MECHANISMS REGULATING HIV-1 PROTEASE ACTIVITY

Marc Potempa

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Microbiology and Immunology in the School of Medicine.

Chapel Hill
2015

Approved by:

Ronald Swanstrom

Stanley M. Lemon

David Margolis

Kristina De Paris

Charlie Carter, Jr.

ABSTRACT

Marc Potempa: Mechanisms Regulating HIV-1 Protease Activity
(Under the direction of Ronald Swanstrom)

The Human Immunodeficiency Virus Type 1 (HIV-1) Protease (PR) has no direct involvement in the early steps of HIV-1 replication. Nonetheless, it is the timely and ordered processing of the viral structural proteins by the HIV-1 PR during virion maturation that facilitates the successful completion of virus entry, reverse transcription, and integration. Though a considerable amount of research has been devoted to deciphering how the enzyme prepares a virus particle for infection, the mechanisms regulating its activities continue to remain incompletely defined.

RNA serves as one putative regulatory factor, since efficient processing of the maturation intermediate p15NC requires RNA *in vitro*. Though previously believed relevant to only p15NC cleavage, I demonstrate that RNA enhances HIV-1 proteolysis reactions in a substrate-independent manner. The increased catalytic activity of the HIV-1 PR results from a direct interaction between RNA and the enzyme, with the magnitude of the effect dependent upon the size of the RNA molecule. Large (>400 base) RNAs accelerated proteolytic processing by over 100-fold under near-physiological conditions. This considerable change stemmed from both improved substrate recognition (K_m) and turnover rate (k_{cat}).

Variability in amino acid sequence also guides HIV-1 PR activity. However, the absence of any overt patterns across HIV-1 cleavage sites has complicated the delineation of why these differences result in diverse processing efficiencies. To address this question, I generated the largest-to-date dataset of globular proteins cleaved by the HIV-1 PR in near-physiological conditions. From these data, I unravel a number of site-specific processing requirements, and identify potentially important relationships shared between multiple cleavage sites. These results additionally enabled the formation of a preliminary conceptual model for explaining processing site amino acid composition.

Though the mind has a habit of blinding itself with all that is amiss, that which we need to appreciate does not always go overlooked. And now, as I complete this arduous-yet-worthwhile experience, I want to mention what I have come to appreciate above all else: the people surrounding me. So it is to you, my family, friends, colleagues, and mentors, that I dedicate this work and my success. Without you, my dreams are just wishes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ART	Antiretroviral Therapy
AVP	Human Adenovirus Proteinase
AZT	Zidovudine
BLAM	Beta-Lactamase
CA	Capsid
CCR5	C-C Chemokine Receptor 5
CD4	Cluster of Differentiation 4
CT	Cytoplasmic Tail
CXCR4	Chemokine (C-X-C motif) Receptor 4
DNA	Deoxyribonucleic Acid
DRV	Darunavir
dsDNA	Double-Stranded Deoxyribonucleic Acid
EC50	Effective concentration at which the response to a stimulus is half maximal
EC50$_{/\text{nt}}$	Effective concentration at which the response is half maximal per nucleotide
eEF1A/G	Eukaryotic Elongation Factor 1 A/G
EM	Electron Microscopy
Env	Envelope
ESCRT	Endosomal Sorting Complex Required for Transport
Gag	Group-specific Antigen
GMCΔ	GST-MA-CA$_{\Delta\text{CTD}}$
gp41	Glycoprotein 41
gp120	Glycoprotein 120

gp160	Glycoprotein 160
HIV	Human Immunodeficiency Virus
IC50	Concentration at 50% Inhibition
IFI16	Interferon, gamma-inducible protein 16
IN	Integrase
k_{cat}	Turnover number; a catalytic constant representative of the maximal rate at which an enzyme can convert an enzyme/substrate complex into enzyme and product
k_{cat}/K_m	Specificity constant or catalytic efficiency; a measure of the overall efficiency of an enzyme in converting substrate to product
K_d	Dissociation constant
kDa	Kilodaltons
K_i	Dissociation constant of an inhibitor
Km	Michaelis Constant; a ratio of the rate constants representing the breakdown of an enzyme-substrate complex and the formation of the complex, also a measure of the substrate concentration
k_{non}	Rate constant for an uncatalyzed hydrolysis reaction
LPV	Lopinavir
LTR	Long Terminal Repeat
M	Molar
MA	Matrix
MA/CA-AAA	Variant of the MA/CA construct where the $K_{26}KQYK_{30}$ sequence in MA has been converted to $A_{26}AQYA_{30}$ sequence of MA
μM	Micro-molar
mM	Milli-molar
NaCl	Sodium Chloride
NC	Nucleocapsid

ng	Nano-grams
nM	Nano-molar
nm	Nano-meter
NNRTI	Non-Nucleoside analog Reverse Transcriptase Inhibitor
NPC	Nuclear Pore Complex
NRTI	Nucleoside analog Reverse Transcriptase Inhibitor
$NS3^{Pro}$	Serine Protease Domain of the Hepatitis C Virus Non-structural 3 Protein
Nup153/358	Nucleoporin 153/358 kilodaltons
ORF	Open Reading Frame
PFV	Prototype Foamy Virus
PI	Protease Inhibitor
PIC	Pre-Integration Complex
pM	Pico-molar
poly(dN)	Homopolymeric deoxyribo(Adenine, Cytosine, Guanine, or Thymine)
poly(rN)	Homopolymeric ribo(Adenine, Cytosine, Guanine, or Thymine)
PR	Protease
RMSE	Root Mean Square Error
RNA	Ribonucleic Acid
RNase	Ribonuclease
RNP	Ribonuceloprotein
RT	Reverse Transcriptase
RTC	Reverse Transcription Complex
SDS	Sodium Dodecyl Sulfate

SP1	Spacer Peptide 1
SP2	Spacer Peptide 2
SQV	Saquinavir
ssDNA	Single Stranded Deoxyribonucleic Acid
TF	Transframe
tRNA	Transfer Ribonucleic Acid
V_{max}	Maximal velocity of an enzymatic reaction; the product of k_{cat} and the enzyme concentration
Vpr	Viral Protein R
VSV	Vesicular Stomatitis Virus

CHAPTER I

THE ROLE OF THE HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 PROTEASE IN VIRAL REPLICATION

A. Classification of the human immunodeficiency virus

Since its discovery in 1983 (1), the World Health Organization estimates more than 70 million people have been infected with the Human Immunodeficiency Virus (HIV) (2). HIV is an enveloped, positive-sense RNA virus belonging to the family *Retroviridae*. The hallmark of retroviruses is their ability to reverse transcribe their genome into DNA and then insert that DNA into the host cell's genome. Once embedded in a chromosome, the virus becomes a permanent part of that cell. While most retroviruses require an actively replicating cell for infection, HIV can productively infect resting cells (3, 4) giving it, and other retroviruses like it, the further classification of Lentivirus. HIV itself is subdivided into HIV-1 and HIV-2, and then also several different groups and subtypes. Group M of HIV-1 is the most common worldwide (5).

While Lentiviruses also encode a number of accessory proteins (6), a trio of open-reading frames (ORF) are common to all retroviruses: *gag*, *pro-pol*, and *env* (Figure 1.1). The *gag* ORF codes for the Gag polyprotein, the main structural unit of immature virus particles. The surface glycoprotein Envelope (Env), which mediates the attachment and fusion of virus particles to target cells, results from the transcription and translation of the *env* ORF. The remaining ORF, *pro-pol*, encodes the viral protease (PR), reverse transcriptase (RT), and integrase (IN) enzymes. The activities of RT and IN have already been prefaced. The former converts an RNA molecule into a double-stranded DNA copy, and the latter implants the viral DNA into the host cell's genome. As for the PR, its activities take place during the maturation step of the virus lifecycle. Maturation converts an immature, non-infectious virus particle that has just been created into one primed to complete the fusion, reverse transcription, and integration steps necessary for infection of a new cell.

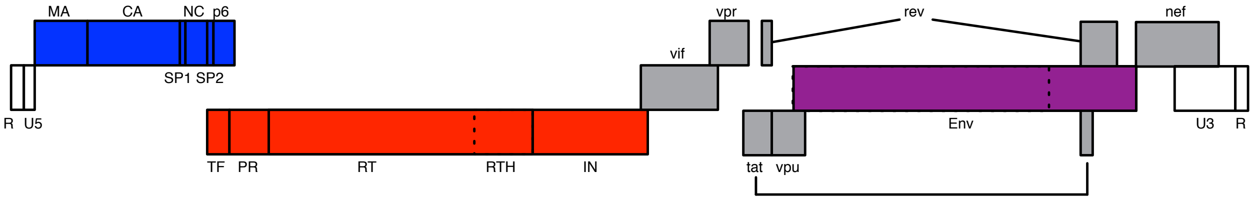

Figure 1.1: Organization of the HIV-1 genome. In blue, red, and purple are the three principal ORFs common to all retroviruses, *gag*, *pro-pol*, and *env*, respectively. The dashed line within *pro-pol* separates the polymerase (RT) and RNase H (RTH) domains of the RT coding region; the dashed line in *env* separates the gp120 and gp41 regions (labels not otherwise provided). In gray are the domains of the HIV-1 accessory proteins, while the uncolored regions are non-protein coding regions.

B. The triple threat of HIV-1 protease inhibitors[1]

1. Introduction

The HIV-1 PR is an indispensable enzyme, responsible for initiating the maturation of newly produced virus particles during the late stages of the HIV-1 replication cycle. The principal substrates for the PR are two HIV-1 polyproteins, Gag and Gag-Pro-Pol, with Gag representing most of the structural proteins of the virion and the Gag-Pro-Pol polyprotein including the viral enzymes used in replication. These proteins are translated from the same viral mRNA, and consequently share the same first 432 amino acids. This shared region contains the structural proteins Matrix (MA), Capsid (CA), and Nucleocapsid (NC), along with a 14-amino acid spacer peptide (SP1) set between CA and NC. While Gag is the predominant translation product, about 5-10% of the time a –1 ribosomal frameshifting event takes place to produce Gag-Pro-Pol instead of just Gag (7). So, while the C terminus of Gag includes a second 16-amino acid spacer peptide (SP2) and a functional domain involved in virus budding called p6, Gag-Pro-Pol instead contains the transframe (TF) region, and monomers of the PR, RT, and IN enzymes.

Gag and Gag-Pro-Pol drive the assembly of new virions through the reorganization of cholesterol-rich lipid raft microdomains on the plasma membrane (8, 9). Gag and Gag-Pro-Pol are targeted to the membrane via a myristate moiety postranslationally attached to their N termini (10-12). Dimers of genomic HIV-1 RNA, which are transported to the plasma membrane through interactions with Gag in the cytoplasm (12, 13), act as scaffolds to facilitate the higher-order multimerization interactions necessary for particle formation (14, 15). Multimerization initiates the budding process, but for efficient completion the host cellular Endosomal Sorting Complex Required for Transport (ESCRT) machinery is co-opted (reviewed in:(16-18)).

[1] This section of Chapter I previously appeared as an article in Current Topics in Microbiology and Immunology. The original citation is as follows: Potempa et al. “The Triple Threat of HIV-1 Protease Inhibitors”. *Current Topics in Microbiology and Immunology*. 2015, 389: 203-241.

Ultimately, released virions contain approximately 2400 Gag (19) and 120-240 Gag-Pro-Pol molecules (7). Immediately after or concomitant with virus budding (20), the HIV-1 PR activates as a result of Gag-Pro-Pol dimerization, and PR functions to convert newly formed virus particles into mature, infectious virions (reviewed in: (18, 21)). This maturation process entails a series of ordered, highly regulated cleavage events that liberate the functional domains from within the Gag and Gag-Pro-Pol polyproteins.

Though maturation is conventionally thought of as the last stage of the HIV-1 lifecycle, these depictions use the host cell as a frame of reference rather than the virus. Consequently, the multifaceted impact the HIV-1 PR exerts on the ability of a virus to successfully complete the so-called early steps in the HIV-1 lifecycle can go underappreciated. Antagonizing the HIV-1 PR can disrupt a number of early events including fusion (22-25), reverse transcription (25-30), and post-reverse transcription steps (25), i.e. nuclear import and/or integration. In this chapter, we review the triple threat of protease inhibitors (PIs): the intermolecular cooperativity that forms the basis of their cooperative dose-response in inhibition; the pleiotropic effects of HIV-1 PR inhibition on the early events of the replication cycle; and the potency associated with being a transition state analog and the considerable degree of improvement PIs can still undergo. Although many of the discoveries described within originally derived from work with other retroviruses, our review will focus on HIV-1. Accordingly, the provided references have been selected to highlight research performed with HIV-1.

2. Molecular mechanisms behind the antiviral activity of PIs

2.1 The HIV-1 PR, the most effective drug target among HIV-1 inhibitors

The HIV-1 PR is a member of the aspartyl proteinase family of enzymes. These enzymes are found as pseudodimers in eukaryotes, (due to an ancient gene duplication/fusion event), but are encoded as a monomer in the retroviral genome. For its activation, two PR monomers must interact to create the catalytic site at their dimerization interface (31-33). The active site formed by this interaction consists of a pair of aspartic acid residues, one from each monomer, and a water molecule to mediate the hydrolysis of peptide bonds (reviewed in: (18)). The initial activation of the HIV-1 PR occurs in the context of Gag-Pro-Pol. This embedded PR dimer is extremely unstable (34), and exhibits much lower enzymatic activity than fully released dimers (35, 36a). It appears that Gag-Pro-Pol active sites adopt the same conformation as the mature PR only a small fraction (3-5%) of the time (34), thereby limiting the embedded PR to intramolecular cleavage events. The first three cleavage events are all intramolecular, first at SP1/NC, then an internal TF site, and lastly the TF/PR site, and succeed in liberating the N termini of the PR monomers (36a, 37-39). These free ends fold into a four-stranded beta-sheet with other amino acids at the C terminus of the PR domain, conferring the stability and catalytic activity necessary for intermolecular cleavage events(40b). The subsequent proteolytic events that completely separate the enzyme monomers are intermolecular, and are performed by the mature PR (41, 42).

Since the N-terminally tethered PR does not function intermolecularly (39), processing of Gag polyproteins occurs subsequent to PR dimer maturation. Just like Gag-Pro-Pol, Gag cleavage follows a specific order of events (Figure 1.2). For simplicity, the five hydrolyzed peptide bonds may be separated into three groups based on their rate of cleavage: fast, medium, and slow. The SP1/NC site is the only member of the fast group. Cleavage at the MA/CA and SP2/p6 sites belong to the medium group (43-45), and these events happen ~10-fold more slowly than the fast cleavage event in an in vitro system using full-length Gag substrates (44). Lastly, the slow cleavages, CA/SP1 and NC/SP2, occur at rates ~400-fold slower than the fast site (43, 45, 46). The mechanisms that guide the PR through the proper sequence of intermolecular cleavage events are still not fully understood. There is only weak amino acid sequence similarity among the different cleavage sites (47), making complex interactions with the amino acid sequence a critical determinant (48, 49). Current models suggest the enzyme recognizes a conserved shape, called the substrate envelope, rather than particular amino acid sequences (48). Nonetheless, contextual cues also appear critical as exemplified by an ~12-fold increase in cleavage rate of the CA/SP1 site when placed into the MA/CA context (50). In any case, even partial disruption of HIV-1 PR activity results in disproportionately large effects on infectivity (26, 28-30, 51).

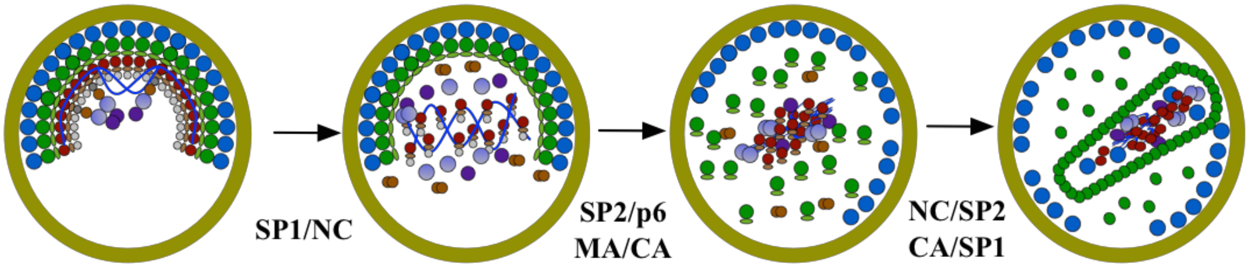

Figure 1.2: A model representation of the step-wise processing of HIV-1 Gag by the HIV-1 Protease. Gag, comprising MA (blue), CA (green), SP1 (light green), NC (red), SP2 (tan), and p6 (gray), is extended in a radial orientation from the membrane (gold), as is Gag-Pro-Pol, which contains the viral enzymes PR (brown), RT (blue-gray), and IN (purple). In the first of three stages, the SP1/NC site is cleaved to remove the NCp15 region comprised of NC/SP2/p6. The genomic RNA dimer increases in stability, but does not yet condense. In the second stage, two cleavage events occur at approximately the same rate. Proteolytic processing of the SP2/p6 releases the p6 domain from NCp9 and induces condensation of the RNA. Cleavage at the MA/CA site releases the CA/SP1 protein from the membrane, dissolving the immature CA lattice. In the final stage, spacer peptides are removed from NC and CA. After SP1 removal, CA forms a fullerene cone-shaped shell that surrounds the ribonucleoprotein core. The precise mechanism by which the CA cone forms (i.e. stochastic or nucleated) is still under investigation. This completed structure constitutes the pre-reverse transcription complex.

The vital role the HIV-1 PR plays in the replication cycle made it an extremely attractive drug target. After the first potent and bioavailable PI was introduced into triple drug regimens, it became apparent that it was possible to fully suppress viral replication and that this led to significant reductions in morbidity and mortality associated with HIV-1 infection (52, 53). These beneficial results further demonstrated the inherent dependency of HIV-1 on PR activity. The plethora of HIV-1 PR crystal structures (reviewed in: (54)) has facilitated the development of several extremely potent PIs by employing "structure-based drug design" (reviewed in: [55]). Currently, nine PIs are used in the treatment of HIV-1 infection. All of these, except for Tipranavir [56] are transition-state analogs that mimic a PR cleavage site, but replace the hydrolysable *P1-P1'* amide bond (Schechter and Berger nomenclature (55)) with a variety of non-hydrolyzable transition-state isosteres (56). PIs bind the wild-type PR enzyme with binding affinities in the nM to pM range (57-61). Comparatively, the binding affinity of the HIV-1 PR for its conventional substrates is in the µM to mM range (62), making PIs several orders of magnitude better interacting partners for the PR active site than their natural substrate. The tight binding of PIs and the need for multiple PR enzyme molecules in each virion to complete maturation (see below) give this class of enzymes distinctive properties among all classes of inhibitors of viral replication.

2.2 HIV-1 PIs display cooperative inhibition of their target enzyme

Once the intravirion space has been sealed off from the host cell cytoplasm, the virus particle must subsist on a limited set of packaged resources until after entry into the next target cell. Based on the estimated number of Gag-Pro-Pol molecules included during virion assembly, and because each enzyme functions as a multimer, the virus must complete maturation, reverse transcription, and integration with a maximum of 125 PR homodimers, 125 RT heterodimers, and 62 IN tetramers, respectively. Results from phenotypic mixing experiments have shown that both the PR (29, 63) and RT (64) enzymes are packaged into virions in excess, because viruses can tolerate some level of catalytically inactivated enzymes without substantial losses in infectivity. However, complete loss of infectivity occurs prior to inactivating 100% of the enzymes, which suggests that multiple copies of these enzymes are required to perform their associated life cycle step. Siliciano and colleagues confirmed this latter conclusion by demonstrating that PI and Non-Nucleoside Reverse Transcriptase Inhibitor (NNRTI) dose response curves for inhibition of infectivity display characteristics typical of cooperative binding reactions (65).

Conventionally, cooperative binding refers to the attachment of ligands to a multivalent receptor, where the attachment of a ligand to the receptor increases the affinity of the receptor for its ligands at other sites. However, although each PR and RT enzyme contains only a single binding site for their respective inhibitors, Shen et al. still found evidence of cooperativity (65). This can be explained by the microenvironment that is formed when a virus particle separates from the host cell. As mentioned above, each particle contains a specific number of enzymes that collectively can complete >100% of the enzymatic activities required for its associated step. If, for example, each PR enzyme contributes 5% to that total ability, and their abilities are, for example's sake, additive, then this theoretical virus would require at least 20 functional PR enzymes to complete maturation. Such enzymes can be thought of as intermolecularly cooperative. Thus, in support of the conclusion from the phenotypic mixing experiments, intermolecular cooperative action requires that multiple enzymes work together to complete a single activity as though they are one enzyme. This concept forms the basis of the "critical subset model" (66). In contrast, integration appears to be non-cooperative, likely requiring only a single catalytically active IN tetramer bound to the ends of the newly synthesized viral DNA to complete the integration step (65-67).

The importance of cooperativity becomes evident when considering the potency of antiretroviral drugs. Whenever the total number or functionality of an enzyme is reduced, the enzymes theoretically become more sensitive to inhibition. Each lost enzyme decreases the total catalytic potential present in the virion, moving the sum closer to falling below the critical threshold required for infectivity (66). Disrupting an infection would therefore require inhibiting one fewer enzyme, and a lower drug concentration, i.e. IC50. Such a prediction has been experimentally proven several times. Henderson et al. found that reducing the amount of functional HIV-1 PR in virions by phenotypic mixing or mutation-induced fitness losses generated an increased sensitivity to PIs (68). Similarly, lowering the amounts of RT by phenotypic mixing (69) or because of PR fitness losses (27, 68) increases RT sensitivity to NNRTIs (68, 69) and zidovudine[2] (AZT) (27, 68).

Unlike non-cooperative enzymes, when the content or the catalytic activity of cooperative enzymes is reduced, the critical threshold is approached more rapidly. In other words, slightly raising the concentration or effectiveness of a drug will result in disproportionately large increases in inhibition. The converse, that minor reductions in drug concentration or effectiveness will have nonlinear decreases in inhibition, would also be true. It was the latter

[2] Although NNRTI dose-response curves show RT functions as a cooperative enzyme, the other class of reverse transcriptase inhibitors, the Nucleoside Reverse Transcriptase Inhibitors (NRTIs), seemingly contradicts this finding (68). This discrepancy has been explained by considering the different targets for these inhibitor classes. NNRTIs seek out and interact specifically with the enzyme RT, whereas NRTIs actually target the elongating viral DNA molecule. Thus, just as integration requires only one enzyme tetramer to catalyze insertion of proviral DNA into the target cell genome, and is thus non-cooperative, only a single NRTI molecule needs to be incorporated into a growing DNA chain to terminate DNA elongation. Therefore, the effectiveness of NRTIs is independent of the number of RT molecules present. One exception does exist: AZT. When AZT is incorporated, it remains in the nucleotide-binding site because the large azido group sterically blocks its transfer to the primer site on RT (340). As a result, RT can excise AZT from its position in the nucleotide binding site using ATP (340, 341). It is this excision activity that appears to be dependent upon the concentration of RT (27, 68), and therefore AZT-mediated inhibition displays some degree of cooperativity.

prediction Sampah et al. demonstrated to support the model (67). When drug resistance mutations were introduced into viruses, and then challenged by the associated antiretroviral drug, differences were observed in the IC50 values for non-cooperative and cooperative enzymatic reactions alike. However, only cooperative enzymes showed changes to the slope of their dose-response curve, which is the mathematical descriptor for cooperativity. This change in slope reflected a much more severe reduction in inhibitory ability.

Siliciano and coworkers determined the theoretical slope value for most of the PIs currently in use, and found that the predicted values (66) had underestimated the actual values (65). The exceptionally high experimentally determined values underscore the superiority of PIs relative to other drugs at inhibiting HIV-1 replication (65, 70). But intriguingly, the slope for each PI varied considerably from the other drugs of the same class. Furthermore, the magnitudes of those differences are accentuated when compared to the intra-class fluctuation in other classes (65). In simplest terms, this result established that there are inherent differences in the maximum level of effectiveness each current HIV-1 PI can attain, and helps explain why certain PIs have been somewhat successful at monotherapy (71), whereas others fail more readily (72). But this PI-to-PI variability only encompasses one noteworthy detail about PI slopes. In addition, the slope values for PIs do not remain constant; as drug concentrations increase, so too do the slopes (70). In contrast, non-cooperative drugs such as NRTIs or IN inhibitors maintain a constant slope as their concentrations increase.

3. The pleiotropic effects of HIV-1 PIs

The implications of the constantly metamorphosing slope were not fully appreciated until the effects of PIs on individual stages of the lifecycle were determined. The changing slope results from the additive effect of interfering with multiple, distinct stages of the lifecycle (25). Considerable evidence exists to support this conclusion: disrupting maturation impairs fusion (22-25), reverse transcription (25, 27, 29, 30, 68), and post-reverse transcription steps (25). In other words, the considerable inhibitory capabilities offered by PIs results, in part, from its ability to perform like multiple drugs at once. Below, we discuss the various stages of the virus lifecycle PIs disturb, briefly reviewing the evidence and commenting on the potential mechanistic basis of the effect.

3.1 PIs antagonize fusion between the viral envelope and target cell membrane

3.1.1 The HIV-1 Env protein mediates fusion of the viral envelope and cellular membrane

The HIV-1 Env protein is translated as a polyprotein precursor, gp160, in the rough endoplasmic reticulum, where it is co-translationally glycosylated (reviewed in: (73)). Following translation, gp160 traffics to the trans-Golgi network, and is cleaved by cellular furin or furin-like proteases into the heterodimer gp120/gp41 (74). These proteins remain non-covalently attached, assemble into trimers of heterodimers with other gp120/gp41 molecules (75), and migrate to the plasma membrane. gp120 is entirely surface-expressed, and contains the CD4 and coreceptor binding sites. The cellular chemokine receptor CCR5 serves as the dominant coreceptor, enabling fusion with CD4+ T cells and macrophages. Later on in infection, Env evolves the ability to utilize the chemokine receptor CXCR4 as an alternative (reviewed in: (76)). gp41 has three distinct domains: the ectodomain that includes the fusion peptide (77), a single-pass transmembrane domain that keeps the Env assemblies tethered to the membrane, and an ~150 amino acid cytoplasmic tail (CT) that is present on the inner face of the viral envelope (78). Despite considerable clustering of Env around Gag assembly sites (79, 80), only about 10-

15 Env trimers get incorporated into each HIV-1 virion (81-83). Nonetheless, these low numbers are sufficient to mediate fusion between the HIV-1 viral membrane envelope and the target cell membrane.

Fusion requires a series of conformational changes and rearrangements in both the Env protein and in the lipid membrane (reviewed in: (84)). Once Env binds CD4, structural changes in gp120 expose the coreceptor binding site in tandem with changes in gp41 that results in the formation of the Pre-Hairpin Complex (85, 86). Presumably, these conformational changes position the gp41 fusion peptide within the target cell membrane (reviewed in: (84, 87)). As a result, a pore forms between the intravirion space and the target cell cytoplasm, although the size of the pore is too small for larger virus assemblies to cross the membrane. The secondary interactions between the Env-CD4 complex and CCR5 or CXCR4 lead to additional structural changes that cause gp120 to dissociate from gp41. Hydrophobic heptad repeat regions in the gp41 ectodomain coalesce into a coiled-coil structure called the six-helix bundle (88, 89), and it is this structure that pulls the viral and host cell membranes together, causing the fusion pore to expand.

3.1.2 "Inside-out" regulation: the HIV-1 PR regulates fusogenicity from within the virion

Comparing the fusogenicity, topology, and stiffness of immature and mature virions has revealed that internal processes affect activities that occur on the exterior side of the envelope. Mature and immature particles show an approximately 10-fold difference in ability to induce syncytia formation (22) or fuse with target cells (23, 24), thus implicating HIV-1 PR activity in conferring Env fusion competence. Since the gp41 CT occupies the intravirion space, this relationship in theory could result from proteolytic cleavage of the CT. However, though other retroviral CTs are truncated by their virus-associated PR (e.g. (90)), no such cleavage has been observed for HIV-1. Instead, a strong, detergent-stable linkage exists between gp120/gp41 and uncleaved Gag, and that relationship is lost in mature particles (91). Evidence that maturation can directly affect Env behavior recently came from direct imaging of mature and immature virus particles by super resolution microscopy. On particles produced in the presence of an inactivated PR, the 10-15 Env trimers were found separated at multiple distinct sites on the virion surface. In contrast, virus particles that have completed maturation appear to have only a single cluster of Env molecules on the surface (92). Interestingly, both Wyma et al. (91) and Chojnacki et al. (92) found that preventing cleavage of the MA/CA site was sufficient for maintaining the immature phenotype of Env molecules, consequently implicating the HIV-1 PR in manipulating Env behavior.

Other lines of evidence have also shown that inhibiting HIV-1 PR affects Env function. When a culture of chronically infected cells was treated with a high concentration of synthetic peptide analogue PIs, Meek and colleagues observed a considerable decrease in the formation of syncytia (93). More recently, use of the BLAM-Vpr assay showed a dose-dependent decrease in virus fusion with primary CD4+ T cells when viruses were produced in the presence of Atazanavir, Darunavir (DRV), or Lopinavir (LPV) (25). Moreover, the ability of viruses pseudotyped with the Vesicular Stomatitis Virus (VSV) G protein to enter cells was completely unaffected by administration of PIs. Collectively, these results demonstrated not just that PIs can inhibit fusion, but also that a specific relationship exists between the HIV-1 Env and the HIV-1 PR, although through an indirect mechanism. Of note, work by Krausslich and colleagues utilizing the same BLAM-Vpr assay contradicted these results. Virus produced in the presence of 2 μM LPV (compared to ~1.2 μM in (25)) did not affect fusion into MT-4 cells (30). However, these authors did report that completely inactivating the PR yielded findings consistent with the other results, and suggested that LPV failed to inhibit fusion because even trace amounts of PR activity might be sufficient for enabling virus entry. Alternatively, the disparity in results may indicate that the ability of HIV-1 PIs to restrict fusion could be a co-receptor-dependent effect (24), a cell type-dependent effect, or even a strain-specific effect. Rabi et al. pseudotyped viruses with samples derived from patients failing PI-based ART regimens, and found that, even in the context of a wild-type PR, Env proteins from 9 of 18 patients were able to confer statistically significant resistance to DRV (25). These results support the latter explanation, that the ability of PIs to disrupt Env fusogenicity may be Env-dependent.

3.1.3 Proteolytic activation of Env fusion competence results from the release of steric restrictions on the gp41 CT

Management of Env fusion activity by the HIV-1 PR is indirect, requiring the use of Gag as an intermediary. Though the cue that enhances Env's fusogenic ability comes from cleavage of the MA/CA site (91, 92), the interaction that directly manages the change in behavior must be between the membrane-associated MA domain of Gag and the CT of gp41. Indeed, ample evidence supports the existence of an interaction between MA and gp41. As previously mentioned, Env cosediments with immature virus cores following treatment with nonionic detergent that should separate membrane proteins from the core (91). Additionally, preferential clustering of Env with high-density lipid rafts requires Gag (79, 80, 94). And furthermore, deletions (95-98) or single amino acid substitutions in MA (97, 99, 100) are sufficient for excluding Env from budding virions. However, no currently published data have pinpointed the site of a direct interaction within either the gp41 CT or MA.

An alternative hypothesis that has gained traction suggests that no specific interaction has been identified because the interactions between MA and gp41 are actually steric (101). Matrix assembles into a hexamer of trimers on membranes enriched in cholesterol and phosphatidylinositol-(4,5)-bisphosphate (102). This arrangement forms gaps between the MA monomers that assemble into trimers, and larger gaps in the middle of the hexamers built from the trimers. Mapping a number of MA mutations that block Env incorporation into virions onto the crystallized mature MA trimer (Figure 1.3) reveals that the blocking mutations congregate near the potential hexameric gap region (101). The compensatory mutation Q62R, which can rescue the Env-incorporation defect exhibited by all four of these mutants, does not map to the same location, but instead represents an amino acid found at the trimer interface (101). Freed and coworkers note that the disparate location argues against Q62R replacing a lost contact, suggesting instead that the compensatory mutation could be adjusting the size and/or spacing of the hexameric pore by manipulating the interactions between trimers (101). In other words, the Q62R mutation compensates for the steric clashes by pinching the MA trimer interface closer together to create more space in the hole formed by the hexamers.

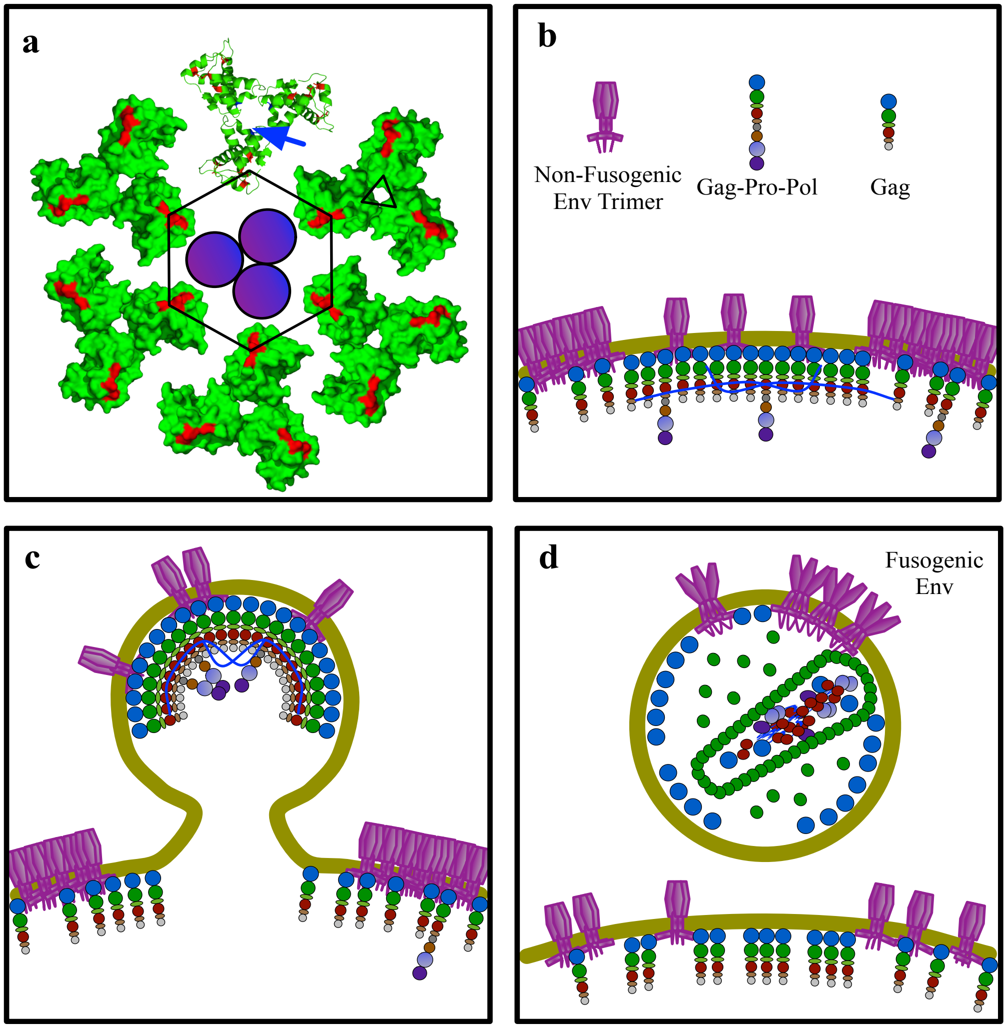

Figure 1.3 A model for the steric restriction and release of Env by Gag and PR. a. Top-down model of the hexamer of trimers that comprise the MA layer interacting with the CT of trimers of HIV-1 Env (purple circles). The mature trimeric form of MA (pdb: 1HIW; (103)) was used for the model since the structure of the immature MA lattice has not been determined at high resolution. Locations of mutations in MA that obstruct Env incorporation are shown in red. The compensatory Q62R mutation is not immediately visible from above. The blue arrow identifies its location near the center of the trimer interface. **b-d.** Model for the activation of Env proteins by proteolysis. Env clusters in high concentrations near sites of Gag assembly (color code consistent with Figure 1.2). The high concentrations effectively immobilize Env, trapping Env in a non-fusogenic conformation. Few Env molecules gain access into the assembly site due to their poor mobility and the limitations of the steric interactions. Those that are packaged are still in the non-fusogenic state due to the steric limitations provided through interactions with the MA layer. After cleavage, the gp41 CT is released, allosterically altering the structure of Env to a fusion-competent state, and potentially leading to the congregation of Env trimers to a single locus.

The CT has a variety of functional activities associated with it, including the allosteric modulation of gp120 conformation (104-107) and control over fusion peptide mobility (80, 108). In theory, by relying on a steric interaction instead of a specific one, the gp41 CT can maintain a moderate degree of sequence variability (109), which may be necessary for controlling the conformational states of a molecule that must undergo frequent change to escape immune pressure. Conceivably, HIV-1 could have evolved this indirect, and non-specific mechanism for regulating fusion competence because maintaining a specific amino acid sequence to serve as an additional HIV-1 processing site may have limited the conformational flexibility and variability in the surface-exposed regions of the Env heterodimer.

Figure 1.3 presents a model of HIV-1 PR-mediated regulation of Env fusogenicity Gag assembly sites manipulate the local membrane composition (9), which creates an environment favored by HIV-1 Env. Env clusters around the assemblies in high concentrations (79, 80), but at the cost of its mobility (110). The relatively immobilized gp41 CT exerts its allosteric control over gp120 (104-107), locking Env in a poorly fusogenic state (80). Owing to its limited mobility, lack of a specific interaction for recruitment (101), and the cramped steric interaction with MA trimers, Env packaging remains a very inefficient process, which accounts for the inclusion of a mere 10-15 Env trimers in the virion (81-83) despite high concentrations of Env around the Gag assembly. When trimers of the gp41 CT successfully interact with the MA lattice, the steric limitations trap Env in its restricted state. Upon activation of the HIV-1 PR, the Env CT is liberated from its trapped state in one of two potential ways: (1) proteolytic cleavage of the MA/CA site disconnects MA from the stable immature CA lattice, reducing the rigidity of the MA lattice, and consequently imparts surface mobility to Env; or (2), MA/CA cleavage could increase the steric clashes between gp41 CT and MA, forcing the Env trimers away from MA-rich regions. In either event, the small clusters seem to rearrange into a single cluster on the surface of the virion (92). And although clustering might have been inhibitory while associated with virion assembly sites (80), the low concentration of Env and conformational freedom granted to the liberated gp41 CT provides the flexibility to both gp120 and the gp41 fusion peptide necessary for fusion.

3.2 Multiple potential mechanisms by which PIs antagonize reverse transcription

3.2.1 RT: the heterodimeric polymerase

The HIV-1 virion contains two plus-strand RNA copies of the viral genome, which must be converted into a single, linear, double-stranded (ds)DNA product for integration into the target cell's genome. The virally encoded enzyme that catalyzes this reaction is RT, a heterodimer comprised of the proteins p66 and p51 (reviewed in: (111, 112)). The p66 subunit provides the enzymatic functions attributed to RT: an RNA-dependent DNA polymerase, a DNA-dependent DNA polymerase, and an RNase H ribonuclease activity. The p51 subunit contains the same first 440 amino acids as p66, but is truncated by the HIV-1 PR at position F440/Y441 to remove a majority of the RNase H domain. Despite the identical amino acid sequence, p51 assumes a distinctly different conformation in which the polymerase active site residues are buried within the protein. Instead of a catalytic function, p51 primarily provides structural support to the p66 subunit, and also contributes to substrate binding (111, 112).

Along with the other viral enzymes, RT originates as a monomer in the Gag-Pro-Pol polyprotein. During assembly and budding, activation of the HIV-1 PR results in the liberation of the full-length p66 RT molecule. Introducing either an L234A (38) or W401A (113) mutation into RT suppresses dimerization and the appearance of p51, suggesting that processing of p66 into p51 requires homodimerization of p66. Interestingly, this dimerization event is asymmetric, requiring one subunit to undergo considerable structural rearrangement (114). Purportedly, this conformational change unravels the RNase H domain, and consequently exposes the RT/RNase H (RT/RH) cleavage site. The mature heterodimer is extremely stable (115, 116), and so the remaining p66 subunit stays in a conformational state that protects it from cleavage within the RNase H domain by the PR (117-119).

Reverse transcription takes place within the aptly named reverse transcription complex (RTC). Though a functional RTC assembles before cellular entry, very little, if any, reverse transcription occurs in the virion, likely due to the absence of nucleotides (120). Until reverse transcription begins, the complex is referred to as the pre-RTC. Thus, the pre-RTC constitutes a ribonucleoprotein (RNP) core surrounded by a CA shell arranged in a fullerene cone structure (121). By electron microscopy (EM), this visualizes as an electron dense nucleoid surrounded by a thinner cone-shaped layer of mature CA (122). Aside from RT and its RNA template, the electron-dense core contains NC (123, 124), IN (123-126), Vpr (123-125), Vif (126), Nef (123, 127, 128), and MA (125, 129). Upon entry into the cell, the presumably active RTC interacts with the cytoskeleton to facilitate its movement toward the nucleus (129, 130). Viral DNA synthesis occurs en route, concomitantly inducing the dissociation of a majority of the CA shell (131-133), and a structural remodeling of the core into the pre-integration complex (PIC) (134-136).

3.2.2 The fragility of reverse transcription to anomalous HIV-1 PR activity

Assembly of the pre-RTC occurs during or immediately after the nascent particle buds away from the cell as part of the virion maturation process. Accordingly, the activity of the HIV-

1 PR and the ability of a virus to perform reverse transcription are intricately linked because the step-wise proteolytic processing of Gag results in pre-RTC formation. Numerous studies have demonstrated the extreme vulnerability of reverse transcription to defects in PR activity by measuring the effects of partially inhibiting the HIV-1 PR. Using sub-optimal amounts of PIs (26, 28, 30) or phenotypic mixing experiments that partially inhibit select processing sites in Gag (29, 30, 51, 137), these groups universally found that minor amounts of certain incompletely processed Gag intermediates had disproportionately large dominant negative effects on infectivity, and that the inhibitory effects frequently manifest somewhere during reverse transcription. For example, introducing a blocking mutation into just 20% of MA/CA cleavage sites completely ablated HIV-1 infectivity, and reduced the amount of early reverse transcription products by 90% compared to wild type (29). At some sub-optimal PI concentrations, and in all the phenotypic mixing experiments, these reverse transcription defects occurred when RT was fully functional and fully processed indicating the inhibitory effect lay in RTC formation.

Even so, the link between PR activity and reverse transcription extends beyond pre-RTC assembly. When PR activity is diminished because of PIs (20, 27, 30) or drug resistance mutation-associated fitness losses (27, 138, 139), a corresponding decrease in the amount and/or functionality of RT in the virion is observed. These differences have been attributed mostly to processing defects (26, 27, 30), but reduced RT packaging has also been suggested (139). Importantly, the amount of functional RT in the virion closely correlates with infectivity (140, 141) because viruses with reduced RT functionality are less efficient in completing reverse transcription (64, 139, 142). Furthermore, particles with reduced RT content or activity display increased sensitivities to NNRTIs (68, 69) and AZT (27, 68), in accordance with the critical subset model (66). Thus, a very complex interplay exists between reverse transcription and HIV-

1 PR activity. Below, we have separated a detailed discussion of several putative mechanisms by which PIs disrupt reverse transcription into two parts: decreased RT activity, and improper condensation of the RNP core.

3.2.3 Inhibiting PR activity decreases virion-associated RT activity

As noted above, nascent viral particles incorporate RT as part of the Gag-Pro-Pol polyprotein. In a gel-based activity assay (143) and in virions generated in the absence of a functional PR (144), the innate catalytic activity of RT while embedded in Gag-Pro-Pol was determined to be 20- to 25-fold less than the fully mature RT heterodimer. This provides one avenue for PIs to inhibit reverse transcription: by trapping RT in a precursor form.

Although some of the first cleavage events in Gag-Pro-Pol are up to 10,000-times more resistant to PIs (37, 39, 145, 146), current evidence suggests sequestering RT within Gag-Pro-Pol or another incompletely processed intermediate is possible. Two PIs, Saquinavir (SQV) and DRV, could effectively inhibit the intramolecular cleavage events in Gag-Pro-Pol in vitro, with IC50 values in the range of 1-2 μM (145, 146). Though these values increased to ~7 μM for DRV and ~10 μM for SQV in cell culture experiments (146), at least DRV has been detected in patient serum at concentrations near 10 μM (147). DRV is the only antiretroviral drug with equivalent potency to a three-drug regimen (71), and its prospective ability to inhibit the intramolecular Gag-Pro-Pol processing events may contribute to its exceptional potency. Furthermore, nM sensitivity to PIs is restored concomitant with the appearance of mature PR functionality (39, 145), and both PR/RT (41) and RT/IN (42) are thought to be targets of the mature PR (Figure 1.4). In support of this, high molecular weight bands of 113 and 107 kilodaltons (kDa) are often observed among cleavage products from virions produced in PI-treated cells (26, 37, 146). These protein species correspond to the PR/RT/IN intermediate with (113 kDa) or without (107 kDa) the abridged TF domain. Since p66 monomers must undergo substantial structural rearrangement to dimerize, particularly within the C-terminal domains of RT (114), there is a strong possibility that the RT/IN linkage accounts for a majority of the restriction on RT function by sterically preventing the structural rearrangements necessary for dimerization. Therefore, the intramolecular cleavage events that are highly resistant to currently available PIs are nonetheless likely to be insufficient for generating fully functional RT enzymes. Unfortunately, the polymerase activities of PR/RT/IN or RT/IN intermediates have not been reported to validate this inference.

