DISSERTATION

UNDERSTANDING REGULATION OF HIV-1 PROTEASE PRECURSOR AUTOPROCESSING

Submitted by

Chih-Feng Tien

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2019

Doctoral Committee:

Advisor: Chaoping Chen

James Bamburg
Olve Peersen
Sandra Quackenbush

ABSTRACT

UNDERSTANDING REGULATION OF

HIV-1 PROTEASE PRECURSOR AUTOPROCESSING

The HIV-1 protease (PR) is initially synthesized as part of the Gag-Pol polyprotein precursor in the infected cell. Protease autoprocessing is generally referred to proteolytic reactions catalyzed by the precursor itself leading to liberation of free, mature PR in a highly regulated manner. We study the precursor autoprocessing mechanism using engineered fusion precursors carrying the p6*-PR miniprecursor sandwiched between various proteins and/or epitope peptides expressed in transfected mammalian cells. The studies reported here examined and identified factors involved in regulation of precursor autoprocessing.

Modulation of precursor autoprocessing activity and outcomes by the 26 amino acid maltose binding protein signal peptide (SigP) mimicking the proviral constructs. A H69D mutation in PR abolished autoprocessing of SigP-containing fusion precursors or Gag processing in viral particles whereas it only partially suppressed autoprocessing of fusion precursors lacking SigP. The mature PRs released from SigP-carrying precursors or associated with the viral particles are both resistant to self-degradation whereas those released from SigP-lacking fusion precursors are prone to self-degradation. Furthermore, the PR-containing autoprocessing intermediate fragments released from a SigP fusion precursor or a proviral constructs showed protease inhibitor response profiles distinct to those released from the corresponding fusion precursor lacking SigP. These findings of context-

dependent modulation reveals the complexity of precursor autoprocessing regulation that most likely accompanies sequence variation imposed by the evolution of the upstream Gag moiety.

We also examined *trans* proteolysis for its functional correlation with precursor dimerization. Fusion enzymes carrying GST, a well-known dimer forming protein, processed the GST-fused substrate *in trans* as expected. Interestingly, positive *trans* processing was also detected between enzyme and substrate precursors carrying maltose binding protein (MBP), a known monomeric tag, or lacking any dimer-inducing tag, suggesting that a dimer-inducing flanking tag is not required for *trans* proteolysis in the transfected cells. Sucrose gradient sedimentation analysis detected dimeric substrates, with or without dimer-inducing GST, as the major complexes in transfected cell lysates. In the presence of a protease inhibitor (PI) at high enough concentrations, dimeric enzymes were predominantly detected. Without PI treatment, fusion enzymes with different tags and varied p6* sequences showed monomers or dimers or mixtures, suggesting modulation of enzyme dimerization by p6* peptides and flanking tags. Precursors carrying two PRs in tandem tethered by a GGS linker demonstrated higher propensities of forming inter-molecular dimers than intra-molecular dimers, indicating a role of p6* peptide in regulating precursor dimerization. Collectively, our results decoupled the requirement of a dimer-inducing tag upstream of the p6*-PR miniprecursor for precursor *trans* proteolysis and demonstrated elements within and beyond p6*-PR miniprecursor that collectively influence precursor dimerization, which revealed additional complexity involved in precursor autoprocessing regulation. In summary, this dissertation highlights complicated regulations and more than one productive pathway involved in HIV-1 protease precursor autoprocessing.

ACKNOWLEDGEMENTS

Throughout the course of my Ph.D study, I received support and assistance from many sources. First, I would like to express my deep gratitude to my advisor, Dr. Chaoping Chen, for teaching me patiently, giving me proper training of research techniques and improving my understanding of the research field overall. Without her help, I could not have met all of the goals I accomplished these years. I am also grateful for Dr. Lilian Huang, who taught me many experimental skills and assisted me with numerous experiments, which resulted in productive progress. I am confident that this training has been instrumental to my Ph.D journey and will continue to impact my future career. I also want to thank my labmates: Satoshi Machihara, Molly Plehaty, Brittney R. Kemp, Tamsen S. Todd, Madison Fitzgerald, Marta Plonski, Wyatt Beyers, Matt O'Malley, Jade Lim, and Megan Gish for technical support.

I would like to thank my committee members: Dr. James Bamburg, Dr. Olve Peersen, and Dr. Sandra Quackenbush for giving me valuable and constructive feedback/comments regarding my research, as well as my future career options. They also provided much help with resolving research issues that arose. I am very grateful for their support and consideration. I would also like to thank the department chair Dr. Laurie Stargell and graduate chair Dr. Jennifer DeLuca for their encouragement and assistance during the course of my Ph.D study.

I want to thank several professors at CSU: Dr. Corey Rosenberg, Dr. Farida Safadi-Chamberlain, Dr. Santiago Di Pietro, Dr. Steven Markus, Dr. Eric Ross, and Dr. Tom Santangelo for giving me the opportunity to teach in their classes. I learned many teaching skills, which improved my communication skills. Thank you to the students in LIFE201B (Spring 2018-2019), LIFE 203

(Spring 2019), LIFE 210 (Fall 2018), and LIFE 212 (Fall 2017-2018) for giving me the chance to be your teaching assistant.

Many Thanks to my classmates: Hazheen Shirnekhi, Amanda Broad, Kenneth Lyon, and Dustin Steele. We supported each other both academically and personally. Thanks to everyone in the department for giving me warm-hearted assistance and kindnesses. Thanks to the Taiwanese Student Association and my Taiwanese friends for helping me to overcome the difficulties in my life. Thanks to Dr. Jeffrey Chang for introducing me to CSU and Dr. Ruey-Yi Chang's encouragement & assistance. Thanks to Professor Chung-Fu Chang, Dr. Elaine Chen, and other Taiwanese seniors for providing me many delicious Taiwanese foods when I missed my home country. Thanks to my uncle James Edenfield and my aunt Kuei-Jen Edenfield for their assistance when I lived here. Thanks to Ministry of Education in Taiwan for providing me a three-year scholarship.

I want to express my profound gratitude to my parents. They always support and encourage me from time to time. I learned diligence and contribution to society from my parents. I also would like to thank my extended family's support and encouragement. I also want to thank my sweetheart for maintaining a wonderful long distance relationship and supporting me in the pursuit of my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 The background of human immunodeficiency virus (HIV)

Human immunodeficiency virus (HIV) belongs to the *Retroviridae*, *Lentivirus* subfamily. It contains two copies of positive single-stranded genomic RNA. As a result of infection, retroviruses insert their genome into the host chromosomes. Two enzymes are involved in this process: reverse transcriptase (RT) and integrase (IN). After virus entry into the host cell, the genomic RNA is transcribed into DNA by reverse transcriptase. The viral DNA is then inserted into the chromosome of the host cell by integrase. The integrated DNA is called proviral DNA. The single-stranded RNA genome is about ~9700 nucleotides in length, including nine open reading frames (ORFs), a 5' cap, and a 3' poly-A tail (1). The genome encodes major structural proteins (Gag and Env), essential enzymes (Pol), regulatory proteins (Tat and Rev) and accessory proteins (Vpu or Vpx, Vpr, Vif, and Nef) (Fig 1.1). The Gag polyprotein is composed of matrix (MA), capsid (CA), SP1, nucleocapsid (NC), SP2, and p6 domain in that order. The 95% Gag polyprotein is translated from unspliced genomic RNA, however, the 5% Gag-Pol polyprotein (Pol reading frame contains the gene of three viral-encoding enzymes) is translated from unspliced genomic RNA by -1 frameshift that is induced by RNA structure. The MA is located at the N-terminus of Gag and it drives Gag binding to the cell membrane (2). The CA, SP1, and NC are involved in Gag assembly, during which Gag and Gag-Pol polyproteins targeted to the cell membrane assemble to produce virus particles (3, 4). The NC interacts with viral RNA during assembly (5). The p6 domain is required for virus release from the infected cell (6). Pol encodes three essential enzymes for virus

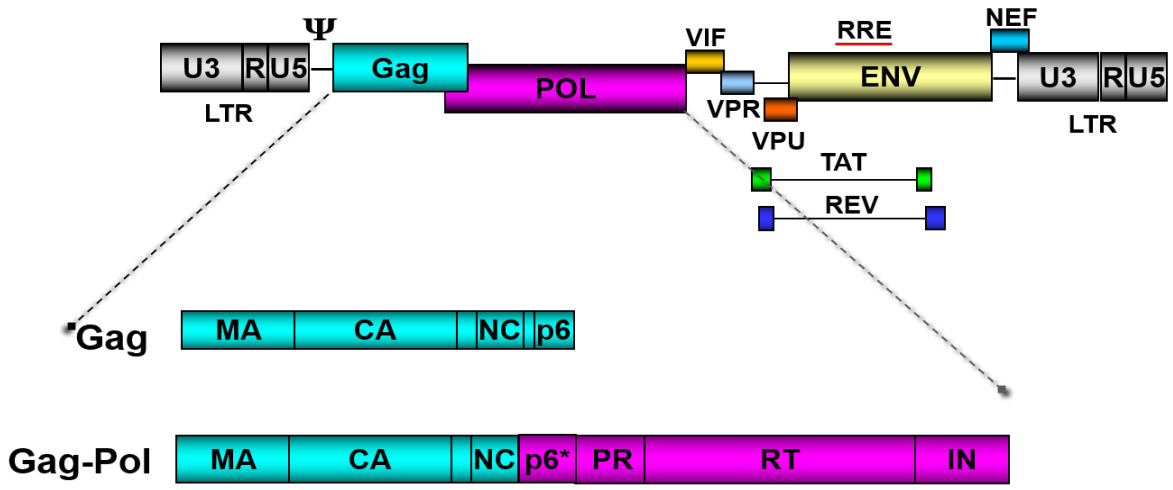

Figure 1.1 The proviral genome of HIV-1. The genome encodes major structural proteins (Gag and Env), essential enzymes (Pol), and accessory proteins (Tat, Rev, Vpu, Vpr, Vif, and Nef). The 95% Gag polyprotein is translated from unspliced genomic RNA, and the remaining 5% GagPol polyprotein is produced by -1 frameshift. The Gag and Gag-Pol have the same N-terminal sequence, and GagPol has three enzymes: protease (PR), reverse transcriptase (RT), and integrase (IN). These enzymes are essential for virus replication. (Courtesy of Dr. Chaoping Chen)

replication: reverse transcriptase (RT), integrase (IN), and protease (PR). PR, an aspartyl protease, processes the Gag and GagPol polyproteins into individual proteins, which is essential for viral infectivity. The Env is a viral spike protein and has two subunits: the surface protein (SU or gp120) and the transmembrane protein (TM or gp41). The gp120 mediates initial virus binding to CD4 receptors; the gp41 mediates virus fusion with the host cell membrane via conformational change in gp41 (7). The virus encodes two gene expression regulatory proteins (Tat and Rev). Tat is the activator for transcription (8), while Rev mediates viral RNAs export from nucleus to cytoplasm. The virus also encodes four accessory proteins (Vpu or Vpx, Vpr, Vif, and Nef). Vpr or Vpx transport the pre-integration complex (PIC) from cytoplasm to nucleus during early infection (9, 10). Vpu facilitates virus release (11), while Nef and Vif promote viral infectivity (12, 13).

There are two main types of HIV: HIV-1 and HIV-2. HIV-1 is the most widespread type in the world; HIV-2 is found mostly in West African countries. The Vpu of HIV-1 is replaced by the Vpx of HIV-2. Moreover, the HIV-2 protease shares only 50% sequence identity with HIV-1 protease (14). Both types are the causative agent of acquired immunodeficiency syndrome (AIDS). There are three clinical phases of HIV-1 infection (15). The first phase is an acute HIV infection. The CD4 protein (cluster of differentiation 4) is a glycoprotein and the primary receptor for HIV infection. CD4+ T cells that carry CD4 receptors on the surface of cells regulate the adaptive immune response against pathogens. At this stage, the virus attacks the immune system and destroys CD4+ T cells, so 50% CD4+ T cells maintains approximately (normal range is 500 to 1500 CD4+ cells per uL of blood) within 2-4 weeks, which is used as an indicator of this phase of HIV infection. The second phase is a chronic HIV infection, in which the virus is maintained at low levels for many years. The late-phase of HIV infection is onset of AIDS. Once an HIV positive

patient's immune system loses functional CD4+ T cells, it becomes too weak to fight against infections and diseases. The HIV patient has certain symptoms of infection, which are defined as AIDS. The levels of CD4+ T cells become low in the blood (<200 cells/ul).

1.2 HIV life cycle

The HIV-1 life cycle can be roughly divided into six stages: a) virus binding and fusion, b) capsid uncapping, c) reverse transcription, d) integration, e) transcription and translation, f) virion assembly and maturation. The HIV-1 virion is an envelope-enclosed spherical particle about ~100 nm in diameter. The envelop (Env) protein is on the surface of the HIV virion. The Env (gp160) is synthesized as a precursor in the endoplasmic reticulum (ER) (7). During translation, glycosylation occurs on gp160. Gp160 oligomerizes into a trimer in the ER and, it is trafficked to the Golgi complex. Then, the gp160 is cleaved by the host protease furin into two proteins: extracellular gp120 and transmembrane gp41. The gp120 trimer associates with a gp41 trimer noncovalently. Gp120 and gp41 form a heterotrimer as Env glycoprotein complexes on the surface of HIV particles.

The viral entry is determined by viral surface protein and cell surface receptor. The initial interaction is between the surface of the virus and the receptor of the cell (16-18). The gp120 has some motifs such as Asn-X-Ser or Asn-X-Thr, and these asparagines are glycosylated by cellular enzymes that are responsible to bind CD4 receptor of the host cell. The CD4 receptors of mature T cells are a primary target for HIV entry. The major function of CD4+ T cell is to activate innate immune system (19). The CD4 receptor serves as an adhesion molecule to recognize antigenic epitopes on the antigen-presenting cells (APCs) to initiation immune responses. The CD4 also

serves as a signal transducer to induce the transcription of cytokine and cytokine receptor gene for immune response. When gp120 interacts with the CD4 receptors, a conformational change of CD4 happens, which exposes the binding site of the G protein-coupled chemokine co-receptors CC chemokine receptor 5 (CCR5) or C-X-C motif chemokine receptor 4 (CXCR4) to gp120 binding. The role of CCR5 in the cell is critical for naïve CD4 T cell differentiation (20), while the role of CXCR4 is to activate the signal transduction inside the cell to regulate cell growth (21). Gp41 has two key domains that facilitate membrane fusion: one is an N-terminal hydrophobic region as a fusion peptide, and the other one is the heptad-repeat regions HR1 & HR2 that form α–helical coiled-coil structures (7). After the co-receptor binding, it triggers the fusion peptide to insert into the cell membrane. The three HR1 motifs and three HR2 motifs within each trimer fold into stable six-helix bundle that brings the cell membrane and viral membrane close with each other to create fusion pore. These events destabilize the membrane, promote membrane fusion between the cell membrane and viral membrane, and produce the fusion pore.

Maraviroc and Enfuvirtide are two recently approved therapeutic agents that act as entry inhibitors. Maraviroc binds to the receptor CCR5, preventing gp120 from interacting with the receptor (22). The disadvantage is that the virus may use other co-receptors, such as CXCR4, to enter into the cell (23). Moreover, viruses resistant to maraviroc utilize mutated gp120 to enter the cell by drug-bound CCR5. Another drug, enfuvirtide, binds to gp41, preventing the viral particle from entering the cell (24). It is an HR-2 derived peptide that blocks the formation of the six-helix bundle. However, the disadvantage is that it is expensive and unavailable in oral form (25).

After a virion fuses with a cell membrane, it releases a capsid core into the cytoplasm of the infected cell. The CA protein itself contains an N-terminal domain (NTD) and a C-terminal domain that are linked by a flexible peptide linker (26-28). The NTDs can form either a hexameric lattice or a pentameric lattice and the CTDs only form homodimers that further assemble into a hexameric lattice. The function of flexible peptide linker is to stabilize capsid core. The capsid core is a closed fullerene cone, which is assembled from linked 252 capsid hexamers primarily and 12 capsid pentamers (28, 29). On the outside of the capsid core are NTDs of CA (hexamers and pentamers), while the CTDs are located towards the inside of the structure. The capsid core houses the nucleocapsid-RNA complexes, integrase (IN), reverse transcriptase (RT), protease (PR), and viral accessory proteins (30).

The capsid core traffics toward the cell nucleus via microtubules, and it undergoes a dissociating process known as uncoating (31, 32). The current model suggests that the capsid core is partially dissociated in the cytoplasm when reverse transcription occurs (33). There are two lines of evidence supporting this model. Firstly, reverse transcription requires host protein to help the synthesis of viral DNA, such as eukaryotic elongation factor 1 subunits (eEF1A and eEF1G) that are mainly cytosolic (34). Secondly, partial dissociation, not complete dissociation, of the capsid core protects the process of reverse transcription and viral genome from the cellular sensors in the cytoplasm. Several studies have shown that some capsids bind to reverse transcription complexes (RTCs) and facilitate RTCs entry into the nucleus (35-37). These fundings suggest that CA may function to protect the viral genome while facilitating its entry into the nucleus. The value of understanding the capsid core disassembly is to develop drugs targeting the capsid protein. Studies show that one Pfizer compound, PF74, could target the NTD of CA and induce premature

uncoating (38). This drug also can disrupt the formation of viral-like particles in the late stage of the HIV life cycle (39). However, a recent study shows that PF74-resistant HIV has several mutations in the CA and the drug stabilize the capsid core in this instance (40).

Reverse transcription occurs at the stage of capsid core uncoating. The reverse transcriptase (RT) is a heterodimer composed of two subunits, p66 and p51. The p66 subunit has two enzymatic activities: DNA polymerase and ribonuclease H (RNase H), while the p51 subunit only has a DNA polymerase domain. The 5' and 3' untranslated regions (UTR) of a full-length RNA genome are critical for reverse transcription (41-43). The 5'UTR has repeats (R), 5' untranslated sequence (U5), and primer binding sites (PBS), while the 3'UTR has a polypurine tract (PPT), 3' untranslated sequence (U3), R and the poly A tails. The R is at the end of viral RNA, which serves as a bridge to synthesize minus-strand DNA. The U3 regions contain the binding sites of cellular transcription factors. The U5 regions have binding sites for Tat and packaging sequences to package the viral RNA genome. The polypurine tract (PPT) is a purine-rich sequence that is essential for providing the plus-strand DNA primer.

The mechanism of reverse transcription from genomic RNA into linear viral DNA is well-characterized (43). Firstly, the host primer $tRNA^{Lys}$ anneals to the PBS of one genomic RNA, which is chaperoned by nucleocapsid (44), and RT synthesizes minus-strand DNA to the 5' end (45). The RNase H then degrades the 5' end of the viral RNA. This minus-strand DNA contains R, U5, and PBS. It subsequently anneals to the R region of the 3'UTR and serves as the primer for synthesis of minus-strand DNA. Then, RNase H degradation removes the template viral RNA, except for PPT. The PPT sequence is resistant to RNase H digestion, which serves as another

primer to synthesize plus-strand DNA. The plus-strand DNA associates with the 3' end of the negative strand DNA and serves as the new primer for synthesizing plus-strand DNA. Finally, RT extends two viral DNA strands, resulting in two viral DNA strands with two flanking long terminal repeats (LTR) at the 5' and 3' ends, which contain U3, R, and U5 regions. These three regions are required for viral DNA integration into the host chromosome by integrase.

Once the double-stranded DNA is synthesized by reverse transcriptase, the integrase multimer will bind to both ends of the linear double-stranded DNA. The integrase contains three domains: the N-terminal domain (NTD), the catalytic core domain (CCD), and the C-terminal domain (CTD) (46, 47). The NTD harbors a conserved pair of His and Cys residues called the HHCC motif, which binds with a Zn^{2+} ion. When a Zn^{2+} ion interacts with the HHCC motif, it promotes the catalytic activity of integrase and the multimerization of integrase (48). The CCD contains the active site of the endonuclease and the polynucleotidyl transferase. Two steps are involved in viral DNA integration: 3'-processing (endonuclease) and strand transfer (polynucleotidyl transferase) (43, 49, 50). Firstly, integrase recognizes the CAGT sequence of the 3 'end of viral DNA and removes a dinucleotides (GT sequences). The IN in the form of a tetramer uses viral DNA CA_{OH}-3' end to target the host DNA 5'-phosphate, which links the 3' end of the viral DNA and the 5' end of the host DNA. This process is referred to as the strand transfer. The cellular proteins then complete the integration by ligating the 5' end of the viral DNA and the 3' end of the host DNA. The HIV IN prefers to insert into active transcription units as the integration sites (51). The integrated viral DNA is called the proviral DNA.

The regulatory protein Tat (transactivating factor) and host cell transcription factors are required to regulate HIV transcription (52). The long terminal repeats (LTR) act as the viral promoter, which contain U3, R, and U5 regions. The Tat functions as a transcription activator through interacting with a 5' stem-loop trans-activating response (TAR) RNA structure at the U3 region (53). The TAR is a highly stable stem-loop structure, which is located between initiation site +1 and +59 site. In a Tat-dependent transcription mechanism, the positive transcription elongation factor (P-TEFb)/7SK snRNP is recruited to the promoter (54). The P-TEFb is a critical factor for HIV elongation and the 7SK snRNP, containing small nuclear RNAs and RNA-binding proteins, controls the elongation activity of P-TEFb. When 7SK snRNP interacts with the P-TEFb, the activity of P-TEFb is inhibited. After the P-TEFb/7SK snRNP complexes recruitment, Tat dirupts the complexes, and P-TEFb is released from 7SK snRNP. The component of P-TEFb that regulates transcription is Cyclin-dependent kinase 9 (CDK9). CDK9 phosphorylates the C-terminal domain of RNA polymerase II to promote elongation (55). Moreover, the expression level and activity of P-TEFb influences the viral genome expression (56). The high expression level and activity of P-TEFb stimulates HIV transcription.

After transcription, the full-length viral RNAs have three fates: (a) The RNAs are spliced to produce sub-genomic RNAs, which then encode for various viral proteins (57, 58). (b) The full-length RNAs are delivered to the cytoplasm, and then packaged into the new virion (5, 59). (c) The full-length RNAs function as the mRNA to translate Gag and Gag-Pol polyproteins. HIV-1 primary transcripts undergo alternative splicing to produce two major spliced mRNAs (52, 60). One group is ~4 kb incompletely spliced (single spliced) mRNAs, which encode for vif, vpr, and env/vpu. The other group is ~1.8 kb completely spliced mRNAs, which encode tat, rev, and nef.

Both the ~9 kb and ~4 kb mRNAs have a Rev-responsive element (RRE) that is a highly structured RNA in the *env* gene. The viral protein Rev, which has a leucine-rich nuclear export signal (NES), interacts with the Rev-responsive element (RRE) of viral RNA (61-63). The nuclear export factor chromosome region maintenance 1 (CRM1) recognizes the protein containing NES such as Rev and interacts with Ran-GTP. Then, the RNA-Rev-CRM1 complex is exported through the nuclear pore complex (NPC). The ~1.8 kb mRNAs don't have RREs and are exported to the cytoplasm by cellular machinery.

After all full-length and spliced RNAs are delivered into the cytoplasm, they serve as mRNA for translation. These mRNAs are translated by several mechanisms such as the canonical cap-dependent mechanism, internal ribosome entry site (IRES)-dependent translation, leaky scanning, and frameshifting (64). The full-length mRNA is translated to Gag and Gag-Pol polyproteins either by canonical cap-dependent mechanism or IRES-dependent translation (65). The cap-dependent mechanism is mediated by the 40s ribosomal subunit with initiation factors. The 40s ribosomal subunit interacts with eukaryotic initiation factors, GTP, and Met-tRNA to form a 43S pre-initiation complex to scan the 5'UTR. However, the highly structured 5'UTR has many stable RNA stem-loops. Thus, the cellular RNA helicase DDX3 unwinds the 5'UTR to assist 43S pre-initiation complex to scan the 5'UTR (66). In contrast, the IRES-dependent mechanism is mediated by the 40s ribosomal subunit with IRES *trans*-acting factors (ITAF). The 5'UTR of the HIV-1 genome harbors two IRESs: one in the 5'UTR, and the other one within the *gag* coding region, called HIV-1 Gag IRES (67, 68). The ribosomal subunits utilize either IRES to express the Gag and Gag-Pol polyproteins. It remains largely unknown how the virus chooses one of the mechanisms for translation. However, studies show that the activity of IRES for translation is

during G2/M phase arrest or stimulated by oxidative stress (69, 70). One possible explanation is that the cell is under stress during HIV infection. Meanwhile, the virus infection also disrupts cellular translation. For example, HIV protease can cleave the initiation factors such as eIF4G (the component of the cap-binding complex) and eIF3 (the component of a 43S pre-initiation complex) (71, 72). The Vif and Vpr expression also induces G2/M cell cycle arrest (73). These studies indicate the reason why the virus switches from cap-dependent translation to IRES-dependent translation. Moreover, HIV-1 uses different methods to translate various proteins from a common mRNA. One method is called leaky scanning for vpu-env translation. The ribosome bypasses the Vpu start codon and translates from the Env start codon of Vpu-Env bicistronic mRNA (64, 74). Another example is translational frameshifting for Gag-Pol translation. From the full-length genome, ~95% translation produces Gag and ~5% translation undergoes a -1 nucleotide ribosomal frameshift producing Gag-Pol. The frameshift efficiency is collectively determined by two major elements: a slippery heptanucleotide sequence (UUUUUUA) and a downstream RNA element called a frameshift stimulatory signal (FSS) containing an eight-nucleotide spacer and a stem-loop structure (75, 76).

After translation, Gag-pol, Gag, Env, and Vpr traffic to the plasma membrane to assemble into a virion. The virus assembly is that viral proteins are concentrated at the assembly site forming a viral bud. The Gag polyproteins contain matrix (MA), capsid (CA), nucleocapsid (NC), p6 domain, and spacer peptides SP1 & SP2. The N-terminal domain (basic residues) of Gag polyprotein's MA binds to the negatively charged lipid, phosphatidylinositol (4,5)-bisphosphate [$PI(4,5)P_2$], of the inner leaflet of the plasma membrane. The MA also recruits Env glycoproteins to form new immature virions (2, 3). The NC domain of Gag polyproteins packages two copies of viral RNAs

by interacting with the packaging signals (Ψ) at the 5' end of UTR during assembly (5, 77). The viral RNA has dimer initiation signals to induce genomic RNA dimerization. The Gag-Gag interaction is mediated by CA and SP1 (78, 79), and it is assemble to immature virion (80). After assembly, the particle is released from the cell surface, which is called virus budding. The release of the virion requires cellular endosomal sorting complexes required for transport (ESCRT) machinery (3, 81-83). The function of ESCRT complexes is to interact with the membrane and ubiquitylated cargo to facilitate the membrane scission away from the cytosol or budding towards the cytosol (84). The ESCRT pathways have four distinct complexes (ESCRT-0, ESCRT-1, ESCRT-2, and ESCRT-3). The p6 domain has two "late domain" motifs that are involved in the ESCRT pathway. The motif Pro-Thr/Ser-Ala-Pro (PTAP) binds the ESCRT-I component tumor susceptibility gene 101 (TSG101) (85), and another motif, Tyr-Pro-X-Leu (YPXL), binds the ESCRT-II and ESCRT-III component ALG2-interacting protein X (ALIX) (86). HIV recruits the components of ESCRT machinery for budding and releasing from the cell surface.

The virion released from the cell in the beginning is immature and noninfectious, containing uncleaved Gag and Gag-Pol. To become mature and infectious, the virion must undergo proteolysis reactions and structural rearrangements of viral proteins (3, 77). It is unknown which signal induces the proteolytic reactions and when the proteolytic reactions occur. Upon or shortly after virion budding, the Gag-Pol polyprotein undergoes autoproteolysis to release free, mature PR. This process is generally referred to as protease autoprocessing. The mature PR is composed of 99 amino acids with its active site formed at the homodimer interfaces. The mature PR can recognize and process all the cleavage sites in the Gag and Gag-Pol polyproteins (82).Each cleavage site has different substrate sequences and cleavage rates. The recombinant mature

protease displays a highly stepwise order to process these (87). These proteolysis reactions cause individual proteins to reorganize or change their conformation. The MA proteins associate with the viral envelope (Env) and enclose a capsid core. The capsid core consists of ~252 CA hexamers and ~12 CA pentamers to house the nucleocapsid-RNA (88). The viral RNA dimerization is stabilized by the nucleocapsid (89). The rearrangement of individual proteins results in the production of mature infectious viruses. Thus, mature PR is required for the conversion from immature (non-infectious) virus to mature (infectious) virus.

1.3 Mature PR and Precursor autoprocessing

In the HIV-1 infected cell, Gag and Gag-Pol polyproteins are translated, they are then targeted to the plasma membrane to assemble into immature virion. Upon virion release, the Gag-Pol precursor undergoes autoproteolysis, leading to liberation of free, mature PR. The precursors are referred to as polyproteins containing protease as part of it, whereas the mature PR is 99 amino acid long without any flanking proteins. There are at least 11 cleavage sites in the Gag and Gag-Pol precursors that can be recognized and processed by the mature PR (90). The precursors can process some, but not all cleavage sites (91). Because proper Gag and Gag-Pol proteolysis are critical for producing mature infectious virus, the mature PR is absolutely required for viral infectivity.

The HIV-1 mature protease (PR) is a homodimer of 99 amino acids, and it belongs to the aspartic protease family. The dimerization of mature PR is essential for the formation of the active site consisting of Asp^{25}-Thr^{26}-Gly^{27} (DTG) (Fig 1.2A) (92). When a substrate peptide comes to the active site, the residue D25 locating the dimer interface utilizes a water molecule to attack the

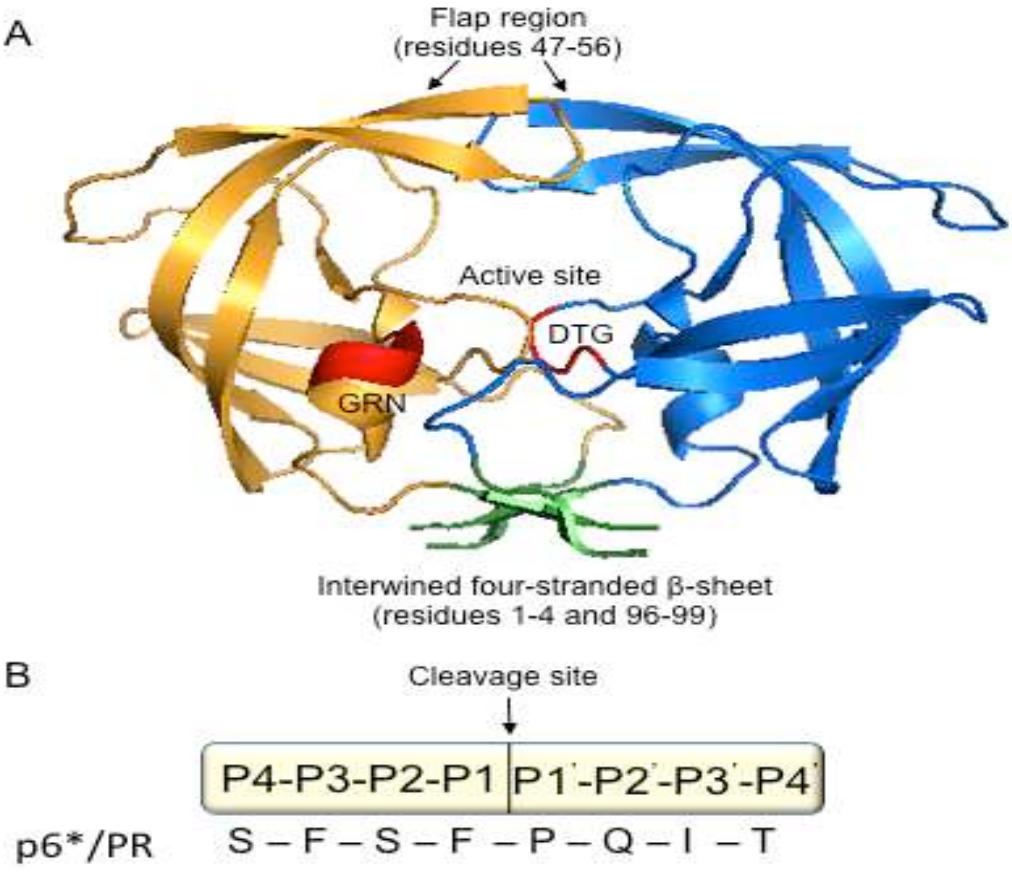

Figure 1.2 The structure of HIV-1 mature protease and the position of residues at the cleavage site. (A) The conserved trails (red), intertwined four stranded β-sheet (green), and flap region are highlighted in the structure of HIV-1 PR (PDB code 4HLA). (B) The residues of a substrate at the cleavage sites are numbered from P4 to P1 and then from P1' to P4'. Four positions each on both sides of the cleavage site (P4-P4') are shown. A scissile bond is between P1 and P1'. The example is the cleavage site between p6* and PR from the HIV-1 NL4-3 strain.

carbonyl carbon of the peptide bond, via acid-base catalysis. This hydrolysis reaction results in cleavage of the peptide bond. If the D25 is mutated to asparagine, the PR^{D25N} will lose its function (93). The substrate residues at the cleavage sites are numbered consecutively from P4 to P1 and then from P1' to P4' with a cleavage site between P1 and P1' (Fig 1.2B). There is no canonical sequence among the known cleavage sites but most cleavage sites are conserved among all HIV-1 group M virus. The HIV-1 group M virus is predominant and responsible for the majority of AIDS pandemic. Cassol *et al.* reported that the last three residues (P3 to P1) of p6* at the p6*/PR site are variable within and among subtypes (94).The substrate specificity of HIV mature PR includes: (i) a preference residue Glu in the P2' position, (ii) large aliphatic and aromatic residues in the P1 and P1' positions, and (iii) a slight preference for a Pro residue in the P1' position (95-97). Each cleavage site has a different cleavage rate, so the processing of Gag and Gag-Pol has a specific order and could be regulated (90, 98). There are three possible factors to regulate the proteolysis kinetics of cleavage sites: (a) the cleavage site sequence, (b) the context of the substrate, (c) the binding affinity between cleavage site and mature PR. When a recombinant purified Gag (MA-CA-SP1-NC-SP2-p6) is processed by mature PR *in vitro* transcription/translation system, SP1/NC is the first cleavage site (98). Then, MA/CA and SP2/p6, which have intermediate rates, are cleaved by PR. The final cleavage sites are CA/SP1 and CA/SP2. Interestingly, when the SP1 domain is deleted, the CA/SP1 site has a 20-fold higher cleavage rate, suggesting that the SP1 domain regulates the cleavage rate of CA/SP1 (99). Thus, the proteolysis processing of Gag polyprotein is temporo-spatially regulated.

