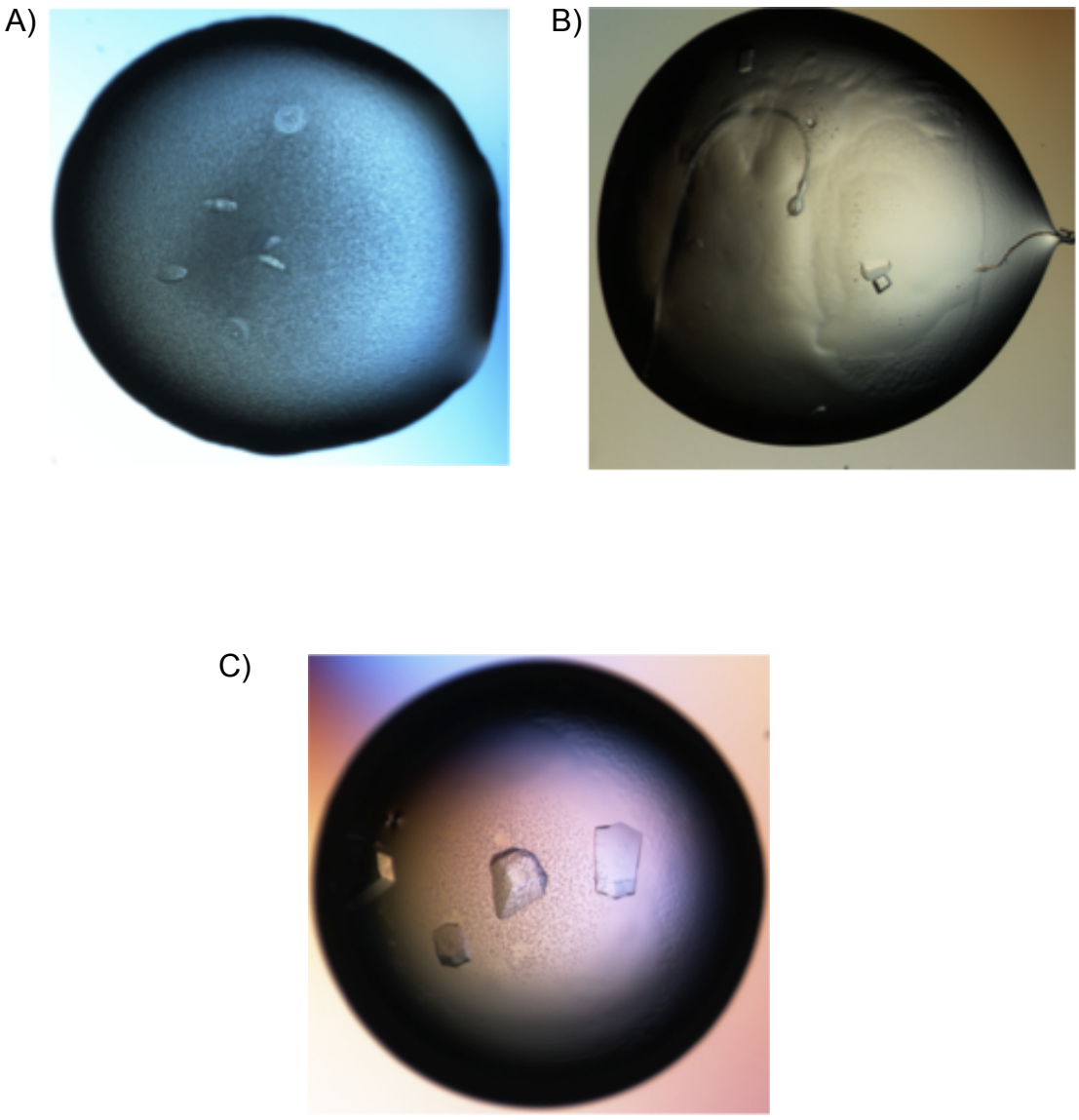



Figure 4-25: Selenomethionine-labeled Ubl123-ICP0 617-627 co-crystals refinement. Sel-Ubl123 and ICP0 617-627 were mixed at a 1:2 molar ratio and incubated at room temperature. A) 0.1 M TRIS hydrochloride pH 8.4, 0.4 M LiCl, 8% (w/v) PEG 8000. B) 0.1 M Sodium cacodylate trihydrate pH 6.4, 0.2 M NaCl, 8% (w/v) PEG 8000. C) 0.1 M Sodium cacodylate trihydrate pH 6.4, 0.2 M NaCl, 4% (w/v) PEG 8000.

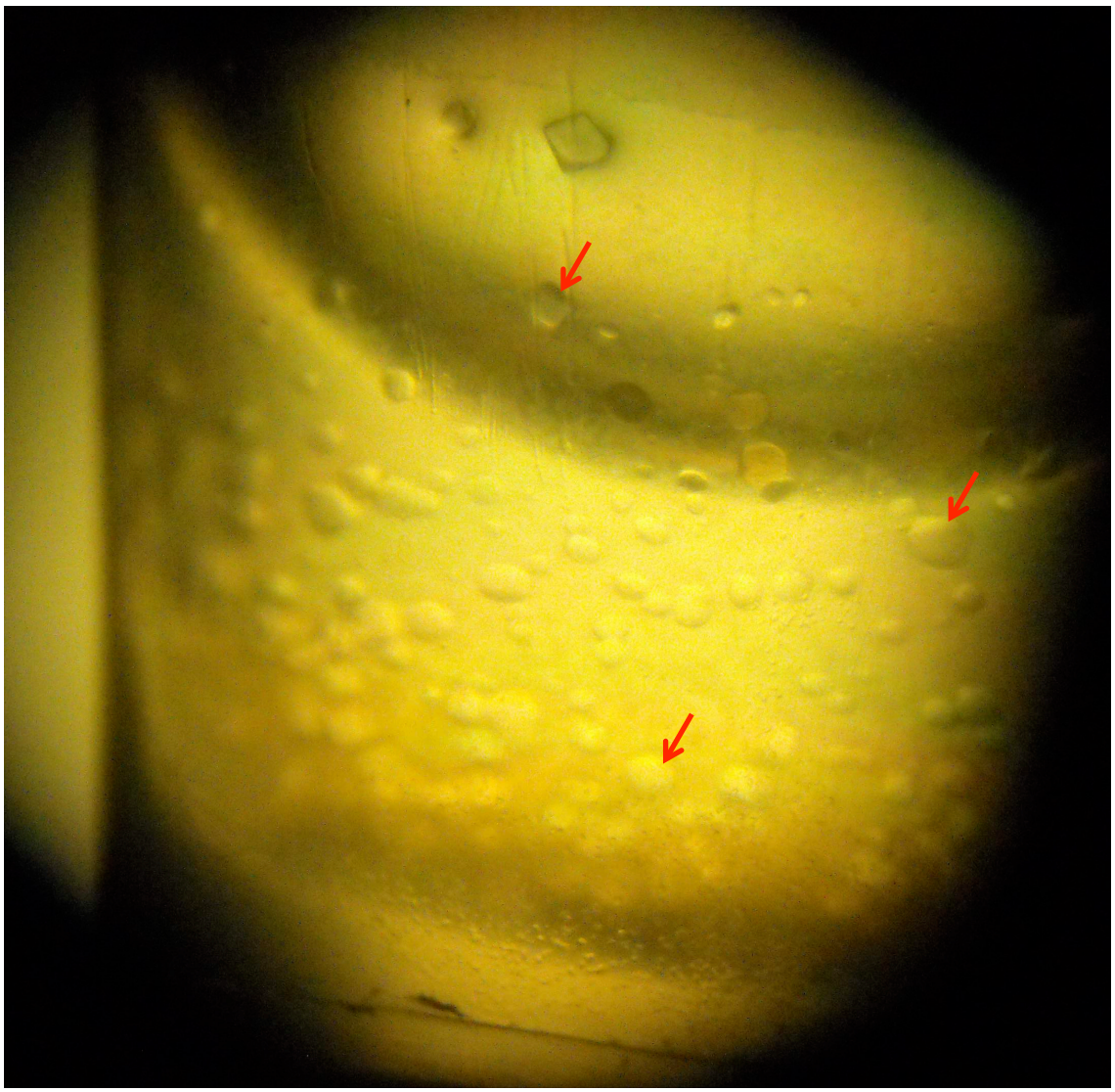


Figure 4-26: Selenomethionine-labeled Ubl123-ICP0 617-627 co-crystallized in a microfuge tube in 167 mM NaCl, 20 mM Tris pH 7.5, 5 mM β-mercaptoethanol. Crystals (arrows) were found in the microfuge tube that was left in room temperature overnight.

