2009-11-13

Co-evolution of HIV-1 Protease and its Substrates: A Dissertation

Madhavi Kolli
University of Massachusetts Medical School Worcester

CO-EVOLUTION OF HIV-1 PROTEASE AND ITS SUBSTRATES

A Dissertation Presented

By

Madhavi Kolli

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

November 13th, 2009

Biochemistry and Molecular Pharmacology

CO-EVOLUTION OF HIV-1 PROTEASE AND ITS SUBSTRATES

A Dissertation Presented

By

MADHAVI KOLLI

The signatures of the Dissertation Defense Committee signifies completion and approval as to style and content of the Dissertation

Celia A. Schiffer, Ph.D., Thesis Advisor

Larry Stern, Ph.D., Member of Committee

Mary Munson, Ph.D., Member of Committee

Mohan Somasundaram, Ph.D., Member of Committee

Francois Clavel, M.D., Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

Phillip Zamore, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school

Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Program in Biochemistry and Molecular Pharmacology

November 13, 2009

ACKNOWLEDGEMENTS

I would like to thank several people for their support, advice, and friendship. First, I would like to thank my graduate advisor, Celia Schiffer, for giving me the opportunity to work on a project that interested me, and also for her support and guidance throughout my graduate studies. I have learned a lot from her the past few years, especially to ask the right questions and to be confident as a scientist. Not only is she a good mentor but is also a wonderful person and I would also like to thank her for her friendship over the past years. She has also brought together a wonderful set of people that make the lab a great environment to do science. I would like to thank all of my lab mates for the lively discussions and their support. I am especially grateful to Ellen Nalivaika, not only for making sure everything in the lab is running smoothly and but also for helping me with an occasional PCR that always works when she does it!! I would like to thank Moses and Madhavi for patiently teaching me x-ray crystallography. I would like to express my gratitude to my bay mates – Rajintha, Keith and Seema for their advice, lively discussions – scientific or otherwise, for sharing your reagents, for the laughs we shared and their friendship. I really enjoyed coming in to lab everyday and I will miss you all. I would like to thank Aysegul for all the fun times, silly laughs and for being such a wonderful friend. I first met her when she visited the Schiffer lab as a summer student, and since then our friendship has grown. Now she's a graduate student here and I truly enjoy the discussions over lunch/ coffee breaks, movie nights and wine with her. I will miss you greatly.

I really enjoyed being part of the Biochemistry department. First, I would like to thank my committee members: Dr. Phillip Zamore, Dr. Mary Munson, Dr. Mohan Somasundaran, Dr.Lawrence Stern and Dr.Francois Clavel for their insightful comments and keeping me focused throughout my graduate career. I would like to thank Dr.C.Robert Matthews for making this department a great place for learning, especially for the graduate students. My gratitude goes out to Tiffanie Covello and Christine Pruitte for all of the support and help, without which life in the lab would not have been so easy. I would like to thank my classmates in the department – Nick, Chaitali, Melonie and Mary, for their advice and support.

To my family – you have always being supportive of me no matter what. I am truly blessed to be a part of such a wonderful family. Thank you Amma and Nanna, for being the best parents and for giving me a life filled with wonderful memories. For teaching me to stand up for my convictions even if you didn't agree with me, to be responsible for my actions and most importantly to be a good person. Thank you Nilu for being such a wonderful sister, especially the past few years when we have come to be such good friends. I would like to thank my in-laws, for welcoming me into their family and for being so supportive of everything I do. I am very fortunate and happy to be a part of your family.

And, last but not least, to my husband Sagar, for always believing in me, for being my best friend and for sticking by me. For all of your support and advice, the scientific discussions through grad school, all the fun and memorable times we have shared through

the years, for all of the times you have made me laugh and for all the little things (and big) you do. With you by my side, life is never dull, and I look forward our future together. Thank you for everything – words will never be enough.

ABSTRACT

Drug resistance is the most important factor that influences the successful treatment of individuals infected with the human immunodeficiency virus type 1 (HIV-1), the causative organism of the acquired immunodeficiency syndrome (AIDS). Tremendous advances in our understanding of HIV and AIDS have led to the development of Highly Active Antiretroviral Therapy (HAART), a combination of drugs that includes HIV-1 reverse transcriptase, protease, and more recently, integrase and entry inhibitors, to combat the virus. Though HAART has been successful in reducing AIDS-related morbidity and mortality, HIV rapidly evolves resistance leading to therapy failure. Thus, a better understanding of the mechanisms of resistance will lead to improved drugs and treatment regimens.

Protease inhibitors (PIs) play an important role in anti-retroviral therapy. The development of resistance mutations within the active site of the protease greatly reduces its affinity for the protease inhibitors. Frequently, these mutations reduce catalytic efficiency of the protease leading to an overall reduction in viral fitness. In order to overcome this loss in fitness the virus evolves compensatory mutations within the protease cleavage sites that allow the protease to continue to recognize and cleave its substrates while lowering affinity for the PIs. Improved knowledge of this substrate co-evolution would help better understand how HIV-1 evolves resistance and thus, lead to improved therapeutic strategies. Sequence analyses and structural studies were performed to investigate co-evolution of HIV-1 protease and its cleavage sites. Though a

few studies reported the co-evolution within Gag, including the protease cleavage sites, a more extensive study was lacking, especially as drug resistance was becoming increasingly severe.

In Chapter II, a small set of viral sequences from infected individuals were analyzed for mutations within the Gag cleavage sites that co-occurred with primary drug resistance mutations within the protease. These studies revealed that mutations within the p1p6 cleavage site coevolved with the nelfinavir-resistant protease mutations. As a result of increasing number of infected individuals being treated with PIs leading to the accumulation of PI resistant protease mutations, and with increasing efforts at genotypic and phenotypic resistance testing, access to a larger database of resistance information has been made possible. Thus in Chapter III, over 39,000 sequences were analyzed for mutations within NC-p1, p1-6, Autoproteolysis, and PR-RT cleavage sites and several instances of substrate co-evolution were identified. Mutations in both the NC-p1 and the p1-p6 cleavage sites were associated with at least one, if not more, primary resistance mutations in the protease.

Previous studies have demonstrated that mutations within the Gag cleavage sites enhance viral fitness and/or resistance when they occur in combination with primary drug resistance mutations within the protease. In Chapter III viral fitness in the presence and absence of cleavage site mutations in combination with primary drug resistant protease mutations was analyzed to investigate the impact of the observed co-evolution. These studies showed no significant changes in viral fitness. Additionally in Chapter III, the

impact of these correlating mutations on phenotypic susceptibilities to various PIs was also analyzed. Phenotypic susceptibilities to various PIs were altered significantly when cleavage site mutations occurred in combination with primary protease mutations.

In order to probe the underlying mechanisms for substrate co-evolution, in Chapter IV, X-ray crystallographic studies were performed to investigate structural changes in complexes of WT and D30N/N88D protease variants and the p1p6 peptide variants. Peptide variants corresponding to p1p6 cleavage site were designed, and included mutations observed in combination with the D30N/N88D protease mutation. Structural analyses of these complexes revealed several correlating changes in van der Waals contacts and hydrogen bonding as a result of the mutations. These changes in interactions suggest a mechanism for improving viral fitness as a result of co-evolution.

This thesis research successfully identified several instance of co-evolution between primary drug resistant mutations in the protease and mutations within NC-p1 and p1p6 cleavage sites. Additionally, phenotypic susceptibilities to various PIs were significantly altered as a result of these correlated mutations. The structural studies also provided insights into the mechanism underlying substrate co-evolution. These data advance our understanding of substrate co-evolution and drug resistance, and will facilitate future studies to improve therapeutic strategies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF THIRD PARTY COPYRIGHTED MATERIAL

Figure I.1 – Nature Publishing Group. License Number - 2330341154043

Figure I.6 – Elsevier. License Number - 2330360707370

Figure I.7 – Elsevier. License Number - 2330361005866

Figure A I.1 – American Society for Microbiology. License Number - 2376090778329

PREFACE

Chapter II has been previously published as:

Kolli, M., Lastere, S., and Schiffer, C.A. Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 cleavage site. J. Virology (2005) 347: 405-409. © Elsevier Inc.

Chapter III has been previously published as:

Kolli, M., Stawiski, E., Chappey, C., and Schiffer, C.A. Human Immunodeficiency Virus Type I Protease-Correlated Cleavage Site Mutations Enhance Inhibitor Resistance. J. Virology (2009) 83(21): 11027-11042. © Elsevier Inc.

CHAPTER I

INTRODUCTION

Drug resistance has become a persistent problem in treating highly communicable diseases such as tuberculosis and acquired immunodeficiency syndrome (AIDS) (Dorman and Chaisson, 2007; Fischbach and Walsh, 2009; Shafer and Schapiro, 2005). Since the discovery of the human immunodeficiency virus type 1 (HIV-1) as the causative agent of AIDS (Chermann et al., 1983), research has led to the development of antiretroviral therapies that have significantly reduced HIV-1-related morbidity and mortality (Chermann et al., 1983; Palella et al., 1998). HIV-1, however, evolves rapidly due to lack of proof-reading mechanism by the reverse transcriptase, rapid replication rate and the selective pressures of antiretroviral therapy, leading to emergence and transmission of drug-resistant viruses resulting in therapy failure (Shafer and Schapiro, 2005).

Current antiretroviral strategies for treating HIV-1 infections include drugs that inhibit the viral enzymes reverse transcriptase, protease, and more recently, integrase. Due to the pivotal role of HIV-1 protease in viral maturation, protease inhibitors are one of the most important components of antiretroviral therapy (Kohl et al., 1988; Kramer et al., 1986). However, the emergence of resistance to protease inhibitors has significantly reduced their efficacy (Shafer and Schapiro, 2005). HIV-1 evolves resistance to protease inhibitors by the selection of mutations not only in the protease (Mammano et al., 2000; Martinez-Picado et al., 1999) but elsewhere in the genome, including the protease substrate sites within Gag and GagPol (Doyon et al., 1996; Zhang et al., 1997).

At the commencement of this study, protease cleavage site mutations had been observed in a few studies as a compensatory mechanism for the deleterious effects of mutations

within the protease (Mammano et al., 2000; Martinez-Picado et al., 1999). Although associations between specific mutations in the protease and its cleavage sites had been reported, they were still fairly uncommon. Therefore, one main goal of my thesis research was to investigate correlations between protease mutations that are the primary cause of resistance to protease inhibitors, and mutations within the substrate-cleavage sites. Another goal was to analyze the impact of these associated mutations on replicative capacity susceptibility to protease inhibitors.

Although biochemical and virological aspects of co-evolving protease-substrate mutations had been studied previously (Maguire et al., 2002; Zennou et al., 1998), the structural basis for this coevolution was not well understood. Therefore, another goal of my research was to investigate conformational changes within both the protease and substrate as a result of the associated mutations using the example of the p1-p6 cleavage site, which is known to co-evolve with the D30N/N88D protease mutations that are a signature of NFV resistance (Kolli, Lastere, and Schiffer, 2006).

In this chapter, I will introduce the epidemiology, treatment and socio-economic impact of HIV-1/AIDS. I will then briefly describe the HIV-1 life cycle and go on to focus on the HIV-1 protease. I will present in detail the structure and function of the protease, mechanism of substrate recognition, and the development of resistance to protease inhibitors. Finally, I will discuss current knowledge on the co-evolution of mutations in HIV-1 protease and its substrates as well as the contribution of this co-evolution to viral fitness and resistance to protease inhibitors.

AIDS

Acquired immunodeficiency syndrome is caused by infection with HIV-1, which primarily infects and kills CD4 T cells, macrophages and dendritic cells. The declining number of CD4 T cells and consequent loss in cell-mediated immunity leaves infected individuals highly susceptible to opportunistic infections, resulting in full-blown AIDS, and eventually death. Infections that commonly affect HIV-infected individuals include tuberculosis, pneumonia and toxoplasmosis. Infected individuals are also prone to several malignancies as a result of co-infections with oncogenic viruses like Epstein-Barr virus, Kaposi's sarcoma-associated herpes virus and human papilloma virus. HIV-1 is transmitted through direct contact of a mucous membrane or blood with a body fluid containing the virus.

Currently, most HIV-1 infected individuals in resource-rich countries receive highly active antiretroviral therapy (HAART) that has successfully slowed the progression to AIDS and improved patients' quality of life. HAART consists of a combination of at least 3 drugs from at least 2 classes: reverse transcriptase and protease inhibitors. These drugs have been successful to an extent, but drug-resistant strains of HIV-1 have emerged due to a combination of factors: the biology of HIV, poor adherence, and anatomical or cellular reservoirs that are inaccessible to the drugs. Resistance has considerably diminished susceptibility of HIV-1 to various drugs and as a result, driving research efforts in discovering new viral and cellular targets. Of equal importance in improving treatment of AIDS are studies on the mechanism of action of various drugs, HIV-1

biology and other factors that contribute to and help improve understanding of drug resistance.

AIDS was first reported in the US in 1981 by the Centers for Disease Control. Since then more than 25 million individuals have died from AIDS. In 2007, the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the World Health Organization (WHO), estimated that 33 million people are now infected worldwide. That year alone, AIDS-related complications caused 2 million deaths and about 2.7 million people were newly infected with HIV-1. Despite better public education, improved understanding of HIV-1 and AIDS, and access to newer treatments, the disease has reached pandemic proportions in sub-Saharan Africa, India, and other regions of South and South-East Asia. The emergence of highly resistant HIV-1 strains has made it harder to keep AIDS in check.

HIV/AIDS has wide-ranging impacts on every aspect of society, especially in the resource-poor countries of Africa and Asia. Families of affected individuals are burdened by the staggering medical costs, loss of income, and social stigma. Agriculture and business are often severely affected due to the loss of productive workers. Healthcare systems do not have the capacity to take on the additional demands of infected individuals battling AIDS-related complications. All these factors directly reduce national economic growth and development.

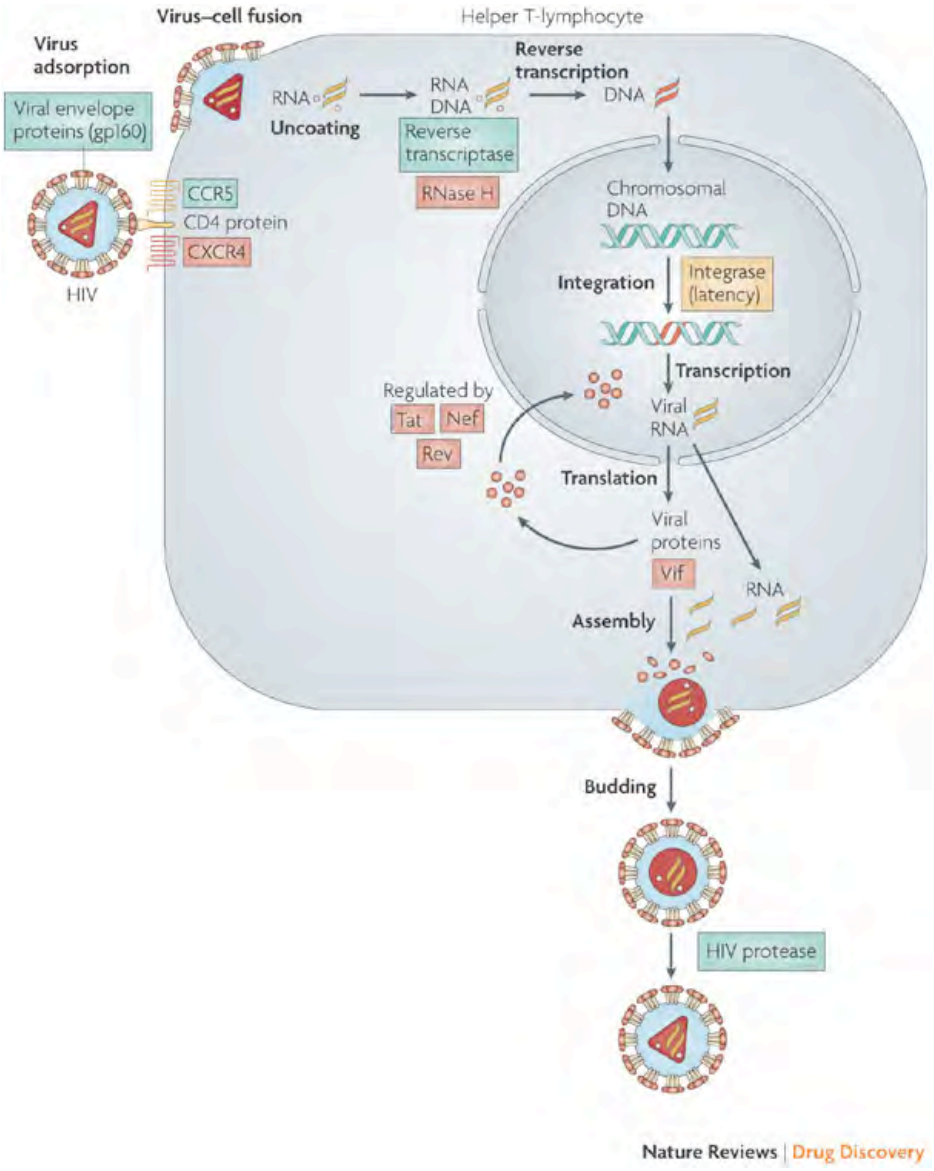
HIV-1 LIFE CYCLE

Human immunodeficiency virus, the causative organism of AIDS, is a member of the family retroviridae and genus lentivirus (Chiu et al., 1985). Lentiviruses or "slow viruses" are so named for the long interval between initial infection and onset of symptomatic disease. HIV-1 has 2 copies of a positive-sense 9.2kb RNA genome that encodes the viral proteins, both structural and functional. The gag gene encodes structural proteins that make up the viral skeleton, the pol gene encodes the three enzymes (protease, reverse transcriptase, and integrase), the env gene encodes the envelope proteins. The HIV-1 genome also has six other genes encoding the regulatory proteins – Tat, Rev, Vif, Vpr, Vpu and Nef, whose functions are yet to be completely understood. Briefly, the viral life cycle consists of the following steps: binding and entry, uncoating and reverse transcription, nuclear entry and integration, nuclear export, budding and maturation (Fig I.1).

Binding and Entry

HIV-1 infects cells of the immune system, mainly CD4+ T-cells, macrophages, and dentritic cells. The virus initiates binding through the interaction of its trimeric envelope complex (gp 160 spike) with the host cell receptor, CD4, a protein important for normal immune recognition (Dalgleish et al., 1984; Sattentau et al., 1993). The gp160 protein consists of a transmembrane domain (TM or gp41) and a surface domain (SU or gp120). gp120 attaches to CD4 receptor with high affinity, however this binding is insufficient for infection, which requires one of several chemokine receptors including CCR5 and

Figure I.1



Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Drug Discovery, Vol. 6, Flexner, HIV drug development – the next 25 years, 959-966, 2007.

Figure I.1. HIV-1 Life Cycle. The virus first interacts with a CD4+ receptor and a chemokine co-receptor through its envelope glycoprotein (**entry**). After binding, the viral core is released into the host cell, uncoated, and reverse-transcribed into the viral cDNA (**reverse transcription**). The viral cDNA is then transported to the nucleus and integrated into the host cell genome (**integration**). After transcription, viral mRNAs are transported to the cytoplasm and translated into the viral proteins (**nuclear export**). The structural proteins and enzymes are transported to the site of budding at the plasma membrane (**budding**). During or after budding, HIV-1 protease cleaves Gag and Gag-Pol polyproteins, leading to a infectious and mature virion (**maturation**). Proteins in green boxes are currently targeted by anti-viral drugs, yellow boxes indicate targets in clinical trials, and red boxes indicate potential drug targets (Pham et al., 2007).

CXCR4 (Clapham and Weiss, 1997). On binding of gp120 to CD4 receptor, conformational changes occur that allow the chemokine binding domains of gp120 to interact with either CCR5 or CXCR4 (Ugolini et al., 1997). This interaction results in structural rearrangements that allow gp41 to penetrate the cell membrane (Gallo, Puri, and Blumenthal, 2001). HR1 and HR2 repeat sequences within gp41 interact to collapse the extracellular part of gp41, bringing the viral and cell membranes in close proximity (Chan and Kim, 1998; Wyatt and Sodroski, 1998). The resulting fusion of viral and cell membranes allows the virus to enter the host cell (Wyatt and Sodroski, 1998).

Uncoating and Reverse Transcription

After fusion of the viral and host membranes, the viral core enters the host cytoplasm and undergoes a poorly understood process called uncoating, during which it forms the reverse transcription complex. Reverse transcription is mediated by the viral reverse transcriptase (RT), which has both RNA-dependent and DNA-dependent DNA polymerase activities (Yarchoan et al., 1985). RT also has a ribonuclease H (RNAse H) domain that degrades the RNA from the RNA-DNA hybrids formed during the reaction (Chandra, Gerber, and Chandra, 1986). Reverse transcription yields a double stranded viral cDNA that is then ready for integration into the host genome (Yarchoan et al., 1985). Due to its very essential role, RT was one of the first targets for anti-retroviral drugs and continues to play an important role in anti-HIV therapy (Broder et al., 1985; Castro et al., 2006; Frankel and Young, 1998).

HIV-1 RT lacks a proofreading mechanism, which causes transcription to be very error prone, leading to a large number of mutations in the newly synthesized cDNA. Additionally, HIV-1 has two copies each of its genomic RNA and because the RT binds to the genomic RNA template with relatively low affinity, transcription jumps from one strand to another resulting in a recombinant DNA genome derived from both parental RNA strands. As a result of these two factors, viral populations in an individual are highly variable forming quasi-species. Thus HIV can rapidly evolve to evade host immune responses and to develop resistance to antiretroviral therapies (Castro et al., 2006).

Nuclear Import and Integration

The newly synthesized viral cDNA is then transported to the nucleus as a nucleoprotein complex called the pre-integration complex (PIC). Integrase forms an important part of this complex and some studies have detected the presence of nucleocapsid, matrix as well as the accessory protein, Vpr (Bukrinsky et al., 1993; Gallay et al., 1995a; Gallay et al., 1995b). Vpr and integrase have been shown to play a role in the nuclear import of the PIC (Zhang et al., 2001) along with other cellular proteins like importin-α and nucleoporins. Once in the nucleus, integrase catalyzes the integration of the viral cDNA into the host genome. After integration, the provirus acts as a template for transcription and synthesis of viral mRNAs, a process that requires the viral transcriptional transactivator protein, Tat (Parada and Roeder, 1996). The viral mRNAs are unspliced (coding for Gag and GagPol polyprotein precursors and are encapsulated as genomic

RNA into the budding virions), partially spliced (coding for Env, Vif, Vpu and Vpr proteins) and highly spliced (coding for Rev, Tat and Nef proteins) (Castro et al., 2006; Emerman and Malim, 1998; Frankel and Young, 1998).

Nuclear Export

The spliced and unspliced mRNAs are exported into the cytoplasm where they are then translated into viral proteins. Cellular mRNAs are fully spliced before being transported to the cytoplasm. Partially spliced mRNAs of HIV-1 are transported with the help of the viral protein Rev, which binds the Rev response element (RRE) within the unspliced viral mRNAs (Emerman and Malim, 1998; Frankel and Young, 1998).

Viral Assembly

After their export from the nucleus, the viral mRNAs are translated into the full complement of viral proteins. The Gag and GagPol precursor polyproteins give rise to structural proteins that form the viral core and viral enzymes, respectively, and play an important role in assembly. The matrix domain of Gag is co-translationally myristolated and is responsible for targeting and binding the plasma membrane (Bryant and Ratner, 1990; Facke et al., 1993). Two copies of single-stranded genomic RNA are then encapsidated, mediated by interactions between the packaging signal and the nucleocapsid (Clever and Parslow, 1997; Laughrea et al., 1997). Following encapsidation, interactions between Gag nucleocapsid and RNA lead to alignment and packing of Gag molecules. The capsid protein has also been shown to play a role in viral

assembly (Gamble et al., 1997). HIV-1 envelope protein precursor, gp160, is synthesized in the endoplasmic reticulum. During its transport through the Golgi apparatus, gp 160 is cleaved by host proteases to generate the mature envelope proteins, gp120 and gp41 (Chan et al., 1997; Earl, Moss, and Doms, 1991). These envelope glycoprotein complexes are rapidly incorporated into viral particles through their interactions with the matrix domain of Gag (Frankel and Young, 1998).

Budding and maturation

During budding, the structural proteins, enzymes and encapsidated RNA genome pinch off from the host membrane by a process known as budding to give rise to an immature and a non-infectious viral particle. During or shortly after budding, the viral protease cleaves Gag and Gag-Pol polyproteins in a specific, sequential manner, producing mature and functional structural proteins and enzymes (Kohl et al., 1988). This protease-mediated cleavage results in significant structural rearrangements of viral proteins and formation of a cone-shaped core, leading to maturation of the viral particle. Because protease is essential for viral maturation, it is one of the most important targets for antiretroviral therapy. However, resistance to one or more protease inhibitors has lowered the efficacy of these drugs and will be discussed in detail in subsequent sections.

HIV-1 PROTEASE

HIV-1 protease, encoded by the pol gene, plays a pivotal role in the viral life cycle. The protease cleaves Gag and GagPol precursor polyproteins into fully functional enzymes and structural proteins, thus playing an essential role in viral maturation (Kohl et al., 1988) (Fig I.2). Mutations that inactivate the protease prevent viral maturation, resulting in a non-infectious viral article (Kramer et al., 1986). Due to its crucial role in the viral life cycle HIV-1 protease is an important target for anti-retroviral therapy (Kramer et al., 1986).

Structure, Mechanism of Action and Function

HIV-1 protease is an aspartyl protease that has the conserved Asp-Thr-Gly sequence as in the case of other retroviral proteases (Seelmeier et al., 1988). This enzyme has 99 amino acids and functions as a homodimer where the Asp-Thr-Gly triad from each monomer contributes to a single active site (Navia et al., 1989; Wlodawer et al., 1989) that is symmetric in the unliganded state. The protease has four main structural regions: the active site (residues 23-28, 81-84), the flap region (residues 40-60), the dimer interface (residues 1-10, 90-99) and the core of each monomer (residues 11-22, 29-39, 61-80, 85-90). The two aspartyl residues present at the base of the active site are essential for catalysis. HIV-1 protease catalyzes hydrolysis of peptide bonds by using an activated water molecule to catalyze nucleophilic attack of the carbonyl group of the scissile amide bond (Navia et al., 1989). The flaps form the roof of the active site and play an

Figure I.2

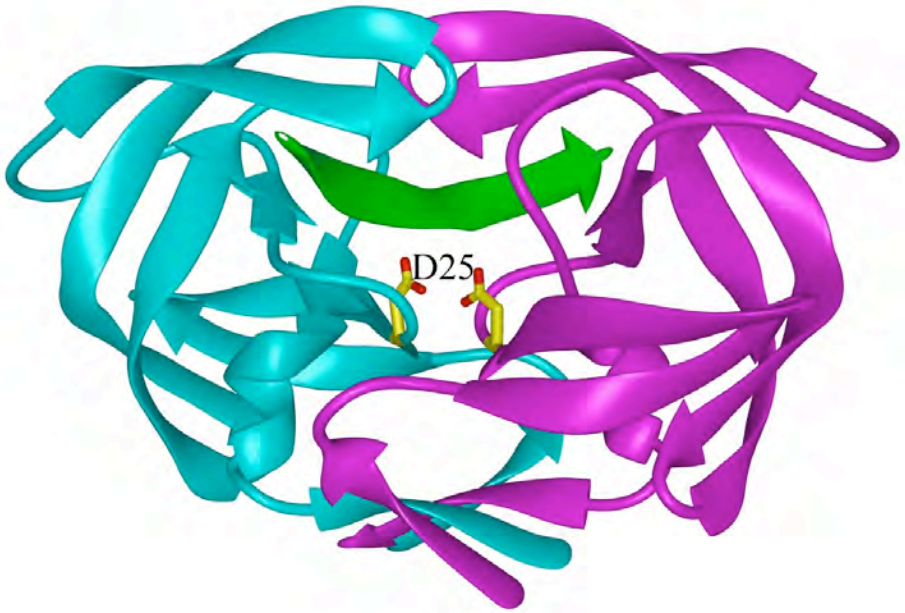


Figure I.2: Structure of HIV-1 Protease. (A.) Structure of HIV-1 protease complexed to a substrate. The two monomers are shown in cyan and magenta, and the catalytic aspartates are shown in yellow. In green is the substrate peptide.

important role in substrate binding. The substrate cleavage sites thread through the active site in an extended manner.

As mentioned previously, Gag and GagPol polyproteins give rise to HIV-1 structural proteins and viral enzymes. The translated polyproteins are targeted and bound to the inner surface of the plasma membrane. The rest of the viral particle is assembled and the new immature viruses are released from the host cell via budding. During or after budding, the HIV-1 protease cleaves at least ten different sites within Gag and GagPol polyproteins giving rise to the structural proteins, matrix (MA), capsid (CA), nucleocapsid (NC) and p6; the viral enzymes, protease (PR), reverse transcriptase (RT) and integrase; and two small spacer peptides, p1 and p2 (Fig I.3) (Kohl et al., 1988). This proteolytic processing of Gag and GagPol leads to the formation of mature viral particles that are now infectious (Kramer et al., 1986).

Proteolytic processing of Gag and GagPol is a sequential and highly ordered process (Pettit et al., 1994a; Wiegers et al., 1998). The order of cleavage within Gag has been shown to be very specific and is important for fitness/infectivity (Pettit et al., 2002; Pettit et al., 1998). The first site to be processed is the CA-p2 cleavage site, with the remaining sites processed as shown in (Fig I.4). Though the exact mechanism of Gag processing is not understood, studies have shown that several factors play a role in determining the order of cleavage, including the cleavage site sequence, presence of spacer domains and the interactions with RNA (Feher et al., 2002; Pettit et al., 2002; Pettit et al., 1994b; Sheng et al., 1997).

Figure I.3.

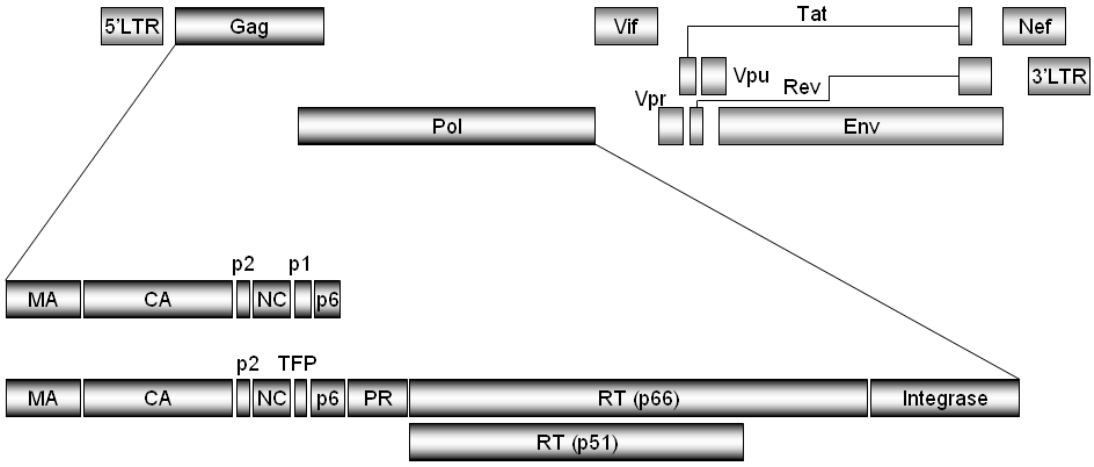

5'LTR
Gag
Vif
Tat
Nef
Vpu
Rev
3'LTR
Vpr
Pol
Env
p2
p1
MA
CA
NC
p6
p2
TFP
MA
CA
NC
p6
PR
RT (p66)
Integrase
RT (p51)

Figure I.3. Schematic Representation of HIV-1 Genome and Organization of Gag and GagPol Polyproteins.

Figure I.4.