The first PR crystal structure was solved using recombinant PR expressed in *E. coli* (100-102). However, the majority of the protease was found in inclusion bodies. Thus, Louis *et al* developed

a method to purify the PR from inclusion bodies under denaturing conditions and refold it *in vitro* (93). Another technical obstacle is that wild-type mature PR tends to undergo self-degradation *in vitro*. To overcome this, several point mutations (Q7K, L33I, and L63I) were introduced to the mature PR to prevent self-degradation (103, 104). Moreover, two mutations, C67A and C95A, were included to prevent PR aggregation by omitting oxidation of the thiol side chain of cysteines. This modified protease bearing five point mutations is called *pseudo* wild-type PR, which is widely used for structural analyses of mature PR as it has similar enzymatic kinetics to wild-type PR (104). Note that the enzymatic kinetics of *pseudo* wild-type PR is not the same as the wild-type PR, suggesting that mature PR activity is sensitive to amino acid alterations.

There are over ~100,000 crystal structures of mature PR including the wildtype & mutated PR with and without protease inhibitors (105, 106). Two highly conserved regions exist in the mature PR (Fig 2.2A): the active site triad Asp^{25}-Thr^{26}-Gly^{27} (DTG) and the C-terminal triad Gly^{86}-Arg^{87}-Asn/Asp^{88} (GRN/D) (107). The active site DTG is located at the dimer interface and forms a hydrogen bond network, stabilizing the center of mature PR (107, 108). Asp^{25} has one hydrogen bond, which interacts with a water molecule, and another hydrogen bond to interact with $Asp^{25'}$ of the other monomer. These interactions are critical for the proteolysis reactions. The NH group of Gly^{27} binds to the CO group of Asp^{25} via hydrogen bonding as well. Thr^{26} has two interactions with Asp^{25} and Leu^{24}. These interactions are called "fireman's grip", which form a rigid network structure (109). The GRN is not located at the catalytic site, and they are hydrogen bonded with other residues (110). When a substrate binds to the protease, Arg^{87} and Asn^{88} interact with a water molecule, separately. The water also bridges other residues to stabilize the conformation. If the R87 is mutated to lysine, the molecular weight of PR_{R87K} indicated a monomer from sedimentation

equilibrium analysis (111). A glycine-rich "flap region" (residues 47-56) creates a dynamic "roof", formed by two β-hairpins. This provides structural flexibility to allow the substrate and inhibitor to enter the active site (97, 112). The terminal residues 1-4 and 96-99 form an intertwined four-stranded β-sheet. The β-sheet contributes the majority of the stabilization energy for the dimerization of mature PR (113).

Mature PR is one of the proven targets for antiretroviral therapy. There are nine US Food and Drug Administration (FDA) approved protease inhibitors on the market. These PR inhibitors (PIs) are non-cleavable substrate analogs (114). That is, the inhibitors mimic substrate peptide binding to the catalytic site of the PR. These PIs bind to the catalytic site of mature PR with high affinity and prevent the formation of infectious progeny virions (115, 116). However, due to a high replication rate and low fidelity of reverse transcriptase, the viral genome carries sepecific mutations. Several studies show that some patients have PI-resistance, which means that the drug's ability to act as a treatment of antiviral therapy is reduced. PIs cannot suppress the activity of drug-resistant PR (117-120), and may thus maintain the correct cleavage order and produce mature virus (121). Interestingly, mutations are found not only within the PR but also at cleavage sites and within other viral proteins of the GagPol and Gag polyprotein (122-124). The drug resistant mechanisms of PR mutations is not fully understood. Recent studies show that the structure of multiple-drug resistant (MDR) protease has an open flap conformation in a dimer form. However, in the presence of DRV, the MDR PR has a closed flap conformation and can bind DRV (120, 125, 126). It does not change the active site binding with DRV. There are two obvious problems for researching drug resistant mechanism: (I) Other viral proteins have drug-resistant mutations, and they may be involved in or regulate this mechanism. It may be misleading to only examine the MDR protease with PIs. For

example, we also demonstrated that the precursor autoprocessing is also involved in this mechanism (127). (II) The purified method for protease is another factor. The available structures of PR are from proteins purified under denaturing conditions followed by refolding in vitro. The conformation of refolded mature PR may show the most favorable and stable structure *in vitro*, but it remains to be demonstrated whether it is also the conformation inside a virion.

Many publications focus on HIV mature PR and its structure, proteolysis mechanism, kinetic analysis, and virion maturation. In contrast, the precursor autoprocessing mechanism is largely unknown. Also, there is no structural information available for the precursor. Moreover, precursor autoprocessing needs to be tightly suppressed during Gag/Gag-Pol assembly and to be activated upon or shortly after virion budding. It is not entirely understood when the precursor autoprocessing is activated or how viral/host proteins are involved in the mechanism (suppression/activation) of protease autoprocessing. The current autoprocessing theory is solely based on the structure of mature PR assuming precursor and mature PR are enzymatically identical. This theory implies precursor dimerization as an obligatory step to activate autoprocessing. The Gag-Pol contains matrix (trimer), capsid (hexamer), nucleocapsid, protease (dimer), reverse transcriptase (dimer), integrase (multimer) and other peptides. How do these proteins coordinate to not form protease dimers during virion assembly and then form a dimeric protease upon virion release? Also, the protease inhibitors are much less effective at suppressing precursor autoprocessing than suppressing mature PRs. The IC_{50} of darunavir for mature PR is 10 nanomolar. However, 7.5 μM darunavir does not inhibit precursor autoprocessing completely (128), suggesting that the protease precursor is enzymatically different from the mature PR. Thus, we

focus on the mechanism of precursor autoprocessing, the structure of the precursor, and drug screening to inhibit precursor autoprocessing.

The Gag-Pol polyprotein is the full-length precursor. When the Gag-Pol precursor undergoes autoprocessing, the Gag-Pol has to cleave two cleavage sites to release the mature PR: the N-terminal cleavage site (between p6* and PR) and the C-terminal cleavage site (between PR and reverse transcriptase). When the C-terminal cleavage site is blocked by mutation, the Gag-Pol precursor autoprocesses into a PR-RT fusion protein (129), which still supports viral replication. In contrast, blocking the N-terminal cleavage site leads to the detection of a p6*-PR fusion protein that is only capable of processing some cleavage sites in the Gag and Gag-Pol polyprotein (91, 130), so these virions are non-infectious. Since blocking the N-terminal cleavage site affects the catalytic activity of PR, this cleavage site is considered a critical step for releasing mature PR. The p6*-PR is thus defined as a miniprecursor PR. The p6* peptide is predicted to be intrinsically disordered (131, 132), indicating that p6* alone does not exhibit any defined structure. However, it may show an ordered structure and exhibit some regulations with its binding partners or in a specific environment. From the Gag-Pol to mature PR, there might be different intermediate precursors all containing the mature PR as part of them. We have used p6*-PR as a model precursor as it is proven to be a simple yet powerful system that has recapitulated all the autoprocessing phenotypes observed with the proviral constructs (96, 127, 133-137).

The PR precursors are proteolysis competent when they are expressed *in vitro* (138, 139), in *E.coli* (133, 134, 140, 141), or in mammalian cells (128, 133-135, 142, 143). Each system uses different PR-containing constructs. Model systems have been developed for the study of precursor

autoprocessing but each has pros and cons. The first model system is an *in vitro* purified PR miniprecursor (TFR-PR). The transframe region (TFR) is equivalent to the p6* peptide, and contains 8 amino acids of the transframe peptide (TFP) and 48 amino acids of $p6^{pol}$ (107). There is a cleavage site between TFP and $p6^{pol}$ (90, 141). Louis *et al* purified the pseudo-wild-type PR miniprecursors carrying R8Q mutation from inclusion bodies of *E.coli* and refolded them to test their autoprocessing activities at various pH levels (141). At pH 4-5, the R8Q miniprecursors process the cleavage sites within the TFR first, then process the cleavage site between the TFR and PR that releases mature PRs. However, R8Q miniprecursors directly process the cleavage site between the TFR and PR to release mature PR at pH 6.5. This indicates that precursor autoprocessing for preferential cleavage is sensitive to pH. This model system has several disadvantages: (i) a small portion of the miniprecursor was already autoprocessed at the cleavage sites within the TFR during precursor purification; (ii) the *in vitro* refolding step may provide a conditions that favors certain conformations that are not necessarily the ones in the infected cell – the Gag-Pol polyprotein does not undergo a denaturing-and-renaturing cycle, which indicates the conformational change from precursor to mature PR is difficult to study using this method; (iii) It is hard to investigate how viral and cellular proteins regulate precursor autoprocessing in this model; (iv) The pseudo wild-type miniprecursor already has several changed residues to become resistant to self-degradation. These changes by themselves may already influence autoprocessing activity (128, 135). Many mutant p6*-PR sequences found in patients with evidence of PI-resistance didn't refold appropriately in this model system. Thus, the information obtained with this model system is limited to the R8Q pseudo-wild-type precursor.

Another model system utilizes a transcription-coupled-translation system to translate PR precursors in rabbit reticulocyte lysates in the presence of [^{35}S]- methionine or cysteine to label precursors during translation (138, 139, 144). The autoprocessed [^{35}S]-labeled cleavage products were analyzed by radiography according to the molecular mass of the products. The full-length Gag-Pol exhibits proteolytic activity and produces several cleavage products. However, the in vitro translated Gag-Pol precursor didn't release the mature PR in this model system. It indicates that rabbit reticulocyte lysates do not provide essential materials for precursor to release the mature PR. In this model, an initial cleavage site occurs at SP1/NC (MA-CA-SP1↓NC-p6*-PR-RT-IN) followed by cleavage within p6* (NC-TFP↓p6^{pol}-PR-RT-IN). This study also showed that the initial cleavages within Gag-Pol are "intramolecular." The dimerization theory is that Gag-Pol must dimerize to have a catalytic site for processing the substrate. Since RT is a dimeric protein, Gag-Pol dimerization may facilitate autoprocessing. However, some reports demonstrate that p6* prevents p6*-PR dimerization and Gag-PR has a higher proteolysis activity by deletion of p6* than wild-type Gag-PR (138, 141). How p6* and PR regulate precursor dimerization is still a mystery. Another argument is that the concentration of Gag-Pol is ~0.2 nM. The dissociation (Kd) of p6*-PR (~700 nM) is much higher than the Kd of mature PR (~5 nM) (141). How does the Gag-Pol polyprotein form a dimer at such low concentrations? An additional study showed that the processing of Gag-Pol carrying a first residue mutation of PR (P1A) has additional cleavage sites (144). This is consistent with other models that change the residue within the PR and find it influences the cleavage preference of precursor (128, 135).

The third model uses a proviral construct that mediates expression of viral components needed for virion assembly and production in transfected mammalian cells. The released virus-like particles

(VLPs) are collected from the culture medium by centrifugation in order to better understand precursor autoprocessing and to further examine virus maturation. However, ~5% GagPol polyproteins are translated by a -1 frameshift and the remaining products are Gag polyproteins. Thus, the detection level of Gag-Pol or mature PR from VLPs is low. Moreover, there is currently no available highly sensitive antibody to detect the mature PR. Instead, this model examines how Gag-Pol, intermediate precursor, and mature PR process the Gag polyproteins. For example, the cleavage site (CA-SP1) is processed by mature PR, but not by precursor. The p25 is an intermediate product (CA-SP1), and the p24 is a final processing product (CA). Therefore, the detection of p24 in VLPs is a readout for the activity of mature PR for Gag processing. For instance, VLPs produced by wild-type Gag-PR contain cleavage product p24 proteins (CA) (135, 143). In contrast, VLPs produced by Gag-PR carrying H69D only contain full-length Gag polyproteins, indicating that H69D in the context of Gag-PR abolishes precursor autoprocessing and impacts protease maturation. Moreover, the PR is capable of processing all cleavage sites whereas the precursor only processes some cleavage sites. A study using this system showed that blocking the cleavage site p6*-PR within Gag-Pol prohibits the release of mature PRs (91). This study found that the PR precursors (p6*-PR) are able to process several cleavage sites (SP1/NC, NC/p6*, PR/RT, and RT/IN). The benefit of this system is that it has viral and cellular proteins to regulate precursor autoprocessing and it can evaluate the virus maturation or immaturation by cleavage products. However, some cleavage sites of Gag polyproteins can process by either mature PR or precursor PR. Also, the ordered processing of Gag-Pol is not fully understood. It is difficult to understand the entire mechanism of proteolytic processes by Gag-Pol, intermediate precursors, and mature PR.

The fourth model, established by our lab, uses fusion precursors expressed in cells to examine autoprocessing activity. A typical fusion precursor contains the miniprecursor PR (p6*-PR) sandwiched between different proteins or small peptide epitopes, such as maltose-binding protein (MBP) or glutathione S-transferase (GST) at the N-terminus and HA or Flag at the C-terminus. This model provides a simple assay to study autoprocessing reactions inside cells, *i.e.* a biologically relevant environment and the participation of cellular proteins. Our studies demonstrate that (i) The H69E abolishes the precursor autoprocessing in the context of proviral construct whereas the H69E has a partial activity to process the substrate (133). Furthermore, the protease inhibitors are less effective at suppressing precursor autoprocessing than suppressing mature PRs (128), suggesting that the precursor is enzymatically different from the mature PR. (ii) Fusion precursors carrying the different residues within the PR affect their autoprocessing activity and cleavage preference (128, 134), and these data are consistent with the result of VLPs in the context of the proviral construct (135). (iii) Different contexts or fusions could modulate the cleavage reactions and display different sensitivities of drug inhibition (128). This means that the context or fusion could modulate catalytic site conformation and further regulate the precursor autoprocessing. (iv) We also found covariance pairs either within the PR or between p6* and PR, and the different variants of covariance pairs could be involved in the regulation of precursor autoprocessing (143). This simple assay can provide a protein detection profile of precursor PRs as well as mature PRs by increasing drug concentration. This model also provides a better understanding of drug effects on the two PRs.

1.4 Thesis Rationale

My research goal is to better understand the regulation mechanism of HIV-1 protease precursor autoprocessing. This goal was built upon the working hypothesis that mature PR and its precursors are enzymatically and conformationally different and precursor autoprocessing must be temporospatially regulated in concert with virion assembly and release. In chapter 2, we demonstrate that precursor autoprocessing activity and its outcomes can be differentially regulated depending on its context. In chapter 3, results are presented of studies to obtain purified fusion precursors from *E. coli* under non-denaturing conditions for biophysical analyses of precursor structures and/or conformations. Work is still required to improve the expression and purification protocols to obtain pure and homogeneous samples. In chapter 4, we report a *trans* cleavage assay to specifically examine precursor enzyme activity in correlation with factors that influence precursor dimerization. This is the first report demonstrating that p6* could regulate the precursor dimerization. In Chapter 5, additional experiments are summarized to further advance our understanding of autoprocessing regulation mechanism.

CHAPTER 2

CONTEXT-DEPENDENT AUTOPROCESSING OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 PROTEASE PRECURSORS

2.1 Introduction

The Human Immunodeficiency Virus type 1 (HIV-1) protease (PR) is one of the three enzymes encoded by the viral genome with the other two being reverse transcriptase (RT) and integrase (IN). In the HIV-infected cell, these enzymes are initially synthesized as part of the Gag-Pol polyprotein precursor which shares the same N-terminus with the Gag structural precursor polyprotein. Within the Gag-Pol precursor, PR is flanked by an upstream peptide sequence and by the downstream RT. The upstream peptide is named transframe region (TFR) or p6* (140, 145) as its coding sequence overlaps with the p6 region of the Gag reading frame. During the late stage of virus replication, Gag and Gag-Pol co-assemble into virus particles that subsequently bud off from the infected cell. Upon or shortly after virion release, the Gag-Pol polyprotein undergoes autoproteolysis and liberates free mature PR – a process generally referred to as PR autoprocessing.

Two cleavage reactions are necessary to release mature PR from its Gag-Pol polyprotein precursor: one at the N-terminus and the other at the C-terminus. Blocking the C-terminal cleavage leads to production of a PR-RT fusion enzyme that still supports productive viral replication (129). In contrast, blocking the N-terminal cleavage leads to production of p6*-PR that exhibits limited proteolytic activity but is incapable of producing infectious virions (91). Removal of the p6* region

is concurrent with the appearance of mature PR activity *in vitro* and p6* deletion also enhances *in vitro* PR autoprocessing (145, 146). Based on several lines of study, the cleavage between p6* and PR is a critical step for liberation of fully active mature PR (96, 107, 136, 147) and the p6*-PR is thus defined as a miniprecursor.

HIV-1 protease autoprocessing is an intriguing process in that the Gag-Pol polyprotein precursor serves as both the substrate and the enzyme prior to liberation of the mature PR. We recently developed a model system to study the autoprocessing mechanism by expressing fusion precursors in transfected mammalian cells (96, 136, 147). A typical fusion precursor consists of the p6*-PR miniprecursor (derived from the NL4-3 strain) sandwiched between GST and a small epitope peptide such as HA derived from the Influenza virus hemagglutinin protein (Fig 1A). This model system allows us to directly examine precursor-mediated liberation of mature PR from the fusion precursors. With this model system, we have reported that precursor autoprocessing is more resistant than mature PR to suppression by known PR inhibitors (136, 147), which is consistent with previous observations made with purified recombinant miniprecursor (148-150) or *in vitro* translated Gag-Pol polyprotein precursor (151). Thus, this model system provides an easy and simple tool for the study of the precursor autoprocessing mechanism.

We previously tested several other tags in the place of GST and found that all the resulting fusion precursors are autoprocessing competent (136). This raised a fundamental question: is p6*-PR miniprecursor autoprocessing autonomous and independent of the flanking sequences? To address this question, we examined fusion precursors carrying slightly different versions of maltose binding protein (MBP) at the N-terminus. Strikingly, our study demonstrated that precursor

autoprocessing outcomes are context-dependent. The mature PRs released from different fusion tags demonstrated distinct self-degradation properties and H69D point mutation displayed different effects on precursor autoprocessing activity when fused to different tags. Furthermore, the MBP derived 22 aa signal peptide (SigP) modulated precursor autoprocessing in a manner similar to a NL4-3 derived proviral construct. Therefore, our study provides evidence suggesting that precursor autoprocessing is a context-dependent process and can be modulated by sequences beyond the p6*-PR region to release autoprocessing products with distinct enzymatic properties.

2.2 Materials and methods

DNA mutagenesis

All the plasmids used in this study were constructed by the standard PCR-mediated mutagenesis and cloning procedures. The parental GST and L-MBP fusion plasmids were previously described (133, 143, 147) and each engineered mutation made for this study was verified by sequencing analysis (see the support text file for details). A standard agreement is followed for any material transfer.

Cell culture and transfection

HEK 293T cells were purchased from ATCC (Manassas, VA) and maintained in DMEM medium containing 10% fetal bovine serum, 100 units/ml of penicillin G sodium salt and 100 μg/ml of streptomycin sulfate. Transfection of HEK293T cells by calcium phosphate was described previously (135, 143, 147). In brief, HEK 293T cells were seeded in a 12-well the day before transfection to achieve 30-40% confluence at the time of transfection. Chloroquine was added into

each well to a final concentration of 25μM. DNA-Calcium mixture was first made by mixing a total of 0.5 μg plasmid DNA in 65.7 μl H_2O with 9.3 μl of 2 M $CaCl_2$. Then 75 μl of 2 x HBS (50 mM Hepes, pH7.04~7.05, 10 mM KCl, 12 mM Dextrose, 280 mM NaCl, and 1.5 mM Na_2HPO_4) was added slowly to the DNA-Calcium mixture with gentle vortex. The resulting mixture was then added to each well dropwise. At 7-11 hours post transfection, the culture medium was replaced with fresh chloroquine-free medium with or without PR inhibitors at the indicated concentrations. At about ~30 h post-transfection, cells were gently rinsed with 1x PBS once, and lysed in situ by adding 100 μl lysis buffer A (Tris-HCl, pH8.0, 150mM NaCl, 1% sodium deoxycholate, and 1% Triton X-100) with protease inhibitor cocktail. The cell lysates were collected and subjected to a brief centrifugation (10,000 x g for 2 min) to remove host chromosomes. The resulting post-nuclear supernatants were directly analyzed by western blot or stored at -20°C.

Immunoprecipitation and N-terminal Sequencing of PR fragments

HEK293T cells transfected with L-MBP-p6*-F56C-PR^{L63P}-HA encoding plasmid grown in 3 x T175 flasks were collected by trypsin-EDTA treatment followed by a brief centrifugation (1000 x g for 5 min). The cell pellet was then lysed in 9 ml of 1x PBSX (1% Triton X-100 made in 1x PBS) containing 1 μM indinavir to prevent autoproteolysis. The cell lysate was first incubated with 120 μL of amylose beads (BioLabs, cat# E8021) at 4°C overnight to absorb MBP-containing fragments including the full length unprocessed precursor. The remaining supernatant was then mixed with 150 μL anti-HA slurry (Sigma, cat# A2095). After overnight incubation at 4°C, the beads were washed three times with 1x PBS buffer and the associated proteins were recovered by boiling the beads in 250 μL 1x SDS loading buffer. The immunoprecipitated proteins were resolved in 13% SDS-PAGE and transferred onto a PVDF membrane that was subsequently

stained with Ponceau S for 5 min followed by two rinses with Nanopure water and air dried. The stained bands were excised from the membrane and loaded onto a 494 Procise Protein Sequencer/140C Analyzer from Applied Biosystems for N-terminal sequencing by Edman degradation.

Microscopic Analysis of Precursor Subcellular Distribution

HeLa cells grown on poly-L-lysine coated cover slips were transfected with the eGFP fusion precursor encoding plasmids with X-treme Gene Transfection Reagent (Roche). After 24h incubation, cells were fixed in 4% formaldehyde, nuclei were stained with Hoechst and preserved with diamond antifade mountant (Invitrogen). Images were captured on an inverted fluorescence/differential-interference contrast (dic) Zeiss Axiovert 200M deconvolving fluorescence microscope operated by AxioVision Version 4.5 (Zeiss) software and deconvolved by using the constrained iterative method (AxioVision).

Viral particle production and collection

HEK293T cells grown in 6-well plates were transfected as previously described (135, 143, 147). The culture medium was replaced at 7-11 h post-transfection and the HIV-1 protease inhibitor indinavir was added to the indicated concentrations. At ~50 h post-transfection, the culture supernatants containing viral particles-were clarified by centrifugation at 20,800 × g for 2 minutes and the particles were then pelleted at 20,800 × g for 2 h through a 20% (wt/vol in phosphate-buffered saline (PBS)) sucrose cushion. The pelleted particles were resuspended in 25 ul of 1.5x SDS loading buffer and subjected to SDS-PAGE analysis followed by western blotting.

SDS-PAGE and Western blotting

Approximately equal volumes of the post-nuclear lysates were resolved by SDS-PAGE followed by protein transfer to a PVDF membrane. Primary antibodies used in this study include rabbit anti-GST (Sigma, cat# G7781), mouse monoclonal anti-HA (Sigma, cat# H9658), anti-flag (Sigma, cat# F1804), anti-GAPDH (Millipore, cat# MAB374), mouse monoclonal anti-p24 (152, 153), and rabbit anti-PR (NIH AIDS Reagent Program, cat# 4105). Secondary antibodies included IR800 fluorescence labeled goat anti-rabbit (Rockland, cat#611-132-003) and IR800 goat anti-mouse (Rockland, cat# 610-132-121). The blots were visualized with an Odyssey infrared dual laser scanning unit (LI-COR Biotechnology, Lincoln, Nebraska). To reduce background noise in some blots, the primary antibody was first absorbed against cell lysates made from untransfected 293T cells that were resolved by SDS-PAGE and transferred onto a PVDF membrane. Then the pre-absorbed supernatant was used to probe the blots containing the test lysates.

2.3 Results

Our prototypic fusion precursor consists of the p6*-PR miniprecursor sandwiched between GST and HA tags. The NL4-3 derived p6*-PR has two autoproteolysis sites (Fig 2.1A). One is between p6* and PR, designated as the proximal (P) site, which is also equivalent to the N-terminal processing site essential for liberation of mature PR. The other one is located at the N-terminal region of p6*, defined as the distal (D) site in this study. Precursor autoprocessing at the proximal site produces GST-p6* and PR (with PQIT at N-terminus); whereas precursor autoprocessing at the distal site released GST and p6*-PR (with PQGK at N-terminus). With GST fusion, these two sites are equally processed as indicated by detection of approximately equal amounts of GST (from

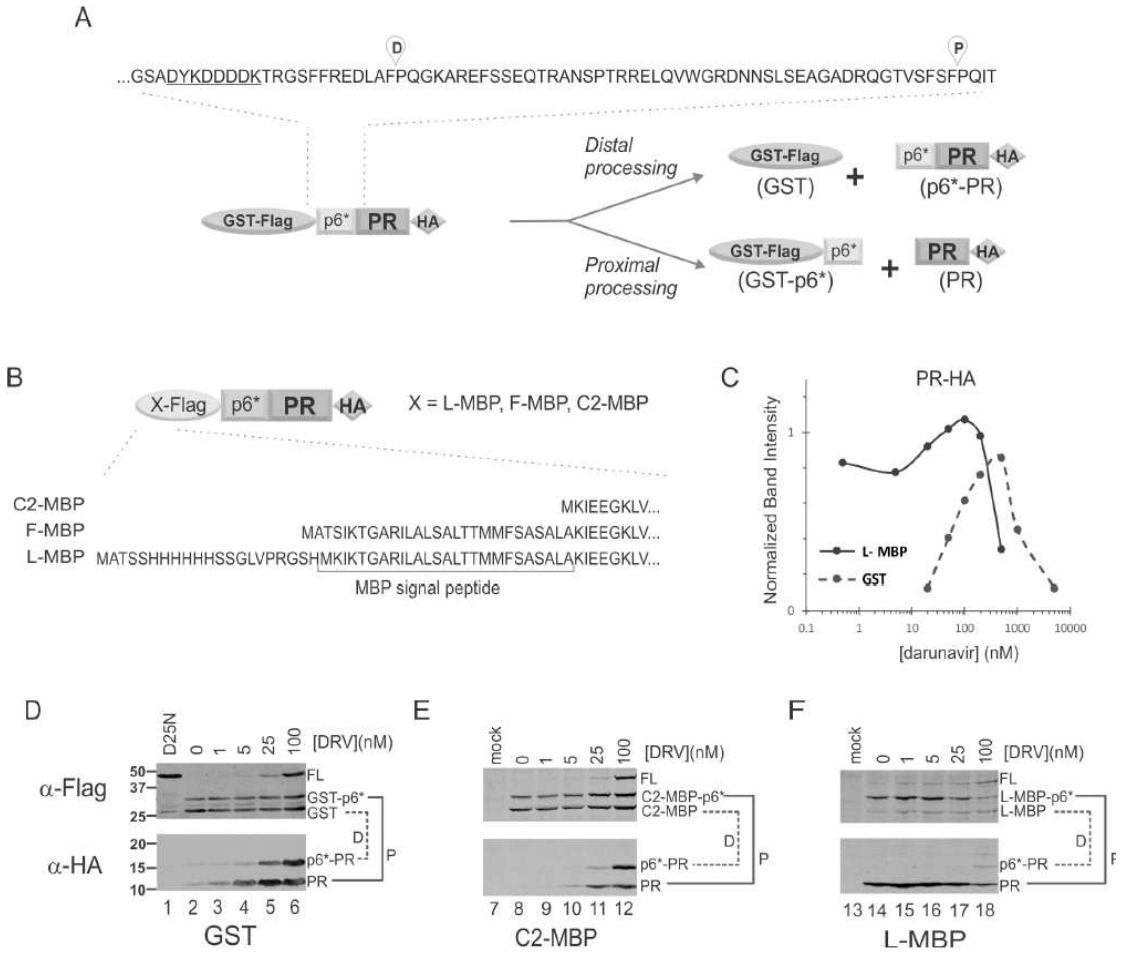

Figure 2.1 MBP fusions modulate autoprocessing products. (A) A schematic diagram illustrating prototypic GST fusion precursor distal and proximal processing sites, and the corresponding products. The Flag epitope sequence is underlined. (B) Domain organization and N-terminal sequences of MBP fusion precursors. (C) Quantification of the mature PRs released from GST- (dashed line) or L-MBP- (solid line) precursors in transfected HEK293T cells. The mature PR was detected with mouse anti-HA antibody and normalized to GAPDH signal to reflect its steady-state amount. The graph is representative of five independent experiments. (D-F) Autoprocessing of various fusion precursors at low DRV concentrations (<100nM). The corresponding proximal (P) and distal (D) processing products are denoted and linked by the solid and dashed lines, respectively.

distal processing) and GST-p6* (from proximal processing) products (Fig 2.1D α-Flag panel as an example). On the other hand, p6*-PR and PR are normally undetectable due to rapid self-degradation in the absence of any PR inhibitor (103, 104). These PR-containing products become detectable when self-degradation is suppressed by PR inhibitor treatment (Fig 2.1D α-HA).

Mature PR released from L-MBP fusion precursor is self-degradation resistant

To investigate whether p6*-PR autoprocessing could be affected by the upstream flanking sequences, we tested maltose binding protein (MBP) as a fusion tag in three slightly different versions (Fig 2.1B). The F-MBP has the full-length coding sequence derived from *E. coli*, including the signal peptide (SigP) responsible for MBP export into the periplasm. The L-MBP has a 6xHis tag fused to the N-terminus of F-MBP through a short linker. The C2-MBP lacks the SigP and is widely used to solubilize fusion proteins in *E. coli* and mammalian cells (154-157).

The C2-MBP precursor displayed similar autoprocessing phenotypes in response to PR inhibitor treatment as the GST precursor (Figs 2.1D-E). Both were autoprocessing competent and processed the P and D sites at approximately equal rates. The mature PRs released from these two fusion precursors exhibited a characteristic bell-shaped detection profile when treated with increasing darunavir (DRV), a potent protease inhibitor (Fig 2.1C, dashed line). The released mature PR is normally undetectable due to self-degradation (103, 104) in the absence of a PR inhibitor. Low DRV concentrations (<500 nM) suppressed PR self-degradation without affecting precursor autoprocessing, resulting in increased detection of the mature PR. As the DRV concentration is increased further, precursor autoprocessing is inhibited leading to less production of the mature

PR. Comparable results were also obtained with saquinavir (SQV), another PR inhibitor (data not shown).

The L-MBP precursor exhibited different autoprocessing phenotypes compared to the C2-MBP and GST precursors. The mature PR released from L-MBP precursor was readily detectable in the absence of PR inhibitor and remained so at low DRV concentrations (<100nM), suggesting that it was not rapidly self-degraded as those released from GST and C2-MBP precursors (Fig 2.1F, α-HA panel). N-terminal sequencing analysis of the mature PR made by the L-MBP precursor verified the PQITL N-terminus as a typical mature PR. Furthermore, F-MBP precursors also released mature PRs that were resistant to self-degradation (data not shown). Therefore, the mature PRs released from L-MBP and F-MBP precursors displayed a distinct self-degradation property unlike those released from GST and C-MBP precursors (Fig 2.1C). Our data suggest that p6*-PR precursor autoprocessing outcomes could be influenced by different contexts, *i.e.*, fusion tags, such that the L-MBP- and the GST- (or C2-MBP-) precursors may have different intrinsic proteolytic activities and therefore release mature PRs with different properties. It should be noted that not all precursors undergo productive autoprocessing with release of anticipated autoprocessing products; a portion of them may directly undergo self-degradation leaving no trace of any defined products. We focused our quantification analysis on the released mature PRs (Fig. 2.1C), which demonstrated different outcomes illustrating the context-dependent characteristic of precursor autoprocessing.

L-MBP fusion abolishes H69D autoprocessing

We previously reported that H69D mutation completely suppresses PR autoprocessing in the context of a NL4-3 proviral construct but only partially inhibits precursor autoprocessing in the context of GST fusion (135, 136, 147). We speculated that this context-dependent discrepancy was indicative of a modulating determinant(s) associated with the NL4-3 proviral sequence but absent in the GST fusion precursor. In the light of different autoprocessing outcomes mediated by L-MBP vs C2-MBP fusion, we examined the possibility that H69D autoprocessing could also be differentially affected by the sequence upstream of p6*. We also included H69Q as another control as it demonstrated wild type autoprocessing activity in our previous reports (135, 136, 147). As shown in Fig 2.2B, with C2-MBP fusion H69D autoprocessing was partially inhibited as indicated by detection of autoprocessing products plus the full length precursor (Fig 2.2B, lane 14), which was similar to the outcome of GST fusion (135, 136, 147). In contrast, with L-MBP fusion H69D completely suppressed precursor autoprocessing as only the full-length precursor was detected and was comparable to the D25N control (Fig 2.2B, lane 5). These results indicate that H69D autoprocessing is differentially affected by the L-MBP vs C2-MBP tag. Therefore, both proviral context and L-MBP fusion abolish H69D autoprocessing although the underlying mechanism remains to be defined.

We next asked whether the p6* peptide has a role in modulating H69D autoprocessing activity by testing L-MBP precursors with various p6* truncations. Our data revealed that the mature PRs released from these L-MBP precursors remained readily detectable (Fig 2.2C, lanes 16-20) and H69D mutation also fully abolished autoprocessing activity of these precursors (Fig 2.2C, lanes 23, 26, 29). The M5 and M6 constructs do not have the Flag tag (Fig 2.2A) and thus are only detectable by HA antibody (Fig 2.2C). The M6 precursor has only the last 3 residues of p6* and was barely active at autoprocessing (Fig 2.2C, lane 27), likely due to disruption of the substrate

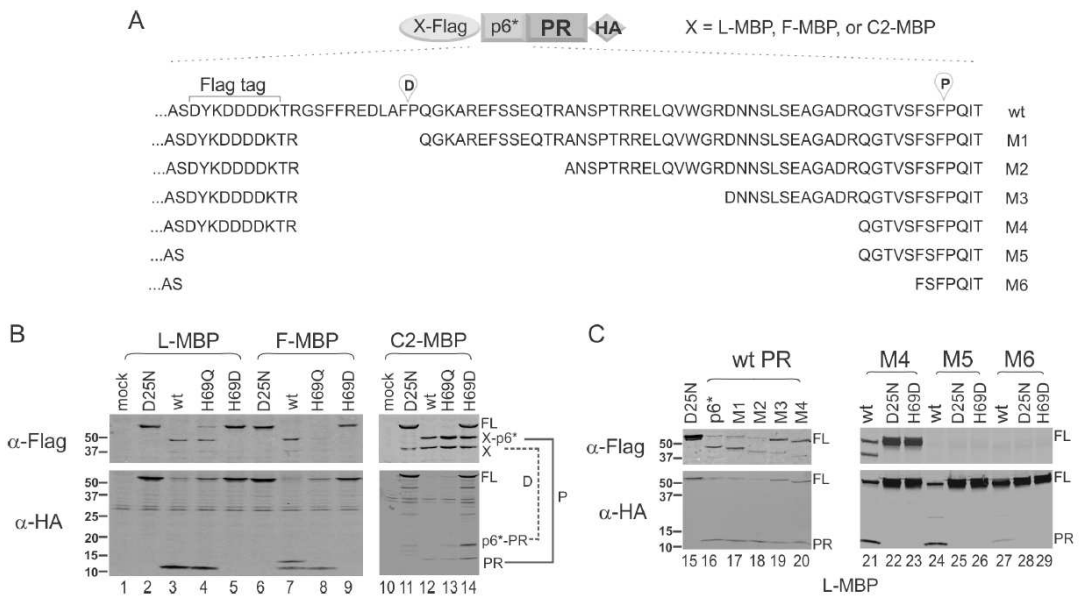

Figure 2.2 H69D mutation abolishes autoprocessing in the context of L-MBP/F-MBP fusion independent of p6* sequences. (A) Schematic illustration of the tested constructs. (B) Autoprocessing of H69 mutants in the context of MBP fusions. Approximately equal amounts of lysate were probed with mouse anti-Flag and anti-HA antibodies in parallel. The corresponding proximal (P) and distal (D) processing products are denoted and connected by the solid and dashed lines, respectively.(C) Autoprocessing of L-MBP fusion precursors carrying p6* truncations.

sequence at the proximal site. In any case, the H69D mutation in the context of the M6 precursor behaved the same as the D25N mutation (lanes 28, 29). Collectively, the L-MBP appeared to influence H69D precursor autoprocessing in a p6*-independent manner.