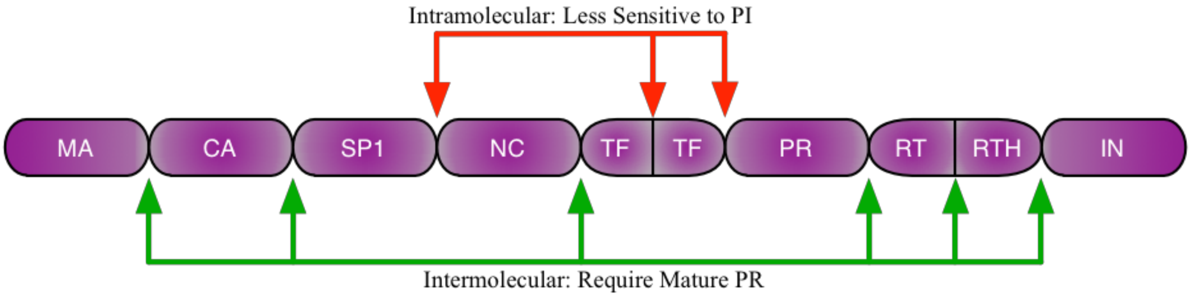
Intramolecular: Less Sensitive to PI
MA
CA
SP1
NC
TF
TF
PR
RT
RTH
IN
Intermolecular: Require Mature PR

Figure 1.4: Locations of intra- and intermolecular cleavage sites targeted by the HIV-1 PR. Diagram of the Gag-Pro-Pol polyprotein, subunits not drawn to scale. Each red arrow identifies a processing site cleaved by the intramolecular, embedded HIV-1 PR. Some PIs, such as DRV and SQV, may be capable of inhibiting these events. However, on the whole, these sites are much less sensitive to PI effects. The green arrows identify target sites for the mature HIV-1 PR. Effective inhibition of HIV-1 PR ability to cleave these sites may occur at nM concentrations of PIs in cell culture.

In addition to p66 excision, cleavage at the RT/RH site also presents an opportunity for HIV-1 PR activity to regulate RT functionality. Sluis-Cremer et al. determined that inhibiting cleavage at the RT/RTH site required lower concentrations of the PI ritonavir (38). Thus, even if PI concentrations are too low to inhibit RT's removal from Gag-Pro-Pol, RT could conceivably be trapped in a p66 homodimer by blocking RT/RTH processing. Both gel-based (148) and in vitro (149) activity assays determined that the homodimeric p66 RT molecule was five-fold worse than p66/p51 heterodimers at catalyzing DNA synthesis. Slowing or reducing viral DNA synthesis could increase exposure of the RTC to cellular restriction factors such as the DNA sensor IFI16 (150). Thus, formation of the fully functional RT heterodimer requires multiple proteolytic cleavage events, each of which is susceptible to PIs. These data strongly argue that PIs can reduce the content or functionality of RT, and therefore the ability to complete reverse transcription, through their inhibition of HIV-1 PR processing activity.

3.2.4 Incompletely processed Gag molecules are dominant negative inhibitors to reverse transcription

Sequential proteolytic processing of the Gag polyproteins controls assembly of the pre-RTC. As briefly described earlier, proteolytic processing of Gag proceeds in a defined order: SP1/NC > SP2/p6 ~ MA/CA > NC/SP2 ~ CA/SP1 (Figure 1.2) (44, 45). These cleavage events generate specific intermediates, each of which has distinct functional abilities. The temporal appearance of each of these intermediates and their associated tasks are critical to the proper assembly of the pre-RTC.

The initial cleavage event between SP1 and NC yields a membrane-bound MA/CA/SP1 intermediate and the NC/SP2/p6 intermediate called p15NC. Completing cleavage of the p15NC intermediate is absolutely essential for maturation of the dimeric RNA genome (151). Though ~10-fold slower than the initial cleavage, the second and third cleavage events occur at approximately the same rate, at least in vitro (44). Cleavage of the MA/CA site dissolves the immature CA lattice (152), and allows the N terminus of CA to form a salt bridge that is essential for the eventual construction of the CA cone (153-155). Most MA remains bound to the membrane, however a small amount relocates into the virus core (125, 129). Meanwhile, removal of the p6 domain from p15NC yields the 71-amino acid NCp9. This protein has many of the same capabilities as the fully mature 55-amino acid NCp7 (reviewed in: (156), but most importantly, potently induces nucleic acid aggregation (134, 157). Thus, upon cleavage of p15NC to NCp9, the RNP core condenses into the iconic electron dense structure found in mature HIV-1 virions. The ultimate fate of p6 remains unknown, as it is not found in the pre-RTC (123). Given that p6 antagonizes condensation while part of p15NC, its exclusion from the pre-RTC is likely necessary.

The remaining cleavage events remove spacer peptides from CA and NC. The mature CA lattice cannot form without CA/SP1 processing (152), and therefore assembly of the CA cone depends upon completion of this step. NC/SP2 processing produces the optimal NC chaperone, NCp7. The strand destabilization activity of NCp7 is similar to NCp9, but its aggregative abilities are inferior to NCp9 (134, 157, 158a). Additionally, NCp7's on/off binding kinetics are very fast compared to the other NC species (158a, 159b). The faster kinetics ensures that NC does not become a physical roadblock that stops RT from sliding along its template (160). For a detailed review on the differences between NC species, see (156).

Partially processed Gag intermediates interfere with proper core assembly. Under ideal cell culture conditions and normal PR activity, EM imaging reveals that the pre-RTC correctly assembles up to 90% of the time (28, 29, 51, 161). The remaining virions display a mostly immature, but occasionally aberrant morphology. However, retention of minor amounts of processing intermediates strongly shifts the distribution toward aberrant and immature assemblages (28-30, 51). For instance, at the IC50 for a PI, 40-50% of particles display an aberrant or immature phenotype (26, 28, 30). However, less than a 10% reduction in CA/SP1 processing in cell-free virions was observed when compared with untreated controls (30). These results suggested small amounts of processing intermediates could act as strong dominant negative inhibitors.

Our laboratory and others (30, 51) explored the relationship between processing at each site in Gag and its dominant negative effect. Progressively interfering with some percentage of processing at basically any Gag cleavage site will eventually eliminate infectivity (29, 30, 51, 137), a phenomenon first reported for Murine Leukemia Virus (162). The sole possible exception is NC/SP2 (see below) (29, 30, 137, 161). The two sites most sensitive to interference are the MA/CA site (29, 30), and the CA/SP1 site (30, 51), suggesting CA intermediates have the strongest dominant negative effect. These results are somewhat surprising since only ~1500 CA molecules need to be fully cleaved for assembly of the CA shell (163) – a number that is only about 60% of the total estimated amount in the virion (19). Considerably higher sensitivity might be expected of NC, since closer to 85% of NCs likely participate in the reverse transcription process (164). Nonetheless, both CA/SP1 and MA/CA/SP1 species are observed in virions collected after sub-optimal PI treatment (26, 28, 30), suggesting the dominant negative effects are valid.

Visualization of virions partially defective for MA/CA (29) or CA/SP1 (51) cleavage found the predominant assembly defects manifested slightly differently for each of these cleavage sites, although the functional block is the same (Figure 1.5). For incomplete MA/CA cleavage, an electron dense sphere consistently appeared immediately adjacent to the virion envelope, which implies at least one CA molecule still tethered to the membrane via MA was included in an assembling cone (29). For incomplete CA/SP1 cleavage, a condensed RNP core also formed, although the majority appeared dissociated from the virus envelope (51). The authors did not comment on the thickness of the envelope, leaving open the question of whether the CA lattice disassembled when various amounts of CA/SP1 cleavage were inhibited. However, as we previously noted, immature lattice disassembly has been reported to occur following cleavage of the MA/CA site (152), suggesting the defect brought on by the CA/SP1 mutants is a failure to assemble the mature lattice. This is in contrast to the maturation inhibitor Bevirimat, which delays CA/SP1 cleavage, but does so by stabilizing the immature CA array (165, 166).

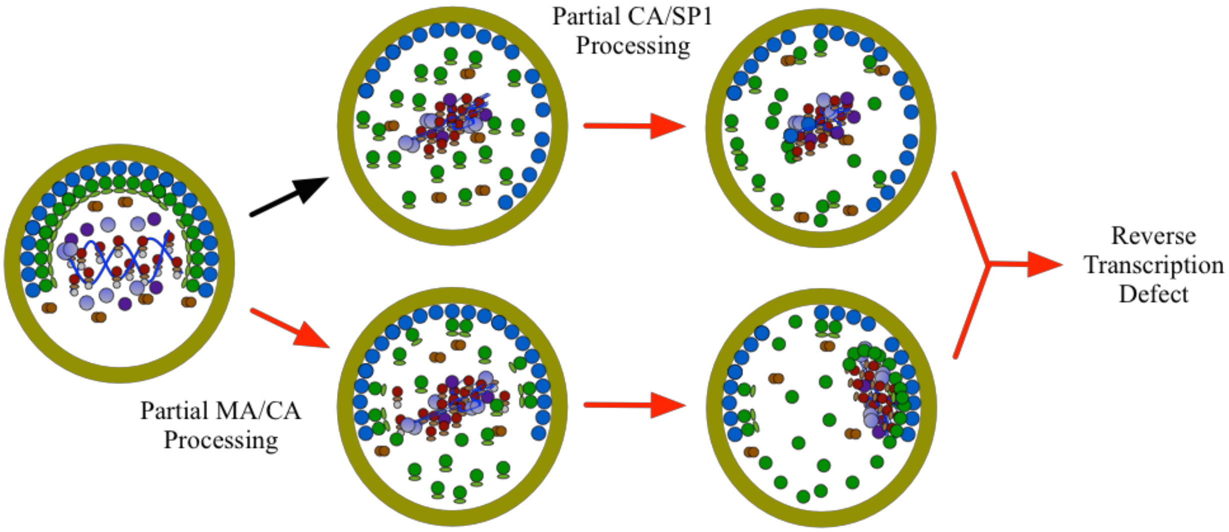

Figure 1.5: Schematic representation of the most common effects of CA/SP1 and MA/CA on virion morphology. For CA/SP1 (top), the core forms normally, but no CA shell encompasses the condensed RNP complex. For MA/CA (bottom), the core forms normally, but generally locates immediately adjacent to the membrane. Such placement likely indicates the CA shell has incorporated a molecule still attached to the membrane because of failed MA/CA cleavage. In both cases, reverse transcription defects are the likely outcome. Gag and Gag-Pro-Pol color scheme is consistent with Figure 1.2.

Both of the dominant negative derivatives of CA prevent the assembly of the fullerene cone. The requirement for a nearly complete, conical CA shell during reverse transcription is still somewhat obscure. CA does not directly interact with nucleic acid, and no reports exist to suggest CA enhances RT activity in vitro. Furthermore, a considerable amount of the CA shell may disassemble soon after cellular entry (128, 167). Nonetheless, HIV-1 replication is severely attenuated and defective for reverse transcription when cones cannot form (155), dissolve too quickly (133, 168-170), or fail to dissociate (170). Several possible reasons for its critical importance have been hypothesized (171): for one, though part of the shell is lost, the remnants may shield the RTC from nucleases or host restriction factors; it may protect the viral reverse transcription products from cytoplasmic innate immune receptors; and potentially, CA may recruit host cellular proteins necessary for the completion of reverse transcription.

The curious case of NC/SP2. Four publications reported that cleavage of the NC/SP2 site is non-essential for infectivity in single round assays (29, 30, 137, 161). Each of these conclusions stemmed from versions of the NL4-3 virus isolate defective only in NC/SP2 processing. Paradoxically, one of these authors reported that partially inhibiting HIV-1 PR activity by sub-optimal concentrations of PI was most effective at inhibiting NC/SP2 processing, and that the amount of NC/SP2 inhibition closely correlated with the total loss in infectivity (30). Furthermore, one group reported that after just four passages, the mutant NC/SP2 cleavage site had reverted back to wild type (137). Another had found that, while apparently non-essential, inhibiting NC/SP2 did reduce virus fitness (29). And lastly, a fifth report contradicts these results with the conclusion that NC/SP2 cleavage does affect viability (172). The principle difference in experimental design from that study was the use of the BH10, instead of NL4-3 like the rest, though they also used a different cell type to measure infectivity.

More indicative of the important role NC/SP2 cleavage plays, mutations in NC/SP2 are frequently observed in vivo as compensatory mutations in PI-resistant viruses (173-175). For example, in the presence of patient-derived PI-resistant viruses, the NC/SP2 mutations A431V and I437V were each individually capable of conferring a fitness advantage to the virus in the presence of PIs, and were just as effective as when the entirety of Gag was supplied (175). These results suggested that NC/SP2 resistance mutations carry the resistance impact for non-PR compensatory mutations, though this conclusion may only be true for particular resistance pathways in the HIV-1 PR (176).

From the available data, we suggest two plausible explanations for these discrepancies: (1) NC/SP2 processing is not essential for infectivity, but specifically for the NL4-3 clone; or (2) NC/SP2 processing is not essential, however, it makes the virus partially defective. We prefer the second option, since if NC/SP2 cleavage was non-essential for NL4-3 only, we would not expect to find the reversion to wild type. The differences in the literature are likely explained, at least in part, by variances in the amount of cleavage required for replication in the different assay conditions. However, the NC/SP2 site overlaps the ribosomal slippery sequence. Thus, it is possible that the rapid reversion occurred because of RNA secondary structure requirements. The fitness advantage of reducing NCp9 to the NCp7 could derive from the faster on/off binding kinetics of NCp7 (158a). NCp9 binds cooperatively, which could interfere with its ability to negotiate the strand transfer reactions during reverse transcription that require rapid nucleic acid rearrangements (159b). Additionally, the slower dissociation of NCp9 from template strands could result in increased pausing and/or dissociation of the RT elongation complex.

3.3 Nuclear import and integration: a far reach

At the completion of reverse transcription, IN molecules bind the LTR regions and engage with each other to form a tetramer (reviewed in: (177). IN proceeds to cleave the DNA ends, in the first of two enzymatic reactions for which it is responsible, at a specific dinucleotide sequence near the end of each long terminal repeat (LTR) region. Generation of the new under-hanging 3'-OH groups represents the final conversion of the RTC into the PIC, the integration-competent nucleoprotein complex containing a complete copy of viral DNA. Before IN performs its second enzymatic function, in which it utilizes the hydroxyl groups created in the first reaction to simultaneously break the phosphodiester bonds in the host DNA and insert the viral DNA, the PIC must first enter the nucleus.

Distinct from other genera of retroviruses, lentiviruses like HIV-1 have evolved mechanisms to infect non-dividing cells (3, 4). This necessitates crossing the nuclear envelope through nuclear pore complexes (NPCs). Molecules that are approximately 10 nm in diameter or more cannot passively diffuse through the NPC (178), and those 40 nm or greater in diameter cannot get through at all without disassembling (179). This seemingly poses a considerable problem for HIV-1, whose PIC has an estimated diameter of 56 nm (180). Although this is still too large to fit through the NPC, it has already shrunk in size from 400 nm in length, and 100 nm width during reverse transcription (129). Therefore, further remodeling to transverse the NPC likely takes place. Vpr has been identified as potentially mediating this process (136), since it has a nucleic-acid binding ability (181, 182) and can bind and fold dsDNA (136). Thus, management of Vpr-nucleic acid interactions might organize the genome for nuclear entry.

Although genetic evidence implicates both CA (183-185) and NC (186, 187) in nuclear import and integration, and MA has frequently been found within the PIC (125, 180, 188-190), the likelihood that interrupting HIV-1 PR activity by PI would have an effect on these later steps without first affecting the completion of reverse transcription seems low. Even the obvious effect of the trapping IN within the Gag-Pro-Pol precursor is more likely to manifest as a reverse transcription block. Viruses generated in the absence of IN or with select mutations fail to complete reverse transcription (191). IN appears to be essential for encapsidation of the electron-dense RNP core by the CA cone (192, 193), and also interacts with the host cellular factors eEF1A and eEF1G, important components of the elongation complex (194). Nonetheless, Rabi et al. mathematically determined that PIs exert some level of inhibition on post-reverse transcription steps, independent of their interference of both entry and reverse transcription (25). Thus, though such an effect seems unlikely, a relationship between PIs and post-reverse transcription functions may occur.

Delayed condensation of the virion core potentially interferes with nuclear import, and not reverse transcription. Nuclear transport of HIV-1 PICs is an active process (3). In quick succession, four different genome-wide RNAi screens identified putative host cellular factors important for HIV-1 replication (195-197), a list that included a considerable number of nuclear transport factors. Several of these putative interacting partners for viral components of the RTC and/or PIC have already been confirmed (reviewed in: (198)). As a consequence of this growing list, the importance of the HIV-1 CA in nuclear entry has come to the forefront. CA was completely absent from early studies identifying the components of the PIC (188-190, 199), leading to the conclusion that the CA shell completely dissociated before the RTC/PIC reached the nuclear pore. However, later studies demonstrated CA as a critical factor in enabling HIV-1 to infect non-dividing cells (183, 200). Concomitantly, evidence accumulated that the CA shell underwent a biphasic process of disassembly (128, 167), and complete uncoating might not occur until at the NPC (201). Now, CA is known to directly interact with several host proteins involved in nuclear entry, including Nup153 (202), Nup358 (203), and Transportin-3 (204).

With the clear importance of CA as an interacting partner of nucleoporins, one attractive mechanism of interference would be causing the misassembly of the fullerene cone. However, we have already discussed the dominant negative effects of Gag processing intermediates on fullerene cone assembly, and they interfere with reverse transcription. While we cannot rule out the possibility that PIs could affect CA assembly in such a way as to allow reverse transcription to occur, but then compromise its ability to facilitate nuclear import, this scenario is at present without support. Alternative possibilities reside with NC processing intermediates. Mutations of the C-terminal domain of Gag that block processing at the SP2/p6 site still allow reverse transcription to occur, albeit at a slightly reduced efficiency (137). These mutants were still considerably less infectious than wild type, even with detection of late reverse transcripts (137, 161). Where 2-LTR circles were quantified, the authors found only one-third as many as in wild-type infections (137). Though there could be alternative explanations, such as decreased stability of the 2-LTR circles, these results supported a defect in nuclear entry.

As for the mechanistic basis of this effect, the p15NC intermediate is incapable of condensing the RNP core because of the presence of p6 (134, 157). This does not frequently prevent formation of a CA shell, although it no longer takes on the conventional fullerene cone shape (161) (Figure 1.6). Instead, maturation of the genomic RNA dimer goes unfinished (151, 172, 205). Even if the NC/SP2 site is processed to release NCp7, the timing of the reaction could ostensibly result in aberrant CA cone formation, or the exclusion of non-structural proteins from the RNP core. Furthermore, since no specific RNA structure is required for the initiation of reverse transcription (205), and p15NC ably interacts with the tRNA^{Lys,3} primer and RT (206), this particular processing intermediate might not interfere with reverse transcription initiation. If the mutant core can still facilitate strand transfer reactions, which becomes more likely if NCp7 is eventually released, then reverse transcription could be completed. However, a defect during the conversion of the RTC into the PIC would halt the infection.

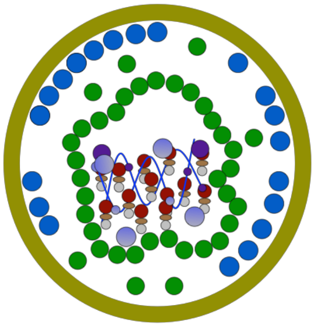

Figure 1.6: Morphology of SP2/p6 processing defect. Schematic representation of the most common morphology found when SP2/p6 cleavage does not occur. Colors consistent with Figure 1.2.

Despite this mechanistic possibility, this block has only been demonstrated when SP2/p6 processing has been artificially prevented by mutagenesis of the cleavage site. Temporally, SP2/p6 is cleaved at approximately the same time as MA/CA cleavage, if not a little ahead (44). Thus, if there is a sufficient concentration of PIs to block SP2/p6 processing, it is highly likely that the virus particle would have the added problem of uncleaved MA/CA. However, selective errors in SP2/p6 processing might become more likely upon a loss of HIV-1 PR fitness during drug resistance selection. In addition to NC/SP2, the SP2/p6 site is one of the most frequently observed locations of compensatory mutations (173, 174, 207). Molecular modeling predicts SP2/p6 protrudes beyond the substrate envelope, and MA/CA much less so (49). Thus, SP2/p6 could be preferentially more sensitive than MA/CA to drug resistance mutations in the PR. Effectively, the aberrant timing of SP2/p6 cleavage could enable the CA cone to assemble before condensation of the core, producing a reverse transcription-competent, but nuclear import-defective virus. Thus, PIs have the potential to exert at least some influence on steps post-reverse transcription, though such an effect may not be immediately apparent.

4. The theoretical potential of PIs that is unique among all inhibitor targets

Beyond the many pleiotropic effects achieved by inhibiting the PR, targeting the active site of the PR itself has an intrinsic advantage in terms of inhibitor binding potential. The theory of transition state affinity, simply put, states that, in order to enhance the rate of a reaction, the affinity of an enzyme for its substrate must increase while changing from the ground state to the transition state by a factor that matches or surpasses the factor by which the enzyme enhances the rate of reaction. Later, the enzyme’s grip relaxes as the products are formed and released. In this way, the enzyme lowers the free energy of activation for the reaction (208). By extension, the inhibitors that bind most tightly are mimetics of the transition state structure. None of the strategies to develop inhibitors of HIV-1 use this most fundamental of inhibitor designs, except for the PIs. These inhibitors invariably contain a hydroxyl group that aligns with the two aspartic acid residues at the active site. The hydroxyl group displaces a water molecule ordinarily used by the aspartic acids to catalyze the hydrolysis of the peptide bond. Thus, the hydroxyl group helps the inhibitor mimic the transition state as it would be shaped while adding water to the peptide bond. For this reason, HIV-1 PIs stand alone in their potency, reaching K_i values that are actually difficult to measure and lie in the low picomolar range (Figure 1.7).

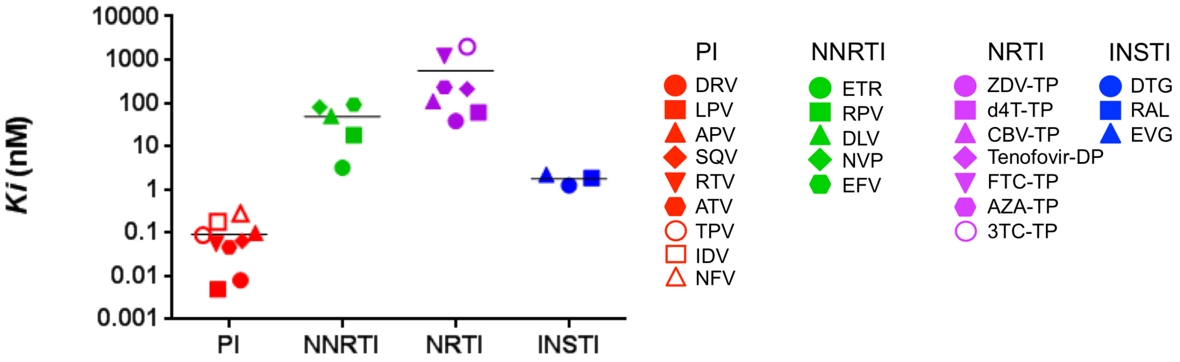

Figure 1.7: Comparison of the inhibitory constants for each inhibitor from four of the antiretroviral drug classes: Protease Inhibitors (PI), Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTI), Nucleoside Reverse Transcriptase Inhibitors (NRTI), and Integrase Strand Transfer Inhibitors (INSTI). All PI values are cited from (209). For NNRTIs, RPV from (210), NVP from (211), EFV from (212), and ETR and DLV from Dr. Nicolas Sluis-Cremer (personal communication). All NRTI values are cited from (213). All INSTI values are cited from (214).

We should not be surprised at the potencies of these PIs, since such high affinities are precisely what is predicted by transition state analog theory. This concept of rate enhancement by inducing the transition state can be made more clearly by comparing the rates of reactions in the presence and absence of enzymes. For this comparison, since reported rate constants for the HIV-1 PR vary dramatically based on the substrate and the reaction conditions, we will use an average value under conditions where the enzyme is especially active, while acknowledging we do not know how this level of catalytic activity compares to the activity of the enzyme during virion maturation. A typical reported rate constant (k_{cat}) for peptide cleavage by the HIV-1 PR is 20 sec^{-1} (215) whereas the rate constant (k_{non}) for uncatalyzed hydrolysis of a model peptide (the glycine-glycine bond of acetylglycylglycine N-methylamide) is 3.6×10^{-11} second^{-1} (216). Thus, the PR enhances the rate of peptide hydrolysis (k_{cat}/k_{non}) by a factor of about 5×10^{11}-fold. The value of K_m for peptide substrates can be as low as 10 μM (215, 217). Therefore, this enzyme's formal affinity for the substrate in the transition state can be described by a dissociation constant equal to the substrate's K_m value divided by k_{cat}/k_{non}, or $2 \text{ X } 10^{-17}$ M. This value argues that even the best current pM inhibitors (e. g. DRV, K_i 10^{-12} M) bind 10^5 fold less tightly than the actual substrate in the transition state. Accordingly, PIs still have a considerable amount of room for improvement. It is certainly true that there is a big difference between good inhibitors and good drugs, and that many good inhibitors never become good drugs. But it is also true that good drugs all started as good inhibitors, so that new and/or improved drugs will have to be built on new concepts in inhibitor design.

There is an important corollary in considering the implications of such tight binding inhibitors. Chemists have a very large array of structures to query in reaching an optimal inhibitor design. In contrast, the virus is limited to the 20 amino acid structures and further limited by the need to maintain binding and function on the normal substrates. Thus, as binding of the inhibitor becomes increasingly tight, the impact of one or two amino acid changes (of the limited choices available) may not allow enough viral replication to occur for further evolution to high levels of resistance. In this case, a very tight binding PI would behave equivalently to highly successful combination therapy, where viral suppression is achieved before resistance appears to drugs that are otherwise easily circumvented. The possibility for very tight binding to be an effective strategy of limiting evolution is most easily proposed for PIs, should they become true transition state analogs.

5. Conclusions and future perspectives

PIs offer two unique and important features in targeting the PR to block viral replication: (i) inhibition of the PR has pleiotropic effects on multiple steps in the viral life cycle, and (ii) PIs are based on transition state analog design which has intrinsically high binding potential. Executing the proteolytic processing pathway during virion maturation requires multiple PR molecules, and thus inhibition has a significant cooperative effect; as the amount of active or functional PR decreases, the virus becomes increasingly sensitive to increasing inhibitor concentrations. Finally, if increased tight binding to PIs can be achieved, this may limit the ability of the virus to evolve biologically significant resistance during the short period when viral replication is decreasing to full suppression. This raises the possibility that highly potent PIs may become legitimate candidates for single drug therapy.

Current PIs/drugs already offer much of this potential. However, rapid metabolism of the current inhibitors typically requires a boosting agent to increase drug levels. While drug resistance to many of the PIs is well described, no reports of drug resistance to the potent inhibitor DRV have been describe in subjects starting DRV who were previously PI-naïve, suggesting generating resistance to DRV de novo is difficult. Furthermore, there has been some success using DRV as a single agent in an induction/maintenance strategy of therapy (71, 218). Thus, DRV appears to have many of the properties we would anticipate for an optimal PI, which also suggests further improvements beyond DRV may take us into truly new territory in HIV-1 drugs and therapy.

C. Dissertation Overview

The infectivity of a HIV-1 particle strongly depends upon the activities of the viral PR during the maturation process. Impaired PR functionality can result in a reduced ability to fuse with a target cell (22-25), in a failure to complete reverse transcription (25-30), and, potentially, in failure to transverse the nuclear envelope (25). For the heavy reliance placed upon the PR, its task is not a simple one. The enzyme must cleave thousands of substrates in a very specific sequence, all of which are technically present at the same time. Yet it is highly effective when unencumbered by human intervention (28, 29, 51, 161).

Despite thirty-plus years of study, the mechanisms that govern the HIV-1 PR to ensure it accomplishes this prodigious feat remain unknown. The diversity of amino acid sequences recognized by the PR as substrate certainly plays a key role (49), yet no one has determined exactly how or why one sequence is superior to any other. Complicating matters are the existence of contextual determinants (50) and putative cofactors (219-221) that obscure the enzyme's specificity based on sequence alone. HIV-1 PR inhibitors are already the most effective treatment for HIV-1 (52, 53, 65, 70, 71), yet even they still have much room for improvement as I have just discussed. Should the mechanisms that control PR function be delineated, the design and effectiveness of these interventions would only grow.

The second chapter of this dissertation discusses the role of RNA as a cofactor for the HIV-1 PR. The long-standing hypothesis suggested RNA accelerated the cleavage rate of a single substrate, p15NC. The original report observed the phenomenon of RNA-dependent enhancement for only this RNA-binding protein and not a secondary substrate also derived from Gag. However, when we included p15NC and a second substrate based on the MA and CA regions of Gag in a single reaction, the ability of RNA to enhance processing rate was not limited to only the RNA-binding protein. This effect, which accelerated processing rate by more than 100-fold, was maintained even in the absence of p15NC or any other RNA-binding protein. These results led to the proposal of a new hypothesis – the interaction responsible for enhanced processing occurs between an allosteric binding site on the HIV-1 PR and RNA. The acceleration of a peptide assay by the addition of RNA offered convincing evidence in favor of this hypothesis, as the peptide was too small to simultaneously interact with the PR and the nucleic acid. Gel-shift assays screening for an interaction between the HIV-1 PR and short single-stranded DNA (ssDNA) oligonucleotides further confirmed the enzyme's ability to interact with nucleic acid. However, a complementary assay with the HIV-2 PR failed. The HIV-2 PR also failed to exhibit an improved processing ability in the presence of RNA. The overt difference between the HIV-1 and HIV-2 enzymes was their overall net charge; the HIV-1 PR was basic, and the HIV-2 PR acidic. This led to the conclusion that the interaction between the HIV-1 PR and RNA was non-specific and dependent upon an electrostatic attraction. An analysis of the kinetic efficiency of the HIV-1 PR in the presence and absence of RNA demonstrated that, mechanistically, both the affinity of the PR for its substrate and its turnover rate increased as a result of the interaction with RNA. While the precise location of the allosteric binding site is yet to be determined, I highlight the flap as a high-priority research target owing to its involvement in both substrate binding and catalysis. I also suggest that the juxtaposition of the HIV-1 PR and virion-packaged RNA could provide an additional regulatory mechanism for PR activity during the maturation of virus particles.

The third chapter sifts through the diversity in HIV-1 processing site amino acid sequence with the goal of identifying the underlying substrate preferences of the PR. Prior efforts have primarily relied upon small catalogs of short peptide substrates, but the pH and ionic strength of those works are not consistent with each other, complicating the interpretation of results. The conditions are also far from physiological, calling into question the accuracy and relevance of the results. To combat these problems, I generated the largest-to-date dataset of globular protein substrates where the cleavage efficiency could be measured under near-physiological conditions. Multiple natural HIV-1 PR cleavage sites were placed into the same context and mutated several times select amino acid positions. The rate of processing for these substrates was measured relative to an internal control protein to provide consistent and comparable measurements across reactions. I then evaluated the effect of the mutations on a site-by-site basis, before applying a series of statistical modeling procedures. The first set of models sought to identify the underlying similarities among all sites, which I accomplished with the use of linear-mixed effects modeling. This procedure accounted for the natural variance in the baseline rates among sites, effectively removing the noise in the data. Important physicochemical properties and amino acid positions were identified. I also attempted to build models with the intention of predicting out-of-sample data. However, these efforts were somewhat unsuccessful as the models were largely ineffective. Nonetheless, I identified an inverse relationship between models that could classify fast and slow sites, and those that could classify active and inactive PR putative cleavage sites. From this I conclude that the identification of an ideal amino acid sequence for the HIV-1 PR will likely require the application of multiple models. More specifically, a model to identify non-functional groups is required, and a second model built to predict rates would also be needed.

The fourth chapter will summarize the results presented in chapters two and three, with a discussion of future research directions.

CHAPTER II

A DIRECT INTERACTION WITH RNA DRAMATICALLY ENHANCES THE CATALYTIC ACTIVITY OF THE HIV-1 PROTEASE IN VITRO[3]

A. Introduction

For HIV-1 to spread from cell-to-cell the virus must assemble a particle capable of leaving its host cell without re-infecting the same cell. HIV-1 accomplishes this by constructing the virion as a rigid (222), non-infectious entity (223), and then later converting the particle into a mature, infectious form (20). The timing of this transfiguration is critical for viral infectivity since prematurely initiating (224, 225) or slowing the kinetics of maturation by reducing the number of active HIV-1 PR molecules (26, 28, 30) both disrupt the production of infectious viruses. This, therefore, requires the virus to employ regulatory mechanisms that manage the assembly, release, and maturation steps of the virus lifecycle. Many of these mechanisms concern the activity of the HIV-1 PR.

During assembly HIV-1 particles consist of two structural polyproteins, Gag and Gag-Pro-Pol. Both share the same first four domains – MA, CA, SP1, and NC – but differ in having either SP2 and a domain called p6 (in Gag) or a TF domain and monomers of the viral enzymes PR, RT, and IN (in Gag-Pro-Pol). The mass assembly of these proteins on the plasma membrane triggers the budding process, which is completed with the help of the ESCRT machinery (16). It is then, concurrent to or immediately after budding, that the HIV-1 PR activates to begin the maturation step of the lifecycle (20).

[3] Published as "A direct interaction with RNA dramatically enhances the catalytic activity of the HIV-1 Protease in vitro." Marc Potempa, Ellen Nalivaika, Debra Ragland, Sook-Kyung Lee, Celia Schiffer, Ronald Swanstrom in the *Journal of Molecular Biology*, Volume 427, Pages 2360-2378, July 17, 2015.

The HIV-1 PR is a dimer of two identical subunits (31-33), necessitating an interaction between a pair of Gag-Pro-Pol molecules to create the active site of the enzyme. The low stability of this interaction (36, 226) and accompanying poor catalytic activity (34, 36) restrict PR activation to budding or budded virions, where a high local concentration of Gag-Pro-Pol provides conditions that favor dimerization. The embedded PR overcomes its limited functionality through a series of intramolecular cleavage events that free the N termini of the monomers (36, 38-40, 227), thereby producing a much more stable enzyme capable of completing intermolecular cleavage events (40, 41, 228). The mature enzyme then proceeds to cleave the remaining structural polyproteins in a step-wise process that must go to near completion (43-45). Even modest under-processing at most sites can result in a non-infectious virus particle (26, 28-30, 51, 162).

Additional regulatory mechanisms exist to control the order and rate of Gag and Gag-Pro-Pol processing. Principally, the cleavage rate is regulated at the level of the processing site amino acid sequence. Each site has a unique sequence, with no obvious pattern connecting them (47). Instead, all the sites can occupy a conformation that fits into the conserved shape, i.e. substrate envelope, recognized by the HIV-1 PR (48, 49). The ability of each site to fill that space therefore defines a key determinant of processing order and rate. For instance, the two processing sites that are cleaved last, CA/SP1 and NC/SP2, are the most dynamic sites, suggesting they frequently shift in and out of conformations that do not mimic the substrate envelope (49). This structural plasticity makes them more difficult to cleave. Secondarily, the rate of cleavage for the CA/SP1, SP2/p6, and SP1/NC processing sites also may exhibit some dependence on contextual determinants (50, 229, 230).

A role for RNA as a cofactor has also been suggested. Maturation generates an intermediate called p15NC, which is comprised of the NC, SP2, and p6 domains of Gag. The next step of processing requires cleavage at the SP2/p6 site, but this does not readily occur in the absence of RNA when examined *in vitro* (134, 220, 221). In the presence of RNA, (or select DNA oligonucleotides), the rate of processing by the HIV-1 PR dramatically increases. In contrast, the cleavage rate of a truncated MA/CA substrate remained unaffected after the removal of RNA from the reaction system (220). Since p15NC contains the principal viral RNA-binding protein, a mechanism was proposed in which an interaction between RNA and p15NC induces a conformational change in the protein that exposes a buried cleavage site, and/or stabilizes the conformation of the SP2/p6 site to make it a more suitable substrate (220). In agreement with this hypothesis, the SP2/p6 site is one of the more dynamic cleavage sites, behind only CA/SP1 and NC/SP2 (49). Given the close proximity of the SP1/NC site to the RNA-binding domains, such a mechanism might also affect SP1/NC processing. Consistent with this, a 30-mer single-stranded DNA molecule was recently shown to increase the rate of SP1/NC processing within a truncated Gag polyprotein (229).

In an effort to study RNA-dependent processing, we established a two-substrate proteolysis system in which cleavage of the p15NC protein by the PR could be measured in tandem to the rate of cleavage of an internal control protein that was purportedly unaffected by nucleic acid. Contrary to prior results, we found both substrates exhibited an increased processing rate in the presence of RNA. Additional single-substrate assays with globular and peptide substrates demonstrated that RNA enhances processing in a substrate-independent manner. This led us to hypothesize that the critical interaction occurs between RNA and the HIV-1 PR, an interaction substantiated with a gel-shift assay. Examination of a panel of HIV-1

PRs demonstrated that this enzyme-RNA interaction is conserved across multiple subtypes, as well as in patient-derived drug-resistant enzymes. In contrast, the HIV-2 PR does not interact with RNA, and does not cleave its substrates more efficiently with RNA present. The interaction between the HIV-1 PR and RNA is primarily electrostatic in nature, although sequence and structural determinants within the polyanion may also play a role. Use of a tethered dimer of the HIV-1 PR revealed RNA-enhanced cleavage does not result from increased dimer stability. While the exact mechanism of enhancement has not yet been identified, we did determine that RNA affects both the K_m and k_{cat}. These findings support the existence of an allosteric binding site on the HIV-1 PR, and raise the possibility that PR activity during assembly could be regulated in part by the juxtaposition of the PR and virion-packaged RNA.

B. Results

1. Multiple substrates of the HIV-1 PR exhibit an enhanced rate of proteolysis in the presence of RNA.

We first documented a two-substrate protease assay where the rate of p15NC processing could be measured relative to an internal control protein. To demonstrate that we could independently measure multiple substrates in a single reaction, we performed a time-course experiment with two substrates containing the canonical MA/CA cleavage site. One substrate consisted of the entirety of the MA and CA domains (MA/CA); the other substrate had a GST-tag fused to the N terminus of MA and a CA region truncated at amino acid 145 (50) (GMCΔ). This truncation occurred between the N- and C-terminal domains of CA, which exist as two separate domains (231, 232), ostensibly leaving the conformation of the MA/CA cleavage site unaffected. In addition, these N- and C-terminal changes allowed the two forms of MA/CA to migrate to different positions in a polyacrylamide gel in both the uncleaved and the cleaved states (Figure 2.1). As we have observed previously (29), both substrates were processed in parallel and at nearly identical rates. Thus, the two-substrate system enables the direct comparison of relative rates of cleavage by the HIV-1 protease using pairs of protein substrates. For this analysis we have relied on coomassie staining of the proteins and the disappearance of substrate over time to provide flexibility in the types of proteins that can be analyzed.

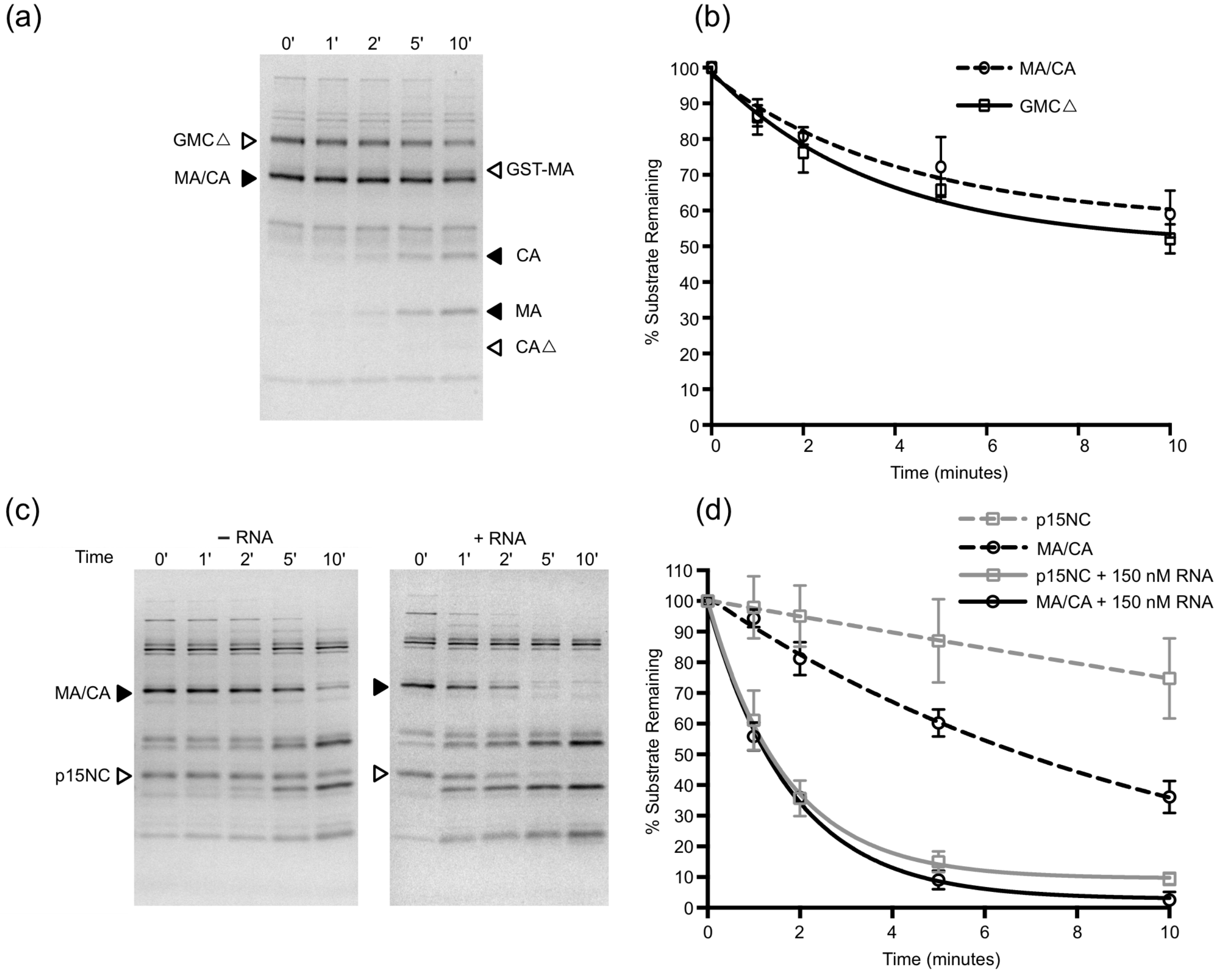

Figure 2.1: Concomitant processing of multiple substrates by the HIV-1 protease. (a) A reaction mixture containing the MA/CA (solid triangle) and GMCΔ (open triangle) substrates in equimolar amounts was incubated at 30ºC, pH 6.5 for 1 hour prior to the addition of the HIV-1 PR. Reactions were run for 10 minutes with the intermittent removal of aliquots that were immediately mixed with SDS to halt the reaction. Reaction substrates and products were separated by SDS-PAGE and stained by coomassie. MA/CA products are indicated with solid, reversed triangles; GMCΔ products with open, reversed triangles. (b) MA/CA (circle) and GMCΔ (square) bands were quantified with imaging software and graphed as substrate remaining versus time. (c) Reactions containing MA/CA (solid triangle) and p15NC (open triangle) in a molar ratio of 1:4 were performed as in Figure 2.1a, +/- 150 nM of a heteropolymeric 532-nucleotide RNA derived from the p15NC region of the HIV-1 NL4-3 genome. (d) Quantification of MA/CA (black circle) and p15NC (grey square) two-substrate assays. Dashed lines show reactions without RNA; reactions with RNA are represented with solid lines. All errors bars represent the standard deviation resulting from three independent experiments.

We sought to determine whether we could observe RNA-dependent rate enhancement of p15NC cleavage by substituting p15NC into the reaction system for the GMCΔ protein. Due to the poor staining profile of p15NC, its final concentration in the reaction was four-fold higher than that of the MA/CA substrate. In the absence of RNA (Figure 1c, left panel), cleavage of both MA/CA and p15NC was observed. Additional reactions performed with or without nuclease pre-treatment were indistinguishable (data not shown), affirming that these reactions were devoid of RNA. The percent substrate remaining was quantified by densitometry and plotted as a function of time (Figure 1d, dashed lines). Product formation for MA/CA was easily observable, however we could not unequivocally identify the products of p15NC cleavage. One stained poorly, and the other product ran to the same location on the gel as a contaminant remaining after protein purification. To confirm the 25% drop in p15NC band intensity we observed was in fact processing of the p15NC substrate by the HIV-1 PR, we performed the reaction in the absence of PR and found that there was no observable change in either substrate (data not shown), suggesting the decrease in p15NC band intensity was due to PR cleavage. Thus, in this system, the MA/CA protein was processed approximately 2.5-fold faster than p15NC in the absence of RNA.

When we performed the two-substrate reaction in the presence of long, heteropolymeric RNA, a 532-base transcript derived from the p15NC region of the HIV-1 genome, we found the rate of substrate disappearance accelerated for both substrates (Figure 2.1 c/d). This result contradicts previously published work that found MA/CA cleavage unaffected by the presence of nucleic acid (220, 229). Nonetheless, we consistently observed a 15-fold increase in the rate of p15NC cleavage in concert with an 8-fold increase in the rate of MA/CA processing. Our results argue that substrates that do not contain NC can also exhibit an enhanced rate of cleavage in the presence of RNA.