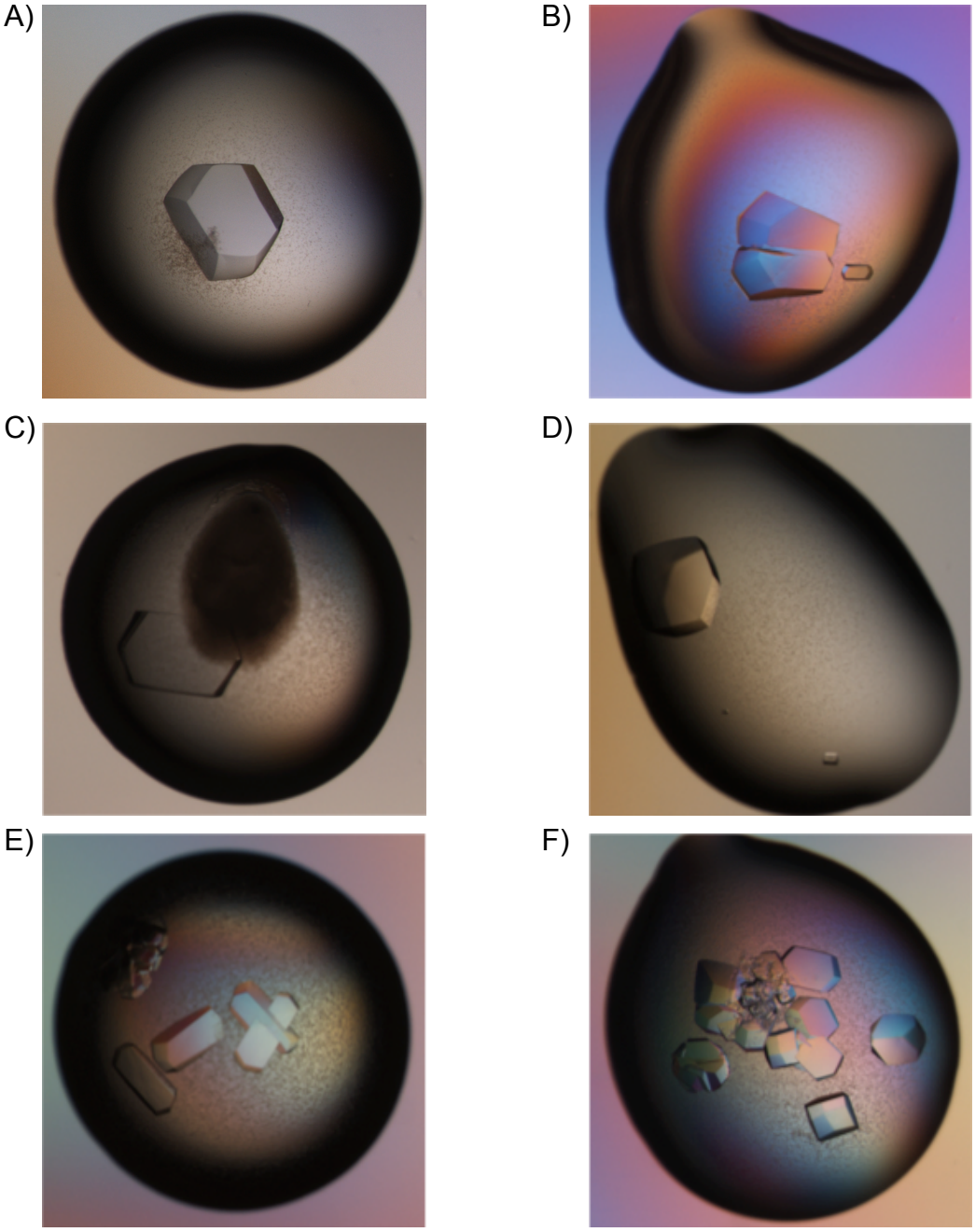



Figure 4-27: Selenomethionine-labeled Ubl123-ICP0 617-627 microfuge tube crystals additive refinement. Ubl123 and ICP0 617-627 were mixed at a 1:2 molar ratio and incubated at room temperature in 167 mM NaCl, 20 mM Tris pH 7.5, 5 mM β-mercaptoethanol. 0.1 µl of additive was added to each Ubl123-ICP0 617-627 0.9 µl drop. Additives used included: A) 0.1 M Phenol B) 1.0 M Glycine C) 0.1 M Betaine Hydrochloride D) 0.1 M Sarcosine E) 1.0 M Guanidine HCl F) 0.01 M GSH, 0.01 M GSSG.

# CHAPTER 5: DISCUSSION

The focus of this thesis was to determine the biochemical basis of interaction between USP7 and ICP0. Results from these experiments reinforced that the ICP0 binding site in USP7 is within residues 535-889, which is composed of three ubiquitin-like domains, Ubl123. Ubl123 strongly interacted with the ICP0 peptide and the dissociation constant for the interaction was 1.3-2.1 μM. In addition, the lysine residues 620 and 624 within the USP7 binding domain of ICP0 were important for this interaction. Mutating these residues abolished the interaction between USP7 Ubl123 and ICP0 peptide $^{615}$PRGPRKCARKTRHAE$^{629}$. Putative Ubl123:ICP0 co-crystals were obtained permitting the future structure determination of the complex.

## 5.1 Protein Expression and Purification

Unlike the USP7 N-terminal (USP7-NTD) and catalytic (USP7-CAT) domains, the USP7 C-terminus (USP7-CTD) is not as well studied functionally or structurally. Bioinformatics analysis predicted that USP7-CTD contains 4-5 Ubiquitin like (Ubl) domains with the characteristic ββαββ fold [57]. An NMR structure of USP7-CTD residues 560-610 confirmed that the predicted Ubl fold for these residues (Figure 1-9). Conducting our own secondary structure prediction of the USP7-CTD (residues 535-1102) using PHYRE, confirmed that there were indeed 5 potential Ubl domains in the C-terminus (Figure 4-1). Varying number and lengths of Ubl domains were cloned into an *E. coli* expression vector in order to express and purify recombinant proteins which were later

used for different biochemical analysis and crystallization trials (Table 4-1). All but one of the recombinant proteins was insoluble, Ubl23 (residues 680-888) (Table 4-2). More than 60% of the constructs had medium to high expression levels. All of the soluble proteins were purified using a combination of nickel affinity and size exclusion chromatography with very good yields sufficient for downstream interaction and crystallization experiments. Some of the recombinant proteins either degraded or aggregated over time suggesting lower stability.

Recombinant protein expression and purification is one of the first steps when studying the biochemistry and determining the three dimensional structure of a protein *in vitro*. Thus, it is very important that the recombinant protein is soluble and produces a good yield. When overexpressing eukaryotic proteins in *E. coli*, solubility is a common problem which could be affected by different factors: what tag (His- or GST-) was fused to the protein, the placement of the tag (N-terminus or C-terminus of the protein), host cell, protein expression induction temperature, size and construct design.

Insoluble protein or inclusion bodies can be due to misfolding from random disulfide bond formation, and exposed hydrophobic groups in the surface of the protein [120]. As we are also dealing with isolated protein domains, the, USP7-CTD Ubls might not be very stable and might require interaction with the surrounding Ubls. Trial and error with multiple construct design helped with issues of construct boundaries.

The recombinant USP7-CTD constructs were expressed as hexahistadine (6XHis-) tagged fusion proteins at the N-terminus. Sometimes, N-terminal or C-terminal placement of the tag helps with protein solubility. Also, the type of tag used can help as well.

Glutathione S-transferase (GST), a very soluble enzyme, can be used as a tag to increase solubility. The Ubl constructs were expressed in BL21(DE3)mgk strain of *E. coli*. Bacteria have their own codon usage [121]. Since these constructs are of human origin, there might be codons that are rare in *E. coli* that could impact solubility and expression [121]. However, BL21(DE3)mgk's mgk plasmid should be able to supplement for the rare tRNAs. The mgk plasmid codes for 3 rare tRNAs (AGG and AGA for Arginine and ATA for Isoleucine). Co-expression with molecular chaperones to help with posttranslational folding can also enhance solubility [122]. Reducing the induction temperature can help with increasing solubility by reducing the frequency of molecular collision and hydrophobic interactions [123].

Degradation of recombinant proteins is another common problem in protein purification. *E. coli* has many proteases that can cleave expressed recombinant proteins [124]. The strain, BL21, is genetically engineered with reduced proteases [124]. The addition of protease inhibitors can help inhibit proteases released from the bacterial cellular compartments. Lowering the temperature of bacterial cultures during protein expression induction and purification buffers can help further reduce protease activity [124].

## 5.2 GST-Pulldown assays

One of the objectives of this thesis was to identify which amino acids and the corresponding Ubl domain(s) in USP7 are required for ICP0 interaction. Initial studies showed that Herpes Simplex virus 1 (HSV-1) viral protein, infected cell protein 0 (ICP0), co-immunoprecipitated with USP7 from HSV-1 infected-cell extract [40,42]. The USP7

interaction domain in ICP0 was then narrowed down to residues 594-632 through *in vitro* binding assays [99]. Within this interaction domain, residues 618-632 contained the essential residues for binding [99]. The ICP0 interaction domain of USP7 was narrowed down to a protease resistant domain within residues 622-801 [58]. Results from these experiments were consistent with the GST-pulldown assay done by a previous graduate student, Hong Zheng. She identified the minimal ICP0 peptide needed to interact with USP7-CTD (residues 560-870) as $^{615}$PRGPRKCARKTRHAE$^{629}$ (Figure 1-12). Unfortunately, the boundaries of this USP7-CTD construct were not ideal as it did not contain all of the Ubl3 residues since it ended at 870 rather than 885. However the identification of which Ubls interacted with ICP0 was still unknown. GST-pulldown assays showed that Ubl123 (residues 535-888) was able to interact with the GST-ICP0 $^{615}$PRGPRKCARKTRHAE$^{629}$ peptide whereas Ubl345 was unable to interact with this same GST-ICP0. Site directed mutagenesis was used to probe the interaction between Ubl123 and ICP0 $^{615}$PRGPRKCARKTRHAE$^{629}$. The importance of the charged residues was investigated by mutating each arginine and lysine to alanine. The two lysine residues, K620 and K624, in ICP0 are absolutely required for interaction with USP7 as mutating either one or both of these lysine residues resulted in loss of Ubl123 binding. However mutation of one or both arginine residues, R619 and R623, did not abolish binding. Combining Everett's (1999) results with mutants K620I and K624I (Appendix A) and our results from the GST-pulldown assays, suggested that ICP0 positions 620 and K624 must be positively charged in order to interact with USP7. As mentioned by Everett et al. (1999), the amino acid sequence 615-629 of HSV-1 ICP0 is also conserved with its HSV-

2 homologue (amino acid sequence alignment shown in Appendix B). This shows the importance of this stretch of amino acids for HSV-1 ICP0/USP7 interaction.