MBP SigP is sufficient to modulate precursor autoprocessing

An obvious difference between L-MBP (or F-MBP) and GST (or C2-MBP) is the MBP SigP (Fig 2.1B). To determine whether the SigP by itself is sufficient to alter the autoprocessing outcome, we first constructed and tested Flag-M1-PR-HA that does not have any bulky fusion tag at the N-terminus and has the distal cleavage site truncated (Fig 2.3A). This precursor autoprocessed effectively like the GST fusion precursor (Fig 2.4) and was used as the parent precursor in the subsequent experiments. The mature PR released from the Flag-M1 precursor displayed a typical bell-shaped detection profile similar to that made by GST precursors (Fig 2.3B, gray line; Fig 2.5). When treated with micromolar indinavir, these two constructs also released additional HA-reactive bands that ran to the positions between mature PR and an intermediate designated as p6*-PR^{b} in a previous report (142) and below. We did not further characterize these bands as they were only detectable at high concentrations of protease inhibitors. Compared to Flag-M1-PR-HA, SigL and SigP have N-terminal extensions both containing the MBP signal peptide (Fig 2.3A). The mature PRs released from SigL and SigP precursors were easily detectable in the absence of PR inhibitor and exhibited detection profiles like those released from L-MBP or F-MBP fusion precursors in response to DRV treatment (Fig 2.3B; Fig 2.5). Our data demonstrated that inclusion of the MBP SigP was sufficient to alter the self-degradation property of the released mature PRs.

We also examined several H69D-containing precursors as another readout of autoprocessing modulation. The WT Flag-M1 was autoprocessing competent and thus showed little or no

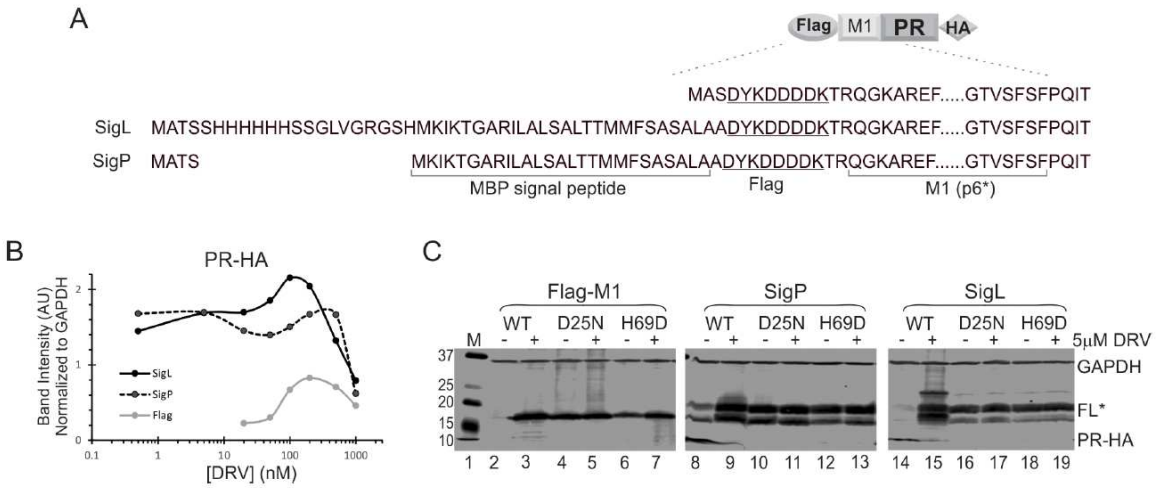

Figure 2.3 The MBP SigP is sufficient at modulating precursor autoprocessing. (A) Schematic diagram of the tested mini-fusion precursors. Flag-M1-PR-HA is the parental construct to which various signal sequences were added N-terminally. (B) Quantitative comparison of the mature PRs released from the indicated precursors. The graph is representative of five independent experiments. (C) Influences of SigL and SigP on H69D autoprocessing activity.

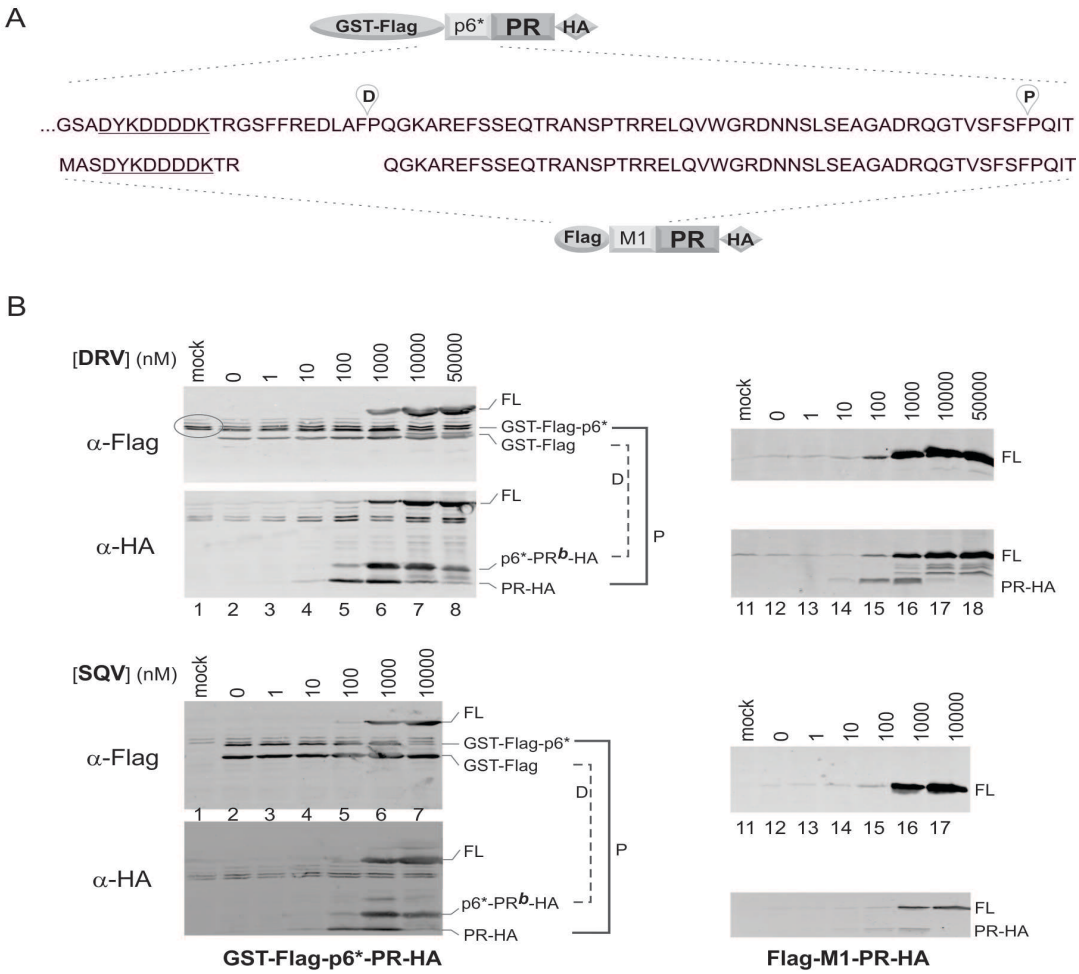

Figure 2.4 Autoprocessing of precursors with or without GST at the N-terminus. (A) Schematic diagram of the fusion precursors. (B) Transfected HEK 293T cells were treated with increasing concentrations of DRV (upper panels) or SQV (lower panels) for 24h. Cell lysates were examined with mouse anti-Flag and anti-HA antibodies. The solid and dotted lines connect the products released from proximal (P) and distal (D) processing, respectively. The circle in panel B denotes indicates a nonspecific band that co-migrated with the GST-Flag-p6* product.

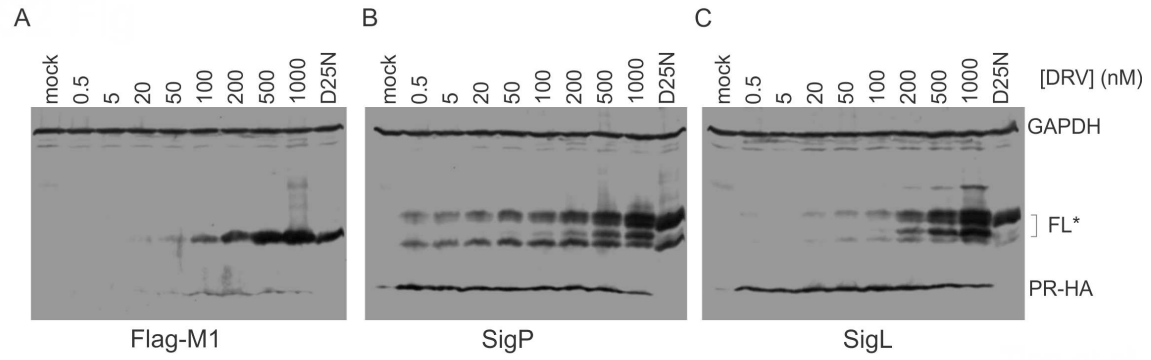

Figure 2.5 DRV sensitivity of mini fusion precursors with or without SigP. Transfected HEK 293T cells were treated with DRV at the indicated concentrations for 24h. Cell lysates were analyzed by SDS-PAGE followed by western blotting with mouse anti-HA and anti-GAPDH antibodies. The asterisks indicate the full-length precursors.

detectable full-length precursor unless treated with 5μM DRV to suppress autoprocessing (Fig 2.3C lanes 2, 3). The D25N protease-deficient Flag-M1 precursor was autoprocessing deficient showing approximately equal amounts of unprocessed precursor with or without 5μM DRV (Fig 2.3C lanes 4, 5). The Flag-M1 precursor bearing the H69D mutation reproducibly exhibited partial activity (Fig 2.3C lanes 6, 7) as about 60% full-length precursor was detected without 5μM DRV compared to with 5μM DRV treatment. In contrast, with SigP or SigL fusion, H69D was mostly autoprocessing-deficient, similar to the D25N controls, showing no difference in detection of the full-length precursor with or without 5 μM DRV (Fig 2.3C lanes 12, 13, 18, 19). Therefore, the MBP SigP alone was sufficient to abolish H69D autoprocessing.

The MBP SigP consists of an N-terminal basic hydrophilic segment (SB) followed by a hydrophobic core (SHC). The last six residues are believed to be recognized by cellular signal peptidases (158). To determine whether there is a specific linear motif responsible for the observed modulation, we engineered mini fusion precursors carrying different segments (SB or SHC) of MBP SigP (Fig 2.6, panel A). The SR2a sequence was previously reported to retain all the essential features required for MBP export in *E. coli* despite missing a portion of the hydrophobic core; the SR2b sequence, carrying a point mutation compared to the SR2a, is deficient in MBP export (159). Both SHC and SR2a precursors maintained the ability to produce mature PRs resistant to self-degradation (Fig 2.6, panel B, lanes 7-10). In contrast, the mature PRs made by SB and SR2b were mostly degraded in the absence of PR inhibitor and only became detectable following 0.2μM DRV treatment (Fig 2.6, panel B, lanes 5, 6, 11, 12).

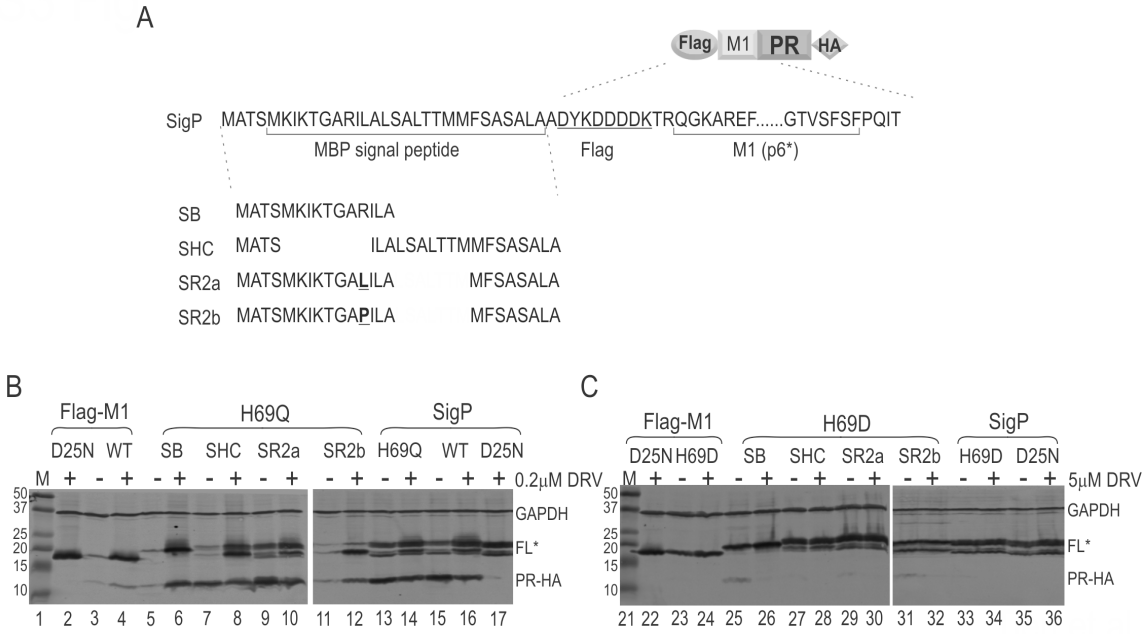

Figure 2.6 Effects of SigP fragments on precursor autoprocessing. (A) Schematic diagram of the tested mini-fusion precursors. Flag-M1-PR-HA is the parental construct to which various signal sequences were added N-terminally. (B-C) Influenced of SigP fragments on mature PR with or without 0.2μM DRV to suppress mature PR self-degradation (*panel B*), and on H69D autoprocessing with or without 5μM DRV to block precursor autoprocessing (*panel C*).

Therefore, SHC and SR2a retained the modulatory effect of releasing self-degradation-resistant PRs whereas SB and SR2b did not. In the context of the SHC and SR2a fusions, H69D autoprocessing was mostly abolished (Fig 2.6, lanes 27-30). These two constructs also showed trace amounts of mature PR, but quantification analysis showed no difference in full-length precursor detection with or without 5μM DRV treatment, indicating that these H69D precursors were as inactive as they were when treated with 5 μM DRV. In the context of SB, H69D autoprocessing was partially active (Fig 2.6, lane 25), which was like the Flag-M1 H69D control (Fig 2.3C, lane 6). In the context of SR2b, H69D demonstrated a phenotype like SR2a in that similar amounts of the FL precursor were detected with or without 5μM DRV. Collectively, our data demonstrated that the hydrophobic core (SHC) alone could induce the observed modulation effects but the N-terminal basic hydrophilic segment (SB) by itself was not. Furthermore, SR2a could recapitulate SHC's effects whereas SR2b displayed a mixed phenotype as it released self-degradation-sensitive mature PR (Fig 2.6, lanes 11, 12) but suppressed H69D autoprocessing (Fig 2.6, lanes 31, 32). Note that SR2a and SR2b differ by only one residue at position 8 (R8L in SR2a and R8P in SR2b) and SR2a is competent in MBP export but SR2b is not (159). Collectively, these results support the notion that precursor autoprocessing can be modulated by functional determinants like SHC or SR2a. However, no new linear motif was identified.

Autoprocessing modulation influenced by SigP position

Within L-MBP and F-MBP fusion precursors, the MBP SigP is separated from the p6*-PR by C2-MBP protein (~38 kD), suggesting that its effect is long range at the primary sequence level. We sought to determine whether the MBP SigP could also exert the same modulation effects on PR when an unrelated protein was inserted between them. GST (~25 kD) and hsp70 (~ 70 kD) were

chosen for this test as they were previously utilized as fusion tags (136). Our data demonstrated that the MBP SigP at the N-terminus retained its ability to modulate autoprocessing outcomes even when it was separated by GST or hsp70 from M1-PR. The released mature PRs were self-degradation resistant (Fig 2.7B, lanes 3-9) and H69D mutation abolished precursor autoprocessing as effectively as a D25N mutation (Fig 2.7C, lanes 21-28). The mature PRs released from the control GST or hsp70 fusion precursor without the MBP SigP were rapidly self-degraded in the absence of protease inhibitor and were only detectable when treated with 0.2μM DRV (Fig 2.7B, lanes 14-17). Also, the H69D fusion precursors without MBP SigP showed partial autoprocessing activities (lanes 33-36). Therefore, the p6*-PR mediated autoprocessing can be modulated by the MBP SigP even when hsp70 was inserted in between.

We then determined if MBP SigP would maintain the modulatory effects when placed in the middle of a fusion precursor immediately upstream of p6*-PR. Surprisingly, although the signal was closer to p6*-PR in the linear sequence, it failed to induce all the previously observed effects. The released mature PRs were susceptible to self-degradation unless suppressed with 0.2μM DRV (lanes 10-13), like the control samples without the MBP SigP (lanes 14-17). On the other hand, the inhibitory effect on H69D autoprocessing was maintained as similar levels of full length precursors were detected with or without 5μM DRV (lanes 29-32). This observation suggested that production of self-degradation resistant mature PR is not necessarily coupled with complete suppression of H69D autoprocessing. Collectively, the results confirmed that the MBP SigP can impact precursor autoprocessing over a long distance to modulate the self-degradation property of the released mature PR when placed at the N-terminus. The MBP SigP can also impede the H69D autoprocessing activity in a position-independent manner.

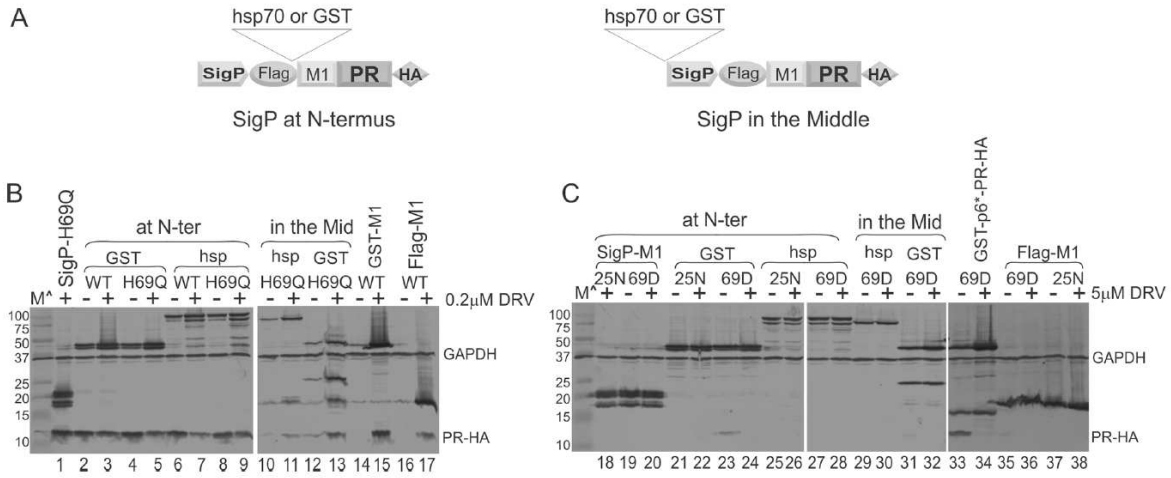

Figure 2.7 Position dependence of MBP SigP on modulation of autoprocessing. (A) Schematic diagram of GST- or hsp70- fusion precursors with MBP SigP placed at the very N-terminus (*left*) or in the middle between the fusion tag and M1-PR (*right*). (B, C) Influences of MBP SigP position on self-degradation of released mature PR (*panel B*) and on H69D autoprocessing activity (*panel C*).

N-terminal SigP alters precursor distribution

To gain further insights into the mechanism underlying the modulation exerted by MBP SigP, we examined the subcellular distribution of fusion precursors tagged with GFP with or without the MBP SigP (Fig 2.8A). Autoprocessing properties of these GFP fusion precursors in transfected HeLa resembled to those observed in transfected HEK293T (Fig 2.8B-C). The mature PR released from the GFP-tagged precursor without MBP SigP was not detectable in the absence of protease inhibitor; it became detectable when treated with 0.2μM DRV (Fig 2.8B, lanes 1, 2). In contrast, the mature PR made by the GFP-tagged precursor with SigP was readily detectable in the absence of protease inhibitor (Fig 2.8B, lane 3). Without SigP, H69D mutation partially suppressed precursor autoprocessing (Fig 2.8C, lanes 5-6); with SigP, H69D abolished precursor autoprocessing as effectively as D25N (Fig 2.8C, lanes 9-12).

Having confirmed that the GFP tag did not alter precursor autoprocessing ability, GFP fusions carrying D25N (to prevent autoprocessing) with or without SigP were examined in transfected HeLa cells by confocal microscopy. Interestingly, the two precursors exhibited distinct distribution patterns. Without SigP, distribution was mostly diffuse; a small percentage of the cells exhibited small puncta (Fig 2.8D). In contrast, precursors with SigP displayed bright clusters that appeared to surround vesicle-like structures (Fig 2.8E). We interpreted this to suggest that SigP mediates precursor targeting to certain membrane vesicles where precursor autoprocessing is indirectly modulated by cellular factors associated with the local environment. The fact that the GFP-tagged fusion precursors were readily detectable excludes their association with organelles such as the lysosome where GFP signals would be quenched by the low pH environment of the lumen. Rather, the results suggest that the subcellular environment modulates precursor autoprocessing.

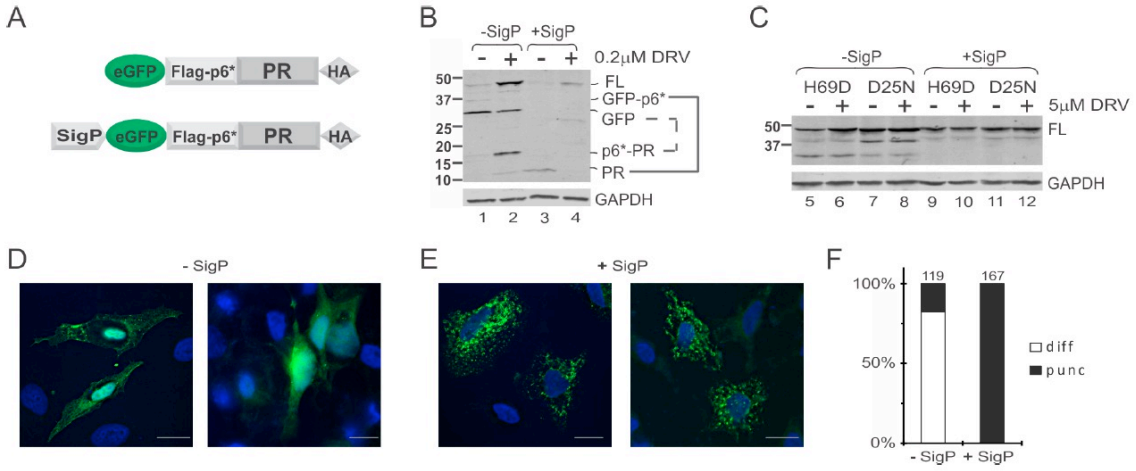

Figure 2.8 MBP SigP alters subcellular distribution of GFP fusion precursors. (A) Schematic diagram of GFP fusion precursors with or without MBP SigP at the N-terminus. (B, C) Western blot assessment of MBP SigP's influences on mature PR self-degradation (*panel B*) and H69D autoprocessing (*panel C*) in transfected HeLa cells. (D, E) Representative images of HeLa cells transfected with plasmids encoding for the D25N GFP fusion without (*panel D*) or with (*panel E*) MBP SigP. Bars: 10 µm. (F) Quantitative analysis of different GFP staining patterns in transfected HeLa cells. The numbers above the bars denote the total GFP positive cells analyzed.

Context-dependent autoprocessing of precursors with mutations identified in variants from HIV-1-infected individuals

We recently described point mutations that emerged in the HIV-1 viruses isolated from a subpopulation of Women's Interagency HIV Study (WIHS) participants where anti-retroviral agents, including indinavir (IDV), had failed to suppress the viral load (160, 161). One of these is a substitution of the C-terminal p6* residue phenylalanine 56 (F56) for cysteine (C; F56C), which thereby alters the P1 residue at the PR N-terminal cleavage site (142). We constructed precursors carrying this point mutation with GST *vs.* L-MBP fusion tags to further test if context-dependent autoprocessing is a general property (Fig 2.9). Our results showed that the mature PR released from the GST fusion precursor displayed a typical bell-shaped detection profile in response to IDV treatment whereas the mature PR released from L-MBP fusion precursor was readily detectable in the absence of any PI and its detection remained steady up to 1μM IDV (Fig 2.9C). This was similar to the wild type precursors (*c.f.*, Fig 2.1C), confirming that the mature PRs released from GST vs L-MBP precursors differed in their self-degradation propensity.

Distal processing of this mutation was also differentially affected by these two tags. The GST precursor produced one distal processing product p6*-PR$^{\mathbf{b}}$ (Fig 2.9A, lower panel) while the L-MBP precursor released two (Fig 2.9B, lower panel). N-terminal sequencing analysis confirmed that both p6*-PR$^{\mathbf{a}}$ and p6*-PR$^{\mathbf{b}}$ had the same N-terminal sequence (PQGKA) despite their different mobility in SDS-PAGE. Of note, the p6*-PR$^{\mathbf{a}}$ fragment was specific to L-MBP fusion (Fig 2.9D) and appeared to be stable over a wide range of IDV concentrations (up to 2μM). The p6*-PR^{b} product displayed a bell-shaped detection profile in response to IDV treatment (Fig 2.9E) as seen with the wild type precursor (136). Consequently, these results support the idea that precursor

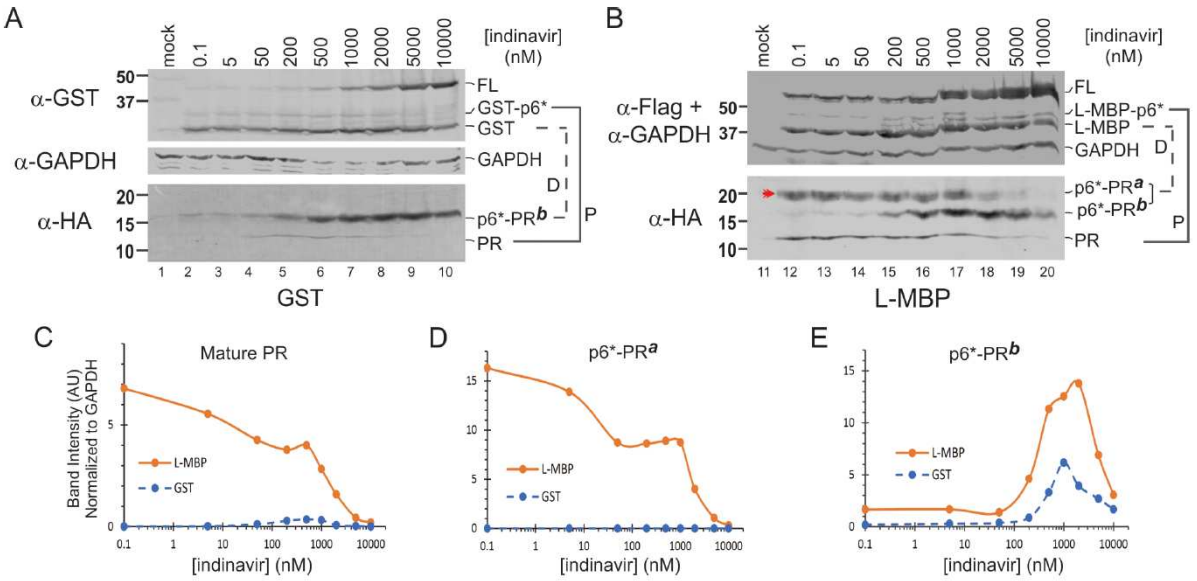

Figure 2.9 Autoprocessing of F56C precursors in the context of GST (A) and L-MBP (B) fusion. Lysates collected from transfected HEK293T cells were resolved by SDS-PAGE and probed with the indicated antibodies for visualization of full-length (FL) precursors and autoprocessing products. The corresponding proximal (P) and distal (D) processing products are denoted and connected with solid or dashed lines, respectively. The double arrow head (panel B, HA blot) indicates a distal processing product unique to the context of L-MBP fusion. Band intensity of mature PR (C), p6*-PR^{a} (D), and p6*-PR^{b} (E) normalized to GAPDH was plotted to IDV concentrations, respectively.

autoprocessing is differentially modulated by GST vs L-MBP tags leading to liberation of products with distinct detection profiles.

Autoprocessing products from the NL4-3 viral particles resemble those released from the L-MBP fusion precursors

In the light of our results showing that precursor autoprocessing is context-dependent, we sought to examine autoprocessing products associated with viral particles released from 293T cells transfected with a NL4-3 derived proviral construct (pNL-MA-HA-ΔIN) (Fig 2.10). The wild type control (Fig 2.10A-B, the upper panels) was compared to a F56C/L63P double mutation in the proviral context. This double mutation displayed phenotypes resembling the F56C single mutation that supported normal Gag processing even with reduced mature PR production due to the suboptimal cleavage site rendered by the F56C mutation at the P1 position (142). The viral particle-associated mature PR and other PR-containing fragments, likely p6*-PR variants, were probed with a rabbit anti-PR antibody (Fig 2.10A). In parallel, the same samples were probed with an anti-HA antibody followed by anti-p24 antibody probing to determine the total Gag proteins and Gag processing (Fig 2.10B). Our results showed that the mature PR in the WT particles was readily detectable and remained so at low IDV concentrations, resembling the mature PR released by the L-MBP fusion precursor. The F56C/L63P double mutation produced less mature PR than the WT control (Fig 2.10C). Using the p6*-PR^{a} and p6*-PR^{b} made by the L-MBP fusion precursor as size references (Fig 2.10A, lane 9), we found several p6*-PR fragments displaying various response profiles to IDV treatment. The p6*-PR^{a} fragment was detected in both constructs with the WT releasing more than the double mutation (Fig 2.10D). Meanwhile, the F56C/L63P produced more p6*-PR^{b} fragment than the WT (Fig 2.10E) but had much less p6*-PR^{c} than the WT (Fig 2.10F).

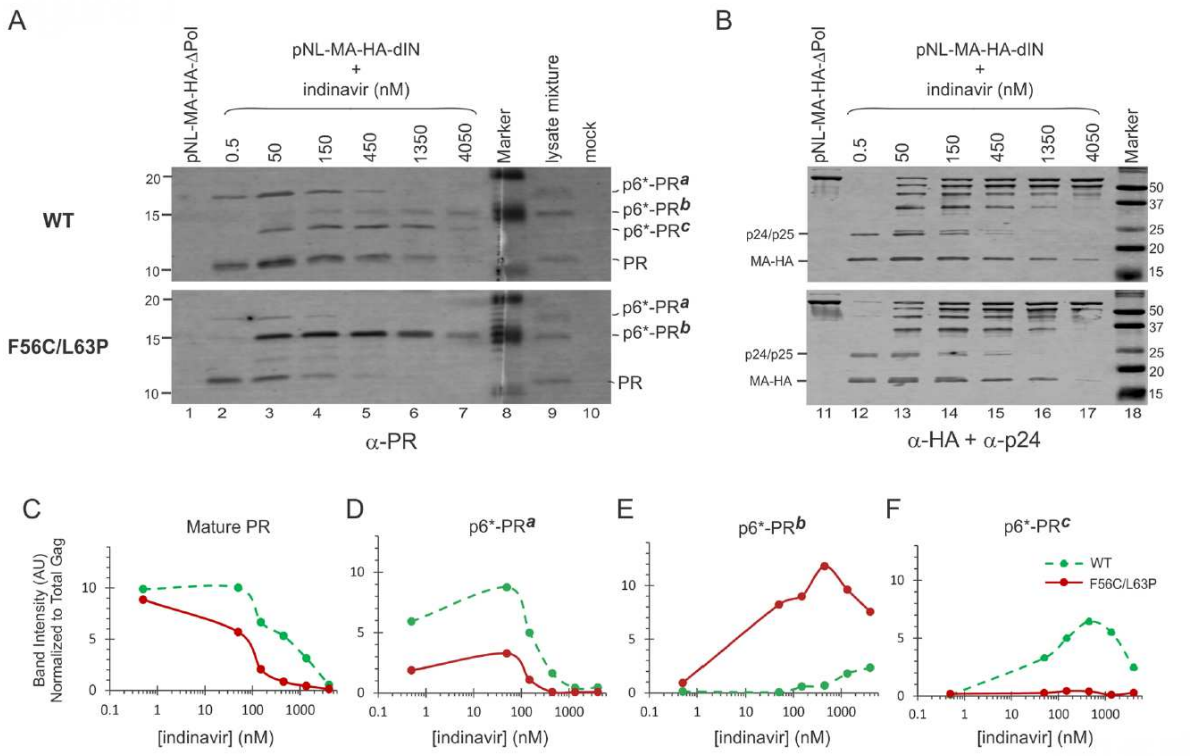

Figure 2.10 Detection and quantification of PRs in viral particles made by the WT and F56C/L63P proviral constructs. (A) The viral particles were probed with a polyclonal PR antibody. A mixture of lysates collected from cells transfected with L-MBP precursors without any C-terminal fusion epitope was included to serve as size references of mature PR, p6*-PR^{a}, and p6*-PR^{b} fragment (lane 9). (B) The same viral particle samples at 5-fold less amounts were resolved in parallel, probed with a HA-antibody first followed by p24 probing. (C-F) quantification of PR-reactive bands normalized to total HA signals representing the total Gag.

Collectively, these data demonstrate the existence of mature PR along with several p6*-PR fragments in viral particles produced in the context of a proviral construct. The SigP-containing precursors therefore appear to recapitulate an autoprocessing property that closely resembles that of the virus.

2.4 Discussion

Mature PRs with diverse self-degradation propensities

Autoproteolysis (self-degradation) is often observed with *in vitro* purified recombinant mature PRs (103, 104, 107, 148, 149) and thus it is widely accepted that all mature PRs are prone to rapid self-degradation. In this report, mature PRs with the same amino acid sequence but distinct (sensitive *vs* resistant) self-degradation properties were detected from precursors with different fusion tags (GST *vs* L-MBP). Accordingly, we suggest that precursor autoprocessing can undergo more than one pathway induced by different contexts leading to production of mature PRs with diverse self-degradation propensities. Of an interesting note, the NL4-3 associated mature PRs were also resistant to self-degradation as the particles were collected 48h post transfection and subjected to a 2h centrifugation step prior to lysis of the particles with SDS in preparation for Western analysis (Fig 2.10A). In this regard, the L-MBP fusion (or the MBP SigP at the N-terminus) is a better mimic of the viral context than the GST or other fusions with regard to liberation of mature PRs that resist self-degradation.