2. RNA-dependent rate enhancement is a substrate-independent phenomenon.

In order to confirm that RNA-dependent enhancement of MA/CA cleavage occurs independently of p15NC, we performed single-substrate proteolysis assays in the presence and absence of RNA. In our single-substrate assays, the MA/CA protein was labeled with the Lumio Green Reagent after mutagenesis of the Cyclophilin A loop in CA to create a fluor binding site (29). This label binds with high specificity and in a one-to-one molar ratio with the substrate, allowing specific and more sensitive detection of the substrate and the CA product compared to the coomassie stain (Figure 2.2). As a result, we could reliably determine the percent product formed (Intensity of CA band x 100% ÷ total fluorescence intensity of all bands in the lane), greatly increasing the accuracy of our data collection. Unfortunately, the label could not be used with the p15NC substrate because the zinc-finger domains in p15NC also bind to the Lumio Green Reagent. With this assay, we could accurately determine changes in the MA/CA cleavage rate of up to 40-fold. However, when comparing the no-RNA to plus-RNA reactions, we detected a fold enhancement beyond that value. Therefore, we only estimate the rate of MA/CA cleavage as 80- to 90-fold faster in the presence of RNA (Figure 2.2b). The magnitude by which MA/CA cleavage was accelerated in single substrate assays was considerably greater than in the two-substrate assay. Several explanations likely account for this discrepancy: the increased sensitivity of the fluorescence-based assay allowed more precise determination of reaction progress while under near steady-state conditions; the potential presence of multiple RNA-binding proteins in the two-substrate assay meant there was competition for RNA, which limited the effect; and, owing to an inability to purify p15NC to high concentrations, the ionic strength of the two-substrate assay was likely higher than anticipated due to a substantially larger contribution of protein storage buffer to the reaction mixture. Notably, raising the ionic strength of the reaction by increasing the salt concentration drastically reduces the magnitude of the RNA-enhancement effect (233), even in single-substrate MA/CA assays (see below).

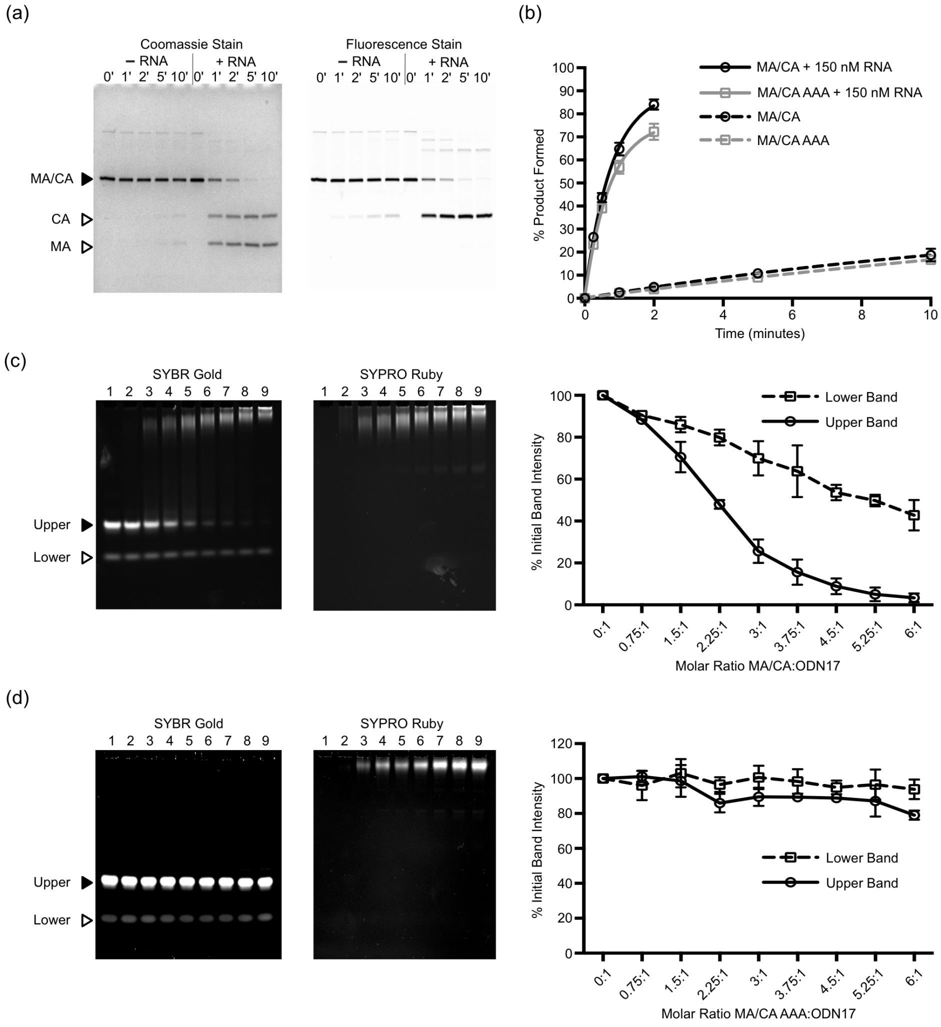

Figure 2.2 RNA accelerates processing independent of the ability of MA/CA to bind nucleic acid. (a) MA/CA proteins were tagged with the Lumio Green Reagent via mutation of the Cyclophilin A binding loop in CA to include a CCPGCC motif. Single-substrate MA/CA proteolysis assays were visualized by coomassie (left) and fluorescence (right). Only CA-containing reactant and product species are present in the fluorescence stain. (b) Reaction progress curves for single-substrate MA/CA (black) and MA/CA-AAA (grey) proteolysis reactions performed in the absence (dashed lines) and presence (solid lines) of RNA. (c) and (d) Binding reactions containing 100 ng (2 μM) of the single-stranded DNA molecule ODN17 and steadily increasing amounts (from 0 μM in lane 1 to 12 μM in lane 9) of MA/CA (c) or MA/CA-AAA (d) were electrophoresed in a 6% polyacrylamide gel under native conditions. Gels were stained with SYBR Gold (left panel) to visualize nucleic acid species. The gels were then washed and restained with SYPRO Ruby (central panel) to view protein. The intensity of each nucleic acid species relative to the respective band in the no-protein reaction (lane 1) was determined by densitometry and plotted as a function of protein:ODN17 concentration (right panel). All errors bars represent the standard deviation resulting from three independent experiments.

Since MA/CA also exhibited RNA-dependent enhancement, we considered the possibility that this effect resulted from an interaction between MA/CA and RNA. MA contains a highly basic region on the globular head of the protein capable of binding nucleic acid (234-238). Because the RNA species used in prior experiments was a 532-base transcript derived from the p15NC region of the HIV-1 genome, it was poorly suited for use in a gel-shift assay. To identify a surrogate nucleic acid, we tested the ability of two short ssDNA oligonucleotides, N5 (21 bases) and ODN17 (17 bases), to accelerate MA/CA processing. Both molecules had previously been reported to enhance p15NC processing similarly to RNA, although not as potently (134, 221). In agreement with previous reports, both N5 and ODN17 increased the rate of processing, but to a much lower degree than RNA (Figure 2.3).

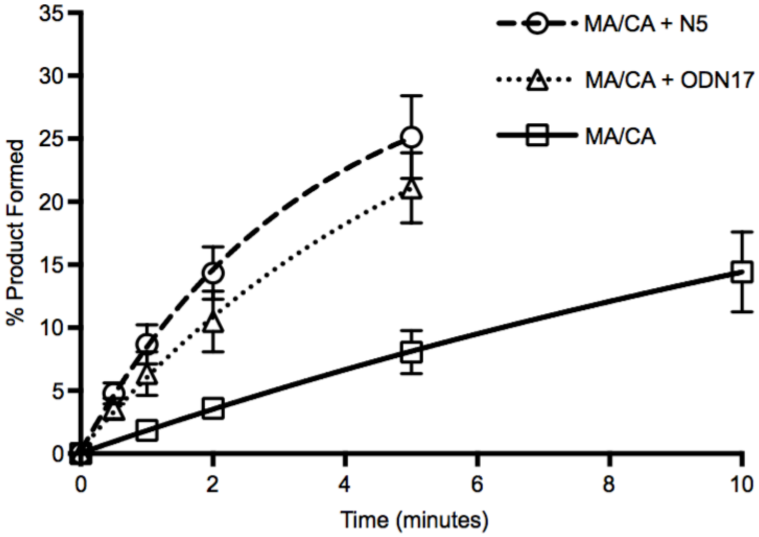

Figure 2.3: Short DNA oligonucleotides previously reported to enhance p15NC processing also increase MA/CA processing rate. Shown are reaction progress curves for single-substrate MA/CA proteolysis assays performed under standard conditions in the absence of nucleic acid (squares), in the presence of a 17-mer single-stranded DNA oligonucleotide (ODN17; triangles), or in the presence of a 21-mer single-stranded DNA oligonucleotide (N5; circles). Both ODN17 and N5 were previously reported to enhance p15NC processing. All errors bars represent the standard deviation resulting from three independent experiments.

Using ODN17, we performed a gel-shift assay to determine the ability of MA/CA to bind to nucleic acid. Under native conditions, ODN17 runs as two species (Figure 2.2c, left panel); presumably, the lower species corresponds to single-stranded molecules, while the upper band may represent a G-quadruplex structure that ODN17 has the potential to form (134, 239). In concert with the addition of increasing amounts of MA/CA to the binding reactions, (from 0 μM in lane 1 to 12 μM in lane 9), both the upper and lower bands progressively diminished (Figure 2.2c, right panel), and a new band appeared toward the top of the gel. The new band overlaps with where the MA/CA protein runs (Figure 2.2c, center panel), signifying that the MA/CA protein interacts with nucleic acid.

Addressing the question of whether this interaction was required for the enhanced rate of MA/CA processing, we replaced the lysines in the $K_{26}KQYK_{30}$ sequence of MA with alanines to generate the mutant protein MA/CA-AAA. In a previous report, mutating these lysines disrupted the residual RNA-binding ability of Gag molecules whose NC domain had been deleted (237). Consistent with those results, MA/CA-AAA was severely attenuated in its ability to interact with ODN17 (Figure 2.2d). If an interaction between MA/CA and nucleic acid was responsible for its increased cleavage rate, then we should have found enhanced processing of MA/CA-AAA to be considerably reduced or nonexistent. However, RNA-dependent enhancement of MA/CA-AAA processing still occurred, and at a nearly identical magnitude (Figure 2.2b, gray lines). These results suggest that RNA-dependent enhancement is independent of the substrate.

In order to confirm the lack of requirement for an interaction between substrate and RNA, we performed proteolysis assays utilizing a 12-amino acid peptide as the substrate. The HIV Protease Substrate 1 (Sigma) is a fluorogenic substrate too small to interact with RNA and simultaneously be cleaved by the HIV-1 PR; conveniently, this peptide also contains the same cleavage site sequence as the MA/CA protein. Adding RNA to the peptide proteolysis reaction still accelerated the rate of the reaction, though only by 20-fold (Figure 2.4). Differences in reaction conditions for the peptide (i.e. pH), and/or the absence of contextual determinants could account for the reduced magnitude of the effect relative to the globular MA/CA substrate. Nonetheless, the results of the peptide assay show that the increased reaction rate observed upon addition of RNA is independent of the substrate.

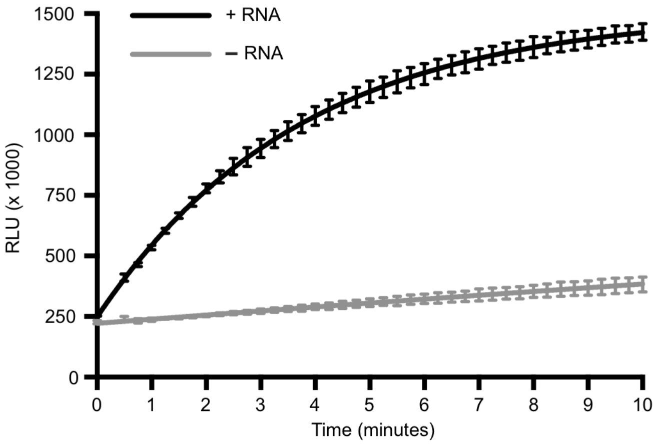

Figure 2.4: The addition of RNA to a reaction accelerates processing of a peptide substrate. A commercially available 12-amino acid peptide substrate was utilized as a substrate for the HIV-1 PR. Processing of the peptide substrate was monitored by an increase in fluorescence resulting from the separation of a fluorophore from a quencher placed on opposing ends of the peptide. Reactions were run at 30ºC, pH 4.8, and performed in the absence (grey) or presence (black) of long, heteropolymeric RNA. All errors bars represent the standard deviation resulting from three independent experiments.

3. The HIV-1 PR can directly interact with nucleic acid.

Since an interaction between RNA and substrate is not the mechanism driving RNA-dependent enhancement, we hypothesized that an interaction was occurring between RNA and the HIV-1 PR. We again used ODN17 as a surrogate for RNA-binding, and performed gel-shift assays utilizing the HIV-1 PR as the protein in each binding reaction. Similar to the wild-type MA/CA protein, adding progressively more PR resulted in the disappearance of the upper oligonucleotide band (Figure 2.5, left panel). In this case the fluorescence intensity of the lower band remained relatively unchanged regardless of the amount of PR present in the binding reaction. Under native conditions, the HIV-1 PR does not enter the gel, likely because of its basic profile (HIV-1 PR has an isoelectric point of 9.1), so there is no overt overlapping band in both stains. Nonetheless, we infer the presence of a PR-ODN17 complex from the selective loss of the upper oligonucleotide band and a low level of fluorescence in the wells of the central lanes of both SYBR gold and SYPRO ruby stains. The fluorescence may disappear from the latter wells because the addition of more HIV-1 PR increases the net charge of the PR-ODN17 complexes so that the complexes flow into the running buffer rather than remain in the well. From this, we conclude that the HIV-1 PR can directly interact with nucleic acids.

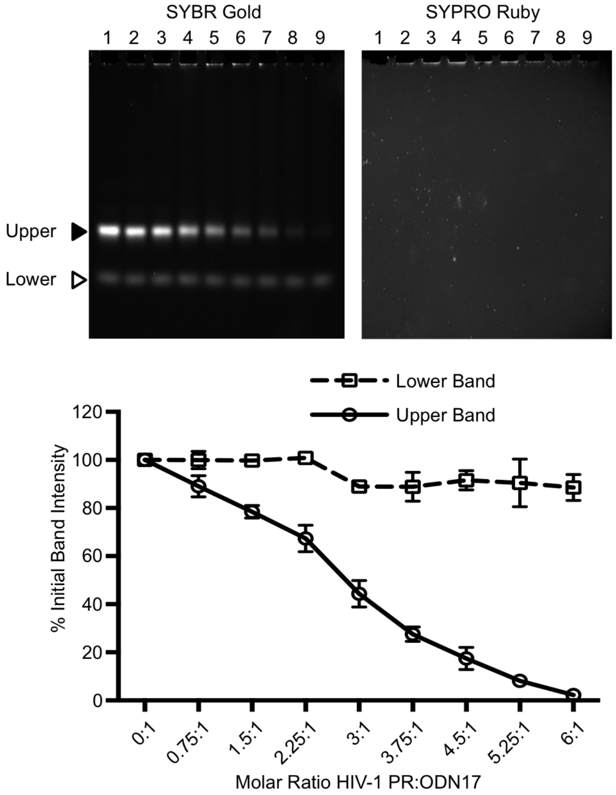

Figure 2.5: The HIV-1 PR can directly interact with nucleic acid. Binding reactions containing 100 ng of ODN17 (2 μM) and steadily increasing amounts (from 0 μM in lane 1 to 12 μM in lane 9) of the HIV-1 PR were electrophoresed in a 6% polyacrylamide gel under native conditions. As previously, gels were stained with SYBR Gold (top left) to visualize the nucleic acid species followed by SYPRO Ruby (top right) for visualization of protein. The relative intensity of each nucleic acid species was determined by densitometry and plotted as a function of protein:ODN17 concentration (bottom). The HIV-1 PR does not enter the gel under native conditions because of its high isoelectric point (pI = 9.1). Complexes can be inferred from the low level fluorescence present in the wells of the central lanes and the selective depletion of the upper nucleic acid species. All errors bars represent the standard deviation resulting from three independent experiments.

4. RNA accelerates processing by HIV-1 PRs from multiple subtypes and from drug resistant variants, but not processing by HIV-2 PR.

Given that we had only shown a single variant of a subtype B HIV-1 PR was capable of interacting with RNA to enhance its activity, we expanded our data set to include HIV-1 PRs of subtype C and CRF01_AE, as well as an HIV-2 PR (Table 2.1). We also tested several patient-derived, drug-resistant subtype B PRs to determine whether the effect is maintained after significant changes to the amino acid sequence (19-26 substitutions from the PR of the SF2 isolate of HIV-1 subtype B used in the prior experiments). All three wild-type HIV-1 PRs exhibited a 100-fold or greater increase in the rate they processed the MA/CA protein in the presence of RNA relative to their respective no-RNA controls (Figure 2.6). The drug-resistant HIV-1 subtype B PRs likewise demonstrated enhanced catalytic activity in the presence of RNA, although there was considerably more variability in the magnitude of the effect for these enzymes. Despite the variability in magnitude, VSL23, the PR whose activity was least accelerated, still cleaved MA/CA 30-fold more quickly. The only enzyme entirely unaffected by the presence of RNA was the HIV-2 PR. For this latter assay the cleavage site of MA/CA was adapted to reflect the canonical site for HIV-2, allowing the substrate to be efficiently cleaved by the enzyme (data not shown). Consistent with these results, we found that the HIV-2 PR did not interact with nucleic acid in a gel-shift assay (Figure 2.7; note the HIV-2 PR with its lower isoelectric point of 5.3 enters the gel). Both the PR-dependent variability in the magnitude of the effect, and the lack of enhancement for the HIV-2 PR further demonstrated that RNA-dependent enhancement results from an interaction between RNA and the HIV-1 PR rather than with the substrate.

Table 2.1: Amino acid sequences and theoretical isoelectric points of HIV PRs utilized in MA/CA processing assays.

Protease	pI	Amino Acid Sequence
HIV-1 Subtype B	9.1	PQITLWKRPLVTIRIGGQLKEALLDTGADDTVLEEMNLPGKWKPKMIGGIGGFIKVRQYDQIPVEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
HIV-1 CRF01_AE	9.1	PQITLWQRPLVTVKIGGQLKEALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIKVRQYDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQIGCTLNF
HIV-1 Subtype C	9.1	PQITLWQRPLVSIRVGGQIKEALLDTGADDTVLEEVNLPGKWKPKMIGGIGGFIKVRQYDQIPIEICGKKAIGTVLVGPTPINIIGRNMLTQLGCTLNF
HIV-2	5.3	PQFSLWKRPVVTAYIEGQPVEVLLDTGADDSIVAGIELGNNYSPKIVGGIGGFINTKEYKNVEIEVLNKKVRATIMTGDTPINIFGRNILTALGMSLNL
HIV-1 SLK19	9.1	PQITLWKRPILTVRIGGQLKEVLLDTGADDTVLEDIDLPGRWKPKMIMGIGGLVKVRQYDQVPIEICGHKVIGSVLVGPTPANVIGRNLLSKIGCTLNF
HIV-1 KY26	9.0	PQITLWKRPVVVVKVGGQLMEALLDTGADDTIFEEMNLPGRWTPKIVGGIGGFMKVRQYENVPIEIYGKKILSTVLIGPTPANIIGRNVMTQIGCTLNF
HIV-1 ATA21	9.1	PQITLWKRPFITVKIGGQQMEALLDTGADDTIVEAINLPGRWKPKIVGGIGGFMKVKQYDQVPVEICGHKAITAVLVGPTPVNVIGRNVMTQIGCTLNF
HIV-1 VEG23	8.8	PQITLWKRPIIKVKIGGQLVEALLDTGADDTIFEGIDLPGRWKPKIVGGIGGFMKVKEYDQIPVEVCGHKVISTVLVGPTPVNVIGRNVMTQIGCTLNF
HIV-1 VSL23	8.8	PQITLWKRPIVTIKIGGQLREALLDTGADDTVFTDIDLPGRWTPKIIVGVGGFSKVKQYDQVPIEICGHKVVGTVLIGPTPANIVGRNLLTQLGCTLNF

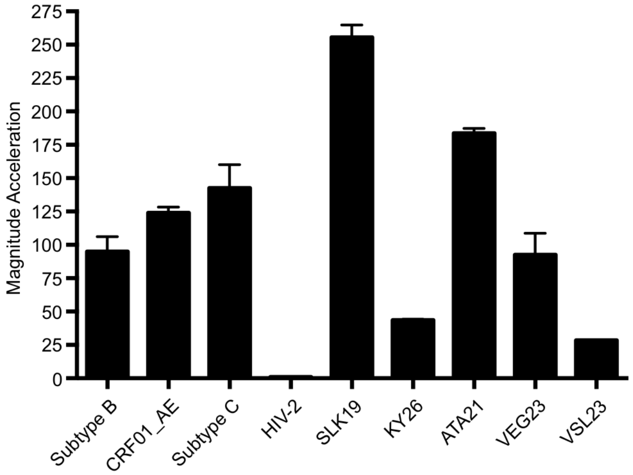

Figure 2.6: The rate of processing by multiple HIV-1 PRs, but not the HIV-2 PR, increases in the presence of RNA. The rate of MA/CA processing by HIV-1 PRs from three different subtypes and an HIV-2 PR were evaluated in the presence and absence of RNA. The MA/CA protein substrate for the HIV-2 PR was mutated at the processing site to better reflect the canonical HIV-2 MA/CA processing site. Five highly mutated drug-resistant HIV-1 subtype B PRs (SLK19, KY26, ATA21, VEG23, and VSL23) were also examined. Reactions were performed with the globular MA/CA substrate under standard conditions with the exception of PR concentration, which was adjusted on an enzyme-to-enzyme basis to achieve 10% cleavage over the course of the reaction in the absence of RNA. SLK19, KY26, and ATA21 required up to 4-fold higher concentrations; VEG23 and VSL23 were more severely attenuated, and required a 20-fold higher concentration of enzyme. Results are reported as the magnitude difference in acceleration of the RNA-plus reaction relative to each enzyme's respective no-RNA control. All errors bars represent the standard deviation resulting from three independent experiments.

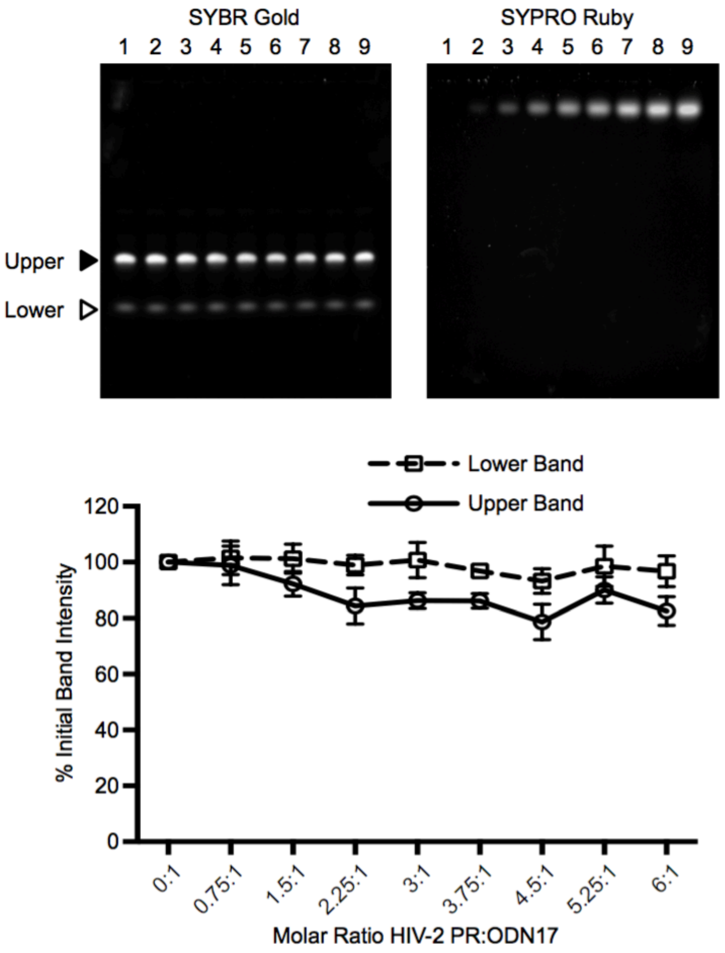

Figure 2.7: The HIV-2 PR does not interact with nucleic acid. Binding reactions containing 100 ng of ODN17 (2 µM) and increasing amounts (from 0 µM in lane 1 to 12 µM in lane 9) of the HIV-2 PR were electrophoresed in a 6% polyacrylamide gel under native conditions. Gels were stained with SYBR Gold (top left) to visualize the nucleic acid species followed by SYPRO Ruby (top right) for visualization of protein. Note the HIV-2 PR with its lower isoelectric point (pI = 5.3) enters the gel. The relative intensity of each nucleic acid species was determined by densitometry and plotted as a function of protein:ODN17 concentration (bottom). All errors bars represent the standard deviation resulting from three independent experiments.

5. Long, heteropolymeric RNA is the most effective enhancer of HIV-1 PR activity, but small ssDNA molecules and tRNA are still functional enhancers.

Whereas MA/CA interacted with both the upper and lower ODN17 bands in the gel-shift assay, the HIV-1 PR demonstrated a selective interaction with only the upper ODN17 species. Additionally, ODN17 and N5 were considerably less potent enhancers than p15 RNA, requiring much higher concentrations to be effective in our earlier assays. These data raised the possibility that a specific interaction might occur between enzyme and nucleic acid, which was fulfilled much more capably by some component of the p15 RNA transcript used in our assays. To determine whether p15 RNA contains some specific feature required for an interaction between the HIV-1 PR and nucleic acid, we generated dose-response curves for multiple different RNA transcripts, yeast tRNA, and the N5 and ODN17 single-stranded DNA molecules. All long (>400 bases), heteropolymeric RNAs accelerated the reaction equivalently (Figure 2.8a), including transcripts that were not derived from the HIV-1 genome (data not shown). The long RNAs also had very similar EC50 values (Table 2.2) that were even closer to identical when adjusted for length ($EC50_{/nt}$). Even though N5 and ODN17 were roughly one-tenth as effective as long heteropolymeric RNA in the magnitude of the enhancement effect, both still accelerated MA/CA processing by about 10-fold. Though the EC50s of N5 and ODN17 were in the μM range, the $EC50_{/nt}$ were reasonably similar to those of long heteropolymeric RNA. Yeast tRNA grouped with the single-stranded DNA molecules regarding the magnitude of enhancement, but the EC50 value was nearer the RNA transcripts. Thus tRNA had the lowest $EC50_{/nt}$, but this value was still only two-fold lower than most other nucleic acids. We conclude that long, heteropolymeric RNAs are the most potent enhancers of HIV-1 PR activity, but because all six nucleic acids tested had very similar $EC50_{/nt}$, the amount of nucleic acid, rather than a specific sequence or structure, appears to be the critical determinant. However, this does not yet address why long heteropolymeric RNA was 10-fold more potent in the magnitude of enhancing HIV-1 PR activity than the other nucleic acids (i.e. 100-fold vs 10-fold enhancement), nor does it explain the selective binding of the oligonucleotides observed in the gel-shift assay.

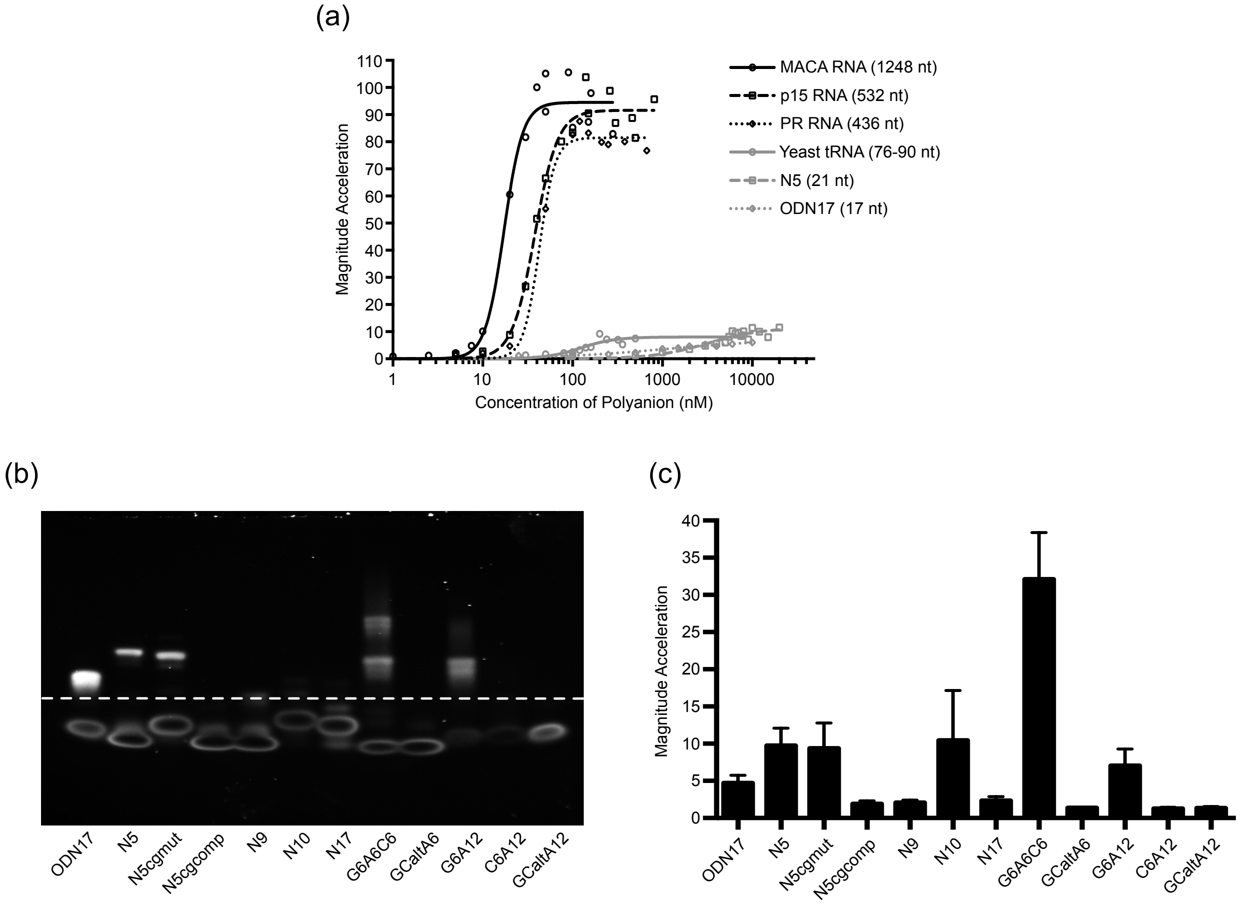

Figure 2.8: Multiple nucleic acid species can enhance HIV-1 PR activity, though potencies vary. (a) Dose response curves were generated for several long heteropolymeric RNAs, yeast tRNA, and the single-stranded DNA oligonucleotides N5 and ODN17. Reactions were performed under standard conditions using the MA/CA protein as the substrate. (b) 100 ng of the indicated single-stranded DNA oligonucleotides were electrophoresed in a 6% polyacrylamide gel under nondenaturing conditions, and then visualized with SYBR gold. The dashed white line distinguishes single-stranded species from oligomeric species. (c) Each single-stranded DNA oligonucleotide was supplied at a final concentration of 10 μM in MA/CA processing reactions and evaluated for its ability to improve HIV-1 PR function. Results are reported as the magnitude of acceleration relative to MA/CA processing in the absence of nucleic acid. All errors bars represent the standard deviation resulting from three independent experiments.

Table 2.2: Length and efficacy of polyanions as enhancers of HIV-1 PR activity.

	Length (nt)	Max. Fold Acceleration	EC50 (nM)	EC50/nt (nt x 10^18/L)
MACA RNA	1248	95	17	13
p15 RNA	532	92	38	12
PR RNA	436	82	43	11
Yeast tRNA	76-90	8.0	108-124	5.7 – 6.7
N5	21	12	3229	41
ODN17	17	7.3	1148	12
Heparin	---	30	175-195	---
Poly(dA)	49	< 2	---	---
Poly(dC)	49	21	569	17
Poly(dG)	49	15	506	15
Poly(dT)	49	18	635	19

As a means of further investigating the apparent selectivity of the HIV-1 PR for the larger ODN17 species, as well as the discrepancy in the magnitude of the effect between long RNA transcripts and single-stranded DNA molecules, we increased our catalogue of oligonucleotides and tested each for their ability to enhance PR activity. Figure 2.8b and 2.8c contain the results from a selection of these molecules, all of which are between 17 and 21 nucleotides in length (Table 2.3). Of the twelve molecules shown, six of them (ODN17, N5, N5cgmut, N10, G6A6C6, G6A12) enhanced the rate of the reaction by at least 5-fold; the other six were ineffective even at concentrations exceeding 10 μM. Among the single-stranded DNA molecules capable of enhancing the reaction, all but the C-rich N10 molecule formed slower migrating species when electrophoresed through a 6% polyacrylamide gel. However, we cannot definitively state whether N10 did or did not form a secondary species because the SYBR gold stain is much less effective at staining C-rich oligonucleotides (e.g. Figure 2.8b, C6A12), and C-rich nucleic acids are capable of forming higher order multimeric species (240). With the possible exception of N10, these results are in accordance with the gel-shift assay where only the larger nucleic acid species interacted with the HIV-1 PR.

Table 2.3: List of single-stranded DNA oligonucleotide sequences and lengths.

Oligonucleotide	Length (nt)	Sequence (5’-3’)
ODN17	17	TTGGGGGGTACAGTGCA
N5	21	GCCCTTTTTCCTAGGGGCCCT
N5cgmut	21	CGCGTTTTTCCTAGGGGCCCT
N5cgcomp	21	CGCGTTTTTCCTAGCGCGCCT
N9	21	GGAAGGCCAGATCTTCCCTAA
N10	21	ATTCCCTGGCCTTCCCTTGTA
N17	21	ATACAGTTCCTTGTCTATCGG
G6A6C6	18	GGGGGGAAAAAACCCCCC
GCaltA6	18	GCGCGCAAAAAAGCGCGC
G6A12	18	GGGGGGAAAAAAAAAAAA
C6A12	18	CCCCCCAAAAAAAAAAAA
GCaltA12	18	GCGCGCAAAAAAAAAAAA
Poly(dA)49	49	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
Poly(dC)49	49	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Poly(dG)49	49	GGGGGGAGGGGGGAGGGGGGAGGGGGGAGGGGGGAGGGGGGAGGGGGGA
Poly(dT)49	49	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

We also noticed a direct correlation between the migration distance in the gel, where observable, and the magnitude of the enhancer effect (Figure 2.8b and 2.8c). Preheating aliquots of the oligonucleotides before use confirmed the importance of these multimeric and/or structured species – only those whose slower migrating species remained after heating retained their enhancer activity (Figure 2.9) – and also strengthened the correlation between migration pattern and effect magnitude. G6A6C6 lost one of its two larger species after heating, with the remaining band migrating to a position similar to the species present in the unheated N5 aliquots. The magnitude by which the heated G6A6C6 preparation enhanced HIV-1 PR activity resembled that of unheated N5, highlighting the proposed relationship. These data suggest that the interaction between the HIV-1 PR and nucleic acid is primarily electrostatic in nature, requiring a polyanion of some particular size or conformation rather than a specific sequence. Also, the potency of an enhancer's effect may be determined by the size and/or conformation of the molecule.

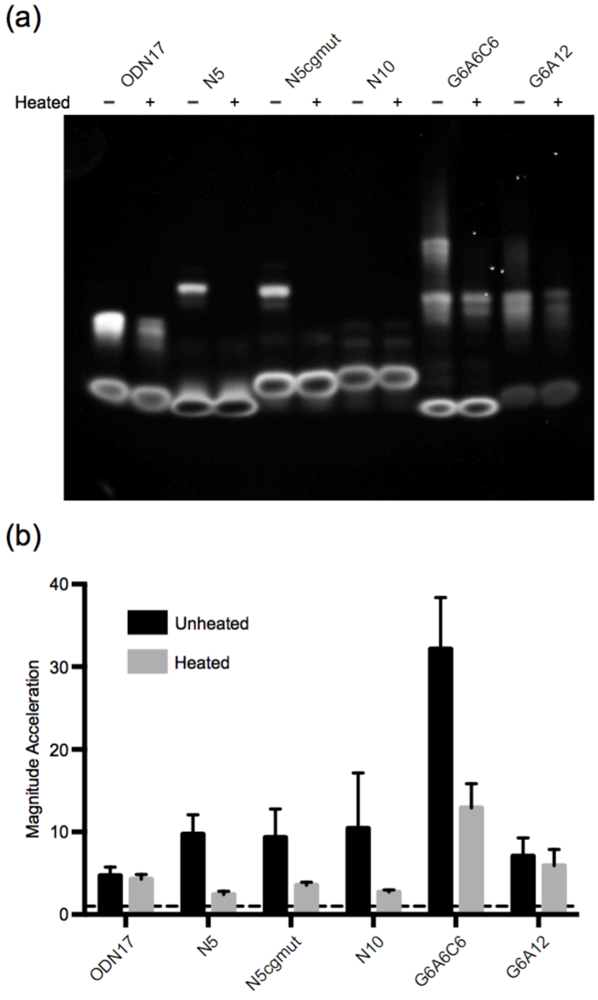

Figure 2.9: Heat-induced loss of slower migrating nucleic acid species corresponds with a loss in effector potency. (a) Aliquots of each indicated nucleic acid species were either left at room temperature or heated at 90ºC for 10 minutes before rapid cooling by centrifugation. Each lane contains 100 ng of DNA from the indicated unheated or heated samples. Bands were visualized with SYBR gold following electrophoresis in a 6% polyacrylamide gel. (b) Unheated (black) and heated (grey) DNA oligonucleotides were provided at a final concentration of 10 μM and examined for their ability to enhance HIV-1 PR-mediated processing of MA/CA under standard reaction conditions. All errors bars represent the standard deviation resulting from three independent experiments.

6. The interaction between the HIV-1 PR and nucleic acid is principally electrostatic, but additional factors affect the magnitude of enhancement.

If the HIV-1 PR-RNA interaction is primarily electrostatic in nature, then a non-nucleic acid polyanion should be sufficient to enhance the enzyme's catalytic activity. To test this, we utilized heparin as the polyanion in single-substrate proteolysis assays and generated a dose-response curve (Figure 2.10a). As expected, heparin enhanced proteolysis, accelerating the rate of the reaction by up to 30-fold. This value put heparin squarely between long, heteropolymeric RNA and the short single-stranded DNA molecules in effectiveness. With the expected size of commercially produced heparin molecules to be 17-19 kDa, an equivalently sized molecule of RNA would be approximately 53-60 nucleotides. As this size fits between the HIV-1 RNA transcripts and the single-stranded DNA molecules, the magnitude of the effect remains consistent with the proposed relationship between size and effectiveness of the polyanion. We also tested a polycation spermine (Sigma) in our system, but it had no effect on HIV-1 or HIV-2 PR activity at all concentrations examined (data not shown). However, we cannot rule out the possibility that its ineffectiveness was the result of the small size of individual spermine molecules.

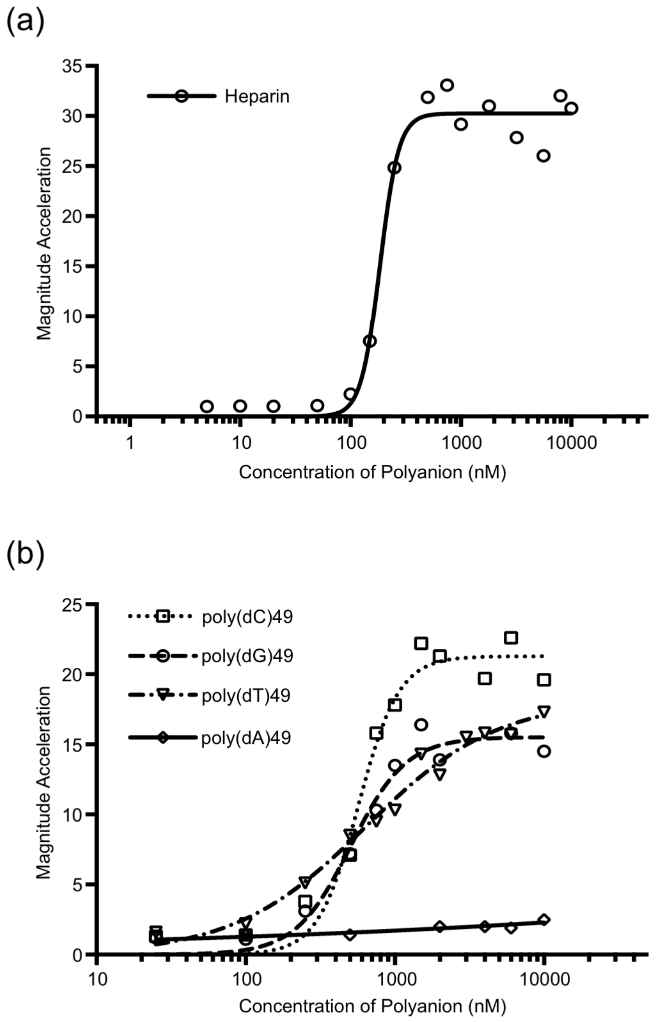

Figure 2.10: A specific nucleic acid sequence or structure is not required for enhancement. Dose-response curves were generated under standard assay conditions for (a) heparin, and (b) deoxynucleotide homopolymers. Reactions were performed with the globular MA/CA substrate under standard conditions.

If an electrostatic interaction were sufficient, we hypothesized that homopolymers of each deoxynucleotide should be equally effective at accelerating the rate of proteolysis. We generated dose response curves for 49-mers of poly(dA), poly(dC), and poly(dT), (Figure 2.10b). Owing to synthesis constraints, the poly(dG) molecule contained an adenosine every seventh base (Table 2.1). We found that only three of the four nucleic acid homopolymers accelerated the rate of MA/CA cleavage, with poly(dA) incapable of enhancing the rate of the reaction by any significant amount under the conditions tested. The three other oligonucleotides yielded similar results in both the magnitude of their effect and their EC50 values (Table 2.2). Furthermore, the 15-20-fold rate enhancement observed was also consistent with the predicted result based upon the size of the molecules. The observation that poly(dA) was ineffective indicates that while a polyanion is necessary, it is not sufficient for enhancing proteolysis.

Additionally, if the PR-RNA interaction is electrostatic in nature, the pH and ionic strength of the reaction mixture should influence the enhancer effect. The intracellular ionic strength of mammalian cells is approximately 0.15 M, which is similar to the conditions of our single-substrate assays. Increasing the ionic strength to 0.2 M by adding NaCl to the reaction buffer reduced the effect to only a 10-fold enhancement. RNA had no effect in reactions with an ionic strength greater than 0.5 M (Figure 2.11a). Though the pH at the site of virus assembly, budding, and maturation has not yet been formally determined, the cytosolic pH of lymphocytes is approximately 7.2 (241). When we performed assays at pH 7.2, select shorter nucleic acids, including tRNA and G6A6C6, still accelerated MA/CA processing albeit to a limited degree; the remaining 17-21 single-stranded DNA molecules became ineffective (data not shown). Long heteropolymeric RNA was also still effective, but similarly to raising the ionic strength, the magnitude of the effect decreased to approximately eightfold enhancement (Figure 2.11b).

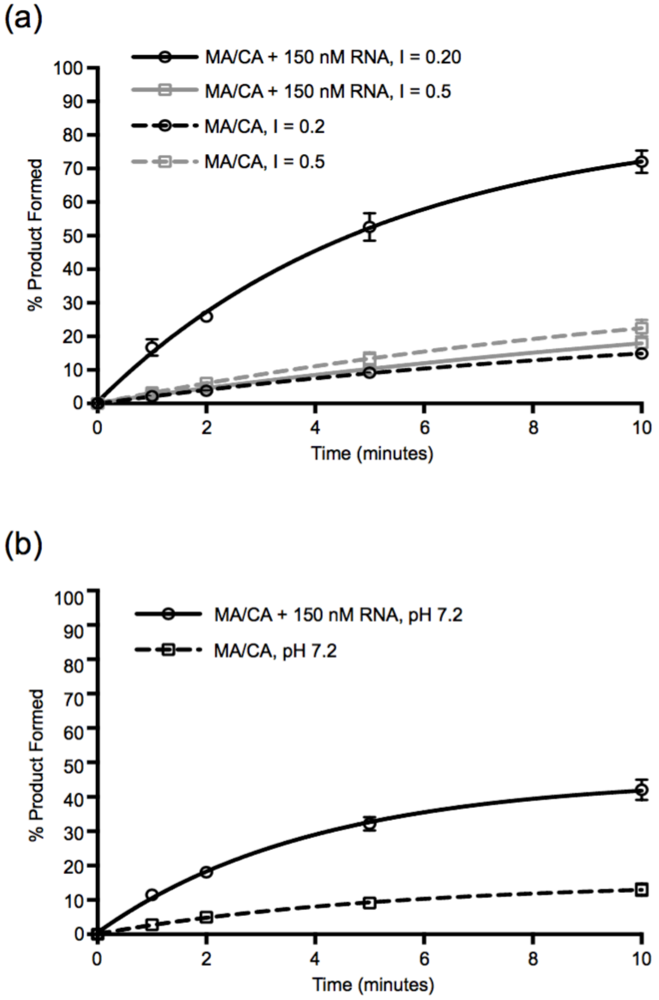

Figure 2.11: The ionic strength and pH of the reaction affect the potency of nucleic acid-dependent enhancement of HIV-1 PR activity. The standard reaction buffer was supplemented with (a) sodium chloride to raise the ionic strength or (b) with sodium hydroxide to increase the pH to 7.2. The ability of long, heteropolymeric RNA to enhance cleavage of MA/CA by the HIV-1 PR was then evaluated. The resulting reaction progress curves are shown. All errors bars represent the standard deviation resulting from three independent experiments.