## 5.3 Fluorescence polarization

Using GST pulldown assays it was shown that ICP0 peptide $^{615}$PRGPRKCARKTRHAE$^{629}$ bound to Ubl123 (residues 535-888) with close to a 1:1 interaction. Fluorescence polarization was then used to determine the dissociation constant between various Ubls and a fluorescein isothiocyanate (FITC)-tagged ICP0 peptide $^{617}$GPRKCARKTRH$^{627}$. The calculated dissociation constants ($K_D$) for the different constructs ranged from 0.3-18.2 μM to no binding (Table 4-4). Ubl123, Ubl123b and Ubl1-5 showed strong interaction with FITC-ICP0. Ubl123b (residues 535-889) had the strongest interaction at 0.3-0.5 μM. Full length USP7 should a much lower dissociation constant than Ubl123b even though the full-length protein has the C-terminal. Conformation of the full-length protein in solution could be affecting the interaction. The negative control, USP7-NTD, did not show any binding to the ICP0 peptide. When USP7-NTD was added to ICP0, it resulted in a negative mP value meaning that the ICP0 peptide did not bind. Ubl345 also showed lack of binding due to an absence of a plateau or saturation region. The high affinity between Ubl123b and ICP0 might be advantageous for HSV-1 during early infection when there is still a low amount of ICP0 present and when there are other cellular proteins trying to interact with USP7. This could also enable USP7 to stabilize ICP0 and reverse its autoubiquitylation thus preventing its degradation.

## 5.4 Protein crystallography

### 5.4.1 USP7-CTD Ubl domains

Crystallization efforts of the USP7-Ubl constructs have been fruitful. Ubl123 and Ubl345, out of 5 constructs plated for crystal trials, crystallized (Table 4-5, Figures 4-22, 4-23). With multiple efforts to refine these crystals, Ubl123 was irreproducible and the X-ray diffraction data for Ubl345 indicated that it was not sufficient for structure determination.

The crystal structure of the USP7 C-terminal (residues 560-1083) was published while Ubl123 and Ubl345 was in refinement [125] (Figure 5-1). The C-terminal indeed has 5 ubiquitin-like (Ubl) domains or HUBL domains with the characteristic ββαββ fold.

### 5.4.2 USP7-Ubl123 (residues 535-888) and ICP0 peptide (residues 617-627) co-crystallization trials

The molecular mechanism of interaction between two proteins can be visualized using protein x-ray crystallography. However, this method has many challenges and factors that could affect the determination of the three dimensional structure of the complex. First of all, the complex must be stable and crystallizable. There are many proteins and complexes that are recalcitrant to crystallization for a variety of reasons including disordered regions, flexible loops, and dynamic motion of the protein [126]. Also, the whole process of trying to produce crystals good enough for structure determination can take months to years.

Since Apo-Ubl123 was a challenge to reproduce and binding assays showed strong interaction with the ICP0 peptide (residues 617-627), this presented a good opportunity to set-up co-crystallization trials. Surprisingly, crystals appeared in over 70 crystallization conditions after three days (Appendix C). The challenge was picking a crystallization condition to refine. Conditions that gave the biggest crystals and had the least components were refined. When additives were used for refinement, some conditions formed 1-2 massive crystals (Figure 4-27). Diffraction data was collected. However, the structure could not be determined.

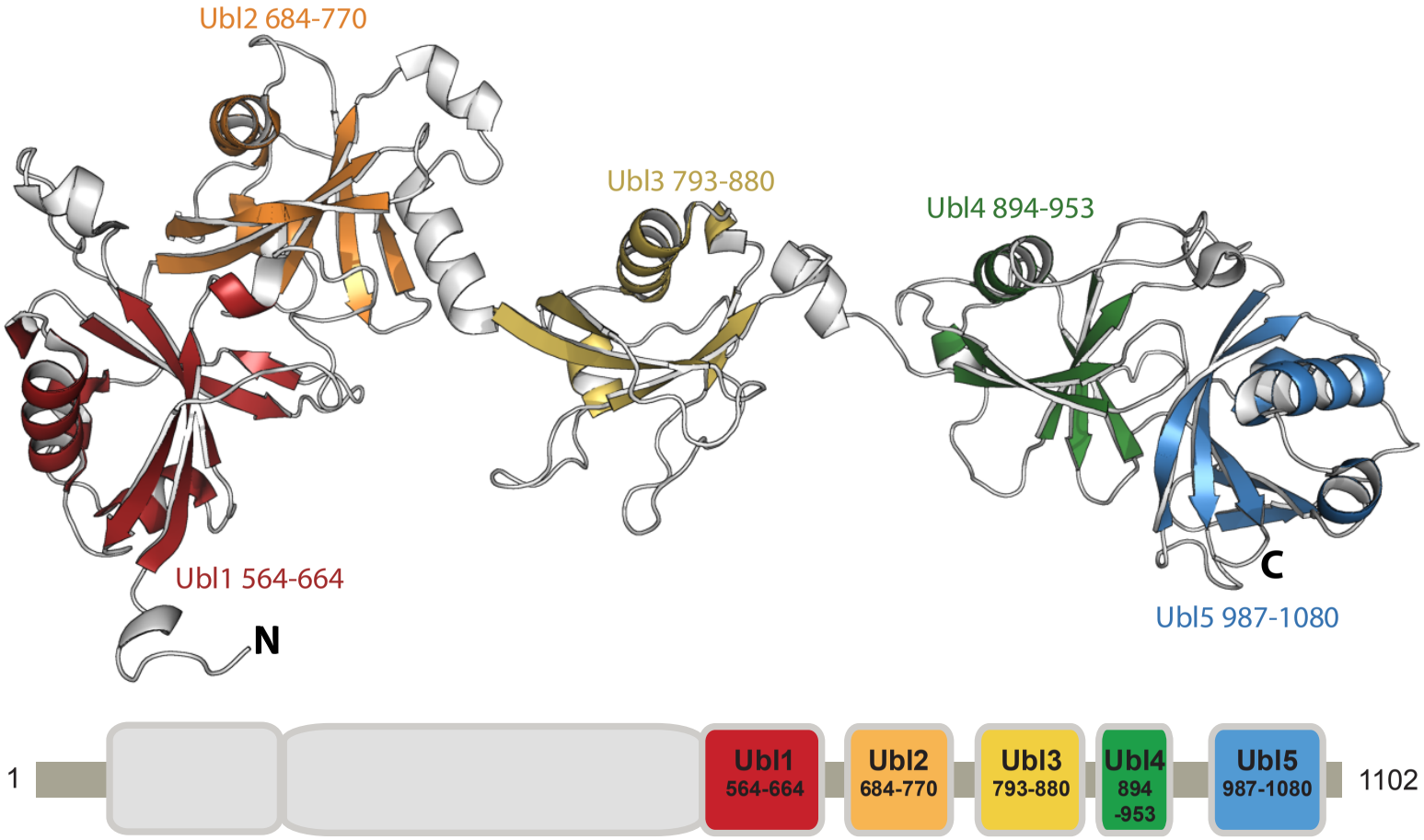



Figure 5-1: USP7 C-terminal structure. The C-terminal has 5 ubiquitin-like (Ubl or HUBL) domains. PDB: 2YLM. Adapted from Faesen *et al*. (2011).

## 5.5 Future Directions

Ubl123 (residues 535-888) co-crystallized with ICP0 peptide (residues 617-627). The initial crystallization screen produced multiple hits (Appendix C). Thus, there are more opportunities to further refine the crystal to obtain a diffraction dataset for structure determination. Once the crystal structure is determined, one would be able to see which amino acids are in direct contact between USP7-Ubl123 and the ICP0 peptide. Mutations in the ICP0 binding site in USP7 can be prepared to determine the role of the amino acids in the interaction. Similar *in vitro* biochemical techniques like pull-down assays and fluorescence polarization done in this thesis can be applied to these mutants. We can also test if ICP0 can affect USP7 deubiquitination activity. The level of deubiquitination activity by full-length USP7 with a di-ubiquitin substrate can be compared with using the same assay but with the addition of increasing amounts of ICP0 peptide. However, most important is to test *in vivo* the effect of these mutations on HSV-1 replication and ICP0 stability during viral infection.

HSV-1 is life long infections resulting to recurring cold sores. Current treatment of cold sores uses either pills or creams with an active compound, acyclovir, that helps inhibit DNA replication during the viral lytic cycle. The ICP0 binding pocket in USP7 can be used to design or screen for active compounds that would specifically target ICP0 or compete against ICP0 binding. This could help for the development of new treatments against HSV-1 infection and maybe a better alternative to the ones currently used.