MBP SigP modulation mechanism

The MBP SigP is known to mediate MBP translocation across the plasma membrane into the periplasmic space in *E coli*. However, its biology in mammalian cells is unclear in terms of whether it would mediate targeting of specific membrane organelles and the subsequent trans-membrane translocation. We speculated that MBP SigP-mediated translation initiation and/or membrane targeting might contribute to the different autoprocessing outcomes between precursors with or without it. Once the SigP is synthesized and emerges from the ribosome, it might interact with mammalian proteins that influence its subcellular distribution and/or folding state. When placed in the middle, the SigP accessibility to these cellular components would be reduced. Consistent with this speculation, our subcellular distribution analysis revealed that MBP SigP targeted the GFP fusion precursor to certain vesicle-like structures, probably at the cytoplasmic site (outer rims); whereas the precursor without SigP was diffuse throughout the cytoplasm (Fig 2.10). Given that Gag/Gag-Pol assembly happens at the cytoplasmic side of the plasma membrane, Gag-Pol precursor autoprocessing could be subjected to modulation by cellular/viral factors that are enriched or specific to the assembly sites as well. Collectively, it appears that different autoprocessing outcomes might be correlated with different subcellular locations varying in molecule composition, pH, redox status etc, revealing that precursor autoprocessing is context-dependent and responsive to different subcellular environments.

Context-dependent production of p6*-PR fragments

Many fusion precursors tested in this study carry both distal (D) and proximal (P) cleavage sites (Fig 2.1). Proximal site processing releases mature PR and distal site processing produces p6*-PR. This report demonstrated that various p6*-PR fragments were detected in different contexts: the GST fusion resulted in production of mainly p6*-PR^{b} fragment (Fig 2.9A), the L-MBP fusion led

to production of both p6*-PR^{a} and p6*-PR^{b} fragments (Fig 2.9B), and the NL4-3 particles contained several p6*-PR fragments (Fig 2.10A). Furthermore, the p6*-PR fragments released from different sequences (*e.g.*, wt *vs* F56C/L63P) also exhibited diverse IDV response profiles (Fig 2.10D-F), indicating that precursor autoprocessing could be influenced by different PR sequences leading to liberation of various products We and others previously reported that a p6*-PR fragment resulting from a mutated proximal site can partially process Gag polyprotein (91), suggesting that these p6*-PR fragments could contribute to Gag processing during virion maturation. In this regard, it is interesting to note that F56C/L63P generated more p6*-PR^{b} (Fig 2.10E) but less mature PR, p6*-PR^{a}, and p6*-PR^{c} than the WT control. The fact that this double mutation displayed normal Gag processing and viral infectivity (161) suggests a tag team strategy by which these p6*-PRs and the mature PR work together to make infectious virions. This strategy could also contribute to development of drug resistance as these p6*-PR fragments were detected at high IDV concentrations and their production was less sensitive to inhibition by PR inhibitors (136, 147, 151).

In summary, this report demonstrated that precursor autoprocessing activity and its outcomes are context-dependent. We speculate that the Gag-Pol precursor is structurally and enzymatically flexible such that it can take on different autoprocessing pathways under different contexts leading to diverse outcomes. In line with this concept, Zybarth et al (146) previously observed that a partial truncation of the nucleocapsid (NC) domain of a forced-frameshift Gag-PR precursor interfered with precursor autoprocessing but deletion of both NC and most p6* rescued and enhanced precursor autoprocessing. As another example, a recombinant MBP-fused precursor expressed in *E. coli* was autoprocessing-deficient but became autoprocessing-competent following an *in vitro*

denaturation/renaturation cycle (162). In this case, different cellular contexts (mammalian vs prokaryote) led to different autoprocessing outcomes, supporting the hypothesis that precursor autoprocessing is regulated by sequences outside of the p6*-PR region *in cis* and/or by other cellular factors *in trans*. Additionally, Yu et al (164) reported that enhanced Gag-Pol autoprocessing induced by replacement of the p6* domain with a leucine zipper was suppressed by the N-terminal tetra-peptide. In addition to revealing the complexity of autoprocessing regulation, our findings underscore the necessity of employing physiologically relevant contexts to study the autoprocessing mechanism.

CHAPTER 3

BIOPHYSICAL CHARACTERIZATION OF RECOMBINANT FUSION PRECURSORS MADE UNDER NON-DENATURING CONDITIONS

3.1 Introduction

The goal of the work reported in this chapter is to purify homogeneous miniprecursor (p6*-PR) under non-denaturing conditions. Currently, there are ~100,000 crystal structures reported for the wildtype & mutant PRs with or without protease inhibitors (105, 106). In distinct contrast, no structural information is available for any precursor; the mature PRs with a 4 amino acid N-terminal extension (^{SFNF}PR) in the presence of protease inhibitors are most relevant ones reported in the literature (163). Well-established approach as for studying mature PR structure are not suitable for studying precursor structure. The recombinant mature PRs are mostly associated with inclusion bodies, from which mature PRs are purified under denaturing conditions and refolded *in vitro*. This approach works well for proteins with defined folding pathways including mature PRs. The HIV-1 p6* is predicted to be intrinsically disordered, meaning it could adapt multiple conformations in a context-dependent manner. In the absence of binding partners, heterogeneous conformations are anticipated upon *in vitro* refolding (148, 164). Therefore, we wanted to explore and establish a precursor purification platform suitable for studying precursor structure. Particularly, we would like to avoid the denature-and-renature cycle to reduce conformation heterogeneity.

Our biochemical and molecular virological analyses (127, 133-137, 143, 161) and those of others (82, 91, 139, 148, 149, 165, 166) collectively demonstrated that the catalytic site PIs are much less effective at suppressing precursor-mediated autoprocessing than inhibiting mature PR activity, suggesting that the mature PR and its precursors are enzymatically different. Therefore, we carried out biophysical analyses of precursors to better understand how they differ. Previously, Sri Rao Boddeda, a former student in the lab, was able to purify maltose-binding protein (MBP) fusion precursors from *E.coli* under non-denaturing conditions. Based on his data, we further constructed and tested additional fusion precursor expression vectors for biophysical analyses.

3.2 Material and methods

Plasmid design

The first construct expressed His-DnaK-TEV-Flag-p6*-PR-strep. The full-length precursors are designed to have His and DnaK at the N-terminus followed by a TEV cleavage site (ENLYFQ/G), a Flag tag, and then p6*-PR precursor tagged with Strep at the C-terminus (Fig 3.1A). The DnaK is a chaperone protein of *E.coli*. The fusion precursor or miniprecursor can be detected by His, Flag, and strep antibodies. This design allowed the use of Ni-NTA affinity column to purify the fusion precursor, and then use TEV protease to process TEV cleavage cut site. Finally, we could purify the cleaved miniprecursor by Strep-tactin resin. The second construct is C2MBP-TEV-Flag-p6*-PR-strep (Fig 3.2E). The method of purification is similar to DnaK fusion precursor except that the first step used amylose beads to purify C2MBP fusion precursor. All plasmids used in this study are listed in Table 3.1.

Expression of DnaK and C2MBP fusion precursor

Table 3.1 The autoprocessing activity of fusion precursors in *E coli*. The calculation of autoprocessing activity was based on Flag antibody.

pET15b-DnaK-derived plasmids	Autoprocessing activity
pET15b-DnaK-TEV-Flag-p6*-PRH69Q-Strep	100%
pET15b-DnaK-TEV-Flag-p6*-PRH69D-Strep	25-50%
pET15b-DnaK-TEV-Flag-p6*-PRD25N-Strep	50%
pET15b-DnaK-TEV-Flag-p6*-PRR87K-Strep	25-50%

pMal-C2MBP (slow+helix)-derived plasmids	Autoprocessing activity
pMal-C2MBP(SH)-Flag-M2MG-PRwt	50%
pMal-C2MBP(SH)-Flag-M2MG-PRwt-Strep	50%
pMal-C2MBP(SH)-Flag-M2MG-PRH69D	50%
pMal-C2MBP(SH)-TEV-Flag-P6*-PRwt-Strep	100%
pMal-C2MBP(SH)-TEV-Flag-P6*-PRH69Q-Strep	100%
pMal-C2MBP(SH)-TEV-Flag-P6*-PRH69D-Strep	75%
pMal-C2MBP(SH)-TEV-Flag-P6*-PRR87K-Strep	75%
pMal-C2MBP(SH)-TEV-Flag-P6*-PR1-95-Strep	75%
pMal-C2MBP(SH)-TEV-Flag-M4-PRwt-Strep	100%
pMal-C2MBP(SH)-TEV-Flag-M4-PRH69Q-Strep	100%
pMal-C2MBP(SH)-TEV-Flag-M4-PRH69D-Strep	50%
pMal-C2MBP(SH)-TEV-Flag-M4-PRR87K-Strep	50%

pMal-C2MBP-derived plasmids	Autoprocessing activity
pMal-C2MBP-Flag-p6*-PRwt-HA	25%
pMal-C2MBP-Flag-p6*-PRH69D-HA	25%
pMal-C2MBP-Flag-p6*-PRwt-Strep	25%
pMal-C2MBP-TEV-Flag-p6*-PRwt-Strep	100%
pMal-C2MBP-TEV-Flag-p6*-PRD25N-Strep	25~50%
pMal-C2MBP-TEV-Flag-p6*-PRH69D-Strep	25~50%
pMal-C2MBP-TEV-Flag-p6*-PRR87K-Strep	25~50%
pMal-C2MBP-TEV-Flag-p6*-PR1-95-Strep	25~50%

The pET15b-DnaK-derived plasmids or pMal-C2MBP-derived plasmids were transformed into E. Coli BL21 (DE3) and a single fresh colony was used to inoculate 1 ml Luria-Bertani (LB) medium with Ampicillin and grown overnight at 37 °C. The overnight culture was diluted 100 fold (20 ul) into 2 ml LB medium with Ampicillin and incubated at 37 °C for 2~2.5 hours. Once the OD_{600} reached 0.4~0.8, isopropyl thiogalactoside (IPTG) was added to a final concentration of 0.4 mM to induce expression. The cells were grown at 37 °C for 3 hours and collected by centrifugation (centrifuge 5417R) at 20800 xg for 2 min. The cells were lysed 120 ul 1.5x SDS loading dye and then analyzed by western blot.

For large scale purification, the overnight culture was diluted 100 fold (15 ml) into 3 liter 2x YT medium with Ampicillin and incubated at 37 °C for 2~2.5 hours. When the OD_{600} reached 0.4~0.8, IPTG was added to a final concentration of 0.4 mM to induce expression. The cells were incubated at 37 °C for 3 hours and collected by centrifugation (Beckman J2-21) at 6800 rpm for 30 min at 4 degree.

The purification of DnaK fusion precursor by Ni-NTA affinity column

The cell paste was suspended in Ni-NTA binding buffer (50 mM sodium phosphate (pH 8.0), 200 mM NaCl, 10% glycerol, and 25 mM imidazole) with 10 mL per gram and lysed by micro fluidizer. The polyethylenimine was added to a final concentration of 0.1 % in the cell lysate. The cell lysate was centrifuged at 15000 xg for 30 min to remove cell debris and the supernatant was filtered by 0.45 μm polyethersulfone membrane. The clarified supernatant was applied to Ni-NTA column (Qiagen, cat# 30760) and this column was washed by 6 column volumes of Ni-NTA binding buffer. The DnaK fusion precursor was eluted stepwise by 2 column volumes of Ni-binding buffer

containing 100 mM, 150 mM, 200 mM, and 250 mM imidazole. The EDTA was added to a final concentration of 2 mM in each elution. About 8 mg DnaK fusion precursor could be purified from 1 liter culture.

The purification of C2-MBP fusion precursor by amylose beads

The cell paste was suspended in MBP-binding buffer (50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 5% glycerol, 1 mM EDTA and 1 mM Phenylmethylsulfonyl Fluoride (PMSF)) with 10 mL per gram and lysed by micro fluidizer. The polyethylenimine was added to a final concentration of 0.1 % in the cell lysate. The cell lysate was centrifuged at 15000 xg for 30 min to remove cell debris and the supernatant was filtered through a 0.45 μm polyethersulfone membrane. The clarified supernatant was mixed with amylose resin (NEB, cat# E8021L) for 15 mins at 4 °C and this lysate-resin mixture was washed by 10 column volumes of MBP-binding buffer. The C2-MBP fusion precursor was eluted by 2 column volumes of MBP-elution buffer (50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 5% glycerol, 1 mM DTT, and 10 mM maltose). About 10 mg C2-MBP fusion precursor could be purified from 1 liter culture.

***In vitro* TEV cleavage assay**

A constant amount (1.35 ug) of DnaK fusion precursor R87K substrate was incubated with increasing amounts of recombinant TEV protease in Ni-binding buffer overnight at 4 °C to determine optimal TEV protease:substrate ratios by SDS-PAGE analysis.

***In vitro* TEV cleavage assay on bead**

The DnaK fusion precursors were bound to Ni-NTA beads (Qiagen, cat# 30210) and washed by 6 column volumes of Ni-binding buffer. Based on Isaiah Jackson's data, the optimal ratio for cleavage reaction on beads in Ni-binding buffer is 3 ug TEV protease to 20 µl beads (50 µg DnaK fusion precursor). The C2-MBP fusion precursor bound to amylose beads was washed by 10 column volumes of MBP-binding buffer. The optimal ratio for cleavage reaction on beads in MBP-binding column is 3.3 µg TEV protease to 20 ul beads (200 µg C2MBP fusion precursor).

The purification of strep-tagged miniprecursor after TEV cleavage assay

After TEV cleavage, the reaction mixture was mixed with strep-tactin sepharose beads (IBA, cat# 2-1201) and settled into an empty column. This column was washed by 5 column volumes of tactin buffer W (100 mM Tris-HCl, 150 NaCl, and pH 8) three times. The Flag-p6*-PR-Strep was eluted with 1 column volume of tactin buffer E (100 mM Tris-HCl, 150 NaCl, 5 mM desthiobiotin, and pH 8) three times.

SDS-PAGE gel and western blotting

The protein samples were run on 8%, 10%, 12.5%, or gradient SDS-PAGE gels for Western blot analysis. The SDS-PAGE gel was stained by imperial blue (Thermo, cat# 24615) or transferred onto PVDF membrane (Millipore, cat# IPVH00010). The membranes were probed with primary antibody either mouse anti-Flag (Sigma, cat# F1804) or mouse anti-strep (Iba, cat# 2-1507-001), respectively. The membranes were then incubated with IR800 mouse anti-mouse (Rockland, cat# 610-132-121). Signals were scanned by Odyssey infrared dual laser scanning unit (LI-COR Biotechnology, Lincoln, Nebraska) and analyzed through Image Studio.

3.3 Results

The DnaK fusion precursor bearing R87K, D25N, or H69D was autoprocessing deficient in *E. coli*

We previously reported that H69D mutation abolished precursor autoprocessing in the context of proviral construct but only partially inhibited autoprocessing of a GST-fused precursor expressed in *E. coli* (136). We here used H69D autoprocessing activity to determine whether an engineered fusion precursor was in the biologically relevant conformation: through its recapitulating the proviral construct phenotype. We engineered DnaK fusion precursor (His-DnaK-TEV-Flag-p6*-PR-strep) carrying the different single mutation (Fig. 3.1A). The DnaK is a heat shock protein of *E.coli*, which could in theory facilitate precursor folding. The D25N and R87K mutations are located at the interface of active site in mature PR. These mutants were shown to disrupt mature PR dimerization by NMR spectroscopy (93, 167). The H69Q mutation was shown to display wildtype autoprocessing (134), serving as a positive control in this experiment. We detected two Flag-reactive bands corresponding to the expected autoprocessing products (His-DnaK-TEV-Flag and His-DnaK-TEV-Flag-p6*) from H69Q precursor, indicating that H69Q fusion precursor was autoprocessing competent when expressed in *E. coli* (Figure 3.1B, lane 4). In contrast, the full-length precursor was predominantly detected by Flag antibody from D25N, H69D, and R87K precursor, suggesting that these three mutants abolished autoprocessing activity (Figure 3.1B, lanes 5-7). Of note, we also detected some bands shorter than the full-length precursor but absent in the H69Q positive control. We reasoned that these might be prematurely terminated translation products or they may have resulted from heterogeneous precursors with some sensitive to *E. coli* proteases. Both are not mutually exclusive given that the p6*-PR is sensitive to context-dependent

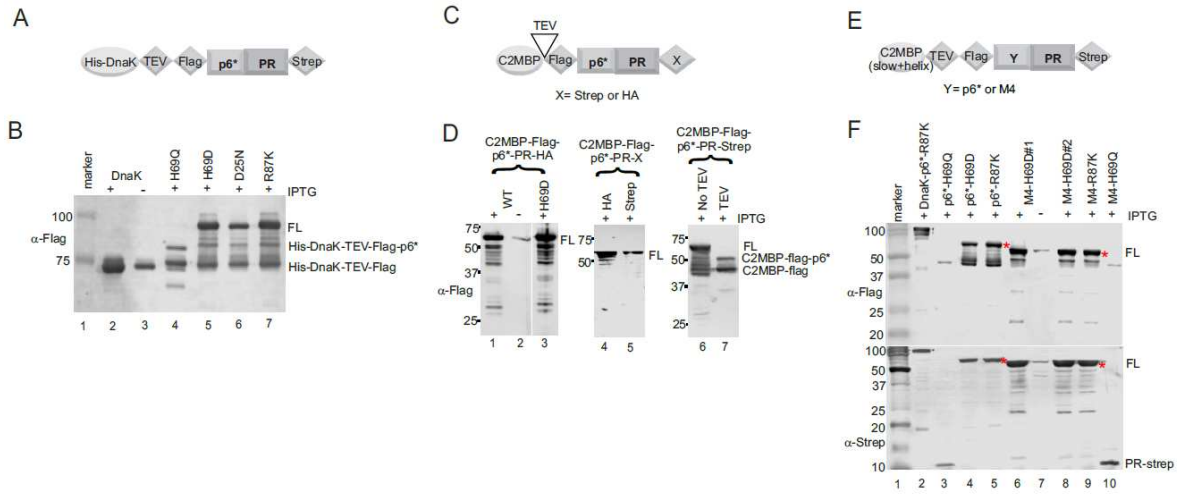

Figure 3.1 Autoprocessing of DnaK- or C2-MBP- fusion precursors carrying single mutation expressed in *E.coli*. (A, C, E) Schematic diagram of DnaK- or C2MBP- fusion precursors with various tags and a single mutant of PR. (B, D, F) Cell lysates collected from *E.coli* BL21 (DE3) were subjected by 8%, 10%, or 12.5% SDS-PAGE and probed with the indicated antibodies. The full-length (FL) and its autoprocessing products are showed at right.

regulation and p6* peptide is intrinsically disordered. Because the full-length precursors were the main species detected in these mutants, we considered the corresponding fusion precursors as good candidates for purification trial.

The C2-MBP fusion precursor bearing mutation R87K, D25N, H69D, or PR1-95 (deletion 96-99) was also autoprocessing deficient within *E. coli*

Additional precursors were examined in the context of C2-MBP fusion. The C2-MBP lacks the N-terminal signal peptide responsible for MBP export into the periplasm and thus expresses as a soluble intracellular protein in *E.coli* (155). We have two versions of C2-MBP. One is the wild-type (Fig 3.1C), and the other one contained slow translation codons and a C-terminal α-helix (Fig 3.1E) that were reported to improve protein solubility and folding homogeneity (168). We also modified the MBP protein containing different surface mutations D82A/K83A/E359A/N362A/D363A and it was truncated after Asn367 (169). These modifications were reported to reduce the surface entropy, which is beneficial to allow crystallization for structural analysis. The C-terminal linker has three alanine residues, serving as a fixed arm to connect target protein. Previously, Sri Rao Boddeda in our lab showed that both the wild-type and H69D precursor was autoprocessing deficient in the context of C2-MBP, suggesting that C2-MBP fusion disrupted folding/conformation of the downstream p6*-PR in *E. coli*. This results was confirmed to ensure reproducibility (Fig 3.1D, lanes 1-3). We also tested whether different C-terminal tags influence their autoprocessing activity. As shown in Fig 3.1D, lane 4, and Table 1, the wildtype C2-MBP fusion precursor carrying either HA or strep at the C-terminus is equally deficient at autoprocessing, confirming that the p6*-PR miniprecursor is not folded into proteolysis conformation when placed downstream of C2-MBP. Interestingly, when the TEV cut site

(ENLYFQG) was interted into the construct, the resulting precursor (C2-MBP-TEV-Flag-p6*-PR-strep) became autoprocessing competent in *E coli* (Fig. 3.1D, lane 7) and two expected autoprocessing flag-containing products (C2-MBP-TEV-Flag-p6* and C2-MBP-TEV-Flag) were detected. Subsequently, we engineered a few more C2-MBP fusion precursors (C2-MBP-TEV-Flag-p6*-PR-strep) carrying various single mutations. Based on the structure of mature PR, the terminal residues 1-4 and 96-99 form an intertwined four-stranded β-sheet, which contributes the majority of the stabilization energy to the dimerization of mature PR (113). Thus, the PR 96-99 (labeled as PR 1-95) in this construct was deleted to examine the impact on the autoprocessing activity of the C-terminal residues. The PR 1-95 is autoprocessing deficient, similar to D25N, H69D, & R87K (Table 3.1). These four constructs are potential candidates for precursor purification if needed. In summary, the C2-MBP fusion appeared to differentially influence the p6*-PR autoprocessing activity with two-fold implications: the p6*-PR is sensitive to modulations and C2-MBP fusion negatively impacts precursor activity if it is immediately upstream of the p6*-PR sequence.

In the other version of C2-MBP (slow translation codons and C-terminal α-helix), the wild-type or H69D is autoprocessing deficient in the context of C2-MBP(SH)-Flag-M2MG-PR. The M2 is a truncated p6*, in which the N-terminal twenty-two amino acids of p6* are deleted. The MG mutation prevented the cleavage between p6* and PR. Although we removed two cleavage sites, these two fusion precursors still have ~50% activity (Table 3.1), indicating that the precursor is flexible to process other potential cleavage sites although the two cleavage sites were removed. In this context, a C-terminal tag Strep was added into the wild-type construct, and their activities were very similar with or without Strep. Next, we focused on the autoprocessing activity of fusion

precursor carrying two cleavage sites. The wild-type and H69Q are autoprocessing competent (Table 3.1 and Fig 3.1F, lane 3) in the context of C2-MBP(SH)-TEV-Flag-p6*-PR-Strep. The cleavage product PR-strep was detected from H69Q control. In contrast, PR1-95, H69D, R87K were autoprocessing deficient in this context (Fig 3.1F, lanes 4-5). We also tested the autoprocessing activity of fusion precursor carrying one cleavage site between p6* and PR. The M4 is a truncated p6* and keeps the 8 C-terminal residues of p6*. The wild-type and H69Q are autoprocessing competent (Table 3.1 and Fig 3.1F, lane 10) in the context of C2-MBP(SH)-TEV-Flag-M4-PR-Strep, whereas H69D and R87K were autoprocessing deficient in this context (Fig 3.1F, lanes 6-9). Interestingly, compared to their autoprocessing activities (Table 3.1), truncated p6* could reduce the activity, suggesting that p6* may facilitate precursor autoprocessing in *E.coli*.

Purification of DnaK R87K fusion precursor by Ni-NTA

DnaK R87K fusion precursor expressed in BL21 (DE3) was purified via Ni-NTA column and eluted with four different concentrations of imidazole (Eluent 1~4). The full-length DnaK R87K and truncated fusion precursors were mainly found in eluent 1 and 2 (Fig. 3.2A, lanes 7-8). Subsequently, pure full-length DnaK R87K was not obtained from single purification (Ni-NTA), and further purification is required.

TEV protease processed the DnaK R87K fusion precursor at multiple sites

The amount of TEV protease required to cleave the fusion precursor was determined using constant amounts of substrate with titration of TEV protease (Fig 3.2C). Unfortunately, the TEV protease tends to crystallize in solution, which decreases its cleavage efficiency. Neither dialyzing the TEV protease nor adding 1 mM DTT prevented TEV protease from crystallization.

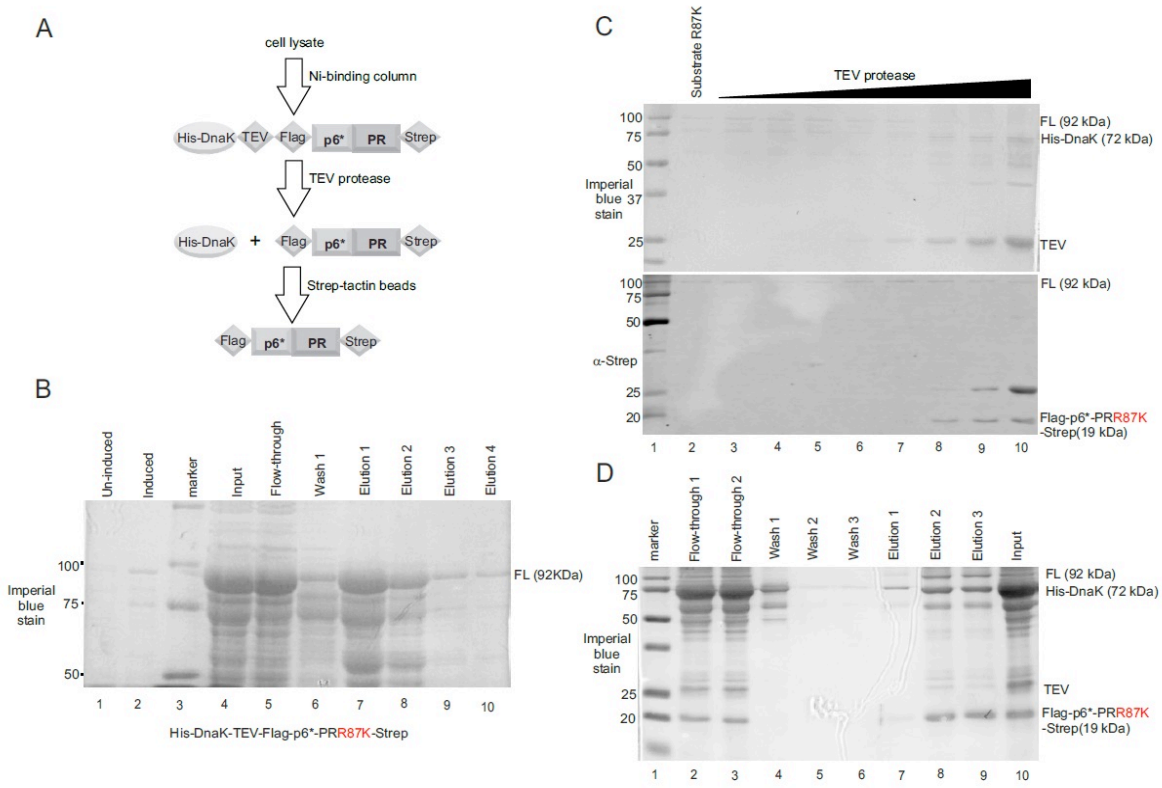

Figure 3.2 The purification for miniprecursor from DnaK fusion precursor using a dual affinity-tags. (A) The schematic flowchart of miniprecursor purification steps. The DnaK fusion precursors were purified with Ni-NTA column followed by a TEV cleavage site, and then miniprecursors were purified by Strep-Tactin beads. (B) DnaK R87K fusion precursor Ni-NTA column purification. After DnaK R87K pass through the column, washed with six column volume of Ni-binding buffer (W1~W3). Then, the Dnak fusion precursor was eluted with two column volume of Ni-binding buffer with 100 (E1), 150 (E2), 200 (E3) and 250 (E4) mM imidazole. The 8% SDS-PAGE was stained with Imperial Protein Stain. (C) DnaK R87K precursor cleaved by TEV. The constant amount R87K substrate (270 ng) was cleaved with increasing amounts of TEV proteases (0.96 ng ~ 2.1 ug). The cleavage reaction samples were resolved by 12.5% SDS-PAGE. The SDS-PAGE (top) was stained with Imperial Protein Stain, and the blot (bottom) was probed with anti-strep. (D) Strep-tactin purification for miniprecursor. The miniprecursor R87K was washed with five column volumes of buffer (W1~W3), and then was eluted with one column volume of elution buffer (E1~E3). The 12.5% SDS-PAGE stained with Imperial Protein Stain.

Nonetheless, the miniprecursor (Flag-P6*-PR-Strep) was detected by strep antibodies at high concentrations of TEV protease (Fig 3.2C, bottom panel, lanes 8-10). We also detected His-DnaK fragment suggesting productive proteolysis at the TEV cleavage site by TEV protease (Fig 3.2C, top panel). Additonally, ~50 kDa and ~40 kDa fragments were also detected at high concentrations of TEV (Fig. 3.2C, top panel, lanes 8-10), indicating that the TEV not only cleaved the TEV cut site but also cleaved at non-specific cut sites.

The Strep-tactin purification of miniprecursor

We attempted to purify the miniprecursor (Flag-p6*-PR-Strep) from the cleaved His-DnaK construct by use of Strep-tactin resin. Because the TEV protease did not cleave all full-length precursors, some full-length precursors were still present (Fig 3.2D). Additionally, although the wash step removed part of His-DnaK (Fig 3.2D, lane 4), miniprecursor and His-DnaK co-eluted in the same elution fractions (lanes 7-9), which suggests that the His-DnaK remains associated with the miniprecursor, likely via non-covalent interactions even after TEV proteolysis.

Purification of C2-MBP fusion precursor by amylose resin

Purification of a few C2-MBP fusion precursors were also examined. The PR1-95, H69D, R87K are autoprocessing deficient in the context of C2-MBP-SH-TEV-Flag-p6*-PR-strep. C2-MBP-SH R87K fusion precursor was expressed and purification was attempted by amylose resin from a large-scale culture. The C2-MBP-SH R87K fusion precursor was eluted using MBP-elution buffer containing 10 mM maltose. The full-length precursor (Fig 3.3B, lanes 5-6) was in both the input and flow-through. However, the Flag-reactive cleavage products were detected as the major band

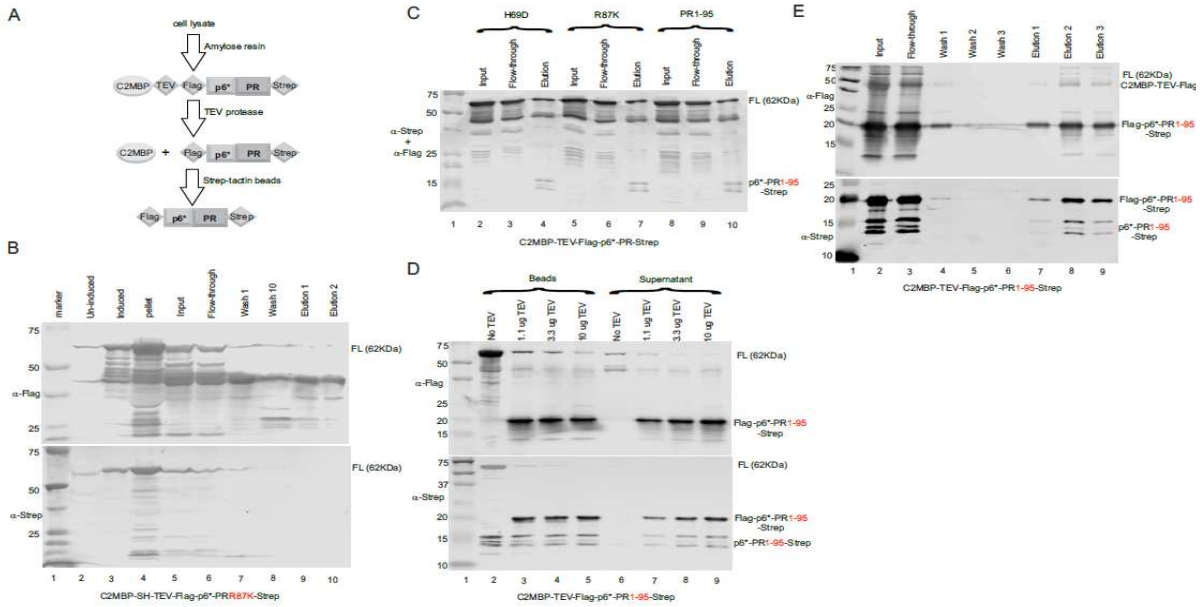

Figure 3.3 The purification for miniprecursor from C2-MBP fusion precursor using a dual affinity-tags. (A) The schematic flowchart of miniprecursor purification steps. The C2-MBP fusion precursors were purified with amylose beads followed by a TEV cleavage site, and then miniprecursors were purified by Strep-Tactin beads. (B) The purification of C2MBP-SH R87K by amylose beads. (C) The purification of C2-MBP H69D, R87K, and PR1-95 by amylose beads. (D) C2-MBP PR1-95 precursor cleaved by TEV on beads. The 20 ul beads associated with C2-MBP PR1-95 precursor (was cleaved with increasing amounts of TEV proteases (1.1, 3.3, and 10 ug). (E) Strep-tactin purification for miniprecursor PR1-95. The samples were resolved by 12.5% SDS-PAGE. The blot were probed with anti-flag and anti-strep.

in eluents (lanes 9-10), indicating that precursor might be processed by itself or cleaved by *E.coli* proteases during purification. Purification was also attempted for C2-MBP fusion precursors carrying H69D, R87K, or PR1-95 from small-scale cultures. However, there were three ~15kDa fragments detected in the eluents (Fig 3.3C, lane 4, 7, & 10). These fragments were analyzed by N-terminal sequencing, and the cleavage sites were mapped to the middle region of p6* peptide. Note that I already included protease inhibitors such as phenylmethylsulfonyl fluoride (PMSF), Pepstatin A or protease inhibitor cocktails into my lysis buffer. Therefore, it is possible that these products were produced by the precursor itself.

The C2-MBP PR^{1-95} fusion precursor on beads cleaved by TEV protease

C2-MBP fusion precursor carrying PR^{1-95} was produced in a large-scale culture (1.5 liter) and bound to amylose beads. TEV protease was used to cleave p6*-PR miniprecursor on the beads to see if degradation could be prevented. Constant amounts of amylose beads were incubated with 1.1, 3.3, and 10 ug TEV protease (Fig 3.3D) before separating beads and supernatant. Maximum change was obtained using a ratio of TEV protease to fusion precursor of 3.3 ug TEV protease to 20 ul beads (200 ug C2-MBP fusion precursor). About 40% of the miniprecursor (Flag-p6*-PR^{1-95}-Strep) was in the supernatant (Fig 3.3D, lane 8), and 60% of the miniprecursor remained associated with the beads (lane 4). However, when fusion precursor associated with amylose beads, three ~15 kDa fragments were produced (Fig 3.3D, lane 2). After TEV cleavage, we still detected three ~15 kDa fragments in the supernatant (lanes 7-9). Thus, TEV cleavage on beads did not avoid precursor degradation. We only obtained 40% miniprecursor in the supernatant. That is, we lost 60% miniprecursor during this treatment. After the on-bead TEV protease cleavage,

miniprecursor (Flag-p6*-PR$^{1-95}$-Strep) was isolated by Strep-tactin resin. Once again, the p6*-PR miniprecursor co-eluted with a ~47 kDa protein (Fig 3.3E, lanes 7-9) and ~15 kDa fragments.