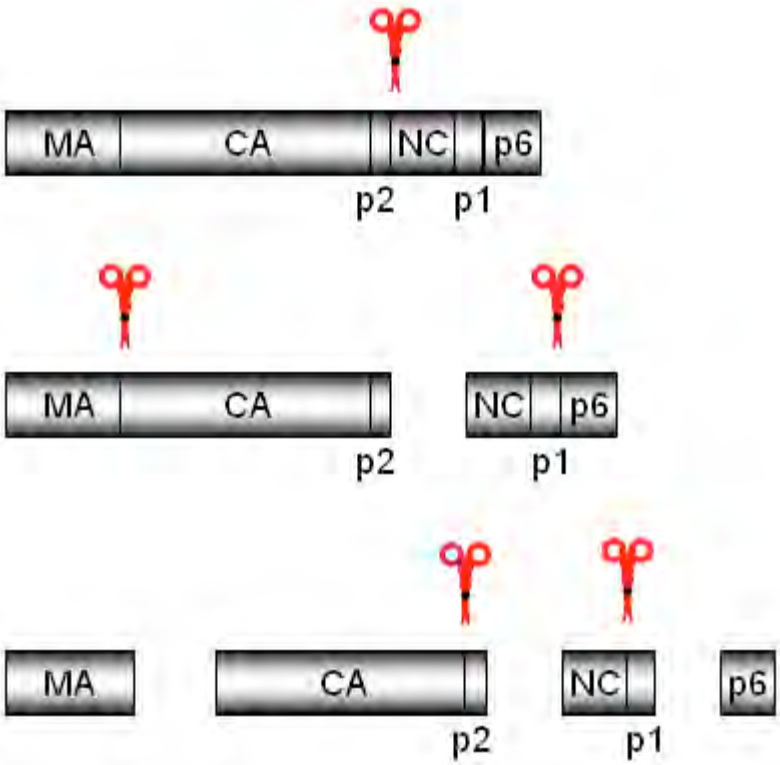


Figure I.4. A Schematic Overview of the Sequential Proteolytic Processing of HIV-1 Gag Polyprotein.

Substrate Recognition, Binding and Cleavage

HIV-1 protease recognizes and cleaves Gag and Gag-Pol polyproteins at least ten different sites resulting in fully functional viral proteins (Fig I.3). The flaps of the protease close down on the active site forming a tunnel through which the substrates are threaded. The specific amino acids in the P4 to P3' positions of the substrate around the scissile bond are essential for substrate recognition (Tözsér et al., 1991a; Weber et al., 1989). The corresponding binding sites within the enzyme include seven sites, S4 through S3'. Though the protease has been shown to cleave peptides as short as 6 amino acids, efficient processing occurs when the peptide is seven or more residues longer(Griffiths et al., 1992; Tözsér et al., 1991b). Amino acids outside the cleavage site, and the context of cleavage sites within the Gag or Gag-Pol polyprotein can influence the rates of cleavage at various sites (Griffiths et al., 1992; Kohl et al., 1988; Tözsér et al., 1991a).

Analysis of Gag cleavage sites from a number of retroviruses showed patterns of amino acid preferences that could broadly be classified into Type 1 and Type 2 cleavage sites (Pettit et al., 1991). Type 1 sites had Pro in their P1' sites whereas Type 2 had Ala, Val or Leu. These two classes also had distinct differences in the flanking amino acids. Type 1 sites were also observed to have a small amino acid at P4, uncharged residues at P3 with Gln observed occasionally and aliphatic amino acids at P2 or Asn. P1 had Tyr, Phe or Leu and P2' had aliphatic residues. Type 2 sites had small amino acids at the P4 position similar to Type 1. P3 amino acids were uncharged with the exception of an

occasional Arg or Lys. P2 had aliphatic residues like the Type 1, however Leu was most frequent at P1 unlike Type 1 sites. P2' and P3' were aliphatic with Ile excluded from P2' and Ala favored at P3'. However, despite other studies to determine and characterize sub-site amino acid preferences, substrate recognition was still poorly understood.

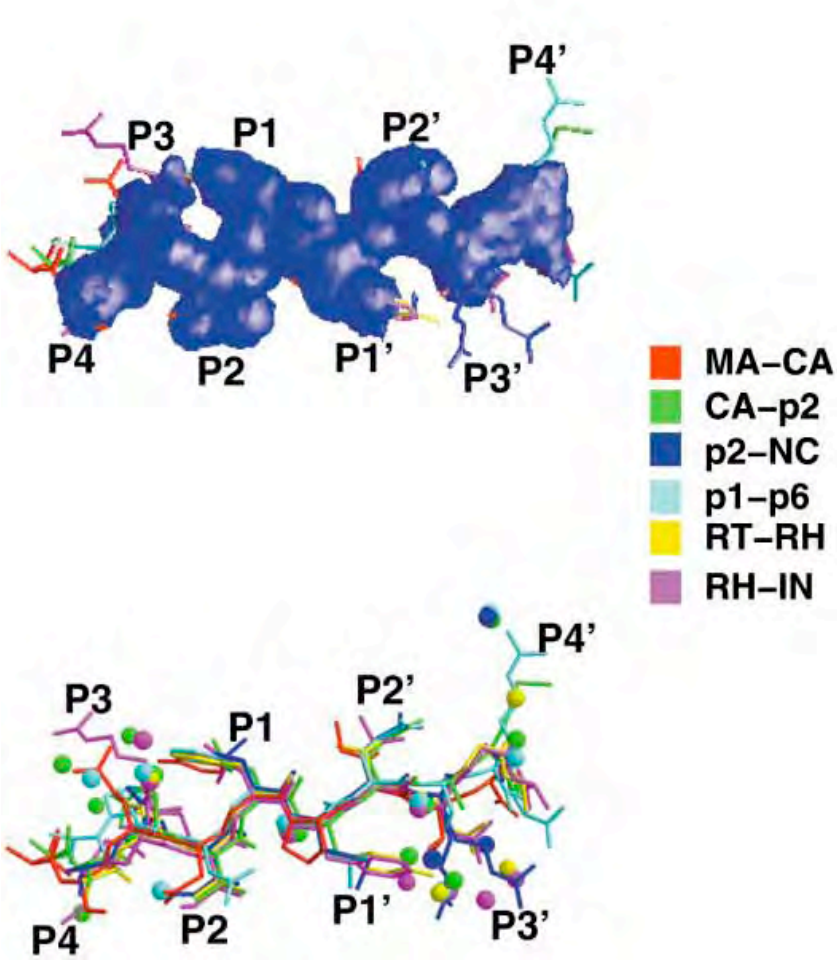
As described in the preceding section, amino acid preferences at the different sites within the cleavage site sequences were known, but how these ten highly variable cleavage site sequences were recognized by the protease was still not understood. From Fig I.5 it is evident that the various cleavage site sequences are non-homologous and asymmetric in both charge and size. These characteristics begged the question as to how a symmetric protease could recognize and cleave an asymmetric substrate. Structural studies from the Schiffer lab have shown that the various cleavage site peptides adopt a conserved shape/volume, which was hypothesized as the basis for recognition of substrate sites by the HIV-1 protease (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002). The overlapping volume of four or more of the substrates within the active site of the protease defines the conserved shape or the “substrate envelope” (Fig1.6). The P1-P3 region forms a toroid, which is thought to be important for specificity, whereas the numerous backbone-to-backbone interactions facilitate binding (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002). The substrate envelope not only explains specificity but also has significant implications for the development of resistance to various protease inhibitors and substrate co-evolution.

Figure I.5.

Cleavage site	Amino acid sequence											
Gag	P5	P4	P3	P2	P1	/	P1'	P2'	P3'	P4'	P5'	*Gag amino acid positions*
MA/CA	V	S	Q	N	Y	/	P	I	V	Q	N	128-137
CA/p2	K	A	R	V	L	/	A	E	A	M	S	359-368
p2/NC	S	A	T	I	M	/	M	Q	R	G	N	373-382
NC/p1	E	R	Q	A	N	/	F	L	G	K	I	428-437
p1/p6^{gag}	R	P	G	N	F	/	L	Q	S	R	P	444-453
Pol	P5	P4	P3	P2	P1	/	P1'	P2'	P3'	P4'	P5'	
NC/TFP	E	R	Q	A	N	/	F	F	R	E	D	
TFP/p6^{pol}	E	D	L	A	F	/	L	Q	G	K	A	
p6^{pol}/PR	V	S	F	N	F	/	P	Q	I	T	L	
PR/RT	C	T	L	N	F	/	P	I	S	P	I	
RT(p51)/RT(p15)	G	A	E	T	F	/	Y	V	D	G	A	
RT/IN	I	R	K	V	L	/	F	L	D	G	I	

Figure I.5. HIV-1 Protease Cleavage Site Sequences. These are based on the HXB2 sequence. Slashes indicate the site where the proteeolytic cleavage occurs (MA:matrix, CA:capsid, NC:nucleocapsid, TFP:transframe protein, PR:protease,RT:reverse transcriptase, IN:integrase).

Figure I.6



Reprinted from Structure, Vol. 10, Prabu-Jeyabalan, Nalivaika, and Schiffer. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes, 369-381, 2002 with permission from Elsevier.

Figure I.6: The Substrate envelope. In blue is the consensus volume occupied by at least four out of six cleavage site peptides when bound to the D25N inactive HIV-1 protease. The six peptides are shown in the same superposition within the consensus volume.

PROTEASE INHIBITORS AND RESISTANCE

The earliest indications for a critical role of the protease in the HIV-1 life cycle were mutational studies that inactivated the protease, generating non-infectious viral particles (Kohl et al., 1988). Further studies led to the development of various types of molecules that bind within the active site of the protease, inhibiting its activity and leading to loss of viral infectivity. Structural studies of HIV-1 protease in combination with these initial inhibitor molecules led to design several clinically successful protease inhibitors that have significantly slowed the progress of HIV-1 infection and delayed the onset of AIDS.

Protease inhibitor design is based on peptidomimetics that target the fact that the protease was an aspartyl protease. Currently, nine Protease inhibitors are being used in the clinic: ritonavir (RTV), saquinavir (SQV), lopinavir (LPV), indinavir (IDV), nelfinavir (NFV), (fos)amprenavir (APV), atazanavir (ATV), darunavir (DRV), and tipranavir (TPV) (Craig et al., 1991; Kempf et al., 1995; Partaledis et al., 1995; Patick et al., 1996; Robinson et al., 2000) (Sham et al., 1998; Turner et al., 1998; Vacca et al., 1994). Except for TPV, all mimic the natural substrates of the protease and bind competitively within its active site. These peptidomimetic inhibitors have a hydroxyl group at a position equivalent to the scissile bond in the natural cleavage sites. The hydroxyl group displaces the active site water and interacts with the active site aspartates, mimicking the transition state. These inhibitors are much shorter than the substrate peptide sequence required for recognition and cleavage. In addition, they have large hydrophobic groups surrounding the hydroxyl group, which form backbone interactions with the protease as well as van der Waals

interactions with residues in the protease sub sites. Together, these interactions enable the protease inhibitors to bind within the active site of HIV-1 protease with high specificity and affinity.

Unfortunately, the efficacy of all protease inhibitors used in the clinic has been quickly reduced by the emergence of resistance. Currently, protease inhibitors are combined with other types of antiretroviral drugs as a part of HAART. Though initially successful in suppressing the virus, the resulting resistance to the various drugs has led to ineffective therapy. The emergence of highly evolved resistant viruses is due to several factors, including the high replication rate of HIV-1, and lack of a proofreading mechanism by the RT. Resistance to protease inhibitors is a very complex process, in which initial mutations arise within the protease gene, but with continued therapy the virus evolves to acquire multiple mutations both within the protease and other regions of the HIV-1 genome (Condra et al., 1995; Kaplan et al., 1994; Kaplan et al., 1993; Molla et al., 1996).

Evolution of Protease Resistance

Viral populations in HIV-1 infected individuals exist as complex and dynamic mixtures of mutant genomes, referred to as quasi-species, with usually one species dominating as result of being the fittest (Domingo and Holland, 1997; Eigen, 1993). Protease inhibitor therapy suppresses viral replication and infectivity resulting in the emergence of other viral species that are more fit. Thus, the selective pressure of protease inhibitor therapy results in a wide spectrum of amino acid substitutions within the protease that is advantageous to the virus.

Under protease inhibitor therapy, a majority of initial mutations arise within the active site of the enzyme, directly affecting inhibitor binding and are the primary cause of resistance to protease inhibitors. Typical primary mutations include D30N, G48V, I50L/V, V82A/F/T, I84V and L90M (Gulnik et al., 1995), some of which have been investigated in my dissertation research. Several primary PI resistance mutations have been described that are a signature of particular protease inhibitors. For example, patients failing NFV therapy develop the D30N protease mutation (Patick et al., 1998), while the I50V and I50L mutations are selected in patients failing amprenavir and atazanavir therapy, respectively (Colonno et al., 2004; Mahalingam et al., 1999). Mutations at protease residue 82 are observed in patients treated with RTV and SQV, and the G48V mutation results in resistance to SQV and ATV(Deeks et al., 1998; Molla et al., 1996). The I84V mutation is one of the most severe primary resistance mutations as it causes cross-resistance to most protease inhibitors (Zolopa et al., 1999). Thus, a range of primary resistance mutations are selected, some of which are unique to a single protease inhibitors, whereas others confer resistance to two or more protease inhibitors. Subsequent to the primary mutations, additional mutations arise that further reduce sensitivity to protease inhibitors. These secondary mutations are often located distant from the active site, although the mechanism by which they enhance resistance is poorly understood.

Primary and secondary protease mutations provide a selective advantage to HIV-1 virus by significantly reducing its susceptibility to protease inhibitors. However, these mutations lower the catalytic activity of the viral protease (Gulnik et al., 1995). Reduced

protease efficacy not only disrupts Gag and GagPol processing as demonstrated by the accumulation of partially cleaved products, but also results in loss of viral fitness as shown by reduced replicative capacities (Croteau et al., 1997; Mahalingam et al., 1999; Nijhuis et al., 1999; Zennou et al., 1998). With prolonged protease inhibitor therapy, however, mutations emerge that compensate for the deleterious effects of primary and secondary mutations in several ways. One way is that compensatory mutations accumulate within the protease and improve viral fitness by mechanisms that are not well understood (Mammano et al., 2000; Martinez-Picado et al., 1999). Conformational changes in the protease as a result of these mutations likely allow the protease to maintain adequate function even as it maintains resistance to inhibitors (King et al., 2004; Prabu-Jeyabalan et al., 2003). Examples of compensatory mutations that have been clearly shown to improve viral fitness are L10I, L63P and A71V (Mammano et al., 2000; Martinez-Picado et al., 1999; Perrin and Mammano, 2003). However, several other mutations arise, and it is not known if they are solely compensatory or if they only contribute to enhance resistance. It is likely that some of these mutations actually contribute to both improvements in viral fitness compensatory and resistance.

Co-evolution of protease cleavage sites

Due to inefficient Gag and GagPol processing compensatory mutations also develop within the substrate cleavage sites following accumulation of resistance mutations within the protease (Doyon et al., 1996; Zhang et al., 1997). Substrate mutations were first reported within the NC-p1 and p1-p6 cleavage sites. The A431V mutation within the

NC-p1 cleavage site and L449F in the p1-p6 cleavage site selected during the evolution of protease inhibitor resistance were observed to correlate with V82A and I50V protease resistance mutations, respectively (Maguire et al., 2002; Zhang et al., 1997). These cleavage site mutations have been demonstrated to be compensatory in nature by improving replicative capacity/processing (Feher et al., 2002; Prabu-Jeyabalan et al., 2004). Other cleavage site mutations, including I437V and P453R, have now been well documented and are associated with several major protease resistance mutations (Maguire et al., 2002; Nijhuis et al., 1999), suggesting that with prolonged protease inhibitor therapy, evolution of protease cleavage sites could be a fairly frequent mechanism for maintaining viral fitness even as the virus evolves resistance to protease inhibitors.

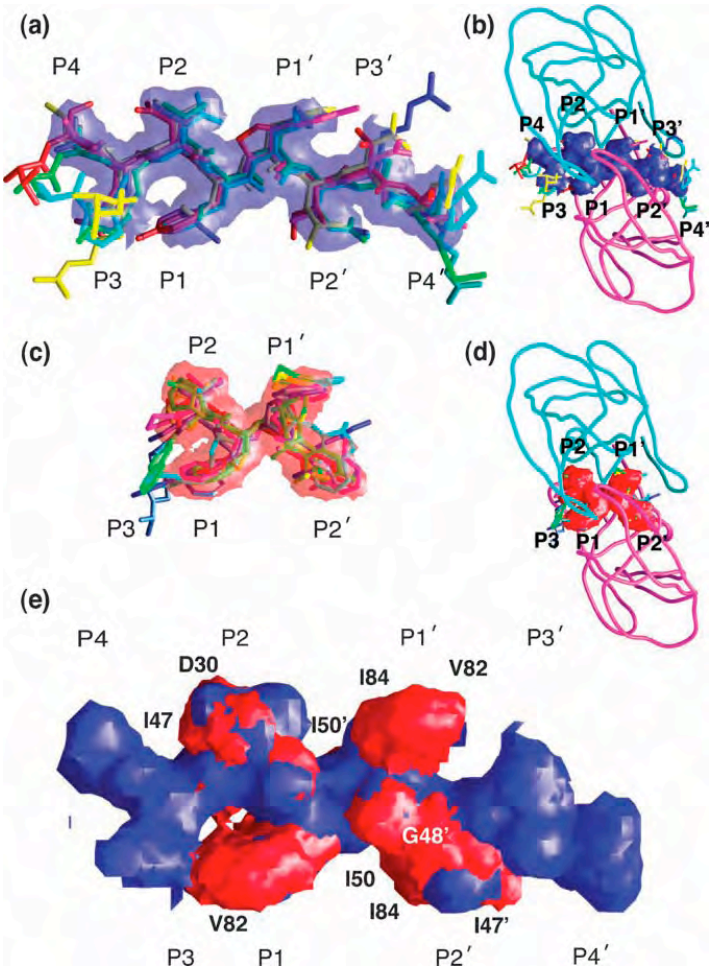
Indeed, Gag processing has been shown to be enhanced by the A431V and I437V mutations within the NC-p1 cleavage site (Dam et al., 2009; Feher et al., 2002). In fact, there were clear structural changes that increased binding of the A431V NC-p1 site with the V82A protease (Prabu-Jeyabalan et al., 2004). Recently though, both A431V and I437V have been shown directly increase resistance, possibly as a result of this enhanced Gag processing (Dam et al., 2009; Nijhuis et al., 2007). Similarly, the L449F mutation within the p1-p6 cleavage site has been shown to increase processing at this cleavage site (Feher et al., 2002; Kolli, Lastere, and Schiffer, 2006; Maguire et al., 2002). Structural modeling studies have observed that the change from a smaller amino acid to a larger Phe likely improves van der Waals contacts contributing improved Gag processing (Kolli, Lastere, and Schiffer, 2006). This suggests a likely mechanism whereby decreased

interactions between cleavage sites and mutant protease can be offset by compensatory mutations within the cleavage sites leading to improved binding and processing.

Drug Resistance – A change in molecular recognition

The development of drug resistance is a major factor for the failure of protease inhibitor therapy. As described previously, the virus evolves to accumulate a multitude of mutations within the protease that prevent protease inhibitors from binding to the protease. More than half the residues within the protease mutate in different combinations and lead to drug resistance. Thus the protease is very plastic which raises the question as to how it continues to recognize and cleave its substrates while preventing the inhibitor from binding. Structural studies from the Schiffer lab have shown that most of the protease inhibitors occupy a similar volume and contact similar resides within the active site of the protease (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002) (Fig I.7). It has been demonstrated that most of the primary resistance associated mutations within the active site of the protease occur at sites where inhibitor atoms contact protease residues beyond the substrate envelope and are thus more important for inhibitor binding than substrate recognition (King et al., 2004; Prabu-Jeyabalan et al., 2006). King et al., (King et al., 2004) also suggested that if, however, the substrate atoms contact the very same residues in the active site of the protease, this could lead to impaired recognition and cleavage of the substrates leading to the evolution of compensatory mutations within the protease cleavage sites. These studies have led to design of inhibitors that fit well within the substrate envelope and contact the same residues within the protease that are

Figure I.7.



Reprinted from Chemistry and Biology, Vol. 11. King, N. *et al*., Combating susceptibility to drug resistance: lessons from HIV-1 protease, 1333-1338, 2004 with permission from Elsevier.

Figure I.7: Substrate and Inhibitor Envelopes of HIV-1 Protease. (A) Substrate envelope in blue with the six substrate peptides superposed. (B) Top view of the substrate envelope within the active site of the protease. (C) Inhibitor envelope in red, calculated from the overlapping volume of five or more of the inhibitor complexes with active HIV-1 protease. (D) Top view of the inhibitor envelope within the active site of the protease. (E) Superposition of the substrate envelope and inhibitor envelope. Also shown are residues contacted by the inhibitor atoms that mutate to cause drug resistance.

required for substrate recognition. Such inhibitors may be more successful in preventing drug resistance, as the likelihood of several substrates co-evolving simultaneously is lower.

Scope of the thesis

Drug resistance is a major obstacle for the success of protease inhibitors in combating HIV-1 infections. Investigating the evolution of resistance to protease inhibitors, including substrate co-evolution, would give insights into the molecular mechanisms underlying drug resistance and lead to design of more robust protease inhibitors. Substrate co-evolution is becoming increasingly common, especially in the NC-p1 and p1p6 sites. Many of these mutations improve viral fitness by enhancing Gag processing. Also, some mutations within Gag cleavage site co-evolve to directly enhance resistance. An in depth study of co-evolution would help identify correlating mutations in the cleavage sites and the protease. The impact of this correlation on viral fitness and drug susceptibility would give insights into the evolution of drug resistance. Such a study would have direct implications for evaluating patients failing protease inhibitor therapy and formulating treatment. X-ray crystallographic studies to study structural changes in the protease as well as the co-evolving substrate peptide will give valuable insights into co-evolution and the underlying mechanisms for drug resistance, leading to design of more robust protease inhibitors.

References

Broder, S., Yarchoan, R., Collins, J. M., Lane, H. C., Markham, P. D., Klecker, R. W., Redfield, R. R., Mitsuya, H., Hoth, D. F., Gelmann, E., and et al. (1985). Effects of suramin on HTLV-III/LAV infection presenting as Kaposi's sarcoma or AIDS-related complex: clinical pharmacology and suppression of virus replication in vivo. *Lancet* **2**(8456)**,** 627-30.

Bryant, M., and Ratner, L. (1990). Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proceedings of the National Academy of Sciences of the United States of America* **87**(2)**,** 523-7.

Bukrinsky, M. I., Sharova, N., McDonald, T. L., Pushkarskaya, T., Tarpley, W. G., and Stevenson, M. (1993). Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proceedings of the National Academy of Sciences of the United States of America* **90**(13)**,** 6125-9.

Castro, H. C., Loureiro, N. I., Pujol-Luz, M., Souza, A. M., Albuquerque, M. G., Santos, D. O., Cabral, L. M., Frugulhetti, I. C., and Rodrigues, C. R. (2006). HIV-1 reverse transcriptase: a therapeutical target in the spotlight. *Current Medicinal Chemistry* **13**(3)**,** 313-24.

Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**(2)**,** 263-73.

Chan, D. C., and Kim, P. S. (1998). HIV entry and its inhibition. *Cell* **93**(5)**,** 681-4.

Chandra, A., Gerber, T., and Chandra, P. (1986). Biochemical heterogeneity of reverse transcriptase purified from the AIDS virus, HTLV-III. *FEBS Letters* **197**(1-2)**,** 84-8.

Chermann, J. C., Barre-Sinoussi, F., Dauguet, C., Brun-Vezinet, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a new retrovirus in a patient at risk for acquired immunodeficiency syndrome. *Antibiotics and Chemotherapy* **32,** 48-53.

Chiu, I. M., Yaniv, A., Dahlberg, J. E., Gazit, A., Skuntz, S. F., Tronick, S. R., and Aaronson, S. A. (1985). Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. *Nature* **317**(6035)**,** 366-8.

Clapham, P. R., and Weiss, R. A. (1997). Immunodeficiency viruses. Spoilt for choice of co-receptors. *Nature* **388**(6639)**,** 230-1.

Clever, J. L., and Parslow, T. G. (1997). Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *Journal of Virology* **71**(5)**,** 3407-14.

Colonno, R., Rose, R., McLaren, C., Thiry, A., Parkin, N., and Friborg, J. (2004). Identification of I50L as the signature atazanavir (ATV)-resistance mutation in treatment-naive HIV-1-infected patients receiving ATV-containing regimens. *J Infect Dis* **189**(10)**,** 1802-10.

Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., Shivaprakash, M., Titus, D., Yang, T.,

Teppler, H., Squires, K. E., Deutsch, P. J., and Emini, E. (1995). *In vivo* emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* **374,** 569-571.

Craig, J. C., Duncan, I. B., Hockley, D., Grief, C., Roberts, N. A., and Mills, J. S. (1991). Antiviral properties of Ro 31-8959, an inhibitor of human immunodeficiency virus (HIV) proteinase. *Antiviral Research* **16**(4)**,** 295-305.

Croteau, G., Doyon, L., Thibeault, D., McKercher, G., Pilote, L., and Lamarre, D. (1997). Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. *J Virol* **71**(2)**,** 1089-96.

Dalgleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**(5996)**,** 763-7.

Dam, E., Quercia, R., Glass, B., Descamps, D., Launay, O., Duval, X., Krausslich, H. G., Hance, A. J., and Clavel, F. (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**(3)**,** e1000345.

Deeks, S. G., Grant, R. M., Beatty, G. W., Horton, C., Detmer, J., and Eastman, S. (1998). Activity of a ritonavir plus saquinavir-containing regimen in patients with virologic evidence of indinavir or ritonavir failure. *AIDS Res Hum Retroviruses* **12,** F97-F102.

Domingo, E., and Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annual Review of Microbiology* **51,** 151-78.

Dorman, S. E., and Chaisson, R. E. (2007). From magic bullets back to the magic mountain: the rise of extensively drug-resistant tuberculosis. *Nature Medicine* **13**(3)**,** 295-8.

Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L., and Lamarre, D. (1996). Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol* **70**(6)**,** 3763-9.

Earl, P. L., Moss, B., and Doms, R. W. (1991). Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein. *Journal of Virology* **65**(4)**,** 2047-55.

Eigen, M. (1993). Viral quasispecies. *Scientific American* **269**(1)**,** 42-9.

Emerman, M., and Malim, M. H. (1998). HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology. *Science* **280**(5371)**,** 1880-4.

Facke, M., Janetzko, A., Shoeman, R. L., and Krausslich, H. G. (1993). A large deletion in the matrix domain of the human immunodeficiency virus gag gene redirects virus particle assembly from the plasma membrane to the endoplasmic reticulum. *Journal of Virology* **67**(8)**,** 4972-80.

Feher, A., Weber, I. T., Bagossi, P., Baross, P., Mahalingam, B., Louis, J. M., Copeland, T. D., Yorshin, I. Y., Harrison, R. W., and Tozser, J. (2002). Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. *J Biochem* **269,** 4114-4120.

Fischbach, M. A., and Walsh, C. T. (2009). Antibiotics for emerging pathogens. *Science* **325**(5944)**,** 1089-93.

Frankel, A. D., and Young, J. A. (1998). HIV-1: fifteen proteins and an RNA. *Annual Review of Biochemistry* **67,** 1-25.

Gallay, P., Swingler, S., Aiken, C., and Trono, D. (1995a). HIV-1 infection of nondividing cells: C-terminal tyrosine phosphorylation of the viral matrix protein is a key regulator. *Cell* **80**(3)**,** 379-88.

Gallay, P., Swingler, S., Song, J., Bushman, F., and Trono, D. (1995b). HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell* **83**(4)**,** 569-76.

Gallo, S. A., Puri, A., and Blumenthal, R. (2001). HIV-1 gp41 six-helix bundle formation occurs rapidly after the engagement of gp120 by CXCR4 in the HIV-1 Env-mediated fusion process. *Biochemistry* **40**(41)**,** 12231-6.

Gamble, T. R., Yoo, S., Vajdos, F. F., von Schwedler, U. K., Worthylake, D. K., Wang, H., McCutcheon, J. P., Sundquist, W. I., and Hill, C. P. (1997). Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* **278**(5339)**,** 849-53.

Griffiths, J. T., Phylip, L. H., Konvalinka, J., Strop, P., Gustchina, A., Wlodawer, A., Davenport, R. J., Briggs, R., Dunn, B. M., and 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**(22)**,** 5193-5200.

Gulnik, S. V., Suvorov, L. I., Liu, B., Yu, B., Anderson, B., Mitsuya, H., and Erickson, J. W. (1995). Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* **34,** 9282-9287.

Kaplan, A. H., Michael, S. F., Wehbie, R. S., Knigge, M. F., Paul, D. A., Everitt, L., Kempf, D. J., Norbeck, D. W., Erickson, J. W., and Swanstrom, R. (1994). Selection of multiple human immunodeficiency virus type 1 variants that encode viral proteases with decreased sensitivity to an inhibitor of the viral protease. *Proc Natl Acad Sci U S A* **91**(12)**,** 5597-601.

Kaplan, A. H., Zack, J. A., Knigge, M., Paul, D. A., Kempf, D. J., Norbeck, D. W., and Swanstrom, R. (1993). Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles. *Journal of Virology* **67**(7)**,** 4050-5.

Kempf, D. J., Marsh, K. C., Denissen, J. F., McDonald, E., Vasavanonda, S., Flentge, C. A., Green, B. E., Fino, L., Park, C. H., Kong, X. P., and et al. (1995). ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proceedings of the National Academy of Sciences of the United States of America* **92**(7)**,** 2484-8.

King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004). Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chem Biol* **11**(10)**,** 1333-8.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., and Sigal, I. S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci U S A* **85,** 4686-4690.

Kolli, M., Lastere, S., and Schiffer, C. A. (2006). Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virology* **347**(2)**,** 405-9.

Kramer, R. A., Schaber, M. D., Skalka, A. M., Ganguly, K., Wong-Staal, F., and Reddy, E. P. (1986). HTLV-III gag protein is processed in yeast cells by the virus pol-protease. *Science* **231**(4745)**,** 1580-4.

Laughrea, M., Jette, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M. A. (1997). Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *Journal of Virology* **71**(5)**,** 3397-406.

Maguire, M. F., Guinea, R., Griffin, P., Macmanus, S., Elston, R. C., Wolfram, J., Richards, N., Hanlon, M. H., Porter, D. J., Wrin, T., Parkin, N., Tisdale, M., Furfine, E., Petropoulos, C., Snowden, B. W., and Kleim, J. P. (2002). Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J Virol* **76**(15)**,** 7398-406.

Mahalingam, B., Louis, J., Reed, C., Adomat, J., Krouse, J., Wang, Y., Harrison, R., and Weber, I. (1999). Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur J Biochem* **263**(1)**,** 238-45.

Mammano, F., Trouplin, V., Zennou, V., and Clavel, F. (2000). Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. *Journal of Virology* **74**(18)**,** 8524-31.

Martinez-Picado, J., Savara, A. V., Sutton, L., and D'Aquila, R. T. (1999). Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *Journal of Virology* **73**(5)**,** 3744-52.

Molla, A., Korneyeva, M., Gao, Q., Vasavanonda, S., Schipper, P. J., Mo, H. M., Markowitz, M., Chernyavskiy, T., Niu, P., Lyons, N., Hsu, A., Granneman, G. R., Ho, D. D., Boucher, C. A., Leonard, J. M., Norbeck, D. W., and Kempf, D. J. (1996). Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat Med* **2**(7)**,** 760-6.

Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **337**(6208)**,** 615-20.

Nijhuis, M., Schuurman, R., de Jong, D., Erickson, J., Gustchina, E., Albert, J., Schipper, P., Gulnik, S., and Boucher, C. A. (1999). Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *Aids* **13**(17)**,** 2349-59.

Nijhuis, M., van Maarseveen, N. M., Lastere, S., Schipper, P., Coakley, E., Glass, B., Rovenska, M., de Jong, D., Chappey, C., Goedegebuure, I. W., Heilek-Snyder, G., Dulude, D., Cammack, N., Brakier-Gingras, L., Konvalinka, J., Parkin, N., Krausslich, H. G., Brun-Vezinet, F., and Boucher, C. A. (2007). A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med* **4**(1)**,** e36.

Palella, F. J., Jr., Delaney, K. M., Moorman, A. C., Loveless, M. O., Fuhrer, J., Satten, G. A., Aschman, D. J., and Holmberg, S. D. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *New England Journal of Medicine* **338**(13)**,** 853-60.

Parada, C. A., and Roeder, R. G. (1996). Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain. *Nature* **384**(6607)**,** 375-8.

Partaledis, J. A., Yamaguchi, K., Tisdale, M., Blair, E. E., Falcione, C., Maschera, B., Myers, R. E., Pazhanisamy, S., Futer, O., Cullinan, A. B., and et al. (1995). In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) isolates with reduced sensitivity to hydroxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. *J Virol* **69**(9)**,** 5228-5235.

Patick, A., Duran, M., Cao, Y., Shugarts, D., Keller, M., Mazabel, E., Knowles, M., Chapman, S., Kuritzkes, D., and Markowitz, M. (1998). Genotypic and phenotypic characterization of human
immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob Agents Chemother* **42**(10)**,** 2637-44.

Patick, A., Mo, H., Markowitz, M., Appelt, K., Wu, B., Musick, L., Kalish, V., Kaldor, S., Reich, S., Ho, D., and Webber, S. (1996). Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrob Agents Chemother* **40**(2)**,** 292-7.