7. RNA-dependent enhancement is not the result of a change in HIV-1 PR monomer-dimer equilibrium.

We attempted to discern the mechanism by which RNA and other polyanions enhance HIV-1 PR catalytic activity. As the HIV-1 PR is a non-tethered dimer, its monomeric and dimeric forms exist in a state of equilibrium (242). We hypothesized that RNA may shift the equilibrium by stabilizing or promoting the dimeric form of the PR, effectively increasing the number of active PR molecules in the reaction. Accordingly, tethering the dimer together should abrogate the RNA-enhancement phenotype. We generated an HIV-1 PR dimer where the C terminus of one monomer is tethered to the N terminus of the second monomer by a flexible five amino acid linker. After controlling for the number of active sites present in the proteolysis reactions, we generated a dose-response curve for the tethered dimer with p15 RNA as the enhancer (Figure 2.12). Compared to the control, no differences were observed between the tethered dimer and wild type PRs. RNA accelerated the rate of both reactions by more than 80-fold, and with similar EC50 values (36 nM for wild type, 35 nM for the tethered dimer). Therefore, RNA-dependent enhancement does not result from promoting dimeric interactions between HIV-1 PR monomers.

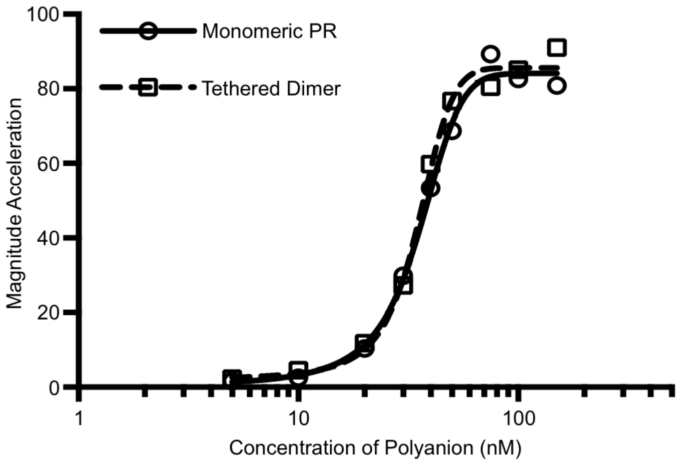

Figure 2.12: RNA acts on the dimeric form of the HIV-1 PR to accelerate MA/CA processing. Dose response curves were created to compare the ability of RNA to enhance the activity of the monomeric HIV-1 PR (circle) and a tethered dimer of the HIV-1 PR (square). Reactions were performed with the globular MA/CA substrate under standard conditions except that the concentration of the tethered PR dimer was reduced to reflect its pre-existing dimerized state.

8. The HIV-1 PR-RNA interaction lowers the K_m and increases the V_{max} of the proteolysis reaction.

In order to determine the effect of RNA on the enzymatic parameters of the PR, we used the fluorogenic peptide substrate to generate Michaelis-Menten plots and determined the effect of RNA on K_m and V_{max}. In the presence of RNA, K_m decreased by almost 4-fold, while V_{max} increased by 3-fold (Figure 2.13 and Table 2.4). The lower K_m indicates that RNA increases the affinity of the HIV-1 PR for its substrates, while the higher V_{max} demonstrates that RNA also increases the rate of the catalysis step. Using V_{max} as a surrogate for k_{cat} we calculated the relative specificity constant (k_{cat}/K_m) for the enzyme with and without RNA present finding that RNA increases the relative k_{cat}/K_m for the peptide reaction by an order of magnitude (Table 2.4). Thus the HIV-1 PR is 10-fold more efficient at cleaving the peptide substrate when interacting with RNA.

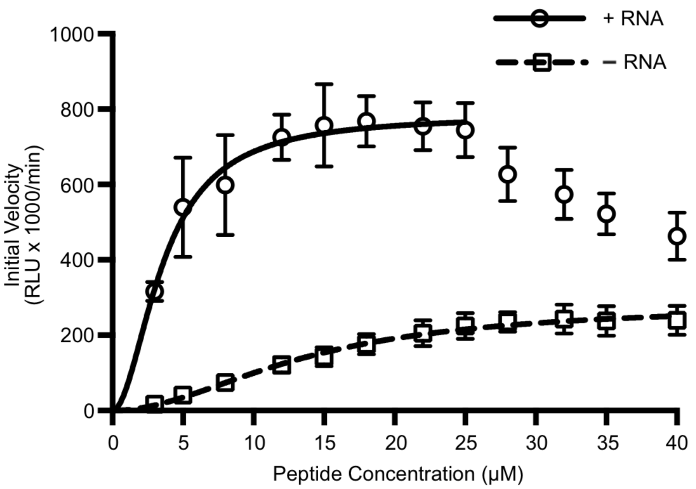

Figure 2.13: Both the affinity of the HIV-1 PR for a peptide substrate and reaction turnover number increase in the presence of RNA. Peptide proteolysis reactions were prepared where the concentration of peptide was varied from 3 to 40 μM. Reaction progress curves were generated for each reaction. The initial velocity of each reaction was determined from the reaction progress curves and plotted as a function of peptide concentration in the absence (square) or presence (circle) of 400 nM long heteropolymeric RNA. At higher peptide concentrations, substrate began outcompeting the HIV-1 PR for binding to the RNA, resulting in reduced initial velocity values. These data points (27 μM and above) were excluded from the plus-RNA curve in the calculation of Km and Vmax. Each point on the curve was calculated twice, with each calculation the result of reactions run in triplicate (i.e. six total reactions per point). Error bars represent the difference in the pair of calculated initial velocity measurements.

Table 2.4: Enzymatic parameters of the HIV-1 PR for processing of a peptide substrate.

	K_m (μM)	V_{max} (RLU/min)	Relative k_{cat}/K_m
Without RNA	13.6	282,000	1x
With RNA	3.6	784,000	10.5x

There were two other notable features of the plots. First, once the peptide substrate concentration reached 25 μM, the initial velocity of the reactions containing RNA began to decline, eventually converging with the minus-RNA curve. The peptide concentration where this reduction began changed depending upon the amount of RNA present in the reaction (data not shown), suggesting that the reduction in effect happened because the substrate was outcompeting the PR for binding to the RNA at these high concentrations. These data points were excluded for the determination of K_m and V_{max} in the presence of RNA. Second, despite the HIV-1 PR having only a single active site, the Hill coefficient was equal to 1.9 for the minus-RNA plot. The hill coefficient for the plus-RNA curve was calculated to be 1.9 as well; however, because of the low K_m, the rapid loss of steady-state conditions at low starting peptide concentrations, and background levels of fluorescence, we did not have enough data points below the K_m to show this with confidence. We did attempt to gather more data by lowering the PR concentration, and in those experiments still found a value for the Hill coefficient to be greater than one (data not shown), suggesting the Hill coefficient is greater than one irrespective of whether RNA is present. We cannot explain this result, though sigmoidal Michaelis-Menten curves can be observed in the absence of cooperativity (243).

C. Discussion

Converting a nascent HIV-1 particle into a mature infectious virion requires the viral PR to cleave the Gag and Gag-Pro-Pol polyproteins in a precise order (43-45). Given the complexity of this process, numerous regulatory mechanisms exist to direct the PR towards specific processing sites at various phases of maturation. These determinants of cleavage include processing site amino acid sequence (47-49), the local structural context (50, 228-230), and cofactors such as RNA or DNA (134, 220, 221, 229). An effect of RNA or other nucleic acids on processing rate has previously been reported only for cleavage sites nearby NC (134, 220, 221, 229), yet we found RNA accelerated the cleavage of substrates completely independent from NC. Moreover, accelerated processing of the MA/CA-AAA and peptide substrates indicated a substrate-RNA interaction was not required. We hypothesized that an enzyme-RNA interaction could enhance PR activity, and found the HIV-1 PR capable of interacting with nucleic acid in a gel-shift assay. The HIV-2 PR lacked this ability, and did not cleave its substrate more efficiently in the presence of RNA, providing corollary evidence. Interactions between the HIV-1 PR and RNA are primarily electrostatic in nature rather than sequence specific, though some additional prerequisites for the polyanion may exist. Mechanistically, RNA both increases the affinity of the HIV-1 PR for its substrates, and accelerates reaction turnover.

Proteolysis reactions that included RNA progressed more rapidly than those without RNA for every substrate tested for cleavage by an HIV-1 PR. This finding is in contrast with the original report, which found the rate of only p15NC processing changed in the presence of RNA (220). Experimental differences could have contributed to overlooking RNA as a general enhancer. In the previous work, most of the assays were limited to the single substrate p15NC, and RNA needed to be removed from the reactions rather than added as a supplement. This carries the inherent risk that some RNA may have remained in the RNA-free reactions. Very low concentrations of long heteropolymeric RNA were sufficient to achieve enhancement (Figure 2.8), so RNA removal would have needed to be exhaustive. Our two-substrate procedure improved upon these limitations by ensuring the reaction conditions were exactly the same for both p15NC and MA/CA substrates, and by precisely controlling the amount of RNA added to the reaction. We also suspect differences in substrates may have impacted the previous results.

Though our MA/CA contained the whole of CA, an unusual truncation of CA within the N-terminal domain (at amino acid 78) was employed previously. This truncation may have contributed by altering the fold of the MA/CA protein in such a way that it affected its ability to serve as a substrate. Thus, our results are consistent with most of the experiments that identified RNA as an effector of p15NC processing. However, our interpretation of these results is very different in that we find the enhancing effect to require an interaction between RNA and the PR rather than RNA and the substrate.

A more recent publication found the rate of SP1/NC processing increased in the presence of a single-stranded DNA molecule (229). The authors attributed this result to an increased accessibility of the SP1/NC processing site after NC bound the target nucleic acid molecule. Importantly, enhanced SP1/NC processing occurred in the absence of accelerated MA/CA or CA/SP1 cleavage. Such a result would argue in favor of RNA selectively enhancing cleavage of the sites nearby NC. There are several significant differences between this study and ours. In order to prevent aggregation of their substrate, the reactions were performed under high salt conditions (300 mM NaCl). We (Figure 2.11) and others (233) have demonstrated that nucleic acid-dependent enhancement of PR activity is tempered by increasing the ionic strength of the reaction. Also, they used a much higher substrate concentration and a different form of substrate, which could also affect cleavage site accessibility. Finally, though we did not directly examine the possibility of site-to-site variability in effect, we note that a modest difference in the magnitude of RNA-dependent enhancement was observed for p15NC and MA/CA in the two-substrate assay. While we have shown a robust effect of RNA on HIV-1 PR activity, it is clear that understanding, and untangling, the effect of RNA on PR and on substrate requires further exploration.

The ability of the HIV-1 PR to bind nucleic acid and have this interaction regulate its catalytic efficiency is not without precedent. Three other viral proteases have been identified that use RNA or DNA as a regulator. The human adenovirus proteinase (AVP) exhibits extremely poor functionality on its own, requiring an 11-amino acid peptide and a non-specific interaction with DNA or other polyanions to achieve its maximal activity (244-246). Prototype foamy virus (PFV) PR utilizes a specific sequence in the PFV genome to facilitate its dimerization and activation (247). In addition, the hepatitis C virus nonstructural 3 protein contains a serine protease domain (NS3^{Pro}) that can directly bind nucleic acid (248, 249). In contrast to the AVP, PFV PR, and HIV-1 PR, this interaction negatively regulates NS3^{Pro} activity (249). Thus, enzymes from very different virus families have been identified that can use nucleic acid as an interacting partner.

In addition to our own work, an interaction between HIV-1 PR and RNA has been suggested by one other study (233), though this report used extremely low ionic strength and low pH conditions. These conditions may have promoted artificial interactions, since they found both polyanions and polycations to be capable of accelerating HIV-1 PR activity; in contrast, we found polycations to be ineffective (data not shown). One additional difference in our results concerns poly(rA), which was reported to enhance PR activity, but we also found ineffective as poly(dA). A possible explanation could be that poly(dA) (250) and poly(rA) (251, 252) exist in different structural states at acidic versus neutral pH. Regardless, we demonstrate a functional interaction between the HIV-1 PR and heteropolymeric RNA can occur in environments with ionic strength and pH conditions likely to be encountered *in vivo*.

Whether a productive interaction between the HIV-1 PR and RNA actually occurs *in vivo* remains unknown, however. While acknowledging that *in vitro* experiments cannot recreate the complex environment within an actual virus particle, in our assays with p15NC the NC-to-PR dimer ratio was 130:1 and the NC-to-nucleotide ratio was 1:8; the latter ratio is noteworthy because the footprint of NC is one molecule per eight nucleotides (253, 254). Thus, in the presence of enough NC to entirely coat the available RNA, and in substrate-to-enzyme conditions that exceed the ratio of NC-to-PR dimer in virus particles (between 20:1 and 40:1), RNA-dependent enhancement was observed. These results would support the possibility that the interaction can occur *in vivo*. On the other hand, the enhancement effect was reduced when the reaction pH was raised from 6.5 to 7.2 (Figure 2.11b), implying the effect might be much more limited than what we observed in our reactions. Despite this potential reduced significance, an important role for the interaction cannot yet be ruled out because the pH at the site of virion biogenesis remains undetermined. Of note, the pH optimum for globular substrates appears to be slightly below neutral (50). Altogether, the currently available information is insufficient for determining whether the interaction between the HIV-1 PR and RNA has a biological role.

The HIV-2 PR was the sole enzyme examined that failed to process its substrate more efficiently in the presence of RNA, and failed to interact with nucleic acid in the gel-shift assay. The HIV-1 and HIV-2 PRs have very similar structures (255-257), so the lack of interaction is probably not structural. The more likely explanation is that the negative charge of the HIV-2 PR (pI = 5.3) prevents electrostatic interactions with RNA. Visualizing the electrostatic potential of the HIV-1 PR and the HIV-2 PR reveals that the HIV-2 PR has fewer positively charged regions on its surface, especially in the flap regions, which would minimize potential interaction sites for polyanions such as RNA (Figure 2.14). Though the HIV-2 PR was the exception among the enzymes we examined, it is not the only retroviral PR with a low isoelectric point. Comparing the isoelectric points of 31 primate lentiviruses, a majority of the HIV and SIV strains resembled the HIV-1 Group M PRs, but almost half had neutral or acidic PRs (data not shown). Orthoretrovirus PRs in general also demonstrate variability in charge, though our limited comparison does not rule out the potential for conservation within specific genera (data not shown). Regardless, the absence of charge conservation among primate lentiviruses implies a functional interaction *in vivo* is not required in all settings, and adds further weight to the argument against a biological role for the interaction between the HIV-1 PR and RNA. That the HIV-1 and HIV-2 PRs have similar catalytic properties with peptides in the absence of RNA (215, 258, 259) also supports this interpretation. Nonetheless, this information is still insufficient for determining whether an interaction between the HIV-1 PR and RNA actually occurs *in vivo*, as RNA could regulate Gag processing in different ways for different retroviruses.

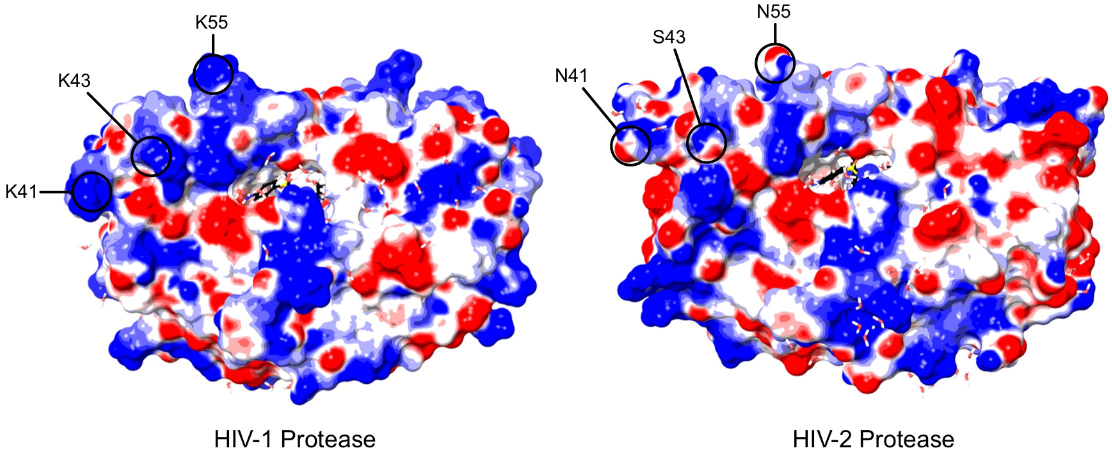

Figure 2.14: Electrostatic potential of the HIV-1 and HIV-2 PRs. Positively charged regions on the surface of the HIV-1 PR (left, PDB: 1T3R) and HIV-2 PR (right, PDB: 3EBZ) are shown in blue; negatively charged regions are shown in red. Both structures were generated in the presence of darunavir. The flap region of the HIV-1 PR appears to have a basic profile, while the flaps of the HIV-2 PR are of a mixed composition. Highlighted are three amino acid positions (41, 43, and 55) involved in binding interactions with putative non-active site inhibitors of the HIV-1 PR, which carry a positive charge in the HIV-1 PR but not the HIV-2 PR.

All long, heteropolymeric RNAs were equally effective as enhancers, suggesting all the RNAs contained a critical sequence and/or structure, or that neither were required. Because no small molecule was equally potent to the long RNAs, yet some could still enhance PR activity, a sequence requirement is unlikely. Most of the short DNA molecules that improved PR function were G-rich, suggesting a G-quadruplex structure could have been necessary. However, the successful enhancement of proteolysis by poly(dC), poly(dT), and heparin makes it unlikely a specific structure is required. Poly(dC) and poly(dT) also accelerated processing to an equivalent extent as poly(dG), implying no nucleotide was preferred either. This leaves electrostatic interactions as the primary means of interaction between enzyme and nucleic acid.

Though an electrostatic attraction appears to be the key requirement, additional determinants also exist that modulate the effectiveness of the interaction. Since the magnitude of enhancement plateaued at lower levels despite higher concentrations of smaller polyanions, simply saturating the binding sites on the HIV-1 PR is not sufficient for achieving a maximum magnitude of the effect. Consequently, the size or length of the polyanion must also be important. Poly(dA) and yeast tRNA point to one other additional requirement, as they were exceptions to this conclusion. Both of these nucleic acids are more rigid than other transcripts: tRNA due to base pairing and base modifications (260), and poly(dA) due to strong base-stacking interactions (261, 262). This suggests that the polyanion must have flexibility to serve as an efficient cofactor, in addition to being of sufficient length.

RNA enhances the catalytic activity of PFV PR by promoting its dimerization (247), so we considered this possibility for the mechanism of HIV-1 PR enhancement. Since, the concentration of HIV-1 PR was 100 nM in our reactions, which is well above the K_d of 6.8 nM determined for similar conditions of pH, ionic strength, and temperature (242), and the estimated half-life of an HIV-1 PR dimer is approximately 30 minutes (242), most of the PR was dimeric throughout the assay regardless of the availability of RNA. Therefore, acceleration by RNA was likely occurring with already intact PR dimers. The tethered dimer, which does not dissociate like the wild-type PR (263), confirms this conclusion because we saw a nearly identical effect of RNA on tethered dimer activity. Additionally, for a particle with a diameter of 120 nm and 120-240 Gag-Pro-Pol molecules, the concentration of the monomeric PR is 220-440 μM, well above the dissociation constant of the PR even while embedded in Gag-Pro-Pol (~680 nM) (36). The monomer-dimer equilibrium should therefore heavily favor the dimeric species in virus particles. Although we have not examined an independent effect on dimerization of the PR, the ability of RNA to enhance PR activity appears unlikely to be the related to dimerization.

RNA increased both the affinity of the HIV-1 PR for a peptide substrate (K_m) and its molecular activity (k_{cat}), collectively increasing the catalytic efficiency of the HIV-1 PR by an order of magnitude for the peptide reaction. Considering RNA affected peptide proteolysis less than cleavage of the globular MA/CA protein, the change in the specificity constant for reactions with the globular substrates would likely be even more substantial. These data do not, however, illuminate the precise mechanistic explanation for RNA-dependent enhancement. As there is a direct interaction, and it does not interfere with substrate binding, RNA more likely interacts with a secondary binding site(s) on the PR. The consistent inferiority in effectiveness of short nucleic acids compared to long heteropolymeric RNA regardless of concentration additionally implies that the polyanion must affect the PR in some way beyond simply saturating the binding site(s) on the PR. Putative allosteric sites have been identified within the flap/hinge region of the PR by means of small molecules (264-266) and existing PR inhibitors (257, 267). As the flaps seem to have key roles in both substrate binding and catalysis (268, 269), it is possible to speculate that RNA interacts with the PR flaps to facilitate the many conformational rearrangements this highly dynamic region must undergo (270-273). Of note, the flap regions of the HIV-1 PR (residues 37-61) contain a trio of basic amino acids (R/K41, K43 (264), and K55 (264, 267)) that are uncharged in the HIV-2 PR (Figure 2.14), and these same residues were identified as a key part of at least some binding interactions.

In summary, we have found that the HIV-1 PR interacts directly with nucleic acid, and this interaction drives the accelerated rate of processing observed for p15NC and other substrates. No specific RNA sequence or structure is necessary for it to serve as an enhancer, but larger and more flexible polyanions are more effective. Though the exact mechanism by which RNA improves the catalytic efficiency of the HIV-1 PR remains undetermined, the net effect on the enzyme is both an increase in substrate-binding affinity and an increase in turnover rate. These data suggest an allosteric binding site may exist on the HIV-1 PR, and argue in favor of viral genomic RNA being an additional regulator of HIV-1 PR activity during virion maturation.

D. Materials and Methods

1. Constructs

The MA/CA and p15NC regions were amplified by PCR from the pBARK plasmid, which contains the entirety of the *gag* and *pro* genes from NL4-3. Primers were designed to add a 6xHis tag to the N terminus of each protein, a termination codon at the C terminus, and flanking NdeI sites. Following digestion with NdeI, the PCR products were cloned into pET-30b (Novagen) to create pET-p15 and pET-MA/CAxTC. The pET-MA/CAxTC plasmid underwent an additional round of mutagenesis to introduce a tetracysteine motif (CCPGCC) in the Cyclophilin A binding loop (His87-Ala92) of CA and create pET-MA/CA. Two additional plasmids coding for the alternative MA/CA substrates, MA/CA-AAA and HIV-2 MA/CA, are derived from pET-MA/CA. For pET-MA/CAaaa, the nucleic acid sequence was altered to change MA amino acids $K_{26}KQYK_{30}$ to $A_{26}AQYA_{30}$. In the pET-MA/CA-HIV2 construct, the coding sequence for the cleavage site was altered from SQNY/PIVQ to the canonical HIV-2 MA/CA sequence of GGNY/PVQQ. The pET-GMCΔ construct was created as previously described (50). The PR region was also amplified out of pBARK and cloned into pET-30b, but without the addition of the 6x-His tag and termination codon, creating pET-PR.

2. Nucleic Acids

All long heteropolymeric RNAs were generated by *in vitro* transcription. The pET-MA/CA, pET-p15NC, and pET-PR plasmids were linearized with Eco*RV*, and purified with the Qiagen PCR purification kit. The MEGAscript T7 high yield transcription kit (Ambion) was utilized to generate RNA from the linearized DNA according to manufacturer's instructions. RNA was purified from the reactions with the Qiagen RNeasy kit and stored short-term in nuclease-free water at -20ºC. All short single-stranded DNA molecules were ordered from Sigma-Aldrich, and resuspended in nuclease-free water. Nucleic acid concentrations were determined with a NanoDrop spectrophotometer (Thermo Scientific).

3. Expression and Purification of Globular HIV-1 PR Substrates

Escherichia coli BL21 DE3 lysogens (Novagen) were transformed with plasmids coding for the p15NC, MA/CA, MA/CA-AAA, HIV-2 MA/CA, or GMCΔ proteins. Starter cultures were grown overnight in 2xYT media, and then used to inoculate MagicMedia (Invitrogen) for protein production. Expression cultures were grown for 8 hours at 37ºC and 225 rpm, before pelleting by centrifugation and freezing overnight at -80ºC. Pellets were resuspended in lysis buffer (TBS pH 7.5, 1% Triton X-100, 2 mM beta-mercaptoethanol) and lysed by sonication. Cellular debris was collected by centrifugation, and the resulting supernatant was applied to Ni-NTA Superflow columns (Qiagen) for purification of the His-tagged proteins by affinity chromatography. Purified proteins were concentrated using Vivaspin Concentrators (GE Healthcare), and underwent buffer exchange into storage buffer (20 mM sodium acetate, 140 mM sodium chloride, 2 mM beta-mercaptoethanol, 10% glycerol, pH 6.5). Sample pH was confirmed using a micro-pH electrode (Thermo Scientific). Purified protein samples were tested for residual nucleic acid with a NanoDrop spectrophotometer (Thermo Scientific), and the levels were found to be negligible.

4. HIV-1 Proteases

Purified HIV-1 proteases were produced as described previously (50, 274). Oligonucleotides for the heavily mutated variants were designed and purchased. Briefly, HIV-1 protease variants were expressed from a pXC35 Escherichia coli plasmid vector. The cell pellets were lysed and the protease was retrieved from inclusion bodies with 100% glacial acetic acid. The protease was separated from higher molecular weight proteins by size-exclusion chromatography on a Sephadex G-75 column. The purified protein was refolded by rapid dilution into a 10-fold volume of 0.05 M sodium acetate buffer at pH 5.5, containing 10% glycerol, 5% ethylene glycol, and 5 mM dithiothreitol (refolding buffer). The tethered dimer gene construct coded for two copies of the HIV-1 monomer linked by the nucleotide sequence that codes for Gly-Gly-Ser-Ser-Gly with unique nucleotide sequences for each monomer (275, 276). The HIV-2 PR (258) was a generous gift from Dr. John M. Louis (NIH). The theoretical isoelectric points of the viral proteases were calculated using the online ExPASy pI/MW tool.

5. Two-substrate Proteolysis Reactions

Two-substrate proteolysis reactions were run in proteolysis buffer (50 mM sodium acetate, 50 mM NaMES, 100 mM Tris, 2 mM beta-mercaptoethanol, pH 6.5). Reactions were 150 μl in volume and pre-incubated at 30ºC for 1 hour before addition of the enzyme. The pre-incubation step was included for consistency, although the its primary role was to allow fluor binding in reactions that included the Lumio Green Reagent (Invitrogen). In the MA/CA and GMCΔ reactions, both substrates began the reaction at concentrations of 2.5 μM. In the MA/CA and p15NC reactions, the initial concentration of MA/CA was 2.5 μM, while the concentration of p15NC was raised to 10 μM due to its poor staining profile in the subsequent analysis. The HIV-1 PR was used at a concentration of 150 nM in the two-substrate assays. RNA was also 150 nM when present. Reaction pH was confirmed as 6.5 using a micro-pH electrode (Thermo Scientific) after the final time point had been collected, and was unaffected by the presence of RNA. To collect time points, 12 μl aliquots were removed from the reactions at the indicated times and added directly to SDS to quench the reaction. The zero minute time point was removed immediately prior to the addition of enzyme. Where applicable, RNA was pre-mixed into the reaction 5 minutes prior to the removal of the zero minute time point. The quenched aliquots were loaded directly into a precast 16% Tris-Glycine gel (Invitrogen), and the substrates and products were then separated by SDS-PAGE at 100V for 2.5 hours before staining with SimplyBlue Safestain (Invitrogen). Band intensities were quantified with molecular imaging software (Carestream), and results were reported as the percent substrate remaining.

6. Single-substrate Proteolysis Reactions

All presented single-substrate proteolysis reactions with globular proteins were run in the proteolysis buffer. For the indicated reactions, the ionic strength was raised to 0.2 M and 0.5 M by the addition of sodium chloride to a final concentration of 50 mM and 350 mM, respectively. Select reactions for data not shown were performed in intracellular buffer (76.6 mM monopotassium phosphate, 60 mM potassium hydroxide, 12 mM sodium bicarbonate, 2.4 mM potassium chloride, 0.8 mM magnesium chloride, pH 7.2). Reaction pH was confirmed after collection of the final time point by a micropH electrode (Thermo Scientific). The final concentrations of MA/CA, MA/CA-AAA, and HIV-2 MA/CA were 2 μM. Reaction mixtures additionally included the Lumio Green Reagent (Invitrogen) to a final concentration of 2.5 μM, and were pre-incubated at 30ºC for one hour prior to initiating proteolysis. The concentration of RNA, where not directly stated, was 150 nM. Heparin and spermine were acquired from Sigma. The HIV-1 and HIV-2 PRs were used at a concentration of 100 nM in the single-substrate assays. The tethered dimer of the HIV-1 PR was used at a concentration of 45 nM, an amount with activity equivalent to the monomeric PR in the absence of RNA. The concentrations of the remaining enzymes were adjusted so that approximately 10% of MA/CA was processed in the absence of RNA after ten minutes. Most of the other enzymes were used at concentrations similar to the HIV-1 PR (75-400 nM); VEG23 and VSL23 required concentrations of 2 μM, however. Time points were collected, quenched, and electrophoresed as for the two-substrate assays. The fluorescently labeled proteins were imaged with a Typhoon 9000 (GE Healthcare/Amersham Biosciences), and quantified by ImageQuant TL (GE Healthcare) software. Results were reported as the percent product formed. The initial rate of the reaction was determined using only the data points collected where the reaction was ≤10% complete, or was estimated based on the first non-zero data point collected.

7. Peptide Proteolysis Reactions

Peptide proteolysis reactions were run in the peptide buffer (100 mM sodium chloride, 30 mM sodium acetate, pH 4.8). The peptide utilized was HIV Protease Substrate 1 (Sigma), a 12-amino acid long peptide containing the canonical HIV-1 MA/CA cleavage site. Substrate master mixes and the PR master mix were aliquoted into separate wells of a 96-well half-area plate (Costar) and pre-incubated in the 30ºC reaction chamber for five minutes, during which time the background level of fluorescence was determined. A multi-channel pipet was used to simultaneously mix the HIV-1 PR into the substrate mixtures to a final concentration of 100 nM. Reactions were followed in real-time on an Envision MultiLabel Reader (PerkinElmer) for 10 minutes with time points collected every 20 seconds. When included in the reaction, the concentration of RNA was 400 nM. Reaction rates were calculated using the data points from only the first 10% of cleavage for each substrate concentration. To determine when 10% cleavage had occurred for each reaction, a standard curve was generated from 50 μM reactions that had been run to completion and diluted to various concentrations. The values were linear from background to the upper limit of detection, a range of 0.25 μM to 7 μM. All values necessary to follow the first 10% of each reaction fell within this range (0.3 μM – 4 μM).

8. Electrophoretic Mobility Shift Assays and Native DNA gels

Binding reactions were 10 μl in size and contained 100 ng of ODN17 (final concentration of ~2 μM). Protein was added to a set of reactions incrementally such that their concentration increased from 0 μM to 12 μM. Where the mixture of nucleic acid and protein was insufficient to reach full volume, storage buffer was used. The pH of the binding reactions was confirmed to be 6.5 by a micropH electrode. After five minutes at room temperature, 1 μl of High-Density TBE Sample Buffer (Novex) was added. Samples were then loaded into a precast 6% DNA retardation gel (Invitrogen), and electrophoresed at 100V for 35 minutes. Gels were stained with SYBR gold (Invitrogen) according to manufacturer's instructions, and viewed with molecular imaging software (Carestream). After three five-minute washes with deionized water, the gels were stained for protein with Sypro Ruby (Invitrogen), also according to manufacturer's instructions. Results were reported as percent band intensity.

Native DNA gels were prepared similarly to the gel-shift assays, but nuclease free water was used in place of storage buffer. Additionally, no protein was present in any sample, and gels were only stained with SYBR gold. Where applicable, aliquots of the stock solutions were heated to 90ºC for 10 minutes, then cooled rapidly by centrifugation prior to dilution. Diluting the samples before heating results in a lower retention of the higher molecular weight species.

CHAPTER III

DISSECTING THE SEQUENCE DIVERSITY OF HIV-1 PROTEASE PROCESSING SITES

A. Introduction

The production of infectious HIV-1 particles requires the virally-encoded PR to cleave the structural polyproteins Gag and Gag-Pro-Pol within a nascent virion. These polyproteins have the same first four N-terminal domains, MA, CA, SP1, and NC, but differ in their remainder. Whereas Gag, which makes up 90-95% of the structural proteins in an immature HIV-1 particle (7), includes a 16-amino acid SP2 region and the late domain p6, Gag-Pro-Pol contains a TF region, and the individual domains of the PR, RT, and IN enzymes. Complete cleavage of the structural proteins by itself does not guarantee the virus particle will be infectious. The order and dynamics with which the PR cleaves Gag and Gag-Pro-Pol also strongly affect the ability of a virus particle to infect a new cell (44, 161, 277, 278).

Given the importance of HIV-1 PR activity in the formation of mature virus particles, it serves as an exceptionally important target for antiretroviral drugs (52, 53, 65, 70, 71). Though a series of very potent PIs are currently in use, the error-prone polymerase activity of HIV-1 RT means drug-resistance is an ever-present danger (279). Thorough understanding of the specificity of the HIV-1 PR would aid in the development of increasingly potent PIs to help combat this problem. However, despite substantial effort, the determinants of HIV-1 PR specificity remain obscure. The principal deterrent is the absence of a consensus amino acid sequence for the HIV-1

PR. Instead, the PR recognizes a conserved molecular shape called the substrate envelope (48, 49). An extremely diverse set of sequences is capable of occupying this conformation (280), providing the major obstacle to delineating enzyme specificity. Beyond sequence recognition, additional factors also complicate understanding. Contextual determinants alter the order of processing from what it may be if it were based on sequence alone (50, 229, 230), and whether a putative interaction between the HIV-1 PR and nucleic acid affects specificity also remains unknown (219).

Aside from a handful of studies employing globular proteins (35, 46, 47, 50, 230) and a few others that attempted molecular modeling (281, 282), the studies investigating HIV-1 PR specificity have followed one of two approaches. In the first approach, the kinetic efficiencies of reactions with short, 6-12 amino acid-long peptides derived from retroviral cleavage sites are determined (215, 283-298). These efficiencies are compared in the attempt to identify the ideal amino acid within any given subsite. Aside from the relative ease of their synthesis, peptides have the advantage of removing most contextual determinants, theoretically making the rate of processing a direct reflection of the enzyme's preferences. However, a few factors confound interpretation of these results: pH and ionic strength conditions vary across studies, none of which recapitulate physiological conditions, and the inherent bias resulting from the use of a limited number of template sequences (often only one) potentially restricts generalizability.

The second approach utilizes bioinformatics. The common practice is to partition a dataset of amino acid sequences into two groups: sequences successfully cleaved by the HIV-1 PR, and sequences that are not cleaved. Following separation, a variety of statistical and machine learning algorithms are used to build a set of rules with the intention of predicting the category to which any given amino acid sequence belongs (299-313). While several of these approaches have been fairly successful in their task (reviewed in: (309)), simply categorizing sites on the basis of cleavability does not necessarily mean the models can discern between well- and poorly-cleaved substrates. To our knowledge, no such analysis has been published to date. Furthermore, the datasets used to derive the cleavage rules are often compilations of the peptide studies, subjecting the bioinformatics analyses to the limitations of those reports as well. A few of the latest attempts have addressed at least this issue (299, 305, 310-312) by expanding their datasets to include a recently published, proteome-derived peptide dataset (280). Proteolysis assays from this study still used low pH conditions, however.

In this report, we attempted to bridge the gap between these methodologies, while simultaneously addressing some of the limitations of the peptide assays. Rather than focus on a single template sequence, we mutagenized six of the HIV-1 Gag and Gag-Pro-Pol processing sites (MA/CA, CA/SP1, SP1/NC, SP2/p6, TF/PR, and RTH/IN. These sites and their derivatives were all housed in the same globular protein background to reduce the impact of varied context. We additionally reduced the contextual effects imposed by the globular protein construct through the introduction of glycine triplets on each side of the cleavage site. This set-up was engineered to judge specificity based as much as possible on processing site amino acid sequence alone, while still housing the cleavage site within a globular substrate. Use of a globular protein enabled the observation of efficient processing by the HIV-1 PR under near-physiological pH and ionic strength conditions (50). The use of an internal control protein as a reference point also allowed us to achieve tremendous consistency in our measurements across all of our reactions. The final dataset consisted of 81 observations (66 cleaved, 15 non-cleaved). We evaluated the effect of the mutations incurred within each site, and also use a series of statistical analyses to identify properties shared across all cleavage sites. In addition, we developed 1570 predictive models with statistical model-building software, and then evaluated them on out-of-sample datasets. We define a number of potentially important relationships between amino acids in each of the six cleavage sites examined, and suggest that predictive models capable of distinguishing between fast and slow sites cannot also distinguish between functional and non-functional cleavage sites.

B. Results

1. A two-substrate system enables the accurate measurement of the relative rates of cleavage for HIV-1 processing sites.

The HIV-1 PR recognizes a conserved structure rather than a particular amino acid sequence (48, 49), enabling the HIV-1 PR to cleave a highly diverse set of amino-acid combinations (280). To determine which sequences the HIV-1 PR preferentially cleaves, we transferred five other eight-amino acid HIV-1 processing sites derived from the NL4-3 or HXB2 strains into the linker region between the MA and CA domains of a globular protein construct. A length of eight amino acids was chosen since this amount conventionally defines a HIV-1 cleavage site (18, 48). The protein construct consisted of an N-terminal GST-tag, the full MA domain, and the N-terminal domain of CA (GMCΔ) (50, 219). In agreement with previously published results, the relative processing rate of the wild-type GMCΔ protein was approximately the same as a protein comprised of full MA and CA domains, when measured in a two-substrate assay (Figure 3.1, black bars). The relative processing rate for each of the alternative cleavage sites revealed which sites the HIV-1 PR preferred. This order was SP1/NC > MA/CA ~ RTH/IN ~ CA/SP1 > TF/PR ~ SP2/p6. Most sites exhibit a relative rate consistent with those determined for when each site is in its natural context (42, 44, 45). However, two of the sites, CA/SP1 and SP2/p6, are out of order relative to that seen previously, with their relative positions switched.

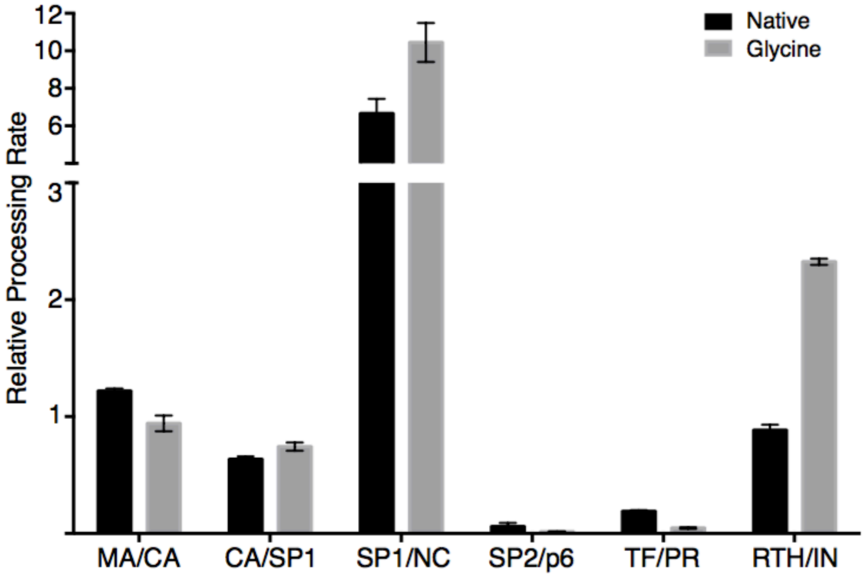

Figure 3.1: The order and relative rates of processing are mostly unaffected by the addition of glycine linkers. Each listed processing site was inserted into the linker region of the GMCΔ substrate. The cleavage rate for each substrate was measured relative to a wild-type version of the MACA substrate included within the reaction as an internal control (black). Glycine triplets were added to all substrates, including the internal control, and the relative rates were re-determined (gray). All reactions were performed at 30ºC in 50 mM NaMES, 50 mM sodium acetate, 100 mM Tris, pH 6.5.

We made additional modifications to the GMCΔ substrates to include glycine triplets on each side of the processing site. The glycines disrupted any local contextual determinants that might be affecting processing, leaving the efficiency of the reaction almost entirely dependent on site sequence. All substrates exhibited a decrease in processing rate due to the introduction of the glycines, though the magnitude of that effect varied depending upon the cleavage site (data not shown). The general decrease in processing rate suggests that for most of the sites the immediate context of the MA/CA site does not affect the rate of cleavage, with the glycines likely providing more mobility and then requiring a greater decrease in entropy when binding to the PR. Since the lack of glycines in the internal control made it a better substrate than nearly all of the processing sites, we switched to a new internal control that also included glycine insertions. When all samples were compared to this internal control, the new order of cleavage was SP1/NC > RTH/IN > MA/CA ~ CA/SP1 > TF/PR ~ SP2/p6 (Figure 3.1, gray bars). The only substrate whose position changed in the order determined when the glycines were absent was that of RTH/IN, which was now cleaved more efficiently relative to the other sites. The increase in the relative rate of cleavage of the RTH/IN indicates that there is a feature of the MA/CA site that is inhibitory to the cleavage of the RTH/IN sequence when placed in that context.

2. The relative rates of processing for mutant substrates differed by as much as 3000-fold.

The number of substrates was expanded to 81 with the addition of 75 mutant sites. Each of the mutants differed from one of the six natural cleavage sites by one or two amino acid substitutions (Table 3.1). We chose these substitutions based on the following criteria: appearance in published HIV-1 subtype B sequences (http://www.hiv.lanl.gov/), ease of generation (i.e. single nucleotide changes), and/or appearance in other HIV-1 processing sites (i.e. partial site exchange). We also had a general interest in choosing mutations that might improve processing rate. Eighteen of the 20 amino acids were included in the dataset, the two exceptions being tryptophan and cysteine. The introduced substitutions were not distributed equally between the *P4, P3, P2, P1, P1', P2', P3'*, and *P4'* positions, (Schechter and Berger nomenclature (55) – the scissile bond connects amino acids *P1* and *P1'*, the new C- and N-terminus, respectively; the integers increase in concert with added distance from the scissile bond), but all eight sites were represented.

Table 1: Cleavage sites and their rates.

Amino Acid Sequence	Relative Rate	log(Relative Rate)
SQVL/FLDG	15.630	1.19
ATIM/FQRG	13.894	1.14
ATIM/MQRG*	10.447	1.02
TTIM/MQRG	9.789	0.99
ATIM/LQRG	7.729	0.89
AAIM/MQRG	6.817	0.83
ATIF/MQRG	6.806	0.83
ATIM/IQRG	6.318	0.80
RQVL/FIDG	6.009	0.78
ATVM/MQRG	5.906	0.77
NTIM/MQRG	4.630	0.67
RQVL/FLDG	4.383	0.64
ARVF/LEAM	4.238	0.63
ARVL/FEAM	3.438	0.54
ATIL/MQRG	2.694	0.43
RKVL/FLDG*	2.325	0.37
ATIM/MQKG	2.166	0.34
RKVL/YLDG	1.930	0.29
SFNF/PQFT	1.924	0.28
SQNF/LIVQ	1.857	0.27
SQNY/LIVQ	1.696	0.23
REVL/FLDG	1.281	0.11
AQVL/AEAM	0.996	0.00
SQNF/PIVQ	0.987	-0.01
SQNY/PIVQ*	0.943	-0.03
SQNY/PIVE	0.926	-0.03
KKVL/FLDG	0.888	-0.05
ATVL/AEAM	0.796	-0.10
RKVL/FLDA	0.796	-0.10
PGNF/FQSR	0.760	-0.12
ARVL/AEAM*	0.743	-0.13
RGVL/FLDG	0.668	-0.18
RKVL/FLNG	0.535	-0.27
ATIM/IQKG	0.520	-0.28
ARIL/AEAM	0.435	-0.36
SFNF/PQIT	0.430	-0.37
ATIM/MIRG	0.384	-0.42
SRNY/PIVQ	0.380	-0.42
SENY/PIVQ	0.375	-0.43
SFSF/PQFT	0.347	-0.46
AKVL/AEAM	0.314	-0.50

Table 1: (cont)

Amino Acid Sequence	Relative Rate	log(Relative Rate)
RKVL/FLEG	0.310	-0.51
ATTM/MQRG	0.273	-0.56
RKVL/FLHG	0.229	-0.64
SHNY/PIVQ	0.225	-0.65
SQNF/LQSR	0.219	-0.66
RKVF/LLDG	0.153	-0.81
ATIM/MLRG	0.149	-0.83
PGNF/FQNR	0.140	-0.85
SQNY/AIVQ	0.112	-0.95
SGNF/LQSR	0.098	-1.01
RKVL/FLDR	0.097	-1.01
PGNF/LQSR*	0.065	-1.19
SFVF/PQIT	0.058	-1.24
PGNF/LQNR	0.053	-1.28
AGNF/LQSR	0.049	-1.31
SFSF/PQIT*	0.047	-1.33
RGNF/LQSR	0.029	-1.54
SLSF/PQIT	0.028	-1.56
SQSF/PQIT	0.019	-1.72
SFSF/PQVT	0.019	-1.72
PANF/LQSR	0.018	-1.74
PGNF/PQSR	0.016	-1.80
PGNY/LQSR	0.011	-1.95
SFGF/PQIT	0.006	-2.24
SKVF/PQIT	0.005	-2.30
SFSL/PQIT	0.000	*NA*
ARVL/PEAM	0.000	*NA*
ARVL/AQAM	0.000	*NA*
ARVL/AKAM	0.000	*NA*
RKNF/LQSR	0.000	*NA*
PQNF/LQSR	0.000	*NA*
PRNF/LQSR	0.000	*NA*
SKSF/PQIT	0.000	*NA*
RKVF/PLDG	0.000	*NA*
SQNI/PIVQ	0.000	*NA*
ARVI/AEAM	0.000	*NA*
ATII/MQRG	0.000	*NA*
PGNI/LQSR	0.000	*NA*
SFSI/PQIT	0.000	*NA*
RKVI/FLDG	0.000	*NA*

Most reactions were 15 minutes in length, approximately the length of time required for 50% of the internal control to be processed by the HIV-1 PR. If a substrate failed to break the limit of detection (2%), it was retested in an extended 2-hour assay. If the reaction failed to register above this detection limit in the extended assay, it was labeled as inactive. Of the 75 mutant cleavage sites, 60 exhibited detectable cleavage by the PR. Rates were calculated using only the data points where ≤10% processing had occurred, or were estimated from the first collected time-point. The difference between the fastest and slowest site was approximately 3000-fold (Figure 3.2). In order to improve interpretability, we present the data after log-transformation.