Using centrifugal filter to estimate the size of miniprecursor complexes

Although never fully purifed, we asked whether the miniprecursor was monomeric, dimeric or oligomeric. We used centrifugal filters with different pore sizes to separate complexes. First, we ran the miniprecursor (Flag-p6*-PR^{R87K}-strep) from the DnaK fusion precursor through a series of centrifugal filters (150 and 30 kDa). The monomeric minprecursor is ~19 kDa from SDS-PAGE immunoblots but could not pass through 30 kDa filter (Fig. 3.4A, lane 6-7), suggesting that the miniprecursor doesn't exist as a monomer, but rather as a complex of at least two miniprecursors. However, some precursors couldn't pass through the 150 kDa filter (Fig. 3.4A, lane 4). The TEV protease didn't cleave fusion precursor completely, so the complexes within one or several uncleaved precursors is larger than 150 kDa. Next, we ran the miniprecursor (Flag-p6*-PR$^{1-95}$-strep) from C2-MBP fusion precursor through a series of centrifugal filters (150 and 100 kDa). The result showed that the full-length precursor was co-purified with the miniprecursor and the miniprecursor couldn't pass through 150 kDa filter (Fig 3.4B, lanes 3-4). It suggesred that the 19kD miniprecursor released formed a high-molecular-weight complex that was larger than 150 kDa.

3.4 Discussion and future directions

Purification of DnaK R87K fusion precursor

The DnaK R87K fusion precursor showed multiple Flag-reactive bands shorter than the full-length and absent in autoprocessing competent H69Q (Fig 3.1B, lane 7), which made it difficult to obtain highly purified DnaK R87K fusion precursor. Such heterogeneity could be intrinsic to this mutant precursor due to structural plasticity of p6*-PR. To test/evaluate this possibility, His-DnaK-TEV-Flag-p6*-Strep without PR should be included as a control to perform Ni-NTA column purification side-by-side and to analyze the eluents by western blot with antibodies of His, Flag, and Strep, respectively. This could help map the cleavage products. Recently, Matt O'Malley in our lab was able to obtain highly purified DnaK fusion precursor by Strep-Tactin beads, providing an effective approach to generate enough materials for biophysical analyses.

About the TEV cleavage reaction, Matt recently repeated the experiment and determined that the optimal ratio for cleavage reaction in Ni-binding column is 27 ug TEV protease to 600 ug DnaK fusion precursor. However, he also observed non-specific cleavages within DnaK R87K fusion precursor. We further confirmed the non-specific cut site is located at DnaK protein. Thus, I would suggest that we could use other protease cleavage site to replace the TEV cut site or we may replace DnaK protein with another fusion protein.

We purified the miniprecursor (Flag-p6*-PR-Strep) from the cleaved His-DnaK construct by use of Strep-tactin resin. Because His-DnaK was co-purified with miniprecursor, a construct without TEV cut site should be use. We could try some small fusion proteins (His-small fusion protein-Flag-p6*-PR-Strep), and show that these fusion precursors are purified from a soluble extract of *E.coli*. My strategy would be to use a Ni-affinity column to

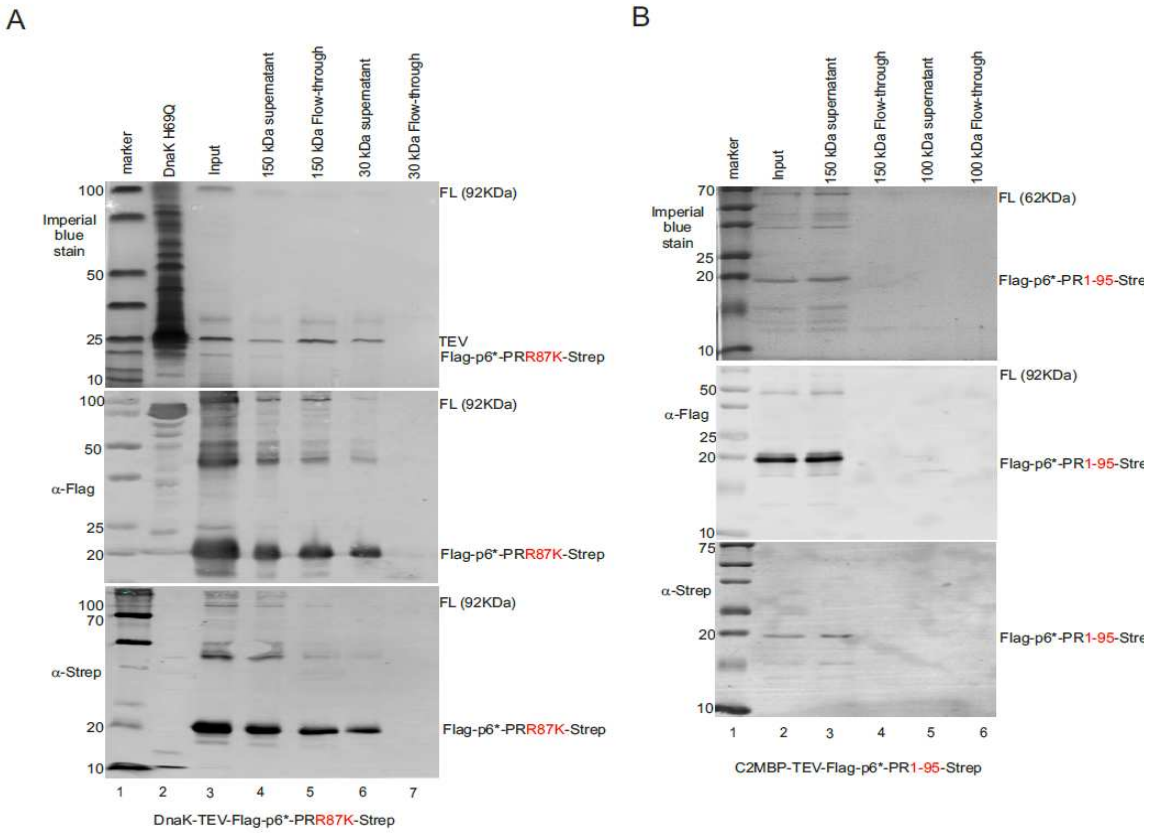

Figure 3.4 The complex of miniprecursor separated by centrifugal filter column. (A) Cleaved miniprecursor from DnaK fusion precursor ran through a series of columns (150 kDa and 30 kDa). (B) Cleaved miniprecursor from C2-MBP fusion precursor ran through a series of columns (150 kDa and 100 kDa). The sample for each fraction was resolved by 13% SDS-PAGE. The SDS-PAGE (top) was stained with Imperial Protein Stain, and the blots (middle & bottom) were transferred to PVDF membrane that probed with anti-Flag and anti-Strep, respectively.

purify the fusion precursor, and then we could purified miniprecursor by Ni-affinity column to get high purity of fusion precursor. We already demonstrated that the precursor autoprocessing is context-dependent. The small tag may influence the conformation of precursor, but we might have a better chance to get highly purity of fusion precursors and analyze them by physical approaches.

Purification of C2-MBP PR^{1-95} fusion precursor

Compared to DnaK fusion precursor, C2-MBP fusion precursor tended to degrade itself during purification although these precursors were autoprocessing deficient. Tris-based buffers were used for purification whereas Sri used PB buffer (2.7 mM KCl, 10 mM Na_2HPO_4, and KH_2PO_4), and no ~15 kDa fragments were observed. The buffer may influence the degradation. Moreover, the C2-MBP-TEV-Flag-p6*-Strep should have been included as a control to perform the purification by amylose beads.

The size of miniprecursor complexes

PR^{1-95} miniprecursor from C2-MBP fusion precursor is a mixture of heterogeneous fragments that do not pass through the 150 kDa cut-off filter. In contrast, those released from DnaK R87K fusion precursors went through 150 kDa cut-off filter. There are three probable means to explain this result: (a) Precursor autoprocessing outcomes are context-dependent (137). Different contexts may produce complexes of different sizes. (b) Mutations R87K or 1-95 may cause altered conformations influencing complex sizes. (c) C2-MBP association with miniprecursor even after TEV cleavage could increase apparent sizes of the complexes. Recently, Matt demonstrated that DnaK full-length precursor is dimeric by sucrose gradient sedimentation. After TEV protease cleavage of DnaK fusion precursor, the DnaK remained associated with miniprecursor illustrated

by immunoprecipitation. As described in Chapter 4, I also examined the apparent size of fusion precursor complexes in the mammalian cell, and demonstrated that precursor dimerization is modulated by various factors. Consequently, I would suggest to make and test ubiquitin-Ubp1 cut site-miniprecursor, and express ubiquitin-fusion precursor and ubiquitin-specific, carboxy-terminal protease (Ubp1) in *E.coli* (170). The Ubp1 could recognize and cleave Ubp1 cut site, thus release miniprecursor, which will facilitate structural characterization of the miniprecursor by X-ray or Cryo-EM.

CHAPTER 4

TRANS PROTEOLYSIS OF HIV-1 FUSION PRECURSORS INDEPENDENT OF A DIMER-INDUCING TAG

4.1 Introduction

Human immunodeficiency virus type 1 (HIV-1) is the causative agent of acquired immunodeficiency syndrome (AIDS), a worldwide epidemic. In the HIV-infected cell, the unspliced genomic RNA also functions as mRNA to direct translation of the Gag and Gag-Pol polyproteins. The Gag-Pol polyprotein has the same N-terminal sequences as Gag polyprotein and is produced as a result of regulated -1 frameshift at the C-terminus of nucleocapsid (NC) domain (75, 171). The *pol* reading frame encodes for protease (PR), reverse transcriptase (RT), and integrase (IN), all of which are essential for productive viral replication. During the late stage of virus replication, Gag and Gag-Pol polyproteins co-assemble into viral particles at the plasma membrane (3, 77, 82, 83, 172). Upon or shortly after virion release from the cell, the Gag-Pol polyprotein precursor undergoes autoproteolysis, leading to liberation of free, mature PR. This process is generally referred to as protease autoprocessing. The mature PR further process Gag and Gag-Pol polyproteins into individual structural and functional proteins, which cause drastic rearrangement of individual proteins within the virions, leading to the production of mature infectious viruses. Therefore, mature PR is absolutely required for the conversion of the immature (non-infectious) virus into the mature (infectious) virus.

Within the Gag-Pol polyprotein, the PR is flanked by peptide p6* at the N-terminus and reverse transcriptase (RT) at the C-terminus. When the Gag-Pol precursor undergoes autoprocessing, two cleavage reactions are essential for liberation of the mature PR: the N-terminal cleavage site (between p6* and PR) and C-terminal cleavage site (between PR and RT). When the C-terminal cleavage site is blocked by mutation, the Gag-Pol precursor autoprocesses into a PR-RT fusion enzyme that can still support productive HIV-1 replication, suggesting that the PR-RT fusion has the essential proteolytic activities to effectively process Gag and Gag-Pol polyproteins (129). In contrast, blocking the N-terminal cleavage site leads to detection of the p6*-PR fragment in the released virion. This fragment can process a few, but not all, cleavage sites within the Gag and Gag-Pol polyproteins (91, 130). Subsequently, the released virions are non-infectious. Thus, the N-terminal cleavage reaction is considered as a critical step for releasing free, fully active mature PR and the p6*-PR is defined as a miniprecursor (96, 107, 134, 136). We have established a cell-based model system to study the autoprocessing mechanism using fusion precursors with the p6*-PR miniprecursor sandwiched between various proteins or small peptide epitopes. Autoprocessing of these fusion precursors in the transfected cells leads to the liberation of various products that can be detected by SDS-PAGE and western blotting analyses. This assay faithfully recapitulated all the phenotypes we tested thus far, which were observed with the proviral constructs, providing a useful tool to gain insights into HIV-1 protease autoprocessing mechanism (127, 133-137, 142, 143).

It is well accepted that HIV-1 PR autoprocessing is temporospatially regulated – suppressed during protein synthesis and virion assembly, and then activated upon or shortly after virion release although the underlying mechanism remains largely undefined. The current autoprocessing theory

is solely built upon the knowledge obtained from the mature PR assuming that the precursor and mature PR are structurally and enzymatically identical. Accordingly, precursor dimerization is believed to be a prerequisite for Gag-Pol autoproteolysis because the catalytic site of the mature PR is formed at the dimer interface (3, 139). Meanwhile, the p6* peptide is known to function preventing precursor dimerization (138, 148). Therefore, the Gag-Pol precursor is hypothesized to start as monomers deficient at autoprocessing; virion assembly increases its local concentration through dimerization and/or oligomerization of protein domains flanking the p6*-PR miniprecursor, which overcomes the inhibitory effect of p6* and induces precursor dimerization and subsequent precursor autoproteolysis. Therefore, it is a fundamental question whether and how the p6*-PR miniprecursor functions as an active enzyme. With our established model system, we engineered substrate and enzyme precursors carrying distinct fusion tags and/or small peptide epitopes, which allowed us to examine factors/determinants required for *trans* proteolysis in transfected cells co-expressing the substrate and enzyme precursors at various ratios and/or treated with indinavir (IDV). In particular, we tested precursors carrying a flanking tag known to form stable dimers (*i.e.*, GST with low nanomolar concentration dissociation constant (173)), or to exist mostly as monomers (MBP), no flanking tag at all. We also constructed and tested various fusion PR precursors with two PRs tethered in tandem with a tripeptide (GGS) to gain further insight into autoprocessing mechanism. Sucrose gradient sedimentation and immunoprecipitation analyses of cell lysates also revealed dimerization status of the substrate and/or enzyme precursors under various conditions. Taken together, our results shed new light into HIV-1 autoprocessing mechanism.

4.2 Materials and methods

DNA mutagenesis

The mammalian expression vector encoding for GST-, L-MBP-, C2-MBP-, Flag-, SigL-, and Myc-tagged fusion precursors were described in our previous studies (128, 137, 142, 143). Additional plasmids used for this study were constructed by standard molecular mutagenesis. Each construct was confirmed by sequencing analysis.

Cell culture and transfection

HEK 293T cells were grown in DMEM medium with 10% fetal bovine serum (FBS), 100 units/ml of penicillin G sodium and, 100 µg/ml of streptomycin sulfate at 37℃. The procedure for calcium phosphate transfection has been described previously (133-135, 137, 142, 143). In brief, the cells were treated with chloroquine to a final concentration of 25 µM. The HEK293T cells (30-40% cell confluence) in 12-well plate were transfected with the 150µl DNA-calcium mixture containing a total of 0.5µg pEBG plasmid DNAs in 65.7µl H_2O, 9.3µl of 2M $CaCl_2$, and 75µl of 2 x HBS (50mM Hepes, 10mM KCl, 12mM Dextrose, 280mM NaCl, and 1.5mM Na_2HPO_4, pH7.04~7.05) per well at 37℃ for 11 hours. The culture medium was replaced with pre-warmed fresh DMEM without chloroquine after 10-11 hours incubation. At ~30 h post transfection, the cells were lysed *in situ* with 80µl lysis buffer (Tris-HCl, pH8.0, 150mM NaCl, 1% sodium deoxycholate, 1% Triton X-100, 1x protease inhibitor cocktail). The resulting supernatant was spun briefly (10,000 x g for 2 min) to remove host chromosomes and cell debris. Each sample was boiled in 1xSDS loading buffer for 5 min and chilled in an ice-water bath prior to SDS-PAGE and western blotting analysis or stored at -20℃ before use.

SDS-PAGE and western blotting

The samples were resolved through 10% or 12.5% SDS-PAGE and subsequently transferred onto PVDF membrane (Immobilon-P). The membranes were probed with various primary antibodies including mouse anti-Flag (Sigma, cat# F1804), mouse anti-HA (Sigma, cat# H9658), mouse anti-Myc (supernatant of Myc 1-9E10.2 hybridoma cells), β-actin (Sigma, cat# A5441), mouse anti-GAPDH (Millipore, cat# MAB374) or rabbit anti-V5 (Rockland, cat# 600-401-378), respectively. The membranes were then incubated with IR700 goat anti-rabbit (Rockland, cat# 611-130-122) or IR800 mouse anti-mouse (Rockland, cat# 610-132-121). Signals were scanned by Odyssey infrared dual laser scanning unit (LI-COR Biotechnology, Lincoln, Nebraska) and analyzed through Image Studio.

Sucrose gradient sedimentation analysis

The procedure was modified from a previous report (174). In brief, the HEK 293T cells in each 10cm petri dish were transfected with 6μg DNA in 900 μl DNA-calcium mixture. At ~30h post transfection, the cells were rinsed with 1x PBS once, collected by gentle pipetting, and pelleted by centrifugation at 500 x *g* for 5 mins. The cell pellet from 2 petri dishes was lysed in 200 μl lysis buffer (0.2% Triton X-100, 0.1% saponin, 0.1% digitonin, 2mM DTT, 1x protease inhibitor cocktail in 1 x PBS) on ice for 10 mins. For precursors expressed at low levels at steady state low levels, the cell pellet from 4 petri dishes were used and lysed in 200 μl lysis buffer (0.4% Triton X-100, 0.1% saponin, 0.2% digitonin, 2mM DTT, 1x protease inhibitor cocktail in 1 x PBS) on ice for 10 mins. After removal of cell debris with a brief centrifuged at 5000 x *g* for 2 mins, the clarified supernatant was loaded onto the top of a 5-50% sucrose gradient made in 1 x PBS. The gradient was spun at 38,000 rpm for 16 h at 4°C in a SW Ti41 rotor of a Beckman L8-70M

Ultracentrifuge. Forty fractions (~120 µL ea) were collected from top to bottom using a Biocomp fractionator and subjected to SDS-PAGE and western blotting.

Immunoprecipitation

HEK 293T cells in each 10 cm petri dish were transfected as described previously. At ~30h post transfection, the cells were collected and centrifuged at 500 xg for 5 mins. The cell pellet was lysed in 1x PBS containing 0.3% Triton X-100, 2 mM DTT, and 1x protease inhibitor cocktail on ice for 10 mins followed by a centrifugation at 5000 xg for 2 mins to remove cell debris. The supernatant containing GST-fused enzyme precursor was incubated with 3 µl anti-Flag M2 antibody (Sigma, cat# F1804) per petri dish at 4℃ on a shaker for 2h. Then, 20 µl protein A/G beads slurry (Thermo cat# 20421) were added to the mixture followed by an incubation at 4℃ on a shaker for 2h. The precursor-antibody-protein A/G beads were washed five times with 1x PBS. For C2-MBP-fused enzyme precursor, the supernatant was incubated with 10 µl anti-HA agarose antibody (Sigma, cat# A2095) per petri dish at 4℃ on a shaker for 4h. The precursor-antibody agarose beads were washed five times with 1x PBS. The final precipitate was boiled in 100 µl 1.5x loading buffer for 5 min and subject to SDS-PAGE and western blotting analyses.

4.3 Results

Positive *trans* proteolysis of GST-fused substrate by GST-fused enzyme

We engineered enzyme and substrate fusion precursors with distinct fusion tags to specifically examine *trans* proteolysis (Fig 4.1A). The enzyme, GST-M2-PSHL-PR-Flag, contains the wild type PR sequence but has PSHL mutation that was previously reported to block the N-terminal

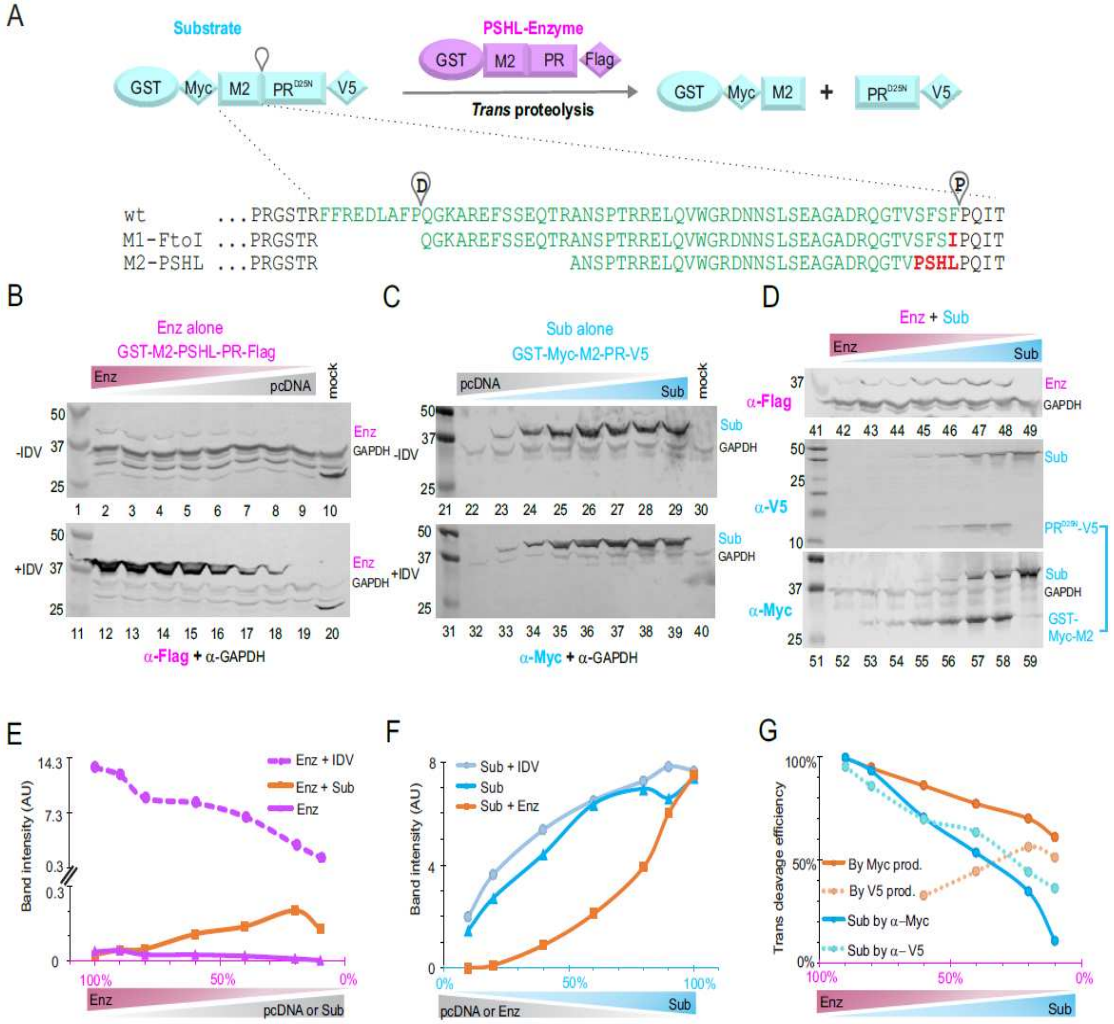

Figure 4.1. *Trans* proteolysis of GST-fused substrate by GST-fused enzyme. (A) Schematic of engineered substrate and enzyme precursors with the p6* amino acids in green and the mutations in red. The distal (D) and proximal (P) cleavage sites were denoted with the drop shapes. (**B & C**) Western blot detection of the enzyme or substrate when expressed alone treated without or with 5 μM indinavir (IDV), a known protease inhibitor. (**D**) Western blot detection of the enzyme and substrate along with *trans* cleavage products at various Enz:Sub input DNA ratios. The image shown is representative of five biological replicates. (**E & F**) Quantification of the enzyme and substrate by band intensity normalized to GAPDH. (**G**) Quantification of *trans* proteolysis efficiencies. The blue lines were plotted using equation (T-U)/T, where U was unprocessed substrate detected by Myc (dark solid blue) or V5 (light dashed blue), T was the total expected substrate computed by sub DNA input ratio in each lane times the substrate intensity at 100% (lane 59) . The orange lines were determined using equation P/(P+U), where P was the proteolysis product detected by Myc (solid line) or V5 (dashed line).

processing and thus results in production of p6*-PR fragment in the released viral particles (91). The M2 is a truncated p6*, in which the N-terminal 22 amino acids of p6* are deleted to remove the distal processing site (Fig 4.1A). Consequently, this GST fused enzyme is incapable of processing itself (autoprocessing deficient) but it could cleave other substrates *in trans*. When expressed alone in transfected cells, this enzyme showed low levels of detection likely due to its self-degradation (Fig 4.1B, upper panel), a well-known property of recombinant HIV protease. When transfected cells were treated with 5 μM indinavir, a known protease inhibitor, enzyme detection was significantly increased by >50-fold and positively correlated with plasmid inputs (Fig 4.1B, lower panel). We also constructed GST-Myc-M2-PR^{D25N}-V5 as a substrate that is catalytically deficient as the active site residue was mutated (D25N) but carries the wild type N-terminal cleavage site that could be *trans* processed by an enzyme. When expressed alone, the substrate showed stable detection in transfected cells with or without IDV treatment, positively correlating with plasmid inputs (Fig 4.1C &4.1F).

We next transfected HEK293T cells with constant amounts of the total input DNA while varying the Enz:Sub ratios to test whether and how the *trans* proteolysis would occur in co-transfected cells. As shown in Fig 4.1D, substrate cleavage by the enzyme was demonstrated by detection of GST-Myc-M2 (bottom panel) and PR^{D25N}-V5 (middle panel) proteolysis products. We quantified the *trans* proteolysis efficiencies by several approaches to ensure analyses consensus (Fig 4.1G). For example, the full-length substrate left unprocessed (U) in transfected cells was determined by Myc or V5 antibody, respectively. The total substrate (T) was calculated according to DNA input ratio relative to 100% substrate because substrate expression is linearly correlated with DNA input (Fig 4.1F). The cleaved substrate was computed by T-U. This approach reported (T-U)/T values

to reflect *trans* proteolysis efficiencies (Fig 4.1G, blue lines), showing a general decline as the Enz:Sub ratio decreased. Another calculation was to determine the amount of cleaved product (P) as a percentage of cleaved (P) plus unprocessed substrate (U). Using Myc-reactive bands (Fig 4.1G, the solid orange line), this calculation was in general agreement with the those calculated using the full-length substrate. However, the *trans* cleavage efficiencies calculated using V5-reactive bands (Fig 4.1G, orange dashed line) were not in the same trend. We speculated this seeming discrepancy could be due to PR-V5 degradation after its liberation from the substrate precursor and/or low detection sensitivity of PR-V5 product. Nevertheless, our data demonstrated productive *trans* cleavage between the GST-fused enzyme and substrate in transfected cells with higher Enz:Sub ratios correlating with higher *trans* proteolysis efficiencies.

GST-fused precursor complexes detected during *trans* proteolysis

An intriguing observation of GST-fused precursor *trans* proteolysis was the increased detection of GST-fused enzyme, compared to enzyme alone (Fig 4.1E orange *vs* purple lines), when the Enz:Sub ratio decreased. We interpreted it to suggest formation of some enzyme/substrate complexes that were proteolysis deficient stabilizing the enzyme by preventing it from self-degradation. To test this possibility, we analyzed post-nucleus cell lysates through sucrose gradient velocity sedimentation. A trace amount of eGFP-V5 encoding plasmid (5% total input) was included for transfection as eGFP is well known to exist as monomers serving as a size reference (Fig 4.2A, lanes 5-6; Fig 4.2D, dark green line). The endogenous GAPDH was also detected as it exists as tetramers (~144 kDa in size, light green line) under non-denaturing conditions, serving as another size reference (Fig 4.2A-C). In the presence 5μM indinavir, GST-fused enzyme (~43 kD each) was mostly dimers (Fig 4.2D, purple line). GST-fused substrate also appeared to be

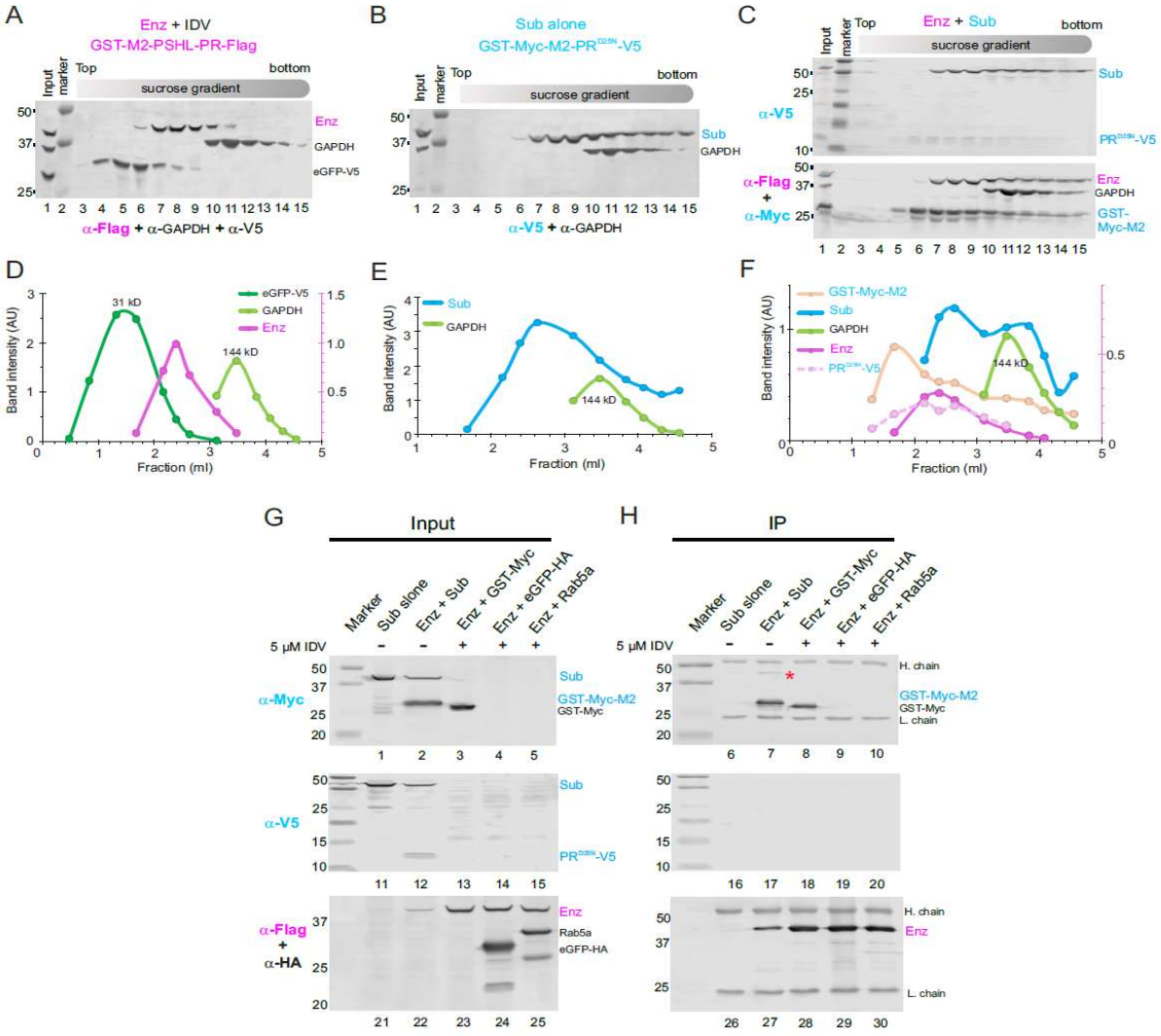

Figure 4.2. Detection of GST-fused precursor complexes by sucrose gradient sedimentation and immunoprecipitation. (**A-C**) Sucrose gradient sedimentation analysis of the indicated cell lysates. At ~30h post transfection, the transfected HEK 293T cells were lysed in the lysis buffer, clarified of cell debris, and loaded onto the top of 5-50% (g/ml) sucrose gradient made in 1xPBS. After ultracentrifugation, 40 fractions (~120µl ea) were collected from top to bottom with a BioComp gradient fractionator and analyzed by western blotting. The image shown is representative of three experiments. (**D-F**) Distribution profiles of the detected proteins. Band intensity was quantified and plotted as a function of fraction volume from the top. (**G-H**) Immunoprecipitation data. The cell lysates of HEK 293T cells transfected with the indicated constructs were incubated with anti-Flag antibody to pull down Flag-tagged enzyme and proteins associated with it. About 3% each lysate (G) was loaded as input control and 30% of each immunoprecipitated sample (H) was examined in parallel. H. chain and L. chain denote antibody heavy and light chain, respectively. Asterisk (*) indicates the full-length substrate. The image shown is representative of two experiments.

mostly dimers (Fig 4.2E, blue line) with an extended tail towards right, suggesting existence of high MW complexes. Because of GST fusion (a dimer-inducing tag), it is not possible to say if the observed dimers were induced by GST alone or in combination with the precursors.

When the enzyme and substrate precursors were co-transfected at 1:4 ratio, productive *trans* processing of substrate was illustrated by detection of GST-Myc-M2 and PR-V5 products (Fig 4.2C). Note that these products along with the substrate and enzyme precursors were detected throughout a wide range of the gradient, suggesting existence of heterogeneous complexes with varying sedimentation velocities. For example, the GST-Myc-M2 product displayed a peak corresponding to expected dimers with a long-extended tail (Fig 4.2F, light orange), hinting its association with high MW complexes even after substrate processing by enzyme *in trans*. The enzyme precursor (solid purple) appeared to be mostly dimers but was also detected in heavier fractions. In addition to the expected dimer peak, the substrate precursor showed more detection in fractions corresponding to complexes >144 kD in size (Fig 4.2F, blue line). The PR-V5 detection was very low overall but nonetheless it showed neither monomers (11kD) nor dimers (22kD). Instead, it spread out across the gradient (pink dashed line), again indicating its association with other proteins Taken together, sucrose gradient sedimentation analysis detected high MW complexes that might contain enzyme, substrate, and proteolysis products in various combinations.

We then performed immunoprecipitation (IP) to investigate if the enzyme was associated with the substrate and/or proteolysis product in high MW complexes. Flag antibody was used to pull down the enzyme along with proteins that were stably associated with it. When the Enz:Sub ratio was at 1:4 without any PI, productive *trans* proteolysis was confirmed by detection of GST-Myc-M2 and

PR-V5 (Fig 4.2G, lanes 2 and 12). The GST-fused enzyme was detectable but at low levels compared to the controls co-expressing the enzyme with other proteins in the presence of 5 μM indinavir (Fig 4.2G, bottom panel). Interestingly, the full-length substrate was co-immunoprecipitated with the enzyme (indicated by a red asterisk in Fig 4.2H top panel), supporting the idea of complex formation between enzyme and substrate. Note that the enzyme also pulled down GST-Myc-M2 product, suggesting that GST-Myc-M2 remained associated with the enzyme after *trans* processing or the released GST-Myc-M2 and the enzyme formed stable dimers mediated by GST tag. Both are not mutually exclusive and the later one was indeed confirmed in cells co-expressing the enzyme and GST-myc such that GST-Myc was pulled down along with GST-M2-PSHL-PR-Flag (Fig 4.2H, lane 8). Meanwhile, neither eGFP-HA nor Rab5a-HA was pulled down by the enzyme (Fig 4.2H, lanes 29 &30), confirming a reasonable degree of immunoprecipitation specificity. Collectively, these results revealed stable association of the enzyme with the full-length substrate and processing product (*e.g.* GST-Myc-M2), which might in part contribute to formation of high MW complexes detected by sucrose gradient sedimentation.