Perrin, V., and Mammano, F. (2003). Parameters driving the selection of nelfinavir-resistant human immunodeficiency virus type 1 variants. *Journal of Virology* **77**(18)**,** 10172-5.

Pettit, S. C., Henderson, G. J., Schiffer, C. A., and 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.

Pettit, S. C., Moody, M. D., Wehbie, R. S., Kaplan, A. H., Nantermet, P. V., Klein, C. A., and Swanstrom, R. (1994a). The p2 domain of human immunodeficiency virus type 1 gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *Journal of Virology* **68**(12)**,** 8017-8027.

Pettit, S. C., Sheng, N., Tritch, R., Erickson-Vitanen, S., and Swanstrom, R. (1998). The regulation of sequential processing of HIV-1 Gag by the viral protease. *Adv Exp Med Biol* **436,** 15-25.

Pettit, S. C., Simsic, J., Loeb, D. D., Everitt, L., Hutchison, C. A. d., and 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**(22)**,** 14539-14547.

Pettit, S. P., Moody, M. D., Wehbie, R. S., Kaplan, A. H., Nantermet, P. V., Klein, C. A., and Swanstrom, R. (1994b). 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**(12)**,** 8017-8027.

Pham, P. A., Hendrix, C. W., Barditch-Crovo, P., Parsons, T., Khan, W., Parish, M., Radebaugh, C., Carson, K. A., Pakes, G. E., Qaqish, R., and Flexner, C. (2007). Amprenavir and lopinavir pharmacokinetics following coadministration of amprenavir or fosamprenavir with lopinavir/ritonavir, with or without efavirenz. *Antivir Ther* **12**(6)**,** 963-9.

Prabu-Jeyabalan, M., King, N. M., Nalivaika, E. A., Heilek-Snyder, G., Cammack, N., and Schiffer, C. A. (2006). Substrate envelope and drug resistance: crystal structure of RO1 in complex with wild-type human immunodeficiency virus type 1 protease. *Antimicrobial Agents and Chemotherapy* **50**(4)**,** 1518-21.

Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., and Schiffer, C. A. (2003). Viability of a drug-resistant HIV-1 protease variant: structural insights for better anti-viral therapy. *J Virol* **77**(2)**,** 1306-15.

Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., and Schiffer, C. A. (2004). Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *Journal of Virology* **78**(22)**,** 12446-54.

Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2002). Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* **10**(3)**,** 369-381.

Robinson, B. S., Riccardi, K. A., Gong, Y. F., Guo, Q., Stock, D. A., Blair, W. S., Terry, B. J., Deminie, C. A., Djang, F., Colonno, R. J., and Lin, P. F. (2000). BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrobial Agents and Chemotherapy* **44**(8)**,** 2093-9.

Sattentau, Q. J., Moore, J. P., Vignaux, F., Traincard, F., and Poignard, P. (1993). Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *Journal of Virology* **67**(12)**,** 7383-93.

Seelmeier, S., Schmidt, H., Turk, V., and von der Helm, K. (1988). Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proceedings of the National Academy of Sciences of the United States of America* **85**(18)**,** 6612-6.

Shafer, R. W., and Schapiro, J. M. (2005). Drug resistance and antiretroviral drug development. *Journal of Antimicrobial Chemotherapy* **55**(6)**,** 817-20.

Sham, H. L., Kempf, D. J., Molla, A., Marsh, K. C., Kumar, G. N., Chen, C. M., Kati, W., Stewart, K., Lal, R., Hsu, A., Betebenner, D., Korneyeva, M., Vasavanonda, S., McDonald, E., Saldivar, A., Wideburg, N., Chen, X., Niu, P., Park, C., Jayanti, V., Grabowski, B., Granneman, G. R., Sun, E., Japour, A. J., Leonard, J. M., Plattner, J. J., and Norbeck, D. W. (1998). ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrobial Agents and Chemotherapy* **42**(12)**,** 3218-24.

Sheng, N., Pettit, S. C., Tritch, R. J., Ozturk, D. H., Rayner, M. M., Swanstrom, R., and 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**(8)**,** 5723-32.

Tözsér, J., Blaha, I., Copeland, T. D., Wondrak, E. M., and Oroszlan, S. (1991a). Comparison of the HIV- and HIV-2 proteinases using oligopeptide substrate representing cleavage sites in Gag and Gag-Pol polyproteins. *FEBS Lett* **281**(1-2)**,** 77-80.

Tözsér, J., Gustchina, A., Weber, I. T., Blaha, I., Wondrak, E. M., and Oroszlan, S. (1991b). Studies on the role of the S4 substrate binding site of HIV proteinases. *FEBS Lett* **279,** 356-360.

Turner, S. R., Strohbach, J. W., Tommasi, R. A., Aristoff, P. A., Johnson, P. D., Skulnick, H. I., Dolak, L. A., Seest, E. P., Tomich, P. K., Bohanon, M. J., Horng, M. M., Lynn, J. C., Chong, K. T., Hinshaw, R. R., Watenpaugh, K. D., Janakiraman, M. N., and Thaisrivongs, S. (1998). Tipranavir (PNU-140690): a potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *Journal of Medicinal Chemistry* **41**(18)**,** 3467-76.

Ugolini, S., Moulard, M., Mondor, I., Barois, N., Demandolx, D., Hoxie, J., Brelot, A., Alizon, M., Davoust, J., and Sattentau, Q. J. (1997). HIV-1 gp120 induces an association between CD4 and the chemokine receptor CXCR4. *Journal of Immunology* **159**(6)**,** 3000-8.

Vacca, J. P., Dorsey, B. D., Schleif, W. A., Levin, R. B., McDaniel, S. L., Darke, P. L., Zugay, J., Quintero, J. C., Blahy, O. M., Roth, E., and et al. (1994). L-735,524: an orally bioavailable human immunodeficiency virus type 1 protease inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **91**(9)**,** 4096-100.

Weber, I. T., Miller, M., Jaskolski, M., Leis, J., Skalka, A. M., and Wlodawer, A. (1989). Molecular modeling of the HIV-1 protease and its substrate binding site. *Science* **243**(4893)**,** 928-31.

Wiegers, K., Rutter, G., Kottler, H., Tessmer, U., Hohenberg, H., and Kräusslich, H.-G. (1998). Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual gag poly protein cleavage sites. *J Virol* **72**(4)**,** 2846-2854.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. H. (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **245,** 616-621.

Wyatt, R., and Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **280**(5371)**,** 1884-8.

Yarchoan, R., Mitsuya, H., Matsushita, S., and Broder, S. (1985). Implications of the discovery of HTLV-III for the treatment of AIDS. *Cancer Research* **45**(9 Suppl)**,** 4685s-4688s.

Zennou, V., Mammano, F., Paulous, S., Mathez, D., and 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. *Journal of Virology* **72**(4)**,** 3300-6.

Zhang, S., Feng, Y., Narayan, O., and Zhao, L. J. (2001). Cytoplasmic retention of HIV-1 regulatory protein Vpr by protein-protein interaction with a novel human cytoplasmic protein VprBP. *Gene* **263**(1-2)**,** 131-40.

Zhang, Y. M., Imamichi, H., Imamichi, T., Lane, H. C., Falloon, J., Vasudevachari, M. B., and Salzman, N. P. (1997). Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol* **71,** 6662-6670.

Zolopa, A. R., Shafer, R. W., Warford, A., Montoya, J. G., Hsu, P., Katzenstein, D., Merigan, T. C., and Efron, B. (1999). HIV-1 genotypic resistance patterns predict response to saquinavir-ritonavir therapy in patients in whom previous protease inhibitor therapy had failed. *Annals of Internal Medicine* **131**(11)**,** 813-21.

CHAPTER II

CO-EVOLUTION OF NELFINAVIR-RESISTANT HIV-1 PROTEASE AND THE p1–p6 SUBSTRATE

The work presented in the following chapter was a collaborative effort. The viral RNA was extracted and sequenced by Stephane Lastere, I designed and performed the analyses and Celia Schiffer and I wrote the paper.

ABSTRACT

The selective pressure of the competitive protease inhibitors causes both HIV-1 protease and occasionally its substrates to evolve drug resistance. We hypothesize that this occurs particularly in substrates that protrude beyond the substrate envelope and contact residues that mutate in response to a particular protease inhibitor. To validate this hypothesis we analyzed substrate and protease sequences for covariation. Using the Chi-Square test we show a positive correlation between the nelfinavir-resistant D30N/N88D protease mutations and mutations at the p1-p6 cleavage site as compared to the other cleavage sites. Both nelfinavir and the substrate p1-p6 protrude beyond the substrate envelope and contact residue 30, thus possibly making the p1-p6 cleavage site more vulnerable to co-evolution.

INTRODUCTION

HIV-1 protease inhibitors bind competitively to the active site of the enzyme (Wlodawer and Erickson, 1993). Mutations in the protease that alter inhibitor binding and cause drug resistance can also affect substrate recognition by changing the enzyme's substrate specificity. To compensate, the virus will be under selective pressure to co-evolve the substrate sequence, thereby allowing the protease to retain activity (Bally et al., 2000; Doyon et al., 1996; Mammano et al, 1998; Feher et al., 2000). Earlier studies from our laboratory have shown that substrate specificity of the protease is based on the shape adopted by the substrate sequences, defined as "the substrate envelope" (Prabu-Jeyabalan et al, 2002). Most primary active-site mutations occur outside the substrate envelope and thereby preferentially impact inhibitor binding over substrate recognition. Therefore most of the substrates do not co-evolve with the protease. However, some substrates protrude beyond the envelope and we observe that they are the ones that co-evolve with the protease.

In this study, we focus on the D30N and N88D protease mutations, which are a signature of nelfinavir (NFV) resistance (Pai and Nahata, 1999; Patick et al., 1998) and their correlation with mutations at the p1-p6 substrate cleavage site. In the crystal structure of wild-type (WT) HIV-1 protease in complex with NFV, Asp30 forms a hydrogen bond with the m-phenol group of NFV (Kaldor et al., 1997) and superposition of NFV on the substrate envelope shows that this group protrudes from the envelope (Fig II.1a), suggesting that mutation at this residue will preferentially impact inhibitor binding over

Figure II.1

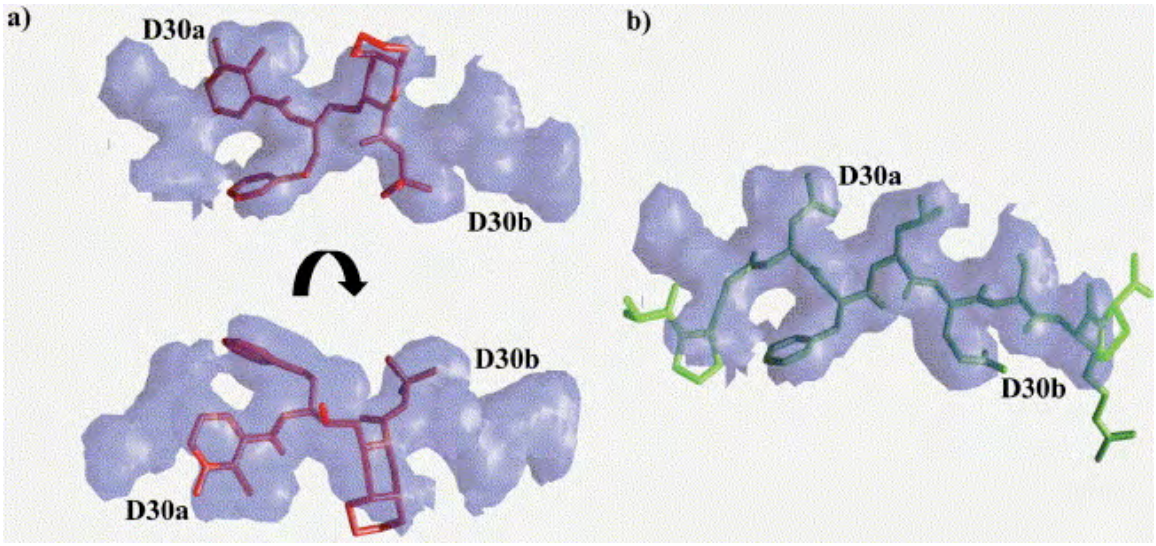


Figure II.1: NFV and p1-p6 both contact D30 outside the substrate envelope. a) Shown in top and bottom views is NFV in red within the substrate envelope in blue. The m-phenol group of NFV protrudes to contact D30 of monomer a (D30a) outside the substrate envelope b) p1-p6 in green within the substrate envelope. GlnP2′ protrudes from the envelope and interacts with D30 of monomer b (D30b).

substrate recognition. The D30N mutation of aspartic acid to an asparagine in the protease likely results in a weaker hydrogen bond that destabilizes NFV binding. In this study we present an analysis of Gag and Pol sequences from viral isolates of patients treated with protease inhibitors that reveals correlations between the D30N/N88D protease mutations and mutations within the corresponding p1-p6 cleavage site, which also makes direct ionic interactions with Asp30.

SEQUENCE ANALYSIS

Viral sequences (N=196) extending from Gag aa 346 to Pol aa 162 were obtained from patients who were part of the NARVAL trial (Meynard et al., 2002). These patients, treated with highly active antiretroviral therapy, had received a median of 3 prescribed protease inhibitors, for an average of 32 months. Plasma viral RNA was extracted and nested RT-PCR was performed to amplify the gag-protease region, including p2, p7, p1, p6 and the whole PR. Direct dideoxynucleoside terminator cycle sequencing of the PCR product was performed and sequencing products were analyzed on an ABI 3100 (Perkin Elmer ABI, Foster City, Calif., USA) instrument, and manually proof read and edited using Sequence Navigator software, in both 5′ and 3′ directions. Polymorphisms were defined as differences in amino acid usage with respect to the HIV-1 B subtype consensus sequence (Los Alamos HIV database, http://hiv-web.lanl.gov). When a mixture of wild-type and mutant residues was detected, the corresponding codon was classified as mutated. Over 90% of these isolates were subtype B. The viral sequences,

which included the protease and 10 substrate cleavage sites, had a median of 6 protease mutations per sequence.

The entire data set was read and entered into a Microsoft Access Database using software written in Visual Basic. Our analysis focused on qualitative changes in cleavage-site sequences, as compared to the consensus HXB2 sequence, that covary with the D30N/N88D drug-resistant substitutions in the protease. For each peptide sequence with the D30N/N88D substitutions in the enzyme, the corresponding substrate cleavage sites were scored for the presence or absence of mutations. Conversely, for every substrate site that scored positive for amino acid substitutions, changes at positions 30 and 88 in the enzyme were scored and tabulated.

The simultaneous occurrence of cleavage-site mutations and the D30N/N88D protease mutations were compared for each of the seven substrates studied (Fig II.2). The percentage of mutated substrate sequences that occurred with the D30N/N88D mutations in the corresponding protease sequence was calculated by dividing the number of mutated cleavage-site sequences by the total number of protease sequences with D30N/N88D mutations (N=21). The percentage of mutated substrate sequences that did not occur with corresponding D30N/N88D protease mutations was calculated by dividing the number of mutated cleavage-site mutations by the total number of protease sequences without D30N/N88D mutations (N=175). The Chi-Square test of significance (after applying the Yates correction) was performed to test for correlation between cleavage-site mutations and the D30N/N88D protease mutations (Table II.1). Statistically significant correlations

Figure II.2

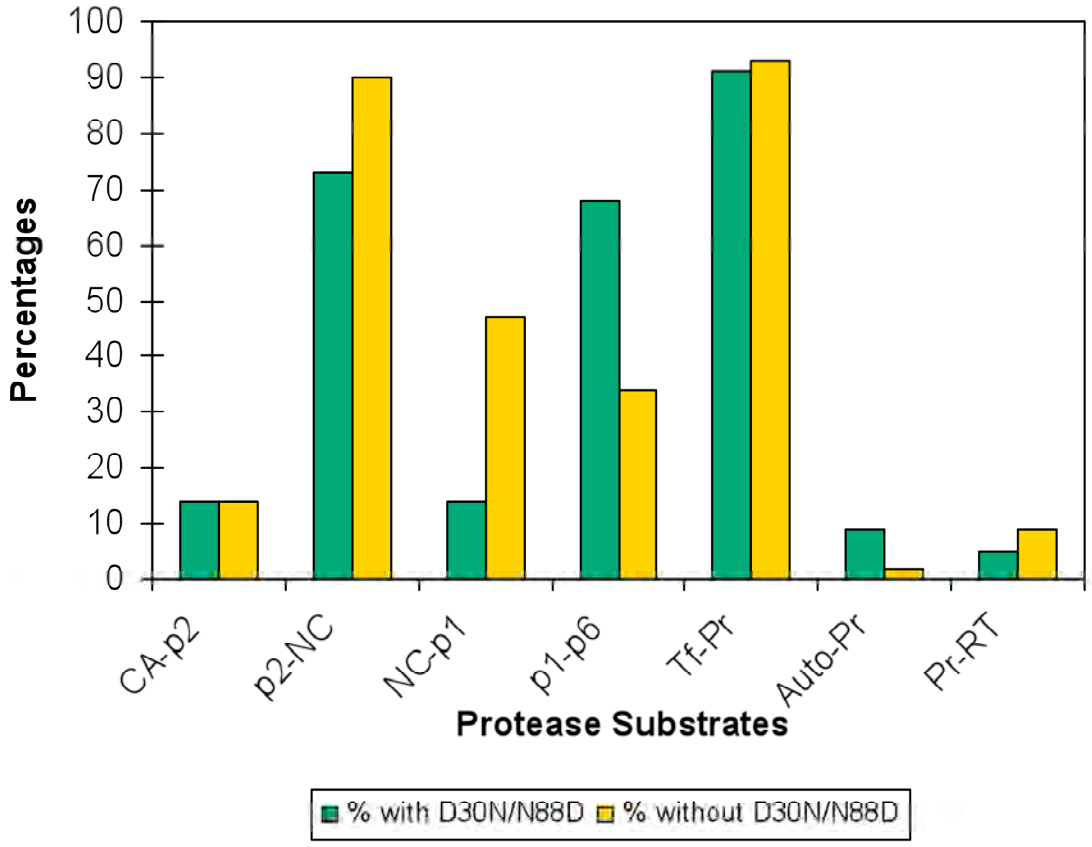

100
90
80
70
60
50
40
30
20
10
0
Percentages
CA-p2
p2-NC
NC-p1
p1-p6
Tf-Pr
Auto-Pr
Pr-RT
Protease Substrates
% with D30N/N88D
% without D30N/N88D

Figure II.2: Mutation Rates at Substrate Cleavage Sites with and without Corresponding D30N/N88D Protease Mutations. Green bars indicate the simultaneous occurrence of cleavage-site mutations and the D30N/N88D protease mutations. Yellow bars represent percentages of isolates in which mutations occur in the cleavage sites in the absence of the D30N/N88D protease mutations.

Table II.1

Substrate Cleavage Site	Number of Mutated Cleavage-Site Sequences[a]	χ^2	*P* value	Phi correlation coefficient
CA-p2	27	0.07	0.79	+0.01
p2-NC	175	2.13	0.14	-0.13
NC-p1	86	6.83	0.01	-0.2
p1-p6	74	10.18	0.001	+0.24
Tf-Pr	184	0.01	0.92	+0.03
Auto-Pr	5	0.16	0.16	+0.15
Pr-RT	16	0.03	0.86	-0.04

Table II.1: Covariation Between Substrate Mutations and D30N/N88D Protease Mutations

Total number of viral sequences isolated from patients = 196

Total number of viral sequences with D30N/N88D protease mutations = 21

Total number of viral sequences without D30N/N88D protease mutations = 175

were those that had p values ≤ 0.05. In addition the phi correlation coefficient was calculated to determine positive or negative correlations.

Substitutions were observed in all 7 substrate sites studied. Initial analysis of the sequences indicated that some substrate cleavage sites are more prone to mutations than others. Among the sequences analyzed, the mutation rates of p2-NC (90.8%) and TF-PR (90.8%) were much higher than those of the CA-p2 (6.8%) and AutoPr (3.4%) cleavage sites, which were almost immutable. We then examined all substrate cleavage sites and the corresponding protease sequences for mutations and covariations.

p1–p6 CLEAVAGE SITE

D30N/N88D and p1-p6 substrate site mutations were observed to be significantly correlated (X^2=10.18, p=0.0014) and the phi value indicates a positive correlation as shown in Table II.1. The majority of isolates with the D30N/N88D protease mutations (71%) also had mutations in the p1-p6 cleavage site. Conversely, only 33% of the corresponding p1-p6 cleavage sites had mutated in the absence of D30N/N88D protease mutations (Fig II.2). There were ten protease sequences that had only D30N, and only two had mutations in the corresponding p1-p6 cleavage site. Most likely, these viruses did not have sufficient time or selective pressure to evolve either the N88D protease mutation or the p1-p6 mutations. Thus, the combined D30N/N88D substitutions in the protease covary with mutations in the p1-p6 substrate.

To examine the possible effects of the D30N mutation on protease recognition and cleavage of p1-p6, structural modeling studies were performed. Asp30 in the active site of WT HIV-1 protease plays an important role in recognizing and binding this particular substrate cleavage site. Asp30 is one of four protease residues that form direct side-chain hydrogen bonds with substrate peptides. Asp30 forms one interaction with substrates sequences MA-CA, CA-p2 and p2-NC but forms three interactions with p1-p6 (Prabu-Jeyabalan et al, 2002). Specifically, Asp30′ OD2 makes a hydrogen bond with GlnP2′ NE2 and ArgP4′ NE of the p1-p6 substrate peptide and Asp30′ OD1 makes a salt bridge with ArgP4′ NE (Fig II.3a). The interactions of p1-p6 with Asp30 are located where this substrate protrudes from the substrate envelope (Fig II.1b) indicating a region in this particular substrate that could be affected by a resistant mutation and thereby propagates co-evolution. Under the selective pressure of NFV, Asp30 mutates to an Asn. Modeling in the D30N substitution suggests that some of these interactions may be altered (Fig II.3b). The salt bridge between Asp30′ in the WT protease and ArgP4′ in p1-p6 may be weakened to a hydrogen bond with Asn, however, in the other substrates that the interactions with Asp30 are unlikely to be modified with the D30N substitution. It is therefore likely that Asp30′ is more crucial to the recognition of the p1-p6 substrate site than to other cleavage sites. Therefore, the D30N protease substitution is more likely to select for compensatory mutations within the p1-p6 substrate cleavage site than with other substrate cleavage sites.

The above observations led us to examine the specific substitutions within the p1-p6 cleavage site. In Table II.2 we present a matrix of the mutational pattern to examine

Figure II.3

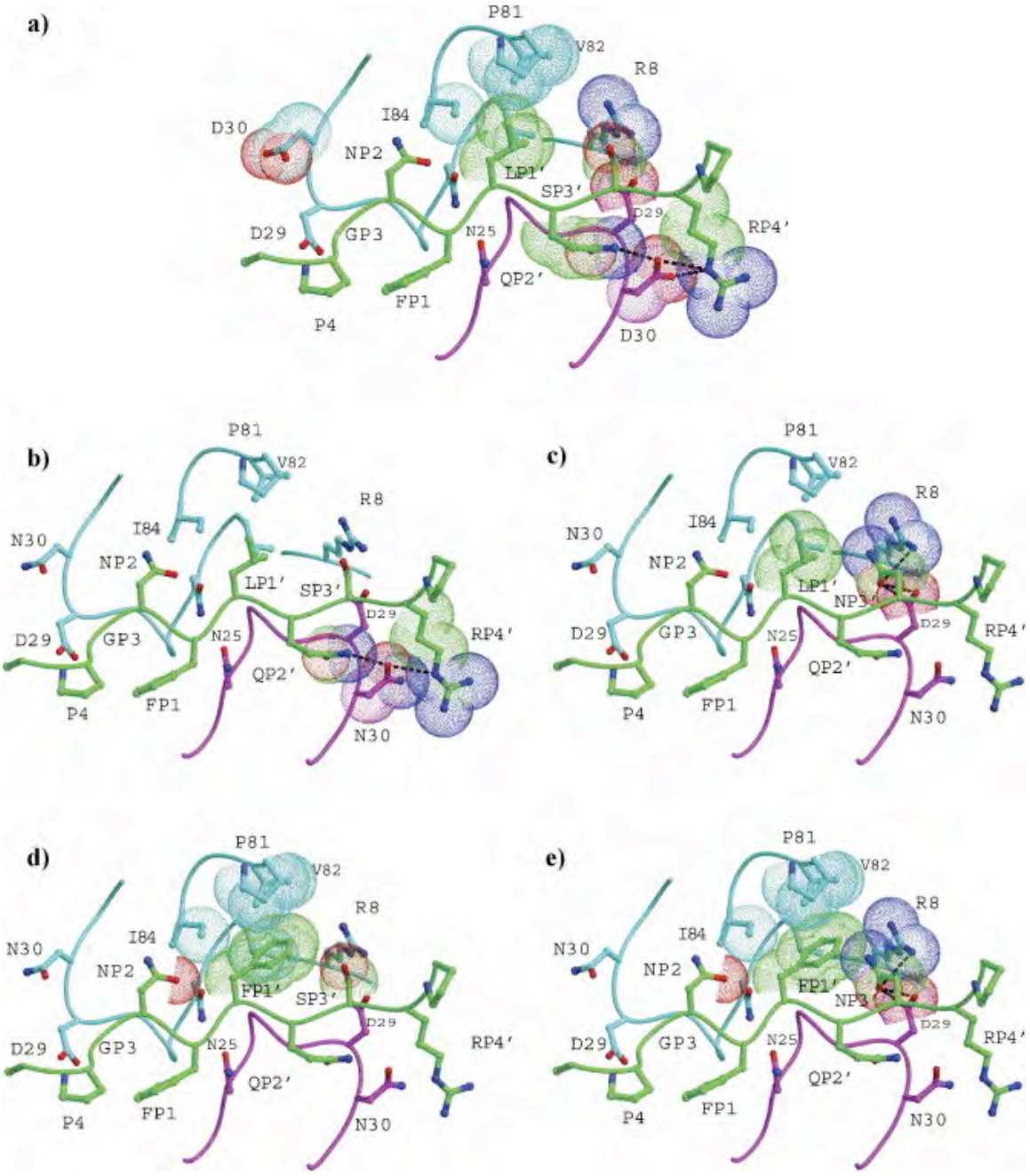

a)
P81
V82
R8
I84
D30
NP2
LP1'
SP3'
D29
RP4'
D29
GP3
N25
QP2'
FP1
P4
D30
b)
P81
V82
R8
N30
I84
NP2
LP1'
SP3'
D29
RP4'
D29
GP3
N25
QP2'
P4
FP1
N30
c)
P81
V82
R8
N30
I84
NP2
LP1'
NP3'
D29
RP4'
D29
GP3
N25
QP2'
P4
FP1
N30
d)
P81
V82
R8
N30
I84
NP2
FP1'
SP3'
D29
RP4'
D29
GP3
N25
QP2'
P4
FP1
N30
e)
P81
V82
R8
N30
I84
NP2
FP1'
NP3'
D29
RP4'
D29
GP3
N25
QP2'
P4
FP1
N30

Figure II.3: Modeling Studies of Complexes of HIV-1 Protease Variants and the p1-p6 Substrate. The two protease monomers are shown in cyan and magenta and the p1-p6 peptide in green, in red is oxygen and in blue is nitrogen. Hydrogen bonds are shown as dashed black lines. The figures were made with the graphics progam MIDAS (Ferrin et al., 1988). a) Crystal structure of HIV (D25N) protease complexed with the p1-p6 substrate peptide (Prabu-Jeyabalan et al, 2002). van der Waals radii are shown for residues 8a, 29a-b, 30a-b, 81a, 82a, 84a, P1′, P2′, P3′ and P4′. Hydrogen bonds between D30b and GlnP2′ and ArgP4′ are shown. b) Probable contacts and hydrogen bonds made by N30b are shown. c) Likely contacts made by AsnP3′ (wt SerP3′) with van der Waals radii shown for residues around it. Also shown are the probable hydrogen bonds d) Probable contacts made by PheP1′ (wt LeuP1′) with van der Waals radii of the surrounding residues. e) van der Waals radii are shown for contacts that are likely made when both P1′ and P3′ are mutated.

Table II.2:

Amino Acid	P	G	N	F	L	Q	S	R
Substrate Subsite	**P4**	**P3**	**P2**	**P1**	**P1′**	**P2′**	**P3′**	**P4′**
P4								
P3								
P2			**3%(D,I)**					
P1								
P1′					**47%(F)**			
P2′						**1%**		
P3′					**5%**		**34%(N,T)**	
P4′					**1%**		**3%**	**5%(S)**

Table II.2: Frequency of Mutations in the p1-p6 Substrate Cleavage Site.

whether multiple sites within the substrate mutate simultaneously. Almost no mutations occur on the unprimed side of the substrate. Interestingly also, GlnP2′ and ArgP4′, each of which interacts with Asp30, mutate very infrequently. However, high variability was observed at the P1′ and P3′ positions in the p1-p6 substrate, though only either P1′ mutates from Leu to the bulkier Phe, or SerP3′ mutates to an Asn, both mutations do not occur simultaneously. Thus the pattern of mutations within the p1-p6 has some complexity.

To further understand the covariation of p1-p6 mutations and the D30N/N88D protease mutations, we performed modeling studies from the crystal structure of with the WT protease bound to the p1-p6 substrate. Phe at P1′ fits the S1′ protease pocket much better than Leu at P1′ (WT) due to more extensive van der Waals contacts (Fig II.3d and II.3a). Similarly, modeling an Asn at the P3′ position may possibly result in interactions with Arg8 and Asp29′ not seen with Ser (Fig II.3c and II.3a). Either mutation may compensate for the potentially weaker interactions at P2′ and P4′ with Asp30. However, simultaneously modeling both of the larger mutations at P1′ and P3′ in the S1′/S3′ pocket of the protease (Fig II.3e) may lead to steric clashes that could explain why the two residues at P1′ and P3′ rarely mutate together.

NC-p1 CLEAVAGE SITE

Mutations at NC-p1 also covaried with the D30N/N88D mutations in the protease (χ^2=6.83, p=0.009). In contrast with the p1-p6 site, mutations within the NC-p1 cleavage site were negatively correlated to the protease D30N/N88D mutations as indicated by a negative phi value (Table II.1). Of the 21 isolates with the D30N/N88D mutations, only 14% had mutations in the NC-p1 cleavage site (Table II.1). Conversely, of the 175 isolates without the D30N/N88D protease mutations, 47% had NC-p1 mutations, 86 of which were AP1V that mostly correlated with the protease mutation V82A. This correlation has been previously observed (Doyon et al., 1996) and explained structurally as the NC-p1 cleavage site makes unique interactions with Val82 (Prabu-Jeyabalan et al., 2004). The protease mutations V82A and D30N do not occur simultaneously in patient sequences (Hoffman et al, 2003; Wu et al., 2003) presumably because the two mutations occurring together would adversely impact protease activity. Since V82A mutates in a correlated manner with NC-p1 but not with D30N/N88D, it is consistent that mutations at NC-p1 would also be anti-correlated with D30N/N88D.

CONCLUSION

Mutations in the HIV-1 protease substrate cleavage-site p1-p6 covary with the D30N/N88D protease mutations. Asp30 is important both to the binding of NFV and also likely to the recognition of the p1-p6 cleavage site. Structural analysis shows that both NFV and p1-p6 have atoms that protrude beyond the substrate envelope and contact

Asp30. Thus, both the inhibitor and the p1-p6 substrate are likely to be affected by D30N mutation. This likely explains the particular co-evolution of the p1-p6 cleavage site with the D30N resistant mutation and also why no other co-evolution with any of the other substrates occurs.

Acknowledgements: We are grateful to Pr. Brun-Vezinet and other members of the scientific committee of the NARVAL trial (NARVAL ANRS 088) for the sequences. We would also like to thank Sagar Kathuria for help with sequence alignments and for many stimulating discussions and Claire Baldwin for editorial assistance. This research was supported by the National Institutes of Health R01GM65347.

References

Bally, F., Martinez, R., Peters, S., Sudre, P., and Telenti, A. (2000). Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: Clinical significance and implications for resistance to protease. *AIDS Res Hum Retrovir* **16**(13)**,** 1209-1213.

Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L., and Lamarre, D. (1996). Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol* **70**(6)**,** 3763-9.

Feher, A., Weber, I.T., Bagossi, P., Baross, P., Mahalingam, B., Louis, J.M., Copeland, T.D., Yorshin, I.Y., Arrison, R.W., Tozser, J. (2002) Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. *J. Biochem*, **269**, 4114-4120.

Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988). The MIDAS display system. *J Mol Graph* **6,** 13-27.