It would be interesting to identify the Ubls that other USP7-binding proteins, like GMPS and FOXO4, interact with. It is possible that they also interact within the same region or a completely different region altogether. The same Ubl constructs can be used for these biochemical assays. The USP7-CTD crystal structure and USP7-CTD interaction with ICP0 can help in elucidating the function of these Ubl domains and encourages further experimentation. Since Ubl domains are one of the most common domains found in USPs [57], one the few questions that arise is the recognition of these Ubl domains. Do these Ubl domains get recognized by DUBs for cleavage? Do they get recognized as a Ubl monomer or Ubl chain? These tandem Ubl domains in USP7 could present a huge variety of binding surfaces for interacting proteins. This could be the reason why USP7 is able to function in so many pathways and regulate a wide variety of proteins. Thus, reinforcing the importance of studying these Ubl domains and the regulation of USP7.

## REFERENCES


1 Ciechanover A, Iwai K. The ubiquitin system: from basic mechanisms to the patient bed. IUBMB Life 2004;56: 193-201.

2 Lindner HA. Deubiquitination in virus infection. Virology 2007;362: 245-56.

3 Baker R, Board P. The Human Ubiquitin Gene Family - Structure of a gene and pseudogenes from the Ub-B Subfamily. Nucleic Acids Res 1987;15: 443-463.

4 Ozkaynak E, Finley D, Solomon MJ, Varshavsky A. The yeast ubiquitin genes: a family of natural gene fusions. EMBO J 1987;6: 1429-39.

5 Hershko A, Heller H, Elias S, Ciechanover A. Components of Ubiquitin-Protein Ligase System - Resolution, Affinity Purification, and Role in Protein Breakdown. J Biol Chem 1983;258: 8206-8214.

6 Weissman A. Themes and variations on ubiquitylation. Nat Rev Mol Cell Biol 2001;2: 169-178.

7 Pickart C, Rose I. Ubiquitin carboxyl-terminal hydrolase acts on ubiquitin carboxyl-terminal amides. J Biol Chem 1985;260: 7903-7910.

8 Ye Y, Rape M. Building ubiquitin chains: E2 enzymes at work. Nat Rev Mol Cell Biol 2009;10: 755-64.

9 Schulman BA, Harper JW. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. Nat Rev Mol Cell Biol 2009; 10: 319–331.

10 Schwartz AL, Ciechanover A. Targeting proteins for destruction by the ubiquitin system: Implications for human pathobiology. Annu Rev Pharmacol Toxicol 2009;49: 73-96.

11 Kawakami T, Chiba T, Suzuki T, Iwai K, Yamanaka K, Minato N, Suzuki H, Shimbara N, Hidaka Y, Osaka F, Omata M, Tanaka K. NEDD8 recruits E2-ubiquitin to SCF E3 ligase. EMBO J 2001;20: 4003-4012.

12 Siepmann T, Bohnsack R, Tokgoz Z, Baboshina O, Haas A. Protein interactions within the N-end rule ubiquitin ligation pathway. J Biol Chem 2003;278: 9448-9457.

13 Huibregtse J, Scheffner M, Beaudenon S, Howley P. A family of proteins structurally and functionally related to the E6-Ap ubiquitin protein ligase. Proc Natl Acad Sci USA 1995;92: 2563-2567.

14 Joazeiro C, Weissman A. RING finger proteins: Mediators of ubiquitin ligase activity. Cell 2000;102: 549-552.

15 Freemont PS. RING for destruction? Curr Biol 2000;10: 84-7.

16 Sun Y. Targeting E3 ubiquitin ligases for cancer therapy. Cancer Biol Ther 2003;2: 623-629.

17 Kerscher O, Felberbaum R, and Hochstrasser M. Modification of proteins by ubiquitin and ubiquitin-like proteins. Annu Rev Cell Dev Biol 2006;22: 159-180.

18 Haglund K, Dikic I. Ubiquitylation and cell signaling. EMBO J 2005;24: 3353-3359.

19 Hicke L. Protein regulation by monoubiquitin. Nat Rev Mol Cell Biol 2001;2: 195-201.

20 Herrmann J, Lerman LO, Lerman A. Ubiquitin and ubiquitin-like proteins in protein regulation. Circ Res 2007;100: 1276-1291.

21 Thrower J, Hoffman L, Rechsteiner M, Pickart C. Recognition of the polyubiquitin proteolytic signal. EMBO J. 2000;19: 94-102.

22 Chastagner P, Israel A, Brou C. Itch/AIP4 mediates Deltex degradation through the formation of K29-linked polyubiquitin chains. EMBO Rep 2006;7: 1147-1153.

23 Chan N, Hill C. Defining polyubiquitin chain topology. Nat Struct Biol 2001;8: 650-652.

24 Kirkpatrick D, Denison C, Gygi S. Weighing in on ubiquitin: the expanding role of mass-spectrometry-based proteomics. Nat Cell Biol 2005;7: 750-757.

25 Saeki Y, Kudo T, Sone T, Kikuchi Y, Yokosawa H, Toh-e A, Tanaka K. Lysine 63-linked polyubiquitin chain may serve as a targeting signal for the 26S proteasome. EMBO J 2009;28: 359-371.

26 Ciechanover A. The ubiquitin proteolytic system: from an idea to the patient bed. Proc Am Thorac Soc 2006;3: 21-31.

27 Hochstrasser M. Origin and Function of Ubiquitin-like Protein Conjugation. Nature 2009;458: 422.

28 Welchman RL, Gordon C, Mayer RJ. Ubiquitin and ubiquitin-like proteins as multifunctional signals. Nat Rev Mol Cell Biol 2005;6: 599-609.

29 Reyes-Turcu FE, Ventii KH, Wilkinson KD. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. Annu Rev Biochem 2009;78: 363-97.

30 Nijman S, Luna-Vargas M, Velds A, Brummelkamp T, Dirac A, Sixma T, and Bernards R. A genomic and functional inventory of deubiquitinating enzymes. Cell 2005;123: 773-786.

31 Amerik AY, Hochstrasser M. Mechanism and function of deubiquitinating enzymes. Biochim Biophys Acta 2004;1695: 189-207.

32 Wilkinson K. (1997). Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. FASEB J 1997;11: 1245-1256.

33 Wilkinson KD, Tashayev VL, O'Connor LB, Larsen CN, Kasperek E, Pickart CM. Metabolism of the polyubiquitin degradation signal: structure, mechanism, and role of isopeptidase T. Biochem 1995;34: 14535–14546.

34 Piotrowski J, Beal R, Hoffman L, Wilkinson KD, Cohen RE, Pickart CM. Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. J Biol Chem 1997;272: 23712-21.

35 Yi JJ, Ehlers MD. Emerging roles for ubiquitin and protein degradation in neuronal function. Pharmacol Rev 200;59: 14-39.

36 Randow F, Lehner PJ. Viral avoidance and exploitation of the ubiquitin system. Nat Cell Biol 2009;11: 527-34.

37 Komander D, Clague MJ, Urbé S. Breaking the chains: structure and function of the deubiquitinases. Nat Rev Mol Cell Biol 2009;10: 550-63.

38 Li M, Brooks CL, Kon N, Gu W. A dynamic role of HAUSP in the p53-Mdm2 pathway. Mol Cell 2004;13: 879–886.

39 Lee HJ, Kim MS, Kim YK, Oh YK & Baek KH. HAUSP, a deubiquitinating enzyme for p53, is polyubiquitinated, polyneddylated, and dimerized. FEBS Lett 2005;579: 4867–4872.

40 Everett RD, Meredith M, Orr A, Cross A, Kathoria M, Parkinson J. A novel ubiquitin-specific protease is dynamically associated with the PML nuclear domain and binds to a herpesvirus regulatory protein. EMBO J 1997;16: 566-77.

41 Fernandez-Montalvan A, Bouwmeester T, Joberty G, Mader R, Mahnke M, Pierrat B, Schlaeppi JM, Worpenberg S, Gerhartz B. Biochemical characterization of USP7 reveals post-translational modification sites and structural requirements for substrate processing and subcellular localization. FEBS J 2007;274: 4256-70.

42 Meredith M, Orr A, Everett R. Herpes simplex virus type 1 immediate-early protein Vmw110 binds strongly and specifically to a 135-kDa cellular protein. Virology 1994;200: 457-69.

43 Li M, Chen D, Shiloh A, Luo J, Nikolaev AY, Qin J, Gu W. Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. Nature 2002;416: 648-53.

44 Meulmeester E, Maurice MM, Boutell C, Teunisse AF, Ovaa H, Abraham TE, Dirks RW, Jochemsen AG. Loss of HAUSP-mediated deubiquitination contributes to DNA damage-induced destabilization of Hdmx and Hdm2. Mol Cell 2005;18: 565–576.

45 Tang J, Qu LK, Zhang J, Wang W, Michaelson JS, Degenhardt YY, El-Deiry WS, Yang X. Critical role for Daxx in regulating Mdm2. Nat Cell Biol 2006;8: 855–862.

46 Song MS, Salmena L, Carracedo A, Egia A, Lo-Coco F, Teruya-Feldstein J, Pandolfi PP. The deubiquitinylation and localization of PTEN are regulated by a HAUSP- PML network. Nature 2008;455: 813-7.

47 van der Horst A, de Vries-Smits AM, Brenkman AB, van Triest MH, van den Broek N, Colland F, Maurice MM, Burgering BM. FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP. Nat Cell Biol 2006;8: 1064-73.

48 Faustrup H, Bekker-Jensen S, Bartek J, Lukas J, Mailand N. USP7 counteracts SCFbetaTrCP- but not APCCdh1- mediated proteolysis of Claspin. J Cell Biol 2009;184: 13-9.

49 Oh YM, Yoo SJ, Seol JH. Deubiquitination of Chfr, a checkpoint protein, by USP7/HAUSP regulates its sta- bility and activity. Biochem Biophys Res Commun 2007;357: 615-9.