A dimer-inducing tag is not required for *trans* proteolysis

We first chose GST as a fusion tag to facilitate precursor dimerization as it forms a stable homodimer with a low nanomolar dissociation constant (173). This choice was based on the autoprocessing theory suggesting precursor dimerization as an obligatory prerequisite. With the *trans* proteolysis platform, we next sought to examine whether GST-mediated dimerization is required for productive *trans* processing. For this purpose, we constructed and examined precursors carrying maltose binding protein lacking its N-terminal signal peptide (C2-MBP), a known monomeric protein, in place of GST. Interestingly, this pair demonstrated *trans* proteolysis

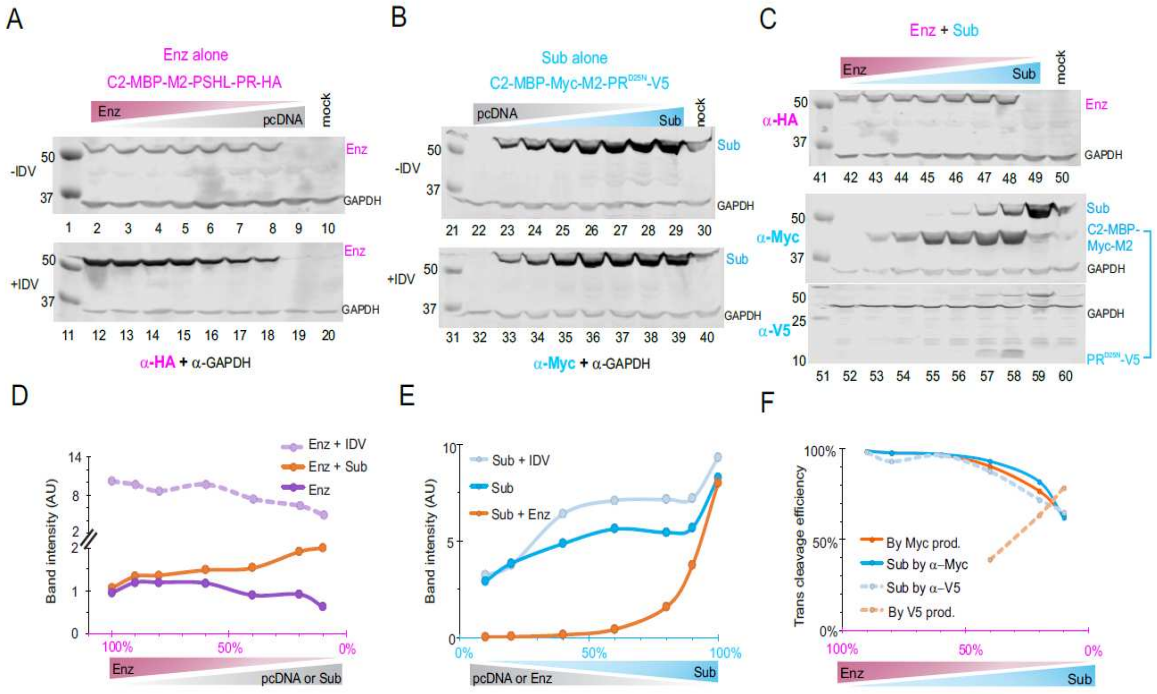

Figure 4.3. *Trans* proteolysis between fusion precursors carrying monomeric C2-MBP tag. (**A-C**) Western blot detection of the enzyme or substrate when expressed alone with or without 5μM IDV treatment or co-expressed at various Enz:Sub ratios. The image shown is representative of two experiments. (**D & E**) Quantification of enzyme and substrate. Band intensity normalized to GAPDH in each lane was plotted as a function of DNA input. (**F**) The *trans* processing efficiencies quantified either by P/(P+U) (orange lines) or by (T-U)/T (blue lines) both using Myc-reactive or V5-reactive bands.

profiles (Fig 4.3) very similar to the GST-fused precursor pair. When expressed alone, the C2-MBP-fused enzyme showed low levels of detection in the absence of any PI treatment and high levels of detection in the presence of 5μM indinavir (Fig 4.3A&D). The C2-MBP substrate expression was positively correlated with plasmid DNA input in the transfected cells with or without PI treatment (Fig 4.3B&E). Also, high *trans* processing efficiencies were correlated with high Enz:Sub ratios (Fig 4.3C&F) with increased enzyme detection when the Enz:Sub ratio decreased (Fig 4.3C&D). Consequently, our data demonstrated that a dimer-inducing tag is not required for positive *trans* proteolysis.

We next carried out sucrose gradient sedimentation analysis to examine whether and what complexes formed in transfected cells expressing the C2-MBP-fused precursors under various conditions. Endogenous proteins such as β-actin and GAPDH were detected in parallel serving as size references. Purified ferritin complexes (~440 kDa in size made of 24 copies of ~20 kD ferritin protein subunit) were spiked into some lysates as a size reference. The C2-MBP-fused substrate showed dimeric complexes as the major form along with oligomeric complexes shown as an extended tail (Fig 4.4A-B); no obvious monomers were detected. Meanwhile, C2- MBP-flag protein was predominantly detected as monomers (Fig 4.5A). These data suggest that precursor dimerization is mediated by the Myc-M2-PR^{D25N}-V5 portion, not the fusion tag. In the absence of any PI, C2MBP-fused enzyme was mostly detected as monomers indicated by a star in Fig 4.4C (pink line). When treated with 5 μM indinavir, the enzyme shifted to dimeric forms,suggesting an induced precursor dimerization by PI treated (dark purple line, triangle). When the enzyme and substrate were mixed at a 1:4 ratio, the enzyme displayed monomers as the predominant form but also showed a small bump overlapping with dimeric complexes (Fig 4.4C, light purple line).

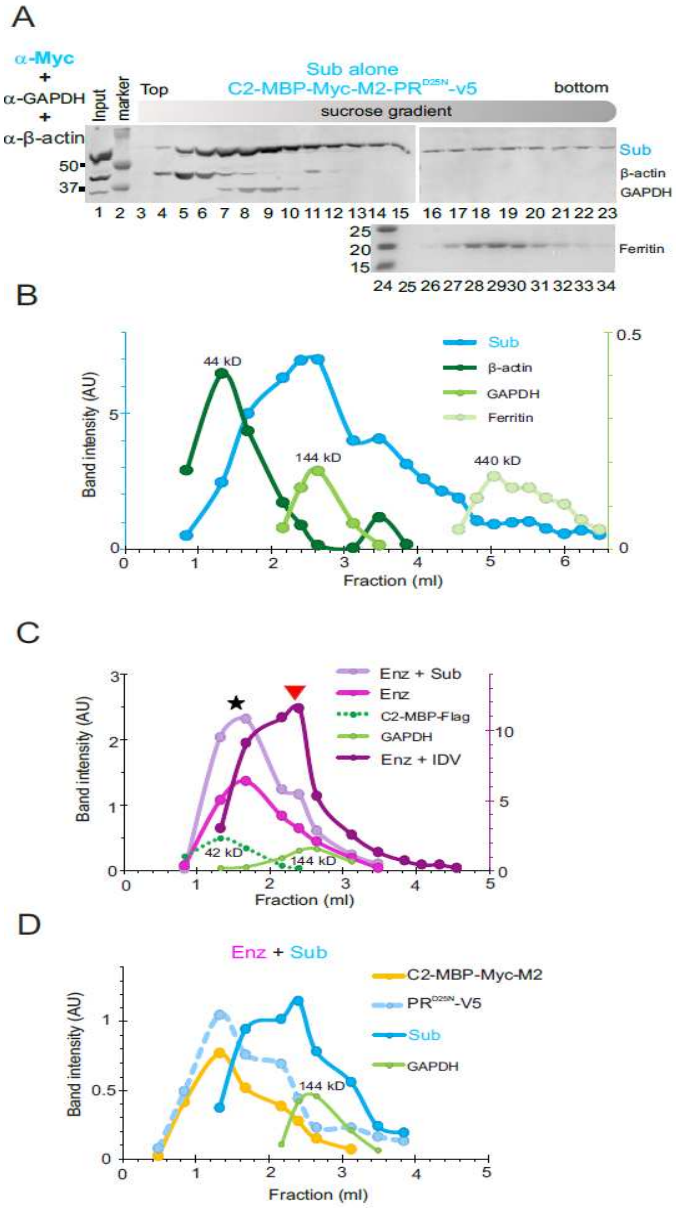

Figure 4.4. Detection of C2-MBP precursor complexes. (A) Distribution of C2-MBP-fused substrate in a 5-50% sucrose gradient. The image shown is representative of three experiments. The substrate and reference proteins were detected by western blotting. **(B)** Band intensity of the substrate and reference proteins quantified and plotted as a function of fraction volume. **(C)** Detection of the C2-MBP-fused enzyme under the indicated conditions (Enz alone, Enz+5μM indinavir, Enz+Sub (at a 1:4 DNA input ratio)). The asterisk and triangle denote monomeric and dimeric complexes, respectively. **(D)** Detection profiles of the C2-MBP-fused substrate along with its proteolysis products in lysates made from HEK 293T cells transfected with a mixture of the enzyme and substrate constructs (at a 1:4 DNA input ratio).

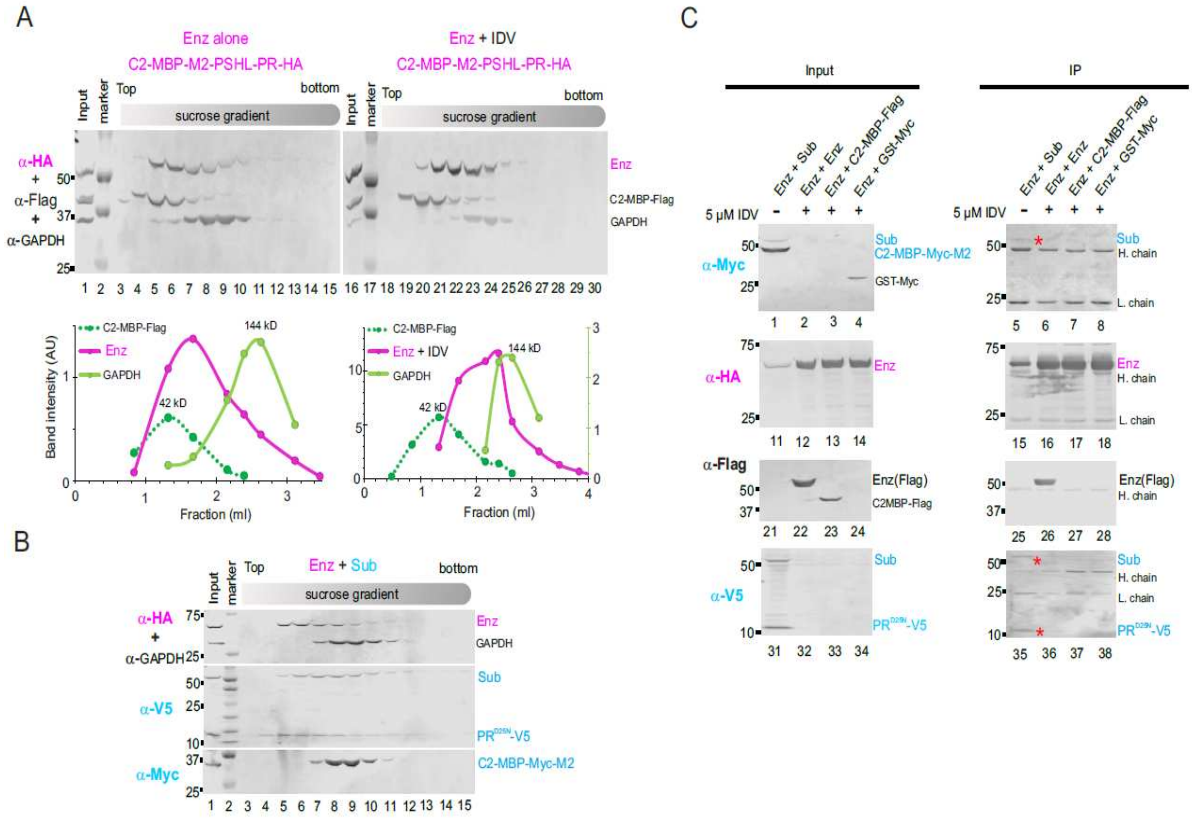

Figure 4.5. Detection of C2-MBP-fused precursor complexes. (**A-B**) Distribution profiles of the C2-MBP enzyme in three conditions analyzed sucrose gradient sedimentation followed by Western blotting. The image shown is representative of three experiments. (**C**) Immunoprecipitation analysis of cell lysates expressing C2MBP-fused enzyme and substrate at a 1:4 Enz:Sub NDA input ratio. Anti-HA antibody conjugated agarose beads were used to pull down HA-tagged enzyme and its associated proteins. About 3% each lysate was loaded as input control (left) and 30% of each immunoprecipitated sample (right) was examined in parallel. H. chain and L. chain denote heavy and light chain, respectively. Red asterisks highlight the co-immunoprecipitated proteins. The image shown is representative of two experiments.

4.4D&4.5B). The C2-MBP-Myc-M2 product exhibited monomers as the major form with some

Meanwhile, the substrate was *trans* processed into C2-MBP-Myc-M2 and PR^{D25N}-V5 (Fig also co-existing in high MW complexes (yellow line). The PR^{D25N}-V5 product is not monomers or dimers as expected. Instead, it seems to be associated with various complexes (dash blue line). Collectively, our data demonstrated effective *trans* proteolysis between C2-MBP-fused precursors, and that the C2-MBP-fused enzyme is mostly monomeric whereas the C2-MBP-fused substrate is mostly dimeric.

We also performed immunoprecipitation analyses to examine whether and what C2-MBP precursor complexes are in the co-transfected cells. The C2-MBP-M2- PSHL-PR-HA was used as bait for IP with HA agarose antibody beads. In the presence 5 μM indinavir, the HA-tagged enzyme pulled down C2-MBP-M2-PSHL-PR-Flag enzyme (Fig 4.5C, lane 16), suggesting dimer formation between the enzymes, which is consistent with sucrose gradient sedimentation results (Figure 4.4C). Meanwhile, the co-expressed C2-MBP-Flag or GST-Myc were not pulled down (Fig 4.5C, lanes 27 & 8), validating IP specificity. From the mixture of C2MBP-fused enzyme and substrate (at 1:4 ratio) in the absence of any PI, the enzyme pulled down both the substrate (Fig 4.5C, lane 5 & 35) and PR^{D25N}-V5 (lane 35), but not C2-MBP-Myc-M2 product (lane 5), suggesting existence of stable complexes containing the enzyme plus substrate and/or PR^{D25N}-V5. This is consistent with the sucrose gradient sedimentation result showing detection of complexes containing all these proteins. Taken together, our data illustrates the existence of various complexes that may be *trans* proteolysis intermediates caught in action at low Enz:Sub ratios.

To further evaluate the requirement of a dimer-inducing fusion tag for *trans* proteolysis, we engineered and analyzed another pair, Flag-M1-FtoI-PR-HA enzyme and Myc-M2-PR^{D25N}-V5

substrate, both lacking any known dimer-inducing fusion tag (Fig 4.6). The FtoI mutation is known to prevent the N-terminal processing from releasing mature PR (142); the M1 is a truncated p6* lacking the first nine N-terminal amino acids (Fig 4.1A). Interestingly, they once again exhibited *trans* proteolysis as effectively as GST- or C2-MBP-fused enzyme and substrate. For example, Flag-M1 enzyme showed low levels of detection in the absence of any PI treatment and high levels of detection in the presence of 5μM indinavir (Fig 4.6B&E). Meanwhile, Myc-M2- substrate was stable in the transfected cells with or without PI treatment (Fig 4.6C&F). High *trans* processing efficiencies correlated with high Enz:Sub ratios (Fig 4.6D&G) as quantified using the full-length substrate (Fig 4.6G, solid line). We were unable to quantify *trans* processing efficiency using Myc-reactive bands as with GST-fused precursor because the Myc-M2 product is too small (~5 kDa) for the standard SDS-PAGE. Quantification of the V5-reactive bands showed lower *trans* processing efficiencies (Fig 4.6G, dashed line), which could be an underestimate as shown in Figure 4.1G. Nevertheless, our results demonstrate positive *trans* processing between this enzyme and substrate pair, suggesting that GST-mediated dimerization is not required for positive *trans* processing.

Inter-molecular dimerization of precursors carrying tethered PRs

Our detection of C2-MBP substrate dimers as the predominant form suggested that the p6*-PR miniprecursor by itself could form inter-molecular dimers in the absence of any dimer-inducing flanking tag. Also, D25N mutation, known to abolish PR catalysis, seemed not to impair precursor dimerization. To further evaluate this property, we engineered and tested precursors carrying two PRs tethered in tandem with a short linker (Fig 4.7A). Note that the "context" of these two PRs were different as the 1st PR had the upstream p6* but the 2nd PR did not. Two versions, one with

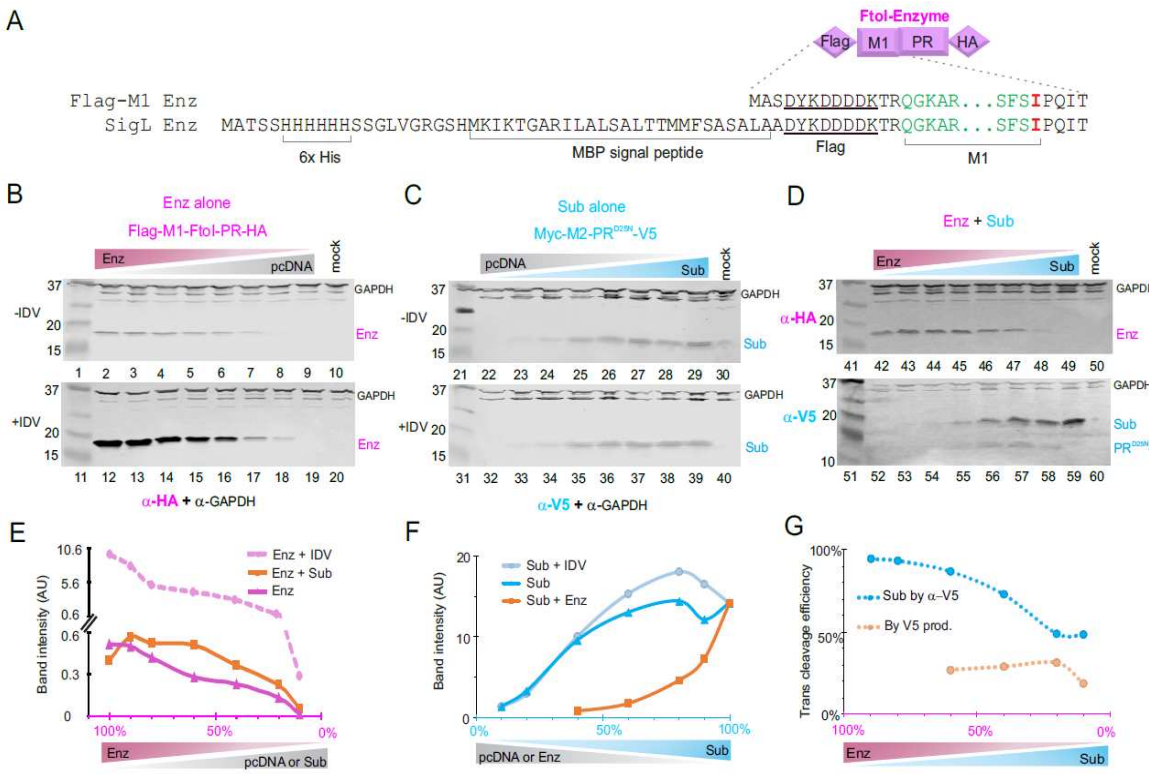

Figure 4.6. ***Trans*** **proteolysis of Flag-M1 precursors. (A**) Schematic of the enzymes lacking any known dimer-inducing tag. (**B & C**) Western blot detection of the indicated enzyme or substrate when expressed alone with or without 5μM IDV treatment. **(D)** Western blot examination of lysates co-expressing the enzyme and substrate at various ratios. The images shown are representative of four experiments. (**E-G**) Quantification of enzyme and substrate under various conditions. Band intensity normalized to GAPDH was determined to reflect the amount of enzyme or substrate. The *trans* processing efficiencies were quantified by either (T-U)/T (dashed blue) or P/(P+U) (dashed orange) using V5-reactive bands.

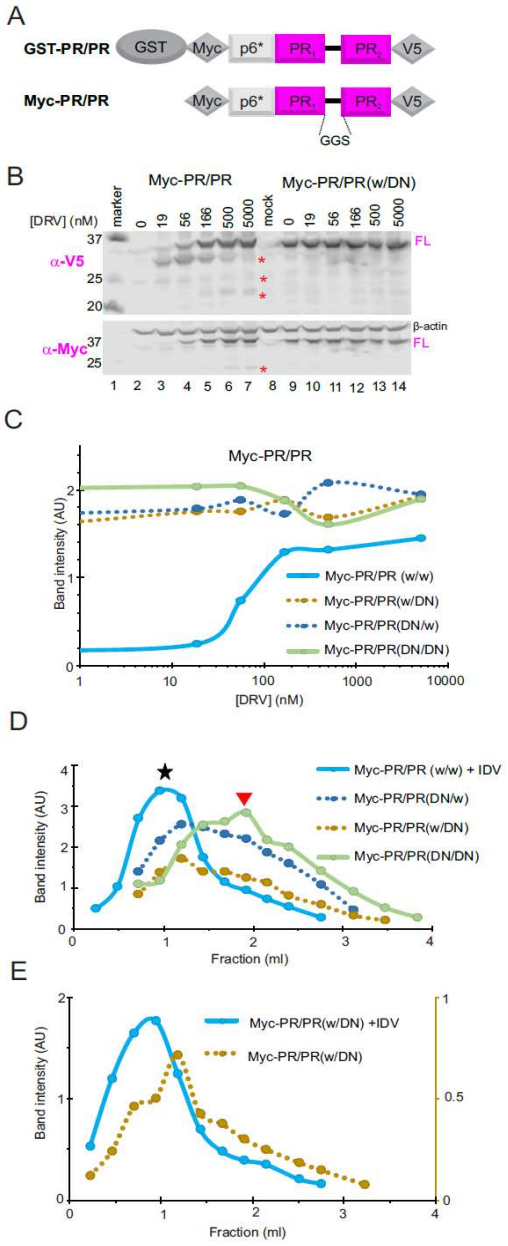

Figure 4.7. Autoprocessing and complexes detection of precursors carrying tethered PRs. (**A**) Schematic diagram of tethered PR precursors used in this study. (**B & C**) Autoprocessing of various tethered PR precursors in response to darunavir (DRV) analyzed by western blotting and followed by quantification of the full-length precursor detected with rabbit anti-V5 and normalized to β-actin signal. (**D & E**) Distributions of various tethered PR precursors in 5-50% sucrose gradient under the indicated conditions. The asterisk (*) and triangle (▼) denote intra-molecular and inter-molecular dimeric complexes, respectively.

GST to facilitate inter-molecular dimerization and one without, were tested to compare effects of GST-induced dimerization (Table 4.1). The Myc-wt/wt precursor (without GST) was autoprocessing competent as indicated by the disappearance of the full-length precursor in the absence of any PI (Fig 4.7B, lane 2). The released autoprocessing products seemed to undergo self-degradation as they were not detectable in the absence of any PI and became detectable when treated with darunavir. Also, there were multiple autoprocessing products (indicated by the asterisks in Fig 4.7B) detected with V5 and Myc antibodies. In addition to autoproteolysis at the expected distal and proximal sites (Fig 4.1A), cleavages inside either the 1st or 2nd PR were detected as well, which complicated quantification of autoprocessing efficiency. Therefore, we used the full-length Myc-wt/wt precursor to reflec proteolysis efficiency as a function of darunavir concentration (Fig 4.7C, solid blue line). Precursors carrying one deficient PR (D25N) or both were mostly defective, suggesting that D25N dominated the overall activity of Myc-PR/PR tethered precursors.

We next examined these Myc-PR/PR tethered precursors by sucrose gradient sedimentation to see whether the tethered PRs form intra- or inter-molecular dimers. For the wt/wt precursor, we treated the transfected cells with 20μM indinavir. Under this condition, the precursor mainly existed as intra-molecular dimers peaked at a position close to eGFP-HA reference (Fig 4.6D & 4.8). This approach, however, couldn't discern whether the intra-molecular dimers were the active conformation stabilized by indinavir binding or the indinavir treatment by itself induced intra-molecular dimerization. On the other hand, inter-molecular dimers seemed to be the predominant forms in DN/DN (Fig 4.7D, solid green line). DN/wt and wt/DN had a mixture of monomeric and dimeric complexes (Fig 4.7D, dash lines). The Myc-wt/DN precursor shifted to more intra-

Table 4.1. Summary of various dimeric precursors and their autoprocessing activities

	GST-dPR	Myc-dPR
WT/WT	+	+
WT/T26A	+	N/A
WT/D25N	+/-	-
D25N/WT	+/-	-
D25N/D25N	-	-

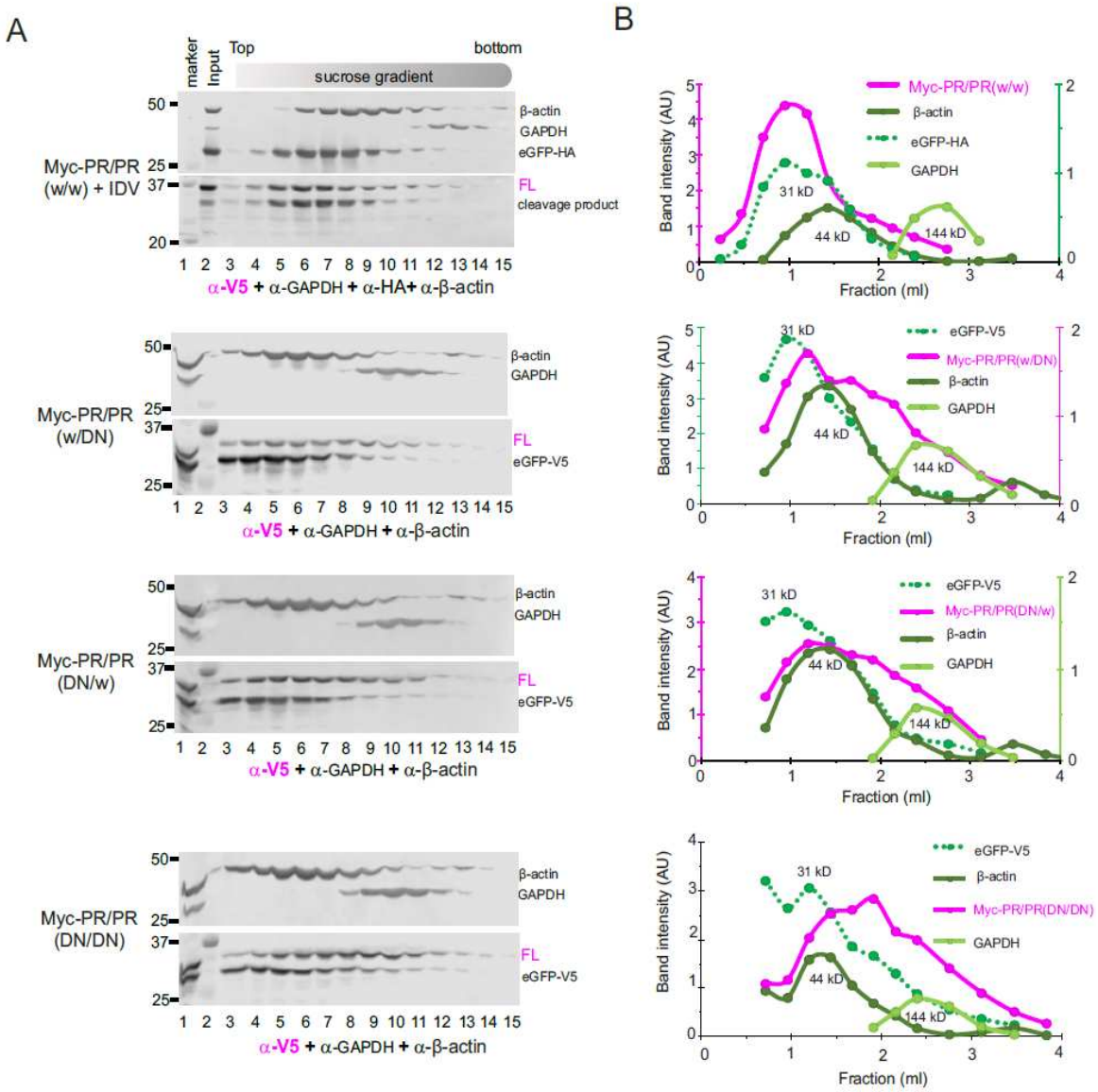

Figure 4.8. Detection of tethered PR precursor complexes. (**A**) Distribution profiles of tethered PR precursors analyzed by sucrose gradient sedimentation followed by Western blotting. The image shown is representative of two experiments. (**B**) Band intensity quantification of each precursor as a function of fraction volume.

molecular dimers when the transfected cells were incubated with 20μM indinavir (Fig 4.7E, blue line). Collectively, sucrose gradient sedimentation analyses revealed that the tested Myc-PR/PR precursors predominantly formed inter-molecular dimers in the absence of any PI, whereas they could also form intra-molecular dimers likely induced by PI treatment.

The autoprocessing data of GST-PR/PR precursors are similar to Myc-PR/PR precursors except that GST-wt/T26A showed wild type autoprocessing activity, whereas GST-wt/DN and GST-DN/wt showed partial activities. We interpreted this phenomenon to suggest that GST induced inter-molecular dimerization such that the two copies of the wild type PR from each precursor formed an active site catalyzing precursor autoprocessing. Only GST-wt/T26A was fully competent at autoprocessing; the wt/DN had partial activities, suggesting an involvement of additional regulations in autoprocessing activity. In this case, it is interesting to note that T26A and D25N in the 2nd position influenced the tethered PR precursors differentially although both were considered similar in abolishing mature PR activity. This result argued for structural and enzymatic plasticity of precursors, which seems to be sensitive to even a single amino acid alteration. Taken together, our analyses of tethered precursors reveal high propensities of inter-molecular dimerization mediated by the p6*-PR miniprecursor and modulations of precursor complexes and activity by flanking sequences such as GST.

The MBP SigP facilitated miniprecursor dimerization

We previously reported that the N-terminal 26 amino acid signal peptide (SigP) of MBP modulated model precursor autoprocessing activities and outcomes mimicking phenotypes observed with the proviral constructs (137). The SigP-containing precursors also exhibited an association with

vesicle-like structures, which was different from the cytosolic distribution pattern of most fusion precursors lacking SigP. To further examine the correlation between precursor dimerization and *trans* processing activity, we engineered and tested another pair carrying the SigL (SigP following an N-terminal 6x His tag) at the N-terminus of the M1-PR miniprecursor (Fig 4.6A). Interestingly, SigL-M1 enzyme displayed similar detection profiles with or without 5μM IDV (Fig 4.9A), positively correlated with the input DNA amount. This was unlike GST-, Flag-M1, and C2-MBP fused enzymes that were barely detectable in the absence of a PI, suggesting that the SigP enzyme might exist in a conformation resistant to self-degradation, which was consistent with our previous report (137). The SigL-Myc-M2-PR^{D25N}-V5 substrate also showed comparable levels of detection when expressed alone with or without 5μM IDV (Fig 4.9B), but not linearly correlative to input DNA amounts. When the SigL enzyme and substrate were co-expressed at various ratios, the PR^{D25N}-V5 product was detected, confirming productive *trans* proteolysis (Fig 4.7C). However, no obvious reduction in sub detection was observed. Subsequent, the calculated *trans* proteolysis efficiencies demonstrated ~30% flat rates in the middle and low values at both ends of the tested Enz:Sub ratios (Fig 4.9F). Nonetheless, positive *trans* proteolysis was once again detected between SigP-containing precursors lacking any known dimer-inducing tag.

We also examined enzyme precursor complexes by sucrose gradient sedimentation (Fig 4.10A & Fig 4.11). The Flag-M1 enzyme formed well-defined dimers in the presence of 5μM IDV (Fig 8, dash orange line); the SigL-M1 enzyme showed a dimer peak with or without PI treatment (grey lines). This was rather interesting considering that the SigL-M1-FtoI-PR enzyme had no known dimer-inducing tag and it was previously shown that p6* peptide plays an inhibitory role in precursor dimerization (146, 148). Also, the C2-MBP-M2-PSHL-PR enzyme was detected as

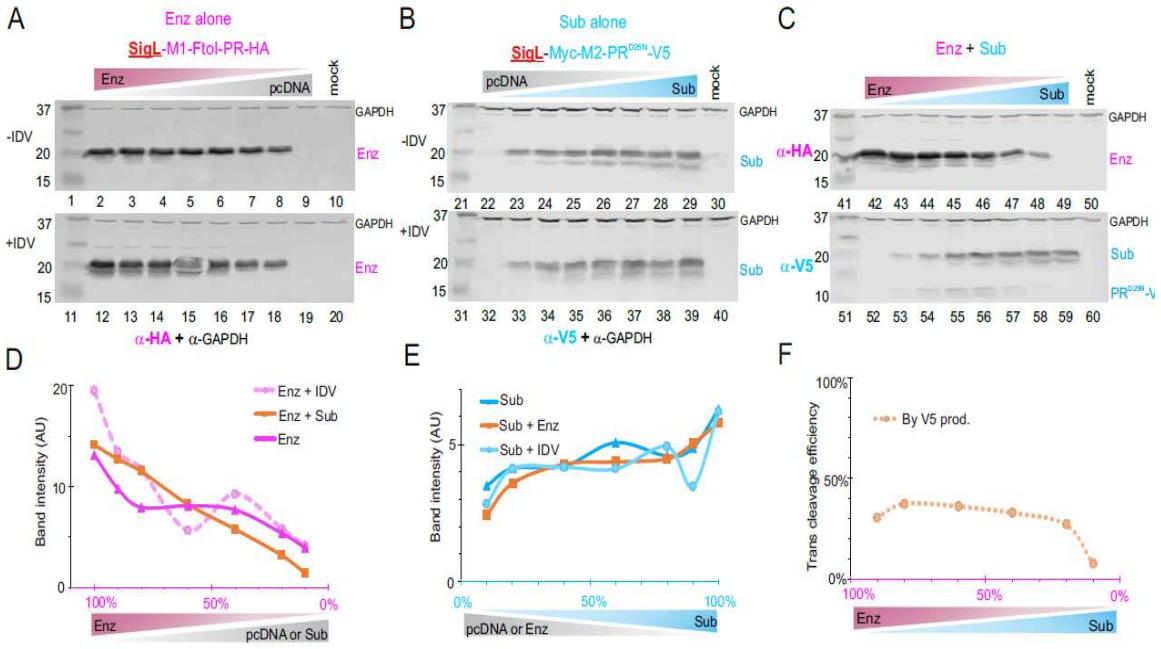

Figure 4.9. ***Trans*** **Proteolysis of precursors carrying MBP signal peptide.** (**A-C**) Western blot detection of the enzyme or substrate when expressed alone with or without 5μM IDV treatment or co-expressed at an Enz:Sub ratio of 1:4 input DNA. The image shown is representative of two experiments. (**D & E**) Quantification of the enzyme or substrate precursors. Band intensity of full-length precursor normalized to GAPDH was quantified to represent the amount of enzyme or substrate. (**F**) Quantification of *trans* processing efficiency by PR^{D25N}-V5/ (FL+ PR^{D25N}-V5).

predominantly monomers by sucrose gradient sedimentation analysis (Fig 4.4C). This could be collectively determined by the differences between these two enzymes: with or without SigP and monomeric C2-MBP, M1 *vs* M2 p6* peptide, PSHL *vs* FtoI mutation to prevent proximal cleavage. To pinpoint the factors that regulate precursor dimerization, we constructed and analyzed a panel of precursors using sucrose gradient sedimentation (Fig 4.12). For precursors containing SigP, it seemed that dimerization was influenced by both the p6* peptide length and proximal site mutations (Fig 4.10B & Fig 4.12I). Whereas M1-FtoI (gray line) showed dimers as the major form, M1-PSHL (yellow line) showed a mixture of monomers and dimers, suggesting that the proximal cleavage site functions more than just a proteolysis substrate; different amino acid alterations had different impact on precursor dimerization as well. From M1-PSHL to M2-PSHL (dash blue line), the enzymes became mostly monomers, suggesting an involvement of p6* peptide precursor dimerization (M1 has 13 more amino acids than M2, Fig 4.1A). Consistent with this assessment, we found all M2-PSHL precursors we tested were mostly monomers in the lysates (Fig 4.10 B-D). On the other hand, the M1-FtoI enzyme by itself did not mediate dimerization as Flag-M1-FtoI-PR existed mainly as monomers; C2-MBP-M1-FtoI-PR has a mixture of monomers and dimers. Only SigL-M1-FtoI-PR was detected mainly as dimers, suggesting a contribution of SigL to enzyme dimerization. Collectively, our data illustrated that precursor dimerization is modulated by multiple factors.