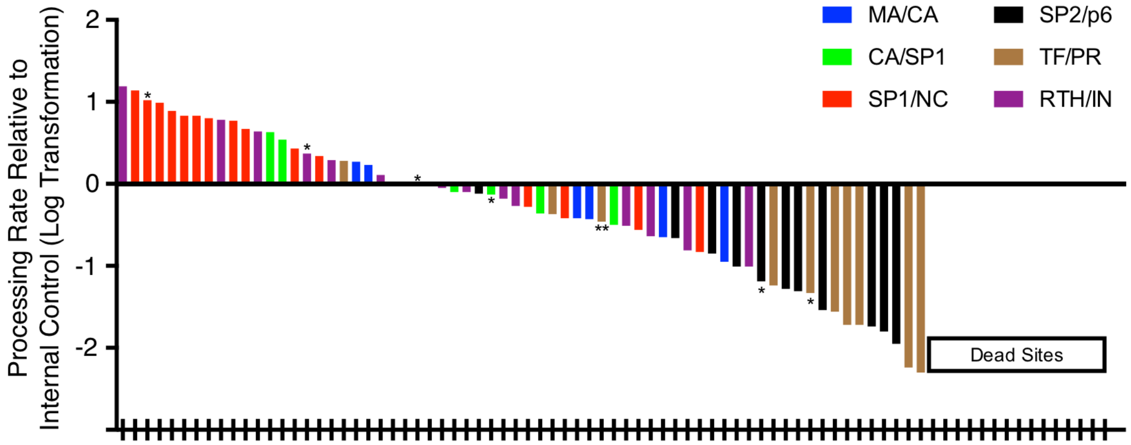

Figure 3.2: The best site was cleaved 3000-times faster than the worst functional site. The rate of each cleavage site is reported relative to the MA/CA internal control, after all values have been log-transformed. The sites are ordered from high-to-low, and are colored according to the natural HIV-1 processing site from which they were derived. Starred bars identify wild-type processing sites. The double starred bar marks the alternative TF/PR reference site. Reaction conditions were the same as figure 3.1: 30ºC in 50 mM NaMES, 50 mM sodium acetate, 100 mM Tris, pH 6.5. The x-axis labels were omitted because of a lack of readability. Site sequences are alternatively shown in Figure 3.3.

3. The MA/CA cleavage site (SQNY/PIVQ) was relatively tolerant to mutations.

The MA/CA site is the natural cleavage site of the GMCΔ construct. Its distinguishing feature is a proline in the *P1'* position, which takes part in a critical structural interaction following cleavage of the site (153-155). Proline is presumably selected at this site because it is required for this interaction, and would otherwise be a sub-optimal amino acid (although it is the *P1'* amino acid in other cleavage sites). We wondered if mutating the proline to the other amino acids found in the *P1'* position of HIV-1 processing sites might improve MA/CA processing. There were mixed effects, with the larger of those being negative. A substitution to leucine, a slightly larger and more hydrophobic amino acid, increased the rate of processing by approximately two-fold (Figure 3.2 top left). Meanwhile, replacing the proline with an alanine was highly detrimental. With the exception of one inactive mutant site, this alanine-containing site was the worst MA/CA site we examined. Its slowed cleavage by 10-fold. Alanine does seem to be a suitable *P1'* amino acid – it appears in the CA/SP1 site, which, independent of context, was cleaved almost equivalently to MA/CA (Figure 3.1) – suggesting that a proline in the MA/CA site actually provides an adequate, if not good, substrate for the HIV-1 PR. However, we have yet to test the effect of the larger *P1'* amino acids (phenylalanine and methionine) on MA/CA processing.

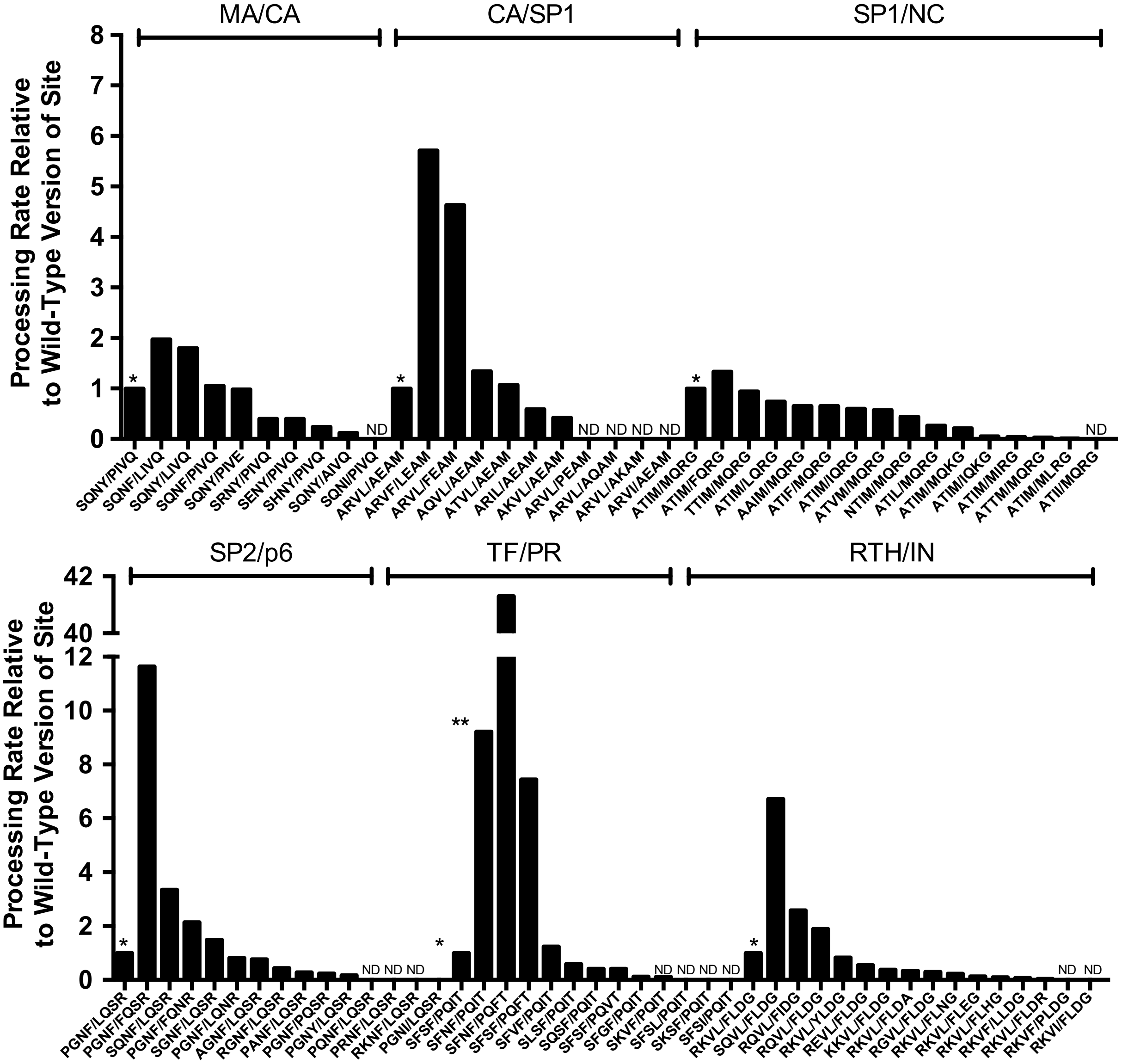

Figure 3.3: Few amino acid substitutions improved substrate processing over the natural site by more than two-fold. Each wild-type HIV-1 processing site was mutated in one or two locations and retested in the two-substrate assay. The cleavage sites are separated as indicated, MA/CA (top left), CA/SP1 (top middle), SP1/NC (top right), SP2/p6 (bottom left), TF/PR (bottom middle), RTH/IN (bottom right). ND = "not detected". Starred bars identify wild-type processing sites. The double starred bar marks the alternative TF/PR reference site. Reaction conditions were the same as figure 3.1: 30ºC in 50 mM NaMES, 50 mM sodium acetate, 100 mM Tris, pH 6.5.

Since charged amino acids occasionally appear in the *P3* position (and the primed counterpart *P3'*), we wondered whether a charged amino acid in the *P3* position would improve processing of the MA/CA site. To examine this, we replaced the wild-type glutamine with glutamic acid, arginine, and histidine. Each reduced the rate of cleavage. Both glutamic acid and arginine caused an approximately 2.5-fold decrease in rate, while the histidine-containing site was further reduced (5-fold loss in relative rate). Since the two-fold change in rate is not much different from the variability observed when simply repeating an assay (for the full dataset, the average was 1.25-fold, range 0- to 2-fold), it would be hard to argue either the glutamic acid or arginine substitution had a significant effect. Thus, our results agree with earlier published data (287-289, 314), which argues that the *P3* position is relatively tolerant to substitutions.

We also examined the effect of a conservative substitution (phenylalanine) for the *P1* tyrosine. No difference in the rate of cleavage was observed either when paired with the wild type *P1'* proline or with a *P1'* leucine substitution. This argues against an important role for the hydroxyl group on the tyrosine.

Only one of the tested amino acid substitutions produced an inactive MA/CA cleavage site. This change was in the *P1* position, but it was a substitution to a beta-branched amino acid, a *P1* substitution already known to block PR cleavage (29, 46, 47, 286).

Altogether, only two of the eight non-inactivating mutations we introduced had a greater than 2.5-fold effect on the rate of MA/CA site processing. We conclude from this that, despite the presence of a proline within the site, the MA/CA site is relatively tolerant to amino acid substitutions.

4. The rate of processing of the CA/SP1 site (ARVL/AEAM) is very sensitive to substitutions.

The CA/SP1 site is very poorly cleaved in its native context (44), but cleaved almost as efficiently as the MA/CA site in the GMCΔ substrate (Figure 3.1). This is despite having the smallest pair of amino acids occupying the two scissile bond positions (leucine/alanine). We asked whether this amino acid pairing was detrimental to the rate of processing by changing the L/A sequence to either L/F or F/L to mimic more conventional *P1/P1'* couplets (Figure 3.3, top middle). In both cases, we found the mutations increased the rate of processing by 5- to 6-fold. We also attempted to place a proline in the *P1'* position, but this cleavage site sequence was inactive. Therefore, the pairing of small amino acids in the *P1/P1'* positions is in fact limiting for processing. The key amino acid is almost certainly the *P1'* alanine, since changing that amino acid to a phenylalanine was sufficient to increase the rate of cleavage by 5-fold.

Since the *P3* position of MA/CA tolerated amino acid substitutions, we investigated whether this was a feature of additional cleavage sites. Two of the substitutions we made in the CA/SP1 *P3* position, arginine to glutamine or threonine, had practically no effect. These data agree with the conclusion that there is some tolerance for different amino acids in the *P3* position of this cleavage site as well. However, there was one noteworthy exception. A conservative substitution of an arginine to a lysine caused the most significant change among the *P3* substitutions in this cleavage site, more than a 3-fold decrease in processing rate. Based on the results of the RTH/IN (see below), we suspect this effect may be specific to the way a *P3* lysine impacts how the cleavage site interacts with the PR.

We tested a total of 10 variant CA/SP1 cleavage sites, four of which were inactive for cleavage. One of these was the aforementioned *P1'* mutation to proline, and another was a *P1* substitution to the beta-branched isoleucine. Introduction of a lysine into the *P2'* position created an inactive site, consistent with the absence of lysine in any of the central four positions (*P2, P1, P1', or P2'*) and with previous analyses (47, 289, 297, 302, 308, 310, 313). The most unexpected result was the change of the glutamic acid in the *P2'* position to glutamine; this conservative substitution produced a dramatic negative effect even though glutamine occurs in *P2'* of other HIV-1 cleavage sites (SP1/NC, SP2/p6, and TF/PR). This reveals that cleavage of the CA/SP1 site is highly dependent upon the interaction of the HIV-1 PR with the negative charge of the *P2'* glutamic acid residue. We suggest that it may be critical for the cleavage of the suboptimal leucine/alanine *P1/P1'* combination. Mutation of *P2'* to glutamine in the presence of the better leucine/phenylalanine *P1/P1'* pairing would provide an important test of this hypothesis.

In summary, the CA/SP1 site exhibited a considerable degree of diversity in response to amino acid substitutions, including several conservative substitutions. For example, the *P3 position* was relatively tolerant to a variety of substitutions, while *P2'* was incredibly sensitive to even highly conservative changes. The volatility in this position may result from having a suboptimal amino acid composition at the *P1/P1'* site of proteolytic cleavage. Testing the same mutations in leucine/alanine and leucine/phenylalanine *P1/P1'* sites would be advantageous in confirming this conclusion.

5. The wild-type SP1/NC site (ATIM/MQRG) is difficult to improve.

The processing site between the SP1 and NC domains is the most efficiently cleaved site in either Gag or Gag-Pro-Pol. Despite testing a number of substitutions, we found only a single substitution improved the rate of processing (Figure 3.3, top right). However, the increase in rate resulting from a change in the *P1'* position from a methionine to a phenylalanine was a modest

1.3-fold. Combined with the results of the CA/SP1 cleavage site, this modest improvement does point to a slight preference of the HIV-1 PR for an aromatic amino acid in the *P1'* position. A pair of alternative substitutions to the *P1'* amino acid, one to leucine and one to isoleucine, were both detrimental to processing rate. Still, the effects of these mutations were only modest reductions of 1.4- and 1.7-fold, respectively. While the current repertoire of substitutions suggests the P1' position of the SP1/NC site does not have a strong influence over the site's processing rate, we note that all mutations so far examined in this position are reasonably similar; each is a large hydrophobic amino acid. Additional tests with the smaller P1' occupants, i.e. proline and alanine, are warranted.

The three substitutions that had a significantly negative, though not inactivating, effect on cleavage all occurred in either the *P2* or the *P2'* position. Each caused at least a 25-fold decrease in processing rate. Two of these substitutions (*P2'* glutamine to isoleucine and to leucine) were to fairly dissimilar amino acids. Isoleucine does appear in the *P2'* position of other cleavage sites, so the unsuitability is in some way impacted by other amino acids in the SP1/NC cleavage site. One possibility is that, since the *P2* position naturally contains an isoleucine, both the *P2* and *P2'* sites cannot contain a beta-branched amino acid. Alternatively, since leucine was also harmful and is not beta-branched, it may be advantageous to have at least one of these amino acid positions be able to form a hydrogen bond with the PR. The third mutation with a negative effect on the rate of cleavage (40-fold) was a substitution of threonine for the *P2* isoleucine. A change of *P2* isoleucine to valine resulted in only a modest 2-fold rate reduction, suggesting the extremely harmful effect of threonine could be the hydroxyl group in its side chain. If this is true, then in conjunction with the *P2'* mutations, the results argue that combinations of like-and-like in *P2* and *P2'* of SP1/NC are disfavored. The optimal layout alternatively pairs a hydrophobic amino acid and a polar amino acid. Switching a polar amino acid into *P2* in conjunction with a hydrophobic amino acid in *P2'* will need to be tested to determine the accuracy of this conclusion. Lastly, just as for the MA/CA site, the only inactive site was the one in which the *P1* amino acid had been altered to isoleucine.

To recap, the SP1/NC site had only a single mutation improve its rate of processing with even a modest effect. All others resulted in a modest to severe loss in rate. The most notable substitutions were those introduced at the *P2* and *P2'* positions. These mutations suggested dissimilar pairs of amino acids in the P2 and P2' positions produce the most efficiently processed sites. Switching a polar amino acid into the SP1/NC P2 position in conjunction with a hydrophobic amino acid in P2' will need to be tested to determine the accuracy of this conclusion. Additionally, interchanging the SP1/NC amino acids into alternative sites will be an interesting test of whether the SP1/NC sequence houses optimal amino acids, or whether its impressive rate of processing results from the sum effect across the cleavage site.

6. The SP2/p6 site (PGNF/LQSR): selection against optimization?

Cleavage at the SP2/p6 processing site must occur prior to the assembly of the CA shell around the viral RNP core (137, 161). This means the SP2/p6 site should be processed ahead of the CA/SP1 site. However, these two processing sites show the opposite pattern (i.e. CA/SP1 > SP2/p6) when cleaved on the basis of sequence alone (Figure 3.1). Thus, there must be contextual determinants, including assembly itself, that impact the rates to make CA/SP1 one of the slowest sites. The role context must play in generating these rates during assembly can be seen by the substitution of leucine to phenylalanine in the *P1'* position, which allowed the SP2/p6 site to be cleaved 12-fold faster (Figure 3.3, bottom left), more in keeping with the observed relative rates of cleavage of these sites in the Gag protein. This is a very infrequent mutation – only 9 of 2254 SP2/p6 cleavage sites from an alignment of Gag/Gag-Pro-Pol contained phenylalanine in the *P1'* position (data not shown). After leucine, proline was the most common *P1'* amino acid, and that substitution in our data made the SP2/p6 processing site 4-fold worse. These results suggest the SP2/p6 site is under selection against optimization.

The otherwise most notable feature of the SP2/p6 site is the presence of an unusual *P4/P3* pairing, a proline and glycine respectively. This doublet is one of two cleavage sites to form an unusual conformation in which the *P4* amino acid actually occupies the position in the HIV-1 PR binding site that normally belongs to the *P3* amino acid (48). Given the only two other occurrences of a proline in a cleavage site are both associated with a critical structural role (153-155, 228) the presence of this proline may be similarly important. Its extremely high level of conservation further supports this conclusion (2247/2254 sequences contained a proline in the *P4* position). Additional evidence that this proline has some other as-yet-unexplained role comes from our mutation data, since mutating only the proline did not have a severe effect on processing rate. Mutating the proline to an alanine or arginine was not problematic, with the worst effect between these two being a 2-fold decrease in cleavage rate. Mutation to a serine actually improved the processing rate by 1.5-fold. Changing only the glycine, on the other hand, induced a considerably larger effect. While alanine could substitute for glycine, albeit ineffectively, both glutamine and arginine exchanges were not tolerated. It is likely that this lack of tolerance reflects the incompatibility of the *P4* proline with almost any other amino acid in the *P3* position, since mutation of both the *P4* and *P3* amino acids to the MA/CA-derived serine-glutamine doublet was favorable, increasing the rate of cleavage by 3.5-fold.

There were four inactive SP2/p6 sites among the mutant cleavage sites we tested. As with all other sites, an isoleucine in the *P1* position prevented cleavage. The other inactive sites were prefaced above, and resulted from a substitution of the glycine at the *P3* position while the proline was still present at *P4*. The final inactive site replaced the P4/P3 sequence with an arginine/lysine pair. While this does replace both the proline and glycine, the arginine/lysine pairing appears to require a beta-branched in the *P2* position (see RTH/IN below). Therefore, the defect in the arginine/lysine site is more likely due to the presence of an asparagine in the *P2* position than the loss of the glycine. This, of course, requires additional testing to confirm.

In conclusion, the SP2/p6 site appears to be under selection against the optimal amino acid sequence, most dramatically at the *P1'* position. Additionally, the proline in the *P4* position likely has a critical, as-yet-unexplained purpose in the virus lifecycle given its high level of conservation and lack of general importance to SP2/p6 processing rate. The glycine, meanwhile, compensates for the presence of the proline to enable processing by the HIV-1 PR.

7. The TF/PR site (SFSF/PQIT): a site with potential.

While we have observed the TF/PR as an intermolecular cleavage event, in the virus it is actually one of three cleavage events that must occur intramolecularly with an immature enzyme at least some of the time (36-39). This cleavage event is essential for releasing HIV-1 PR from the Gag-Pro-Pol precursor to achieve full enzymatic activity (35, 36, 40). To avoid premature activation of the enzyme, which can interfere with particle production (224, 225), this site might be expected to remain under some negative regulation. There is some evidence for this, though not necessarily as compelling as the evidence in the SP2/p6 site.

Two single amino acid changes allowed a 7- and 9-fold increase in rate (Figure 3.3, bottom middle). One of these was an isoleucine to phenylalanine change in the *P3'* position. Though this substitution did not occur a single time in the alignment (data not shown), its absence is almost certainly for a structural reason. The *P1'-P4'* amino acids are all part of the PR. After processing, these amino acids participate in a beta-sheet with the C-terminal amino acids of the PR, and it is the formation of this beta-sheet that gives the HIV-1 PR the conformational stability necessary for robust enzyme activity (40). It is possible that a phenylalanine in this position is not tolerated because it fails to adequately assemble this beta-sheet.

The other substitution to bring about a significant increase in processing rate is the serine to asparagine change in the *P2* position. While the serine is more frequent in the alignment (1232/2254), *P2* asparagine also occurs a substantial amount of the time (779/2254). Thus, there does not appear to be much of a restriction on this substitution from occurring. Considering the embedded HIV-1 PR only occupies the wild-type conformation an estimated 3-5% of the time (34), it's possible the immature PR active site might have a slightly different sequence preference than the mature version. Therefore, the 9-fold difference in rate we observe when cleaved *in trans* may not exist when TF/PR is cleaved intramolecularly. Alternatively, this sequence does overlap with the p6 domain, which could be where the selective pressure is being applied.

Among the mutated TF/PR sites tested, three resulted in an inactive site – two if we exclude the isoleucine substitution in the *P1* position. One of these substitutions was a lysine in the *P3* position, providing yet another example of the damaging effect of this mutation. We do note that accompanying the lysine with a valine in the *P2* position rescued the site to a very low level of cleavage, pointing toward a requirement for a beta-branched *P2* amino acid with a *P3* lysine. However, we note that this site just barely broke the limit of detection in the two-hour assay. The remaining site with an undetectable amount of processing altered the *P1* amino acid from phenylalanine to leucine. This paired a proline in the *P1'* position with a leucine in the *P1* position. The only other occurrence of the L/P doublet in our dataset was in the CA/SP1 site, and it too was defective.

In summary, the strongest conclusion that can be drawn from this data is that a serine in the *P2* position is not an optimal amino acid in the substrate for the mature HIV-1 PR. The larger asparagine provides a more favorable substrate, though whether this remains true for the embedded, immature PR is unknown. The effect of this mutation on p6 function must also be considered.

8. The RTH/IN site (RKVL/FLDG) has a suboptimal amino acid in the P3 position.

The RTH/IN site takes on an unusual conformation when bound to the PR. Like the SP2/p6 site, the *P4* amino acid occupies the position within the PR that is normally occupied by the *P3* amino acid (48). For this particular site though, the *P4* amino acid is arginine, and the *P3* amino acid is lysine. The presence of a basic amino acid in the *P3* position does not necessarily cause this unusual conformation, as the CA/SP1 site includes an arginine in the *P3* position but takes on the conventional processing site layout when bound to the PR. We suspect this effect results specifically from an odd, non-optimal interaction between the HIV-1 PR and the *P3* lysine. In support of this concept, all of the mutant RTH/IN sites that were better substrates for the PR did not have the lysine in them (Figure 3.3, bottom right). Additionally, the four instances in which we placed a lysine into the P3 position all had harmful effects. Two of these

(SKSF/PQIT and RKNF/PQSR) were completely inactive; one of them (SKVF/PQIT) was barely detectable even after extending the length of the cleavage assay, having been rescued by the inclusion of a beta-branched amino acid in the *P2* position; and the last was a conservative substitution in the CA/SP1 site (arginine to lysine) that still caused a 2.5-fold reduction in processing rate. Altogether, these results point toward lysine in the *P3* position as a sub-optimal amino acid.

Continuing to look at sites containing a lysine in the *P3* position, there was a definite pattern present in order to optimize a site with that amino acid. Prefaced in the last paragraph, lysine must be paired with a beta-branched amino acid to avoid inactivity. While it can be paired with an aromatic amino acid in the *P1*, these sites were poorer substrates than when the *P1* amino acid was leucine. For instance, flipping the RTH/IN *P1/P1'* amino acids so that a phenylalanine is in the *P1* position and a leucine is in the *P1'* position resulted in a 15-fold decrease in processing rate. We conclude that when lysine is in the *P3* position, the optimal amino acid arrangement in a site has a beta-branched amino acid at the *P2* position, and a non-aromatic residue at *P1*.

The mutated RTH/IN sites had only one inactive site among those tested, aside from the *P1* isoleucine. This site contained a phenylalanine in the *P1* position and a proline at *P1'*. The failure of this site likely results from the conflicting requirements of the proline and lysine. Lysine's preferences have just been documented. Proline prefers large aromatic amino acids in the *P1* position, and is compromised by a beta-branched amino acid at *P2* (288). Therefore, the central amino acids and external amino acids require opposite features, which makes the site uncleavable.

In summary, the *P3* lysine appears to be a problematic. We hypothesize that it is the key amino acid in creating the unusual conformation the RTH/IN site takes on in order to be recognized by the HIV-1 PR as substrate. For the optimal site including a *P3* lysine, a beta-branched amino acid in *P2* may be an absolute requirement, while a non-aromatic amino acid in the *P1* site is preferred.

9. Mixed-effects modeling enabled the identification of patterns and important predictors common to all cleavage sites.

While simple substitution allowed us to identify important patterns within individual cleavage sites, we sought also to identify patterns that permeated throughout all sites. Because each of the original sites was not cleaved at the same rate, the rate for each cleavage site tended to group with other sites derived from the same natural HIV-1 processing site (Figure 3.2). In order to account for these baseline differences in rate when considering the effect of a mutation, we used linear-mixed effects modeling. In this modeling procedure, observations are not considered independent, but are grouped categorically by original cleavage site. We confirmed that this was an important distinction by comparing the null standard ordinary least squares model with the null linear mixed-effects model (different only in the inclusion of the categorical separation of observations). A likelihood ratio test to compare the effectiveness of the two models at describing the data found the mixed-effects model was significantly better at doing so ($\chi^2(1) = 32.95$, $p < 0.0001$).

Our next step was to define each cleavage site by a specific set of criteria. Eight physicochemical properties were chosen (Table 3.2). The first model we developed considered each site as a single unit and did not concern itself with where any mutation was made. When all variables were included in this model, the residual distribution did not meet the requirements of the normality assumption (Figure 3.4A). This was indicative of a missing term from the model, likely that of an interaction.

Table 3.2: Amino acid scales utilized in model building.

Amino Acid		Hydrophobicity[1] (kcal/mol)	Nonpolar ASA[2] (Å^2)	Polar ASA[2] (Å^2)	Polarizability[3] (a.u.)	van der Waals Volume[4] (Å^3)	Molecular Weight (Da)	Side Chain Length[#] (Å)	Charge at pH 6.5
Alanine	A	0.17	67.2	0	55.86	90	89	3.3	0
Arginine	R	0.81	71.1	122.3	115.57	194	174	9.2	1
Asparagine	N	0.42	31	81.5	79.77	124.7	132	5.7	0
Aspartic Acid	D	1.23	32.6	73.2	76.35	117.3	133	5.12	-1
Cysteine	C	-0.24	92.9	0	74.96	113.7	121	4.9	0
Glutamine	Q	0.58	50.6	88.5	91.22	149.4	146	7	0
Glutamic Acid	E	2.02	52.4	79.3	90.42	142.2	147	6.32	-1
Glycine	G	0.01	37.4	0	44.25	64.9	75	2.3	0
Histidine	H	0.17	100.6	42.8	102.59	160	155	6.6	0.24*
Isoleucine	I	-0.31	133	0	95.24	163.9	131	5.8	0
Leucine	L	-0.56	137	0	94.49	164	131	5.8	0
Lysine	K	0.99	98.6	67.4	101.2	167.3	146	8.3	1
Methionine	M	-0.23	145.2	0	102.14	167	149	7.5	0
Phenylalanine	F	-1.13	164.1	0	122.9	191.9	165	7.3	0
Proline	P	0.45	98.7	0	73.49	122.9	115	4.5	0
Serine	S	0.13	43.4	35.4	61.24	95.4	105	4.5	0
Threonine	T	0.14	71.3	28.3	73.7	121.5	119	4.6	0
Tryptophan	W	-1.85	177	26	157.76	228.2	204	8.4	0
Tyrosine	Y	-0.94	130.6	49.4	129.6	197	181	7.92	0
Valine	V	0.07	110.2	0	81.49	139	117	4.6	0

The values for the side chain length were determined in PyMol. The distance between the Cα and furthest atom on the side chain was measured. One van der Waals radius for that atom was then added to the measurement, and this is the reported value.

* The theoretical ratio of neutral Histidine to charged Histidine at pH 6.5 is ~3:1. The charge for Histidine was set to 0.24 to reflect this ratio.

1. Reference 327; 2. Reference 328; 3. Reference 329; 4. Reference 330.

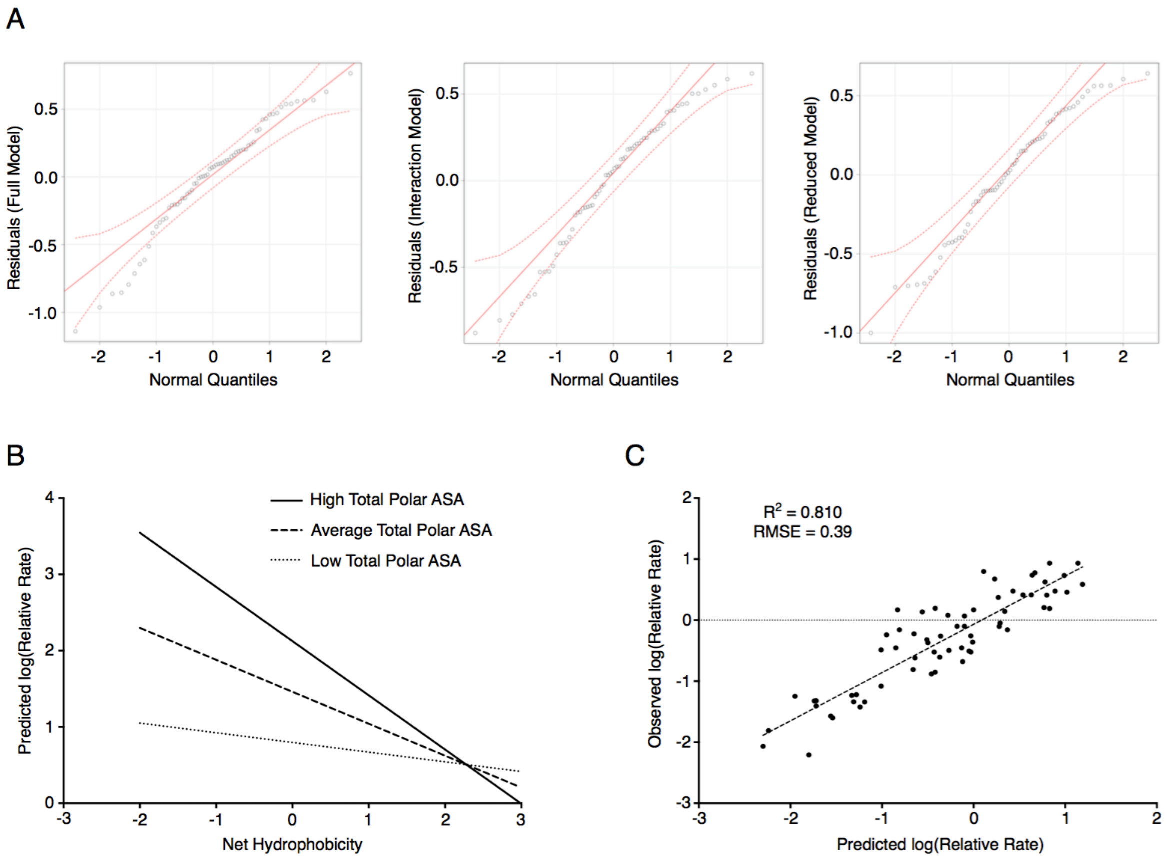

Figure 3.4: Mixed-effects modeling accounts for different baseline conditions to build a superior model than ordinary least squares analysis. (A) Quantile-quantile plots for the three stages of building a multiple-property mixed-effects model. The residuals for the full model without the interaction are plotted in the left panel, those for the full model with the interaction term in the central panel, and the final reduced model in the right panel. The solid red line represents the theoretical normal distribution of residuals. The dashed red lines define the 95% confidence envelope. (B) The theoretical relationship between the net hydrophobicity of a cleavage site and relative rate of cleavage when cleavage sites contain different amounts of polar accessible surface area. (C) A plot of the relative processing rates predicted by the reduced mixed-effects model versus the actual observed values.

After examining each possible two-way interaction among terms, one model was found to be significantly superior. The interaction included in this model suggested the total hydrophobicity of a site had varying effects on reaction rate when different amounts of polar surface area were present in the site (Figure 3.4b). In short, sites that were more hydrophobic but contained a higher total quantity of polar surface area produced faster sites. However, when sites were more hydrophilic, the amount of polar surface area did not matter.

We desired the model of maximum parsimony, and to obtain that model we performed a series of likelihood ratio tests to identify and remove non-significant terms. The final reduced model contained only five terms (Table 3.3). Its effectiveness at describing our dataset was not significantly different from this original model ($\chi^2(3) = 2.18$, $p = 0.5358$), but was still considerably superior to the null model ($\chi^2(5) = 46.41$, $p < 0.0001$). The root mean square error (RMSE) was 0.39 logs, meaning this model predicted most of the source data to within 2.5-fold of its observed value.

Table 3.3: Properties of the reduced full-site model.

	Coefficient	Chi Square (Df)	Significance
Intercept	-8.586 ± 2.029	18.01 (4)	$p = 0.0012$
Net Hydrophobicity	0.671 ± 0.285	7.40 (4)	$p = 0.1161$
Polar ASA	0.007 ± 0.002	15.00 (4)	$p = 0.0047$
Polarizability	0.010 ± 0.003	12.53 (4)	$p = 0.0138$
Net Charge	-0.738 ± 0.109	36.05 (4)	$p < 0.0001$
Hydrophobicity*Polar ASA	-0.004 ± 0.001	15.79 (4)	$p = 0.0033$

Coefficients are non-standardized.
ASA: Accessible Surface Area

10. Individual predictor models identified potential relationships between amino acid positions within cleavage sites.

Though effective, evaluating processing sites as a singular whole fails to identify key characteristics of specific positions or interactions between positions. Unfortunately, including all of the positional and interaction terms was not feasible. Such models would suffer from both over-fitting (where a model describes the noise in the data rather than the key relationships) and multicollinearity (where predictor variables are too highly correlated to separate their effects). Moreover, the categorical separation of cleavage sites further reduced the effective sample size, making the number of variables we could use before falling victim to over-fitting even smaller. Thus, to evaluate the prospective importance of variables on a position-by-position basis, we considered each of our eight physicochemical properties independently of the others.

Seven of eight criteria described the data significantly better than the null model, the lone exception being amino acid charge. Judging by RMSE, none of the individual models were as effective as the model built using all the criteria. Among the individual models, two interaction terms were identified. One reflected an interaction between *P2* and *P4'* in the hydrophobicity model ($\chi^2(1) = 7.40$, $p = 0.0065$). Processing sites with slightly hydrophilic *P2* amino acids were cleaved more efficiently when *P4'* was hydrophobic, though a variety of different amino acids could be accommodated in *P4'* when *P2* was hydrophobic (Figure 3.5A). The other interaction was between *P2* and *P2'* in the polarizability model ($\chi^2(1) = 12.2$, $p = 0.0005$). Here, sites containing *P2* amino acids with lower polarizability faired better when *P2'* was slightly higher in polarizability (Figure 3.5B). The sensitivity of this interaction to the change in polarizability in the *P2'* position was impressive. On a scale with a range of 113 atomic units (a.u.), a difference of only 5 a.u. in the P2' site could cause as much as a 100-fold effect.

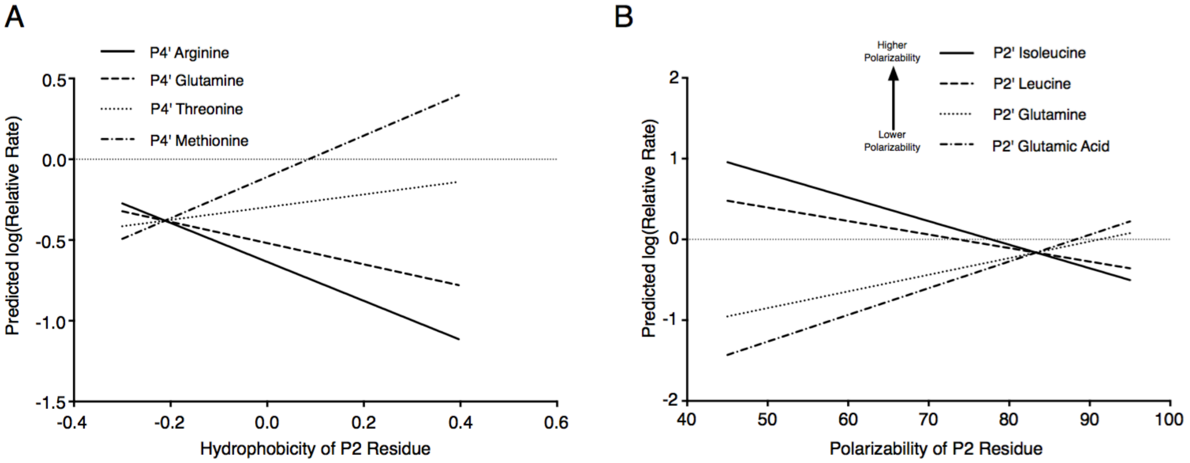

Figure 3.5: The theoretical relationships for the interactions identified within individual mixed-effects models. (A) A plot of the relationship between hydrophobicity of the P2 amino acid and processing rate when different amino acids occupy the P4' position. (B) A plot of the relationship between the polarizability of the P2 amino acid and processing rate when different amino acids are found in the P2' position.

The models were simplified by removing variables that were non-significant in likelihood ratio tests, and some patterns stood out in the variables that remained (Table 3.4). Three positions were removed from nearly all of the models: *P3*, *P1*, and *P3'*. The *P3* and *P3'* results are consistent with our earlier observations, with the exception of the *P3* lysine. The *P1* result, on the other hand, was not. We suspect *P1* was on this list not because of its lack of importance, but because only cleavage sites containing four very effective *P1* amino acids were included in our dataset. The positions on which rate was most dependent, according to these models, were *P2*, *P1'*, *P2'*, and *P4'*. The *P4* amino acid was important under only select circumstances, whereas these other four positions were present in nearly all models.

Table 3.4: Individual variable, reduced model coefficients.

Criteria	Intercept	P4	P3	P2	P1	P1'	P2'	P3'	P4'	Interaction
Hydrophobicity	-0.81 ± 0.48	-1.05 ± 0.47		**0.74 ± 0.57***		-0.79 ± 0.19	0.78 ± 0.29		**-1.85 ± 0.67**	5.13 ± 1.81
Molecular Weight	-15.8 ± 3.2			0.033 ± 0.008		0.018 ± 0.004	0.074 ± 0.020		-0.011 ± 0.004	
Polarizability	-222 ± 67	-0.015 ± 0.005		**2.67 ± 0.77**		0.019 ± 0.004	**2.40 ± 0.74**		-0.011 ± 0.005	-0.029 ± 0.008
Non-Polar ASA	-1.81 ± 0.54					0.012 ± 0.004				
Polar ASA	-0.58 ± 0.30			0.013 ± 0.005				0.008 ± 0.004	-0.013±0.004	
vdW Volume	8.42 ± 3.43					0.013 ± 0.003	-0.062 ± 0.022		-0.009±0.003	
Side Chain Length	-11.3 ± 2.1			0.54 ± 0.13		0.31 ± 0.07	1.01 ± 0.27			

*Non-significant term that was retained in the model due to its involvement in the interaction term.
Bolded terms are included in the interaction.

11. Building models for prediction.

Mixed-modeling procedures are effective tools for building a descriptive model. However, if we desire a model that may be applied to the prediction of out-of-sample cleavage sites, the categorical distinction between cleavage sites must be discarded and ordinary least squares methodologies applied. This is because all possible cleavage sites ($n = 20^8$) cannot be grouped into one of the six categories within our dataset. We explored a variety of model selection techniques, including best subsets and stepwise regression, ridge, elastic net, and lasso regression, and partial least squares regression. (Short descriptions of each are in the methods.) We continued to evaluate the processing sites on the basis of the eight criteria used in during the mixed-effects modeling procedures. A ninth criterion of total side chain surface area was also evaluated. We developed models where all eight amino acids were treated as one unit, and where each amino acid was considered individually. For the latter, we built models that included measurements from one, two, or three of the nine criteria (i.e. one, two, or three pieces of information for each of the eight amino acids). Models were built where interactions were excluded, and where they were included. In total, 1570 models were generated.

Approximately one-fifth (308/1570) had a better RMSE than that of the principal mixed effects model, though the first mixed-effects model was considerably more parsimonious (and therefore statistically preferred). All 15 of the top models used three criteria to evaluate the sites, included interaction terms, and made use of the stepwise selection procedure (Table 3.5). One concern for these models was the number of variables. We allowed up to 25 in the selection process; 11 of 15 made full use of that allotment, and all 15 of them contained more than 20. This points toward the models overfitting the data. The lack of consistency between the variables found in the mixed-effects models and the models built for prediction was also suggestive of a problematic fit. The positions deemed unimportant when the different sites were analyzed as part of groups (*P3* and *P3'*) were the very same positions that appeared frequently in the prediction models. For example, the *P3/P3'* interaction term was the most common element within these models (Figure 3.6).

Table 3.5: Characteristics of the top fifteen models for describing the source data, and for predicting out-of-sample rates.

Rank	Criteria			INT	Procedure	RMSE	RMSEoos
1	Hydrophobicity	Nonpolar ASA	Side Chain Length	Y	Stepwise	0.10	9.57
2	Molecular Weight	Nonpolar ASA	Side Chain Length	Y	Stepwise	0.12	3.25
3	Hydrophobicity	Molecular Weight	Nonpolar ASA	Y	Stepwise	0.12	2.92
4	Hydrophobicity	Polarizability	Nonpolar ASA	Y	Stepwise	0.12	2.45
5	Polar ASA	Total ASA	Side Chain Length	Y	Stepwise	0.12	11.2
6	Nonpolar ASA	Total ASA	Side Chain Length	Y	Stepwise	0.12	2.61
7	Hydrophobicity	Polar ASA	Side Chain Length	Y	Stepwise	0.12	2.91
8	Hydrophobicity	Nonpolar ASA	Polar ASA	Y	Stepwise	0.12	13.4
9	Hydrophobicity	Molecular Weight	Polar ASA	Y	Stepwise	0.13	9.26
10	Hydrophobicity	Nonpolar ASA	Total ASA	Y	Stepwise	0.13	1.82
11	Molecular Weight	Nonpolar ASA	Polar ASA	Y	Stepwise	0.13	3.18
12	Nonpolar ASA	Side Chain Length	van der Waals Volume	Y	Stepwise	0.13	4.29
13	Polarizability	Polar ASA	Total ASA	Y	Stepwise	0.13	6.28
14	Polarizability	Polar ASA	Side Chain Length	Y	Stepwise	0.13	3.31
15	Nonpolar ASA	Polar ASA	Side Chain Length	Y	Stepwise	0.13	3.73
1354	Polarizability	---	---	N	Ridge.o	0.57	0.96
1367	Polarizability	---	---	N	Elastic Net.o	0.58	0.96
1564	Full Site	---	---	N	OLS	0.94	0.98
1568	Full Site	---	---	N	Best Subset	0.96	1.00
1337	Polarizability	---	---	N	Lasso.m	0.56	1.00
1310	Polarizability	---	---	N	Ridge.m	0.56	1.00
673	Charge	Polarizability	van der Waals Volume	N	Ridge.m	0.47	1.02
1300	Polarizability	---	---	N	Elastic Net.m	0.56	1.03
1496	Polarizability	---	---	Y	Lasso.o	0.62	1.04
1338	Polarizability	van der Waals Volume	---	N	Lasso.o	0.56	1.04
1298	Polarizability	---	---	N	OLS	0.56	1.06
1301	Polarizability	---	---	N	PLSR	0.56	1.07
1491	Polarizability	van der Waals Volume	---	Y	Lasso.o	0.62	1.07
890	Polar ASA	---	---	Y	Ridge.o	0.50	1.07
1489	Charge	van der Waals Volume	---	Y	Ridge.o	0.62	1.07

Abbreviations: INT – Interactions; OLS – Ordinary Least Squares; PLSR – Partial Least Squares
Other notations: __ .m/o reflects whether the lambda used in the penalty term of the methodology listed was the minimum lambda (_.m), or the value of lambda that was one standard error away from the minimum (_.o).