Hoffman, N., Schiffer, C. A., and Swanstrom, R. (2003). Covariation of amino acid positions in HIV-1 protease. *Virology* **314,** 536-548.

Kaldor, S. W., Kalish, V. J., Davies, J. n., Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., Dressman, B. A., Hatch, S. D., Khalil, D. A., Kosa, M. B., Lubbehusen, P. P., Muesing, M. A., Patick, A. K., Reich, S. H., Su, K. S., and Tatlock, J. H. (1997). Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J Med Chem* **40**(24)**,** 3979-3985.

Mammano, F., Petit, C., and Clavel, F. (1998). Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol* **72**(9)**,** 7632-7.

Meynard, J. L., Vray, M., Morand-Joubert, L., Race, E., Descamps, D., Peytavin, G., Matheron, S., Lamotte, C., Guiramand, S., Costagliola D, Brun-Vezinet, F., Clavel, F., Girard, P. M., and Group., N. T. (2002). Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. *AIDS* **16**(5)**,** 727-36.

Pai, V. B., and Nahata, M. C. A. P., 33: 325-339 (1999). Nelfinavir mesylate: a protease inhibitor. *Ann Pharmacotherapy* **33**(3)**,** 325-39.

Patick, A., Duran, M., Cao, Y., Shugarts, D., Keller, M., Mazabel, E., Knowles, M., Chapman, S., Kuritzkes, D., and Markowitz, M. (1998). Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob Agents Chemother* **42**(10)**,** 2637-44.

Prabu-Jeyabalan, M., King, N. M., Nalivaika, E., and Schiffer, C. A. (2004). Co-evolution of the Nucleocapsid-p1 cleavage site with the V82A mutation in HIV-1 protease preserves substrate recognition. *J Virol* **78**(22)**,** 12446-54.

Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2002). Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* **10**(3)**,** 369-381.

Wlodawer, A., and Erickson, J. W. (1993). Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem* **62,** 543-585.

Wu, T. D., Schiffer, C. A., Gonzales, M., R., K., Chou, S., Israelski, D., Zolopa, A., Fessel, J., and Shafer, R. W. (2003). Mutation patterns and structural correlates in HIV-1 protease following varying degrees of protease inhibitor treatment. *J Virol* **77,** 4836-4847.

CHAPTER III

HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 PROTEASE-CORRELATED CLEAVAGE SITE MUTATIONS ENHANCE INHIBITOR RESISTANCE

The work presented in the following chapter was a collaborative effort. The susceptibility data was acquired from Monogram Biosciences, Inc. Eric Stawiski queried the database and ran the statistical analysis software, Celia Schiffer and I conceived, designed and performed the analyses, Eric Stawiski and I made the figures, Celia Schiffer and I wrote the paper, with inputs from Eric Stawiski and Colombe Chappey.

ABSTRACT

Drug resistance is an important cause of anti-retroviral therapy failure in HIV-infected patients. Mutations in the protease render the virus resistant to protease inhibitors (PIs). Gag cleavage sites also mutate, sometimes correlating with resistance mutations in the protease, but their contribution to resistance has not been systematically analyzed. The present study examines mutations in Gag cleavage sites that associate with protease mutations and the impact of these associations on drug susceptibilities. Significant associations were observed between mutations in the NC-p1 and p1-p6 cleavage sites and various PI resistance-associated mutations in protease. Several patterns were frequently observed, including: mutations in the NC-p1 cleavage site in combination with I50L, V82A and I84V within the protease; and mutations within the p1-p6 cleavage site in combination with D30N, I50V and I84V within the protease. For most patterns, viruses with mutations both in the protease and in either cleavage site were significantly less susceptible to specific PIs than viruses with mutations in the protease alone. Altered PI resistance in HIV-1 was found to be associated with the presence of Gag cleavage site mutations. These studies suggest that associated cleavage site mutations may contribute to PI susceptibility in highly specific ways depending on the particular combinations of mutations and inhibitors. Thus, cleavage site mutations should be considered when assessing the level of PI resistance.

INTRODUCTION

Combination antiretroviral treatment regimens containing PIs often do not completely and durably suppress HIV-1 infection, resulting in therapy failure. Selective drug pressure of competitive PIs has resulted in accumulation of combinations of mutations within both Gag and GagProPol polyproteins that render the virus highly resistant to multiple drugs (Boden and Markowitz, 1998; Molla et al., 1998; Schinazi, Larder, and Mellors, 1997). In addition, there is an increasing incidence of new infections with viruses that are already drug-resistant (Brenner et al., 2002).

Drug resistance provides an escape mechanism for the virus by lowering the affinity of target enzymes for inhibitors while still maintaining efficient processing, thus leading to therapy failure (Kantor et al., 2002; Rhee et al., 2003). This paradox maybe explained using the concept of 'substrate envelope' which states that substrate specificity is based not on a particular amino acid sequence but on a conserved shape (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002). The consensus volume of the majority of the substrates defines the substrate envelope. Although most of the volume of any given inhibitor also fits within this envelope, particular inhibitor atoms occasionally protrude beyond the envelope to contact surrounding protease residues. Mutations at these residues preferentially impact inhibitor binding over substrate recognition and cleavage (King et al., 2004a). Often, these resistance-associated mutations reduce the catalytic efficiency of the protease resulting in immature or non-infectious viruses (Croteau et al., 1997). To compensate for this loss in efficiency, secondary mutations develop within the protease

(Nijhuis et al., 1999). In addition, mutations develop within the cleavage sites complementing the changes in the resistant protease. All genotype-based algorithms used to predict the susceptibility of patient virus to PIs only assess mutations within the protease (Gulnik et al., 1995; Kaplan et al., 1994; Nijhuis et al., 1999). However, resistance-associated mutations in Gag, especially within the cleavage sites, are often associated with protease mutations (Bally et al., 2000; Doyon et al., 1996; Feher et al., 2002; Kolli, Lastere, and Schiffer, 2006; Mammano, Petit, and Clavel, 1998; Zhang et al., 1997) and may be important for resistance testing.

Gag A431V, within the nucleocapsid-p1 (NC-p1) cleavage site, was one of the first cleavage site mutations to be associated with the V82A protease mutation (Zhang et al., 1997), and more recently I437V was shown also to be associated with 82 (Nijhuis et al., 2007). Association was reported between mutations within the p1-p6 cleavage site and the I50V protease mutation (Maguire et al., 2002). In earlier studies (Kolli, Lastere, and Schiffer, 2006) we showed an association between the occurrence of mutations within the p1-p6 cleavage site and the signature protease mutations of Nelfinavir (NFV) resistance, D30N/N88D (Pai and Nahata, 1999; Patick et al., 1998). The present study focuses on the primary resistance-associated mutations at residues 30, 50, 82, 84, 88 and 90 in the protease, and their association with mutations within the Gag cleavage sites, NC-p1 and p1-p6 (Fig III.1a-e).

To further study the impact of these associations, the susceptibility of viruses to various PIs was examined by mining a large database of linked genotypes and phenotypic

Figure III.1

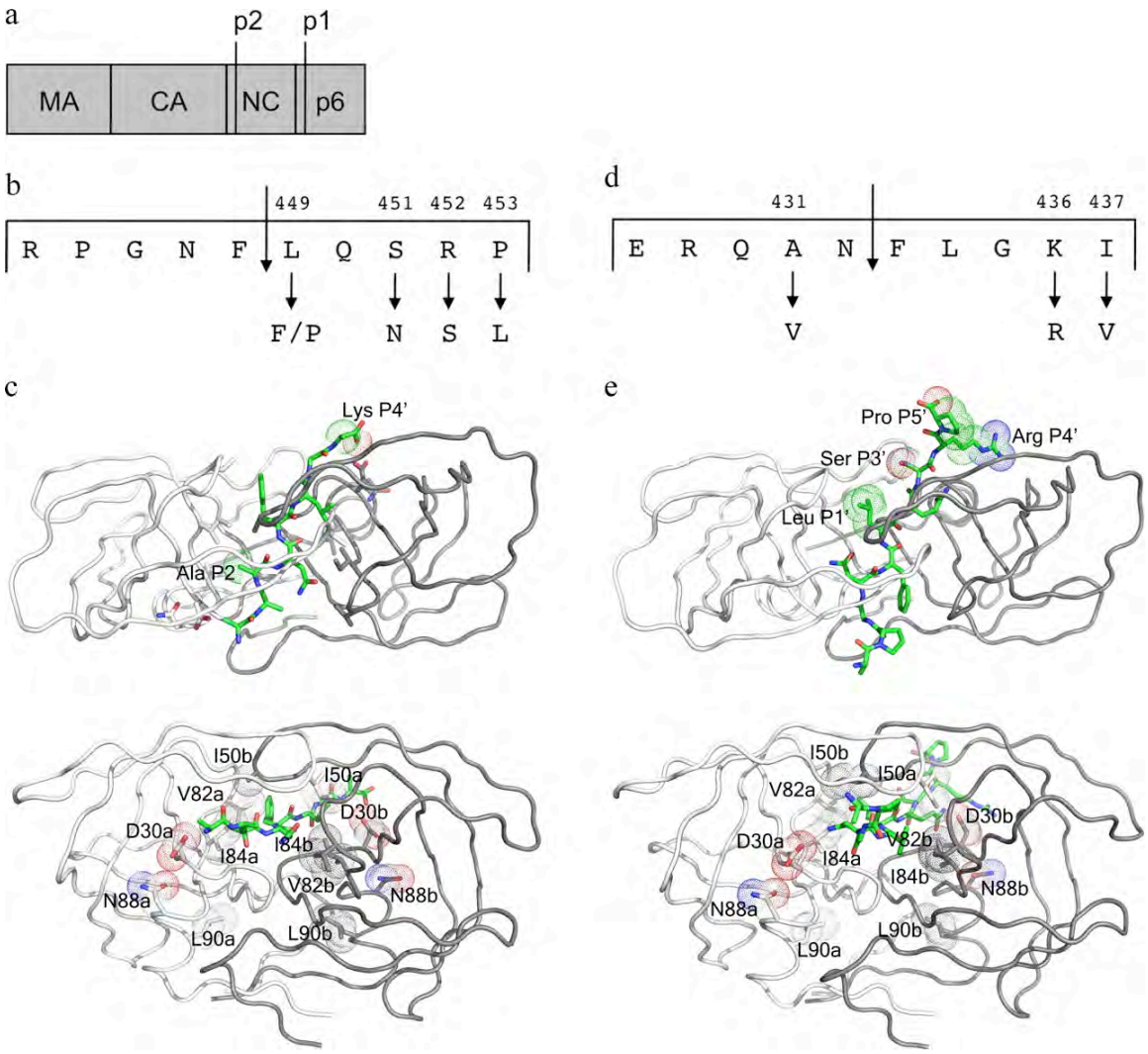

a
p2
p1
MA
CA
NC
p6
b
449
451 452 453
R P G N F L Q S R P
F/P N S L
d
431
436 437
E R Q A N F L G K I
V R V
c
Lys P4'
Ala P2
I50b
I50a
V82a
D30b
D30a
I84b
I84a
V82b
N88b
N88a
L90a
L90b
e
Pro P5'
Arg P4'
Ser P3'
Leu P1'
I50b
I50a
V82a
D30b
D30a
V82b
I84a
I84b
N88b
N88a
L90b
L90a

Figure III.1: Overview of the NC-p1 and p1-p6 Cleavage Sites. (a) Overview of Gag showing the constituent proteins. (b) Details of the p1-p6 cleavage site. Also shown are the two residues, AP2 and LP4′, which mutate frequently. (c) Two different views of the crystal structure of HIV-1 protease complexed with NC-p1 cleavage site (King et al., 2004b) (PDB ID:1TSU). In green is the substrate and in white and grey are the two monomers. Above is the top view showing van der Waals surfaces on cleavage site residues that mutate frequently. Shown below is the side view with van der Waals surfaces on protease residues that are the focus of this study. (d) Details of the NC-p1 cleavage site. Residues that mutate frequently are shown. (e) Two different views of the crystal structure of HIV-1 protease complexed with p1-p6 cleavage site (King et al., 2004b) (1KJF). In green is the substrate and in white and grey are the two monomers. Above is the top view showing van der Waals surfaces on cleavage site residues that mutate frequently. Shown below is the side view with van der Waals surfaces on protease residues that are the focus of this study. Images were generated using PyMol (DeLano, 2008).

measures of viral susceptibility to antiretroviral drugs. Several of these patterns significantly alter the virus susceptibility to PIs, indicating that the evolution of drug resistance not only occurs within the protease but also involves multiple regions of Gag and GagProPol polyproteins in an interdependent manner.

METHODS

Sequence Database. HIV-1 protease (PR, 1-99), reverse transcriptase (RT, 1-305) and Gag (Gag, 418-500) sequences were obtained from samples submitted to Monogram Biosciences for routine phenotype and genotype drug resistance testing. Viral sequences from infected patients were determined by the GeneSeq™ HIV assay (Petropoulos et al., 2000). A data set of 39,152 subtype B sequences from the Monogram Biosciences database was used. Wild type viruses i.e. showing none of the known resistance mutations in either protease or reverse transcriptase were excluded. RT mutations that were included are M41L, K65R, D67N, T69, K70ER, L74X, V75A/M/S/T, A98G, L100I, K101P, K103N/S, V106A/M, Y115F, Q151M, Y181X, M184X, Y188X, G190X, L210W, T215F/Y, K219X, P225X, F227X, M230L, P236L (X stands for any mutation). PR mutations that were included are L10X, K20X, L23X, L24X, D30X, V32X, L33X, M46X, I47X, G48X, I50X, I54X, A71X, G73X, L76X, V82X, I84X, N88X, L90X. Only one sample per patient was included. Gag sequence alignments were performed using the profile Hidden Markov Model (HMM) software HMMER (http://hmmer.janelia.org). The profile HMM was previously constructed from 1,152 gag sequences (data not shown). Amino acid positions were numbered through the alignment with the HXB2

consensus sequence. All statistical analyses were performed using the R software environment, version 2.4.1 (http://www.r-project.org).

Covariation Analysis. Sequences were analyzed with respect to presence or absence of the following primary mutations within the protease, either individually or in combination: D30N, I50V, I50L, V82A, I84V, N88D and L90M. Samples with mixtures at particular positions and mutations other than those specified were excluded. The corresponding NC-p1 and p1-p6 cleavage sites were screened for mutations relative to HXB2. The association of cleavage site mutations with a specific protease mutation was computed as the proportion of the total number of sequences with that particular protease mutation. These proportions were compared to the case when the cleavage site mutations occurred in the absence of any primary protease mutation. The significance of the association was tested by the Chi square test (R function chisq.test). Adjustments were made for multiple testing using the Bonferroni method. The corrected p-values < 0.05 indicated statistical significance.

Replication Capacities (RC). Relative fitness of the virus was determined with the Replication Capacity^{TM} and expressed as a percentage relative to that of a reference virus, NL4-3 (Deeks et al., 2001). RCs for viruses including the protease mutations of interest with and without Gag cleavage site mutations were extracted and analyzed. Difference in mean RC was tested using the Mann-Whitney test (R function wilcoxon.test). Adjustments were made for multiple testing using the Bonferroni method. The corrected p-values < 0.05 indicated statistical significance.

Drug Susceptibilities. Virus susceptibility to Amprenavir (APV), Atazanavir (ATV), Indinavir (IDV), Lopinavir (LPV), Nelfinavir (NFV), Ritonavir (RTV), Saquinavir (SQV) and Tipranavir (TPV) was determined with the PhenoSense™ assay (Parkin et al., 2004; Petropoulos et al., 2000) (Monogram Biosciences, South San Francisco CA). Results are expressed as fold change (http://www-fbsc.ncifcrf.gov/HIVdb) in IC50 of the patient virus relative to that of the drug-sensitive reference, NL4-3. FC distribution for groups of samples containing certain protease mutations with and without Gag cleavage site mutations were extracted and analyzed. Samples with mixtures at particular positions and mutations other than those specified were considered not to meet the profile and were excluded. The difference in median IC50 FC was tested using the Mann-Whitney test (R function wilcoxon.test). Adjustments were made for multiple testing using the Bonferroni method. The corrected p-values < 0.05 indicated statistical significance.

Additional Validation Studies. To help to ensure the validity of our studies the following additional studies were performed – 1) the average number of secondary protease mutations in viruses with or without particular Gag mutations was computed and the difference was evaluated using the t-test. 2) Sequences with and without Gag mutations in the absence of protease mutations were extracted from the database. The difference in mean FC in IC50s to various protease inhibitors was evaluated using the t-test. 3) Two sets of Gag sequences (418-500) were aligned. The first group consists of non-WT samples that were used for the susceptibility analyses and the second group consists of sequences that were WT for protease. The mutation frequency at each Gag

position was calculated for both sets and the frequency difference between the two groups was plotted.

RESULTS

The NC-p1 and p1-p6 cleavage sites (Fig III.1b, d) are essential in the critical final steps of viral budding (Sheng et al., 1997) and maturation. To examine the mutational patterns in these two key cleavage sites and their association with HIV-1 protease mutations, data was extracted from 39,152 HIV-1, subtype B viral sequences from infected patients. These sequences had a median of one PI-associated protease mutations (range 0-8). Sequences were examined for associated mutations between D30N, D30N/N88D(/L90M), I50V/L, V82A(/L90M), I84V(/L90M) and L90M primary drug-resistance mutations in protease and mutations within Gag in either the NC-p1 or the p1-p6 cleavage sites. Gag A431 within the NC-p1 cleavage site was the most common site for associated mutations. The A431V mutation was observed to associate with the V82A protease mutation, as previously reported (Doyon et al., 1996; Zhang et al., 1997). Mutations at Gag 431 also associated with either I50L or I84V protease mutations. Mutations at Gag 436, another site within NC-p1, associated with the V82A protease mutation. Gag I437V mutation within NC-p1 recently shown to enhance PI resistance (Nijhuis et al., 2007) was observed in this study to significantly associate with both V82A and I84V protease mutations. Previously, we observed that mutations at the p1-p6 cleavage site associate with the D30N/N88D NFV-resistance protease mutations (Kolli, Lastere, and Schiffer, 2006). In the present study p1-p6 mutations were also observed to

associate with N88D alone. In addition, p1-p6 cleavage site mutations associated with either I50V or I84V protease mutations, independently. The particular patterns of mutations within the p1-p6 cleavage site varied specifically depending on the particular protease mutation. D30N/N88D and I84V protease mutations associated with mutations at Gag L449, S451, R452 and P453 while I50V protease mutation associated with mutations at Gag L449, R452 and P453. Thus, associated mutations were observed in either the NC-p1 and/or the p1-p6 cleavage sites in the presence of all the primary active site resistance mutations.

These patterns of associations between cleavage site mutations and particular primary protease resistance mutations were further analyzed in terms of RC and drug susceptibility. The effects of additional protease mutations (L10, K20, L23, L24, V32, L33, M46, I47, G48, I54, A71, G73 and L76) have not been excluded as attempts to exclude them limited the sample size and did not significantly alter the observed associations. However, additional control studies, as detailed below, were performed to understand other factors that may likely contribute to alterations in phenotypic susceptibility to various inhibitors in addition to the role of Gag mutations.

Groups of viruses with particular protease mutations in the presence and absence of Gag mutations were compared for the number of secondary protease mutations (Supplemental Table III.1). With the exception of viruses with V82A/L90M all viruses with other protease mutations in combination with Gag A431V had significantly more secondary mutations in the protease. Of these, only D30N and I50V viruses did not show increased

FCs. Viruses with V82A and V82A/L90M in combination with mutation at Gag 449 had slightly more secondary protease mutations and showed higher FCs in IC50 compared to viruses without the Gag 449 mutation. Viruses with mutations at Gag 452 had increased number of secondary mutations when in combination with the I84V, I84V/L90M, V82A and V82A/L90M protease mutations and also showed a significant in FCs. A similar trend was observed in viruses that had mutations at Gag 453 in combination with V82A and V82A/L90M protease mutations.

Alterations in phenotypic susceptibilities to protease inhibitors as a result of various Gag cleavage site mutations in the absence of protease mutations were analyzed (Fig III.2). With the exception of IDV and NFV, there were small but statistically significant increases in susceptibilities to various protease inhibitors as a result of cleavage site mutations alone in the absence of any resistance-associated protease mutations. However, both groups of viruses were found to be highly susceptible to all PIs except to NFV where a slight reduction in susceptibility was observed.

Mutation frequencies within Gag were higher in the case of non-WT samples that were used in the phenotypic susceptibility analyses as compared to WT sequences (Fig III.3). Specifically, Gag 431 and 453, within the NC-p1 and p1-p6 sites respectively, mutated much more frequently in non-WT samples as compared to WT sequences. Increased mutational frequencies were also observed at 436, 437, 449 and 452 in the case of the non-WT samples as compared to WT samples. Small increases in mutation frequencies were also observed in the PTAP region in the non-WT samples.

Figure III.2

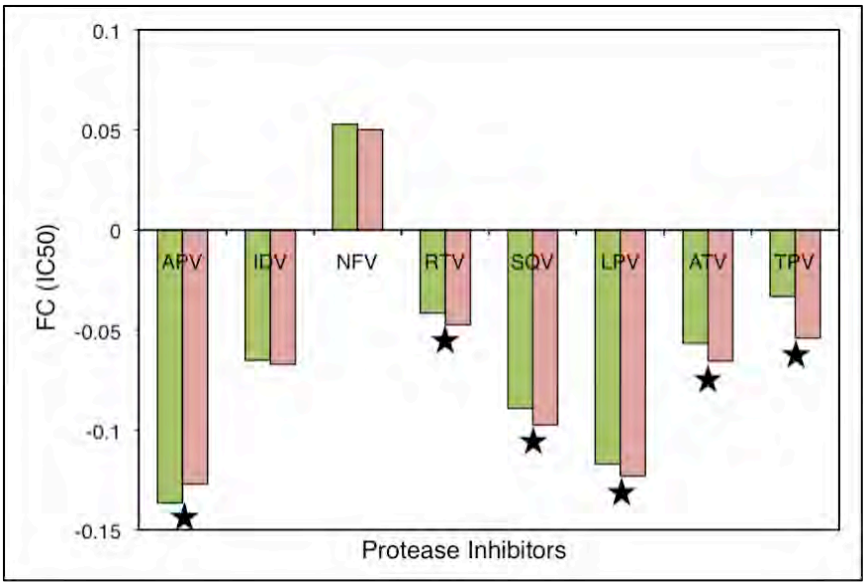

0.1
0.05
0
-0.05
-0.1
-0.15
FC (IC50)
APV
IDV
NFV
RTV
SQV
LPV
ATV
TPV
Protease Inhibitors

Figure III.2: PI susceptibility of viruses with and without Gag mutations in the absence of protease resistance mutations. Mean FCs in IC50 are shown on a log scale on the y-axis and the various PIs tested against are shown on the x-axis. Green bars represent viruses with Gag mutations and pink bars represent viruses without Gag mutations. Stars indicate significant differences in susceptibility of the viruses to a particular PI. p-values ≤ 0.05 are significant.

Figure III.3

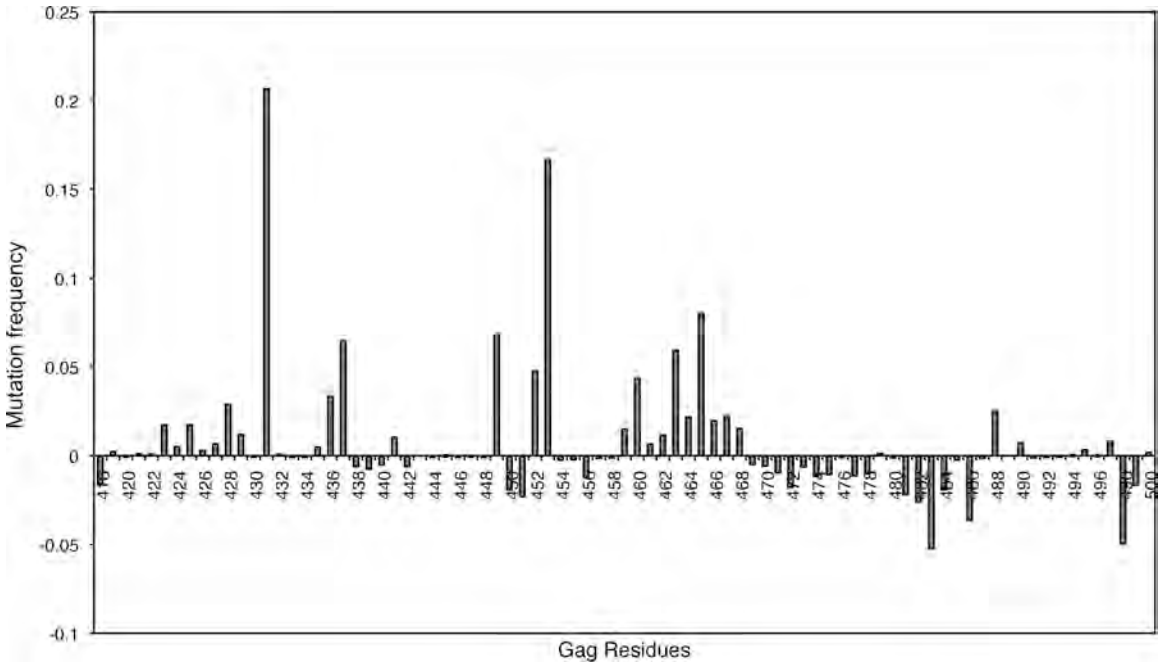

0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
Mutation frequency
420
422
424
426
428
430
432
434
436
438
440
442
444
446
448
452
454
458
460
462
464
466
468
470
474
476
478
480
488
490
492
494
496
500
Gag Residues

Figure III.3: Differential mutation rates within Gag from viruses with and without PR resistance mutations. Mutation frequencies are shown on the y-axis and Gag residues numbering 418 to 500 are shown on the x-axis.

Replication Capacities. In this study, fitness of groups of viruses with particular protease mutations in the presence and absence of Gag mutations were compared to a reference virus. Interestingly, in most instances, viruses with Gag mutations in combination with particular protease mutations did not improve fitness as compared to that of viruses with mutations in the protease alone (Fig III.4 and Supplemental Table III.2). In fact, in the few cases where significant differences were observed, Gag mutations actually decreased fitness as compared to viruses that have protease mutations in the absence of Gag mutations. Both V82A and V82A/L90M viruses had lower RCs in combination with mutations at Gag 431 as compared to those without this mutation (Fig III.4h and i). Similarly, mutation at Gag 431 also lowered viral fitness in the presence of I84V protease mutation (Fig III.4j). In addition, viruses with I84V or I84V/L90M protease mutations in combination with mutations at Gag 452 also had lower RCs as compared to viruses with the protease mutation alone (Fig III.4j and k). Viruses with L90M protease mutation in combination with mutations at either Gag 431, 437 and 453 were less fit as compared to viruses with L90M alone (Fig III.4l).

D30N. D30N is a signature mutation for NFV often in combination with N88D or occasionally N88D/L90M (Mitsuya et al., 2006; Patick et al., 1998). D30N, D30N/N88D and N88D protease mutations were not associated (Fig III.5a, b and d) with mutations in NC-p1. However, there was a significant decrease in phenotypic susceptibilities to all protease inhibitors in viruses where A431V mutation occurred in combination with D30N/N88D(/L90M), N88D or N88D/L90M (Fig III.6a, d and g). No corresponding changes in PI sensitivity were seen with mutations at Gag 436, whereas mutations at Gag

Figure III.4

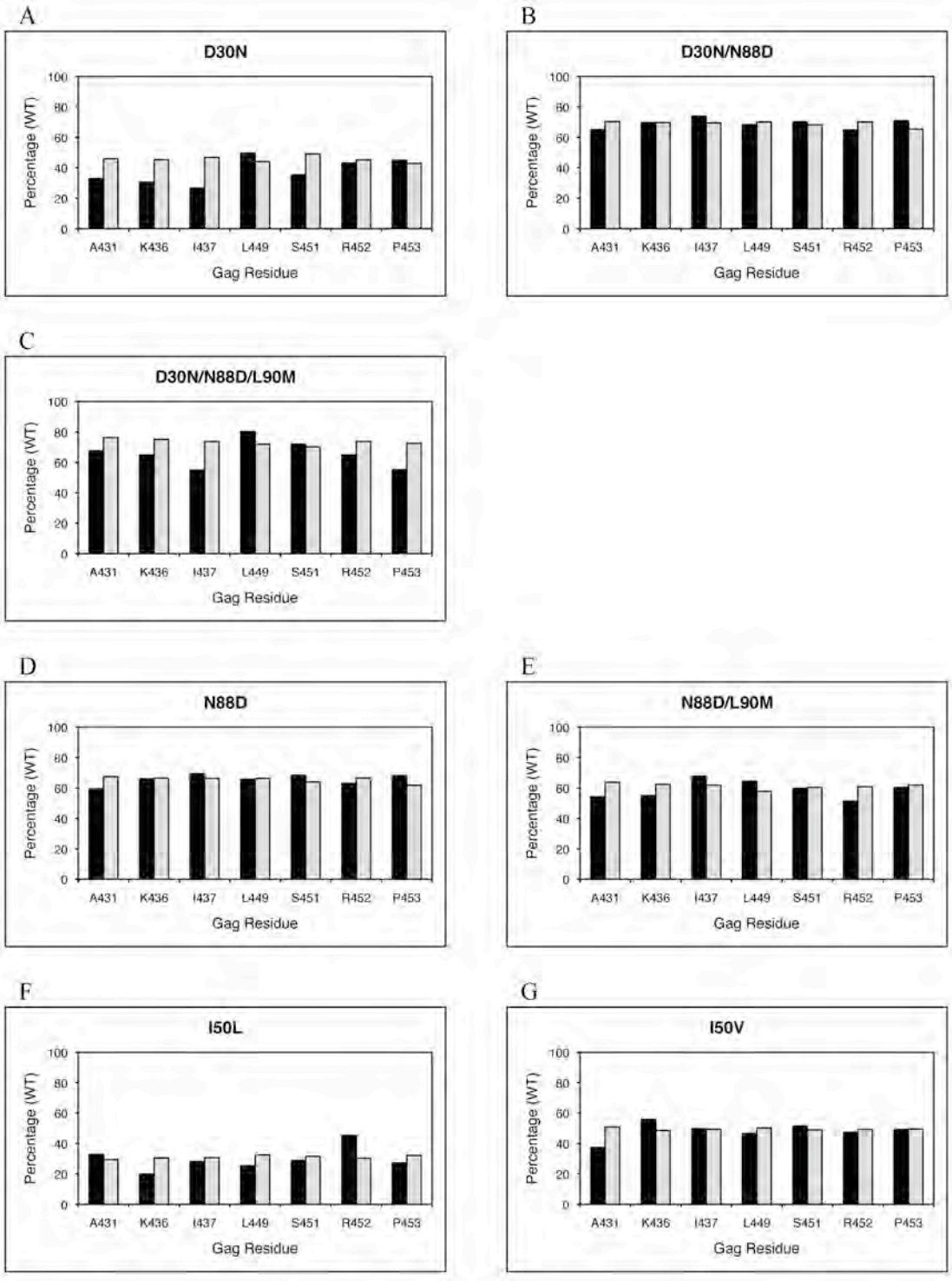

A
D30N
B
D30N/N88D
C
D30N/N88D/L90M
D
N88D
E
N88D/L90M
F
I50L
G
I50V
Percentage (WT)
0
20
40
60
80
100
A431
K436
I437
L449
S451
R452
P453
Gag Residue

Figure III.4 (*contd.*)

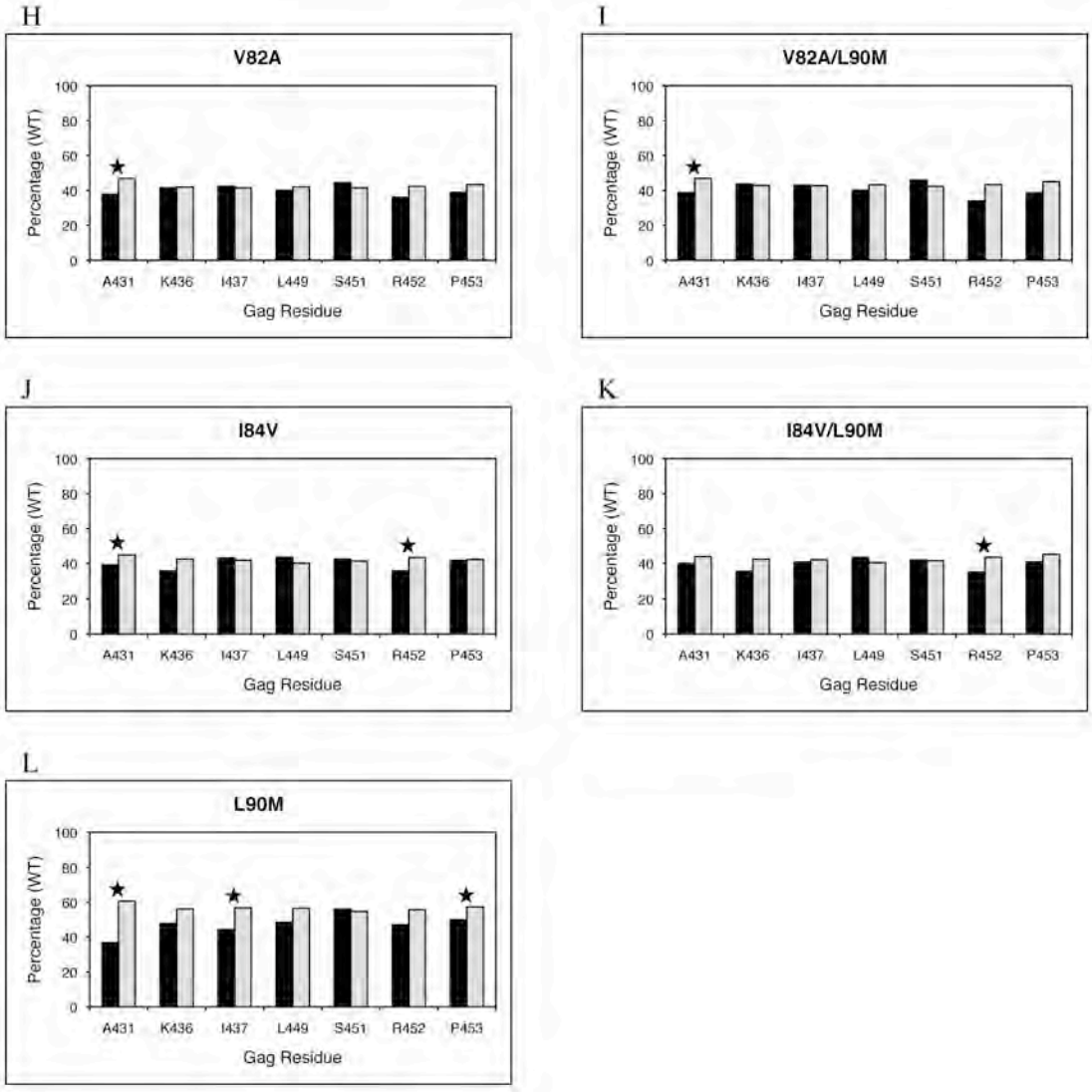


Figure III.4: Replicative Capacities of viruses with primary drug-resistant protease mutations and associated NC-p1 and p1-p6 cleavage site mutations. RCs of viruses tested are shown as a percentage of the WT reference virus (100%) on the y-axis and on the y-axis are Gag residues within NC-p1 and p1-p6 that mutate frequently. Closed bars represent viruses with both the particular protease and cleavage site mutations present. Open bars represent viruses with only the particular protease mutations present. Stars indicate significant difference in RCs for a particular set. p-values ≤ 0.05 are significant.