50 van der Knaap JA, Kumar BR, Moshkin YM, Langenberg K, Krijgsveld J, Heck AJ, Karch F, Verrijzer CP. GMP synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. Mol Cell 2005;17: 695-707.

51 Sarkari F, Sanchez-Alcaraz T, Wang S, Holowaty MN, Sheng Y, Frappier L. EBNA1-mediated recruitment of a histone H2B deubiquitylating complex to the Epstein-Barr virus latent origin of DNA replication. PLoS Pathog 2009;5: e1000624.

52 Hong S, Kim SJ, Ka S, Choi I, Kang S. USP7, a ubiquitin-specific protease, interacts with ataxin-1, the SCA1 gene product. Mol Cell Neurosci 2002;20: 298-306.

53 Nathan JA, Sengupta S, Wood SA, Admon A, Markson G, Sanderson C, Lehner PJ. The ubiquitin E3 ligase MARCH7 is differentially regulated by the deubiquitylating enzymes USP7 and USP9X. Traffic 2008;9: 1130-45.

54 Saridakis V, Sheng Y, Sarkari F, Holowaty MN, Shire K, Nguyen T, Zhang RG, Liao J, Lee W, Edwards AM, Arrowsmith CH, Frappier L. Structure of the p53 binding domain of HAUSP/USP7 bound to Epstein-Barr nuclear antigen 1 implications for EBV-mediated immortalization. Mol Cell 2005;18: 25-36.

55 Sheng Y, Saridakis V, Sarkari F, Duan S, Wu T, Arrowsmith CH, Frappier L. Molecular recognition of p53 and MDM2 by USP7/HAUSP. Nat Struct Mol Biol 2006;13: 285-91.

56 Hu M, Li P, Li M, Li W, Yao T, Wu JW, Gu W, Cohen RE, Shi Y. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. Cell 2002;111: 1041-54.

57 Zhu X, Ménard R, Sulea T. High incidence of ubiquitin-like domains in human ubiquitin-specific proteases. Proteins 2007;69: 1-7.

58 Holowaty MN, Sheng Y, Nguyen T, Arrowsmith C, Frappier L. Protein interaction domains of the ubiquitin-specific protease, USP7/HAUSP. J Biol Chem 2003;278: 47753-61.

59 Akhtar J, Shukla D. Viral entry mechanisms; cellular and viral mediators of herpes simplex virus entry. FEBS J 2009; 276: 7228-36.

60 Vahlne A, Svennerholm B, Lycke E. Evidence for herpes simplex virus type-selective receptors on cellular plasma membranes. J Gen Virol 1979;44: 217-25.

61 Morgan C, Rose WM, Mednis B. Electron microscopy of herpes simplex virus I entry. J Virol 1968;2: 507–516.

62 Salameh S, Sheth U, Shukla D. Early events in herpes simplex virus lifecycle with implications for an infection of lifetime. Open Virol J 2012;6: 1-6.

63 Roizman B, Sears AE. An inquiry into the mechanisms of herpes simplex virus latency. Annu Rev Microbiol 1987;41: 543-71.

64 Latchman DS. Molecular biology of Herpes simplex virus 1 latency. J Exp Pathol 1990;71: 133–141.

65 Lee HR, Toth Z, Shin YC, Lee JS, Chang H, Gu W, Oh T, Kim MH, Jung JU. Kaposi's sarcoma-associated herpesvirus viral interferon regulatory factor 4 targets MDM2 to deregulate the p53 tumor suppressor pathway. J Virol 2009;83: 6739–6747.

66 Rawlins DR, Milman G, Hayward SD, Hayward GS. Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA1) to clustered sites in the plasmid maintenance region. Cell 1985;42: 859–868.

67 Lee HR, Kim MH, Lee JS, Liang C, Jung JU. Viral interferon regulatory factors. J Interferon Cytokine Res 2009;29: 621-7.

68 Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 1994;266: 1865–1869.

69 Lee HR, Choi WC, Lee S, Hwang J, Hwang E, Guchhait K, Haas J, Toth Z, Jeon YH, Oh TK, Kim MH, Jung JU. Bilateral inhibition of HAUSP deubiquitinase by a viral interferon regulatory factor protein. Nat Struct Mol Biol. 2011;18: 1336-44.

70 Rainbow L, Platt GM, Simpson GR, Sarid R, Gao SJ, Stoiber H, Herrington CS, Moore PS, Schulz TF. The 222- to 234-kilodalton latent nuclear protein (LNA) of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) is encoded by orf73 and is a component of the latency-associated nuclear antigen. J Virol 1997;71: 5915–5921.

71 Lim C, Sohn H, Lee D, Gwack Y, Choe J. Functional dissection of latency-associated nuclear antigen 1 of Kaposi's sarcoma-associated herpesvirus involved in latent DNA replication and transcription of terminal repeats of the viral genome. J Virol 2002;76: 10320-31.

72 Jäger W, Santag S, Weidner-Glunde M, Gellermann E, Kati S, Pietrek M, Viejo-Borbolla A, Schulz TF. The ubiquitin-specific protease USP7 modulates the replication of Kaposi's sarcoma-associated herpesvirus latent episomal DNA. J Virol 2012;86: 6745-57.

73 Whitley RJ, Roizman B. Herpes simplex virus infections. Lancet 2001;357: 1513-8.

74 Elion GB, Furman PA, Fyfe JA, de Miranda P, Beauchamp L, Schaeffer HJ. Selectivity of action of an antiherpetic agent, 9-(2-hydroxyethoxymethyl) guanine. Proc Natl Acad Sci U S A 1977;74: 5716–5720.

75 Honess RW, Roizman B. Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. J Virol 1974;14: 8-19.

76 Watson RJ, Preston CM, Clements JB. Separation and characterization of herpes simplex virus type 1 immediate-early mRNA's. J Virol. 1979 Jul;31(1):42-52.

77 Dubbs DR, Kit S. Mutant strains of herpes simplex deficient in thymidine kinase-inducing activity. Virology 1964;22: 493-502.

78 Holland LE, Anderson KP, Shipman C Jr, Wagner EK. Viral DNA synthesis is required for the efficient expression of specific herpes simplex virus type 1 mRNA species. Virology 1980;101 :10-24.

79 Everett RD. A detailed analysis of Vmw110, a trans-acting transcriptional activator encoded by herpes simplex virus type 1. EMBO J 1987;6: 2069-2076.

80 Maringer K, Elliott G. Recruitment of herpes simplex virus type 1 immediate-early protein ICP0 to the virus particle. J Virol 2010;84: 4682-96.

81 Everett RD, Maul GG. HSV-1 IE protein Vmw110 causes redistribution of PML. EMBO J 1994;13: 5062–5069.

82 Lopez P, Van Sant C, Roizman B. Requirements for the nuclear- cytoplasmic translocation of infected-cell protein 0 of herpes simplex virus 1. J Virol 2001;75: 3832–3840.

83 Preston CM, Nicholl MJ. Induction of cellular stress overcomes the requirement of herpes simplex virus type 1 for immediate-early protein ICP0 and reactivates expression from quiescent viral genomes. J Virol 2008;82: 11775-83.

84 Everett RD. ICP0, a regulator of herpes simplex virus during lytic and latent infection. Bioessays 2000;22: 761-70.

85 Smith MC, Boutell C, Davido DJ. HSV-1 ICP0: paving the way for viral replication. Future Virol 2011;6: 421-429.

86 Stow ND, Stow EC. Isolation and characterization of a herpes simplex virus type 1 mutant containing a deletion within the gene encoding the immediate early polypeptide Vmw110. J Gen Virol 1986;67: 2571-2585.

87 Sacks WR, Schaffer PA. Deletion mutants in the gene encoding the herpes simplex virus type 1 immediate-early protein ICP0 exhibit impaired growth in cell culture. J Virol 1987;61: 829-839.

88 Everett RD, Preston CM, Stow ND. Functional and genetic analysis of the role of Vmw110 in herpes simplex virus replication. In Wagner, EK. ed; The control of herpes simplex virus gene expression. Boca Raton. CRC Press Inc. 1991. p50-76.

89 Stow EC, Stow ND. Complementation of a herpes simplex virus type 1 Vmw110 deletion mutant by human cytomegalovirus. J Gen Virol 1989;70: 695-704.

90 Samaniego LA, Neiderhiser L, DeLuca NA. Persisitence and expression of the herpes simplex virus genome in the absence of immediate-early proteins. J Virol 1998;72: 3307-3320.

91 Harris RA, Everett RD, Zhu X, Silverstein S, Preston CM. The HSV immediate early protein Vmw110 reactivates latent HSV type 2 in an in vitro latency system. J Virol 1989;63: 3513-3515.

92 Halford WP, Schaffer PA: ICP0 is required for efficient reactivation of herpes simplex virus type 1 from neuronal latency. J. Virol 2001;75: 3240-3249.

93 Leib DA, Coen DM, Bogard CL, Hicks KA, Yager DR, Knipe DM, Tyler KL, Schaffer PA. Immediate early regulatory gene mutants define different stages in the establishment and reactivation of herpes simplex virus latency. J Virol 1989;63: 759-768.

94 Cai W, Astor TD, Liptak LM, Cho C, Coen D, Schaffer PA. The herpes simplex virus type 1 regulatory protein ICP0 enhances replication during acute infection and reactivation from latency. J Virol 1993;67: 7501-7512.