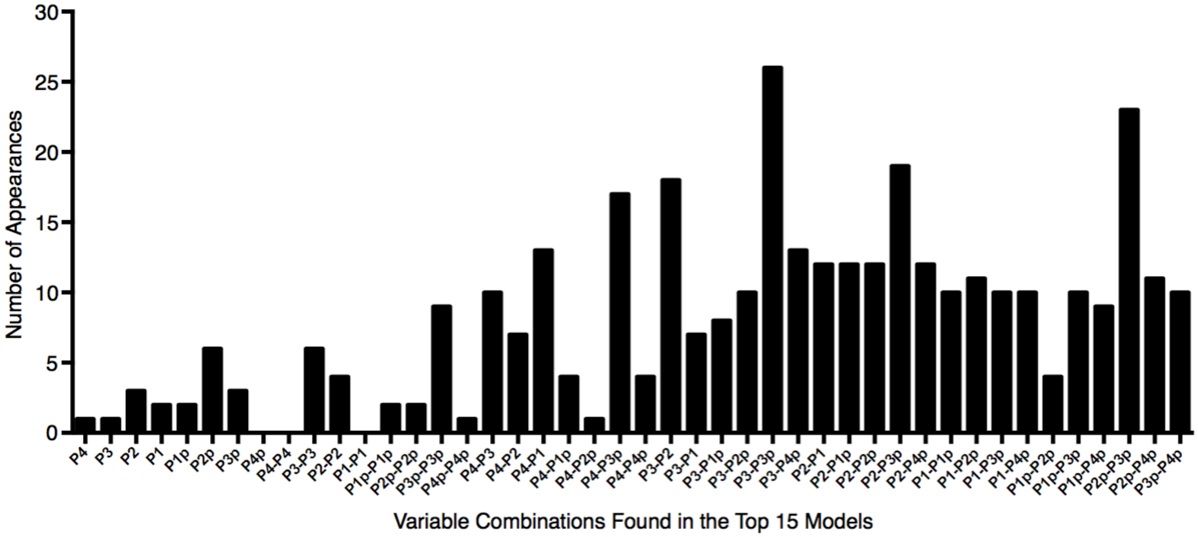

Figure 3.6: The distribution of variables within the top 15 models developed for prediction. All 15 models were the result of step-wise selection to reduce three-property models that included two-way interaction terms. An upper limit on the number of variables that could be included in each model was supplied, but no guidance on which variables to include was supplied. Labels that duplicate the same position imply an interaction between two properties at that position.

12. No correlation exists with rates determined by peptide analysis at low pH.

We compiled a small dataset from published results that had reported cleavage rates for substrates relative to a wild-type MA/CA sample (Table 3.6). We wished to apply our models to this dataset to judge their ability to predict processing rates. Within this dataset, 14 observations of 10 cleavage sites overlapped with our own. These were removed and directly compared to our results (Figure 3.7). The dashed line represents a perfect 1:1 correlation between rates. Only one sample falls very near this line (PQNF/LQSR), but that sample was inactive in our assay and barely detectable elsewhere (215). Boxed data points are results for the same site (the pair of boxed black points differed in context). The Pearson product-moment correlation was trending towards significance, but ultimately was not (r = 0.470, CI = -0.081 to 0.801, *p* = 0.0899), implying no relationship exists between our results and the previously published results. This suggested our models would not likely be a good predictor of the out-of-sample data. Our results confirmed that they were not. Using RMSE as a comparator, all of the best models at describing our data were quite terrible at predicting the out-of-sample data. The lowest RMSE among these top 15 models was 1.82 logs, which meant the predicted rates were off by 66-fold of the published value on average. Three of the predictive models had RMSEs below one (Table 3.5), but not by much (lowest RMSE = 0.96). Furthermore, two of the top four models were among the worst models at describing the source data. In summary, none of the 1570 models could predict of the precise rates for the out-of-sample dataset with a high accuracy.

Table 3.6: Out of Sample Rate Data

Source	pH	Ionic	Sequence	log(Rate)
Tözser 97	5.6	2.25 M	SDTY/YIVQ	-0.39
Tözser 97	5.6	2.25 M	SQTY/YIDQ	-0.43
Tözser 97	5.6	2.25 M	SDAY/YTDS	-0.53
Tözser 97	5.6	2.25 M	SDAY/YADS	-0.54
Tözser 97	5.6	2.25 M	SDIY/YTDS	-0.65
Tözser 97	5.6	2.25 M	SDTY/YLDS	-0.68
Tözser 97	5.6	2.25 M	SQNY/YTVQ	-0.82
Tözser 97	5.6	2.25 M	SDTY/YADS	-0.97
Tözser 97	5.6	2.25 M	SQNY/YNQS	-1.07
Tözser 97	5.6	2.25 M	SQNY/PTVQ	-1.27
Tözser 97	5.6	2.25 M	SQTY/YTVQ	-1.28
Tözser 97	5.6	2.25 M	SQTY/PIVQ	-1.43
Tözser 97	5.6	2.25 M	SDTY/YTDS	-1.45
Tözser 97	5.6	2.25 M	SFTY/YTDS	-1.47
Tözser 97	5.6	2.25 M	SDNY/PIVQ	-1.49
Tözser 97	5.6	2.25 M	SQNY/YTDQ	-1.66
Tözser 97	5.6	2.25 M	SDLY/YTDS	-1.68
Tözser 97	5.6	2.25 M	SDTY/YTFS	-1.69
Tözser 97	5.6	2.25 M	SGTY/YTDS	-1.69
Tözser 97	5.6	2.25 M	SDTY/YTGS	-1.74
Tözser 97	5.6	2.25 M	SDEY/YTDS	-1.78
Tözser 97	5.6	2.25 M	SQTY/YTDS	-1.82
Tözser 97	5.6	2.25 M	SDTY/YTQS	-1.89
Tözser 97	5.6	2.25 M	SQTY/YTDQ	-1.89
Tözser 97	5.6	2.25 M	SQTY/YIVQ	-1.9
Tözser 97	5.6	2.25 M	SQNY/YIVQ	-1.96
Tözser 97	5.6	2.25 M	SQNY/PIDQ	-2
Tözser 97	5.6	2.25 M	ATAM/MATA	-2.04
Tözser 97	5.6	2.25 M	SDTY/YEDS	-2.1
Tözser 97	5.6	2.25 M	SQTY/YTQS	-2.22
Tözser 97	5.6	2.25 M	SGTY/YTGS	-2.31
Tözser 97	5.6	2.25 M	SDTY/YTLS	-2.43
Tözser 97	5.6	2.25 M	SDTY/YTDQ	-2.45
Tözser 97	5.6	2.25 M	SDEY/YEDS	-2.48
Tözser 97	5.6	2.25 M	SDTY/YGDS	-2.96
Tözser 97	5.6	2.25 M	SDGY/YTDS	-3.05
Tözser 97	5.6	2.25 M	SLTY/YTDS	-3.05

Table 3.6: (cont)

Source	pH	Ionic	Sequence	log(Rate)
Tözser 91A	5.6	2.25 M	PQNYPIVQ	-0.3
Tözser 91A	5.6	2.25 M	GQNYPIVQ	-0.42
Tözser 91A	5.6	2.25 M	NQNYPIVQ	-0.75
Tözser 91A	5.6	2.25 M	DQNYPIVQ	-0.84
Tözser 91A	5.6	2.25 M	AQNYPIVQ	-1.01
Tözser 91A	5.6	2.25 M	RQNYPIVQ	-1.33
Tözser 91A	5.6	2.25 M	TQNYPIVQ	-1.36
Tözser 91A	5.6	2.25 M	MQNYPIVQ	-1.66
Tözser 91A	5.6	2.25 M	KQNYPIVQ	-1.7
Tözser 91B	5.6	2.25 M	RKILFLDG	0.65
Tözser 91B	5.6	2.25 M	SLNLPVAK	0.39
Tözser 91B	5.6	2.25 M	RQVLFLEK	0.11
Tözser 91B	5.6	2.25 M	TLNFPISP	-0.27
Tözser 91B	5.6	2.25 M	GLAAPQFS	-0.49
Tözser 91B	5.6	2.25 M	AETFYVDG	-0.66
Tözser 91B	5.6	2.25 M	GGNYPVQH	-1.05
Tözser 91B	5.6	2.25 M	ARLMAEAL	-1.29
Tözser 91B	5.6	2.25 M	PFAAAQQR	-1.61
Tözser 91B	5.6	2.25 M	PRNFPVAQ	-1.88
Bagossi 98	5.6	2.25 M	SQLYPIVQ	-1.09
Bagossi 98	5.6	2.25 M	LQNYPIVQ	-2.29
Bagossi 05	5.6	2.25 M	SQAYPIVQ	-0.3
Bagossi 05	5.6	2.25 M	SQVYPIVQ	-0.7
Bagossi 05	5.6	2.25 M	SQGYPIVQ	-1
Bagossi 05	5.6	2.25 M	SQIYPIVQ	-1
Bagossi 05	5.6	2.25 M	SQFYPIVQ	-1.52
Margolin 90	6.0	0.025M	SQNYPAVQ	-1.43
Pettit 02	7.0	2.25 M	SQNMPIVQ	-0.6
Pettit 02	7.0	2.25 M	SQNLPIVQ	-1

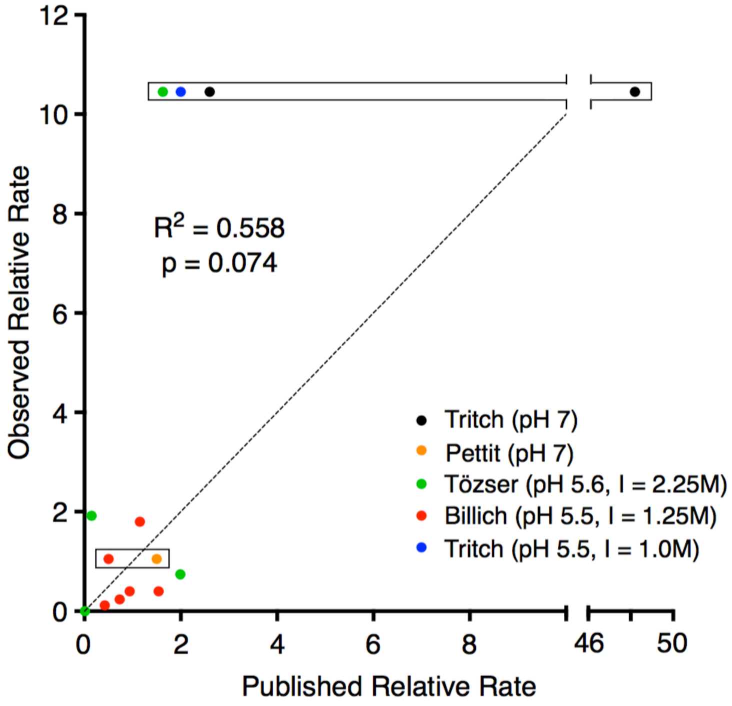

Figure 3.7: No correlation exists between processing rate in the two-substrate assay and those reported by prior publications. Rates are reported as relative to the MA/CA cleavage site on a linear scale. The dashed line denotes a perfect 1:1 correlation. The boxes group data points with identical cleavage sites. Data is referenced from: 214, 229, 284, 294, 295.

13. The predictive models were more effective as classifiers of fast and slow sites.

In lieu of being able to predict precise rate values, we ascertained whether the models could simply make a distinction between sites cleaved faster or slower than a reference value. Though the ideal choice would have been the rate of cleavage of our internal control protein, only 3/66 samples from the out-of-sample data set had rates faster than MA/CA. We did proceed with this analysis, in which we generated receiver operating characteristic (ROC) curves and calculated the area under the curve (AUC) as a measurement of accuracy. With these guidelines, several of our models were able to separate both our own data and the out-of-sample data very successfully (Figure 3.8A). Two models scored above 95% with both datasets. However, we felt choosing a value that introduced more parity would better evaluate the models. So, a rate of one-quarter that of MA/CA (-0.6 logs) became the classification point. One dozen models perfectly partitioned the source data, and almost 300 models succeeded at a better than 95% success rate (Figure 3.8B). Unfortunately, separation of the out-of-sample data was far less effective. Only 24 models sorted the sites with better than 70% accuracy, a milestone of adequacy, not success. There were three models that scored better than 95% on the source data and better than 70% with the out-of-sample data. While they did correspond to some of the models with the lowest RMSE when evaluating the source data, they were all on the bottom half of the models when predicting the out-of-sample data. Thus, even though these models were reasonable classifiers, they were still some of the poorer predictors of precise rates.

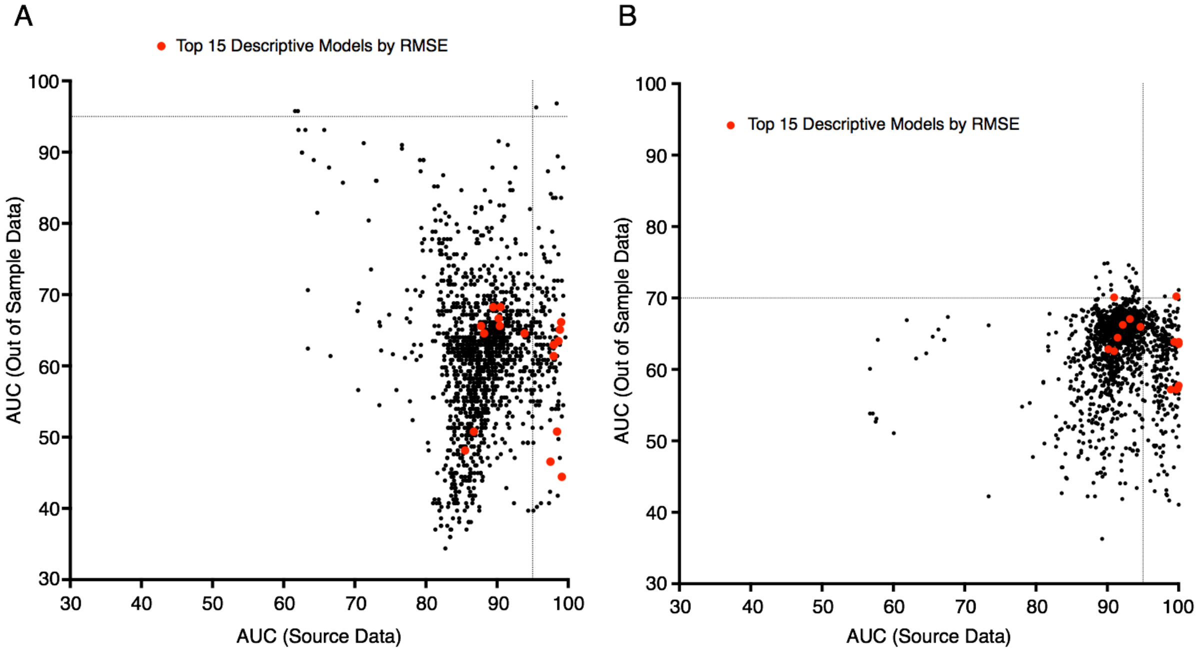

Figure 3.8: Few prediction models can separate out-of-sample data into fast and slow sites with good accuracy. ROC curves were generated for each prediction model to distinguish between well and poorly cleaved processing sites. The reference point chosen was either (A) the rate of MA/CA processing or (B) one-quarter the rate of MA/CA processing. The second grouping was chosen to increase the parity in the out-of-sample data. Gray lines denote 95% accuracy (x-axis) and (A) 95% or (B) 70% percentile (y-axis). Red points denote the top fifteen of the prediction models at describing the source dataset according to the minimization of RMSE.

14. Models built to predict processing rates could not distinguish between active and inactive sites.

Most studies attempting to determine HIV-1 PR cleavage site specificity focus on the difference between functional and inactive sites (299-313). These studies operate under the premise that developing a set of cleavability rules will define cleavage site preferences. However, we questioned whether models that could discern between active and inactive sites were also models that capable of distinguishing between fast and slow sites.

In order to examine this question, we assembled three more datasets. The first dataset, which we labeled the DEAD set, contained all 81 samples tested in the two-substrate proteolysis system (66 live, 15 dead). The second dataset was a derivative of the combined Schilling (280) and Impens (311) datasets that had been cleaned and made available by Rögnvaldsson, You, and Garwicz (311). In this dataset, all cysteine- and tryptophan-containing cleavage sites were removed since our own results lacked these amino acids. We labeled this data the NOCW set, and it contained 491 active and 3143 inactive sites. There were no duplicates with our own data, and only two sequences with as many as five amino acids in the same position. The third dataset was a subset of NOCW, consisting of only the sequences with leucine, methionine, phenylalanine, or tyrosine in the *P1* position (FLMY). The FLMY set had 684 sequences, of which 271 marked as cleaved.

Examining first the ability of our models to separate the functional and non-functional sites of our own data, we found limited success. The best model at partitioning the DEAD set achieved an accuracy of 84%, but this was one of only six to exceed 80% accuracy (Figure 3.9A). However, its AUC with the FLMY data was only 54%. Essentially, this model was no better at predicting whether a site in the FLMY data was active or inactive than if it had guessed randomly. Allowing the models the chance to separate the NOCW data, they performed about as well as they did on the FLMY dataset (data not shown). The average difference in AUC for each model was only 3.51 ± 2.78%. Thus, the added diversity in the *P1* position did not make a difference in prediction ability.

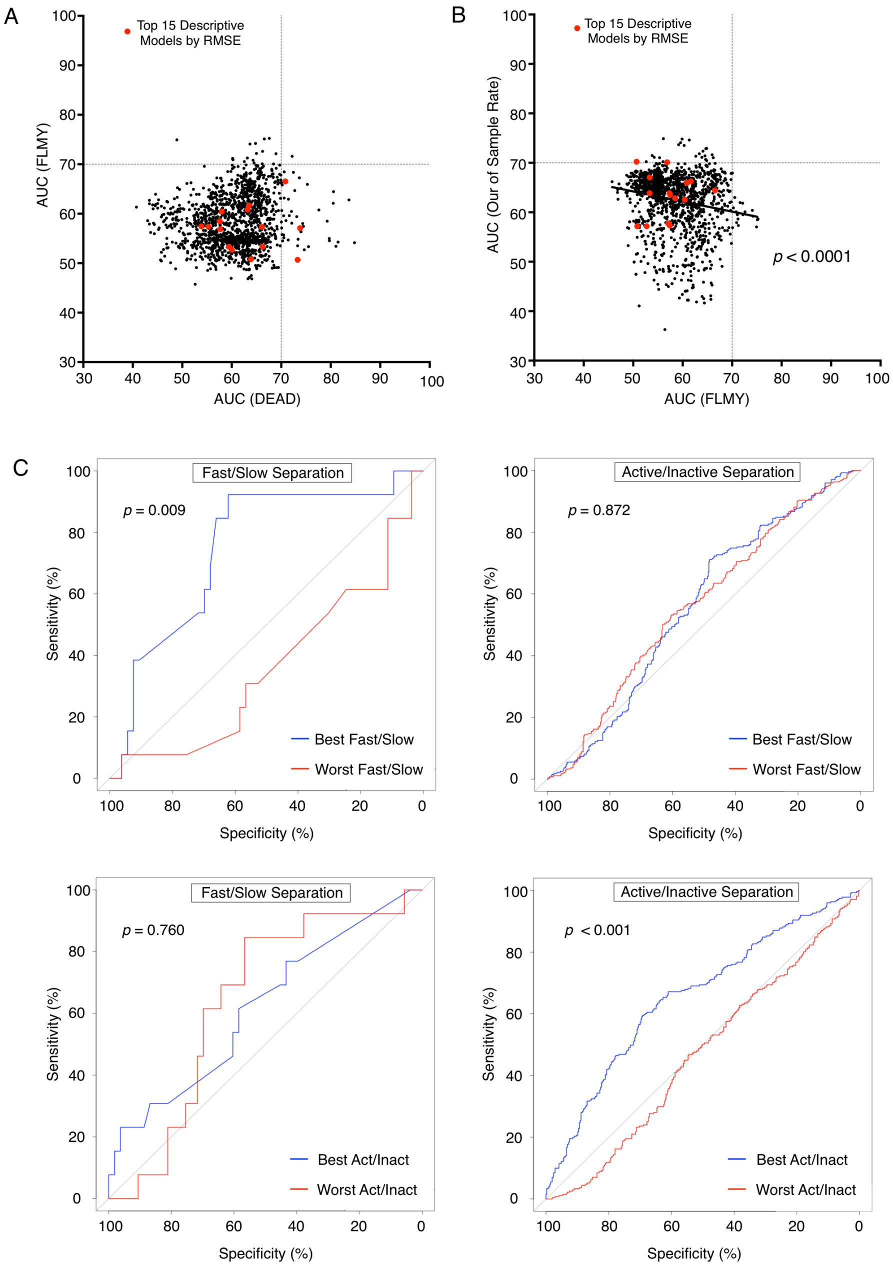

Figure 3.9: Models built for rate prediction were ineffective at distinguishing between active and inactive sites. (A) A comparison of the accuracy of the prediction models at separating active and inactive sites from the source and FLMY dataset. Red points denote the top fifteen prediction models at describing the source data according to the minimization of RMSE. (B) A comparison of the ability of each model to classify cleavage sites as faster or slower than one-quarter the rate of MA/CA (y-axis) and the ability to classify functional and non-functional sites (x-axis). The trend line and significance of the correlation is shown. (C) ROC curves for the best and worst fast/slow classifiers compared for their accuracy as separators of fast and slow sites (top left panel) and active and inactive sites (top right panel). Similarly, the best and worst models at partitioning live and dead sites were compared on the basis of their ability to separate fast and slow sites (bottom left panel) and active and inactive sites (bottom right panel).

When we compared the ability of our collection of models to separate out-of-sample data on the basis of both rate and activity, we found a significant negative correlation Figure 3.9B. As a more targeted illustration of this relationship, we performed a statistical test to compare ROC curves. Two pairs of ROC curves were drawn from the data. The first pair represented the best and the worst model at predicting whether a cleavage site was faster or slower. The second pair was the best and worst model at separating active from inactive sites (FLMY dataset). While there was a significant difference between the model pairs in accomplishing the task for which they were chosen, when asked to do the other task, the models were not statistically different (Figure 3.9C). Thus, our models, which were based off the physicochemical properties of the cleavage sites, could not discern between active and inactive sites as well as between fast and slow sites. This suggests that simply defining the requirements for processing will not also define the content of optimal HIV-1 PR processing sites.

C. Discussion

The successful maturation of HIV-1 particles into infectious virions depends upon the order and dynamics with which the PR cleaves Gag and Gag-Pro-Pol. Though context and putative cofactors also affect HIV-1 PR specificity (50, 219-221, 228-230), the primary determinant of where and when the PR functions is the eight-amino acid processing site sequence. In this chapter, we generated the largest-to-date dataset of mutant HIV-1 processing sites assayed in the context of a globular protein. Six different wild-type HIV-1 processing sites were mutated with various substitutions, and from these mutants we were able to identify a number of important relationships among amino acids across the eight positions that represent the processing site. We also applied statistical model-building programs to identify patterns common to all sites and to build a series of predictive models. While there was good correlation between the patterns found in the site-by-site analysis and the descriptive mixed-effects models, the predictive models were dissimilar in what they highlighted. We tested these models on an out-of-sample dataset, and found the results to be disappointing. Utilizing the models as classifiers was slightly more effective, although performance was still at best only adequate. Very few models could separate both fast from slow sites and active from inactive sites to even a low level of success.

Six HIV-1 processing sites were placed into the MA/CA linker region of a globular protein substrate, and their rates of processing were measured relative to the wild-type MA/CA site. We found the order of cleavage to be SP1/NC > MA/CA ~ RTH/IN ~ CA/SP1 > TF/PR ~ SP2/p6. This order agrees with the previously published results for this assay with the exception of the SP1/NC site. We previously reported the SP1/NC site to have a relative rate of approximately half that of the MA/CA site (50), whereas we now report a rate of ~10-fold greater than MA/CA. The difference stems from our use of the HXB2 version of the MA/CA cleavage site (ATIM/MQRG) rather than the version found in the NL4-3 clone of HIV-1 (ATIM/IQKG). The relative rate of the Met-Met site to the internal control was more consistent with the findings of Tritch et al. (230), and represents a rate much closer to that observed by Pettit et al. when the site was in its natural context (44). In addition, the expected rates of the CA/SP1 and SP2/p6 sites differed from those observed when in their natural context. Both of these sites are believed to be under the influence of contextual determinants, which would account for their different rates.

The addition of glycine triplets to the natural HIV-1 substrates caused a global reduction in processing rate, but had only one effect on the order in which the cleavage sites were processed. In the presence of glycines, the RTH/IN site was processed more efficiently relative to the other cleavage sites than when the glycines were absent. The observation of this change in order implies there is a feature of the MA/CA context that negatively regulates the RTH/IN site. One other report has suggested a restrictive effect of the MA/CA linker region on a non-native cleavage site, though that was the SP1/NC site (230). That restriction was similarly enforced on the upstream portion of the cleavage site; alteration of the *P5* amino acid was necessary for the SP1/NC site to exhibit its characteristic processing rate. We did observe an increase in the relative processing rate of SP1/NC when placed in the context of the glycine linkers,, though it was not as noteworthy a change as that previously published. Thus, despite the lack of apparent structure in the MA/CA linker region (315), the MA-adjacent portion has some influence over the cleavage site structure. This effect could influence orienting the normal MA/CA processing site, which necessarily contains a proline in the *P1'* position. Substrates must form a beta-strand like structure (48), and more rigidity in the linker region could mitigate the typical disruptive effect of proline on secondary structure.

In the analysis of each cleavage site, the MA/CA (with a *P1'* proline being fixed) and SP1/NC cleavage sites distinguished themselves as lacking significantly suboptimal amino acids. Neither of them could be significantly improved with a single amino acid substitution, nor did any of the collective 22 substitutions introduced result in an inactive site (excluding the expected *P1* isoleucines). We conclude that these two processing sites are already largely optimized. This inability to improve the SP1/NC and MA/CA sites could reflect the importance the virus places in controlling the timing of their cleavage. Proteolytic processing begins at the membrane before particle release (20), and excessive or early PR activity interferes with virus particle production (224). Any increase in processing rate of the initial target site of the HIV-1 PR (SP1/NC) could ostensibly lead to excessive processing and a destabilization of the virus particle before release can be achieved. As for MA/CA, cleavage at its junction releases steric constraints on the HIV-1 envelope protein to improve its fusogenicity (22-25). Earlier processing of MA/CA could cause virus particles to gain the ability to fuse with target cells while still in the immediate vicinity of its cell of origin, increasing the risk of superinfecting the original cell. Thus, the virus protects itself from potential breakdowns in the lifecycle by using cleavage sites that cannot be improved any further. They are still somewhat sensitive to negative alterations, though this may be less of a concern.

Each of the other four sites could be improved with single amino acid substitutions. For CA/SP1, the absence of this change to an improved rate of cleavage can likely be related to secondary structure. While in the immature CA lattice, the CA/SP1 site assembles into an alpha-helical bundle (316-321). Though phenylalanine is not a disfavored amino acid in alpha-helices, alanine (at *P1'*) has a higher helical propensity (322). An additional concern is whether the *P1'* alanine is part of the interaction face between Gag molecules, as the phenylalanine could be disruptive to the formation of the helical bundle. In the absence of a high-resolution structure of this interaction, this possibility remains unknown. Alternatively, if the phenylalanine does accelerate processing of the CA/SP1 site to such an extent that the CA shell begins to form before condensation of the RNP core completes, an improperly formed reverse transcription complex might be more susceptible to interference by host antiviral proteins, could fail to complete reverse transcription, or even may exhibit defects in nuclear entry (Chapter 1).

We have already given some discussion to the structural reason that might otherwise deter the TF/PR site from making the *P3'* isoleucine to *P3'* phenylalanine change that potently improved TF/PR processing. However, to reiterate, the *P1'-P4'* amino acids are all apart of a critical beta-sheet formed following cleavage at the TF/PR site. This beta-sheet imparts stability to the HIV-1 PR that is absolutely essential for processing cleavage sites *in trans*. The structure of the HIV-1 PR (323) shows the *P3'* amino acid points toward the interior of the HIV-1 PR. The orientation of this side chain is consistent with the idea that the phenylalanine, a larger and longer amino acid than isoleucine, might sterically conflict with other amino acids. This would either disrupt the interior of the HIV-1 PR and/or the stability of the beta sheet. Thus, the isoleucine to phenylalanine change does not occur.

As for the higher prevalence of serine in the TF/PR *P2* position despite asparagine's clear superiority as a substrate, overlap between Gag and Gag-Pro-Pol reading frames may provide the explanation. The coding region for the ALIX binding motif in the p6 domain (LYPx_nLxxL) (324) entirely overlaps the TF/PR processing site. The difference between a serine and asparagine in the TF/PR site corresponds to a difference of alanine or threonine in the x_2 position of the binding motif, respectively. That *P2'* overlaps with a non-conserved position presumably allows its diversity, but that same lack of conservation in the x_2 amino acid does not necessarily mean that the x_2 amino acid's identity does not affect ALIX binding. If an alanine in this position does help virus particle release efficiency, or if a threonine hurts it, it could explain serine's predominance.

Similar to the CA/SP1 and TF/PR sites, both SP2/p6 and RTH/IN allowed substitutions that improved the processing rate. However, we are unaware of any potential structural explanation for selection against these substitutions. For the SP2/p6 site, a leucine to phenylalanine substitution in the *P1'* position improved rate by 12-fold. The relative scarcity of this mutation in our cleavage site alignment (9/2254) becomes even more glaring when recognizing that this substitution is a frequent compensatory mutation within Gag after drug resistance mutations have developed in the HIV-1 PR (173). One simple explanation is that the cleavage site would simply become too good. The SP2/p6 site was cleaved roughly 1.5-fold faster than the MA/CA site in an in vitro assay with full-length Gag (44). If this 12-fold effect is projected on to that 1.5-fold increase, then cleavage of the SP2/p6 site would occur with approximately the same efficiency as cleavage of the SP1/NC site, if not faster. Thus, should viruses produce SP2/p6 sites with a phenylalanine/phenylalanine pairing in the *P1/P1'* positions, the order of processing may be disrupted and the pre-reverse transcription complex would not assemble correctly.

The improved processing rate of the RTH/IN site by the *P3* lysine to glutamine mutation was not as beneficial as that of the SP2/p6 *P1'* substitution. Though IN is important to the encapsulation of the RNP core within the reassembling CA protein during virion maturation (192, 193), the importance of the timing of its release by cleavage at the RTH/IN site is unknown. IN can be included in viable viruses solely via a Vpr-IN fusion protein (325), implying IN facilitates the inclusion of the RNP within the CA shell after cleavage at the RTH/IN. However, whether this is actually necessary remains untested. Furthermore, why a mutation that would accelerate the release of IN from RT might interfere with such a role is unclear. Looking at a crystal structure of the HIV-1 heterodimeric RT (326), the RTH/IN cleavage site would occur immediately after the end of an alpha helix. One could speculate that if the lysine is in fact the causative agent for the unusual conformation the RTH/IN site adopts, it does not mutate because this shape is necessary to ensure the helix break and/or to stably position the site away from the body of the heterodimer.

Regardless of whether the lysine causes the *P4* amino acid to assume the conformation of the RTH/IN site, there was a definite pattern in the makeup of the site. Lysine required a beta-branched amino acid at the *P2* position and strongly preferred a non-aromatic amino acid in the *P1* site. It was not absolutely prohibitive to a proline in the *P1'* site, though cleavage of the site by the HIV-1 PR was exceptionally poor. Since proline has the exact opposite preferences, the near total incompatibility is not altogether surprising. These results support the concept of "blueprint" amino acids. In short, these blueprint amino acids are the amino acids around which the site is built. According to this concept, they are chosen because of an external requirement, i.e. for a function unrelated to their role as a substrate. For the *P1'* proline sites, these are the stabilization of CA monomers (MA/CA) (153-155) and the regulation of intramolecular cleavage order (TF/PR) (228). For lysine, this is the ensured termination of the C-terminal alpha helix of the RNase H domain and/or the stable positioning of the cleavage site coming out of the helix. This blueprint concept is key to unraveling processing site composition.

The second half of understanding the makeup of each cleavage site comes from the identification of interdependent amino acid positions. Recall that the mixed-effects models we generated included several interactions. Two of these interactions concerned specific positions within the cleavage sites, namely the *P2* and *P4'* positions, and the *P2* and *P2'* positions. Importantly, the overlap between these interactions links all three amino acids together. This increases the likelihood that if we know one of these positions, we can predict the optimal amino acids for both of the other sites. Here, we immediately return to our current catalog of blueprint amino acids for a simple example. Both proline and lysine influenced the composition of a pair of other amino acids in the cleavage site, and one of these is the *P2* amino acid. As stated above, because we know one of the three amino acids in this interdependent group, we can limit the potential amino acids that can occupy the other two sites. Effectively, we are proposing that processing sites are built by a cascading selection process. Depending on the location of the blueprint amino acid, it will limit the potential content of certain other positions in the cleavage site. At least one of these positions will overlap with a different set of linked amino acids, and therefore will limit the repertoire of amino acids that can fill those positions. This continues until all sites are filled.

In addition to our descriptive models, we built 1570 predictive models. Unfortunately, they were extremely poor predictors of out-of-sample data. There are multiple possible explanations for this deficiency. First, the models were likely overfit to the data. We allowed a certain number of variables in each of the models, and all of the models were very near that limit. Another potential problem is that our dataset did not represent a true random sample. There are 25.6 billion (20^8) possible eight amino acid combinations, yet our dataset comprised a limited set of highly related sequences. The results are therefore very biased to these particular groupings of amino acids. Furthermore, because the null mixed-effects model was significantly better than the null standard linear regression model meant there was a significant amount of noise in our data due to the differences in baseline cleavage rates. By viewing each data point as an independent sample during the model creation process, the models were trained on that noise. As a result, the wrong variables were likely included. For example, the *P3/P3'* interaction was the most frequent term found in the top 15 prediction models. However, the *P3* position was highly amenable to substitution in both the MA/CA and CA/SP1 sites, with the sole possible exception of lysine. In addition to our own work, a considerable amount of experimental (47, 287-289, 314) and bioinformatics evidence (308, 313) sharing that conclusion has been generated. Thus, the limitations of our dataset could be the major reason the predictive models failed to perform as well as hoped.

Some fault also likely lies with the out-of-sample data. First, not all of the reactions were performed under the same pH and ionic strength conditions (215, 283-298). This is a complication among HIV-1 PR cleavage studies that confounds the interpretation of results across studies. Both substrate binding (K_{m}) and turnover rate (k_{cat}) are affected by pH and ionic strength (294), making the comparison of HIV-1 PR specificity across reactions difficult. The specificity constant for the MA/CA peptide across these studies is sourced from the same initial experiment (215, 284, 295, 296), yet when this peptide was retested (283), the specificity constant was nearly twice as large as in the original publication (45.3 $\mathrm{mM}^{-1}\mathrm{s}^{-1}$ versus 82.44 $\mathrm{mM}^{-1}\mathrm{s}^{-1}$). Thus, the relative rates of processing we derived from these studies potentially have additional complications. Fortunately, our own results do not suffer from this concern because all reactions included a reference site. Altogether, the limited ability of our prediction models to identify the rates of processing for out-of-sample data does not necessarily mean the models are wrong. However, this initial modeling attempt is not encouraging, and more data should be gathered before a firm conclusion is made.

In addition to the prediction of precise rates, we examined the models as simple classifiers of fast and slow sites. Here, at least, we achieved a modicum of success. Because our out-of-sample dataset contained primarily rates that were much slower than the MA/CA site, we chose to evaluate our samples relative to a processing rate of one-quarter that of MA/CA. Our predictive models were successful at separating our source data, but were again considerably less effective utilizing the out-of-sample data. However, not all models were failures. Two-dozen models could partition the out-of-sample results to an acceptable level (>70% accuracy), and each of these models was at least 88% effective with the source data. As far as we are aware, we are the first to attempt a classification between fast and slow sites. Therefore, there is no benchmark to which we can compare our results.

In contrast, many have built cleavage rules to distinguish between active and inactive sites. These algorithms perform very well (85% or better) on data sets much larger than our own (309, 311), making our best active/inactive predictor model (75%) non-competitive. As a result of this analysis, we noticed an inverse correlation between the ability of a model to separate fast and slow cleavage sites and its ability to predict which cleavage sites are active versus inactive. This is not particularly surprising. Take, for example, the defective sites that were identical to the wild-type substrates save for an isoleucine in the *P1* position. The value of isoleucine on each of the physicochemical scales we used are nearly identical to that of leucine. While leucine is a very capable amino acid in the *P1* position – it was actually the *P1* amino acid of the fastest site we measured – isoleucine is guaranteed to block cleavage. Thus, our models likely did not separate these samples successfully because they would appear to be nearly identical numerically. It is possible that separation could be achieved with the addition of a categorical variable to distinguish beta-branched amino acids from non-beta-branched amino acids, however the lack of variation in the data will complicate the application of the statistical procedures we employed. We alternatively suggest that models wishing to identify ideal amino acid combinations ought to select those combinations in two phases. In one phase, the precise rates of processing are predicted; in the other, the sites that violate a cleavability rule are removed from contention.

In this report, we measured the cleavage rate of more than 80 substrates of the HIV-1 PR relative to the cleavage rate of the MA/CA processing site. To our knowledge, these results represent the largest-to-date dataset of globular substrates for the HIV-1 PR cleaved under near-physiological pH and ionic strength conditions. We discussed the amino acid preferences for six of the HIV-1 cleavage sites found in Gag and Gag-Pro-Pol. For the MA/CA and SP1/NC cleavage sites, we were unable to provide a significant improvement over their wild-type sequences with only one or two amino acid substitutions. The other four cleavage sites were all much more amenable to cleavage rate increases with amino acid changes. In addition, we identified lysine in the *P3* position as a very influential amino acid to its surrounding context, and suggest it join *P1'* proline with the distinction of being a "blueprint" amino acid. These blueprint positions help dissect processing site sequences by limiting the content of other amino acid positions, that in turn link to still more parts of a cleavage site. Lastly, we applied a large number of predictive models to out-of-sample data with the hope of predicting their cleavage rates. However, we were largely unsuccessful. A few of the models were at least adequate as classifiers, though no model could classify both active and inactive, and fast and slow sites with good accuracy. This leads us to conclude that the identification of an ideal cleavage site sequence must satisfy the terms of two models: one that predicts rate, and one that predicts viability.

D. Materials and Methods

1. Constructs

The MA/CA protein was generated as in Chapter Two. Briefly, MA/CA was amplified from the pBARK plasmid, which contains the entirety of the *gag* and *pro* genes from NL4-3. A 6xHis tag was added to the N terminus, a termination codon at the C terminus, and flanking NdeI sites were introduced to enable cloning into the pET-30b (Novagen) vector. An additional round of mutagenesis introduced a tetracysteine motif (CCPGCC) in the Cyclophilin A binding loop (His87-Ala92) of CA.

The original pET-GMCΔ construct was created in a similar manner, where the GST-tag was added to the N-terminus of MA by overlapping PCR. The full procedure is published in (50). Each alternative cleavage site was introduced by a two-step procedure: first generation of the insert by overlapping PCR centered on the cleavage site, followed by subcloning into the pET-GMCΔ vector utilizing the NdeI cleavage sites that had been previously introduced. The glycine insertion mutants for each of the natural HIV-1 cleavage sites were also generated by this procedure. All other mutations made within each processing site were introduced by site-directed mutagenesis. The decision for which mutations to be made were dependent upon the presence of mutations within the HIV sequence compendium (http://www.hiv.lanl.gov/), their presence in alternative cleavage sites, and/or due to their ability to be made with a single nucleotide substitution. Primers were obtained through Sigma-Aldrich.

2. Expression and purification of the HIV-1 PR and globular HIV-1 PR substrates

Growth and expression of globular substrates proceeded as described in Chapter Two. Briefly, *E. coli* BL21 DE3 lysogens (Novagen) were transformed with pET-MA/CA or pET-GMCΔ. Following growth overnight in 2xYT media, starter cultures were used to inoculate MagicMedia (Invitrogen) for protein production. Expression cultures were grown for 8 hours at 37ºC and 225 rpm, pelleted by centrifugation and frozen overnight at -80ºC. Lysis was performed by sonication in TBS pH 7.5, 1% Triton X-100, 2 mM beta-mercaptoethanol. Cellular debris was collected by centrifugation, and the protein collected by affinity chromatography using the Ni-NTA Superflow columns (Qiagen). Purified proteins were concentrated using Vivaspin Concentrators (GE Healthcare), and underwent buffer exchange into storage buffer (20 mM sodium acetate, 140 mM sodium chloride, 2 mM beta-mercaptoethanol, 10% glycerol, pH 6.5). The pH was confirmed to within 0.2 units with a micro-pH electrode (Thermo Scientific).

Purified HIV-1 proteases were produced as described in Chapter Two.

3. Two-substrate proteolysis reactions

Two-substrate proteolysis reactions were run in proteolysis buffer (50 mM sodium acetate, 50 mM NaMES, 100 mM Tris, 2 mM beta-mercaptoethanol, pH 6.5). Reactions were 150 μl in volume and pre-incubated at 30ºC for 1 hour before addition of the enzyme to allow the Lumio Green Reagent (Invitrogen) to bind the CCPGCC motif in the CA region of each protein. Both substrates were included at an initial concentration of 1.2 μM. The HIV-1 PR was used at a concentration of 150 nM in the two-substrate assays where the substrates lack the glycine insertions, and 400 nM when the glycines were present to make up for the minor drop in processing rate. Aliquots were collected at specific time points throughout the course of the reaction and added directly to SDS to halt the reaction. The zero minute time point was removed immediately prior to the addition of enzyme. Most reactions were limited to 15 minutes, the timeframe required for the internal control to reach ~50% processing, though only data points generated within the first 10% of processing were used to generate relative rates. Sites exhibiting <10% cleavage after 15 minutes were retested in extended, 120-minute assays. If cleavage was not still observed after 120 minutes, the site was deemed defective. After the final time point was collected, reaction pH was confirmed as 6.5 using a micro-pH electrode (Thermo Scientific). Substrates and products were separated by SDS-PAGE using precast 16% Tris-Glycine gels (Invitrogen). The fluorescently labeled proteins were then imaged with a Typhoon 9000 (GE Healthcare/Amersham Biosciences), and quantified by ImageQuant TL (GE Healthcare) software. Results were reported as the percent product formed. The initial reaction rate for each substrate was determined using only the data points collected where the reaction was ≤10% complete, or was estimated based on the first non-zero data point collected. To determine the relative rate of processing, the ratio of initial velocities was compared using the internal control as the denominator, and the value recorded. The relative rate for each mutant was determined in at least two separate reactions. The average variance in estimated rate between the two reactions was 1.25-fold, with a range of 0 to 2-fold. Overall, these values differed by over 3000-fold, and were log-transformed for ease of interpretation.

4. Statistical modeling procedures

All statistical procedures were performed using "R" software, version 3.2.0 (327), and the RStudio version 0.99.447 interface. The physicochemical properties used to analyze the data include hydrophobicity (328), nonpolar, polar, and total side-chain accessible surface area (ASA) (329), total amino acid polarizability – a volumetric measurement of an amino acid's response to an external electric field (330), total amino acid van der Waals volume (331), total amino acid molecular weight, side-chain length, and amino acid charge. The side-chain length was determined using PyMol. The distance between the alpha-carbon and furthest atom was determined with the "Distance" function provided within the standard PyMol software. To this value was added one van der Waals radius (332) according to the identity of the atom used in distance calculation. Unless required by the statistical procedure, the data was not standardized. Predictions for the out of sample data were generated using the *predict* function that accompanied each statistical procedure.

Out of sample data: Three out of sample datasets were generated. The first dataset was generated from a series of previously published studies where comparisons of the observed rates were made to a wild-type version of the MA/CA processing site (46, 215, 283, 284, 290, 295, 296). Most samples (63/66) were recovered from reactions performed in similar pH and ionic strength conditions, though they were spread across five publications and the reference value for the MA/CA processing site appears to have only been determined once. The second dataset (NOCW) is a compilation of the Schilling (280) and Impens (311) datasets, after all cleavage sites containing cysteine and tryptophan were removed because they did not appear in our reactions. Rognvaldsson et al. (311) originally organized this data for use and posted it online for others to access. No rates are given, only the classification of whether the processing site was cleaved or not. There are 3634 samples in this dataset, of which 491 are cleaved. The third dataset (FLMY) is a smaller version of the NOCW set where only the processing sites containing one of the four common amino acids at the P1 position were included. This contained 684 sites, 271 cleaved.