Figure III.5

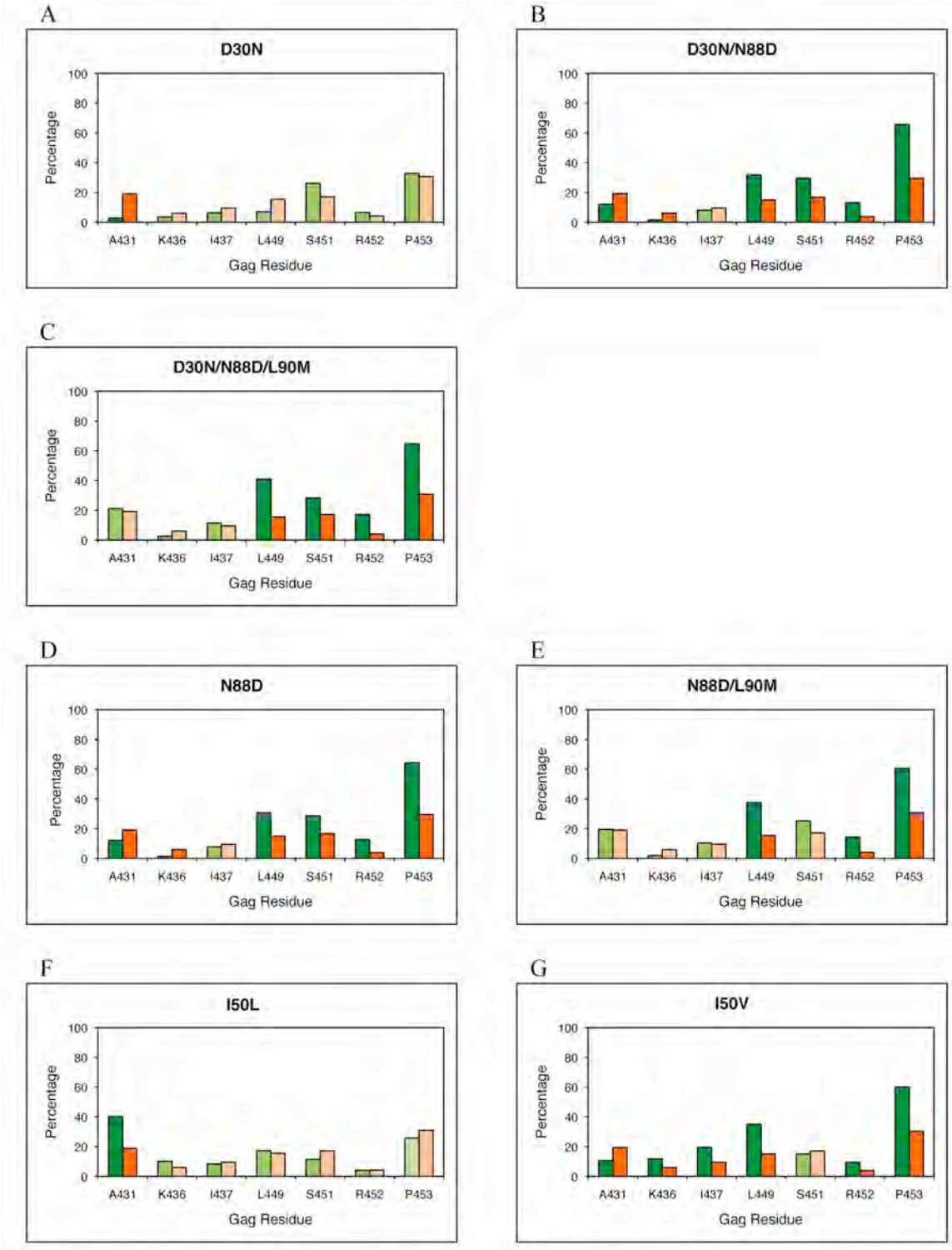

A
D30N
B
D30N/N88D
C
D30N/N88D/L90M
D
N88D
E
N88D/L90M
F
I50L
G
I50V
Percentage
0
20
40
60
80
100
A431
K436
I437
L449
S451
R452
P453
Gag Residue

Figure III.5 (*contd.*)

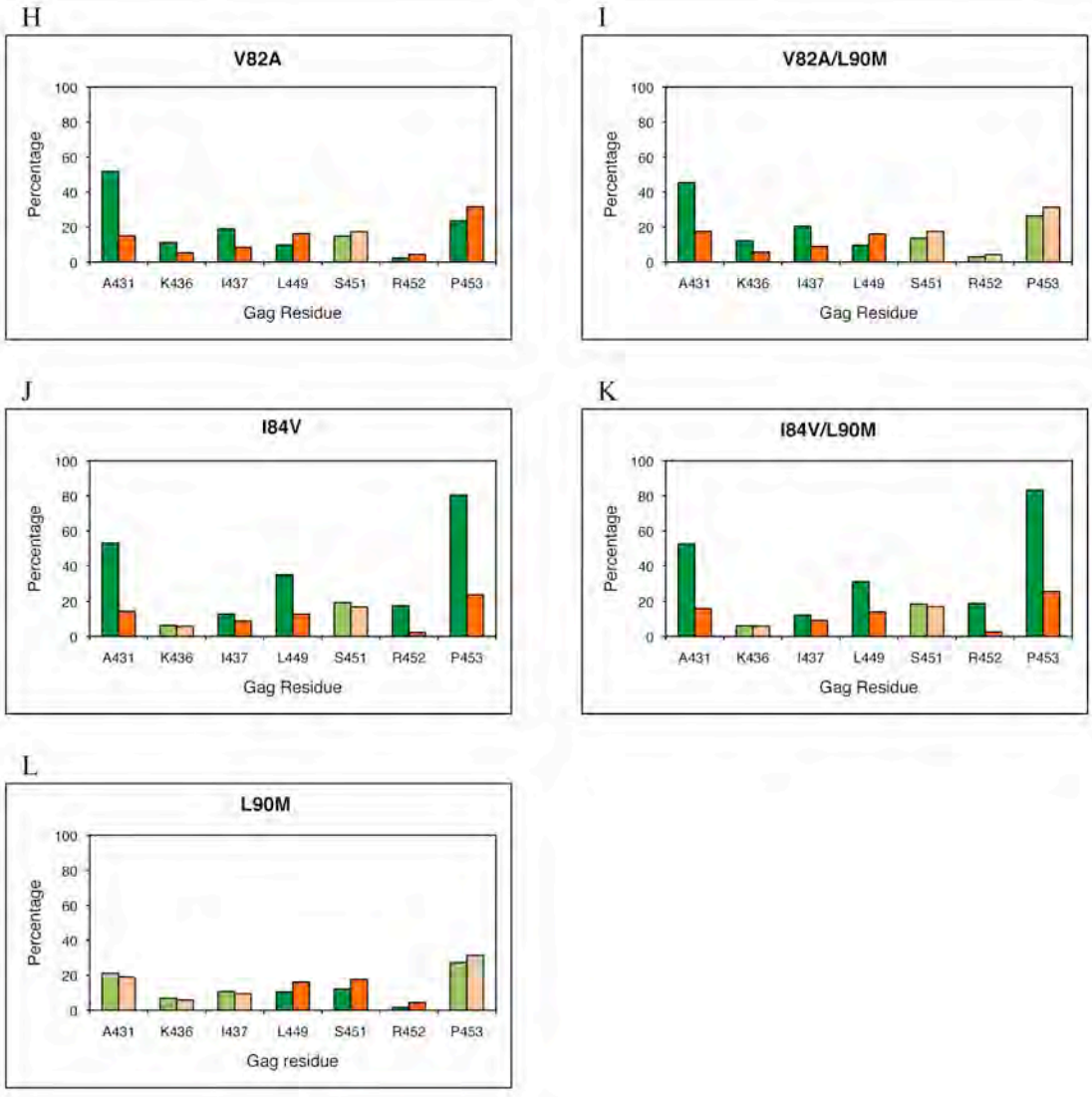


Figure III.5: Mutations within the NC-p1 and p1-p6 cleavage sites associated with primary resistance-associated mutations in the protease at residues 30, 50, 82, 84, 88 and 90. The frequency of cleavage site mutations are shown as percentages on the y-axis. On the x-axis are the amino acid residues (Gag numbering) within the NC-p1 and p1-p6 cleavage sites. In green are cleavage sites that mutate in the presence of the primary protease mutation and in orange are cleavage sites that mutate in the absence of the given protease mutations. Dark green and dark orange indicate significant association of mutations. p-values ≤ 0.05 are significant.

Figure III.6

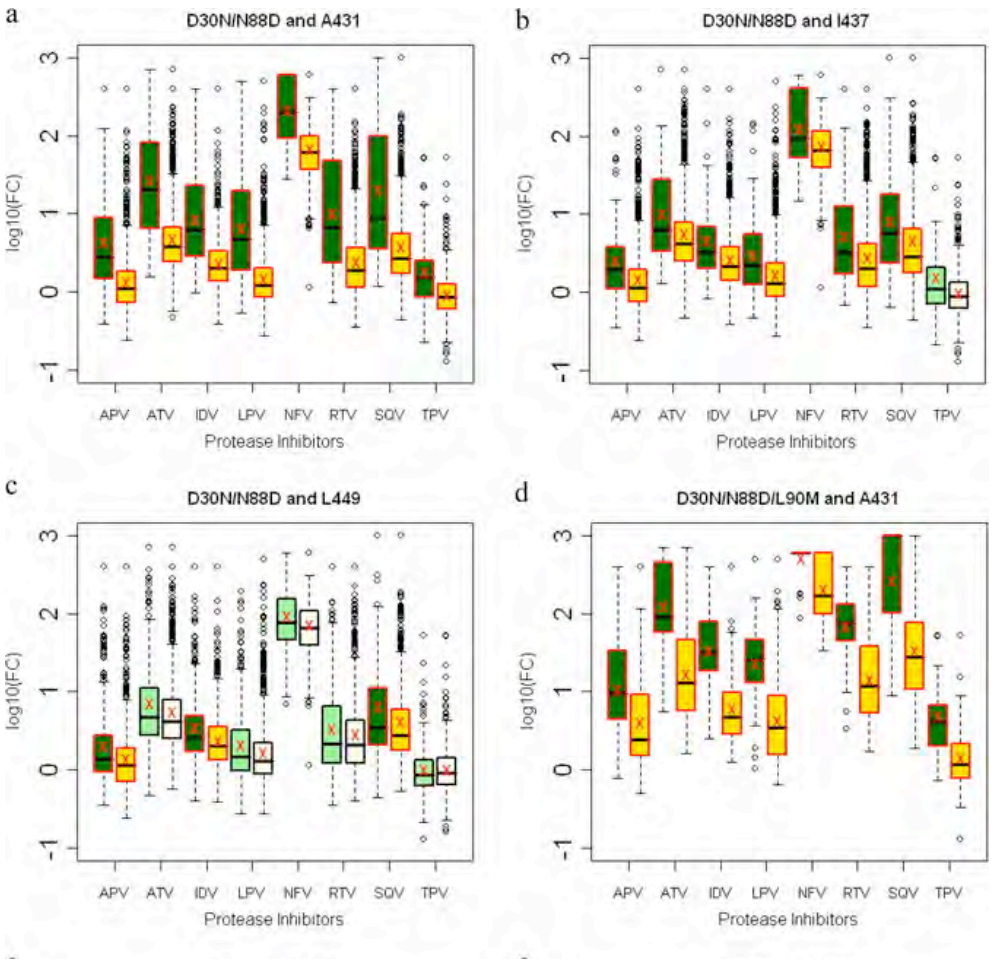

a
D30N/N88D and A431
b
D30N/N88D and I437
c
D30N/N88D and L449
d
D30N/N88D/L90M and A431
log10(FC)
-1
0
1
2
3
APV ATV IDV LPV NFV RTV SQV TPV
Protease Inhibitors

Figure III.6 *(cont.)*

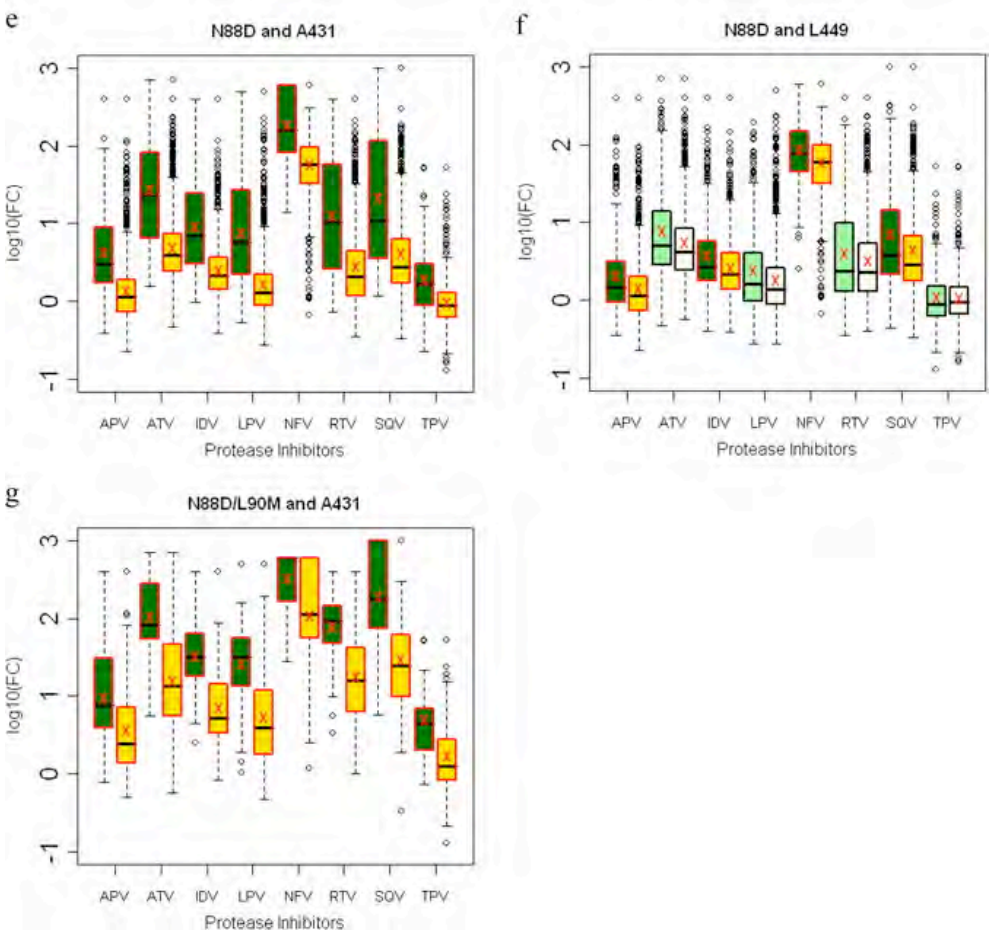


Figure III.6: Box plots (ratio of median FCs on a log scale on the Y-axis) showing susceptibilities of NFV-resistant viruses to various PIs (on the X-axis) in the presence (green boxes) or absence (yellow boxes) of associating mutations within the NC-p1 and p1-p6 cleavage sites. The boxes represent the inter quartile range between the first and third quartile. The upper and lower lines represent 1.5 times the upper and lower quartile limits. The black line represents the median and the red X represents the mean of the distribution. Outliers are shown as open circles. Dark green and gold boxes indicate significant effect of associating mutations within Gag cleavage sites on PI susceptibilities.

(a) Susceptibilities of D30N/N88D viruses with and without mutations at Gag 431.

(b) Susceptibilities of D30N/N88D viruses with and without mutations at Gag 437.

(c) Susceptibilities of D30N/N88D viruses with and without mutations at Gag 449.

(d) Susceptibilities of D30N/N88D/L90M viruses with and without mutations at Gag 431.

(e) Susceptibilities of N88D viruses with and without mutations at Gag 431.

(f) Susceptibilities of N88D viruses with and without mutations at Gag 449.

(g) Susceptibilities of N88D/L90M viruses with and without mutations at Gag 431.

p-values ≤ 0.05 are significant.

437 in combination with D30N/N88D protease mutations resulted in viruses with reduced sensitivity to APV, ATV, IDV, LPV, NFV, RTV and SQV (Fig III.6b). In contrast, as we've previously seen, D30N/N88D protease mutations were significantly associated with mutations within p1-p6 cleavage site where 32, 30 and 66% of sequences had mutations at L449, S451 and P453 respectively whereas only 15, 17 and 30% of sequences had mutations at these positions in the absence of D30N/N88D (Fig III.5b and Supplemental Table III.3) confirming our earlier studies (Kolli, Lastere, and Schiffer, 2006). Although not as frequent, Gag R452 was also observed to mutate in an associated manner with D30N/N88D (Fig III.5b). A similar trend was observed in D30N/N88D/L90M viruses (Fig III.5c). Viruses with D30N/N88D in combination with mutations at L449 showed small changes in FC across APV, IDV and SQV (Fig III.6c). Mutations within p1-p6 are also associated with the N88D and N88D/L90M protease mutations (Fig III.5d and e), and decreased sensitivity to APV, IDV, NFV, SQV and TPV was observed in viruses with Gag 449 mutation in combination with N88D (Fig III.6f). Thus D30N/N88D and N88D protease mutations associate with mutations within p1-p6, in contrast with mutations in NC-p1, although both combinations modulate susceptibilities to the PIs.

I50V and I50L. I50, positioned at the flap tips of the protease, plays an essential role in substrate recognition and cleavage by locking down the active site (Mahalingam et al., 1999; Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002). This key residue did not mutate in response to the first generation PIs, however subsequent PIs result in I50 mutating in a drug dependent manner. The I50L mutation results in severe resistance to ATV, but

causes hypersusceptibility to most other PIs (Colonno et al., 2004; Yanchunas et al., 2005) whereas the I50V mutation results in resistance to APV and to a lesser extent, DRV (Delaugerre et al., 2007; Partaledis et al., 1995; Pazhanisamy et al., 1998). The I50L protease mutation was observed to associate with mutations at Gag A431 within the NC-p1 cleavage site (Fig III.5f). Previous studies showed that Gag L449 and P453 within the p1-p6 cleavage site mutated in an associated manner with the I50V protease mutation (Maguire et al., 2002; Maguire et al., 2001) and was also observed here (Fig III.5g). Less frequently, mutations at Gag R452 were also associated with the I50V mutation. In addition, I50V protease mutation was also associated with mutations at K436 and I437 within NC-p1 cleavage site (Fig III.5g). Altered susceptibilities to various drugs were observed with both I50L and I50V protease mutations in the presence of mutations in NC-p1 and p1-p6 cleavage sites respectively. Viruses with I50L, a signature mutation of ATV resistance, often show hypersusceptibility to other protease inhibitors (Sista et al., 2008). Viruses with I50L in combination with Gag A431V lose this hypersusceptibility and show decreased susceptibility to APV, ATV, IDV, LPV, NFV and RTV (Fig III.7a). I50V viruses showed significantly reduced susceptibilities to IDV, LPV, NFV and SQV in combination with K436R mutation (Fig III.7b), in combination with mutations at Gag 437 showed reduced susceptibility to all PIs (Fig III.7c) whereas I50V viruses in combination with Gag L449F showed increased susceptibilities to IDV, LPV and especially to RTV (Fig III.7d). Thus, the impact of mutations at I50 is modulated by the addition of particular cleavage site mutations.

Figure III.7

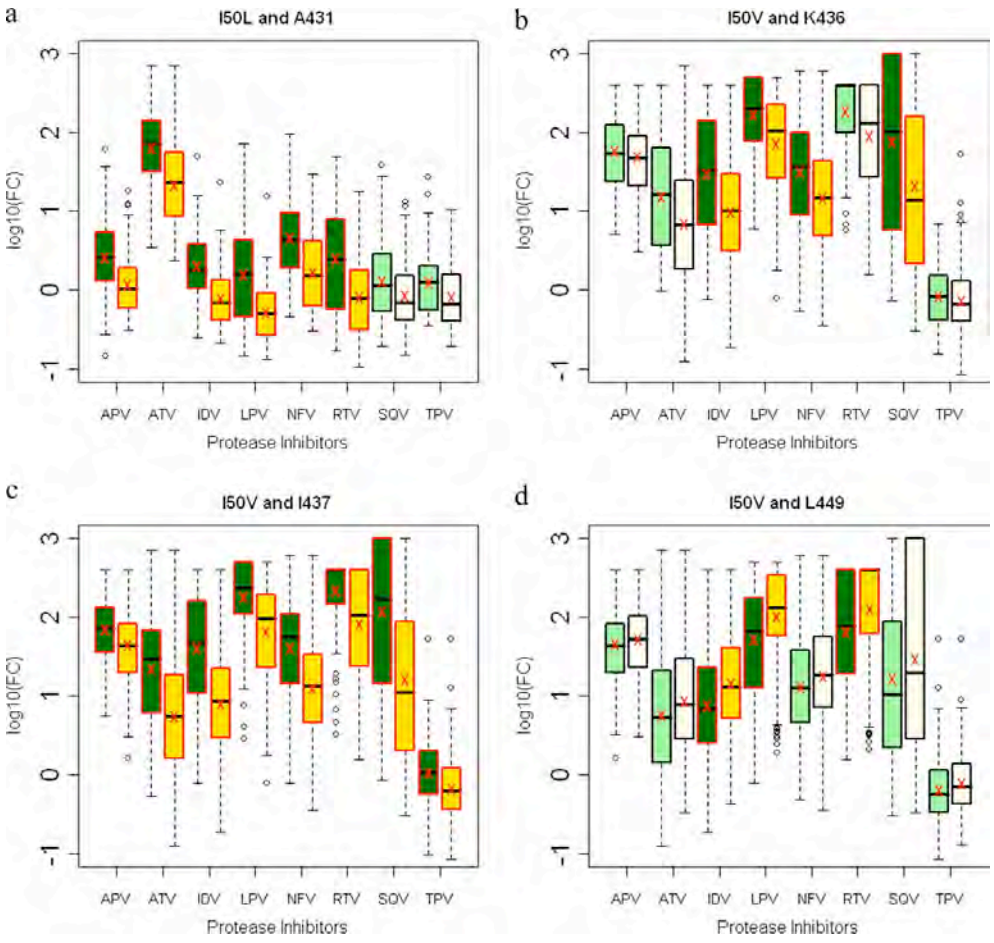

a
I50L and A431
b
I50V and K436
c
I50V and I437
d
I50V and L449
log10(FC)
-1
0
1
2
3
APV
ATV
IDV
LPV
NFV
RTV
SQV
TPV
Protease Inhibitors

Figure III.7: Box plots (ratio of median FCs on a log scale on the Y-axis) showing susceptibilities of I50V and I50L viruses to various PIs (on the X-axis) in the presence (green boxes) or absence (yellow boxes) of associating mutations within the NC-p1 and p1-p6 cleavage sites. The boxes represent the inter quartile range between the first and third quartile. The upper and lower lines represent 1.5 times the upper and lower quartile limits. The black line represents the median and the red X represents the mean of the distribution. Outliers are shown as open circles. Dark green and gold boxes indicate significant effect of associating mutations within Gag cleavage sites on PI susceptibilities.

(a) Susceptibilities of I50L viruses with and without mutations at Gag 431.

(b) Susceptibilities of I50V viruses with and without mutations at Gag 436.

(c) Susceptibilities of I50V viruses with and without mutations at Gag 437.

(d) Susceptibilities of I50V viruses with and without mutations at Gag 449.

p-values ≤ 0.05 are significant.

V82A. The V82A protease mutation has been well documented to associate with Gag A431V (Prabu-Jeyabalan et al., 2006; Zhang et al., 1997). 52% of viruses have this combination of mutations whereas only 15% have the A431V mutation in the absence of the V82A protease mutation (Fig III.5h and Supplemental Table III.3). Those viruses with both mutations exhibited lower susceptibility to IDV and LPV but were hypersusceptible to TPV (Fig III.8a). However, a significant increase in susceptibility to RTV and TPV, to a lesser extent, was observed in viruses with the combination of V82A/L90M and Gag A431V mutations (Fig III.8b). Mutations at Gag 436 and 437 within the NC-p1 were also associated with both V82A and V82A/L90M protease mutations (Fig III.5h and i). Mutations at Gag K436 were observed to associate with the V82A protease mutation, where Gag K436 mutated in 11% and 12% of the viruses in the presence while only 5% and 6% mutated in the absence of V82A and V82A/L90M respectively (Supplemental Table III.3). However, no associated changes in susceptibilities were observed. The Gag 437 mutations, which occurred 19% and 20% of the time with V82A and V82A/L90M respectively, resulted in reduced sensitivity to all PIs (Fig III.8c and d). Mutations within the p1-p6 site were significantly not associated with either V82A or V82A/L90M protease mutations. Only 10% of viruses had Gag L449F mutation and V82A or V82A/L90M protease mutations as opposed to 16% that had mutations only in the cleavage site (Fig III.5h, i and Supplemental Table III.3). Mutations at Gag L449 in combination with V82A or V82A/L90M decreased susceptibilities to all protease inhibitors tested (Fig III.8e and f). Mutations at Gag P453 with V82A viruses also showed significantly decreased susceptibilities to all PIs (Fig

Figure III.8

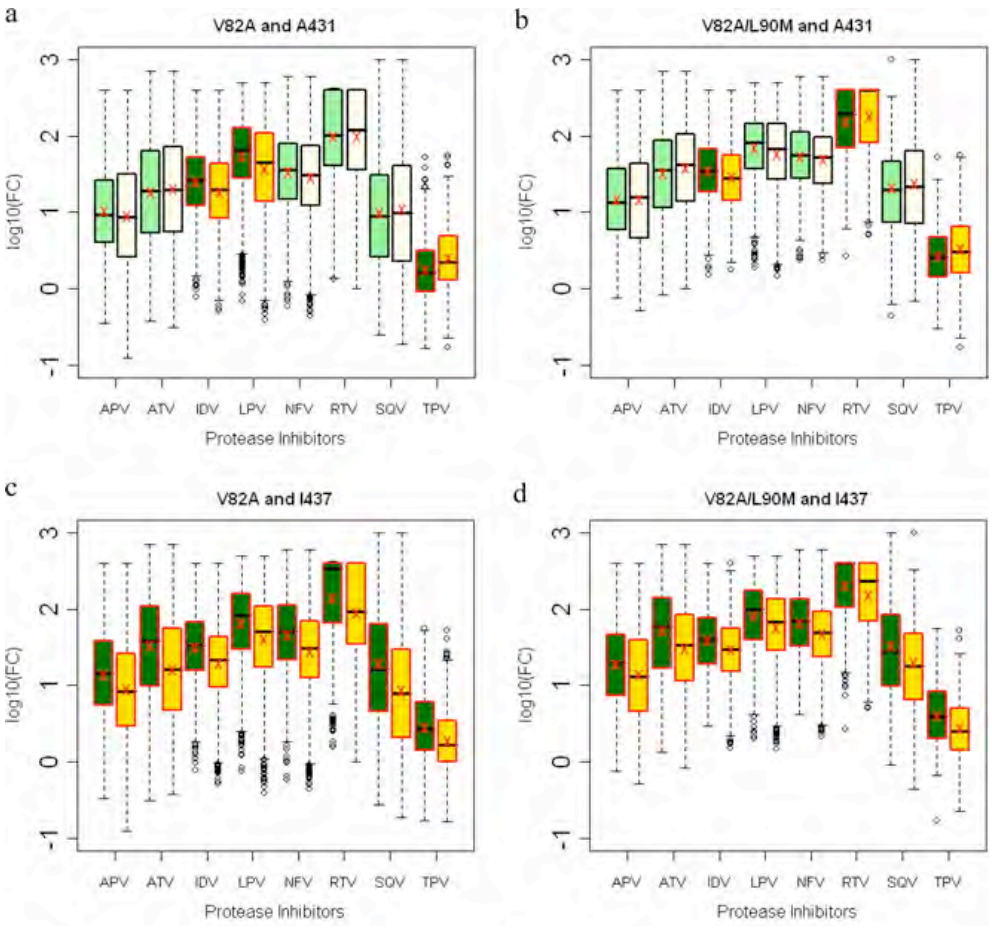

a
V82A and A431
b
V82A/L90M and A431
c
V82A and I437
d
V82A/L90M and I437
log10(FC)
-1
0
1
2
3
APV
ATV
IDV
LPV
NFV
RTV
SQV
TPV
Protease Inhibitors

Figure III.8 *(cont.)*

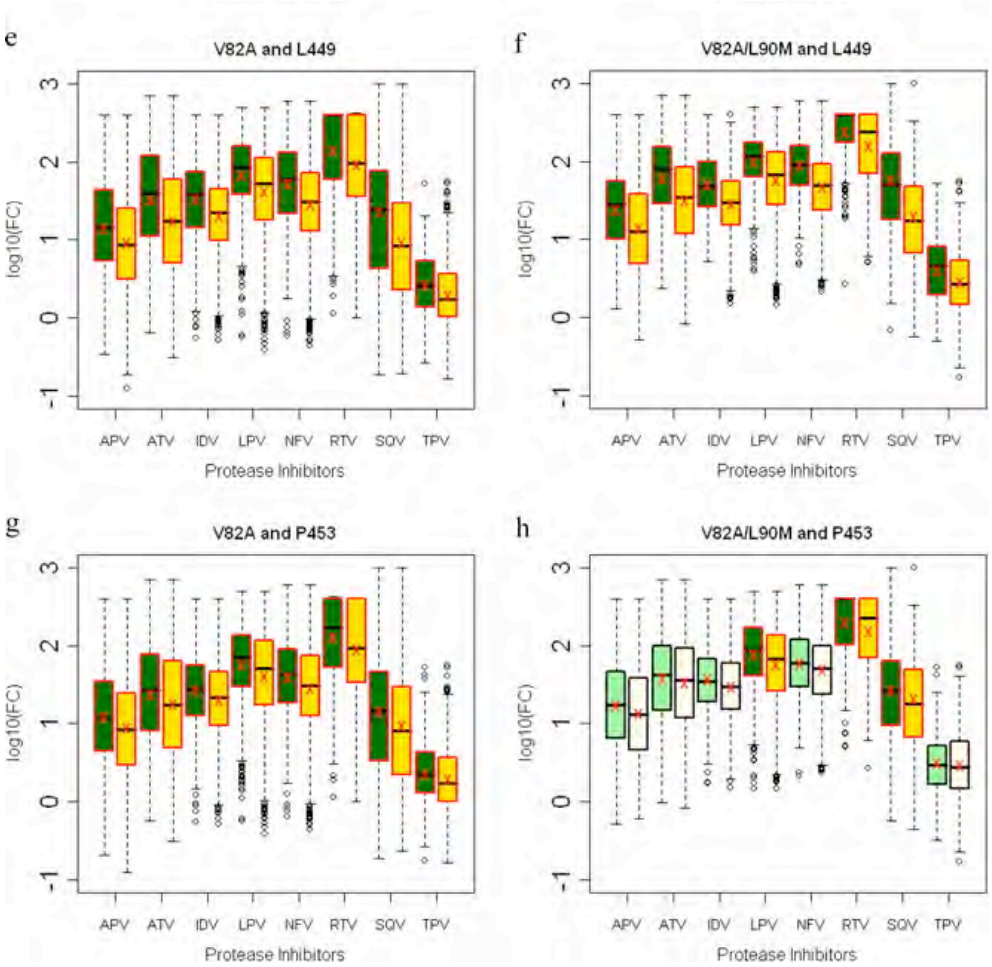


Figure III.8: Box plots (ratio of median FCs on a log scale on the Y-axis) showing susceptibilities of V82A and V82A/L90M viruses to various PIs (on the X-axis) in the presence (green boxes) or absence (yellow boxes) of associating mutations within the NC-p1 and p1-p6 cleavage sites. The boxes represent the inter quartile range between the first and third quartile. The upper and lower lines represent 1.5 times the upper and lower quartile limits. The black line represents the median and the red X represents the mean of the distribution. Outliers are shown as open circles. Dark green and gold boxes indicate significant effect of associating mutations within Gag cleavage sites on PI susceptibilities.