Trans proteolysis among various precursors at varying efficiencies

The preceding sections tested *trans* proteolysis between homotypic enzyme and substrate pairs. We next examined whether *trans* proteolysis occurred among combinations of enzyme and substrate precursors that differ in their dimerization propensity and subcellular distribution pattern.

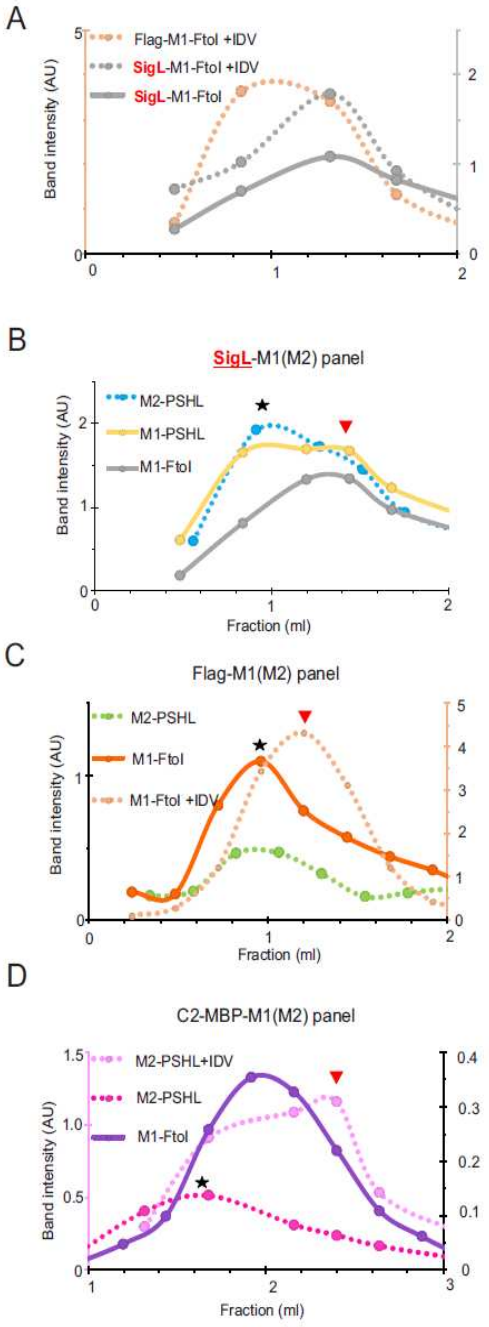

Figure 4.10. Examination of factors influencing precursor dimerization. Cell lysates of the indicated precursors were subjected to 5-50% sucrose gradient sedimentation followed by Western blotting analysis. Band intensity was quantified and plotted as a function of fraction volume from to top to bottom. The asterisk (*) and triangle (▼) denote monomeric and dimeric complexes, respectively.

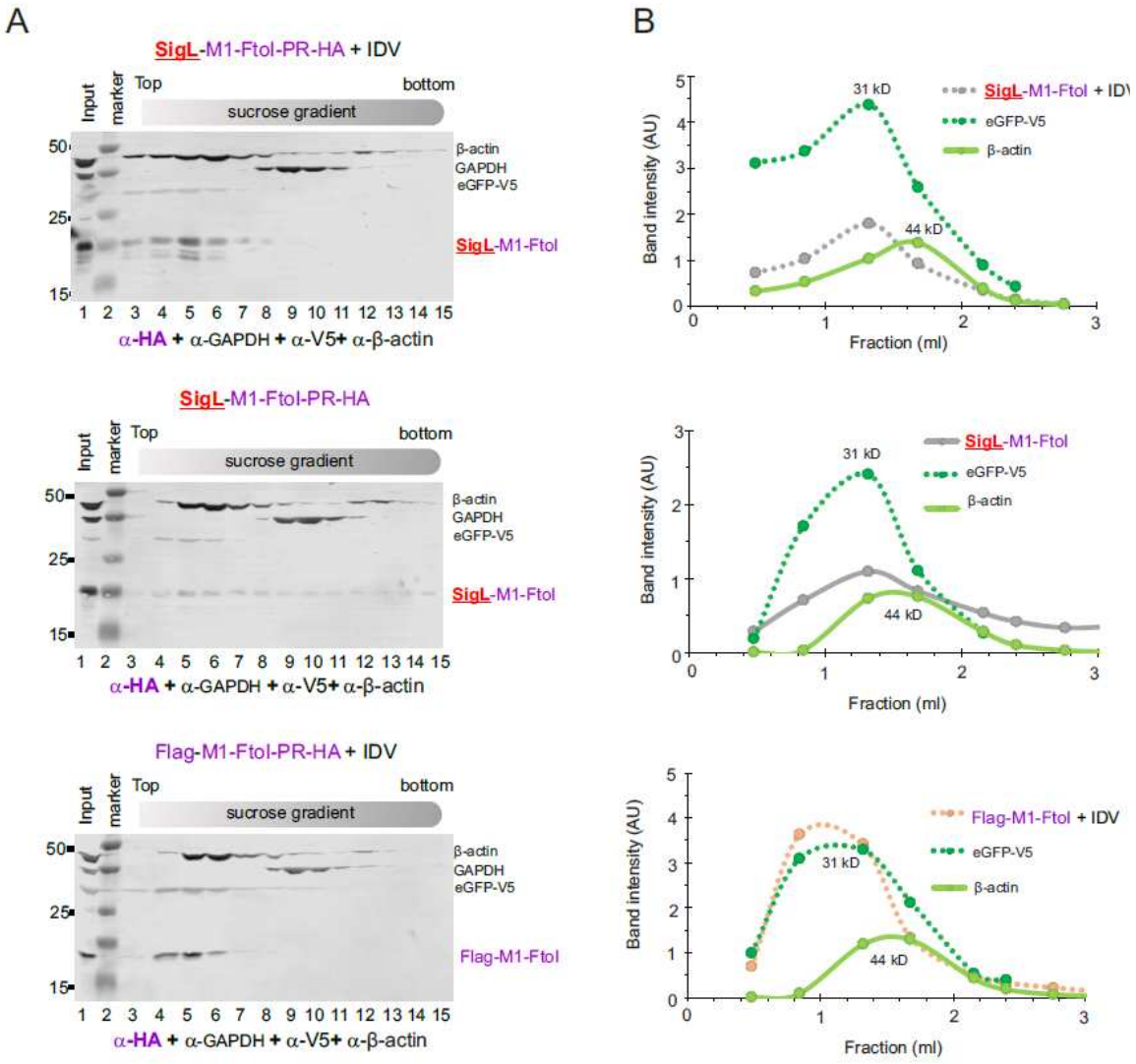

Figure 4.11. Detection of precursor complexes carrying SigL peptide. (**A**) Distribution profiles of the indicated precursors analyzed by sucrose gradient sedimentation followed by Western blotting. The image shown is representative of two experiments. (**B**) Band intensity quantification of each precursor as a function of fraction volume.

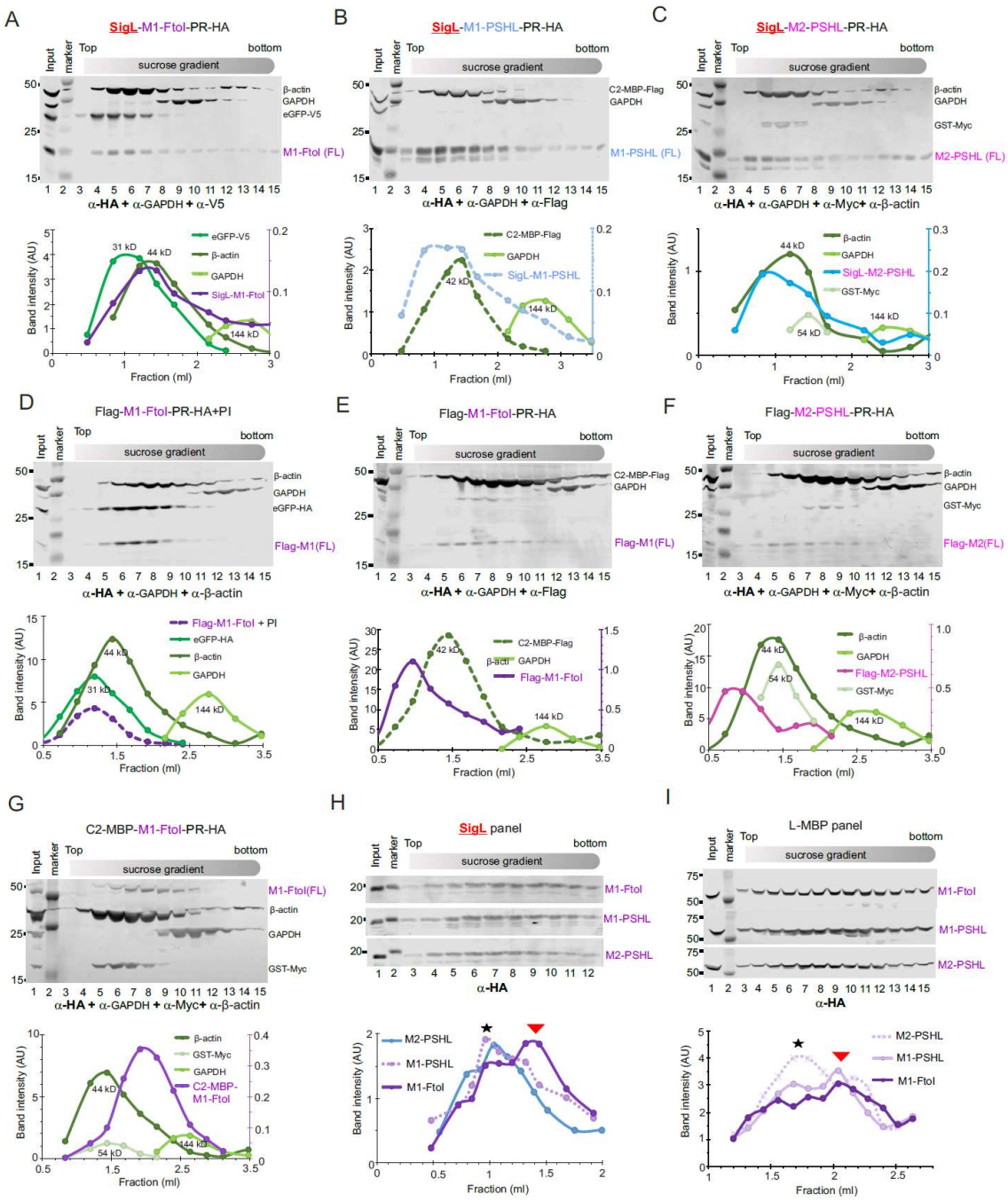

Figure 4.12. Detection of various precursor complexes in 5-50% sucrose gradients. (A-C) SigL-containing precursors with varied p6* sequences. (D-F) Flag-(M1) precursors with varied p6* sequences. (G) C2-MBP-M1-FtoI-PR-HA precursor. (H-I) SigL-containing precursors without (H) or with (I) C2-MBP domain. The asterisk (*) and triangle (▼) denote monomeric and dimeric complexes, respectively.

Flag-M1-FtoI enzyme and GST-fused substrate with or without N-terminal SigL or SigP were chosen for this test (Fig 4.13A). The use of GST-Myc fused substrate facilitated quantification of *trans* proteolysis efficiency by Myc antibody probing. Overall, positive *trans* proteolysis was detected in all the combinations with varied efficiencies (Fig 4.13 B-C, E-F). Higher efficiencies were observed with the enzyme/substrate pairs having the same expected subcellular distribution. With the cytosolic GST substrate, Flag-M1 enzyme worked better than SigL-M1 enzyme (Fig 4.13D); whereas with the membrane-associated sGST substrate, SigL-M1 enzyme processed more that Flag-M1 enzyme (Fig 4.13G). Nevertheless, positive *trans* proteolysis was detected in all the tested combinations by the enzymes lacking a dimer-inducing tag, confirming that a dimer-inducing tag flanked precursor enzyme is not required for *trans* proteolysis at least in our test model.

4.4 Discussion

HIV-1 protease autoprocessing is a fascinating process in that it must be suppressed during Gag and Gag-Pol synthesis and virion assembly, and then activated upon or shortly after virion release. Also, the Gag-Pol polyprotein precursor must function as both enzyme and substrate prior to liberation of any mature PR and it remains elusive whether the initial proteolysis occurs in *trans* (two precursors form an active enzyme to process a substrate precursor) or in *cis* (two precursors form an active enzyme to process itself). Built upon on a previously reported model system we developed for autoprocessing study, we here described characterization of *trans* proteolysis using various fusion precursors.

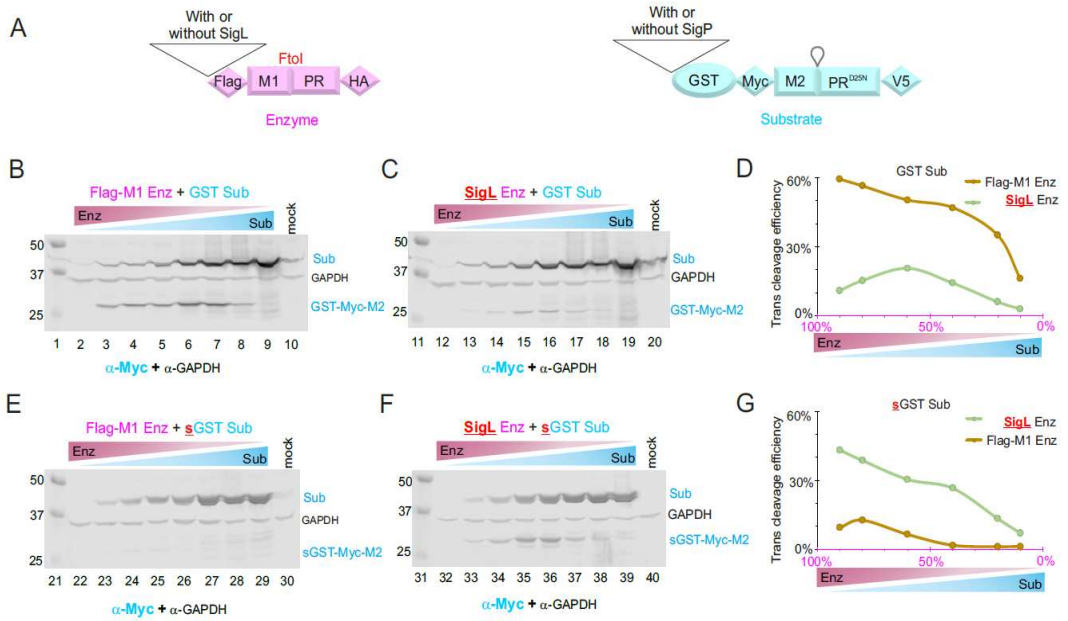

Figure 4.13. *Trans* proteolysis among precursors with various fusion tags. (**A**) Schematic of enzyme and substrate precursors. (**B & C**) Western blot analyses of *trans* proteolysis of GST-fused substrate. The image shown is representative of three experiments. (**D**) Quantification of *trans* proteolysis efficiency by P/(P+U) with Myc-reactive bands. (**E & F**) Western blot analyses of *trans* proteolysis of SigP-GST-fused substrate (**s**GST Sub). The image shown is representative of three experiments. (**G**) Quantification of *trans* proteolysis efficiency by P/(P+U) with Myc-reactive bands.

Precursor dimerization decoupled with dimer-inducing GST tag – Here we demonstrate detection of precursor dimers in transfected cells independent of a dimer-inducing tag, such as GST, upstream of the p6*-PR miniprecursor. For example, stable dimers were reproducibly detected with D25N substrate precursors with (Fig 4.2E) or without (Fig 4.7D) GST fusion. Furthermore, even when fused to a known monomeric tag (C2-MBP), the resulting substrate was still mainly dimeric (Fig 4.4B). Additionally, SigL-M1 enzyme precursor, not carrying any known dimer-inducing tag, was also predominantly dimeric (Fig 4.10A). These data collectively argued that the p6*-PR miniprecursor by itself is dimerization competent in the absence of a dimer-inducing tag. This finding is contradictory to the general assumption that p6*-PR should be mostly monomeric. We took the dimerization dissociation constant into consideration to reconcile this seemingly discrepancy. Louis et al reported many years ago that a purified recombinant model p6*-PR precursor had an apparent dimerization dissociation constant of 670nM (148), based on which it was suggested that precursor dimerization is solely driven by the precursor concentration that increases as the Gag and Gag-Pol assemble into viral particles. Therefore, our data could be interpreted to suggest that the fusion precursors were all expressed at high concentrations (*i.e.*, sub-micromolars) that induced precursor dimerization not requiring any dimer-inducing tag. This is certainly possible for most transiently expressed proteins and is indirectly supported by data showing that PIs at micromolar concentrations are needed to suppress precursor autoprocessing (127, 133-137, 142, 143, 149). Of course, determination of dissociation constants of these dimeric complexes would be critical to validate this concept, but this remains technically challenging using the cell-based model. Meanwhile, it is worth noting that these dimers remain associated during sucrose gradient sedimentation when a reduction in concentration is anticipated. Therefore, we postulate that factors other than precursor concentration are also involved in stabilization of dimer

complexes. In agreement with this speculation, we report that dimer detection is differentially influenced by various p6* peptide sequences and/or lengths (Fig 4.10 & Fig 4.12). Additionally, the reported decoupling doesn't exclude an involvement of a dimer-inducing tag in autoprocessing regulation. Instead, it is plausible that a dimer-inducing flanking tag is needed to facilitate precursor dimerization especially when precursor concentration is below dimerization dissociation constants as is with the Gag-Pol precursors made by *in vitro* transcription-coupled-translation (151). Autoprocessing activities of the tethered PR precursors with or without GST fusion also support this idea of flanking sequences contributing to regulation of autoprocessing (Table 4.1). Taken together, our results suggest that p6*-PR miniprecursor is able to undergo dimerization by itself but is subject to modulations by factors within and beyond its coding region.

***Trans* proteolysis uncoupled with enzyme dimerization –** Our data demonstrated positive *trans* proteolysis by a variety of enzyme precursors, some of which showed detection of dimers (Figs 4.2F & 4.10A) and some monomers (Fig 4.4C). Given that precursor dimerization is an obligatory step to form an active site, our results argue that *trans* proteolysis is not directly correlated with detection of stable dimers. It is intriguing that C2-MBP fused enzyme is predominantly monomeric but fully competent at *trans* proteolysis. We speculate that the C2-MBP enzyme exists at a dynamic equilibrium between dimers and monomers in the transfected cell whereas monomers were mostly detected by sucrose gradient sedimentation. In accordant with this speculation, we also detect mixtures of monomers and dimers from various enzyme precursors (Figs 4.7D, 4.10B-D, & 4.12). Based on our data showing that positive *trans* proteolysis is observed regardless of detection of stable dimers, we argue that stable dimerization is not linearly correlated *trans* proteolysis activity. Some enzymatically active precursors (e.g., C2-MBP enzyme) could form dimers only transiently,

which was previously illustrated using a recombinant model precursor (175). On the other hand, stable dimers also exist under certain precursor contexts (Fig 4.10A, & 4.12G&H). Consequently, our results indicate the existence of more than one active form of precursor enzyme capable of *trans* proteolysis. Different contexts (p6*-PR sequences along with flanking tags) appeared to modulate formation of these different conformations (stable or unstable dimers), revealing precursor conformation plasticity and complicated regulation involved in autoprocessing reactions. Additional structural and biophysical analyses of precursor are essential to advance our understanding of the precursor autoprocessing mechanism.

CHAPTER 5

SUMMARY AND FUTURE PERSPECTIVE

5.1 Summary

My dissertation focuses on the regulatory mechanism of precursor autoprocessing (chapter 2), biophysical characterization of the miniprecursor under native conditions (chapter 3), and the *trans* cleavage mechanism of precursor (chapter 4). These studies demonstrate that precursor dimerization and precursor autoprocessing activity are subject to various regulation. For example, autoprocessing activity of the H69D mutation differs under different "context": abolished in SigP-carrying precursors and proviral constructs, and partially suppressed in GST- or C2-MBP- fused precursors. The mature PRs released from SigP-containing precursors or associated with the viral particles are resistant to self-degradation, whereas those released from GST- or C2-MBP- fused precursors are prone to rapid self-degradation. Furthermore, my data reveals that positive *trans* proteolysis between engineered enzyme and substrate precursors is independent of the dimer-inducing GST tag. We also examined the complex formation of these fusion precursors. Surprisingly, the monomeric enzyme precursor is a monomer, but it had similar cleavage processing efficiency to the dimeric enzyme precursor. This finding indicates that transient dimerization is sufficient for *trans* processing, suggesting that the dimer-monomer transition is involved in the *trans* process of monomeric enzyme precursor. Precursor dimerization is also context-dependent regulated by both sigP and p6* peptide.

The autoprocessing activity of fusion precursor expressed in *E.coli* is also explored. DnaK fusion precursor carrying D25N, H69D, or R87K mutant is autoprocessing deficient. On the other hand, C2MBP fusion precursor bearing D25N, H69D, R87K, or PR1-95 is autoprocessing deficient, indicating that D25N, H69D, R87K, or PR1-95 are possibly good candidates in different contexts for purification. In summary, this dissertation highlights the mechanism of precursor autoprocessing and demonstrates the complex formation of the precursor. The precursor needs to form dimers to undergo autoprocessing, which is regulated by its context and p6* peptide.

5.2 Understanding the role of viral protein matrix (MA) in regulating precursor autoprocessing

This was a side project of chapter 2. The MBP SigP modulated precursor autoprocessing in a way that mimicked the proviral context and targeted the fusion precursors to vesicle-like structures (137). Accordingly, we speculated that the local membrane environment (lipid composition, charge property, and/or membrane associated proteins) of these vesicles might be involved in modulating the precursor activity. Because viral matrix (MA) protein is known to mediate Gag targeting to the plasma membrane where Gag and Gag-Pol assembly occur (176), two versions of MA-fused precursors were engineered to examine their role in precursor autoprocessing regulation. One version is derived from HIV-1 MA (hMA), and the other one is from equine infectious anemia virus (EIAV) MA (eMA). Both MAs have HA epitopes at the C-terminus and a G2A mutation, which abolishes co-translational myristoylation of MA and thus prevents plasma membrane targeting (177). Both fusion precursors contain two cleavage sites: the cleavage site 1 (SQNY/PIVQ for hMA or SEEY/PIMI for eMA) between MA & capsid and the proximal cleavage

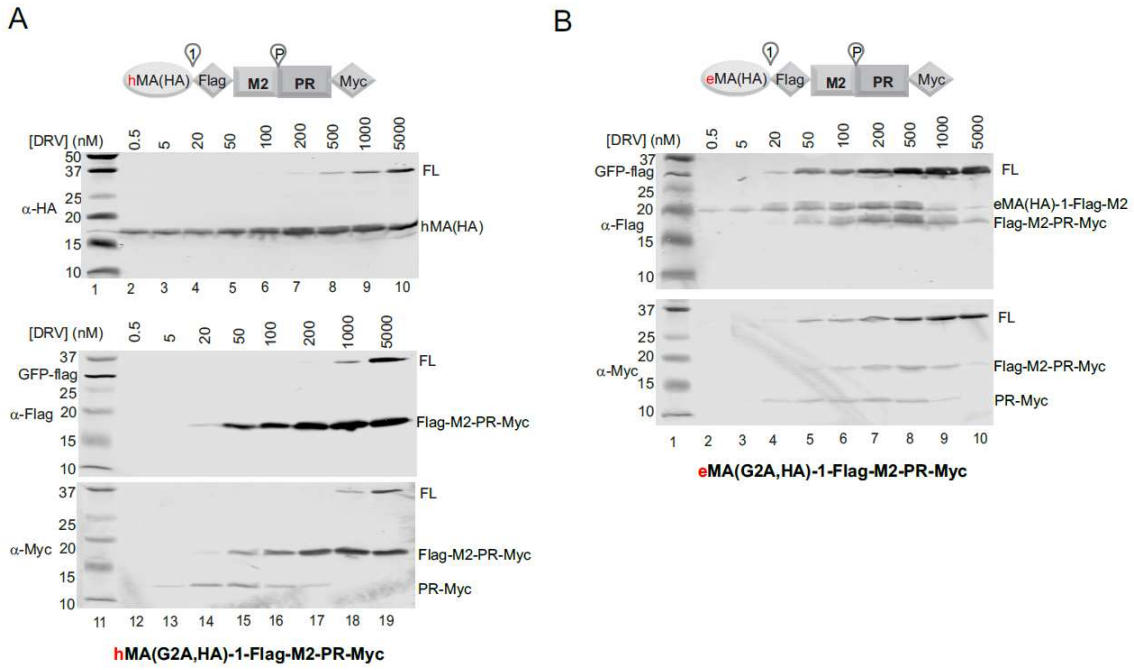

Fig 5.1 Autoprocessing of fusion precursor in the context of matrix (MA). Schematic diagram of the HIV-1 MA (a) or equine infectious anemia virus (EIAV) MA (b) fusion precursors. The cell lysates collected from HEK 293T cells transfected with MA fusion precursor, and were examined with mouse anti-Flag, anti-HA and anti-myc antibodies.

site between p6* & PR. M2 is a truncated p6*, which lacks the first twenty N-terminal amino acids (Fig 2.2A). Autoprocessing of these MA-fused precursors was tested with increasing concentrations of darunavir (DRV). The released mature PR and M2-PR were prone to self-degradation from both fusion precursors (Fig 5.1A, lane 12, and Fig 5.1B, lane 2, bottom panel). Autoprocessing of hMA precursor was much less sensitive to darunavir inhibition; cleavage products hMA and Flag-M2-PR-myc were still detected at 5 μM DRV (Fig 5.1A, lane 10 & 19). Also, the hMA-Flag-M2 product, as expected from precursor processing the proximal site, was not detected at all, indicating that this reaction didn't happen while precursor processing at site 1 dominated. In comparison, autoprocessing of eMA precursor was more responsive to darunavir inhibition and approximately equal processing frequencies at the two cleavage sites. It is interesting to note that the EIAV MA↓CA cleavage site was effectively by HIV precursor in this experiment, but nonetheless showed a trend similar to hMA precursor. This is another example showing that autoprocessing activity (and PI sensitivity) of p6*-PR precursor is modulated by different fusion tags. However, this result also argues against the postulation of membrane-targeting as an underlying mechanism of the SigP phenotype. Additional experiments need to be carried out to better understand the mechanism. For example, I'd like to examine/determine subcellular distribution patterns of these MA-fused precursors as it is possible that the MA domain by itself is insufficient to mediate membrane targeting such that they simply functioned as a fusion tag. Also, it will be interesting to compare the MA-fusion precursors carrying G2A mutation to further define whether and how MA domain modulate p6*-PR autoprocessing activities.

5.3 Developing *in vitro trans* cleavage assay

Proteolysis of fusion substrate by fusion enzyme was studied in transfected cells in chapter 4. Sucrose gradient sedimentation analyses identified stable substrate dimers and various enzyme complexes (monomers, dimers, or mixtures of the two) in the cell lysates. From the substrate part, GST fusion precursor and C2MBP fusion precursor carrying D25N were dimeric proteins. The enzyme complexes are dependent on its context and the residue of p6*. For example, in the context of SigP, the majority of M1-FtoI-PR is a dimer whereas M2-PSHL-PR is a monomer. The enzyme of C2MBP fusion precursor (M2-PSHL-PR) is a monomer. The substitution of FtoI is at the last amino acid of p6*; the substitution of PSHL is at the last four amino acids of p6* replacing SFSF. The *in vivo trans* cleavage assay showed that C2MBP fusion precursor enzyme (M2-PSHL-PR) is able to process the C2MBP fusion precursor substrate, indicating that transient dimers of enzyme is required to precursor autoprocessing in the context of C2MBP. Based on these results, I hope to develop an *in vitro trans* proteolysis assay and proposed three phases towards this goal.

Phase I will test *trans* proteolysis using cell lysates made from transfected cells individually expressing the enzyme or substrate. These tests will hopefully answer questions like 1) whether *trans* proteolysis requires a reducing condition as inside cells; 2) whether energy consumption is involved for the implicated conformational change during precursor autoprocessing; 3) whether substrate and enzyme concentration matter. Phase II will study *trans* proteolysis using enzyme and substrate precursors partially purified from sucrose gradient sedimentation. We collect each faction and analyze them by western blot. The advantage of this phase is that we can use specific fractions with known complex properties to sort out components essential for *trans* proteolysis. Phase III will use substrate and enzyme precursors purified from *E coli* to define minimal essential players involved in *trans* proteolysis and test candidates that modulate the reaction. This assay

will help us better understand how an enzyme precursor processes a substrate precursor in a great detail.

5.4 Examine the miniprecursor (p6*-PR) associated with viral particles by sucrose gradient sedimentation

In chapter 4, I demonstrated that the majority of SigP-M1-FtoI-PR existed as dimers (Fig 4.8), whereas the majority of M2-PSHL-PR existed as monomers. The p6* has 56 amino acids, and it is predicted to be intrinsically disordered (131, 132), suggesting that p6* itself does not exhibit any defined structure. Theoretically, it might exhibit ordered structures when it is associated with other viral components or host factors. Our previous results shows that SigP precursor had similar autoprocessing outcomes as in the context of proviral construct (137), suggesting that the SigP is mimicking proviral context to regulate precursor autoprocessing. Based on these results, I would like to engineer the FotI and PSHL mutants into the proviral construct to determine whether the p6*-PR fragments released from these Gag-Pol exist as monomers or dimers. Unfortunately, we cannot delete the upstream region of p6* to study whether the length affects the precursor dimerization in this setting. The upstream region of p6* contains slippery region and frameshift stimulatory signal (FSS) (75, 76). If we delete this region, the Gag-Pol cannot be produces. Moreover, the p6* of Gag-Pol polyprotein overlaps with the SP2 and p6 of Gag polyprotein. The p6 is responsible for viral assembly and viral budding (85, 86). It is the challenge to delete the p6*, but it does not affect the function of p6. Thus, we will engineer FotI and PSHL proviral constructs and transfect HEK293T cell with these constructs. Then the virus like particles (VLP) will be collected and lysed to release its components. One perceivable problem is whether the lysis buffer

used for sucrose gradient sedimentation could "dissolve" the capsid (CA) core to release the p6*-PR complexes. We already know that PSHL blocks the cleavage site between p6* and PR, so the viral particle from PSHL proviral construct only releases p6*-PSHL-PR fragment (91). The p6*-PR only processes some cleavage sites of Gag and Gag-Pol polyproteins, and the p25 cannot be processed to produce p24. As a result, the PSHL or FtoI viral particles are expected not to be fully mature. Thus, our lysis buffer might be able dissolve them. However, when the wild-type proviral construct is used as a control, things might change as a completely enclosed CA core is anticipated. This needs to be tested using different lysis buffers followed by sucrose gradient velocity sedimentation to determine lysis conditions. The next problem is existence of heterogeneous PR-containing intermediate, and the p6*-PR may associate with other intermediate precursors. If these potential issues can be solved, this work will shed new light into precursor autoprocessing mechanism.

REFERENCES

1. **Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL, Weeks KM.** 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature **460:**711-716.
2. **Hearps AC, Jans DA.** 2007. Regulating the functions of the HIV-1 matrix protein. AIDS Res Hum Retroviruses **23:**341-346.
3. **Freed EO.** 2015. HIV-1 assembly, release and maturation. Nat Rev Microbiol **13:**484-496.
4. **Gamble TR, Yoo S, Vajdos FF, von Schwedler UK, Worthylake DK, Wang H, McCutcheon JP, Sundquist WI, Hill CP.** 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science **278:**849-853.
5. **Hellmund C, Lever AM.** 2016. Coordination of genomic rna packaging with viral assembly in HIV-1. Viruses **8**.
6. **Demirov DG, Orenstein JM, Freed EO.** 2002. The late domain of human immunodeficiency virus type 1 p6 promotes virus release in a cell type-dependent manner. J Virol **76:**105-117.
7. **Checkley MA, Luttge BG, Freed EO.** 2011. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. J Mol Biol **410:**582-608.
8. **Ruben S, Perkins A, Purcell R, Joung K, Sia R, Burghoff R, Haseltine WA, Rosen CA.** 1989. Structural and functional characterization of human immunodeficiency virus tat protein. J Virol **63:**1-8.
9. **Le Rouzic E, Benichou S.** 2005. The Vpr protein from HIV-1: distinct roles along the viral life cycle. Retrovirology **2:**11.

10. **Bouzar AB, Guiguen F, Morin T, Villet S, Fornazero C, Garnier C, Gallay K, Gounel F, Favier C, Durand J, Balleydier S, Mornex JF, Narayan O, Chebloune Y.** 2003. Specific G2 arrest of caprine cells infected with a caprine arthritis encephalitis virus expressing vpr and vpx genes from simian immunodeficiency virus. Virology **309:**41-52.

11. **Klimkait T, Strebel K, Hoggan MD, Martin MA, Orenstein JM.** 1990. The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release. J Virol **64:**621-629.

12. **Andrew A, Strebel K.** 2014. HIV-1 accessory proteins: Vpu and Vif. Methods Mol Biol **1087:**135-158.

13. **Miller MD, Warmerdam MT, Gaston I, Greene WC, Feinberg MB.** 1994. The human immunodeficiency virus-1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages. J Exp Med **179:**101-113.