Linear mixed-effects modeling: The linear mixed-effects models were generated with the *lmer* function from the “lme4” package (333). Each processing site was grouped according to their relationship with one of the cleavage sites found within the Gag or Gag-Pro-Pol polyproteins, and this grouping was used as the random effects term of a random-intercept mixed-effects model. Likelihood ratio tests were performed using the *anova* function from the standard “stats” package (327). To enable valid comparisons between models, full maximum likelihood estimation was utilized instead of restricted/residual maximum likelihood. Selection of the most parsimonious model was accomplished by step-wise elimination of the least significant (i.e. highest *p* value) term as determined by the likelihood ratio test. One variable was removed at each step until further removal resulted in a significant change to the model's effectiveness.

Ordinary least squares (OLS) modeling: All ordinary least squares regression analyses were performed using the included *lm* function in the basic R package. The analyses were performed in only the absence of interactions. Given that no selection was performed with these standard models, and the upper limit of variables was set at 25 (roughly 1/3rd the number of observations), interaction terms could not be included without violating this limit. Any singularities (i.e. multicollinearity among variables) were automatically removed from the modeling procedure by the R software. Combinations of one, two, and three physicochemical properties were considered when individual amino acids were used as the model criteria.

Best subset and step-wise selection: Best subset selection compares all possible combinations of variables at every possible level. Step-wise selection begins with a base model, and then attempts to add or remove a variable at every step of selection. All variables are tested at each step for their affect on model significance. Both best subset selection and step-wise selection were performed using the *regsubsets* function of the "leaps" package (334). The exhaustive search is the default, and may be used when the total number of possible variables is less than 40. It was applied to any model lacking interactions, and also the one-property model with interactions. The two- and three-property models with interactions were generated with the step-wise selection procedure. The maximum number of variables allowed in each model was 25 – one for each predictor variable in a three-property OLS model without interactions plus one more for the intercept. After the selection process, the model with the lowest Bayesian information criterion (BIC) was picked for use in OLS analysis.

Ridge, elastic net, and lasso regression: These ‘shrinkage’ methodologies are useful when the number of potential variables is larger than the total amount of observations. These methods are similar to OLS analysis, but a penalty term/shrinkage factor (lambda) is introduced. In ridge regression, no variables are removed but the coefficients of insignificant variables are reduced closer and closer to zero. Lasso regression uses a slightly different penalty term that results in the removal of insignificant variables. Elastic net regression compromises between ridge and lasso, behaving more like one or the other depending upon the value chosen for a so-called tuning parameter. All three methodologies were performed using the *glmnet* function of the “glmnet’ package (335). For ridge regression, the tuning factor (alpha) was set to zero; for lasso regression, alpha was set to one; and for elastic net regression the alpha was calculated with the *train* function of the “caret” package by repeated cross-validation (336). All variables for all shrinkage methodologies were necessarily standardized to avoid improperly weighting of the regression coefficients. The programs do return unstandardized coefficients, however. Lambdas were chosen by cross-validation after the value for alpha had been chosen using the cross validation function *cv.glmnet* provided by the “glmnet” package. Both the minimum lambda and the lambda one standard error away from the minimum were used to generate models. The minimum lambda is often not sufficiently parsimonious, which is why a second lambda that was not significantly different from the minimum lambda was also used.

Partial least squares regression (PLSR): Combinatory methods like principal component analysis and partial least squares regression take a multi-dimensional model and reduce the dimensionality into a smaller number of new components (i.e. combinations of the original variables). These new variables may then be used in standard OLS regression. PLSR differs from principal component analysis by considering the variance in the response variable in addition to the variance among the predictor variables. This makes PLSR much more advantageous for constructing predictive models as compared to principal component analysis. “caret” (336) and “pls” (337) packages were required for completing PLSR. The number of components for each model was chosen with the *train* function from the “caret” package making use of repeated cross validation. Variables were again standardized, and as in the best-subset selection methodologies, the maximum number of components allowed was 25. The actual PLSR fit was performed with the *plsr* function of the “pls” package, relying upon cross-validation as to choose the models.

Receiver operator characteristic (ROC) curves: ROC curves measure the performance of a classifier at distinguishing between one of two categories. The curve is a reflection of the sensitivity (true positive/(true positive + false negative)) and specificity (true negative/(true negative + false positive)). Accuracy is reflected in the area under the curve (AUC). Above 90% is excellent, above 80% is good, 70% is adequate, 60% is poor, and below 60% indicates the model is useless as a classifier. All ROC curves and AUC determinations were made with the *roc* function in the “pROC” package (338). The statistical tests to compare ROC curves, i.e. the DeLong test, were also within the “pROC”.

CHAPTER IV

CONCLUSION

A two-substrate enzymatic assay provides an elegant solution to overcome the potential inconsistencies between reactions that might otherwise result when observations are made separately. From the use of a two-substrate system, I discovered that the enhanced rate of processing exhibited by the HIV-1 PR in the presence of RNA was a general phenomenon rather than one specific to the p15NC substrate. This result contradicted the prevailing hypothesis that a substrate-RNA interaction was necessary (220, 221). After the initial discovery, I utilized a globular substrate with a severely compromised ability to interact with nucleic acid, as well as a peptide substrate too small to simultaneously interact with RNA and the HIV-1 PR to resoundingly disprove the long-standing hypothesis. Both of these reactions continued to exhibit RNA-dependent rate enhancement, leading to the conclusion that the substrate's ability to bind nucleic acid was irrelevant or at the very least optional.

As an alternative hypothesis, I suggested RNA-mediated enhancement resulted from a direct interaction between the RNA and an allosteric binding site on the HIV-1 PR. The reactions where I used the peptide substrate already supported this conclusion, but to directly demonstrate the interaction I performed a series of gel-shift assays. The progressive addition of a higher concentration of the HIV-1 PR to a constant amount of nucleic acid – in this case short ssDNA molecules also capable of enhancing PR activity – resulted in the selective disappearance of high molecular weight nucleic acid species when observed in a gel. Unfortunately, the prototypical “shift” was not observed in these assays. Though the unbound species could no longer be detected, the overall basic charge of the HIV-1 PR meant it was incapable of migrating into the gel under native conditions and no new band appeared near the top of the gel. Nevertheless, the selective disappearance of a nucleic acid species was a strong indicator of a direct interaction. Additional support for the hypothesis stemmed from work with the HIV-2 PR. This alternative enzyme was completely unaffected by the addition of RNA to its processing reactions, and also failed to interact with nucleic acid in the gel-shift assays. The lack of effect given that the substrate (save the cleavage site) was the same strongly demonstrated the effect was enzyme-dependent.

The selective interaction of the HIV-1 PR with high molecular weight nucleic acid species and the HIV-2 PR’s lack of interaction suggested specific requirements must be satisfied for the interaction to occur. The HIV-2 PR had a net negative charge, in contrast to the net positive charge of the HIV-1 PR, immediately pointing to an electrostatic determinant. The failure of the positively charged polymer spermine, and the success of the negatively charged polymer heparin at mimicking the effect of RNA in proteolysis assays supported this hypothesis. Additionally, the similarity in the magnitude of the enhancement effect caused by equal-length cytosine, thymine, and guanine homopolymers additionally suggested that the size of the polyanion was critical, whereas the structure or sequence was not. Larger sized polyanions will have a larger net-negative charge, and this will improve the strength of the interaction between enzyme and enhancer. Work with the adenosine homopolymer and tRNA actually intimated a rigid structure in the polyanion was unfavorable.

Also of interest was the mechanistic explanation for how RNA improved the efficiency of the HIV-1 PR. To gauge whether RNA altered the binding affinity of the enzyme for its substrate, the rate of product formation post-substrate binding, or potentially both, I performed a series of peptide cleavage assays to determine the values for k_{cat} and K_m in the presence and absence of RNA. The results found the k_{cat} increased and the K_m decreased, collectively demonstrating that both substrate binding and reaction turnover improved as a result of the interaction between RNA and PR.

The discovery that the enhanced reaction rate in the presence of RNA results from the interaction between the enzyme and RNA opens up an abundance of future research opportunities. Foremost among them is the investigation of the phenomenon's relevance *in vivo*. All assays I conducted were *in vitro*, and, despite a strong similarity to physiological conditions, the reactions can never perfectly recapture the environment within a lymphocyte. The most straightforward option to address *in vivo* relevance would be to identify the precise binding site(s) on the HIV-1 PR, mutagenize them to block the interaction, and then determine the fitness of the virus relative to the wild-type version. There are, of course, potential complications to this strategy, as the mutations that disrupt the interaction might additionally damage the functionality of the PR even in the absence of RNA. But because RNA is a required structural element of virus particles (14), any antagonism from that direction is likely out of the question. One alternative strategy would be to create a hybrid HIV-1/HIV-2 strain that contains the HIV-2 PR and its preferred cleavage site sequences in Gag and Gag-Pro-Pol. Provided such a virus is viable, growth competition assays with the wild-type virus would be recommended. If the virus were not viable, a number of tests to determine the defect would be the correct course of action; a BLAM-Vpr fusion assay or PCR screen for early and late reverse transcription products for example. Direct imaging of virus particles by electron microscopy and the identification of processing intermediates by western blot would also likely be worthwhile.

The most likely location on the HIV-1 PR for the interaction with RNA is the flap. The flap region is key to both substrate binding and hydrolysis of the scissile bond (268-273), which satisfies both of the criteria established by the enzyme kinetics assays. This region of the HIV-1 PR also contains a large number of basic amino acids while its counterpart in the HIV-2 PR does not. The HIV-1 and HIV-2 PRs have very similar tertiary structures (255-257), so simply replacing the corresponding amino acids from the HIV-2 PR into the HIV-1 enzyme might be sufficient to disrupt the interaction with RNA while also minimizing the effect on enzyme functionality. Certain positions with the HIV-1 PR have been mutated before without a major effect on enzyme activity (339) and this may be a good place to start, albeit the reactions that identified these locations were almost certainly RNA-rich.

Other possible research goals concern antiretroviral inhibitors. If we presume the effect is significant *in vivo*, then the important question becomes ‘how does the interaction with RNA affect drug resistance?’ We know from Chapter Two that several highly drug-resistant HIV-1 enzymes all retained the ability to interact with RNA *in vitro*. In most cases, the effect of this interaction was substantial enough that these extremely inefficient versions of the enzyme became comparably functional to the wild-type PR. Furthermore, we know the virus requires multiple copies of the HIV-1 PR to complete maturation (25); one enzyme simply does not have sufficient functionality to complete the process. Since RNA dramatically increases the catalytic ability of the HIV-1 PR, this could mean that a smaller number of enzymes is required to complete maturation. Thus, RNA-mediated enhancement could mean that higher drug concentrations are required to effectively inhibit the virus. Once an HIV-1 enzyme unresponsive to RNA has been identified, a study that compares the inhibitory concentrations of PIs between the mutant and wild-type version would be valuable.

Regardless of whether the PR-RNA interaction has a prominent role in vivo, the fact remains that an allosteric binding site exists on the enzyme. Thus, another inhibitor-oriented research proposal would be to identify small-molecule binding partners for the interaction site. Several have already identified putative flap-binding molecules (257, 264-267), giving some credence to the proposal. One of the more difficult tasks would be designing a system in which flap-binding inhibitors could be differentiated from substrate mimetics. That RNA binding affected K_m implies RNA interacts with the unbound version of the PR. Thus, inhibitors would likely need to interact with a PR that is not engaging a substrate. A fluorescence anisotropy-based competition assay between putative inhibitors and labeled polyanionic species might provide one solution. However, the physiological, low-salt conditions in which one would want to run the assay might cause aggregation of the polyanion, in turn interfering with detection. Structure-based drug design following the identification of the binding site could be a more lucrative strategy.

With regard to the study of HIV-1 PR sequence specificity, it has suffered from a litany of problems. There are, of course, the inherent difficulties: the existence of a wide variety of amino acid sequences capable of occupying the substrate envelope (280); the different contextual effects in place around each of the natural cleavage sites (50); and, as determined in Chapter Two of this dissertation, the existence of cofactors that affect substrate binding and reaction turnover. There are also issues resulting from a lack of coordination among researchers; specifically, an inconsistency in the reaction conditions utilized in different publications. Additionally, few, if any, have explored specificity in an environment where the pH and ionic strength is remotely similar to that encountered by the PR *in vivo*. Unfortunately, this latter issue was often necessary because peptides were used as substrates. However, unlike peptides, the

HIV-1 PR can efficiently process globular substrates under more normal pH conditions *in vitro* (50). In Chapter Three, I discussed the results of assays run using a series of globular proteins as substrates under near-physiological conditions. Each of these proteins contained a different cleavage site sequence, but all cleavage sites were placed within the same protein construct. I again employed the two-substrate system to take advantage of the internal control protein. This secondary substrate offered a reference point from which I could compare the relative rates of processing for each cleavage site with an exceptional level of accuracy and consistency.

Six cleavage sites were drawn from different locations throughout Gag and Gag-Pro-Pol. When examined in the two-substrate system, a relative order of processing was determined that differed from the order determined by *in vitro* assays where each site was in its natural context (42, 44). This was not unexpected as a previous publication found a very similar relationship among the sites (50). I ensured the present context was not responsible for this alteration in order by introducing a triplet of glycine residues on either side of the cleavage sites. The flexibility of these glycine segments disrupted any residual contextual cues. When re-examining the glycine versions, there was a minor, though noticeable effect on one particular site, RTH/IN. This result indicated there was some feature in the protein construct that negatively regulated the RTH/IN site. Since the natural site in this context obligatorily contains a proline, which is disruptive to the beta-strand conformation the enzyme requires for substrate recognition, the minor degree of influence this context imparts on the cleavage site might be essential to properly position the amino acids.

The existence of proline in this site is not a unique event. Three of the natural HIV-1 processing sites I examined contain a proline, though only two of them in the *P1'* position. This particular amino acid arrangement has been used to differentiate processing sites into two groups

– *P1'* proline and non-*P1'* proline sites. Other work has already concluded that the *P1'* proline sites prefer large, aromatic residues in the *P1* position (46, 47, 287, 295, 298), and non-beta-branched amino acids in the *P2* position (288). From my dataset of 81 substrates, I confirmed these results, and further suggest a second site-defining amino acid to be a lysine in the *P3* position. Lysine was associated with a number of inactive or poorly cleaved sites. I suspect this was because it attempts to force the processing site into an unconventional structural arrangement where the *P4* amino acid occupies the position normally expected of the *P3* amino acid, an arrangement that is particular to only a pair of sites (48). I also noticed a pattern in the composition of efficiently cleaved sites containing a lysine in the *P3* position. These sites required a beta-branched amino acid in the *P2* position, and strongly preferred a non-aromatic amino acid in the *P1* site. Notably, these preferences conflicted with those of sites with a *P1'* proline. Because the presence of a lysine in the *P3* position could be used to predict the optimal layout of a site, I labeled it as a site-defining amino acid.

Although several patterns could be identified by eye, a more detailed examination of the dataset was gained by the application of statistical modeling procedures. In order to compare each site, I applied nine different physicochemical properties to each amino acid. Having utilized multiple HIV-1 processing sites as a starting point, the effect of each amino acid substitution was partially hidden by the baseline rate of the cleavage site from which it was derived. As a means of overcoming this obstacle, I used linear mixed-effects modeling. This procedure grouped the data according to whichever cleavage site was used as its reference site. In doing this, more than 50% of the variability was removed from the data, all of which was effectively noise.

The first analysis considered each site as a whole entity – in other words, the location of any differences had no bearing on the result. This model was actually the most effective descriptive analysis among all the models that I built, at least when the principal of maximum parsimony is also considered. The most important physicochemical properties included hydrophobicity, net charge, total polarizability (a volumetric measurement of an object's response to an external electric field, i.e. the acquisition of a dipole moment), and the amount of polar surface area available. I also found an interaction existed between the hydrophobicity and amount of polar surface area. Substrates that were more hydrophilic than hydrophobic did not vary much in rate, but those that were hydrophobic tended to be preferred cleavage sites if they had more polar surface area. Thus, better sites were comprised of strongly hydrophobic amino acids intermixed with polar amino acids.

Additional mixed-effects modeling analyses looking at individual amino acid residues within the sites were more complicated to perform. While the dataset was large compared to similar studies previously published, it was still only of modest size. This limited the number of variables a model could contain before causing an error in the statistical procedures. Grouping the data further limited this number. The solution I decided upon was to look at each property individually. From these analyses, the *P2*, *P1'*, *P2'*, and *P4'* positions were strongly implicated as important to the determination of processing rate. Furthermore, a relationship existed between the hydrophobicity of the *P2* and *P4'* amino acids, and the polarizability of the *P2* and *P2'* amino acids. The overlap of these two interactions suggests we may think of these three amino acids as a unit, and that a limited number of optimal or functional combinations may exist. These units are not necessarily site-defining amino acids, as both the P1' proline and P3 lysine were the amino acids that imposed restrictions on the P2 amino acid (and by extension the P2' and P4' prime amino acid), rather than the reverse. Nevertheless, their interrelatedness simplifies otherwise pattern-deficient HIV-1 cleavage sites.

The positional interactions were not only on opposite sides of the scissile bond, but were also *trans*-interactions. This implied that the interaction of an amino acid with one monomer of the HIV-1 PR might actually affect the relationship of the *other* PR monomer with an amino acid in the cleavage site. These results highlighted not just the interdependence of amino acids within a cleavage site, but also the interplay between monomers of the HIV-1 PR. This give-and-take between halves of the enzyme would ostensibly impart the asymmetry the homodimeric PR requires to recognize an amino acid sequence (49, 274).

While a descriptive analysis was useful, ultimately I sought to build predictive models that could be utilized to identify an idealized amino acid sequence. For this, each observation was treated independently and non-mixed-effects modeling techniques were employed. Given there exist 25.6 billion (20^8) possible amino acid combinations as potential substrates for the HIV-1 PR, I could not use mixed-effects techniques because only a very small percentage of the total catalog of substrates could be grouped with any of the current set of cleavage sites. The resulting models suffered accordingly, as it is more likely they modeled the noise in the data than the actual patterns. Analyzing the distribution patterns in the top 15 (of 1570) predictive models revealed that most of the terms in the models included positions deemed relatively unimportant by the mixed-effects modeling procedure (e.g. *P4*, *P3*, *P3'*). As a result, the ability to predict out-of-sample data, albeit data derived from multiple sources and obtained under low pH/high salt conditions, was extremely poor. The models were slightly more effective as classifiers of fast and slow sites, however.

Many others have applied statistical and machine learning algorithms to the task of separating active and inactive cleavage sites. As one final analysis, I employed my set of statistical models to this alternative classification test as well. While none of the models were competitive with any of the previously published results, I did note an inverse correlation between the ability of my models to be a rate-classifier and their ability to be an active/inactive classifier. Since all of the naturally occurring cleavage sites only require one amino acid substitution to become inactive (*P1* isoleucine), this result is not altogether surprising. At least on a physicochemical basis, the ability of a model to distinguish between a functional site with eight non-ideal amino acids and a non-functional site with seven ideal amino acids and one damaging amino acid could be limited. Thus, the identification of an ideal amino acid sequence by physicochemical descriptors will require a two-step procedure: one step in which all cleavage sites violating a set of cleavability rules are removed, and one in which the rate of each site is predicted.

The results presented in Chapter Three were the product of my first attempt at building descriptive and predictive models. The immediate future directions for this research entail the revision and reapplication of the methods. All of the predictive models were generated independently of the mixed-effect modeling results, and they were inferior to those models according to most statistical criteria. It is possible to force the models to include the terms identified by the mixed-effects modeling procedures, and that is an avenue worth exploring given the failure of the current set of predictive models. An expansion of the dataset is also warranted. I made a number of site-specific mutations, so expanding the dataset with sites based on these predictions would also provide a more worthwhile out-of-sample dataset, especially since the results would be generated under the same conditions as the data used to create the models. Additionally, the new information will help refine and improve the models when they are eventually refit to the entire set that is collected. I also identified several potential patterns within characteristic of cleavage sites in general (i.e. the ability to predict site composition by knowing the *P1'*, *P3*, or even one of the *P2-P2'-P4'* amino acids), and a more targeted set of mutations to test those patterns would help in confirming those results.

Once the predictive modeling procedures produce a functional model, use of this model in an algorithm that cycles through each of the 25.6 billion permutations of eight amino acids will identify candidate combinations for the ideal cleavage site sequence. When this sequence has been identified, it will naturally be tested. Following that, the long-term goal of this research is its application to drug design. With an idealized sequence, small molecule substrate mimetics could be designed and tested for their inhibitory abilities. Additional refinements to already-existent HIV-1 PIs could also be an application for the results.

A series of outstanding questions about the basic virology of HIV-1 could also be addressed using the information gathered in this dataset. While the order of processing was defined long ago (44, 45), the actual length of time available to complete each step remains unknown. This could be monitored by the adjustment of the relative rates of processing. Additionally, special conditions could be present during the assembly and budding process that promote ordered processing. A test of whether or not slowing the kinetics of processing is detrimental to infectivity – even if the relative order is maintained – is another interesting research question. Furthermore, do the enzymes have a structural role in generating the mature virus architecture, or does the timing of cleavage at the RTH/IN site, for example, have no bearing on infectivity? IN has been implicated in encapsulation of the RNP core by the CA shell (192, 193), suggesting the timing of release for IN may be important for virus infectivity.

In Chapter One, I discussed the key role the HIV-1 PR plays in the virus lifecycle. This enzyme completes a highly complicated series of cleavage events within a newly created virus particle to prepare it for infection of a new cell. The two chapters thereafter detailed a pair of studies where I uncovered potential mechanisms controlling HIV-1 PR activity throughout this process. The second chapter identified RNA as a prospective binding partner for the HIV-1 PR, an interaction that greatly enhances the catalytic efficiency of the enzyme *in vitro*. In the third chapter, I explored the sequence specificity of the HIV-1 PR. The physicochemical factors affecting processing rate were defined under conditions more similar to those encountered by the PR in the cell, providing a necessary upgrade to the specificity studies of previously published work. Lastly, in this fourth chapter, I reviewed my results and presented a selection of possible applications of the results to future research. In conclusion, I have addressed several unresolved questions about the complex interplay between the HIV-1 PR and its substrate, and I have additionally developed a number of exciting new research directions worthy of future investigation.

REFERENCES

1. **Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C.** 1983. Isolation of T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). Science **220:**868-871.

2. **World Health Organization.** July 2015 2015. HIV/AIDS Fact Sheet No. 360. http://www.who.int/mediacentre/factsheets/fs360/en/.

3. **Bukrinsky MI, Sharova N, Dempsey MP, Stanwick TL, Bukrinskaya AG, Haggerty S, Stevenson M.** 1992. Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. Proc Natl Acad Sci **89:**6580-6584.

4. **Gartner S, Markovits P, Markovitz DM, Kaplan MH, Gallo RC, Popovic M.** 1986. The role of mononuclear phagocytes in HTLV-III/LAV infection. Science **233:**215-219.

5. **Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM.** 2008. The Challenge of HIV-1 Subtype Diversity. N Engl J Med **358:**1590-1602.

6. **Malim MH, Emerman M.** 2008. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe **3:**388-398.

7. **Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE.** 1988. Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. Nature **331:**280-283.

8. **Ono A, Freed EO.** 2001. Plasma membrane rafts play a critical role in HIV-1 assembly and release. Proc Natl Acad Sci **98:**13925-13930.

9. **Hogue IB, Grover JR, Soheilian F, Nagashima K, Ono A.** 2011. Gag induces the coalescence of clustered lipid rafts and tetraspanin-enriched microdomains at HIV-1 assembly sites on the plasma membrane. J Virol **85:**9749-9766.

10. **Göttlinger HG, Sodroski JG, Haseltine WA.** 1989. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proc Natl Acad Sci USA **86:**5781-5785.

11. **Bryant M, Ratner L.** 1990. Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. Proc Natl Acad Sci USA **87:**523-527.

12. **Kutluay SB, Bieniasz PD.** 2010. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. PLoS Pathog **6:**e1001200.

13. **Moore MD, Nikolaitchik OA, Chen J, Hammarskjöld ML, Rekosh D, Hu WS.** 2009. Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses. PLoS Pathog **5:**e1000627.

14. **Muriaux D, Mirro J, Harvin D, Rein A.** 2001. RNA is a structural element in retrovirus particles. Proc Natl Acad Sci **98:**5246-5251.

15. **Khorchid A, Halwani R, Wainberg MA, Kleiman L.** 2002. Role of RNA in facilitating gag/gag-pol interaction. J Virol **76:**4131-4137.

16. **Bieniasz PD.** 2009. The cell biology of HIV-1 virion genesis. Cell Host Microbe **5:**550-558.

17. **Weiss ER, Göttlinger H.** 2011. The role of cellular factors in promoting HIV budding. J Mol Biol **410:**525-533.

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

19. **Carlson L-A, Briggs JAG, Glass B, Riches JD, Simon MN, Johnson MC, Müller B, Grünewald K, Kräusslich H-G.** 2008. Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. Cell Host Microbe **4:**592-599.

20. **Kaplan AH, Manchester M, Swanstrom R.** 1994. The activity of the protease of human immunodeficiency virus type 1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency. J Virol **68:**6782-6786.

21. **Swanstrom R, Wills JW.** 1997. Synthesis, assembly, and processing of viral proteins. *In* Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

22. **Murakami T, Ablan S, Freed EO, Tanaka Y.** 2004. Regulation of human immunodeficiency virus type 1 env-mediated membrane fusion by viral protease activity. J Virol **78:**1026-1031.

23. **Wyma DJ, Jiang J, Shi J, Zhou J, Lineberger JE, Miller MD, Aiken C.** 2004. Coupling of human immunodeficiency virus type 1 fusion to virion maturation: a novel role of the gp41 cytoplasmic tail. J Virol **78:**3429-3435.

24. **Jiang J, Aiken C.** 2006. Maturation of the viral core enhances the fusion of HIV-1 particles with primary human T cells and monocyte-derived macrophages. Virology **346:**460-468.

25. **Rabi SA, Laird GM, Durand CM, Laskey S, Shan L, Bailey JR, Chioma S, Moore RD, Siliciano RF.** 2013. Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. J Clin Invest **123:**3848-3860.

26. **Kaplan AH, Zack JA, Knigge M, Paul DA, Kempf DJ, Norbeck DW, Swanstrom R.** 1993. Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles. J Virol **67:**4050-4055.

27. **de la Carriére LC, Paulous S, Clavel F, Mammano F.** 1999. Effects of human immunodeficiency virus type 1 resistance to protease inhibitors on reverse transcriptase processing, activity, and drug sensitivity. J Virol **73:**3455-3459.

28. **Moore MD, Fu W, Soheilian F, Nagashima K, Ptak RG, Pathak VK, Hu WS.** 2008. Suboptimal inhibition of protease activity in human immunodeficiency virus type 1: effects on virion morphogenesis and RNA maturation. Virology **379:**152-160.

29. **Lee SK, Harris J, Swanstrom R.** 2009. A strongly transdominant mutation in the human immunodeficiency virus type 1 gag gene defines an Achilles heel in the virus life cycle. J Virol **83:**8536-8543.

30. **Muller B, Anders M, Akiyama H, Welsch S, Glass B, Nikovics K, Clavel F, Tervo HM, Keppler OT, Krausslich HG.** 2009. HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 infectivity. J Biol Chem **284:**29692-29703.

31. **Pearl LH, Taylor WR.** 1987. A structural model for retroviral proteases. Nature **329:**351-354.

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

33. **Navia MA, Fitzgerald PMD, McKeever BM, Leu C, 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.

34. **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.

35. **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.

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

37. **Lindhofer H, von der Helm K, Nitschko H.** 1995. *In vivo* processing of Pr160gag-pol from human immunodeficiency virus type 1 (HIV) in acutely infected, cultured human T-lymphocytes. Virology **214:**624-627.

38. **Sluis-Cremer N, Arion D, Abram ME, Parniak MA.** 2004. Proteolytic processing of an HIV-1 pol polyprotein precursor: insights into the mechanism of reverse transcriptase p66/p51 heterodimer formation. Int J Biochem Cell Biol **36:**1836-1847.

39. **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.

40. **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.

41. **Wondrak EM, Louis JM.** 1996. Influence of flanking sequences on the dimer stability of human immunodeficiency virus type 1 protease. Biochemistry **35:**12957-12962.

42. **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:**66-71.

43. **Erickson-Viitanen S, Manfredi J, Viitanen P, Tribe DE, Tritch R, Hutchison III CA, Loeb DD, Swanstrom R.** 1989. Cleavage of HIV-1 *gag* polyprotein synthesized in vitro: sequential cleavage by the viral protease. AIDS Res Hum Retrov **5:**577-591.

44. **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.

45. **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.

46. **Pettit SC, Henderson GJ, Schiffer CA, Swanstrom R.** 2002. Replacement of the P1 amino acid of human immunodeficiency virus type 1 gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. J Virol **76:**10226-10233.

47. **Pettit SC, Simsic J, Loeb DD, Everitt LE, Hutchison III CA, Swanstrom R.** 1991. Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. J Biol Chem **266:**14539-14547.

48. **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.

49. **Ozen A, Haliloglu T, Schiffer CA.** 2011. Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. J Mol Biol **410:**726-744.

50. **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.

51. **Checkley MA, Luttge BG, Soheilian F, Nagashima K, Freed EO.** 2010. The capsid-spacer peptide 1 Gag processing intermediate is a dominant-negative inhibitor of HIV-1 maturation. Virology **400:**137-144.

52. **Gulick RM, Mellors JW, Havlir D, Eron JJ, Gonzalez C, McMahon D, Richman DD, Valentine FT, Jonas L, Meibohm A, Emini EA, Chodakewitz JA.** 1997. Treatment with indinavir, zidovudine, and lamivudine in adults with human

immunodeficiency virus infection and prior antriretroviral therapy. N Engl J Med **337:**735-739.

53. **Palella FJ, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, Aschman DJ, Holmberg SD, Investigators HOS.** 1998. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. N Engl J Med **338:**853-860.

54. **Miller M.** 2010. The early years of retroviral protease crystal structures. Biopolymers **94:**521-529.

55. **Schechter I, Berger A.** 1967. On the size of the active site in proteases. I. Papain. 1967. Biochem Biophys Res Commun **27:**157-162.

56. **Roberts NA, Martin JA, Kinchington D, Broadhurst AV, Craig JC, Duncan IB, Galpin SA, Handa BK, Kay J, Kröhn A, Lambert RW, Merrett JH, Mills JS, Parkes KEB, Redshaw S, Ritchie AJ, Taylor DL, Thomas GJ, Machin PJ.** 1990. Rational design of peptide-based HIV proteinase inhibitors. Science **248:**358-361.

57. **Kröhn A, Redshaw S, Ritchie JC.** 1991. Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isotere. J Med Chem **34:**3340-3342.

58. **Kempf DJ, Marsh KC, Denissen JF, McDonald E, Vasavanonda S, Flentge CA, Green BE, Fino L, Park CH, Kong XP, Wideburg NE, Saldivar A, Ruiz L, Kati WM, Sham HL, Robins T, Stewart KD, Hsu A, Plattner JJ, Leonard JM, Norbeck DW.** 1995. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability. Proc Natl Acad Sci USA **92:**2484-2488.

59. **Kaldor SW, Kalish VJ, Davis JF, Shetty BV, Fritz JE, Appelt K, Burgess JA, Campanale KM, Chirgadze NY, Clawson DK, Dressman BA, Hatch SD, Khalil DA, Kosa MB, Lubbehusen PP, Muesing MA, Patick AK, Reich SH, Su KS, Tatlock JH.** 1997. Viracept (Nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. J Med Chem **40:**3979-3985.

60. **Thaisrivongs S, Strohbach JW.** 1999. Structure-based discovery of Tipranavir disodium (PNU-140690E): a potent, orally bioavailable, nonpeptidic HIV proteased inhibitors. Biopolymers **51:**51-58.

61. **Stoll V, Qin W, Stewart KD, Jakob C, Park C, Walter K, Simmer RL, Helfrich R, Bussiere D, Kao J, Kempf D, Sham HL, Norbeck DW.** 2002. X-ray crystallographic stucture of ABT-378 (Lopinavir) bound to HIV-1 protease. Bioorg Med Chem **10:**2803-2806.

62. **Dreyer GB, Metcalf BW, Tomaszek Jr. TA, Carr TJ, Chandler III AC, Hyland L, Fakhoury SA, Magaard VW, Moore ML, Strickler JE, Debouck C, Meek TD.** 1989. Inhibition of human immunodeficiency virus 1 protease *in vitro*: rational design of substrate analogue inhibitors. Proc Natl Acad Sci USA **86:**9752-9756.

63. **Babé LM, Rosé J, Craik CS.** 1995. Trans-dominant inhibitory human immunodeficiency virus type 1 protease monomers prevent protease activation and virion maturation. Proc Natl Acad Sci **92:**10069-10073.

64. **Julias JG, Ferris AL, Boyer PL, Hughes SH.** 2001. Replication of phenotypically mixed human immunodeficiency virus type 1 virions containing catalytically active and catalytically inactive reverse transcriptase. J Virol **75:**6537-6546.

65. **Shen L, Peterson S, Sedaghat AR, McMahon MA, Callender M, Zhang H, Zhou Y, Pitt E, Anderson KS, Acosta EP, Siliciano RF.** 2008. Dose-response curve slope sets class-specific limits on inhibitory potential of anti-HIV drugs. Nature Med **14:**762-766.

66. **Shen L, Rabi SA, Sedaghat AR, Shan L, Lai J, Xing S, Siliciano RF.** 2011. A critical subset model provides a conceptual basis for the high antiviral activity of major HIV drugs. Sci Transl Med **3:**91ra63-91ra63.

67. **Sampah MES, Shen L, Jilek BL, Siliciano RF.** 2011. Dose-response curve slope is a missing dimension in the analysis of HIV-1 drug resistance. Proc Natl Acad Sci **108:**7613-7618.

68. **Henderson GJ, Lee SK, Irlbeck DM, Harris J, Kline M, Pollom E, Parkin N, Swanstrom R.** 2012. Interplay between single resistance-associated mutations in the HIV-1 protease and viral infectivity, protease activity, and inhibitor sensitivity. Antimicrob Agents Chemother **56:**623-633.

69. **Ambrose Z, Julias JG, Boyer PL, KewalRamani VN, Hughes SH.** 2006. The level of reverse transcriptase (RT) in human immunodeficiency virus type 1 particles affects susceptibility to nonnucleoside RT inhibitors but not to Lamivudine. J Virol **80:**2578-2581.

70. **Jilek BL, Zarr M, Sampah ME, Rabi SA, Bullen CK, Lai J, Shen L, Siliciano RF.** 2012. A quantitative basis for antiretroviral therapy for HIV-1 infection. Nature Med **18:**446-451.

71. **Arribas JR, Horban A, Gerstoft J, Fätkenheuer G, Nelson M, Clumeck N, Pulido F, Hill A, van Delft Y, Stark T, Moecklinghoff C.** 2010. The MONET trial: darunavir/ritonavir with or without nucleoside analogues, for patients with HIV RNA below 50 copies/ml. AIDS **24:**223-230.

72. **Bierman WFW, van Agtmael MA, Nijhuis M, Danner SA, Boucher CAB.** 2009. HIV monotherapy with ritonavir-boosted protease inhibitors: a systematic review. AIDS **23:**279-291.

73. **Wilen CB, Tilton JC, Doms RW.** 2012. HIV: cell binding and entry. Cold Spring Harb Perspect Med **2:**a006866-a006866.

74. **Hallenberger S, Bosch V, Angliker H, Shaw E, Klenk HD, Garten W.** 1992. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. Nature **360:**358-361.

75. **White TA, Bartesaghi A, Borgnia MJ, Meyerson JR, de la Cruz MJV, Bess JW, Nandwani R, Hoxie JA, Lifson JD, Milne JLS, Subramaniam S.** 2010. Molecular architectures of trimeric SIV and HIV-1 envelope glycoproteins on intact viruses: strain-dependent variation in quaternary structure. PLoS Pathog **6:**e1001249.

76. **Arrildt KT, Joseph SB, Swanstrom R.** 2012. The HIV-1 env protein: a coat of many colors. Curr HIV/AIDS Rep **9:**52-63.

77. **Freed EO, Myers DJ, Risser R.** 1990. Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. Proc Natl Acad Sci USA **87:**4650-4654.

78. **Haffar OK, Dowbenko DJ, Berman PW.** 1988. Topogenic analysis of the human immunodeficiency virus type 1 envelope glycoprotein, gp160, in microsomal membranes. J Cell Biol **107:**1677-1687.

79. **Muranyi W, Malkusch S, Müller B, Heilemann M, Kräusslich H-G.** 2013. Super-resolution microscopy reveals specific recruitment of HIV-1 envelope proteins to viral assembly sites dependent on the envelope C-terminal tail. PLoS Pathog **9:**e1003198.

80. **Roy NH, Chan J, Lambele M, Thali M.** 2013. Clustering and mobility of HIV-1 env at viral assembly sites predict its propensity to induce cell-cell fusion. J Virol **87:**7516-7525.

81. **Chertova E, Bess JW, Crise BJ, Sowder Ii RC, Schaden TM, Hilburn JM, Hoxie JA, Benveniste RE, Lifson JD, Henderson LE, Arthur LO.** 2002. Envelope glycoprotein incorporation, not shedding of surface envelope glycoprotein (gp120/SU), is the primary determinant of SU content of purified human immunodeficiency virus type 1 and simian immunodeficiency virus. J Virol **76:**5315-5325.

82. **Zhu P, Chertova E, Bess J, Lifson JD, Arthur LO, Liu J, Taylor KA, Roux KH.** 2003. Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. Proc Natl Acad Sci **100:**15812-15817.

83. **Zhu P, Liu J, Bess J, Chertova E, Lifson JD, Grisé H, Ofek GA, Taylor KA, Roux KH.** 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature **441:**847-852.

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

85. **Sattentau QJ, Moore JP.** 1991. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. J Exp Med **174:**407-415.

86. **Liu J, Bartesaghi A, Borgnia MJ, Sapiro G, Subramaniam S.** 2008. Molecular architecture of native HIV-1 gp120 trimers. Nature **455:**109-113.

87. **Checkley MA, Luttge BG, Freed EO.** 2011. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. J Mol Biol **410:**582-608.

88. **Chan DC, Fass D, Berger JM, Kim PS.** 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell **89:**263-273.

89. **Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC.** 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature **387:**426-430.

90. **Rein A, Mirro J, Haynes JG, Ernst SM, Nagashima K.** 1994. Function of the cytoplasmic domain of a retroviral transmembrane protein: p15E-p2E cleavage activates

the membrane fusion capability of the murine leukemia virus env protein. J Virol **68:**1773-1781.

91. **Wyma DJ, Kotov A, Aiken C.** 2000. Evidence for a stable interaction of gp41 with Pr55gag in immature human immunodeficiency virus type 1 particles. J Virol **74:**9381-9387.

92. **Chojnacki J, Staudt T, Glass B, Bingen P, Engelhardt J, Anders M, Schneider J, Muller B, Hell SW, Krausslich HG.** 2012. Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. Science **338:**524-528.

93. **Meek TD, Lambert DM, Dreyer GB, Carr TJ, Tomaszek TA, Moore ML, Strickler JE, Debouck C, Hyland LJ, Matthews TJ, Metcalf BW, Petteway SR.** 1990. Inhibition of HIV-1 protease in infected T-lymphocytes by synthetic peptide analogues. Nature **343:**90-92.

94. **Bhattacharya J, Repik A, Clapham PR.** 2006. Gag regulates association of human immunodeficiency virus type 1 envelope with detergent-resistant membranes. J Virol **80:**5292-5300.

95. **Yu X, Yuan X, Matsuda Z, Lee TH, Essex M.** 1992. The matrix protein of human immunodeficiency virus type 1 is required for incorporation of viral envelope protein into mature virions. J Virol **66:**4966-4971.

96. **Dorfman T, Mammano F, Haseltine WA, Göttlinger HG.** 1994. Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. J Virol **68:**1689-1696.

97. **Mammano F, Kondo E, Sodroski J, Bukovsky A, Göttlinger HG.** 1995. Rescue of human immunodeficiency virus type 1 matrix protein mutants by envelope glycoproteins with short cytoplasmic domains. J Virol **69:**3824-3830.

98. **Murakami T, Freed EO.** 2000. Genetic evidence for an interaction between human immunodeficiency virus type 1 matrix and α-helix 2 of the gp41 cytoplasmic tail. J Virol **74:**3548-3554.

99. **Freed EO, Martin MA.** 1995. Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. J Virol **69:**1984-1989.

100. **Brandano L, Stevenson M.** 2012. A highly conserved residue in the C-terminal helix of HIV-1 matrix is required for envelope incorporation into virus particles. J Virol **86:**2347-2359.

101. **Tedbury PR, Ablan SD, Freed EO.** 2013. Global rescue of defects in HIV-1 envelope glycoprotein incorporation: implications for matrix structure. PLoS Pathog **9:**e1003739.

102. **Alfadhli A, Barklis RL, Barklis E.** 2009. HIV-1 matrix organizes as a hexamer of trimers on membranes containing phosphatidylinositol-(4,5)-bisphosphate. Virology **387:**466-472.

103. **Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sudquist WI.** 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. Proc Natl Acad Sci USA **93:**3099-3104.

104. **Edwards TG, Wyss S, Reeves JD, Zolla-Pazner S, Hoxie JA, Doms RW, Baribaud F.** 2002. Truncation of the cytoplasmic domainiInduces exposure of conserved regions in the ectodomain of human immunodeficiency virus type 1 envelope protein. J Virol **76:**2683-2691.

105. **Kalia V, Sarkar S, Gupta P, Montelaro RC.** 2005. Antibody neutralization escape mediated by point mutations in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41. J Virol **79:**2097-2107.

106. **Wyss S, Dimitrov AS, Baribaud F, Edwards TG, Blumenthal R, Hoxie JA.** 2005. Regulation of human immunodeficiency virus type 1 envelope glycoprotein fusion by a membrane-interactive domain in the gp41 cytoplasmic tail. J Virol **79:**12231-12241.

107. **Joyner AS, Willis JR, Crowe Jr JE, Aiken C.** 2011. Maturation-induced cloaking of neutralization epitopes on HIV-1 particles. PLoS Pathog **7:**e1002234.

108. **Abrahamyan LG, Mkrtchyan SR, Binley J, Lu M, Melikyan GB, Cohen FS.** 2005. The cytoplasmic tail slows the folding of human immunodeficiency virus type 1 env from a late prebundle configuration into the six-helix bundle. J Virol **79:**106-115.

109. **Steckbeck JD, Craigo JK, Barnes CO, Montelaro RC.** 2011. Highly conserved structural properties of the C-terminal tail of HIV-1 gp41 protein despite substantial sequence variation among diverse clades: implications for functions in viral replication. J Biol Chem **286:**27156-27166.

110. **Lucas TM, Lyddon TD, Grosse SA, Johnson MC.** 2010. Two distinct mechanisms regulate recruitment of murine leukemia virus envelope protein to retroviral assembly sites. Virology **405:**548-555.

111. **Hu WS, Hughes SH.** 2012. HIV-1 Reverse Transcription. Cold Spring Harb Perspect Med **2:**a006882-a006882.

112. **Le Grice SFJ.** 2012. Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise. J Biol Chem **287:**40850-40857.

113. **Wapling J, Moore KL, Sonza S, Mak J, Tachedjian G.** 2005. Mutations that abrogate human immunodeficiency virus type 1 reverse transcriptase dimerization affect maturation of the reverse transcriptase heterodimer. J Virol **79:**10247-10257.

114. **Zheng X, Pedersen LC, Gabel SA, Mueller GA, Cuneo MJ, DeRose EF, Krahn JM, London RE.** 2014. Selective unfolding of one ribonuclease H domain of HIV reverse transcriptase is linked to homodimer formation. Nucleic Acids Res doi:10.1093/nar/gku143.

115. **Venezia CF, Howard KJ, Ignatov ME, Holladay LA, Barkley MD.** 2006. Effects of Efavirenz binding on the subunit equilibria of HIV-1 reverse transcriptase. Biochemistry **45:**2779-2789.

116. **Venezia CF, Meany BJ, Braz VA, Barkley MD.** 2009. Kinetics of association and dissociation of HIV-1 reverse transcriptase subunits. Biochemistry **48:**9084-9093.

117. **Lowe DM, Aitken A, Bradley C, Darby GK, Larder BA, Powell KL, Purifoy DJM, Tisdale M, Stammers DK.** 1988. HIV-1 reverse transcriptase: crystallization and analysis of domain structure by limited proteolysis. Biochemistry **27:**8884-8889.

118. **Bathurst IC, Moen LK, Lujan MA, Gibson HL, Feucht PH, Pichuantes S, Craik CS, Santi DV, Barr PJ.** 1990. Characterization of the human immunodeficiency virus type-1 reverse transcriptase enzyme produced in yeast. Biochem Biophys Res Commun **171:**589-595.

119. **Chattopadhyay D, Evans DB, Deibel Jr MR, Vosters AF, Eckenrode FM, Einspahr HM, Hui JO, Tomasselli AG, Zurcher-Neely HA, Heinrikson RL, Sharma SK.** 1992. Purification and characterization of heterodimeric human immunodeficiency virus type 1 (HIV-1) reverse transcriptase produced by *in vitro* processing of p66 with recombinant HIV-1 protease. J Biol Chem **267:**14227-14232.

120. **Zhang H, Dornadula G, Pomerantz RJ.** 1996. Endogenous reverse transcription of human immunodeficiency virus type 1 in physiological microenvironments: an important stage for viral infection of nondividing cells. J Virol **79:**2809-2824.

121. **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.

122. **Gelderblom HR, Hausmann EHS, Özel M, Pauli G, Koch MA.** 1987. Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. Virology **156:**171-176.