(a) Susceptibilities of V82A viruses with and without mutations at Gag 431.

(b) Susceptibilities of V82A/L90M viruses with and without mutations at Gag 431.

(c) Susceptibilities of V82A viruses with and without mutations at Gag 437.

(d) Susceptibilities of V82A/L90M viruses with and without mutations at Gag 437.

(e) Susceptibilities of V82A viruses with and without mutations at Gag 449.

(f) Susceptibilities of V82A/L90M viruses with and without mutations at Gag 449.

(g) Susceptibilities of V82A viruses with and without mutations at Gag R453.

(h) Susceptibilities of V82A/L90M viruses with and without mutations at Gag 453.

p-values ≤ 0.05 are significant.

III.8g) whereas in combination with V82A/L90M viruses, decreased susceptibilities were observed with LPV, RTV and SQV (Fig III.8h). Thus, with a few exceptions, mutations within the NC-p1 and p1-p6 cleavage sites, often modulated by L90M, are seen to decrease the susceptibility of V82A protease containing viruses.

I84V. Several Gag cleavage site mutations result in viruses with significant decreases in susceptibilities when combined with the I84V protease mutation. Within NC-p1, mutation at Gag A431 occurred 53% of the viruses in presence of I84V or I84V/L90M as compared to 14% and 16% respectively, in the absence of these protease mutations (Fig III.5j, k and Supplemental Table III.3). Decreased susceptibilities to all protease inhibitors were observed when mutations at Gag 431 occurred with either I84V alone or I84V/L90M protease mutations (Fig III.9a and b). Even though no associations were observed, mutations at Gag 436 in combination with I84V resulted in decreased susceptibility to RTV, whereas in combination with I84V/L90M decreased susceptibility to ATV and RTV (Fig III.9c and d). Mutations at I437 within the NC-p1 were also associated with these protease mutations (Fig III.5j and k). In the case of I84V viruses with mutations at Gag 437, a significant reduction to all PIs was observed (Fig III.9e) whereas with I84V/L90M viruses, reduced sensitivity was observed to all PIs except SQV (Fig III.9f). Significant association of mutations was also observed between I84V and mutations within the p1-p6 cleavage site. Mutations at Gag L449 and P453 were the most common, and mutations at R452 were less frequent. Gag 449 mutations associated with I84V and I84V/L90M, in 35% and 31% of the viruses respectively, as compared to 13% and 44% respectively, in the absence of I84V and I84V/L90M mutations (Fig III.5j,

Figure III.9

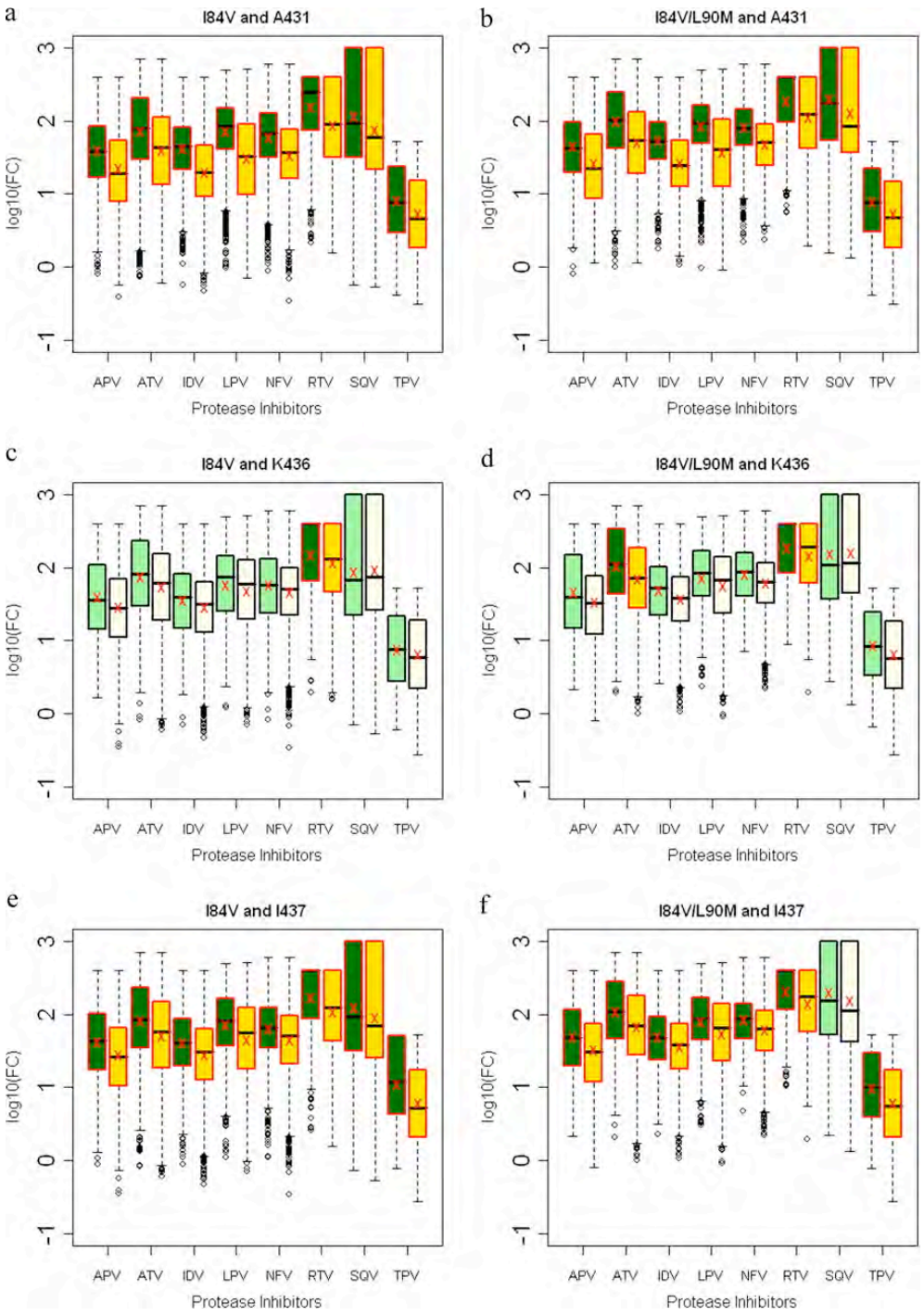

a
I84V and A431
b
I84V/L90M and A431
c
I84V and K436
d
I84V/L90M and K436
e
I84V and I437
f
I84V/L90M and I437
log10(FC)
-1
0
1
2
3
APV ATV IDV LPV NFV RTV SQV TPV
Protease Inhibitors

Figure III.9 *(cont.)*

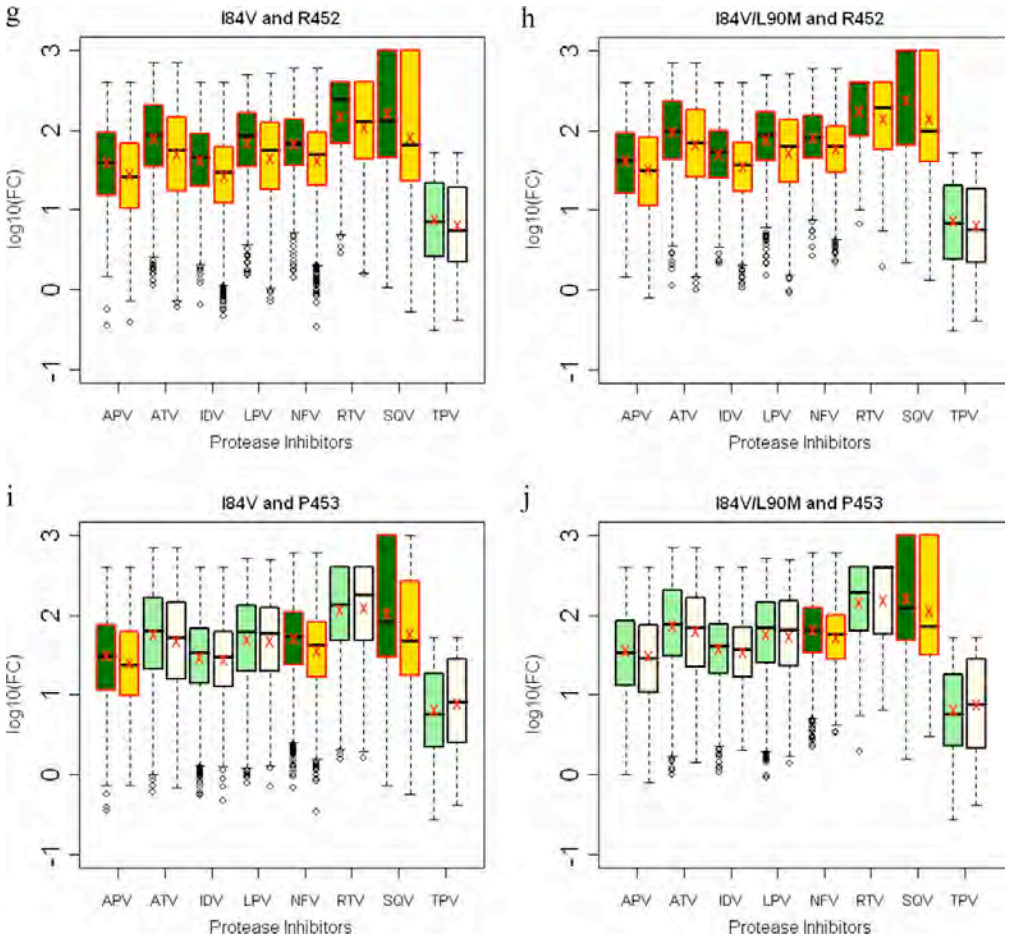


Figure III.9: Box plots (ratio of median FCs on a log scale on the Y-axis) showing susceptibilities of I84V and I84V/L90M viruses to various PIs (on the X-axis) in the presence (green boxes) or absence (yellow boxes) of associating mutations within the NC-p1 and p1-p6 cleavage sites. The boxes represent the inter quartile range between the first and third quartile. The upper and lower lines represent 1.5 times the upper and lower quartile limits. The black line represents the median and the red X represents the mean of the distribution. Outliers are shown as open circles. Dark green and gold boxes indicate significant effect of associating mutations within Gag cleavage sites on PI susceptibilities.

(a) Susceptibilities of I84V viruses with and without mutations at Gag 431.

(b) Susceptibilities of I84V/L90M viruses with and without mutations at Gag 431.

(c) Susceptibilities of I84V viruses with and without mutations at Gag 436.

(d) Susceptibilities of I84V/L90M viruses with and without mutations at Gag 436.

(e) Susceptibilities of I84V viruses with and without mutations at Gag 437.

(f) Susceptibilities of I84V/L90M viruses with and without mutations at Gag 437.

(g) Susceptibilities of I84V viruses with and without mutations at Gag 452.

(h) Susceptibilities of I84V/L90M with and without mutations at 452.

(i) Susceptibilities of I84V viruses with and without mutations at Gag 453.

(j) Susceptibilities of I84V/L90M with and without mutations at Gag 453.

p-values ≤ 0.05 are significant.

k and Supplemental Table III.3). Mutations at Gag 452 were also associated, to a lesser extent, with both I84V and I84V/L90M protease mutations (Fig III.5j and k). No corresponding modulations in drug susceptibilities were observed in either I84V or I84V/L90M viruses with mutations at Gag 449. Significant decreases in susceptibilities were observed with APV, ATV, IDV, LPV, NFV and especially RTV and SQV, when I84V or I84V/L90M associated with mutations at Gag R452 (Fig III.9g and h). Viruses with mutations at Gag 453 showed reduced susceptibilities to APV, NFV and SQV in combination with I84V (Fig III.9i), whereas with I84V/L90M reduced susceptibilities were observed with NFV and SQV (Fig III.9j). Thus, several Gag cleavage site mutations, in combination with I84V, modulate the level of susceptibility to various protease inhibitors.

L90M. L90M protease mutation is most commonly associated with other primary PI-resistant protease mutations leading to an increase in resistance. Mutations within both NC-p1and p1p6 cleavage sites were not associated when L90M occurred independent of other primary protease mutations (Fig III.5l). Mutations at Gag 431 and 436 decreased susceptibilities to all PIs in combination with L90M protease mutation (Fig III.10a and b). Decreased susceptibilities to APV, ATV, NFV and SQV were observed with mutations at Gag 449; to ATV and SQV with mutations at Gag 451; to SQV with mutations at Gag 452 and to all PIS with mutations at Gag 453 when they occurred in combination with the L90M protease mutation (Fig III.10c-f). Thus, decreased sensitivity to various PIs were observed in combination with mutations in both NC-p1 and p1-p6 cleavage sites.

Figure III.10

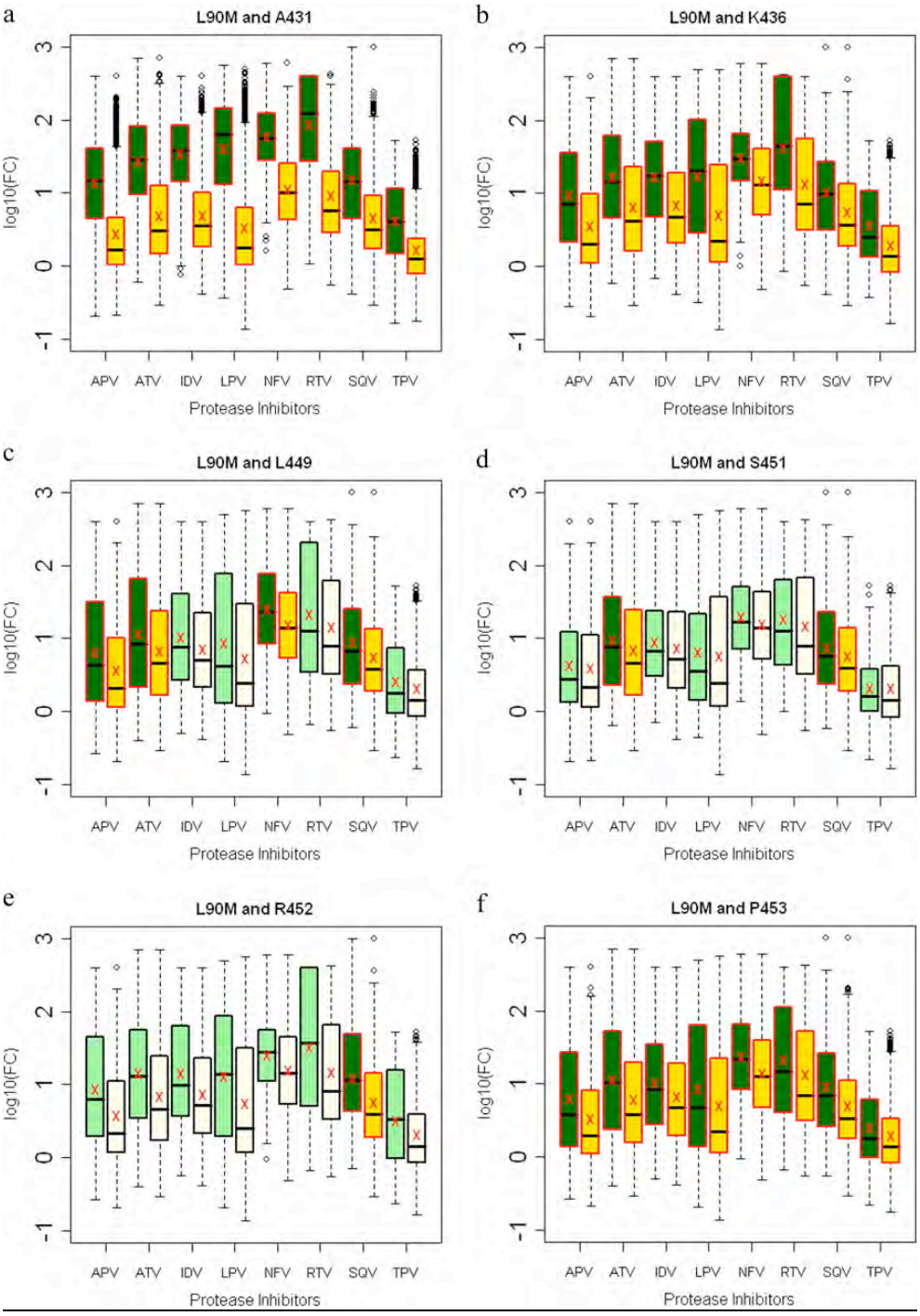


Figure III.10: Box plots (ratio of median FCs on a log scale on the Y-axis) showing susceptibilities of L90M viruses to various PIs (on the X-axis) in the presence (green boxes) or absence (yellow boxes) of associating mutations within the NC-p1 and p1-p6 cleavage sites. The boxes represent the inter quartile range between the first and third quartile. The upper and lower lines represent 1.5 times the upper and lower quartile limits. The black line represents the median and the red X represents the mean of the distribution. Outliers are shown as open circles. Dark green and gold boxes indicate significant effect of associating mutations within Gag cleavage sites on PI susceptibilities.

(a) Susceptibilities of L90M viruses with and without mutations at Gag 431.

(b) Susceptibilities of L90M viruses with and without mutations at Gag 436.

(c) Susceptibilities of L90M viruses with and without mutations at Gag 449.

(d) Susceptibilities of L90M viruses with and without mutations at Gag 451.

(e) Susceptibilities of L90M viruses with and without mutations at Gag 452.

(f) Susceptibilities L90M viruses with and without mutations at Gag 453.

p-values ≤ 0.05 are significant.

DISCUSSION

NC-p1 and p1-p6 cleavage sites, two critical sites in Gag processing, were systematically analyzed for mutations associated with primary PI resistance mutations in the protease. For every single active site protease mutation studied, at least one, often novel, associated mutation was observed in either of these two critical cleavage sites. In this study we show that a number of these cleavage site mutations are associated with drug resistant mutations within the protease. While previous studies have shown an increased resistance and/or improved fitness for some pairs of protease-cleavage site mutations (Bally et al., 2000; Dam et al., 2009; Doyon et al., 1996; Feher et al., 2002; Kolli, Lastere, and Schiffer, 2006; Maguire et al., 2002; Mammano, Petit, and Clavel, 1998) our analyses reveal that even though only some of these alter fitness, a majority of these cleavage site mutations significantly modulate, usually increasing, resistance.

NC-p1 mutations occurred frequently at Gag 431, 436 and 437, often in combination with several highly resistant protease mutations. Earlier studies have shown that mutations at Gag 436 and 437 enhance Gag processing (Dam et al., 2009; Maguire et al., 2002; Nijhuis et al., 2007) and increase resistance, although the exact mechanism is not clearly understood. Our study observed that K436R increased resistance to certain PIs in combination with I50V, I84V and I84V/L90M mutations in protease and I437V increased resistance to all PIs with I50V, V82A and I84V mutations in protease. Gag 436 and 437 are further from the catalytic D25 than Gag 431, thus making structural predictions of their potential mechanism of action less reliable. The A431V mutation results in a more

efficient cleavage of the NC-p1 cleavage site irrespective of the protease variant (Feher et al., 2002) as a result of additional contacts with the protease (Prabu-Jeyabalan et al., 2004). A recent study showed that in viruses from heavily treated patients, A431V also appeared to directly enhance resistance as a result of this improved NC-p1 processing (Dam et al., 2009; Nijhuis et al., 2007). This non-specific improvement in processing of the NC-p1 cleavage site is possibly why this mutation was observed frequently in combination with V82A/F, I84V and I50L protease mutations in our study. Even in combination with the D30N protease mutation, that is not associated with A431V, an increased level resistance to multiple PIs in the 12% of viruses that have this combination was observed. In the validation studies, unlike with most of the cleavage site mutations, generally more secondary mutations were observed in the protease in viruses with A431V (Supplemental Table III.1), perhaps contributing to the increased resistance and warranting further study. Overall our study demonstrates that mutations within NC-p1 result in higher levels of resistance when coupled with specific protease mutations.

Mutations within the p1-p6 cleavage site occurred frequently in combination with D30N, D30N/N88D, I84V and I84V/L90M protease mutations sometimes enhancing resistance to various PIs. However mutations within this cleavage site are negatively associated with V82A and V82A/L90M, even while enhancing resistance. Mutations at Gag 449 and 453 alone, in the absence of protease mutations, were previously shown not to reduce susceptibility to PIs (Maguire et al., 2002), however mutated peptides corresponding to these cleavage sites are better substrates for a variety of protease variants (Feher et al., 2002; Maguire et al., 2002). L449F may enhance catalytic efficiency as a result of

improved van der Waals contacts, as we and others have previously suggested (Feher et al., 2002; Kolli, Lastere, and Schiffer, 2006). The overall improved catalytic efficiency from mutations in p1-p6 in combination with specific protease mutations may be part of the potential mechanism by which the observed increase in resistance to various PIs occurs.

Other underlying causes for the observed increased levels of resistance were evaluated and seem not alter our conclusions. Systematic analysis of the impact of these associated NC-p1 and p1-p6 cleavage site mutations on fitness and susceptibility to PIs showed these processes not to be directly coupled. There were no significant improvements in viral fitness due to the cleavage site mutations, even as significant alterations in sensitivity to various PIs were observed. Overall, in resistant viruses, most of the changes in this region of Gag (AA 418-500) occurred within the cleavage sites, though increased mutation rates were also observed in the PTAP region (Fig III.3) but to a lesser extent. The number of secondary protease mutations observed in viruses with mutation in protease with and without Gag cleavage site mutations generally did not show a trend (Supplemental Table III.1) except with A431. Thus the interdependency that leads to modulation of resistance appears to be specific between the mutations in protease and the mutations in the NC-p1 and p1-p6 cleavage sites.

This study strongly suggests that association of mutations within Gag cleavage sites and the protease contributes to PI susceptibility in an inhibitor specific manner. Mutations within NC-p1 and p1-p6 cleavage sites, in combination with specific protease mutations,

were observed to significantly modulate drug susceptibilities. Frequently, these specific combinations enhance PI resistance, while in a few cases increased susceptibility was observed. Modulated sensitivity to PIs due to these mutations may be the result of alterations in polyprotein processing and structural adaptations of both the protease and Gag proteins (Dam et al., 2009; Ho et al., 2008; Maguire et al., 2002; Nijhuis et al., 1999; Nijhuis et al., 2007). In mutations in either NC-p1 or p1-p6, cases occur that make the cleavage site a better substrate, complementing the mutated protease even with protease mutations not associated with it and thereby causing increased levels of resistance. Further studies are warranted to determine the exact mechanisms and extent to which these cleavage site mutations contribute to PI resistance. The inclusion of Gag during genotypic or phenotypic resistance testing may lead to improved accuracy in the measurement of viral resistance and have direct implications for evaluating patients failing PI therapy.

Acknowledgements: We acknowledge Dr. Neil. T. Parkin for his assistance in obtaining and evaluating the Monogram Biosciences, Inc. data. This research was supported by the National Institutes of Health (R01- GM65347).

Supplemental Table III.1

Gag	Protease Profile	Mean P-P	Mean P-A	PP - PA	P-value
A431	D30N	4.6	2.0	2.6	1.7E-03
	D30N/N88D	5.4	3.8	1.6	0.0E+00
	I50L	5.5	3.5	2.0	1.5E-10
	I50V	7.8	6.8	1.0	4.7E-04
	V82A	6.7	6.5	0.3	3.8E-04
	V82A/L90M	7.5	7.5	0.1	1.9E-01
	I84V	7.3	6.6	0.8	0.0E+00
	I84V/L90M	7.7	7.0	0.7	0.0E+00
	N88D	5.5	3.9	1.6	0.0E+00
	N88D/L90M	7.3	5.9	1.5	4.0E-08
	L90M	6.8	4.0	2.8	0.0E+00
K436	D30N	1.6	2.2	-0.6	2.5E-01
	D30N/N88D	4.3	4.0	0.3	1.8E-01
	I50L	5.4	4.2	1.1	2.6E-02
	I50V	7.4	6.9	0.5	1.4E-01
	V82A	6.5	6.6	-0.1	2.4E-01
	V82A/L90M	7.3	7.5	-0.2	4.4E-02
	I84V	7.1	6.9	0.1	1.0E-01
	I84V/L90M	7.5	7.4	0.2	1.2E-01
	N88D	4.3	4.1	0.2	2.1E-01
	N88D/L90M	5.4	6.2	-0.8	3.2E-01
	L90M	5.9	4.4	1.4	2.2E-16
I437	D30N	1.9	2.1	-0.2	6.4E-01
	D30N/N88D	4.3	4.0	0.3	2.5E-01
	I50L	4.8	4.2	0.6	3.3E-01
	I50V	7.8	6.7	1.1	3.8E-07
	V82A	7.0	6.5	0.4	1.7E-08
	V82A/L90M	7.7	7.4	0.3	2.7E-03
	I84V	7.2	7.0	0.2	7.0E-05
	I84V/L90M	7.8	7.3	0.5	4.1E-08
	N88D	4.5	4.1	0.4	1.1E-01
	N88D/L90M	6.6	6.1	0.5	1.8E-01
	L90M	6.1	4.4	1.8	0.0E+00

Supplemental Table III.1 *(cont.)*

L449	D30N	2.4	2.1	0.2	8.2E-01
	D30N/N88D	4.2	3.9	0.3	3.0E-01
	I50L	3.8	4.3	-0.5	1.1E-01
	I50V	6.5	7.2	-0.7	4.2E-06
	V82A	7.1	6.5	0.6	1.3E-08
	V82A/L90M	8.1	7.4	0.7	1.7E-09
	I84V	6.7	7.1	-0.4	1.8E-08
	I84V/L90M	7.4	7.3	0.0	5.7E-01
	N88D	4.3	4.0	0.3	7.9E-02
	N88D/L90M	6.3	6.0	0.3	1.4E-01
	L90M	5.2	4.5	0.7	7.2E-07
S451	D30N	2.3	2.1	0.2	7.0E-01
	D30N/N88D	4.0	4.1	-0.1	3.1E-01
	I50L	3.9	4.3	-0.4	2.6E-01
	I50V	7.4	6.9	0.5	3.0E-02
	V82A	6.5	6.6	-0.1	8.3E-01
	V82A/L90M	7.7	7.5	0.2	6.0E-03
	I84V	6.8	7.0	-0.1	2.8E-02
	I84V/L90M	7.3	7.4	-0.1	1.5E-01
	N88D	4.0	4.2	-0.2	9.2E-02
	N88D/L90M	6.1	6.2	-0.1	9.2E-01
	L90M	4.8	4.6	0.3	2.2E-02
R452	D30N	2.7	2.1	0.6	4.8E-01
	D30N/N88D	4.4	4.0	0.4	4.2E-03
	I50L	4.9	4.2	0.6	3.5E-01
	I50V	7.4	6.9	0.5	1.1E-01
	V82A	7.4	6.6	0.8	1.2E-04
	V82A/L90M	7.9	7.5	0.4	4.2E-02
	I84V	7.4	6.9	0.5	7.7E-12
	I84V/L90M	7.7	7.3	0.4	3.0E-07
	N88D	4.4	4.1	0.4	1.4E-02
	N88D/L90M	6.4	6.1	0.3	3.2E-01
	L90M	5.6	4.5	1.0	3.1E-03

Supplemental Table III.1 *(cont.)*

P453	D30N	2.4	2.1	0.3	4.7E-01
	D30N/N88D	4.1	4.0	0.1	8.9E-02
	I50L	4.0	4.3	-0.3	3.4E-01
	I50V	6.9	6.9	0.0	9.8E-01
	V82A	7.0	6.5	0.6	2.0E-13
	V82A/L90M	7.8	7.4	0.4	4.4E-07
	I84V	7.0	7.0	0.0	8.9E-01
	I84V/L90M	7.4	7.4	0.0	6.4E-01
	N88D	4.2	4.1	0.1	2.5E-01
	N88D/L90M	6.4	5.8	0.6	2.1E-02
	L90M	5.3	4.3	1.0	0.0E+00

Supplemental Table III.1: Average number of secondary mutations
P-P -- Samples with specified mutations present in both the protease and the cleavage site
P-A -- Samples with specified mutations present in the protease but absent the in cleavage site
p-values < 0.05 are significant

Supplemental Table III.2

Protease Mutations	Gag	P-P	P-A	Mean P-P	Mean P-A	Median P-P	Median P-A	Std dev P-P	Std dev P-A
D30N	A431	4	118	32.9	46.0	27.3	39.0	24.7	38.3
	K436	3	120	30.4	45.3	38.7	35.2	16.5	38.3
	I437	8	114	26.6	46.8	29.2	39.5	23.4	38.5
	L449	8	111	49.6	44.1	48.0	34.9	26.5	38.8
	S451	32	84	35.3	49.1	32.6	39.5	29.5	40.8
	R452	7	111	43.1	45.2	59.2	34.9	34.1	38.4
	P453	39	70	45.0	42.8	38.6	37.1	36.8	34.1
D30N/N88D	A431	122	852	65.1	70.3	61.2	61.2	43.3	48.4
	K436	14	954	69.6	69.7	64.3	61.7	54.1	47.2
	I437	80	882	73.9	69.5	56.8	62.0	55.4	46.6
	L449	296	649	68.4	70.0	61.2	61.7	47.1	47.3
	S451	272	622	70.2	68.3	63.7	59.0	46.5	48.7
	R452	124	808	64.8	70.0	57.4	62.0	46.9	47.0
	P453	590	312	70.9	65.4	61.3	60.7	48.5	43.6
D30N/N88D/L90M	K436	37	125	64.9	75.1	67.2	70.6	30.4	49.3
	I437	4	163	55.1	73.7	57.2	68.1	44.0	46.7
	L449	19	146	80.3	72.1	77.8	68.0	58.4	45.2
	S451	67	93	72.0	70.2	69.4	59.7	41.9	48.4
	R452	44	110	64.9	73.9	56.9	68.0	40.8	47.6
	P453	27	128	55.3	72.5	45.0	70.7	38.4	45.1
	A431	103	55	67.6	76.3	54.1	70.6	46.4	46.4