95 Everett RD. Analysis of the functional domains of herpes simplex virus type 1 immediate-early polypeptide Vmw110. J Mol Biol 1988;202: 87-96.

96 Freemont PS. The RING finger. A novel protein sequence motif related to the zinc finger. Ann N Y Acad Sci 1993;684: 174-92.

97 Everett RD, Barlow P, Milner A et al. : A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. J Mol Biol 1993;234: 1038-1047.

98 Mullen MA, Ciufo DM, Hayward GS. Mapping of intracellular localization domains and evidence for colocalization interactions between the IE110 and IE175 nuclear transactivator proteins of herpes simplex virus. J Virol 1994; 68: 3250–3266.

99 Everett RD, Meredith MR, Orr A. The ability of herpes simplex virus type 1 immediate-early protein Vmw110 to bind to a ubiquitin- specific protease contributes to its roles in the activation of gene expression and stimulation of virus infection. J Virol 1999;73: 417-426.

100 Ciufo DM, Mullen MA, Hayward GS. Identification of a dimerization domain in the C-terminal segment of the IE110 transactivator protein from herpes simplex virus. J Virol 1994;68: 3267-82.

101 Boutell C, Sadis S, Everett RD. Herpes simplex virus type 1 immediate-early protein ICP0 and is isolated RING finger domain act as ubiquitin E3 ligases in vitro. J Virol 2002;76: 841–850.

102 Chelbi-Alix MK, de The H. Herpes virus induced proteasome-dependent degradation of the nuclear bodies-associated PML and Sp100 proteins. Oncogene 1999;18: 935–941.

103 Everett RD, Freemont P, Saitoh H, et al. The disruption of ND10 during herpes simplex virus infection correlates with the Vmw110- and proteasome-dependent loss of several PML isoforms. J Virol 1998;72: 6581–6591.

104 Muller S, Dejean A. Viral immediate-early proteins abrogate the modification by SUMO-1 of PML and Sp100 proteins, correlating with nuclear body disruption. J Virol 1999;73: 5137–5143.

105 Maul GG, Guldner HH, Spivack JG. Modification of discrete nuclear domains induced by herpes simplex virus type 1 immediate early gene 1 product (ICP0). J Gen Virol 1993;74: 2679–2690.

106 Maul GG. Nuclear domain 10, the site of DNA virus transcription and replication. Bioessays 1998;20: 660–667.

107 Everett RD, Chelbi-Alix MK. PML and PML nuclear bodies: implications in antiviral defence. Biochimie 2007;89: 819-30.

108 Regad T, Chelbi-Alix MK. Role and fate of PML nuclear bodies in response to interferon and viral infections. Oncogene 2001;20: 7274-86.

109 Everett RD, Parada C, Gripon P, Sirma H, Orr A. Replication of ICP0- null mutant herpes simplex virus type 1 is restricted by both PML and Sp100. J Virol 2008;82: 2661–2672.

110 Everett RD, Rechter S, Papior P, Tavalai N, Stamminger T, Orr A. PML contributes to a cellular mechanism of repression of herpes simplex virus type 1 infection that is inactivated by ICP0. J Virol 2006;80: 7995–8005.

111 Sarkari F, Wang X, Nguyen T, Frappier L. The herpesvirus associated ubiquitin specific protease, USP7, is a negative regulator of PML proteins and PML nuclear bodies. PLoS One 2011;6: e16598.

112 Boutell C, Canning M, Orr A, Everett RD. Reciprocal activities between herpes simplex virus type 1 regulatory protein ICP0, a ubiquitin E3 ligase, and ubiquitin-specific protease USP7. J Virol 2005;79: 12342–12354.

113 Canning M, Boutell C, Parkinson J, Everett RD. A RING finger ubiquitin ligase is protected from autocatalyzed ubiquitination and degradation by binding to ubiquitin-specific protease USP7. J Biol Chem 2004;279: 38160–38168.

114 Daubeuf S, Singh D, Tan Y, Liu H, Federoff HJ, Bowers WJ, Tolba K. HSV ICP0 recruits USP7 to modulate TLR-mediated innate response. Blood 2009;113: 3264-75.

115 Benvenuti M, Mangani S. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. Nat Protoc 2007;2: 1633-51.

116 Heidner E. Protein crystallizations: The functional dependence of the nucleation rate on the protein concentration and the solubility. J Crystal Growth 1978;44: 139-144.

117 Judge RA, Jacobs RS, Frazier T, Snell EH, Pusey ML. The effect of temperature and solution pH on the nucleation of tetragonal lysozyme crystals. Biophys J 1999;77: 1585–1593.

118 Fukuchi S, Homma K, Minezaki Y, Nishikawa K. Intrinsically disordered loops inserted into the structural domains of human proteins. J Mol Biol 2006;355: 845-57.

119 Garman EF, Doublié S. Cryocooling of macromolecular crystals: optimization methods. Methods Enzymol 2003;368: 188-216.

120 Wang W. Protein aggregation and its inhibition in biopharmaceutics. Int J Pharm 2005;289: 1-30.

121 Chen D, Texada DE. Low-usage codons and rare codons of *Escherichia coli*. Gene Ther Mol Biol Vol 2006;10: 1-12.

122 de Marco A. Molecular and chemical chaperones for improving the yields of soluble recombinant proteins. Methods Mol Biol 2011;705: 31-51.

123 Speed MA, King J, Wang DI. Polymerization mechanism of polypeptide chain aggregation. Biotechnol Bioeng 1997;54: 333–343.

124 Ryan BJ, Henehan GT. Overview of approaches to preventing and avoiding proteolysis during expression and purification of proteins. Curr Protoc Protein Sci. 2013;5: 5.25.

125 Faesen AC, Dirac AM, Shanmugham A, Ovaa H, Perrakis A, Sixma TK. Mechanism of USP7/HAUSP activation by its C-terminal ubiquitin-like domain and allosteric regulation by GMP-synthetase. Mol Cell. 2011; 44: 147-59.

126 Gu J and Hilser VJ. The significance and impacts of protein disorder and conformational variants. In Structural Bioinformatics, Second Edition. Ed. J. Gu and P.E. Bourne, John Wiley & Sons 2009.

127 Hicke L, Schubert HL, Hill CP. Ubiquitin-binding domains. Nat Rev Mol Cell Biol. 2005;6: 610-21.

**Appendix A : Summary of USP7 binding to GST-tagged ICP0 mutants (M) (aa. 553-712) and wild-type (WT) ICP0 (aa. 594-775). There are three positively charged doublets.**

| Mutation | Partial ICP0 sequence | USP7 binding |
|---|---|---|
| WT | 615 619 620 623 624 626 627 629<br>P R G P R K C A R K T R H A E | + |
| M5 | P R G P I K C A R K T R H A E | - |
| M4 | P R G P R I C A R K T R H A E | -- |
| M2 | P R G P R K C A L K T R H A E | - |
| M1 | P R G P R K C A L I T R H A E | -- |
| M7 | P R G P R K C A R K T L H A E | - |
| M6 | P R G P R K C A R K T R Y A E | + |

(+) binding (-) reduced binding (--) lack of binding. R626 (M7) and H627 (M6) was not mutated in our experiment. Adapted from Everett (1999).