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

124. **Nermut MV, Fassati A.** 2003. Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes. J Virol **77:**8196-8206.

125. **Iordanskiy S, Berro R, Altieri M, Kashanchi F, Bukrinsky M.** 2006. Intracytoplasmic maturation of the human immunodeficiency virus type 1 reverse transcription complexes determines their capacity to integrate into chromatin. Retrovirology **3:**4.

126. **Carr JM, Coolen C, Davis AJ, Burrell CJ, Li P.** 2008. Human immunodeficiency virus 1 (HIV-1) virion infectivity factor (Vif) is part of reverse transcription complexes and acts as an accessory factor for reverse transcription. Virology **372:**147-156.

127. **Kotov A, Zhou J, Flicker P, Aiken C.** 1999. Association of Nef with the human immunodeficiency virus type 1 core. J Virol **73:**8824-8830.

128. **Forshey BM, Aiken C.** 2003. Disassembly of human immunodeficiency virus type 1 cores in vitro reveals association of nef with the subviral ribonucleoprotein complex. J Virol **77:**4409-4414.

129. **McDonald D, Vodicka MA, Svitkina TM, Borisy GG, Emerman M, Hope TJ.** 2002. Visualization of the intracellular behavior of HIV in living cells. J Cell Biol **159:**441-452.

130. **Bukrinskaya A, Brichacek B, Mann A, Stevenson M.** 1998. Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton. J Exp Med **188:**2113-2125.

131. **Arfi V, Lienard J, Nguyen XN, Berger G, Rigal D, Darlix JL, Cimarelli A.** 2009. Characterization of the behavior of functional viral genomes during the early steps of human immunodeficiency virus type 1 infection. J Virol **83:**7524-7535.

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

133. **Yang Y, Fricke T, Diaz-Griffero F.** 2013. Inhibition of reverse transcriptase activity increases stability of the HIV-1 core. J Virol **87:**683-687.

134. **Mirambeau G, Lyonnais S, Coulaud D, Hameau L, Lafosse S, Jeusset J, Borde I, Reboud-Ravaux M, Restle T, Gorelick RJ, Le Cam E.** 2007. HIV-1 protease and reverse transcriptase control the architecture of their nucleocapsid partner. PLoS One **2:**e669.

135. **Suzuki Y, Craigie R.** 2007. The road to chromatin — nuclear entry of retroviruses. Nature Rev Microbiol **5:**187-196.

136. **Lyonnais S, Gorelick RJ, Heniche-Boukhalfa F, Bouaziz S, Parissi V, Mouscadet J-F, Restle T, Gatell JM, Le Cam E, Mirambeau G.** 2013. A protein ballet around the viral genome orchestrated by HIV-1 reverse transcriptase leads to an architectural switch: from nucleocapsid-condensed RNA to vpr-bridged DNA. Virus Res **171:**287-303.

137. **Coren LV, Thomas JA, Chertova E, Sowder RC, Gagliardi TD, Gorelick RJ, Ott DE.** 2007. Mutational analysis of the C-terminal gag cleavage sites in human immunodeficiency virus type 1. J Virol **81:**10047-10054.

138. **Zennou V, Mammano F, Paulous S, Mathez D, Clavel F.** 1998. Loss of viral fitness associated with multiple gag and gag-pol processing defects in human immunodeficiency virus type 1 variants selected for resistance to protease inhibitors in vivo. J Virol **72:**3300-3306.

139. **Bleiber G, Munoz M, Ciuffi A, Meylan P, Telenti A.** 2001. Individual contributions of mutant protease and reverse transcriptase to viral infectivity, replication, and protein maturation of antiretroviral drug-resistant human immunodeficiency virus type 1. J Virol **75:**3291-3300.

140. **García Lerma JG, Yamamoto S, Gómez-Cano M, Soriano V, Green TA, Busch MP, Folks TM, Heneine W.** 1998. Measurement of human immunodeficiency virus type 1

plasma virus load based on reverse transcriptase (RT) activity: evidence of variabilities in levels of virion-associated RT. J Infect Dis **177:**1221-1229.

141. **Marozsan AJ, Fraundorf E, Abraha A, Baird H, Moore D, Troyer R, Nankja I, Arts EJ.** 2004. Relationships between infectious titer, capsid protein levels, and reverse transcriptase activities of diverse human immunodeficiency virus type 1 isolates. J Virol **78:**11130-11141.

142. **Wang J, Bambara RA, Demeter LM, Dykes C.** 2010. Reduced fitness in cell culture of HIV-1 with nonnucleoside reverse transcriptase inhibitor-resistant mutations correlates with relative levels of reverse transcriptase content and RNase H activity in virions. J Virol **84:**9377-9389.

143. **Lori F, Scovassi AI, Zella D, Achilli G, Cattaneo E, Casoli C, Bertazzoni U.** 1998. Enzymatically active forms of reverse transcriptase of the human immunodeficiency virus. AIDS Res Hum Retroviruses **4:**393-398.

144. **Kawamura M, Shimano R, Inubushi R, Amano K, Ogasawara T, Akari H, Adachi A.** 1997. Cleavage of gag precursor is required for early replication phase of HIV-1. FEBS Lett **415:**227-230.

145. **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 **108:**9072-9077.

146. **Davis DA, Soule EE, Davidoff KS, Daniels SI, Naiman NE, Yarchoan R.** 2012. Activity of human immunodeficiency virus type 1 protease inhibitors against the initial autocleavage in gag-pol polyprotein processing. Antimicrob Agents Chemother **56:**3620-3628.

147. **Sekar V, Lavreys L, Van de Casteele T, Berckmans C, Spinosa-Guzman S, Vangeneugden T, De Pauw M, Hoetelmans R.** 2010. Pharmacokinetics of Darunavir/Ritonavir and Rifabutin coadministered in HIV-negative healthy volunteers. Antimicrob Agents Chemother **54:**4440-4445.

148. **Schatz O, Cromme FV, Grüninger-Leitch F, Le Grice SFJ.** 1989. Point mutations in conserved amino acid residues within the C-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. FEBS Lett **257:**311-314.

149. **Fletcher RS, Holleschak G, Nagy E, Arion D, Borkow G, Gu Z, Wainbeg MA, Parniak MA.** 1996. Single-step purification of recombinant wild-type and mutant HIV-1 reverse transcriptase. Protein Expr Purif **7:**27-32.

150. **Monroe KM, Yang Z, Johnson JR, Geng X, Doitsh G, Krogan NJ, Greene WC.** 2014. IFI16 DNA sensor is required for death of lymphoid CD4 T cells abortively infected with HIV. Science **343:**428-432.

151. **Shehu-Xhilaga M, Kraeusslich HG, Pettit S, Swanstrom R, Lee JY, Marshall JA, Crowe SM, Mak J.** 2001. Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. J Virol **75:**9156-9164.

152. **de Marco A, Müller B, Glass B, Riches JD, Kräusslich HG, Briggs JAG.** 2010. Structural analysis of HIV-1 maturation using cryo-electron tomography. PLoS Pathog **6:**e1001215.

153. **von Schwedler UK, Stemmler TL, Klishko VY, Li S, Albertine KH, Davis DR, Sundquist WI.** 1998. Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. EMBO J **17:**1555-1568.

154. **Fitzon T, Leschonsky B, Bieler K, Paulus C, Schröder J, Wolf H, Wagner R.** 2000. Proline residues in the HIV-1 NH2-terminal capsid domain: structure determinants for proper core assembly and subsequent steps of early replication. Virology **268:**294-307.

155. **Tang S, Murakami T, Agresta BE, Campbell S, Freed EO, Levin JG.** 2001. Human immunodeficiency virus type 1 N-terminal capsid mutants that exhibit aberrant core morphology and are blocked in initiation of reverse transcription in infected cells. J Virol **75:**9357-9366.

156. **Mirambeau G, Lyonnais S, Gorelick RJ.** 2010. Features, processing states, and heterologous protein interactions in the modulation of the retroviral nucleocapsid protein function. RNA Biol **7:**724-734.

157. **Mirambeau G, Lyonnais S, Coulaud D, Hameau L, Lafosse S, Jeusset J, Justome A, Delain E, Gorelick RJ, Le Cam E.** 2006. Transmission electron microscopy reveals an optimal HIV-1 nucleocapsid aggregation with single-stranded nucleic acids and the mature HIV-1 nucleocapsid protein. J Mol Biol **364:**496-511.

158. **Cruceanu M, Urbaneja MA, Hixson CV, Johnson DG, Datta SA, Fivash MJ, Stephen AG, Fisher RJ, Gorelick RJ, Casas-Finet JR, Rein A, Rouzina I, Williams MC.** 2006. Nucleic acid binding and chaperone properties of HIV-1 gag and nucleocapsid proteins. Nucleic Acids Res **34:**593-605.

159. **Cruceanu M, Gorelick RJ, Musier-Forsyth K, Rouzina I, Williams MC.** 2006. Rapid kinetics of protein–nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. J Mol Biol **363:**867-877.

160. **Wu T, Datta SAK, Mitra M, Gorelick RJ, Rein A, Levin JG.** 2010. Fundamental differences between the nucleic acid chaperone activities of HIV-1 nucleocapsid protein and gag or gag-derived proteins: biological implications. Virology **405:**556-567.

161. **de Marco A, Heuser AM, Glass B, Krausslich HG, Muller B, Briggs JAG.** 2012. Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity. J Virol **86:**13708-13716.

162. **Rulli SJ, Jr., Muriaux D, Nagashima K, Mirro J, Oshima M, Baumann JG, Rein A.** 2006. Mutant murine leukemia virus Gag proteins lacking proline at the N-terminus of the capsid domain block infectivity in virions containing wild-type Gag. Virology **347:**364-371.

163. **Briggs JAG, Simon MN, Gross I, Kräusslich H-G, Fuller SD, Vogt VM, Johnson MC.** 2004. The stoichiometry of gag protein in HIV-1. Nature Struct Mol Biol **11:**672-675.

164. **Ganser-Pornillos BK, Yeager M, Sundquist WI.** 2008. The structural biology of HIV assembly. Curr Opin Struct Biol **18:**203-217.

165. **Li F, Goila-Gaur R, Salzwedel K, Kilgore NR, Reddick M, Matallana C, Castillo A, Zoumplis D, Martin DE, Orenstein JM, Allaway GP, Freed EO, Wild CT.** 2003. PA-457: A potent HIV inhibitor that disrupts core condensation by targeting a late step in gag processing. Proc Natl Acad Sci **100:**13555-13560.

166. **Keller PW, Adamson CS, Heymann JB, Freed EO, Steven AC.** 2011. HIV-1 maturation inhibitor Bevirimat stabilizes the immature gag lattice. J Virol **85:**1420-1428.

167. **Forshey BM, Shi J, Aiken C.** 2005. Structural requirements for recognition of the human immunodeficiency virus type 1 core during host restriction in owl monkey cells. J Virol **79:**869-875.

168. **Stremlau M, Owens CM, Perron MJ, Klessling M, Autissler P, Sodroski J.** 2004. The cytoplasmic body component TRIM5α restricts HIV-1 infection in old world monkeys. Nature **427:**848-853.

169. **Stremlau M, Perron M, Lee M, Li Y, Song B, Javanbakht H, Diaz-Griffero F, Anderson DJ, Sundquist WI, Sodroski J.** 2006. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5 restriction factor. Proc Natl Acad Sci **103:**5514-5519.

170. **Forshey BM, von Schwedler U, Sundquist WI, Aiken C.** 2002. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. J Virol **76:**5667-5677.

171. **Fassati A.** 2012. Multiple roles of the capsid protein in the early steps of HIV-1 infection. Virus Res **170:**15-24.

172. **Kafaie J, Dolatshahi M, Ajamian L, Song R, Mouland AJ, Rouiller I, Laughrea M.** 2009. Role of capsid sequence and immature nucleocapsid proteins p9 and p15 in human immunodeficiency virus type 1 genomic RNA dimerization. Virology **385:**233-244.

173. **Doyon L, Croteau G, Thibeault D, Poulin F, Pilote L, Lamarre D.** 1996. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. J Virol **70:**3763.

174. **Bally F, Martinez R, Peters S, Sudre P, Telenti A.** 2000. Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors. AIDS Res Hum Retroviruses **16:**1209-1213.

175. **Dam E, Quercia R, Glass B, Descamps D, Launay O, Duval X, Kräusslich HG, Hance AJ, Clavel F, Group AS.** 2009. Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. PLoS Pathog **5:**e1000345.

176. **Kolli M, Stawiski E, Chappey C, Schiffer CA.** 2009. Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. J Virol **83:**11027-11042.

177. **Krishnan L, Engelman A.** 2012. Retroviral integrase proteins and HIV-1 DNA integration. J Biol Chem **287:**40858-40866.

178. **Paine PL, Moore LC, Horowitz SB.** 1975. Nuclear envelop permeability. Nature **254:**109-114.

179. **Panté N, Kann M.** 2002. Nuclear pore complex is able to transport macromolecules with diameters of ~39 nm. Mol Biol Cell **13:**425-434.

180. **Miller MD, Farnet CM, Bushman FD.** 1997. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. J Virol **71:**5382-5390.

181. **de Rocquigny H, Petitjean P, Tanchou V, Decimo D, Drouot L, Delaunay T, Darlix JL, Roques BP.** 1997. The zinc fingers of HIV nucleocapsid protein NCp7 direct interacts with the viral regulatory protein vpr. J Biol Chem **272:**30753-30759.

182. **Zhang S, Pointer D, Singer G, Feng Y, Park K, Zhao LJ.** 1998. Direct binding to nucleic acids by vpr of human immunodeficiency virus type 1. Gene **212:**157-166.

183. **Yamashita M, Perez O, Hope TJ, Emerman M.** 2007. Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. PLoS Pathog **3:**e156.

184. **Qi M, Yang R, Aiken C.** 2008. Cyclophilin A-dependent restriction of human immunodeficiency virus type 1 capsid mutants for infection of nondividing cells. J Virol **82:**12001-12008.

185. **Yamashita M, Emerman M.** 2009. Cellular restriction targeting viral capsids perturbs human immunodeficiency virus type 1 infection of nondividing cells. J Virol **83:**9835-9843.

186. **Buckman JS, Bosche WJ, Gorelick RJ.** 2003. Human immunodeficiency virus type 1 nucleocapsid Zn2+ fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. J Virol **77:**1469-1480.

187. **Thomas JA, Gagliardi TD, Alvord WG, Lubomirski M, Bosche WJ, Gorelick RJ.** 2006. Human immunodeficiency virus type 1 nucleocapsid zinc-finger mutations cause defects in reverse transcription and integration. Virology **353:**41-51.

188. **Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M.** 1993. Association of integrase, matrix, and reverse transcriptase antigens of human

immunodeficiency virus type 1 with viral nucleic acids following acute infection. Proc Natl Acad Sci USA **90:**6125-6129.

189. **Karageorgos L, Li P, Burrell C.** 1993. Characterization of HIV replication complexes early after cell-to-cell infection. AIDS Res Hum Retroviruses **9:**817-823.

190. **Heinzinger NK, Bukrinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, Gendelman HE, Ratner L, Stevenson M, Emerman M.** 1994. The vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. Proc Natl Acad Sci **91:**7311-7315.

191. **Wu X, Liu H, Xiao H, Conway JA, Hehl E, Kalpana GV, Prasad V, Kappes JC.** 1999. Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through interactions with the nucleoprotein reverse transcription complex. J Virol **73:**2126-2135.

192. **Engelman A, Englund G, Orenstein JM, Martin MA, Craigie R.** 1995. Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. J Virol **69:**2729-2736.

193. **Shehu-Xhilaga M, Hill M, Marshall JA, Kappes J, Crowe SM, Mak J.** 2002. The conformation of the mature dimeric human immunodeficiency virus type 1 RNA genome requires packaging of pol protein. J Virol **76:**4331-4340.

194. **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 **109:**9587-9592.

195. **Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ.** 2008. Identification of host proteins required for HIV infection through a functional genomic screen. Science **319:**921-926.

196. **König R, Zhou Y, Elleder D, Diamond TL, Bonamy GMC, Irelan JT, Chiang C-y, Tu BP, De Jesus PD, Lilley CE, Seidel S, Opaluch AM, Caldwell JS, Weitzman MD, Kuhen KL, Bandyopadhyay S, Ideker T, Orth AP, Miraglia LJ, Bushman FD, Young JA, Chanda SK.** 2008. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell **135:**49-60.

197. **Yeung ML, Houzet L, Yedavalli VSRK, Jeang KT.** 2009. A genome-wide short hairpin RNA screening of Jurkat T-cells for human proteins contributing to productive HIV-1 replication. J Biol Chem **284:**19463-19473.

198. **Matreyek KA, Engelman A.** 2013. Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. Viruses **5:**2483-2511.

199. **Fassati A, Goff SP.** 2001. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. J Virol **75:**3626-3635.

200. **Yamashita M, Emerman M.** 2004. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. J Virol **78:**5670-5678.

201. **Arhel NJ, Souquere-Besse S, Munier S, Souque-P., Guadagnini S, Rutherfod S, Prévost MC, Allen TD, Charneau P.** 2007. HIV-1 DNA flap formation promotes uncoating of the pre-integration complex at the nuclear pore. EMBO J **26:**3025-3037.

202. **Matreyek KA, Engelman A.** 2011. The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection Is determined by the viral capsid. J Virol **85:**7818-7827.

203. **Schaller T, Ocwieja KE, Rasaiyaah J, Price AJ, Brady TL, Roth SL, Hué S, Fletcher AJ, Lee K, KewalRamani VN, Noursadeghi M, Jenner RG, James LC, Bushman FD, Towers GJ.** 2011. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. PLoS Pathog **7:**e1002439.

204. **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.

205. **Ohishi M, Nakano T, Sakuragi S, Shioda T, Sano K, Sakuragi Ji.** 2011. The relationship between HIV-1 genome RNA dimerization, virion maturation and infectivity. Nucleic Acids Res **39:**3404-3417.

206. **Barat C, Schatz O, Le Grice S, Darlix JL.** 1993. Analysis of the interactions of HIV1 replication primer tRNALys,3 with nucleocapsid protein and reverse transcriptase. J Mol Biol **231:**185-190.

207. **Mammano F, Petit C, Clavel F.** 1998. Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and *gag* coeveloution in protease inhibitor-treated patients J Virol **72:**7632-7637.

208. **Wolfenden R.** 1972. Analog approaches to the structure of the transition state in enzyme reactions. Acc Chem Res **5:**10-18.

209. **Altman MD, Ali A, Reddy KK, Nalam MNL, Anjum SG, Cao H, Chellappan S, Kairys V, Fernandes MX, Gilson MK, Schiffer CA, Rana TM, Tidor B.** 2008. HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. J Am Chem Soc **130:**6099-6113.

210. **Singh K, Marchand B, Rai DK, Sharma B, Michailidis E, Ryan EM, Matzek KB, Leslie MD, Hagedorn AN, Li Z, Norden PR, Hachiya A, Parniak MA, Xu HT, Wainberg MA, Sarafianos SG.** 2012. Biochemical mechanism of HIV-1 resistance to Rilpivirine. J Biol Chem **287:**38110-38123.

211. **Schuckmann MM, Marchand B, Hachiya A, Kodama EN, Kirby KA, Singh K, Sarafianos SG.** 2010. The N348I mutation at the connection subdomain of HIV-1 reverse transcriptase decreases binding to Nevirapine. J Biol Chem **285:**38700-38709.

212. **Braz VA, Holladay LA, Barkley MD.** 2010. Efavirenz binding to HIV-1 reverse transcriptase monomers and dimers. Biochemistry **49:**601-610.

213. **White KL, Margot NA, Ly JK, Chen JM, Ray AS, Pavelko M, Wang R, McDermott M, Swaminathan S, Miller MD.** 2005. A combination of decreased NRTI incorporation and decreased excision determines the resistance profile of HIV-1 K65R RT. AIDS **19:**1751-1760.

214. **Quashie PK, Mesplede T, Han YS, Veres T, Osman N, Hassounah S, Sloan RD, Xu HT, Wainberg MA.** 2013. Biochemical analysis of the role of G118R-linked Dolutegravir drug resistance substitutions in HIV-1 integrase. Antimicrob Agents Chemother **57:**6223-6235.

215. **Tozser J, Blaha I, Copeland TD, Wondrak EM, Oroszlan S.** 1991. Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in gag and gag-pol polyproteins. FEBS **281:**77-80.

216. **Radzicka A, Wolfenden R.** 1996. Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. J Am Chem Soc **118:**6105-6109.

217. **Darke PL, Nutt RF, Brady SF, Garsky VM, Ciccarone TM, Leu CT, Lumma PK, Freidinger RM, Veber DF, Sigal IS.** 1988. HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins. Biochem Biophys Res Commun **156:**297-303.

218. **Arribas JR, Clumeck N, Nelson M, Hill A, Delft Y, Moecklinghoff C.** 2012. The MONET trial: week 144 analysis of the efficacy of darunavir/ritonavir (DRV/r) monotherapy versus DRV/r plus two nucleoside reverse transcriptase inhibitors, for patients with viral load < 50 HIV-1 RNA copies/mL at baseline. HIV Med **13:**398-405.

219. **Potempa M, Nalivaika E, Ragland D, Lee SK, Schiffer CA, Swanstrom R.** 2015. A Direct Interaction with RNA Dramatically Enhances the Catalytic Activity of the HIV-1 Protease In Vitro. J Mol Biol **427:**2360-2378.

220. **Sheng N, Erickson-Viitanen S.** 1994. Cleavage of p15 protein in vitro by human immunodefiency virus type 1 protease is RNA dependent. J Virol **68:**6207-6214.

221. **Sheng N, Pettit SC, Tritch R, Ozturk DH, Rayner MM, Swanstrom R, Erickson-Viitanen S.** 1997. Determinants of the human immunodeficiency virus type 1 p15NC-RNA interaction that affect enhanced cleavage by the viral protease. J Virol **71:**5723-5732.

222. **Kol N, Shi Y, Tsvitov M, Barlam D, Shneck RZ, Kay MS, Rousso I.** 2007. A stiffness switch in human immunodeficiency virus. Biophys J **92:**1777-1783.

223. **Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS.** 1988. Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci USA, **85:**4686-4690.

224. **Krausslich HG.** 1991. Human immunodeficiency virus proteinase dimer as component of the viral polyprotein prevents particle assembly and viral infectivity. Proc Natl Acad Sci USA, **88:**3213-3217.

225. **Park J, Morrow CD.** 1991. Overexpression of the gag-pol precursor from human immunodeficiency virus type 1 proviral genomes results in efficient proteolytic processing in the absence of virion production. J Virol **65:**5111-5117.

226. **Agniswamy J, Sayer JM, Weber IT, Louis JM.** 2012. Terminal interface conformations modulate dimer stability prior to amino terminal autoprocessing of HIV-1 protease. Biochemistry **51:**1041-1050.

227. **Lindhofer H, von der Helm K, Nitschko H.** 1995. In vivo processing of Pr160$^{gag-pol}$ from human immunodeficiency virus type 1 (HIV) in acutely infected, cultured human T-lymphocytes. Virology **214:**624-627.

228. **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.

229. **Deshmukh L, Ghirlando R, Clore GM.** 2015. Conformation and dynamics of the Gag polyprotein of the human immunodeficiency virus 1 studied by NMR spectroscopy. Proc Natl Acad Sci USA, doi:10.1073/pnas.1501985112.

230. **Tritch R, Cheng YE, Yin FH, Erickson-Viitanen S.** 1991. Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 *gag* polyprotein. J Virol **65:**922-930.

231. **Dorfman T, Bukovsky A, Ohagen A, Hoglund S, Gottlinger HG.** 1994. Functional domains of the capsid protein of human immunodefiency virus type 1. J Virol **68:**8180-8187.

232. **Momany C, Kovari LC, Prongay AJ, Keller W, Gitti RK, Lee BM, Gorbalenya AE, Tong L, McClure J, Ehrlich LS, Summers MF, Carter C, Rossmann MG.** 1996. Crystal structure of dimeric HIV-1 capsid protein. Nat Struct Biol **3:**763-770.

233. **Porter DJT, Hanlon MH, Carter III LH, Danger DP, Furfine ES.** 2001. Effectors of HIV-1 protease peptidolytic activity. Biochemistry **40:**11131-11139.

234. **Alfadhli A, McNett H, Tsagli S, Bachinger HP, Peyton DH, Barklis E.** 2011. HIV-1 matrix protein binding to RNA. J Mol Biol **410:**653-666.

235. **Chukkapalli V, Oh SJ, Ono A.** 2010. Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. Proc Natl Acad Sci USA, **107:**1600-1605.

236. **Lochrie MA, Waugh S, Pratt Jr. DG, Clever J, Parslow TG, Polisky B.** 1997. In vitro selection of RNAs that bind to the human immunodeficiency virus type-1 gag polyprotein. Nucleic Acids Res **25:**2902-2910.

237. **Ott DE, Coren LV, Gagliardi TD.** 2005. Redundant roles for nucleocapsid and matrix RNA-binding sequences in human immunodeficiency virus type 1 assembly. J Virol **79:**13839-13847.

238. **Purohit P, Dupont S, Stevenson M, Green MR.** 2001. Sequence-specific interaction between HIV-1 matrix protein and viral genomic RNA revealed by in vitro genetic selection. RNA **7:**576-584.

239. **Lyonnais S.** 2003. G-quartets direct assembly of HIV-1 nucleocapsid protein along single-stranded DNA. Nucleic Acids Res **31:**5754-5763.

240. **Gueron M, Leroy J.** 2000. The i-motif in nucleic acids. Curr Opin Struct Biol **10:**326-331.

241. **Deutsch C, Taylor JS, Wilson DF.** 1982. Regulation of intracellular pH by human peripheral blood lymphocytes as measured by ^{19}F NMR. Proc Natl Acad Sci USA, **79:**7944-7948.

242. **Darke PL, Jordan SP, Hall DL, Zugay JA, Shafer JA, Kuo LC.** 1994. Dissociation and association of the HIV-1 protease dimer subunits: equilibria and rates. Biochemistry **33:**98-105.

243. **Copeland RA.** 2000. Cooperativity in enzyme catalysis, p 367-384. *In* Copeland RA (ed), Enzymes: a practical introduction to structure, mechanism, and data analysis, Second ed. Wiley-VCH, New York, NY, USA.

244. **Blainey PC, Graziano V, Perez-Berna AJ, McGrath WJ, Flint SJ, San Martin C, Xie XS, Mangel WF.** 2013. Regulation of a viral proteinase by a peptide and DNA in one-dimensional space: IV. viral proteinase slides along DNA to locate and process its substrates. J Biol Chem **288:**2092-2102.

245. **Mangel WF, McGrath WJ, Toledo DL, Anderson CW.** 1993. Viral DNA and a viral peptide can act as cofactors of adenovirus virion proteinase activity. Nature **361:**274-275.

246. **McGrath WJ, Baniecki ML, Li C, McWhirter SM, Brown MT, Toledo DL, Mangel WF.** 2001. Human adenovirus proteinase: DNA binding and stimulation of proteinase activity by DNA. Biochemistry **40:**13237-13245.

247. **Hartl MJ, Bodem J, Jochheim F, Rethwilm A, Rosch P, Wohrl BM.** 2011. Regulation of foamy virus protease activity by viral RNA: a novel and unique mechanism among retroviruses. J Virol **85:**4462-4469.

248. **Beran RK, Serebrov V, Pyle AM.** 2007. The serine protease domain of hepatitis C viral NS3 activates RNA helicase activity by promoting the binding of RNA substrate. J Biol Chem **282:**34913-34920.

249. **Ray U, Das S.** 2011. Interplay between NS3 protease and human La protein regulates translation-replication switch of Hepatitis C virus. Sci Rep **1:**1.

250. **Chakraborty S, Sharma S, Maiti PK, Krishnan Y.** 2009. The poly dA helix: a new structural motif for high performance DNA-based molecular switches. Nucleic Acids Res **37:**2810-2817.

251. **Rich A, Davies DR, Crick FHC, Watson JD.** 1961. The molecular structure of polyadenylic acid. J Mol Biol **3:**71-86.

252. **Zimmerman SB, Davies DR, Navia MA.** 1977. An ordered single-stranded structure for polyadenylic acid in denaturing solvents. An X-ray fiber diffraction and model building study. J Mol Biol **116:**317-330.

253. **Drummond JE, Mounts P, Gorelick RJ, Casas-Finet JR, Bosche WJ, Henderson LE, Waters DJ, Arthur LO.** 1997. Wild-type and mutant HIV type 1 nucelocapsid proteins increase the proportion of long cDNA transcripts by viral reverse transcriptase. AIDS RES HUM RETROV **13:**533-543.

254. **Khan R, Giedroc DP.** 1994. Nucleic acid binding properties of recombinant Zn2 HIV-1 nucleocapsid protein are modulated by COOH-terminal processing. J Biol Chem **269:**22538-22546.

255. **Chen Z, Li Y, Chen E, Hall DL, Darke PL, Culberson C, Shafer JA, Kuo LC.** 1994. Crystal structure at 1.9-Å resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. J Biol Chem **269:**26344-26348.

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

257. **Kovalevsky AY, Louis JM, Aniana A, Ghosh AK, Weber IT.** 2008. Structural evidence for effectiveness of darunavir and two related antiviral inhibitors against HIV-2 protease. J Mol Biol **384:**178-192.

258. **Louis JM, Ishima R, Aniana A, Sayer JM.** 2009. Revealing the dimer dissociation and existence of a folded monomer of the mature HIV-2 protease. Protein Sci **18:**2442-2453.

259. **Pichuantes S, Babe LM, Barr PJ, DeCamp DL, Craik CS.** 1990. Recombinant HIV2 protease processes HIV1 Pr53^{gag} and analogous junction peptides in vitro. J Biol Chem **265:**13890-13898.

260. **Motorin Y, Helm M.** 2010. tRNA stabilization by modified nucleotides. Biochemistry **49:**4934-4944.

261. **Alderfer JL, Smith SL.** 1971. A proton magnetic resonance study of polydeoxyriboadenylic acid. J Am Chem Soc **93:**7305-7314.

262. **Ke C, Humeniuk M, S-Gracz H, Marszalek PE.** 2007. Direct Measurements of Base Stacking Interactions in DNA by Single-Molecule Atomic-Force Spectroscopy. Phys Rev Lett **99**.

263. **Cheng YE, Yin FH, Foundling S, Blomstrom D, Kettner CA.** 1990. Stability and activity of human immunodeficiency virus protease: comparison of the natural dimer with a homologous, single-chain tethered dimer. Proc Natl Acad Sci USA, **87:**9660-9664.

264. **Judd DA, Nettles JH, Nevins N, Snyder JP, Liotta DC, Tang J, Ermolieff J, Schinazi RF, Hill CL.** 2001. Polyoxometalate HIV-1 protease inhibitors. A new mode of protease inhibition. J Am Chem Soc **123:**886-897.

265. **Sperka T, Pitlik J, Bagossi P, Tozser J.** 2005. Beta-lactam compounds as apparently uncompetitive inhibitors of HIV-1 protease. Bioorg Med Chem Lett **15:**3086-3090.

266. **Ung PM, Dunbar JB, Jr., Gestwicki JE, Carlson HA.** 2014. An allosteric modulator of HIV-1 protease shows equipotent inhibition of wild-type and drug-resistant proteases. J Med Chem **57:**6468-6478.

267. **Kovalevsky AY, Liu F, Leshchenko S, Ghosh AK, Louis JM, Harrison RW, Weber IT.** 2006. Ultra-high resolution crystal structure of HIV-1 protease mutant reveals two binding sites for clinical inhibitor TMC114. J Mol Biol **363:**161-173.

268. **Baca M, Kent SBH.** 1993. Catalytic contribution of flap-substrate hydrogen bonds in "HIV-1 protease" explored by chemical synthesis. Proc Natl Acad Sci USA, **90:**11638-11642.

269. **Torbeev VY, Raghuraman H, Hamelberg D, Tonelli M, Westler WM, Perozo E, Kent SB.** 2011. Protein conformational dynamics in the mechanism of HIV-1 protease catalysis. Proc Natl Acad Sci USA, **108:**20982-20987.

270. **Freedberg DI, Ishima R, Jacob J, Wang YX, Kustanovich I, Louis JM, Torchia DA.** 2002. Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations. Protein Sci **11:**221-232.

271. **Karthik S, Senapati S.** 2011. Dynamic flaps in HIV-1 protease adopt unique ordering at different stages in the catalytic cycle. Proteins **79:**1830-1840.

272. **Scott WRP, Schiffer CA.** 2000. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. Structure **8:**1259-1265.

273. **Torbeev VY, Raghuraman H, Mandal K, Senapati S, Perozo E, Kent SBH.** 2009. Dynamics of "flap" structures in three HIV-1 protease/inhibitor complexes probed by total chemical synthesis and pulse-EPR spectroscopy. J Am Chem Soc **131**.

274. **Prabu-Jeyabalan M, Nalivaika E, Schiffer CA.** 2000. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. J Mol Biol **301:**1207-1220.

275. **Alvizo O, Mittal S, Mayo SL, Schiffer CA.** 2012. Structural, kinetic, and thermodynamic studies of specificity designed HIV-1 protease. Protein Sci **21:**1029-1041.

276. **Bhat TN, Baldwin ET, Liu B, Cheng YE, Erickson JW.** 1994. Crystal structure of a tethered dimer of HIV-1 proteinase complexed with an inhibitor. Nat Struct Biol **1:**552-556.

277. **Mattei S, Anders M, Konvalinka J, Krausslich HG, Briggs JA, Muller B.** 2014. Induced maturation of human immunodeficiency virus. J Virol **88:**13722-13731.

278. **Zhou J, Yuan X, Dismuke D, Forshey BM, Lundquist C, Lee KH, Aiken C, Chen CH.** 2003. Small-Molecule Inhibition of Human Immunodeficiency Virus Type 1 Replication by Specific Targeting of the Final Step of Virion Maturation. Journal of Virology **78:**922-929.

279. **Ali A, Bandaranayake RM, Cai Y, King NM, Kolli M, Mittal S, Murzycki JF, Nalam MN, Nalivaika EA, Ozen A, Prabu-Jeyabalan MM, Thayer K, Schiffer CA.** 2010. Molecular Basis for Drug Resistance in HIV-1 Protease. Viruses **2:**2509-2535.

280. **Schilling O, Overall CM.** 2008. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat Biotechnol **26:**685-694.

281. **Chaudhury S, Gray JJ.** 2009. Identification of structural mechanisms of HIV-1 protease specificity using computational peptide docking: implications for drug resistance. Structure **17:**1636-1648.

282. **Ozer N, Haliloglu T, Schiffer CA.** 2006. Substrate specificity in HIV-1 protease by a biased sequence search method. Proteins **64:**444-456.

283. **Bagossi P, Cheng YE, Oroszlan S, Tozser J.** 1998. Comparison of the specificity of homo- and heterodimeric linked HIV-1 and HIV-2 proteinase dimers. Protein Eng **11:**439-445.

284. **Bagossi P, Sperka T, Feher A, Kadas J, Zahuczky G, Miklossy G, Boross P, Tozser J.** 2005. Amino acid preferences for a critical substrate binding subsite of retroviral proteases in type 1 cleavage sites. J Virol **79:**4213-4218.

285. **Billich A, Winkler G.** 1991. Analysis of subsite preferences of HIV-1 proteinase using MA/CA junction peptides substituted at the P3-P1' positions. Arch Biochem Biophys **290:**186-190.

286. **Dunn BM, Gustchina A, Wlodawer A, Kay J.** 1994. Subsite preferences of retroviral proteinases. Meth Enzymol **241:**254-278.

287. **Eizert H, Bander P, Bagossi P, Sperka T, Miklossy G, Boross P, Weber IT, Tozser J.** 2008. Amino acid preferences of retroviral proteases for amino-terminal positions in a type 1 cleavage site. J Virol **82:**10111-10117.

288. **Griffiths JT, Phylip LH, Konvalinka J, Strop P, Gustchina A, Wlodawer A, Davenport RJ, Briggs R, Dunn BM, Kay J.** 1992. Different requirements for productive interaction between the active site of HIV-1 proteinase and substrates containing -hydrophobic*hydrophobic- or -aromatic*pro- cleavage sites. Biochemistry **31:**5193-5200.

289. **Konvalinka J, Strop P, Velek J, Cerna V, Kostka V, Phylip LH, Richards AD, Dunn BM, Kay J.** 1990. Sub-site preferences of the aspartic proteinase from the human immunodeficiency virus, HIV-1. FEBS Lett **268:**35-38.

290. **Margolin N, Heath W, Osborne E, Lai M, Vlahos C.** 1991. Substitutions at the P2' site of gag p17-p24 affect cleavage efficiency by HIV-1 protease. Biochem Biophys Res Commun **167:**554-560.

291. **Richards AD, Phylip LH, Farmerie WG, Scarborough PE, Alvarez a, Dunn BM, Hirel PH, Konvalinka J, Strop P, Pavlickova L.** 1990. Sensitive, soluble chromogenic substrates for HIV-1 proteinase. J Biol Chem **265:**7733-7736.

292. **Ridky TW, Cameron CE, Cameron J, Leis J, Copeland TD, Wlodawer A, Weber IT, Harrison RW.** 1996. Human immunodeficiency virus, type 1 protease substrate specificity is limited by interactions between substrate amino acids bound in adjacent enzyme subsites. J Biol Chem **271:**4709-4717.

293. **Ridky TW, Kikonyogo A, Leis J, Gulnik S, Copeland T, Erickson J, Wlodawer A, Kurinov I, Harrison RW, Weber IT.** 1998. Drug-resistant HIV-1 proteases identify enzyme residues important for substrate selection and catalytic rate. Biochemistry **37:**13835-13845.

294. **Szeltner Z, Polgar L.** 1996. Rate-determining steps in HIV-1 protease catalysis. J Biol Chem **271:**32180-32184.

295. **Tozser J, Bagossi P, Weber IT, Louis JM, Copeland TD, Oroszlan S.** 1997. Studies on the symmetry and sequence context dependence of the HIV-1 proteinase specificity. J Biol Chem **272:**16807-16814.

296. **Tözser J, Gustchina A, Weber IT, Blaha I, Wondrak EM, Oroszlan S.** 1991. Studies on the role of the S4 substrate binding site of HIV proteinases. FEBS Lett **279:**356-360.

297. **Tozser J, Weber IT, Gustchina A, Blaha I, Copeland TD, Louis JM, Oroszlan S.** 1992. Kinetic and modeling studies of S3-S3' subsites of HIV proteinases. Biochemistry **31:**4793-4800.

298. **Urban J, Konvalinka J, Stehlikova J, Gregorova E, Majer P, Soucek M, Andreansky M, Fabry M, Strop P.** 1992. Reduced-bond tight-binding inhibitors of HIV-1 protease. FEBS **298:**9-13.

299. **Gök M, Özcerit AT.** 2012. A new feature encoding scheme for HIV-1 protease cleavage site prediction. Neural Computing and Applications **22:**1757-1761.

300. **Jaeger S, Chen S.** 2010. Information fusion for biological prediction. J Data Sci **8:**269-288.

301. **Kim G, Kim Y, Lim H, Kim H.** 2010. An MLP-based feature subset selection for HIV-1 protease cleavage site analysis. Artif Intell Med **48:**83-89.

302. **Kontijevskis A, Wikberg JE, Komorowski J.** 2007. Computational proteomics analysis of HIV-1 protease interactome. Proteins **68:**305-312.

303. **Li X, Hu H, Shu L.** 2010. Predicting human immunodeficiency virus protease cleavage sites in nonlinear projection space. Mol Cell Biochem **339:**127-133.

304. **Nanni L, Lumini A.** 2009. Using ensemble of classifiers for predicting HIV protease cleavage sites in proteins. Amino Acids **36:**409-416.

305. **Newell NE.** 2011. Cascade detection for the extraction of localized sequence features; specificity results for HIV-1 protease and structure-function results for the Schellman loop. Bioinformatics **27:**3415-3422.

306. **Ogul H.** 2009. Variable context Markov chains for HIV protease cleavage site prediction. Biosystems **96:**246-250.

307. **Ozturk O, Aksac A, Elsheikh A, Ozyer T, Alhajj R.** 2013. A consistency-based feature selection method allied with linear SVMs for HIV-1 protease cleavage site prediction. PLoS One **8:**e63145.

308. **Rognvaldsson T, You L.** 2004. Why neural networks should not be used for HIV-1 protease cleavage site prediction. Bioinformatics **20:**1702-1709.

309. **Rognvaldsson T, You L, Garwicz D.** 2007. Bioinformatic approaches for modeling the substrate specificity of HIV-1 protease: an overview. Expert Rev Mol Diagn **7:**435-451.

310. **Rognvaldsson T, Etchells TA, You L, Garwicz D, Jarman I, Lisboa PJ.** 2009. How to find simple and accurate rules for viral protease cleavage specificities. BMC Bioinformatics **10:**149.

311. **Rognvaldsson T, You L, Garwicz D.** 2015. State of the art prediction of HIV-1 protease cleavage sites. Bioinformatics **31:**1204-1210.

312. **Song J, Tan H, Perry AJ, Akutsu T, Webb GI, Whisstock JC, Pike RN.** 2012. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. PLoS One **7:**e50300.

313. **You L, Garwicz D, Rognvaldsson T.** 2005. Comprehensive bioinformatic analysis of the specificity of human immunodeficiency virus type 1 protease. J Virol **79:**12477-12486.

314. **Tozser J, Zahuczky G, Bagossi P, Louis JM, Copeland TD, Oroszlan S, Harrison RW, Weber IT.** 2000. Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. Eur J Biochem **267:**6287-6295.

315. **Tang C, Ndassa Y, Summers MF.** 2002. Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. Nat Struct Biol **9:**537-543.

316. **Accola MA, Hoglund S, Gottlinger HG.** 1998. A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 gag precursor is crucial for viral particle assembly. J Virol **72:**2072-2078.

317. **Bharat TA, Castillo Menendez LR, Hagen WJ, Lux V, Igonet S, Schorb M, Schur FK, Krausslich HG, Briggs JA.** 2014. Cryo-electron microscopy of tubular arrays of

HIV-1 Gag resolves structures essential for immature virus assembly. Proc Natl Acad Sci U S A **111:**8233-8238.

318. **Briggs JA, Riches JD, Glass B, Bartonova V, Zanetti G, Krausslich HG.** 2009. Structure and assembly of immature HIV. Proc Natl Acad Sci U S A **106:**11090-11095.

319. **Schur FK, Hagen WJ, Rumlova M, Ruml T, Muller B, Krausslich HG, Briggs JA.** 2015. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 A resolution. Nature **517:**505-508.

320. **Woodward CL, Cheng SN, Jensen GJ.** 2015. Electron cryotomography studies of maturing HIV-1 particles reveal the assembly pathway of the viral core. J Virol **89:**1267-1277.

321. **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.

322. **Pace CN, Scholtz JM.** 1998. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J **75:**422-427.

323. **Mittal S, Cai Y, Nalam MN, Bolon DN, Schiffer CA.** 2012. Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. J Am Chem Soc **134:**4163-4168.

324. **Lee S, Joshi A, Nagashima K, Freed EO, Hurley JH.** 2007. Structural basis for viral late-domain binding to Alix. Nat Struct Mol Biol **14:**194-199.

325. **Liu H, Wu X, Xiao H, Conway JA, Kappes J.** 1997. Incorporation of functional human immunodeficiency virus type 1 integrase into virions independent of the Gag-Pol precursor protein. J Virol **71:**7704-7710.

326. **Lapkouski M, Tian L, Miller JT, Le Grice SF, Yang W.** 2013. Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. Nat Struct Mol Biol **20:**230-236.

327. **Team RC.** 2015. R: A language and environment for statistical computing., R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

328. **Wimley WC, White SH.** 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol **3:**842-848.

329. **Wimley WC, Creamer TP, White KL.** 1996. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry **35:**5109-5124.

330. **Kassimi NE-B, Thakkar AJ.** 2009. A simple additive model for polarizabilities: Application to amino acids. Chemical Physics Letters **472:**232-236.

331. **Tsai J, Taylor R, Chothia C, Gerstein M.** 1999. The packing density in proteins: standard radii and volumes. J Mol Biol **290:**253-266.

332. **Bondi A.** 1964. van der Waals volumes and radii. J Phys Chem **68:**441-451.

333. **Bates D, Macechler M, Bolker B, Walker S.** 2015. Fitting linear mixed-effects models using lme4. J Stat Software.

334. **Lumley TuFcbAM.** 2009. leaps: regression subset selection, vR package version 2.9. http://CRAN.R-project.org/package=leaps.

335. **Friedman J, Hastie T, Tibshirani R.** 2010. Regularization paths for generalized linear models via coordinate descent. J Stat Software **33:**1-22.

336. **Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, COoper T, Z. M, Kenkel B, Team RC, Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y, Candan C.** 2015. caret: classification and regression training, vR package version 6.0-52. http://CRAN.R-project.org/package=caret.

337. **Mevik BH, Wehrens R, Liland KH.** 2013. pls: partial least squares and principal component regression, vR package version 2.4-3. http://CRAN.R-project.org/package=pls.

338. **Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M.** 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics **12:**77.

339. **Loeb DD, Swanstrom R, Everitt L, Manchester M, Stamper SE, Hutchinson III CA.** 1989. Complete mutagenesis of the HIV-1 protease. Nature **340:**397-400.

340. **Boyer PL, Sarafianos SG, Arnold E, Hughes SH.** 2001. Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase. J Virol **75:**4832-4842.

341. **Meyer PR, Matsuura SE, Mian AM, So AG, Scott WA.** 1999. A mechanism of AZT resistance: an increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase. Mol Cell **4:**35-43.