14. **Gustchina A, Weber IT.** 1991. Comparative analysis of the sequences and structures of HIV-1 and HIV-2 proteases. Proteins **10:**325-339.

15. **Hill AL, Rosenbloom DIS, Nowak MA, Siliciano RF.** 2018. Insight into treatment of HIV infection from viral dynamics models. Immunol Rev **285:**9-25.

16. **Doms RW, Moore JP.** 2000. HIV-1 membrane fusion: targets of opportunity. J Cell Biol **151:**F9-14.

17. **Blumenthal R, Durell S, Viard M.** 2012. HIV entry and envelope glycoprotein-mediated fusion. J Biol Chem **287:**40841-40849.

18. **Jakobsdottir GM, Iliopoulou M, Nolan R, Alvarez L, Compton AA, Padilla-Parra S.** 2017. On the Whereabouts of HIV-1 Cellular Entry and Its Fusion Ports. Trends Mol Med **23:**932-944.

19. **Zhu J, Paul WE.** 2008. CD4 T cells: fates, functions, and faults. Blood **112:**1557-1569.

20. **Murphy PM.** 2019. Clinial Immunology, Fifth ed.

21. **Sun X, Cheng G, Hao M, Zheng J, Zhou X, Zhang J, Taichman RS, Pienta KJ, Wang J.** 2010. CXCL12 / CXCR4 / CXCR7 chemokine axis and cancer progression. Cancer Metastasis Rev **29:**709-722.

22. **Pugach P, Ketas TJ, Michael E, Moore JP.** 2008. Neutralizing antibody and anti-retroviral drug sensitivities of HIV-1 isolates resistant to small molecule CCR5 inhibitors. Virology **377:**401-407.

23. **Westby M, Smith-Burchnell C, Mori J, Lewis M, Mosley M, Stockdale M, Dorr P, Ciaramella G, Perros M.** 2007. Reduced maximal inhibition in phenotypic susceptibility assays indicates that viral strains resistant to the CCR5 antagonist maraviroc utilize inhibitor-bound receptor for entry. J Virol **81:**2359-2371.

24. **Lalezari JP, Eron JJ, Carlson M, Cohen C, DeJesus E, Arduino RC, Gallant JE, Volberding P, Murphy RL, Valentine F, Nelson EL, Sista PR, Dusek A, Kilby JM.** 2003. A phase II clinical study of the long-term safety and antiviral activity of enfuvirtide-based antiretroviral therapy. AIDS **17:**691-698.

25. **LaBonte J, Lebbos J, Kirkpatrick P.** 2003. Enfuvirtide. Nat Rev Drug Discov **2:**345-346.

26. **Gitti RK, Lee BM, Walker J, Summers MF, Yoo S, Sundquist WI.** 1996. Structure of the amino-terminal core domain of the HIV-1 capsid protein. Science **273:**231-235.

27. **Jiang J, Ablan SD, Derebail S, Hercik K, Soheilian F, Thomas JA, Tang S, Hewlett I, Nagashima K, Gorelick RJ, Freed EO, Levin JG.** 2011. The interdomain linker region

of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. Virology **421:**253-265.

28. **Pornillos O, Ganser-Pornillos BK, Yeager M.** 2011. Atomic-level modelling of the HIV capsid. Nature **469:**424-427.

29. **Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI.** 1999. Assembly and analysis of conical models for the HIV-1 core. Science **283:**80-83.

30. **Welker R, Hohenberg H, Tessmer U, Huckhagel C, Krausslich HG.** 2000. Biochemical and structural analysis of isolated mature cores of human immunodeficiency virus type 1. J Virol **74:**1168-1177.

31. **Lukic Z, Dharan A, Fricke T, Diaz-Griffero F, Campbell EM.** 2014. HIV-1 uncoating is facilitated by dynein and kinesin 1. J Virol **88:**13613-13625.

32. **Pawlica P, Berthoux L.** 2014. Cytoplasmic dynein promotes HIV-1 uncoating. Viruses **6:**4195-4211.

33. **Ambrose Z, Aiken C.** 2014. HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. Virology **454-455:**371-379.

34. **Warren K, Wei T, Li D, Qin F, Warrilow D, Lin MH, Sivakumaran H, Apolloni A, Abbott CM, Jones A, Anderson JL, Harrich D.** 2012. Eukaryotic elongation factor 1 complex subunits are critical HIV-1 reverse transcription cofactors. Proc Natl Acad Sci U S A **109:**9587-9592.

35. **Hulme AE, Perez O, Hope TJ.** 2011. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. Proc Natl Acad Sci U S A **108:**9975-9980.

36. **Zhou L, Sokolskaja E, Jolly C, James W, Cowley SA, Fassati A.** 2011. Transportin 3 promotes a nuclear maturation step required for efficient HIV-1 integration. PLoS Pathog **7:**e1002194.

37. **Matreyek KA, Yucel SS, Li X, Engelman A.** 2013. Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. PLoS Pathog **9:**e1003693.

38. **Shi J, Zhou J, Shah VB, Aiken C, Whitby K.** 2011. Small-molecule inhibition of human immunodeficiency virus type 1 infection by virus capsid destabilization. J Virol **85:**542-549.

39. **Blair WS, Pickford C, Irving SL, Brown DG, Anderson M, Bazin R, Cao J, Ciaramella G, Isaacson J, Jackson L, Hunt R, Kjerrstrom A, Nieman JA, Patick AK, Perros M, Scott AD, Whitby K, Wu H, Butler SL.** 2010. HIV capsid is a tractable target for small molecule therapeutic intervention. PLoS Pathog **6:**e1001220.

40. **Xu H, Franks T, Gibson G, Huber K, Rahm N, Strambio De Castillia C, Luban J, Aiken C, Watkins S, Sluis-Cremer N, Ambrose Z.** 2013. Evidence for biphasic uncoating during HIV-1 infection from a novel imaging assay. Retrovirology **10:**70.

41. **Hu WS, Hughes SH.** 2012. HIV-1 reverse transcription. Cold Spring Harb Perspect Med **2**.

42. **Cimarelli A, Darlix JL.** 2014. HIV-1 reverse transcription. Methods Mol Biol **1087:**55-70.

43. **Skalka AM.** 2014. Retroviral DNA Transposition: Themes and Variations. Microbiol Spectr **2:**MDNA300052014.

44. **Darlix JL, Godet J, Ivanyi-Nagy R, Fosse P, Mauffret O, Mely Y.** 2011. Flexible nature and specific functions of the HIV-1 nucleocapsid protein. J Mol Biol **410:**565-581.

45. **Goldschmidt V, Rigourd M, Ehresmann C, Le Grice SF, Ehresmann B, Marquet R.** 2002. Direct and indirect contributions of RNA secondary structure elements to the initiation of HIV-1 reverse transcription. J Biol Chem **277:**43233-43242.

46. **Choi E, Mallareddy JR, Lu D, Kolluru S.** 2018. Recent advances in the discovery of small-molecule inhibitors of HIV-1 integrase. Future Sci OA **4:**FSO338.

47. **Engelman AN, Singh PK.** 2018. Cellular and molecular mechanisms of HIV-1 integration targeting. Cell Mol Life Sci **75:**2491-2507.

48. **Zheng R, Jenkins TM, Craigie R.** 1996. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. Proc Natl Acad Sci U S A **93:**13659-13664.

49. **Engelman A, Cherepanov P.** 2014. Retroviral Integrase Structure and DNA Recombination Mechanism. Microbiol Spectr **2:**1-22.

50. **Roth MJ, Schwartzberg PL, Goff SP.** 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. Cell **58:**47-54.

51. **Ciuffi A.** 2016. The benefits of integration. Clin Microbiol Infect **22:**324-332.

52. **Karn J, Stoltzfus CM.** 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med **2:**a006916.

53. **Marcello A, Zoppe M, Giacca M.** 2001. Multiple modes of transcriptional regulation by the HIV-1 Tat transactivator. IUBMB Life **51:**175-181.

54. **Mousseau G, Valente ST.** 2017. Role of Host Factors on the Regulation of Tat-Mediated HIV-1 Transcription. Curr Pharm Des **23:**4079-4090.

55. **Taube R, Peterlin M.** 2013. Lost in transcription: molecular mechanisms that control HIV latency. Viruses **5:**902-927.

56. **Ramakrishnan R, Dow EC, Rice AP.** 2009. Characterization of Cdk9 T-loop phosphorylation in resting and activated CD4(+) T lymphocytes. J Leukoc Biol **86:**1345-1350.

57. **Stoltzfus CM.** 2009. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. Adv Virus Res **74:**1-40.

58. **Zhang X, Aida Y.** 2009. HIV-1 Vpr: a novel role in regulating RNA splicing. Curr HIV Res **7:**163-168.

59. **Kuzembayeva M, Dilley K, Sardo L, Hu WS.** 2014. Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. Virology **454-455:**362-370.

60. **Langer S, Sauter D.** 2016. Unusual Fusion Proteins of HIV-1. Front Microbiol **7:**2152.

61. **Phuphuakrat A, Auewarakul P.** 2003. Heterogeneity of HIV-1 Rev response element. AIDS Res Hum Retroviruses **19:**569-574.

62. **Fernandes J, Jayaraman B, Frankel A.** 2012. The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. RNA Biol **9:**6-11.

63. **Kuss SK, Mata MA, Zhang L, Fontoura BM.** 2013. Nuclear imprisonment: viral strategies to arrest host mRNA nuclear export. Viruses **5:**1824-1849.

64. **Guerrero S, Batisse J, Libre C, Bernacchi S, Marquet R, Paillart JC.** 2015. HIV-1 replication and the cellular eukaryotic translation apparatus. Viruses **7:**199-218.

65. **Ohlmann T, Mengardi C, Lopez-Lastra M.** 2014. Translation initiation of the HIV-1 mRNA. Translation (Austin) **2:**e960242.

66. **Yang Q, Del Campo M, Lambowitz AM, Jankowsky E.** 2007. DEAD-box proteins unwind duplexes by local strand separation. Mol Cell **28:**253-263.

67. **Brasey A, Lopez-Lastra M, Ohlmann T, Beerens N, Berkhout B, Darlix JL, Sonenberg N.** 2003. The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. J Virol **77:**3939-3949.

68. **Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF.** 2001. The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. J Virol **75:**181-191.

69. **Vallejos M, Deforges J, Plank TD, Letelier A, Ramdohr P, Abraham CG, Valiente-Echeverria F, Kieft JS, Sargueil B, Lopez-Lastra M.** 2011. Activity of the human immunodeficiency virus type 1 cell cycle-dependent internal ribosomal entry site is modulated by IRES trans-acting factors. Nucleic Acids Res **39:**6186-6200.

70. **Gendron K, Ferbeyre G, Heveker N, Brakier-Gingras L.** 2011. The activity of the HIV-1 IRES is stimulated by oxidative stress and controlled by a negative regulatory element. Nucleic Acids Res **39:**902-912.

71. **Jager S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K, Hernandez H, Jang GM, Roth SL, Akiva E, Marlett J, Stephens M, D'Orso I, Fernandes J, Fahey M, Mahon C, O'Donoghue AJ, Todorovic A, Morris JH, Maltby DA, Alber T, Cagney G, Bushman FD, Young JA, Chanda SK, Sundquist WI, Kortemme T, Hernandez RD, Craik CS, Burlingame A,**

Sali A, Frankel AD, Krogan NJ. 2011. Global landscape of HIV-human protein complexes. Nature **481:**365-370.

72. **Castello A, Franco D, Moral-Lopez P, Berlanga JJ, Alvarez E, Wimmer E, Carrasco L.** 2009. HIV- 1 protease inhibits Cap- and poly(A)-dependent translation upon eIF4GI and PABP cleavage. PLoS One **4:**e7997.

73. **Sharma A, Yilmaz A, Marsh K, Cochrane A, Boris-Lawrie K.** 2012. Thriving under stress: selective translation of HIV-1 structural protein mRNA during Vpr-mediated impairment of eIF4E translation activity. PLoS Pathog **8:**e1002612.

74. **Gale M, Jr., Tan SL, Katze MG.** 2000. Translational control of viral gene expression in eukaryotes. Microbiol Mol Biol Rev **64:**239-280.

75. **Mouzakis KD, Lang AL, Vander Meulen KA, Easterday PD, Butcher SE.** 2013. HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. Nucleic Acids Res **41:**1901-1913.

76. **Staple DW, Butcher SE.** 2005. Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. J Mol Biol **349:**1011-1023.

77. **Sundquist WI, Krausslich HG.** 2012. HIV-1 assembly, budding, and maturation. Cold Spring Harb Perspect Med **2:**a006924.

78. **Krausslich HG, Facke M, Heuser AM, Konvalinka J, Zentgraf H.** 1995. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. J Virol **69:**3407-3419.

79. **Accola MA, Strack B, Gottlinger HG.** 2000. Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. J Virol **74:**5395-5402.

80. **Wright ER, Schooler JB, Ding HJ, Kieffer C, Fillmore C, Sundquist WI, Jensen GJ.** 2007. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. EMBO J **26:**2218-2226.

81. **Jouvenet N, Simon SM, Bieniasz PD.** 2011. Visualizing HIV-1 assembly. J Mol Biol **410:**501-511.

82. **Lee SK, Potempa M, Swanstrom R.** 2012. The choreography of HIV-1 proteolytic processing and virion assembly. J Biol Chem **287:**40867-40874.

83. **Lingappa JR, Reed JC, Tanaka M, Chutiraka K, Robinson BA.** 2014. How HIV-1 Gag assembles in cells: Putting together pieces of the puzzle. Virus Res **193:**89-107.

84. **Schoneberg J, Lee IH, Iwasa JH, Hurley JH.** 2017. Reverse-topology membrane scission by the ESCRT proteins. Nat Rev Mol Cell Biol **18:**5-17.

85. **Pornillos O, Alam SL, Davis DR, Sundquist WI.** 2002. Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. Nat Struct Biol **9:**812-817.

86. **Zhai Q, Fisher RD, Chung HY, Myszka DG, Sundquist WI, Hill CP.** 2008. Structural and functional studies of ALIX interactions with YPX(n)L late domains of HIV-1 and EIAV. Nat Struct Mol Biol **15:**43-49.

87. **Pettit SC, Sheng N, Tritch R, Erickson-Viitanen S, Swanstrom R.** 1998. The regulation of sequential processing of HIV-1 Gag by the viral protease. Adv Exp Med Biol **436:**15-25.

88. **Campbell EM, Hope TJ.** 2015. HIV-1 capsid: the multifaceted key player in HIV-1 infection. Nat Rev Microbiol **13:**471-483.

89. **Paillart JC, Shehu-Xhilaga M, Marquet R, Mak J.** 2004. Dimerization of retroviral RNA genomes: an inseparable pair. Nat Rev Microbiol **2:**461-472.

90. **Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R.** 2005. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. Retrovirology **2:**66.

91. **Ludwig C, Leiherer A, Wagner R.** 2008. Importance of protease cleavage sites within and flanking human immunodeficiency virus type 1 transframe protein p6* for spatiotemporal regulation of protease activation. J Virol **82:**4573-4584.

92. **Brik A, Wong CH.** 2003. HIV-1 protease: mechanism and drug discovery. Org Biomol Chem **1:**5-14.

93. **Louis JM, Ishima R, Nesheiwat I, Pannell LK, Lynch SM, Torchia DA, Gronenborn AM.** 2003. Revisiting monomeric HIV-1 protease. Characterization and redesign for improved properties. J Biol Chem **278:**6085-6092.

94. **de Oliveira T, Engelbrecht S, Janse van Rensburg E, Gordon M, Bishop K, zur Megede J, Barnett SW, Cassol S.** 2003. Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? J Virol **77:**9422-9430.

95. **Tykvart J, Barinka C, Svoboda M, Navratil V, Soucek R, Hubalek M, Hradilek M, Sacha P, Lubkowski J, Konvalinka J.** 2015. Structural and biochemical characterization of a novel aminopeptidase from human intestine. J Biol Chem **290:**11321-11336.

96. **Huang L, Chen C.** 2013. Understanding HIV-1 protease autoprocessing for novel therapeutic development. Future Med Chem **5:**1215-1229.

97. **Konvalinka J, Krausslich HG, Muller B.** 2015. Retroviral proteases and their roles in virion maturation. Virology **479-480:**403-417.

98. **Wiegers K, Rutter G, Kottler H, Tessmer U, Hohenberg H, Krausslich HG.** 1998. Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. J Virol **72:**2846-2854.

99. **Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA, Swanstrom R.** 1994. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. J Virol **68:**8017-8027.

100. **McKeever BM, Navia MA, Fitzgerald PM, Springer JP, Leu CT, Heimbach JC, Herbert WK, Sigal IS, Darke PL.** 1989. Crystallization of the aspartylprotease from the human immunodeficiency virus, HIV-1. J Biol Chem **264:**1919-1921.

101. **Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SB, Wlodawer A.** 1989. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. Science **246:**1149-1152.

102. **Navia MA, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP.** 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature **337:**615-620.

103. **Rose JR, Salto R, Craik CS.** 1993. Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. J Biol Chem **268:**11939-11945.

104. **Mildner AM, Rothrock DJ, Leone JW, Bannow CA, Lull JM, Reardon IM, Sarcich JL, Howe WJ, Tomich CS, Smith CW, et al.** 1994. The HIV-1 protease as enzyme and

substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. Biochemistry **33:**9405-9413.

105. **Palese LL.** 2017. Analysis of the conformations of the HIV-1 protease from a large crystallographic data set. Data Brief **15:**696-700.

106. **Palese LL.** 2017. Conformations of the HIV-1 protease: A crystal structure data set analysis. Biochim Biophys Acta Proteins Proteom **1865:**1416-1422.

107. **Louis JM, Ishima R, Torchia DA, Weber IT.** 2007. HIV-1 protease: structure, dynamics, and inhibition. Adv Pharmacol **55:**261-298.

108. **Pearl L, Blundell T.** 1984. The active site of aspartic proteinases. FEBS Lett **174:**96-101.

109. **Strisovsky K, Tessmer U, Langner J, Konvalinka J, Krausslich HG.** 2000. Systematic mutational analysis of the active-site threonine of HIV-1 proteinase: rethinking the "fireman's grip" hypothesis. Protein Sci **9:**1631-1641.

110. **Prabu-Jeyabalan M, Nalivaika E, Schiffer CA.** 2002. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. Structure **10:**369-381.

111. **Ishima R, Ghirlando R, Tozser J, Gronenborn AM, Torchia DA, Louis JM.** 2001. Folded monomer of HIV-1 protease. J Biol Chem **276:**49110-49116.

112. **Sadiq SK, De Fabritiis G.** 2010. Explicit solvent dynamics and energetics of HIV-1 protease flap opening and closing. Proteins **78:**2873-2885.

113. **Bannwarth L, Reboud-Ravaux M.** 2007. An alternative strategy for inhibiting multidrug-resistant mutants of the dimeric HIV-1 protease by targeting the subunit interface. Biochem Soc Trans **35:**551-554.

114. **Wensing AM, van Maarseveen NM, Nijhuis M.** 2010. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. Antiviral Res **85:**59-74.

115. **la Porte CJ.** 2009. Saquinavir, the pioneer antiretroviral protease inhibitor. Expert Opin Drug Metab Toxicol **5:**1313-1322.

116. **McKeage K, Perry CM, Keam SJ.** 2009. Darunavir: a review of its use in the management of HIV infection in adults. Drugs **69:**477-503.

117. **Kuiper BD, Keusch BJ, Dewdney TG, Chordia P, Ross K, Brunzelle JS, Kovari IA, MacArthur R, Salimnia H, Kovari LC.** 2015. The L33F darunavir resistance mutation acts as a molecular anchor reducing the flexibility of the HIV-1 protease 30s and 80s loops. Biochem Biophys Rep **2:**160-165.

118. **Vickrey JF, Logsdon BC, Proteasa G, Palmer S, Winters MA, Merigan TC, Kovari LC.** 2003. HIV-1 protease variants from 100-fold drug resistant clinical isolates: expression, purification, and crystallization. Protein Expr Purif **28:**165-172.

119. **Weber IT, Kneller DW, Wong-Sam A.** 2015. Highly resistant HIV-1 proteases and strategies for their inhibition. Future Med Chem **7:**1023-1038.

120. **Agniswamy J, Louis JM, Roche J, Harrison RW, Weber IT.** 2016. Structural studies of a rationally selected multi-drug resistant HIV-1 protease reveal synergistic effect of distal mutations on flap dynamics. PLoS One **11:**e0168616.

121. **Louis JM, Deshmukh L, Sayer JM, Aniana A, Clore GM.** 2015. Mutations proximal to sites of autoproteolysis and the alpha-helix that co-evolve under drug pressure modulate the autoprocessing and vitality of HIV-1 protease. Biochemistry **54:**5414-5424.

122. **Cote HC, Brumme ZL, Harrigan PR.** 2001. Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir. J Virol **75:**589-594.

123. **Flynn WF, Chang MW, Tan Z, Oliveira G, Yuan J, Okulicz JF, Torbett BE, Levy RM.** 2015. Deep sequencing of protease inhibitor resistant HIV patient isolates reveals patterns of correlated mutations in Gag and protease. PLoS Comput Biol **11:**e1004249.

124. **Parry CM, Kolli M, Myers RE, Cane PA, Schiffer C, Pillay D.** 2011. Three residues in HIV-1 matrix contribute to protease inhibitor susceptibility and replication capacity. Antimicrob Agents Chemother **55:**1106-1113.

125. **Roche J, Louis JM, Bax A.** 2015. Conformation of inhibitor-free HIV-1 protease derived from NMR spectroscopy in a weakly oriented solution. Chembiochem **16:**214-218.

126. **Louis JM, Roche J.** 2016. Evolution under drug pressure remodels the folding free-energy landscape of mature HIV-1 protease. J Mol Biol **428:**2780-2792.

127. **Huang L, Li L, Tien C, LaBarbera DV, Chen C.** 2019. Targeting HIV-1 protease autoprocessing for high-throughput drug discovery and drug resistance assessment. Sci Rep **9:**301.

128. **Huang L, Li Y, Chen C.** 2011. Flexible catalytic site conformations implicated in modulation of HIV-1 protease autoprocessing reactions. Retrovirology **8:**79.

129. **Cherry E, Liang C, Rong L, Quan Y, Inouye P, Li X, Morin N, Kotler M, Wainberg MA.** 1998. Characterization of human immunodeficiency virus type-1 (HIV-1) particles that express protease-reverse transcriptase fusion proteins. J Mol Biol **284:**43-56.

130. **Tessmer U, Krausslich HG.** 1998. Cleavage of human immunodeficiency virus type 1 proteinase from the N-terminally adjacent p6* protein is essential for efficient Gag polyprotein processing and viral infectivity. J Virol **72:**3459-3463.

131. **Beissinger M, Paulus C, Bayer P, Wolf H, Rosch P, Wagner R.** 1996. Sequence-specific resonance assignments of the 1H-NMR spectra and structural characterization in solution of the HIV-1 transframe protein p6. Eur J Biochem **237:**383-392.

132. **Ishima R, Torchia DA, Lynch SM, Gronenborn AM, Louis JM.** 2003. Solution structure of the mature HIV-1 protease monomer: insight into the tertiary fold and stability of a precursor. J Biol Chem **278:**43311-43319.

133. **Huang L, Sayer JM, Swinford M, Louis JM, Chen C.** 2009. Modulation of human immunodeficiency virus type 1 protease autoprocessing by charge properties of surface residue 69. J Virol **83:**7789-7793.

134. **Huang L, Chen C.** 2010. Autoprocessing of human immunodeficiency virus type 1 protease miniprecursor fusions in mammalian cells. AIDS Res Ther **7:**27.

135. **Huang L, Hall A, Chen C.** 2010. Cysteine 95 and other residues influence the regulatory effects of Histidine 69 mutations on Human Immunodeficiency Virus Type 1 protease autoprocessing. Retrovirology **7:**24.

136. **Huang L, Li Y, Chen C.** 2011. Flexible catalytic site conformations implicated in modulation of HIV-1 protease autoprocessing reactions. Retrovirology **8:**79.

137. **Tien C, Huang L, Watanabe SM, Speidel JT, Carter CA, Chen C.** 2018. Context-dependent autoprocessing of human immunodeficiency virus type 1 protease precursors. PLoS One **13:**e0191372.

138. **Partin K, Zybarth G, Ehrlich L, DeCrombrugghe M, Wimmer E, Carter C.** 1991. Deletion of sequences upstream of the proteinase improves the proteolytic processing of human immunodeficiency virus type 1. Proc Natl Acad Sci U S A **88:**4776-4780.

139. **Pettit SC, Everitt LE, Choudhury S, Dunn BM, Kaplan AH.** 2004. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. J Virol **78:**8477-8485.

140. **Louis JM, Weber IT, Tozser J, Clore GM, Gronenborn AM.** 2000. HIV-1 protease: maturation, enzyme specificity, and drug resistance. Adv Pharmacol **49:**111-146.

141. **Louis JM, Clore GM, Gronenborn AM.** 1999. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. Nat Struct Biol **6:**868-875.

142. **Watanabe SM, Simon V, Durham ND, Kemp BR, Machihara S, Kemal KS, Shi B, Foley B, Li H, Chen BK, Weiser B, Burger H, Anastos K, Chen C, Carter CA.** 2016. The HIV-1 late domain-2 S40A polymorphism in antiretroviral (or ART)-exposed individuals influences protease inhibitor susceptibility. Retrovirology **13:**64.

143. **Counts CJ, Ho PS, Donlin MJ, Tavis JE, Chen C.** 2015. A functional interplay between human immunodeficiency virus type 1 protease residues 77 and 93 involved in differential regulation of precursor autoprocessing and mature protease activity. PLoS One **10:**e0123561.

144. **Pettit SC, Clemente JC, Jeung JA, Dunn BM, Kaplan AH.** 2005. Ordered processing of the human immunodeficiency virus type 1 GagPol precursor is influenced by the context of the embedded viral protease. J Virol **79:**10601-10607.

145. **Zybarth G, Carter C.** 1995. Domains upstream of the protease (PR) in human immunodeficiency virus type 1 Gag-Pol influence PR autoprocessing. J Virol **69:**3878.

146. **Partin K, Zybarth G, Ehrlich L, DeCrombrugghe M, Wimmer E, Carter C.** 1991. Deletion of sequences upstream of the proteinase improves the proteolytic processing of human immunodeficiency virus type 1. Proc Natl Acad Sci USA **88:**4776-4780.

147. **Huang L, Chen C.** 2010. Autoprocessing of human immunodeficiency virus type 1 protease miniprecursor fusions in mammalian cells. AIDS Res Ther **7:**27.

148. **Louis JM, Clore GM, Gronenborn AM.** 1999. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. NatStructBiol **6:**868.

149. **Louis JM, Aniana A, Weber IT, Sayer JM.** 2011. Inhibition of autoprocessing of natural variants and multidrug resistant mutant precursors of HIV-1 protease by clinical inhibitors. Proc Natl Acad Sci USA **108:**9072-9077.

150. **Park JH, Sayer JM, Aniana A, Yu X, Weber IT, Harrison RW, Louis JM.** 2016. Binding of clinical inhibitors to a model precursor of a rationally selected multidrug resistant HIV-1 Protease is significantly weaker than that to the released mature enzyme. Biochemistry **55:**2390-2400.

151. **Pettit SC, Everitt LE, Choudhury S, Dunn BM, Kaplan AH.** 2004. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. J Virol **78:**8477.

152. **Toohey K, Wehrly K, Nishio J, Perryman S, Chesebro B.** 1995. Human immunodeficiency virus envelope V1 and V2 regions influence replication efficiency in macrophages by affecting virus spread. Virology **213:**70-79.

153. **Wehrly K, Chesebro B.** 1997. p24 antigen capture assay for quantification of human immunodeficiency virus using readily available inexpensive reagents. Methods **12:**288-293.

154. **Sun P, Tropea JE, Waugh DS.** 2011. Enhancing the solubility of recombinant proteins in Escherichia coli by using hexahistidine-tagged maltose-binding protein as a fusion partner. Methods Mol Biol **705:**259-274.

155. **Raran-Kurussi S, Waugh DS.** 2012. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. PLoS One **7:**e49589.

156. **Waugh DS.** 2016. Crystal structures of MBP fusion proteins. Protein Sci **25:**559-571.

157. **Reuten R, Nikodemus D, Oliveira MB, Patel TR, Brachvogel B, Breloy I, Stetefeld J, Koch M.** 2016. Maltose-Binding Protein (MBP), a secretion-enhancing tag for mammalian protein expression systems. PLoS One **11:**e0152386.

158. **Fikes JD, Barkocy-Gallagher GA, Klapper DG, Bassford PJ, Jr.** 1990. Maturation of Escherichia coli maltose-binding protein by signal peptidase I in vivo. Sequence requirements for efficient processing and demonstration of an alternate cleavage site. J Biol Chem **265:**3417-3423.

159. **Fikes JD, Bankaitis VA, Ryan JP, Bassford PJ, Jr.** 1987. Mutational alterations affecting the export competence of a truncated but fully functional maltose-binding protein signal peptide. J Bacteriol **169:**2345-2351.

160. **Kemal KS, Burger H, Mayers D, Anastos K, Foley B, Kitchen C, Huggins P, Schroeder T, Picchio G, Back S, Gao W, Meyer WA, 3rd, Weiser B.** 2007. HIV-1 drug resistance in variants from the female genital tract and plasma. J Infect Dis **195:**535-545.

161. **Watanabe SM, Chen MH, Khan M, Ehrlich L, Kemal KS, Weiser B, Shi B, Chen C, Powell M, Anastos K, Burger H, Carter CA.** 2013. The S40 residue in HIV-1 Gag p6

impacts local and distal budding determinants, revealing additional late domain activities. Retrovirology **10:**143.

162. **Louis JM, McDonald RA, Nashed NT, Wondrak EM, Jerina DM, Oroszlan S, Mora PT.** 1991. Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in Escherichia coli. Eur J Biochem **199:**361-369.

163. **Agniswamy J, Shen CH, Aniana A, Sayer JM, Louis JM, Weber IT.** 2012. HIV-1 protease with 20 mutations exhibits extreme resistance to clinical inhibitors through coordinated structural rearrangements. Biochemistry **51:**2819-2828.

164. **Louis JM, Wondrak EM, Kimmel AR, Wingfield PT, Nashed NT.** 1999. Proteolytic processing of HIV-1 protease precursor, kinetics and mechanism. J Biol Chem **274:**23437-23442.

165. **Lee SK, Potempa M, Kolli M, Ozen A, Schiffer CA, Swanstrom R.** 2012. Context surrounding processing sites is crucial in determining cleavage rate of a subset of processing sites in HIV-1 Gag and Gag-Pro-Pol polyprotein precursors by viral protease. J Biol Chem **287:**13279-13290.

166. **Humpolickova J, Weber J, Starkova J, Masinova E, Gunterova J, Flaisigova I, Konvalinka J, Majerova T.** 2018. Inhibition of the precursor and mature forms of HIV-1 protease as a tool for drug evaluation. Sci Rep **8:**10438.

167. **Ishima R, Torchia DA, Louis JM.** 2007. Mutational and structural studies aimed at characterizing the monomer of HIV-1 protease and its precursor. J Biol Chem **282:**17190-17199.

168. **Zalucki YM, Jones CE, Ng PS, Schulz BL, Jennings MP.** 2010. Signal sequence non-optimal codons are required for the correct folding of mature maltose binding protein. Biochim Biophys Acta **1798:**1244-1249.

169. **Moon AF, Mueller GA, Zhong X, Pedersen LC.** 2010. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci **19:**901-913.

170. **Gohara DW, Ha CS, Kumar S, Ghosh B, Arnold JJ, Wisniewski TJ, Cameron CE.** 1999. Production of "authentic" poliovirus RNA-dependent RNA polymerase (3D(pol)) by ubiquitin-protease-mediated cleavage in Escherichia coli. Protein Expr Purif **17:**128-138.

171. **Telenti A, Martinez R, Munoz M, Bleiber G, Greub G, Sanglard D, Peters S.** 2002. Analysis of natural variants of the human immunodeficiency virus type 1 gag-pol frameshift stem-loop structure. **76:**7868-7873.

172. **Bukrinskaya AG.** 2004. HIV-1 assembly and maturation. ArchVirol **149:**1067-1082.

173. **Fabrini R, De Luca A, Stella L, Mei G, Orioni B, Ciccone S, Federici G, Lo Bello M, Ricci G.** 2009. Monomer-dimer equilibrium in glutathione transferases: a critical re-examination. Biochemistry **48:**10473-10482.

174. **Chen C, Sheng S, Shao Z, Guo P.** 2000. A dimer as a building block in assembling RNA. A hexamer that gears bacterial virus phi29 DNA-translocating machinery. J Biol Chem **275:**17510-17516.

175. **Tang C, Louis JM, Aniana A, Suh JY, Clore GM.** 2008. Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. Nature **455:**693-696.

176. **Alfadhli A, Mack A, Ritchie C, Cylinder I, Harper L, Tedbury PR, Freed EO, Barklis E.** 2016. Trimer enhancement mutation effects on HIV-1 matrix protein binding activities. J Virol **90:**5657-5664.

177. **Li H, Dou J, Ding L, Spearman P.** 2007. Myristoylation is required for human immunodeficiency virus type 1 Gag-Gag multimerization in mammalian cells. J Virol **81:**12899-12910.

LIST OF ABBREVIATIONS

AIDS	Acquired immunodeficiency syndrome
ALIX	ALG2-interacting protein X
APCs	Antigen-presenting cells
CA	Capsid
CCD	Catalytic core domain
CCR5	CC chemokine receptor 5
CDK9	Cyclin-dependent kinase 9
CRM1	Chromosome region maintenance 1
CTD	C-terminal domain
CXCR4	C-X-C motif chemokine receptor 4
DRV	Darunavir
eEF	Eukaryotic elongation factor
EIAV	Equine infectious anemia virus
Env (gp160)	Envelop
ESCRT	Endosomal sorting complexes required for transport
FDA	Food and drug administration
FSS	Frameshift stimulatory signal
GST	Glutathione S-transferase
HIV	Human Immunodeficiency Virus
HSP	Heat shock protein

IDV	Indinavir
IN	Integrase
IP	Immunoprecipitation
IRES	Internal ribosome entry site
ITAF	IRES *trans*-acting factors
LTR	Long terminal repeats
MA	Matrix
MBP	Maltose-binding protein
MDR	multiple-drug resistant
NC	Nucleocapsid
NES	Nuclear export signal
NPC	Nuclear pore complex
NTD	N-terminal domain
PBS	Primer binding sites
PIC	Pre-integration complex
PPT	Polypurine tract
PR	Protease
P-TEFb	Positive transcription elongation factor
R	Repeats
RNase H	Ribonuclease H
RRE	Rev-responsive element
RT	Reverse transcriptase
RTCs	Reverse transcription complexes

SB	Basic hydrophilic segment
SigP	Signal peptide
SHC	Hydrophobic core
SQV	Saquinavir
SU (gp 120)	Surface protein
TFP	Transframe peptide
TFR	Transframe region
TM (gp 41)	Transmembrane protein
TSG101	Tumor susceptibility gene 101
U3	3’ untranslated sequence
U5	5’ untranslated sequence
Ubp1	Ubiquitin-specific, carboxy-terminal protease
UTR	Untranslated regions
VLPs	Virus-like particles
WIHS	Women’s Interagency HIV Study