# Appendix B: HSV-1 and HSV-2 ICP0 ClustalW 2.1 sequence alignment.

```
HSV-1           MEPRPGASTRR---PEGRPQREP----APDVWVFPCDRDLPDSSDSEAETEVGGRGDADH 53
HSV-2           MEPRPGTSSRADPGPERPPRQTPGTPAAPHAWGMLNDMQWLASSDSEEETEVG-ISDDDL 59
                ******:*:*    **  *:: *    **..* :  * :   ***** *****  .* *

HSV-1           HDDDSASEADSTDTELFETGLLGPQGVDGG--AVSGGSPPREEDPGSCGGAPPRED--GG 109
HSV-2           HR-DSTSEAGSTDTEMFEAGLMDAATPPARPPAERQGSPTPADAQGSCGGGPVGEEEAEA 118
                *  **:***.*****:**:**:..    .   *   ***.  :  *****.*  *:   .

HSV-1           SDEGDVCAVCTDEIAPHLRCDTFPCMHRFCIPCMKTWMQLRNTCPLCNAKLVYLIVGVTP 169
HSV-2           GGGGDVCAVCTDEIAPPLRCQSFPCLHPFCIPCMKTWIPLRNTCPLCNTPVAYLIVGVTA 178
                .. ************ ***::***:* *********: *********: :.*******.

HSV-1           SGSFSTIPIVNDPQTRMEAEEAVRAGTAVDFIWTGNQRFAPRYLTLGGHTVRALSPTHPE 229
HSV-2           SGSFSTIPIVNDPRTRVEAEAAVRAGTAVDFIWTGNQRTAPRSLSLGGHTVRALSPTPPW 238
                *************:**:*** ***************** *** *:************ *

HSV-1           PTTDEDDDDLDDADYVPPAPRRTPRAPTRRGAAAPPVTGGASHAAPQPAAARTAPPSAPI 289
HSV-2           PGTDDEDDDLADVDYVPPAPRRAP----RRGGGG----AGATRGTSQPAATRPAPPGAPR 290
                * **::**** *.*********:*    ***...    .**::.:.****:*.***.**

HSV-1           GPHGSSNTNTTTNSSGGGGGXXSRPAAPRGASGP--SGGVGVGVGVVEAEAGRPRGRTGP 347
HSV-2           SSSSGGAPLRAGVGSGSGGGPAVAAVVPRVASLPPAAGGGRAQARRVGEDAAAAEGRTPP 350
                .. ... .  :  .**.***    ...** ** *  :**  . .  *  :*. ..*** *

HSV-1           LVNRPAPLANNRDPIVISDSPPASPHRPP----------AAPMPGSAPRPG--PPASSAA 395
HSV-2           AG---QPRAAQEPPIVISDSPPPSPRRPAGPGPLSFFSSSSAQVSSGPGGGGLPQSSGRA 407
                      * * :. *********.**:**.          ::.  .*.*  *  * :*. *

HSV-1           SGPARPXXXXXXXXX--XXXXXXXXXAPAPGAEPAARPADARRVPQSHSSLAQAANQEQS 453
HSV-2           ARPRAAVAPRVRSPPRAAAAPVVSASADAAGPAPPAVPVDAHRAPRSRMTQAQTDTQAQS 467
                : *  .                    * *.*. *.* *.**:*.*:*: : **: .* **

HSV-1           LCRARATVARGSGGPGVEGGHGPSRGAAPSGAAPLPSAVSVEQEAAVRPRKRRGS----- 508
HSV-2           LGRAGATDARGSGGPGAEGGPGVPRGTNTPGAAPHAAEG-----AAARPRKRRGSDSGPA 522
                * ** ** ********.*** * .**: ..**** .:       **.********

HSV-1           -GQENPSPQSTRPPLAP--AGAKRAATHPPSDSGPGGRGQG-------GPGTP------- 551
HSV-2           ASSSASSSAAPRSPLAPQGVGAKRAAPRRAPDSDSGDRGHGPLAPASAGAAPPSASPSSQ 582
                 ... .*. :.*.****  .******.: ..**..*.**:*       *...*

HSV-1           -----LTSSAASASSSSASSSSAPTPAGAASSAAGAASSSASASS-----GGAVGALGGR 601
HSV-2           AAVAAAASSSSSAASSSSSAASSSSSAASSSSAASSSSSSSSASSSAGGAGGSVASASGA 642
                      :**::**:***:*:::*:.:.*.::****.::***:****     **:*.: .*

HSV-1           QE--ETSLGPRAASGPRGPRKCARKTRHAETS----GAAPAGGLTRYLPISGVSSVVALS 655
HSV-2           GERRETSLGPRAAA-PRGPRKCARKTRHAEGGPEPGARDPAPGLTRYLPIAGVSSVVALA 701
                 *  *********: ********************  .    .  ** ********:********:

HSV-1           PYVNKTITGDCLPILDMETGNIGAYVVLVDQTGNMATRLRAAVPGWSRRTLLPETAGNHV 715
HSV-2           PYVNKTVTGDCLPVLDMETGHIGAYVVLVDQTGNVADLLRAAAPAWSRRTLLPEHARNCV 761
                ******:******:******:*************:*  ****.*.********* * * *

HSV-1           MPPEYPTAPASEWNSLWMTPVGNMLFDQGTLVGALDFRSLRSRHPWSGEQGASTRDEGKQ 775
HSV-2           RPPDYPTPPASEWNSLWMTPVGNMLFDQGTLVGALDFHGLRSRHPWSREQGAPAPAGDAP 821
                 **:***.*****************************:.******** ****.:   .

HSV-1           -----
HSV-2           AGHGE 826
```

# Appendix C: Summary of crystallization screen hits.

| Suite | Well (96 well-plates) | Salt 1 | Salt 2 | Buffer | Precipitant 1 | Precipitant 2 | Precipitant 3 | Final pH |
|---|---|---|---|---|---|---|---|---|
| **UBL123SEL-ICP0B** | | | | | | | | |
| JCSG I | A03 | 0.05 M Lithium sulfate | 0.05 M Sodium sulfate | 0.05 M Tris-HCl pH 8.5 | 30 %(w/v) PEG 400 | | | |
| JCSG I | A07 | | | 0.2 M tri-Potassium citrate | 20%(w/v) PEG 3350 | | | |
| JCSG I | A08 | | | 0.2 M tri-Sodium citrate | 20%(w/v) PEG 3350 | | | |
| JCSG I | E03 | 0.2 M Sodium chloride | | 0.1 M Na/K phosphate pH 6.2 | 10 %(w/v) PEG 8000 | | | |
| JCSG I | G01 | | | 0.18 M tri-Ammonium citrate | 20%(w/v) PEG 3350 | | | |
| JCSG + | A3 | 0.18 M tri-Ammonium citrate | | | 20 %(w/v) PEG 3350 | | | |
| JCSG + | E8 | 1 M di-Ammonium phosphate | | 0.1 M Sodium acetate pH 4.5 | | | | |
| JCSG + | F7 | 0.8 M Succinic acid pH 7.0 | | | | | | |
| JCSG + | F10 | 1.1 M Sodium malonate | | 0.1 M HEPES pH 7.0 | 0,5 %(v/v) Jeffamine ED-2001 | | | 7.0 |
| PEGs II | A2 | | | 0.1 M MES pH 6.5 | 15 %(w/v) PEG 400 | | | |
| PEGs II | A4 | 0.2 M tri-Sodium citrate | | 0.1 M TRIS pH 8.5 | 15 %(w/v) PEG 400 | | | |
| PEGs II | A6 | 0.2 M Lithium sulfate | | 0.1 M TRIS pH 8.5 | 25 %(w/v) PEG 400 | | | |
| PEGs II | A12 | 0.2 M tri-Sodium citrate | | 0.1 M TRIS pH 8.5 | 30 %(w/v) PEG 400 | | | |
| PEGs II | E1 | | | 0.1 M TRIS pH 8.5 | 8 %(w/v) PEG 4000 | 0.8 M Lithium chloride | | |
| PEGs II | E3 | | | 0.1 M tri-Sodium citrate pH 5.6 | 10 %(w/v) PEG 4000 | 10 %(w/v) Isopropanol | | |
| PEGs II | E8 | 0.2 M Ammonium sulfate | | 0.1 M tri-Sodium citrate pH 5.6 | 15 %(w/v) PEG 4000 | | | |
| PEGs II | E9 | 0.2 M Ammonium sulfate | | 0.1 M HEPES pH 7.5 | 16 %(w/v) PEG 4000 | 10 %(w/v) Isopropanol | | |
| PEGs II | G3 | 0.1 M Potassium chloride | | 0.1 M TRIS pH 8.5 | 3 %(w/v) PEG 6000 | | | |
| PEGs II | H5 | 0.2 M Lithium chloride | 0.05 M Magnesium sulfate | | 8 %(w/v) PEG 8000 | | | |
| PEGs II | H8 | 0.2 M Calcium acetate | | 0.1 M HEPES pH 7.5 | 10 %(w/v) PEG 8000 | | | |
| PEGs II | H9 | 0.05 M Magnesium acetate | 0.1 M Sodium acetate | | 10 %(w/v) PEG 8000 | | | |
| PEGs II | H10 | 0.2 M Magnesium acetate | | | 10 %(w/v) PEG 8000 | | | |
| Classics II | B10 | | | 0.8 M Succinic acid pH 7.0 | | | | |
| Classics II | C4 | | | 0.56 M Sodium citrate pH 7.0 | | | | |
| Classics II | C7 | 0.8 M Sodium/Potassium tartrate | | 0.1 M TRIS pH 8.5 | 0.5 %(w/v) PEG 5000 MME | | | |
| Classics II | F1 | 0.2 M L-Proline | | 0.1 M HEPES pH 7.5 | 10 %(w/v) PEG 3350 | | | |
| Classics II | F3 | 0.064 M Sodium citrate pH 7.0 | | 0.1 M HEPES pH 7.0 | 10 %(w/v) PEG 5000 MME | | | |
| **UBL123HisCut-ICP0B** | | | | | | | | |
| JCSG I | D07 | 0.2 M di-Ammonium tartrate | | | 20%(w/v) PEG 3350 | | | |
| **UBL123** | | | | | | | | |
| JCSG I | G03 | 1.0 M Lithium chloride | | 0.1 M Citric Acid pH 5.0 | 10 %(w/v) PEG 6000 | | | 5.0 |

## UBL345

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| JCSG III | F10 | 0.2 M Potassium/Sodium tartrate | 0.1 M Sodium citrate pH 5.6 | 2.0 M Ammonium sulfate | | | |

## UBL345HisCut

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| JCSG III | F08 | | 0.1 M MES pH 5.0 | 2.4 M Ammonium sulfate | | | 6.0 |
| JCSG III | F10 | 0.2 M Potassium/Sodium tartrate | 0.1 M Sodium citrate pH 5.6 | 2.0 M Ammonium sulfate | | | |