Supplemental Table III.2 *(cont.)*

I50L	A431	58	81	32.9	29.6	26.0	23.2	29.2	25.0
	K436	14	123	19.8	30.6	10.3	25.9	30.3	24.8
	I437	13	135	28.1	30.7	23.3	58.0	26.4	48.0
	L449	26	124	25.4	32.6	27.2	24.0	19.5	27.7
	S451	16	128	28.9	31.6	20.2	25.8	21.8	27.2
	R452	5	148	45.3	30.4	47.9	23.3	29.4	26.3
	P453	40	112	27.2	32.2	22.4	25.4	22.2	27.8
I50V	A431	56	485	37.4	50.9	33.5	45.2	30.7	37.2
	K436	60	455	55.8	48.6	53.5	43.3	36.3	37.0
	I437	104	428	49.7	49.3	47.4	43.2	34.8	37.1
	L449	181	336	46.7	50.2	36.2	44.9	38.6	33.3
	S451	79	449	51.5	48.9	45.1	43.4	37.6	36.7
	R452	47	468	47.5	49.3	43.4	43.6	30.8	35.7
	P453	295	195	49.3	49.5	43.9	37.9	34.1	42.2
V82A	A431	1775	1636	37.9	46.8	31.3	38.6	30.5	35.6
	K436	372	2952	41.6	41.9	33.6	33.7	33.3	33.1
	I437	635	2698	42.4	41.5	36.0	33.1	32.7	33.4
	L449	321	3041	40.1	42.1	32.5	34.0	32.8	33.2
	S451	487	2831	44.4	41.5	37.0	33.3	33.5	33.2
	R452	75	3404	36.1	42.3	29.6	34.5	32.6	33.3
	P453	759	2495	38.9	43.4	32.0	35.5	31.8	33.8

Supplemental Table III.2 *(cont.)*

V82A/L90M	A431	871	1039	38.8	46.9	32.2	39.0	30.5	35.6
	K436	227	1628	43.7	42.9	36.2	34.7	33.7	33.3
	I437	378	1475	43.1	42.7	36.4	34.2	33.3	33.6
	L449	179	1690	40.1	43.2	32.4	35.6	31.6	33.4
	S451	247	1605	45.9	42.3	37.0	34.7	33.8	33.3
	R452	55	1884	34.0	43.4	29.5	35.9	32.1	33.5
	P453	465	1343	38.6	45.1	32.5	36.9	30.3	34.6
N88D	A431	138	954	59.4	67.5	53.7	57.0	44.0	48.5
	K436	15	1068	66.1	66.5	50.2	57.2	54.0	47.6
	I437	87	989	69.5	66.4	54.6	57.7	55.3	47.1
	L449	320	739	65.7	66.4	57.6	56.7	47.4	47.8
	S451	295	709	68.4	64.0	59.7	53.2	46.0	48.9
	R452	136	910	63.1	66.6	55.0	57.4	47.1	47.5
	P453	646	361	68.3	61.7	57.6	56.5	49.3	43.0
N88D/L90M	A431	48	181	54.3	63.8	55.9	50.6	34.4	48.8
	K436	4	229	55.1	62.4	57.2	50.6	44.0	47.1
	I437	24	204	67.8	61.9	48.8	50.8	57.5	45.9
	L449	86	139	64.4	57.8	59.4	46.3	44.4	47.0
	S451	56	162	59.7	60.4	51.8	48.2	41.0	47.6
	R452	32	188	51.4	60.8	44.6	50.8	38.8	45.6
	P453	134	86	60.4	62.0	46.3	54.0	47.1	45.3

Supplemental Table III.2 *(cont.)*

I84V	A431	1957	1703	39.2	44.9	31.8	36.8	31.3	35.0
	K436	229	3514	35.7	42.5	28.3	34.7	29.3	33.6
	I437	495	3290	43.1	41.8	37.2	33.7	31.7	33.5
	L449	1276	2354	43.5	40.2	35.3	33.0	34.3	32.3
	S451	703	2970	42.5	41.4	33.2	34.1	35.1	32.5
	R452	649	3063	35.6	43.4	26.9	35.7	30.3	33.8
	P453	2948	684	41.7	42.3	33.9	34.7	33.3	33.3
I84V/L90M	A431	1439	1262	39.9	44.0	32.4	36.2	31.5	34.0
	K436	167	2607	35.5	42.4	27.6	34.9	28.5	33.1
	I437	347	2454	40.9	42.1	35.5	34.4	30.4	33.2
	L449	837	1843	43.5	40.5	35.6	33.2	33.1	32.4
	S451	491	2230	42.0	41.5	33.8	34.2	34.0	32.3
	R452	517	2226	35.1	43.6	26.7	36.3	28.2	33.4
	P453	2255	444	41.0	45.2	33.7	36.1	32.2	34.9
L90M	A431	602	2220	36.7	60.6	27.6	49.8	32.1	46.7
	K436	189	2584	47.7	56.1	35.4	44.7	40.1	45.2
	I437	305	2481	44.4	56.7	36.0	45.3	36.1	45.9
	L449	297	2494	48.4	56.6	39.5	45.0	39.5	45.5
	S451	330	2419	56.0	54.7	42.8	43.9	45.5	44.4
	R452	50	2795	47.0	55.7	37.2	44.3	35.7	44.9
	P453	722	1902	49.9	57.3	38.5	46.5	42.6	45.5

Supplemental Table III.2: Replicative Capacities for associating mutations within the protease and NC-p1 and p1-p6 cle

P-P -- Samples with specified mutations present in both the protease and the cleavage site

P-A -- Samples with specified mutations present in the protease but absent the in cleavage site

p-values < 0.05 are significant

Supplemental Table III.3

Gag Residue (Cleavage Site)	Gag Mutation	Protease Mutations	P-P %	A-P %	P-P	P-A	A-P	A-A	Chi2	p-value
A431 (NC-p1)	A > V	D30N	3	19	4	141	6589	28134	23.7	1.9E-02
		D30N/N88D	12	19	133	975	6460	27300	35.1	5.2E-05
		D30N/N88D/L90M	21	19	40	149	6434	27141	0.4	1.0E+00
		I50L	40	19	64	95	6738	29150	46.3	1.7E-07
		I50V	10	19	63	538	7092	29576	29.3	1.0E-03
		V82A	52	15	1986	1844	4722	26984	3046.1	<10.0E-12
		V82A/L90M	45	17	969	1167	5614	26624	1009.0	<10.0E-12
		I84V	53	14	2161	1921	4456	26551	3504.2	<10.0E-12
		I84V/L90M	53	16	1582	1429	4914	26030	2381.3	<10.0E-12
		N88D	12	19	149	1085	6447	27288	38.1	1.1E-05
		N88D/L90M	20	19	51	210	6426	27167	0.0	1.0E+00
		L90M	21	19	669	2487	5805	24803	9.1	1.0E+00
K436 (NC-p1)	K > R	D30N	3	6	5	140	2008	32160	1.1	1.0E+00
		D30N/N88D	2	6	17	1078	1996	31222	37.3	1.7E-05
		D30N/N88D/L90M	3	6	5	187	1970	31042	3.3	1.0E+00
		I50L	10	6	16	142	2066	33324	4.5	1.0E+00
		I50V	12	6	67	505	2121	34155	33.7	1.1E-04
		V82A	11	5	416	3314	1641	29609	208.6	<10.0E-12
		V82A/L90M	12	6	252	1822	1764	29977	149.8	<10.0E-12
		I84V	6	6	263	3913	1765	28594	1.5	1.0E+00
		I84V/L90M	6	6	192	2902	1795	28511	0.4	1.0E+00
		N88D	1	6	18	1200	2000	31195	43.2	8.3E-07
		N88D/L90M	2	6	5	258	1972	31057	7.0	1.0E+00
		L90M	7	6	213	2885	1762	28344	5.1	1.0E+00

Supplemental Table III.3 *(cont.)*

I437 (NC-p1)	I > V	D30N	6	9	9	136	3246	31062	1.4	1.0E+00
		D30N/N88D	8	9	87	1005	3168	30193	2.7	1.0E+00
		D30N/N88D/L90M	12	10	22	169	3160	29989	0.7	1.0E+00
		I50L	8	9	14	156	3369	32135	0.2	1.0E+00
		I50V	19	9	115	476	3432	32957	66.3	6.5E-12
		V82A	19	8	709	3033	2619	28779	437.4	<10.0E-12
		V82A/L90M	20	9	423	1653	2828	29062	296.9	<10.0E-12
		I84V	13	9	543	3673	2730	27739	66.1	7.1E-12
		I84V/L90M	12	9	381	2736	2817	27608	28.5	1.6E-03
		N88D	8	9	94	1120	3165	30172	4.0	1.0E+00
		N88D/L90M	10	10	27	232	3158	30009	0.1	1.0E+00
		L90M	11	9	336	2784	2846	27374	5.8	1.0E+00
L449 (p1-p6)	L > F	D30N	7	15	10	132	5240	28699	7.0	1.0E+00
		D30N/N88D	32	15	339	728	4911	28103	225.1	<10.0E-12
		D30N/N88D/L90M	41	15	76	110	5053	27758	89.4	<10.0E-12
		I50L	17	15	30	143	5414	29600	0.3	1.0E+00
		I50V	35	15	200	371	5404	30590	171.9	<10.0E-12
		V82A	10	16	364	3410	5014	25944	109.8	<10.0E-12
		V82A/L90M	10	16	200	1890	5045	26460	61.3	8.2E-11
		I84V	35	13	1416	2634	3880	26359	1337.9	<10.0E-12
		I84V/L90M	31	14	936	2053	4231	25960	617.9	<10.0E-12
		N88D	31	15	365	824	4899	28090	220.0	<10.0E-12
		N88D/L90M	38	15	96	160	5046	27780	93.1	<10.0E-12
		L90M	10	16	328	2800	4801	25068	66.9	4.7E-12

Supplemental Table III.3 *(cont.)*

S451 (p1-p6)	S > N	D30N	26	17	36	102	5571	27130	7.3	1.0E+00
		D30N/N88D	30	17	299	710	5308	26522	115.1	<10.0E-12
		D30N/N88D/L90M	28	17	50	127	5413	26201	14.5	1.2E-02
		I50L	11	17	19	148	5776	28044	3.4	1.0E+00
		I50V	15	17	88	497	5893	28827	1.4	1.0E+00
		V82A	15	17	553	3174	5156	24606	14.3	1.0E+00
		V82A/L90M	14	17	282	1788	5273	25041	19.1	2.0E-01
		I84V	19	17	795	3309	4858	24084	16.8	7.0E-01
		I84V/L90M	18	17	560	2477	4944	23991	3.4	1.0E+00
		N88D	29	17	322	804	5310	26496	107.8	<10.0E-12
		N88D/L90M	25	17	62	183	5424	26204	10.8	1.0E+00
		L90M	12	18	373	2696	5090	23632	60.0	1.6E-10
R452 (p1-p6)	R > S	D30N	6	4	9	131	1420	33353	1.4	1.0E+00
		D30N/N88D	13	4	139	920	1290	32564	224.6	<10.0E-12
		D30N/N88D/L90M	17	4	31	150	1367	32242	74.2	<10.0E-12
		I50L	4	4	7	169	1474	34478	0.0	1.0E+00
		I50V	9	4	54	518	1455	35497	42.8	1.0E-06
		V82A	2	4	87	3808	1365	30345	37.5	1.5E-05
		V82A/L90M	3	4	64	2096	1354	30905	7.5	1.0E+00
		I84V	17	2	724	3419	711	30287	2146.6	<10.0E-12
		I84V/L90M	19	3	576	2483	826	30102	1833.7	<10.0E-12
		N88D	13	4	151	1030	1286	32545	231.8	<10.0E-12
		N88D/L90M	14	4	36	215	1369	32257	63.6	2.6E-11
		L90M	2	4	53	3125	1345	29267	53.3	4.9E-09

Supplemental Table III.3 *(cont.)*

P453 (p1-p6)	P > L	D30N	33	31	42	86	10127	22916	0.2	1.0E+00
		D30N/N88D	66	30	671	351	9498	22651	605.9	<10.0E-12
		D30N/N88D/L90M	65	31	119	65	9830	22127	96.9	<10.0E-12
		I50L	26	31	45	130	10490	23407	2.0	1.0E+00
		I50V	60	30	326	217	10574	24269	219.7	<10.0E-12
		V82A	24	32	861	2788	9526	20609	97.8	<10.0E-12
		V82A/L90M	26	31	531	1484	9612	21071	21.6	5.5E-02
		I84V	81	24	3266	784	6975	22319	5397.7	<10.0E-12
		I84V/L90M	83	26	2507	503	7502	21778	4241.0	<10.0E-12
		N88D	64	29	731	403	9467	22657	629.2	<10.0E-12
		N88D/L90M	61	31	152	99	9823	22144	102.3	<10.0E-12
		L90M	27	31	797	2127	9152	20065	20.4	1.1E-01

Supplemental Table III.3: Association between protease mutations and mutations within the NC-p1 and p1-p6 cleavage si

P-P -- Samples with specified mutations present in both the protease and the cleavage site

P-A -- Samples with specified mutations present in the protease but absent the in cleavage site

p-values < 0.05 are significant

Supplemental Table III.4

Protease Mutations	Gag Residue (Cleavage Site)	PI	P-P	P-A	Mean P-P	Mean P-A	Median P-P	Median P-A	Std dev P-P	Std dev P-A	p-value
D30N/N88D	A431 (NC-p1)	APV	133	975	4.7	14.2	2.7	1.1	38.9	27.1	<10.0E-12
		ATV	133	975	16.9	112.4	20.4	3.9	202.1	68.3	<10.0E-12
		IDV	133	975	5.2	26.8	6.2	2.0	64.1	26.3	<10.0E-12
		LPV	133	975	4.1	21.9	4.7	1.2	53.9	21.2	<10.0E-12
		NFV	133	975	117.2	320.7	199.7	61.8	243.0	159.9	<10.0E-12
		RTV	133	975	9.0	44.5	6.7	1.9	87.2	38.0	<10.0E-12
		SQV	133	975	27.3	196.8	9.1	2.7	372.1	135.1	<10.0E-12
		TPV	84	501	1.5	5.0	1.5	0.9	11.0	4.3	6.3E-07
D30N/N88D	I437 (NC-p1)	APV	87	1005	9.1	5.1	1.9	1.1	22.7	26.8	2.1E-05
		ATV	87	1005	42.7	26.6	6.4	4.2	127.6	94.8	9.0E-05
		IDV	86	1004	17.8	6.7	3.3	2.2	61.9	27.8	3.0E-06
		LPV	87	1005	9.6	5.6	2.3	1.3	25.0	26.8	3.9E-06
		NFV	87	1005	221.7	137.2	93.7	66.7	228.4	181.2	1.3E-04
		RTV	87	1005	21.9	12.4	3.3	2.1	62.7	45.4	2.2E-05
		SQV	87	1005	64.4	47.9	5.8	2.9	210.7	190.5	8.3E-05
		TPV	45	526	5.3	1.7	1.1	0.9	13.2	4.4	1.3E-02
D30N/N88D	L449 (p1-p6)	APV	339	728	4.4	8.4	1.4	1.1	34.2	22.6	3.5E-05
		ATV	339	728	23.2	39.4	4.7	4.2	120.7	86.8	1.0E+00
		IDV	339	728	5.9	11.0	2.6	2.0	40.8	27.2	1.1E-06
		LPV	339	728	5.6	6.6	1.5	1.3	19.3	25.8	3.4E-01
		NFV	339	728	131.6	170.9	77.5	65.1	204.9	175.7	6.6E-02
		RTV	339	728	12.2	14.9	2.2	2.1	47.1	45.7	1.0E+00
		SQV	339	728	39.3	72.0	3.5	2.8	234.8	171.2	7.8E-03
		TPV	178	389	1.7	2.6	0.9	0.9	8.0	4.8	1.0E+00
D30N/N88D/L90M	A431 (NC-p1)	APV	40	149	28.3	21.3	9.5	2.4	63.5	66.6	1.1E-02
		ATV	40	149	258.2	52.1	89.3	13.2	271.1	120.1	3.8E-09
		IDV	39	148	70.5	17.4	32.2	4.7	104.5	56.7	5.1E-10
		LPV	40	149	45.5	15.5	26.0	3.5	80.0	49.0	3.3E-08
		NFV	40	149	555.1	305.9	600.0	172.1	137.1	236.7	6.6E-06
		RTV	40	149	121.4	38.3	74.5	12.1	127.1	79.0	6.9E-08
		SQV	40	149	572.2	135.8	1000.0	27.7	459.2	292.7	3.2E-07
		TPV	27	80	11.7	3.8	4.0	1.2	17.9	10.0	1.6E-04

Supplemental Table III.4 *(cont.)*

D30N/N88D/L90M	L449 (p1-p6)	APV	76	110	25.3	17.9	3.3	2.7	67.2	55.1	1.0E+00
		ATV	76	110	103.7	91.6	30.4	20.4	188.4	182.6	1.0E+00
		IDV	75	109	35.5	23.3	7.7	5.2	80.8	66.7	1.0E+00
		LPV	76	110	20.6	18.7	6.6	4.2	36.6	52.6	1.0E+00
		NFV	76	110	391.4	341.0	600.0	214.8	242.6	241.2	1.0E+00
		RTV	76	110	51.9	56.8	27.8	18.3	88.6	98.8	1.0E+00
		SQV	76	110	263.7	217.7	43.1	35.2	401.4	373.0	1.0E+00
		TPV	44	64	6.7	5.1	1.2	1.7	15.0	11.1	1.0E+00
D30N/N88D/L90M	S451 (p1-p6)	APV	50	127	17.5	20.5	2.6	3.9	57.6	53.6	1.0E+00
		ATV	50	127	101.7	96.6	20.1	26.2	176.3	186.8	1.0E+00
		IDV	49	127	25.0	31.6	6.5	6.5	60.8	79.2	1.0E+00
		LPV	50	127	25.6	17.3	4.1	6.4	73.8	31.7	1.0E+00
		NFV	50	127	343.7	372.3	205.5	600.0	241.9	244.8	1.0E+00
		RTV	50	127	65.1	53.0	18.9	20.5	99.7	95.6	1.0E+00
		SQV	50	127	262.7	223.2	42.1	34.4	399.3	379.0	1.0E+00
		TPV	25	77	4.9	6.5	1.2	1.6	10.9	13.8	1.0E+00
D30N/N88D/L90M	R452 (p1-p6)	APV	31	150	44.0	18.9	8.0	2.8	99.6	57.9	1.0E+00
		ATV	31	150	210.1	73.7	58.0	16.7	296.0	146.5	1.0E+00
		IDV	31	148	96.8	14.4	20.6	5.0	154.8	22.7	1.0E+00
		LPV	31	150	56.7	15.2	18.4	4.5	99.5	44.1	1.0E+00
		NFV	31	150	424.9	339.9	600.0	219.5	241.3	241.3	1.0E+00
		RTV	31	150	120.5	42.1	43.4	17.2	158.1	70.9	1.0E+00
		SQV	31	150	301.4	215.0	59.1	33.1	421.3	371.9	1.0E+00
		TPV	18	82	22.6	2.3	4.0	1.3	25.0	3.0	6.5E-02
D30N/N88D/L90M	P453 (p1-p6)	ATV	119	65	111.0	72.5	31.6	15.7	200.7	153.3	1.0E+00
		APV	119	65	26.1	17.9	3.2	2.9	73.4	52.5	1.0E+00
		IDV	118	64	39.0	10.2	6.8	6.0	89.2	10.1	1.0E+00
		LPV	119	65	27.7	12.1	6.8	3.7	70.8	22.0	1.0E+00
		NFV	119	65	372.5	338.4	600.0	215.6	241.4	240.9	1.0E+00
		RTV	119	65	63.5	43.1	22.2	17.8	107.5	77.5	1.0E+00
		SQV	119	65	272.0	142.8	39.8	33.6	408.2	302.3	1.0E+00
		TPV	75	34	7.4	2.1	1.6	1.1	14.9	3.7	5.2E-01
I50V	K436 (NC-p1)	APV	67	505	108.1	97.5	55.0	47.4	125.7	126.4	1.0E+00
		ATV	67	505	48.9	27.9	16.1	6.8	74.4	80.9	9.7E-02
		IDV	67	505	96.4	37.2	32.5	10.1	134.2	80.7	1.5E-03
		LPV	67	505	273.4	171.3	199.6	106.3	200.0	179.1	5.7E-03
		NFV	67	505	93.1	40.1	36.7	14.6	154.9	76.6	3.9E-02
		RTV	67	505	275.9	204.1	400.0	131.5	157.6	175.1	1.3E-01
		SQV	67	505	453.5	244.0	103.2	14.1	482.7	408.2	4.7E-02
		TPV	48	325	1.4	1.5	0.9	0.7	1.6	4.3	1.0E+00

Supplemental Table III.4 *(cont.)*

I50V	I437 (NC-p1)	APV	115	476	124.0	87.7	71.4	43.9	133.2	117.7	5.2E-05
		ATV	115	476	65.7	19.1	29.5	5.5	119.7	52.2	<10.0E-12
		IDV	115	468	119.8	24.7	47.3	8.8	151.1	51.4	<10.0E-12
		LPV	115	476	286.4	156.2	233.4	96.0	196.5	169.0	1.2E-10
		NFV	115	476	104.4	30.2	56.1	13.6	156.7	48.4	<10.0E-12
		RTV	115	476	305.0	191.4	400.0	106.6	147.5	173.3	2.6E-10
		SQV	115	476	516.7	197.6	169.2	11.1	482.5	374.1	<10.0E-12
		TPV	68	313	2.4	1.2	1.1	0.6	6.5	3.2	8.5E-04
I50V	L449 (p1-p6)	APV	200	371	93.0	100.2	43.4	52.7	125.8	123.8	1.0E+00
		ATV	200	371	25.4	31.9	5.4	7.9	74.1	82.5	1.0E+00
		IDV	200	371	30.8	51.5	7.0	13.3	70.8	100.7	3.7E-03
		LPV	200	371	147.4	200.6	65.4	132.3	180.4	180.2	7.9E-04
		NFV	200	371	33.7	52.2	12.7	18.7	59.6	102.4	1.0E+00
		RTV	200	371	163.3	241.5	77.3	400.0	166.7	171.1	4.4E-05
		SQV	200	371	193.7	297.1	10.6	19.7	368.4	437.9	1.0E+00
		TPV	130	242	1.5	1.4	0.6	0.7	4.8	3.6	1.0E+00
I50L	A431 (NC-p1)	APV	62	94	5.5	2.0	2.6	1.0	9.5	2.8	9.3E-04
		ATV	62	94	123.4	42.8	71.5	22.8	166.6	76.5	8.5E-05
		IDV	62	94	3.9	1.3	1.8	0.7	6.7	2.5	5.9E-06
		LPV	62	94	4.8	0.9	1.6	0.5	10.4	1.6	2.4E-04
		NFV	62	94	9.9	3.5	4.3	1.6	17.5	4.9	1.9E-04
		RTV	62	94	9.0	1.9	2.5	0.8	14.2	3.2	4.3E-03
		SQV	62	94	3.1	1.6	1.1	0.7	6.1	2.3	1.0E+00
		TPV	45	60	2.5	1.3	1.2	0.7	4.7	1.8	1.0E+00
V82A	A431 (NC-p1)	APV	1986	1841	31.6	28.9	9.3	8.7	64.2	66.9	1.0E+00
		ATV	1986	1841	74.6	64.0	19.4	20.0	125.2	144.0	1.0E+00
		IDV	1986	1841	41.4	47.9	26.6	20.1	70.1	70.2	1.1E-12
		LPV	1986	1841	90.4	102.2	64.8	45.6	117.5	121.7	1.1E-12
		NFV	1986	1841	65.9	68.5	36.3	31.0	102.2	105.7	2.5E-01
		RTV	1986	1841	201.1	184.1	101.2	120.8	163.1	171.1	1.0E+00
		SQV	1986	1841	96.6	65.6	9.0	9.9	207.3	257.7	1.0E+00
		TPV	1243	1109	4.4	2.9	1.6	2.2	4.4	6.6	<10.0E-12
V82A	I437 (NC-p1)	APV	709	3033	37.4	28.7	14.3	8.4	71.9	64.8	<10.0E-12
		ATV	708	3031	109.3	58.0	38.7	16.6	176.2	118.8	<10.0E-12
		IDV	699	3010	64.5	40.5	33.7	21.5	91.8	64.5	<10.0E-12
		LPV	709	3033	130.2	89.1	83.3	51.7	143.6	114.1	<10.0E-12
		NFV	709	3033	97.2	59.2	51.7	30.5	135.9	92.9	<10.0E-12
		RTV	709	3033	237.6	178.7	342.7	93.2	167.3	164.8	<10.0E-12
		SQV	709	3033	147.6	63.9	16.2	7.9	321.1	204.0	<10.0E-12
		TPV	460	1840	5.2	3.2	2.7	1.7	7.7	5.0	<10.0E-12

Supplemental Table III.4 *(cont.)*

V82A	L449 (p1-p6)	APV	364	3407	29.1	37.1	14.5	8.6	68.2	65.5	6.8E-06
		ATV	364	3407	62.7	99.7	39.5	17.4	152.4	127.4	4.8E-10
		IDV	364	3407	40.0	75.5	37.9	22.1	107.7	61.6	1.2E-10
		LPV	364	3407	91.0	129.0	84.3	52.9	137.4	115.9	3.9E-09
		NFV	364	3407	59.8	119.1	59.0	30.9	163.5	91.9	<10.0E-12
		RTV	364	3407	182.6	243.4	400.0	96.1	169.3	165.5	1.6E-07
		SQV	364	3407	69.0	160.9	24.6	8.3	326.8	214.4	<10.0E-12
		TPV	237	2063	3.4	4.4	2.6	1.8	5.3	5.5	1.1E-04
V82A	P453 (p1-p6)	APV	861	2785	28.3	34.8	11.9	8.4	70.0	64.3	4.4E-07
		ATV	861	2785	65.5	75.3	27.2	17.3	137.7	131.1	3.7E-04
		IDV	861	2785	41.0	56.3	27.8	22.0	86.9	63.6	7.0E-06
		LPV	861	2785	91.7	110.7	70.2	51.1	125.6	117.8	2.6E-07
		NFV	861	2785	61.7	83.0	44.1	31.1	124.0	95.3	2.6E-07
		RTV	861	2785	179.2	223.9	171.0	93.2	168.1	165.3	2.0E-10
		SQV	861	2785	77.3	91.5	14.4	8.1	244.6	230.3	5.2E-08
		TPV	543	1689	3.5	3.7	2.2	1.7	5.5	5.5	7.0E-03
V82A/L90M	A431 (NC-p1)	APV	969	1165	42.8	41.1	13.6	15.8	81.6	78.3	1.0E+00
		ATV	969	1165	101.4	84.0	35.5	42.4	141.8	163.4	1.0E+00
		IDV	969	1165	52.9	61.6	34.6	28.2	82.1	77.4	4.5E-04
		LPV	969	1165	115.2	121.4	81.4	68.0	130.1	134.3	5.1E-01
		NFV	969	1165	88.4	94.7	55.4	53.1	122.9	121.1	1.0E+00
		RTV	969	1165	264.7	233.3	192.9	400.0	160.9	160.5	1.3E-02
		SQV	969	1165	122.9	97.9	19.6	21.4	249.1	281.8	1.0E+00
		TPV	610	710	5.6	4.3	2.6	3.1	5.6	7.7	3.9E-02
V82A/L90M	I437 (NC-p1)	APV	423	1653	48.4	40.7	19.7	13.2	85.5	79.7	3.7E-05
		ATV	423	1651	130.1	82.0	59.5	34.2	187.1	141.0	3.3E-10
		IDV	416	1639	75.1	52.5	39.8	29.4	99.6	74.1	3.5E-07
		LPV	423	1653	147.6	111.1	98.8	68.7	151.9	128.4	5.2E-07
		NFV	423	1653	120.0	82.6	69.8	50.2	151.8	111.5	2.4E-08
		RTV	423	1653	284.1	238.0	400.0	230.2	154.9	162.3	3.9E-07
		SQV	423	1653	284.1	238.0	400.0	230.2	154.9	162.3	3.9E-07
		TPV	282	1005	6.9	4.5	3.9	2.5	9.1	6.2	1.8E-09
V82A/L90M	L449 (p1-p6)	APV	200	1888	40.7	52.1	28.1	12.9	82.5	80.2	5.9E-06
		ATV	200	1888	85.0	129.5	80.3	34.6	160.4	146.7	5.0E-09
		IDV	200	1888	50.8	93.4	48.0	29.4	114.3	70.8	1.5E-09
		LPV	200	1888	111.1	157.5	114.6	68.9	145.1	129.0	2.8E-08
		NFV	200	1888	82.0	154.9	89.1	48.7	178.1	110.6	<10.0E-12
		RTV	200	1888	240.1	311.6	400.0	242.9	142.5	162.6	1.6E-06
		SQV	200	1888	95.0	233.5	50.5	17.5	375.7	244.9	<10.0E-12
		TPV	127	1148	4.8	6.1	4.6	2.7	6.3	6.8	6.6E-03

Supplemental Table III.4 *(cont.)*

V82A/L90M	P453 (p1-p6)	APV	531	1482	40.8	43.9	17.7	13.0	78.0	81.2	5.5E-02
		ATV	531	1482	89.2	94.8	42.4	36.5	154.9	148.3	1.0E+00
		IDV	531	1482	52.6	68.7	34.9	29.9	96.7	72.8	1.4E-01
		LPV	531	1482	112.3	132.9	86.0	68.7	137.2	130.7	1.8E-03
		NFV	531	1482	85.2	107.6	60.0	51.7	144.1	112.6	3.5E-01
		RTV	531	1482	238.1	276.8	400.0	228.5	155.1	162.8	5.9E-04
		SQV	531	1482	107.8	121.9	25.9	18.0	276.7	264.3	3.3E-02
		TPV	327	912	5.1	4.8	3.0	2.7	6.4	6.9	1.0E+00
I84V	A431 (NC-p1)	APV	2157	1919	89.4	66.7	38.4	18.9	124.5	115.5	<10.0E-12
		ATV	2157	1919	178.1	114.6	79.3	42.9	226.2	185.9	<10.0E-12
		IDV	2157	1919	77.0	43.9	44.9	19.9	98.6	73.3	<10.0E-12
		LPV	2157	1919	125.9	80.7	84.4	32.7	130.7	122.0	<10.0E-12
		NFV	2157	1919	117.3	67.5	66.8	37.8	154.1	102.7	<10.0E-12
		RTV	2157	1919	239.8	180.7	249.8	90.6	163.2	167.9	<10.0E-12
		SQV	2157	1919	403.7	302.9	93.6	60.2	454.5	419.0	1.4E-12
		TPV	1515	1174	16.4	12.6	7.5	4.6	18.0	17.0	<10.0E-12
I84V	K436 (NC-p1)	APV	263	3907	110.8	75.9	36.1	28.3	148.9	118.2	3.7E-01
		ATV	263	3907	202.0	145.7	82.6	60.9	252.0	208.8	1.6E-01
		IDV	263	3907	79.3	60.5	39.3	31.7	107.9	88.2	1.0E+00
		LPV	263	3907	128.5	102.9	74.5	60.0	152.6	127.6	1.0E+00
		NFV	263	3907	125.5	92.4	58.5	51.2	171.0	133.4	1.0E+00
		RTV	263	3907	254.7	207.9	400.0	131.9	169.5	167.5	4.6E-02
		SQV	263	3907	357.9	355.3	67.8	74.1	446.7	440.5	1.0E+00
		TPV	185	2568	16.2	14.5	7.7	5.9	18.0	18.0	1.0E+00
I84V	I437 (NC-p1)	APV	543	3673	99.1	74.8	43.8	26.5	130.7	118.7	3.0E-12
		ATV	543	3667	192.9	143.1	84.1	58.2	236.2	207.9	5.2E-11
		IDV	533	3636	84.3	58.3	40.6	30.9	112.2	84.9	1.4E-09
		LPV	543	3673	136.5	99.7	82.6	56.9	148.1	125.7	<10.0E-12
		NFV	543	3673	118.2	92.0	66.6	50.4	154.2	135.0	1.2E-10
		RTV	543	3672	261.8	202.6	400.0	123.6	161.9	167.7	<10.0E-12
		SQV	543	3673	413.8	345.8	93.3	71.0	459.0	437.3	8.7E-04
		TPV	392	2396	21.3	13.5	11.8	5.2	20.3	16.8	<10.0E-12
I84V	R452 (p1-p6)	APV	724	3413	92.3	75.6	39.8	26.5	127.7	119.4	3.1E-08
		ATV	724	3413	176.1	144.6	84.8	56.8	219.0	211.9	2.0E-10
		IDV	724	3413	81.9	57.9	46.2	30.3	103.6	86.6	<10.0E-12
		LPV	724	3413	134.3	98.8	84.1	55.6	144.7	126.2	<10.0E-12
		NFV	724	3413	128.8	88.6	68.5	49.1	164.9	130.6	0.0E+00
		RTV	724	3413	237.8	205.3	242.0	128.0	165.1	168.0	3.2E-05
		SQV	724	3413	466.5	329.2	132.8	65.2	464.1	431.5	<10.0E-12
		TPV	504	2240	16.1	14.5	7.1	5.6	18.0	18.0	1.0E+00