## UBL123-ICP0B

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| JCSG I | A07 | | 0.2 M tri-Potassium citrate | 20%(w/v) PEG 3350 | | | |
| JCSG I | A08 | | 0.2 M tri-Sodium citrate | 20%(w/v) PEG 3350 | | | |
| JCSG I | A09 | | 0.2 M tri-Lithium citrate | 20%(w/v) PEG 3350 | | | |
| JCSG I | A12 | | 0.2 M Magnesium acetate | 20%(w/v) PEG 3350 | | | |
| JCSG I | B03 | | 0.1 M HEPES pH 7.5 | 10 %(w/v) PEG 8000 | | | |
| JCSG I | B07 | 0.2 M di-Sodium tartrate | | 20%(w/v) PEG 3350 | | | |
| JCSG I | D02 | 0.2 M Potassium chloride | | 20%(w/v) PEG 3350 | | | |
| JCSG I | D07 | 0.2 M di-Ammonium tartrate | | 20%(w/v) PEG 3350 | | | |
| JCSG I | D10 | | 0.1 M HEPES pH 7.5 | 10 %(w/v) PEG 6000 | 5 %(v/v) MPD | | |
| JCSG I | F08 | | 0.1 M Sodium citrate pH 5.5 | 20 %(w/v) PEG 3000 | | | |
| JCSG I | G01 | | 0.18 M tri-Ammonium citrate | 20%(w/v) PEG 3350 | | | |
| JCSG I | G03 | 1.0 M Lithium chloride | 0.1 M Citric Acid pH 5.0 | 10 %(w/v) PEG 6000 | | | 5.0 |
| JCSG II | A02 | 0.2 M Sodium chloride | 0.1 M CHES pH 9.5 | 1.26 M Ammonium sulfate | | | |
| JCSG II | A05 | | 0.1 M Bicine pH 9.0 | 10%(w/v) PEG 20000 | 2%(v/v) 1,4-Dioxane | | |
| JCSG II | A08 | | 0.1M Tris pH 8.5 | 5%(w/v) PEG 8000 | 20%(v/v) PEG 300 | 10%(v/v) Glycerol | |
| JCSG II | A12 | | 0.1 M Tris-HCl pH 8.5 | 8%(w/v) PEG 8000 | | | |
| JCSG II | C10 | | 0.1 M HEPES pH 6.5 | 10%(w/v) PEG 6000 | | | 7.0 |
| JCSG II | D04 | 0.2 M Potassium sulfate | | 20%(w/v) PEG 3350 | | | |
| JCSG III | A07 | 0.2 M di-Sodium hydrogen phosphate | | 20%(w/v) PEG 3350 | | | |
| JCSG III | B11 | 0.2 M di-Ammonium hydrogen phosphate | | 20%(w/v) PEG 3350 | | | |
| JCSG III | D08 | | 0.1 M HEPES pH 6.5 | 5%(w/v) PEG 6000 | | | 7.0 |
| JCSG III | F11 | 0.17 M Ammonium acetate | 0.085 M Sodium citrate pH 5.6 | 25.5%(w/v) PEG 4000 | 15%(v/v) Glycerol | | |
| JCSG + | A2 | | 0.1 M tri-Sodium citrate pH 5.5 | 20 %(w/v) PEG 3000 | | | |
| JCSG + | A3 | 0.18 M tri-Ammonium citrate | | 20 %(w/v) PEG 3350 | | | |
| JCSG + | A5 | 0.2 M Magnesium formate | | 20 %(w/v) PEG 3350 | | | |
| JCSG + | B3 | | 0.1 M BICINE pH 8.5 | 20 %(w/v) PEG 6000 | | | 9.0 |
| JCSG + | B4 | | 0.1 M HEPES pH 7.5 | 10 %(w/v) PEG 8000 | 8 %(v/v) Ethylene glycol | | |
| JCSG + | B12 | 0.2 M tri-Potassium citrate | | 20 %(w/v) PEG 3350 | | | |
| JCSG + | C4 | | 0.1 M HEPES pH 6.5 | 10 %(w/v) PEG 6000 | | | 7.0 |
| JCSG + | C10 | | 0.1 M BICINE pH 9.0 | 10 %(w/v) PEG 20000 | 2 %(v/v) 1,4-Dioxane | | |
| JCSG + | E10 | | 0.1 M BICINE pH 9.0 | 10 %(w/v) PEG 6000 | | | 9.0 |
| JCSG + | E12 | | 0.1 M Imidazole pH 8.0 | 10 %(w/v) PEG 8000 | | | |
| JCSG + | F7 | 0.8 M Succinic acid pH 7.0 | | | | | |
| JCSG + | F10 | 1.1 M Sodium malonate | 0.1 M HEPES pH 7.0 | 0,5 %(v/v) Jeffamine ED-2001 | | | 7.0 |
| JCSG + | F11 | 1 M Succinic acid | 0.1 M HEPES pH 7.0 | 1 %(w/v) PEG MME 2000 | | | 7.0 |
| JCSG + | G2 | 0.02 M Magnesium chloride | 0.1 M HEPES pH 7.5 | 22 %(w/v) Polyacrylic acid 5100 sodium salt | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| JCSG + | G6 | 0.24 M Sodium malonate pH 7.0 | | 20 %(w/v) PEG 3350 | | |
| JCSG + | G8 | 0.15 M DL-Malic acid pH 7.0 | | 20 %(w/v) PEG 3350 | | |
| PACT | C11 | 0.2M Calcium chloride | 0.1M Hepes pH 7 | 20% w/v PEG 6000 | | |
| PACT | D12 | | 0.1M Tris pH 8 | 20% w/v PEG 6000 | | |
| PACT | E11 | 0.2M Sodium citrate | | 20% w/v PEG 3350 | | |
| PACT | E12 | 0.2M Sodium malonate | | 20% w/v PEG 3350 | | |
| PACT | F8 | 0.2M Sodium sulphate | 0.1M Bis Tris propane pH 6.5 | 20% w/v PEG 3350 | | |
| PACT | F11 | 0.2M Sodium citrate | 0.1M Bis Tris propane pH 6.5 | 20% w/v PEG 3350 | | |
| PACT | F12 | 0.2M Sodium malonate | 0.1M Bis Tris propane pH 6.5 | 20% w/v PEG 3350 | | |
| PACT | G8 | 0.2M Sodium sulphate | 0.1M Bis Tris propane pH 7.5 | 20% w/v PEG 3350 | | |
| PACT | G10 | 0.2M Sodium/potassium phosphate | 0.1M Bis Tris propane pH 7.5 | 20% w/v PEG 3350 | | |
| PACT | G11 | 0.2M Sodium citrate | 0.1M Bis Tris propane pH 7.5 | 20% w/v PEG 3350 | | |
| PACT | H10 | 0.2M Sodium/potassium phosphate | 0.1M Bis Tris propane pH 8.5 | 20% w/v PEG 3350 | | |
| PACT | H11 | 0.2M Sodium citrate | 0.1M Bis Tris propane pH 8.5 | 20% w/v PEG 3350 | | |
| PACT | H12 | 0.2M Sodium malonate | 0.1M Bis Tris propane pH 8.5 | 20% w/v PEG 3350 | | |
| Classics II | D3 | | 0.1 M HEPES pH 7.0 | 30 %(v/v) Jeffamine ED-2001 | | 7 |
| Classics II | F1 | 0.2 M L-Proline | 0.1 M HEPES pH 7.5 | 10 %(w/v) PEG 3350 | | |
| Classics II | F6 | 0.2 M Ammonium sulfate | 0.1 M BIS-TRIS pH 5.5 | 25 %(w/v) PEG 3350 | | |
| Classics II | G2 | 0.2 M Lithium sulfate | 0.1 M BIS-TRIS pH 5.5 | 25 %(w/v) PEG 3350 | | |
| Classics II | G3 | 0.2 M Lithium sulfate | 0.1 M BIS-TRIS pH 6.5 | 25 %(w/v) PEG 3350 | | |
| Classics II | G5 | 0.2 M Lithium sulfate | 0.1 M TRIS pH 8.5 | 25 %(w/v) PEG 3350 | | |
| Classics II | H3 | | 0.24 M Sodium malonate pH 7.0 | 20 %(w/v) PEG 3350 | | |
| Classics II | H4 | | 0.2 M Ammonium citrate pH 7.0 | 20 %(w/v) PEG 3350 | | |
| Classics II | H8 | 0.1 M Magnesium formate | | 15 %(w/v) PEG 3350 | | |
| Classics II | H10 | 0.2 M Sodium citrate | | 20 %(w/v) PEG 3350 | | |
| PEGs II | C5 | 0.1 M Sodium acetate | 0.1 M HEPES pH 7.5 | 12 %(w/v) PEG 4000 | | |
| PEGs II | C6 | | 0.1 M TRIS pH 8.5 | 12 %(w/v) PEG 4000 | | |
| PEGs II | C8 | 0.2 M Sodium acetate | 0.1 M TRIS pH 8.5 | 16 %(w/v) PEG 4000 | | |
| PEGs II | E4 | | 0.1 M HEPES pH 7.5 | 10 %(w/v) PEG 4000 | 5 %(w/v) Isopropanol | |
| PEGs II | E12 | 0.1 M tri-Sodium citrate | | 20 %(w/v) PEG 4000 | 5 %(w/v) Isopropanol | |
| PEGs II | F1 | 0.1 M tri-Sodium citrate | | 20 %(w/v) PEG 4000 | 20 %(w/v) Isopropanol | |
| PEGs II | F6 | 0.2 M Ammonium sulfate | 0.1 M tri-Sodium citrate pH 5.6 | 25 %(w/v) PEG 4000 | | |
| PEGs II | G4 | 0.01 M Magnesium chloride | | 10 %(w/v) PEG 6000 | | |
| PEGs II | G7 | 0.05 M Potassium chloride | 0.01 M Magnesium chloride | 15 %(w/v) PEG 6000 | | |
| PEGs II | G8 | 0.01 M tri-Sodium citrate | | 16 %(w/v) PEG 6000 | | |
| PEGs II | H4 | | | 4 %(w/v) PEG 8000 | | |
| PEGs II | H6 | | 0.1 M TRIS pH 8.5 | 8 %(w/v) PEG 8000 | | |
| PEGs II | H11 | | 0.1 M HEPES pH 7.5 | 10 %(w/v) PEG 8000 | 10 %(w/v) Ethylene Glycol | |