Supplemental Table III.4 *(cont.)*

I84V	P453 (p1-p6)	APV	3264	780	81.4	68.3	31.2	24.1	122.8	112.4	2.6E-02
		ATV	3264	780	154.4	142.2	62.8	53.4	215.5	211.9	1.0E+00
		IDV	3264	780	63.9	57.7	33.3	29.8	92.5	86.4	1.0E+00
		LPV	3264	780	107.1	101.7	62.5	59.4	132.0	127.6	1.0E+00
		NFV	3264	780	101.7	78.3	55.1	41.8	143.5	124.0	3.2E-08
		RTV	3264	780	209.1	222.8	134.8	182.3	167.6	170.1	1.0E+00
		SQV	3264	780	372.6	276.0	82.0	48.2	445.1	409.7	3.9E-12
		TPV	2160	516	14.2	17.1	5.7	8.2	17.0	19.0	3.8E-01
I84V/L90M	A431 (NC-p1)	APV	1580	1427	98.4	76.0	42.0	22.6	131.1	123.7	<10.0E-12
		ATV	1580	1427	205.6	131.3	103.8	54.7	263.4	195.6	<10.0E-12
		IDV	1580	1427	90.0	51.0	53.4	24.6	105.6	77.5	<10.0E-12
		LPV	1580	1427	136.5	91.7	92.7	40.3	135.1	129.6	<10.0E-12
		NFV	1580	1427	137.1	81.1	79.8	50.5	162.4	112.2	<10.0E-12
		RTV	1580	1427	263.3	202.9	400.0	123.5	156.0	168.0	<10.0E-12
		SQV	1580	1427	497.0	374.7	172.7	85.8	462.3	440.9	<10.0E-12
		TPV	1090	857	16.1	12.4	7.7	4.7	18.0	16.0	5.6E-10
I84V/L90M	K436 (NC-p1)	APV	192	2898	126.7	84.1	39.9	32.7	159.0	125.0	1.0E+00
		ATV	192	2898	236.5	166.7	114.1	72.7	260.1	219.2	2.0E-02
		IDV	192	2898	91.5	70.0	51.8	38.8	110.1	93.9	1.3E-01
		LPV	192	2898	145.9	112.8	85.3	67.7	161.1	132.7	1.0E+00
		NFV	192	2898	152.9	108.3	86.1	64.0	182.9	142.0	6.1E-02
		RTV	192	2898	280.9	230.1	400.0	192.9	161.5	164.6	4.9E-02
		SQV	192	2898	448.2	435.1	110.6	115.1	464.5	455.5	1.0E+00
		TPV	134	1871	17.4	14.2	8.3	5.8	19.0	17.0	1.0E+00
I84V/L90M	I437 (NC-p1)	APV	381	2736	108.8	83.9	47.7	31.0	136.4	126.2	1.7E-08
		ATV	381	2732	220.9	164.7	107.5	71.0	246.4	218.1	3.0E-08
		IDV	375	2736	94.8	68.0	47.4	38.5	118.8	90.4	4.1E-05
		LPV	381	2736	140.4	111.3	88.5	65.4	145.7	132.6	2.4E-07
		NFV	381	2736	141.2	108.1	86.0	63.0	169.4	142.8	1.0E-07
		RTV	381	2735	284.1	225.7	400.0	177.2	153.3	165.3	8.3E-11
		SQV	381	2736	504.3	426.7	156.2	111.5	466.7	454.1	9.4E-03
		TPV	268	1758	19.1	13.5	10.2	5.5	19.3	16.8	2.2E-08
I84V/L90M	R452 (p1-p6)	APV	576	2479	96.6	85.9	42.8	31.4	131.1	127.8	2.9E-03
		ATV	576	2479	194.8	167.0	99.6	69.7	225.8	223.0	1.8E-05
		IDV	576	2479	92.0	67.3	54.6	37.0	108.0	92.0	8.8E-11
		LPV	576	2479	138.1	110.6	88.7	64.0	142.8	133.7	7.3E-08
		NFV	576	2479	143.9	105.4	81.8	62.8	171.5	139.8	2.3E-09
		RTV	576	2479	255.4	229.2	400.0	192.2	160.4	165.6	4.7E-02
		SQV	576	2479	546.8	408.3	1000.0	100.5	463.3	450.2	5.3E-10
		TPV	407	1587	15.8	14.2	6.9	5.7	18.0	17.0	1.0E+00

Supplemental Table III.4 *(cont.)*

I84V/L90M	P453 (p1-p6)	APV	2505	501	89.3	81.5	34.1	29.0	129.2	123.7	1.0E+00
		ATV	2505	501	176.4	162.0	77.3	70.2	226.1	222.0	1.0E+00
		IDV	2505	501	73.6	67.7	40.4	36.9	98.3	93.3	1.0E+00
		LPV	2505	501	115.9	117.9	70.2	64.9	135.1	142.3	1.0E+00
		NFV	2505	501	116.7	98.4	66.6	57.4	151.0	139.5	1.3E-02
		RTV	2505	501	230.9	247.6	193.1	400.0	164.3	167.3	1.0E+00
		SQV	2505	501	446.0	364.3	123.4	71.2	456.6	443.1	4.4E-05
		TPV	1639	316	13.9	16.9	5.8	7.5	17.0	19.0	1.0E+00
N88D	A431 (NC-p1)	APV	149	1085	14.1	5.1	3.0	1.1	37.5	28.5	<10.0E-12
		ATV	149	1085	106.7	17.2	23.5	3.9	192.2	65.6	<10.0E-12
		IDV	149	1085	27.3	5.8	7.0	2.1	61.0	25.5	<10.0E-12
		LPV	149	1085	25.5	5.6	5.7	1.3	53.6	27.0	<10.0E-12
		NFV	149	1085	293.0	109.0	157.9	57.5	243.5	154.9	<10.0E-12
		RTV	149	1085	58.2	13.4	10.5	2.1	105.9	51.8	<10.0E-12
		SQV	149	1085	185.0	30.2	10.9	2.8	354.3	141.6	<10.0E-12
		TPV	92	554	5.1	1.9	1.7	0.9	11.0	5.4	<10.0E-12
N88D	A449 (p1-p6)	APV	365	824	8.26	4.96	1.46	1.15	33.1	25.7	1.8E-06
		ATV	365	824	40.1	22.72	5.05	4.21	117.4	82.2	4.5E-02
		IDV	365	824	11.89	6.51	2.72	2.17	39.9	26.1	1.3E-06
		LPV	365	824	8.68	7.44	1.61	1.37	22.1	32.0	1.2E-01
		NFV	365	824	164.59	119.52	76.04	59.48	201.1	168.8	3.6E-05
		RTV	365	824	23.16	16.9	2.37	2.26	68.5	58.6	1.0E+00
		SQV	365	824	76.19	39.72	3.85	2.84	237.3	168.4	1.7E-04
		TPV	365	824	76.19	39.72	3.85	2.84	8.7	5.4	1.7E-04
N88D/L90M	A431 (NC-p1)	APV	51	210	26.0	18.4	7.6	2.4	57.7	62.6	3.9E-04
		ATV	51	210	217.9	46.1	81.9	13.4	252.5	104.1	1.5E-11
		IDV	51	210	62.9	17.3	31.3	5.2	93.7	48.6	2.0E-11
		LPV	51	210	48.7	20.1	31.4	4.0	73.1	56.7	2.1E-09
		NFV	51	210	451.0	230.0	600.0	113.6	234.4	236.2	7.7E-05
		RTV	51	210	139.0	53.1	93.1	15.7	138.7	101.6	7.0E-09
		SQV	51	210	471.0	125.0	177.2	24.8	451.9	279.7	2.7E-08
		TPV	32	115	11.0	4.5	4.5	1.3	17.0	10.0	3.7E-04
L90M	A431 (NC-p1)	APV	669	2486	44.1	17.2	14.67	1.7	85.3	58.7	<10.0E-12
		ATV	669	2486	77.76	25.62	28.76	3.04	140.7	81.0	<10.0E-12
		IDV	669	2486	78.48	16.52	37.8	3.6	107.4	47.9	<10.0E-12
		LPV	669	2486	108.87	25.56	62.7	1.77	136.1	75.2	<10.0E-12
		NFV	669	2486	114.55	29.7	55.56	10.08	155.6	73.8	<10.0E-12
		RTV	669	2486	200.53	49.64	124.87	5.72	174.7	114.4	<10.0E-12
		SQV	669	2486	79.98	26.4	14.3	3.14	225.4	126.3	<10.0E-12
		TPV	424	1396	10.31	4.59	4.08	1.26	15.0	10.0	<10.0E-12

Supplemental Table III.4 *(cont.)*

L90M	K436 (NC-p1)	APV	213	2884	44.8	20.88	7.2	2.01	96.2	62.3	<10.0E-12
		ATV	213	2884	56.97	34.66	14.37	4.15	111.7	95.7	<10.0E-12
		IDV	213	2884	47.38	27.8	17.16	4.81	84.5	67.7	<10.0E-12
		LPV	213	2884	69.62	40.46	20.65	2.22	107.9	95.4	<10.0E-12
		NFV	213	2884	63.37	45.94	29.69	13.17	104.0	100.9	<10.0E-12
		RTV	213	2884	134.29	77.05	45.19	7.09	161.7	138.7	<10.0E-12
		SQV	213	2884	56.83	35.48	9.96	3.7	189.3	149.6	4.1E-09
		TPV	138	1661	10.22	5.59	2.54	1.37	16.0	11.0	2.5E-06
L90M	L449 (p1-p6)	APV	328	2799	38.49	21.45	4.28	2.09	88.1	63.3	2.0E-04
		ATV	328	2799	69.05	33.06	8.48	4.54	147.0	90.3	8.0E-04
		IDV	328	2799	50.94	27.43	7.72	5.14	103.2	64.4	5.1E-02
		LPV	328	2799	69.26	40.93	4.16	2.46	133.1	93.6	1.0E-01
		NFV	328	2799	81.39	44.12	23.25	13.76	148.3	95.0	1.6E-06
		RTV	328	2799	115.37	78.65	12.57	7.77	165.1	138.6	2.8E-01
		SQV	328	2799	66.24	33.22	6.67	3.85	214.1	141.5	2.2E-04
		TPV	199	1612	6.97	5.86	1.79	1.45	12.0	12.0	1.0E+00
L90M	S451 (p1-p6)	APV	373	2695	17.61	23.61	2.8	2.15	47.9	67.9	1.0E+00
		ATV	373	2695	41.45	36.94	7.72	4.59	103.9	100.0	7.6E-03
		IDV	373	2695	34.59	29.71	6.75	5.21	83.2	69.3	4.7E-01
		LPV	373	2695	45.13	43.33	3.57	2.45	107.9	96.3	1.0E+00
		NFV	373	2695	53.35	47.84	16.66	14.08	106.8	103.4	3.2E-01
		RTV	373	2695	86.9	81.67	12.54	7.85	145.1	141.4	3.6E-01
		SQV	373	2695	28.79	37.95	5.66	3.94	105.8	156.6	1.2E-02
		TPV	210	1549	4.98	6.09	1.6	1.42	9.4	12.0	1.0E+00
L90M	R452 (p1-p6)	APV	53	3124	44.18	22.55	6.4	2.14	93.2	65.7	1.2E-01
		ATV	53	3124	64.34	36.16	13.17	4.66	139.5	97.1	3.3E-01
		IDV	53	3124	66.57	29.15	9.8	5.19	119.5	68.4	7.2E-01
		LPV	53	3124	80.87	42.65	14.14	2.52	138.5	96.1	6.1E-01
		NFV	53	3124	68.01	47.3	27.83	14.25	120.6	101.1	1.0E+00
		RTV	53	3124	146.54	80.7	37.2	8.02	178.5	140.7	7.9E-01
		SQV	53	3124	49.2	36.69	11.36	3.95	140.1	152.1	2.3E-02
		TPV	32	1808	8.53	5.85	3.39	1.44	11.0	11.0	1.0E+00
L90M	P453 (p1-p6)	APV	797	2127	38.45	16.64	3.86	1.94	90.5	50.5	<10.0E-12
		ATV	797	2127	54.27	32.29	10.37	3.84	122.3	93.6	<10.0E-12
		IDV	797	2127	40.71	27.49	8.49	4.79	84.8	67.2	3.7E-10
		LPV	797	2127	58.59	38.4	4.71	2.25	112.1	91.8	2.0E-09
		NFV	797	2127	70.62	41.41	21.44	12.58	132.1	92.0	0.0E+00
		RTV	797	2127	101.3	75.45	14.94	7.01	151.7	137.0	3.0E-08
		SQV	797	2127	53.7	33.83	7.05	3.37	180.4	149.8	0.0E+00
		TPV	502	1195	7.21	5.56	1.76	1.38	13.0	11.0	1.6E-02

Supplemental Table III.4: Drug susceptibility data for associating mutations within the protease and NC-p1 and p1-p6 cleavage sites
P-P -- Samples with specified mutations present in both the protease and the cleavage site
P-A -- Samples with specified mutations present in the protease but absent the in cleavage site
p-values < 0.05 are significant

References:

Bally, F., Martinez, R., Peters, S., Sudre, P., and Telenti, A. (2000). Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: Clinical significance and implications for resistance to protease. *AIDS Res Hum Retrovir* **16**(13)**,** 1209-1213.

Boden, D., and Markowitz, M. (1998). Minireview: Resistance to human immunodeficiency virus type 1 protease inhibitors. *Antimicrob Agents Chemother* **42,** 2775-2783.

Brenner, B. G., Routy, J. P., Petrella, M., Moisi, D., Oliveira, M., Detorio, M., Spira, B., Essabag, V., Conway, B., Lalonde, R., Sekaly, R. P., and Wainberg, M. A. (2002). Persistence and fitness of multidrug-resistant human immunodeficiency virus type 1 acquired in primary infection. *J Virol* **76**(4)**,** 1753-61.

Colonno, R., Rose, R., McLaren, C., Thiry, A., Parkin, N., and Friborg, J. (2004). Identification of I50L as the signature atazanavir (ATV)-resistance mutation in treatment-naive HIV-1-infected patients receiving ATV-containing regimens. *J Infect Dis* **189**(10)**,** 1802-10.

Croteau, G., Doyon, L., Thibeault, D., McKercher, G., Pilote, L., and Lamarre, D. (1997). Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. *J Virol* **71**(2)**,** 1089-96.

Dam, E., Quercia, R., Glass, B., Descamps, D., Launay, O., Duval, X., Krausslich, H. G., Hance, A. J., and Clavel, F. (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**(3)**,** e1000345.

Deeks, S. G., Wrin, T., Liegler, T., Hoh, R., Hayden, M., Barbour, J. D., Hellmann, N. S., Petropoulos, C. J., McCune, J. M., Hellerstein, M. K., and Grant, R. M. (2001). Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia. *New England Journal of Medicine* **344**(7)**,** 472-80.

The PyMol Molecular Graphics System: DeLano, W.L. The PyMOL Molecular Graphics System. (2008) DeLano Scientific LLC, Palo Alto, CA, USA. http://www.pymol.org

Delaugerre, C., Mathez, D., Peytavin, G., Berthe, H., Long, K., Galperine, T., and de Truchis, P. (2007). Key amprenavir resistance mutations counteract dramatic efficacy of darunavir in highly experienced patients. *AIDS* **21**(9)**,** 1210-3.

Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L., and Lamarre, D. (1996). Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol* **70**(6)**,** 3763-9.

Feher, A., Weber, I. T., Bagossi, P., Baross, P., Mahalingam, B., Louis, J. M., Copeland, T. D., Yorshin, I. Y., Harrison, R. W., and Tozser, J. (2002). Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. *J Biochem* **269,** 4114-4120.

Gulnik, S. V., Suvorov, L. I., Liu, B., Yu, B., Anderson, B., Mitsuya, H., and Erickson, J. W. (1995). Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* **34,** 9282-9287.

Ho, S. K., Coman, R. M., Bunger, J. C., Rose, S. L., O'Brien, P., Munoz, I., Dunn, B. M., Sleasman, J. W., and Goodenow, M. M. (2008). Drug-associated changes in amino acid residues in Gag p2, p7(NC), and p6(Gag)/p6(Pol) in human immunodeficiency virus type 1 (HIV-1) display a dominant effect on replicative fitness and drug response. *Virology* **378**(2)**,** 272-81.

http://www-fbsc.ncifcrf.gov/HIVdb.

Kantor, R., Fessel, W. J., Zolopa, A. R., Israelski, D., Shulman, N., Montoya, J. G., Harbour, M., Schapiro, J. M., and Shafer, R. W. (2002). Evolution of primary protease inhibitor resistance mutations during protease inhibitor salvage therapy. *Antimicrobial Agents and Chemotherapy* **46**(4)**,** 1086-92.

Kaplan, A. H., Michael, S. F., Wehbie, R. S., Knigge, M. F., Paul, D. A., Everitt, L., Kempf, D. J., Norbeck, D. W., Erickson, J. W., and Swanstrom, R. (1994). Selection of multiple human immunodeficiency virus type 1 variants that encode viral proteases with decreased sensitivity to an inhibitor of the viral protease. *Proc Natl Acad Sci U S A* **91**(12)**,** 5597-601.

King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004a). Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chem Biol* **11**(10)**,** 1333-8.

King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004b). Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chemistry and Biology* **11**(10)**,** 1333-8.

Kolli, M., Lastere, S., and Schiffer, C. A. (2006). Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virology* **347**(2)**,** 405-9.

Maguire, M. F., Guinea, R., Griffin, P., Macmanus, S., Elston, R. C., Wolfram, J., Richards, N., Hanlon, M. H., Porter, D. J., Wrin, T., Parkin, N., Tisdale, M., Furfine, E., Petropoulos, C., Snowden, B. W., and Kleim, J. P. (2002). Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J Virol* **76**(15)**,** 7398-406.

Maguire, M. F., MacManus, S., Griffin, P., Guinea, C., Harris, W., Richards, N., Wolfram, J., Tisdale, M., Snowden, W., and Kleim, J. P. (2001). Interaction of HIV-1 protease and *gag* gene mutations in response to amprenavir-selective pressure exerted in amprenavir-treated subjects - contribution of Gag p6 changes L449F and P453L. *Antivir Ther* **6 (Suppl 1),** 48.

Mahalingam, B., Louis, J., Reed, C., Adomat, J., Krouse, J., Wang, Y., Harrison, R., and Weber, I. (1999). Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur J Biochem* **263**(1)**,** 238-45.

Mammano, F., Petit, C., and Clavel, F. (1998). Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol* **72**(9)**,** 7632-7.

Mitsuya, Y., Winters, M. A., Fessel, W. J., Rhee, S. Y., Hurley, L., Horberg, M., Schiffer, C. A., Zolopa, A. R., and Shafer, R. W. (2006). N88D facilitates the co-occurrence of D30N and L90M and the development of multidrug resistance in HIV type 1 protease following nelfinavir treatment failure. *AIDS Res Hum Retroviruses* **22**(12)**,** 1300-5.

Molla, A., Granneman, G. R., Sun, E., and Kempf, D. J. (1998). Recent developments in HIV protease inhibitor therapy. *Antiviral Res* **39,** 1-23.

Nijhuis, M., Schuurman, R., de Jong, D., Erickson, J., Gustchina, E., Albert, J., Schipper, P., Gulnik, S., and Boucher, C. A. (1999). Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *Aids* **13**(17)**,** 2349-59.

Nijhuis, M., van Maarseveen, N. M., Lastere, S., Schipper, P., Coakley, E., Glass, B., Rovenska, M., de Jong, D., Chappey, C., Goedegebuure, I. W., Heilek-Snyder, G., Dulude, D., Cammack, N., Brakier-Gingras, L., Konvalinka, J., Parkin, N., Krausslich, H. G., Brun-Vezinet, F., and Boucher, C. A. (2007). A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med* **4**(1)**,** e36.

Pai, V. B., and Nahata, M. C. A. P., 33: 325-339 (1999). Nelfinavir mesylate: a protease inhibitor. *Ann Pharmacotherapy* **33**(3)**,** 325-39.

Parkin, N. T., Hellmann, N. S., Whitcomb, J. M., Kiss, L., Chappey, C., and Petropoulos, C. J. (2004). Natural variation of drug susceptibility in wild-type human immunodeficiency virus type 1. *Antimicrob Agents Chemother* **48**(2)**,** 437-43.

Partaledis, J. A., Yamaguchi, K., Tisdale, M., Blair, E. E., Falcione, C., Maschera, B., Myers, R. E., Pazhanisamy, S., Futer, O., Cullinan, A. B., and et al. (1995). In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) isolates with reduced sensitivity to hydroxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. *J Virol* **69**(9)**,** 5228-5235.

Patick, A., Duran, M., Cao, Y., Shugarts, D., Keller, M., Mazabel, E., Knowles, M., Chapman, S., Kuritzkes, D., and Markowitz, M. (1998). Genotypic and phenotypic characterization of humanimmunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob Agents Chemother* **42**(10)**,** 2637-44.

Pazhanisamy, S., Partaledis, J., Rao, B., and Livingston, D. (1998). *In vitro* selection and characterization of VX-478 resistant HIV-1 variants. *Adv Exp Med Biol* **436,** 75-83.

Petropoulos, C. J., Parkin, N. T., Limoli, K. L., Lie, Y. S., Wrin, T., Huang, W., Tian, H., Smith, D., Winslow, G. A., Capon, D. J., and Whitcomb, J. M. (2000). A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob Agents Chemother* **44**(4)**,** 920-8.

Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., and Schiffer, C. A. (2004). Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *Journal of Virology* **78**(22)**,** 12446-54.

Prabu-Jeyabalan, M., Nalivaika, E. A., Romano, K., and Schiffer, C. A. (2006). Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *J Virol* **80**(7)**,** 3607-16.

Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2002). Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* **10**(3)**,** 369-381.

Rhee, S. Y., Gonzales, M. J., Kantor, R., Betts, B. J., Ravela, J., and Shafer, R. W. (2003). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* **31**(1)**,** 298-303.

Schinazi, R. F., Larder, B. A., and Mellors, J. W. (1997). Mutations in retroviral genes associated with drug resistance. *Internatl Antiviral News* **5,** 129-142.

Sheng, N., Pettit, S. C., Tritch, R. J., Ozturk, D. H., Rayner, M. M., Swanstrom, R., and 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**(8)**,** 5723-32.

Sista, P., Wasikowski, B., Lecocq, P., Pattery, T., and Bacheler, L. (2008). The HIV-1 protease resistance mutation I50L is associated with resistance to atazanavir and susceptibility to other protease inhibitors in multiple mutational contexts. *Journal of Clinical Virology* **42**(4)**,** 405-8.

Yanchunas, J., Jr., Langley, D. R., Tao, L., Rose, R. E., Friborg, J., Colonno, R. J., and Doyle, M. L. (2005). Molecular basis for increased susceptibility of isolates with atazanavir resistance-conferring substitution I50L to other protease inhibitors. *Antimicrob Agents Chemother* **49**(9)**,** 3825-32.

Zhang, Y. M., Imamichi, H., Imamichi, T., Lane, H. C., Falloon, J., Vasudevachari, M. B., and Salzman, N. P. (1997). Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol* **71,** 6662-6670.

CHAPTER IV

STRUCTURAL INSIGHTS INTO SUBSTRATE CO-EVOLUTION

The work presented in the following chapter was a collaborative effort. Celia Schiffer and I conceived and designed the study, Michelle Banfill performed site-directed mutagenesis to introduce the D30N and N88D protease mutations, Aysegul Ozen expressed and purified the D30N/N88D mutant protease, I set up the crystal trials, solved and refined the structures to completion, and I performed the structural analyses with inputs from Celia Schiffer.

ABSTRACT

HIV-1 protease is one of the main targets for anti-retroviral therapy due to its important role in the viral life cycle. However, the selective pressure of protease inhibitor therapy results in a wide spectrum of mutations within the protease that renders the virus resistant to protease inhibitors (PIs). In addition, the virus evolves mutations within the Gag cleavage that sometimes correlate with resistance mutations in the protease. Previously we showed that Gag L449F and S451N mutations within p1-p6 cleavage site co-evolved with the nelfinavir (NFV)-resistant D30N/N88D protease mutations. This chapter describes X-ray crystallographic studies performed to examine structures of WT and NFV-resistant protease in complex with p1-p6 peptide variants. In the WT protease, Asp30 interacts with Arg 452 and Gln 450 within the p1-p6 cleavage site. With the D30N mutation, interactions between 30 and Arg 452 are disrupted as a result of Arg 452 adopting a new orientation. This loss of interaction is compensated by the co-evolving L449F and S451N mutations in the p1p6 cleavage site. These mutations significantly increase the overall interactions with the protease, both by means of hydrogen bonds and van der Waals interactions. Increased interactions were also observed between Asp 29 and Ile 50 of the protease and the p1-p6 cleavage peptide. The structural changes in the protease as well as the p1-p6 peptide as a result of the co-evolution suggest an interdependency of interactions between the protease and the p1-p6 peptide. This study provides a rationale for co-evolution of cleavage sites within Gag with resistance mutations within the protease.

INTRODUCTION

Since the identification of HIV-1 protease and the pivotal role it plays in viral maturation, the protease has been an important target for anti-HIV drugs. Protease inhibitors were first developed in 1995 and are currently an integral component of highly active antiretroviral therapy (HAART). This combination therapy, which includes both reverse transcriptase and protease inhibitors, and more recently, integrase and entry inhibitors, has significantly improved prognosis for HIV-1 infected individuals.

There are nine FDA-approved protease inhibitors (PIs) that are currently being used to treat HIV-1 infections. Despite the tremendous success of PIs in treating HIV infections, the evolution of resistance to PIs is a major cause for failure of antiretroviral therapy. A number of mutations develop within and around the active site of the protease, including D30N, G48V, I50V/L, V82A/F/T and I84V in response to PI therapy (Shafer and Schapiro, 2008). These mutations result in altered interactions between the protease and the inhibitors, leading to decreased binding affinity of the inhibitor to the protease.

Nelfinavir mesylate (NFV), a peptidomimetic inhibitor, was discovered using structure-based drug design in 1997 (Kaldor et al., 1997; Kempf et al., 1995; Pai and Nahata, 1999). Though initially approved for use in monotherapy, it was later successfully used in combination therapy with RT inhibitors (Patick et al., 1996). Early studies showed that the appearance of D30N protease mutation, frequently in association with N88D mutation, resulted in severe resistance to NFV (Patick et al., 1998). Mutations at residues 35, 36, 46 and 71 within the protease were also observed in combination with the D30N

mutation (Clemente et al., 2003; Johnson et al., 2006; Patick et al., 1998). In addition to the N88D protease mutation, compensatory mutations were also observed in the p1-p6 cleavage site in combination with the NFV-resistant D30N mutation (Kolli, Lastere, and Schiffer, 2006; Kolli et al., 2009).

The D30N protease mutation is selected specifically in response to NFV both in viral cultures and in patients being treated with NFV (Patick et al., 1998). Structural studies have revealed that Asp 30 interacts with NFV through a sidechain-mediated hydrogen bond that contributes to inhibitor affinity (Kaldor et al., 1997). However, recent studies on protease variants carrying the D30N mutation have shown that this hydrogen bond is still maintained (Clemente et al., 2003; Kozisek et al., 2007). The loss of affinity for NFV has been attributed to a large decrease in conformational entropy on binding of the inhibitor, loss in hydrogen bonding strength and acid-base interactions (Clemente et al., 2003; Kozisek et al., 2007).

Asp 30 is not only important for inhibitor binding but is also essential for recognition of the p1-p6 cleavage site (Kolli, Lastere, and Schiffer, 2006; Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002b). Alteration in affinity for NFV as a result of the D30N mutation (Kozisek et al., 2007) is also likely to compromise p1-p6 recognition and binding leading to its co-evolution (Chapter II). Mutations within p1-p6 cleavage site co-evolve with the D30N/N88D protease mutations associated with NFV resistance (Chapter II and III). Specifically, Gag 449, 451 and 453 within the p1-p6 cleavage site were observed to

mutate frequently. Mutations in p1-p6 cleavage site decrease susceptibility of the virus to NFV when present in combination with D30N88D protease mutations (Chapter III).

Within the p1-p6 cleavage site, mutations were observed at either Gag 449 or 451 but not simultaneously, in combination with D30N/N88D protease mutations (chapter II); (Kolli, Lastere, and Schiffer, 2006). Gag 449 mutated frequently from Leu to Phe, whereas Ser at Gag 451 mutated frequently to an Asn. In both cases, the mutations resulted in a larger amino acid interacting with the surrounding protease residues. Modeling studies described in Chapter II suggested that the mutation to a larger Phe or Asn likely improves van der Waals interactions with the surrounding protease residues, thus compensating for the loss of interactions with Asp 30. However, the simultaneous occurrence of both mutations is infrequent, suggesting the likelihood of van der Waals clashes leading to inefficient processing of the p1-p6 cleavage site. Several studies have shown that Phe at Gag 449 in p1-p6 makes it a better substrate for the wild type (WT) protease, thus enhancing processing at this site (Feher et al., 2002; Maguire et al., 2002). The observed interdependency of molecular interactions likely plays a role in the co-evolution of the p1-p6 cleavage site.

This chapter focuses on elucidating the structural rationale for the co-evolution of p1-p6 cleavage site with the NFV-resistant D30N/N88D HIV-1 protease. Crystal structures of WT and D30N/N88D HIV-1 proteases in complex with WT and mutant p1-p6 substrate variants were determined. Structural changes in the protease as well as the p1-p6 peptide

as a result of the mutations indicate an interdependency of interactions between the protease and the p1-p6 peptide.

MATERIALS AND METHODS

Nomenclature. HIV-1 protease variants (WT or D30N/N88D) and the p1-p6 substrate variants will be distinguished as subscripts throughout this chapter. For example, WT protease in complex with WT p1-p6 is denoted by PR_{WT} – $p1\text{-}p6_{WT}$ and mutant protease in complex with S451 mutation in p1-p6 is denoted by $PR_{D30N/N88D}$ – $p1\text{-}p6_{S451N}$.

Substrate Peptides. Decameric peptides corresponding to the p1-p6 processing site within the Gag polyprotein (AA 444-453) and its variants were purchased from Quality Controlled Biochemicals, Inc., Hopkinton, MA. The following peptide variants were used in the x-ray crystallographic structural studies described in the subsequent sections: (1) $p1\text{-}p6_{WT}$ – RPGNFLQSRP (2) $p1\text{-}p6_{L449F}$ – RPGNFFQNRP (3) $p1\text{-}p6_{S451N}$ – RPGNFLQNRP and (Collaborative-Computational-Project) $p1\text{-}p6_{L449F/S451N}$ – RPGNFFQNRP.

D30N/N88D Protease Gene Construction. The clade B wild-type (WT) protease gene was constructed synthetically using codons optimized for protein expression in Escherichia coli as previously described with the Q7K mutation introduced to prevent autoproteolysis (Rose, Salto, and Craik, 1993). The D25N was introduced to inactivate the protease and prevent substrate cleavage. The D30N/N88D protease mutations were then introduced sequentially into the gene. Mutations were introduced using the Quik-

Change™ site-directed mutagenesis kit (Stratagene ®, La Jolla, CA) and confirmed by sequencing.

Protein Expression and Purification. Expression and purification of the inactive WT and D30N/N88D protease variants was performed as previously described (Prabu-Jeyabalan et al., 2004). Briefly, the protease gene was sub-cloned into the heat-inducible pXC35 expression vector (American Type Culture Collection (ATCC), Manassas, VA) and transformed into *E. coli* TAP-106 cells. Transformed cells were grown at 30℃ till A_{600}= 0.6 and induced by raising the temperature to 42℃. Cells were allowed to grow for 4 hours and lysed to release the protease-containing inclusion bodies. The protease was extracted with 50% acetic acid and purified by gel filtration on a 2.1L Sephadex G-75 superfine (Sigma Chemicals) column equilibrated with 50% acetic acid. The purified protease was refolded by rapid 10-fold dilution into a buffering solution (refolding buffer) containing 0.05 M sodium acetate (pH 5.5), 10% glycerol, 5% ethylene glycol, and 5mM dithiothreitol. The refolded protein was concentrated and dialyzed to remove any residual acetic acid. The protease was further purified with a Pharmacia Superdex 75 fast performance liquid chromatography (FPLC) column equilibrated with refolding buffer.

Crystallization and Data Collection. Crystals used in this study were obtained under multiple conditions. Protease solutions between 1.0 and 2.2 mg mL^{-1} were equilibrated with a ten-fold molar excess of the p1-p6 peptide variants, individually, for 1 hour on ice. Crystals used were grown by the hanging drop vapor diffusion method over a reservoir

solution consisting of one of the following three buffer solutions – 1) 126-mM sodium phosphate buffer (pH 6.2), 63-mM sodium citrate, and 20% to 32% ammonium sulfate 2) 0.1-0.5M MES monohydrate buffer (pH 6.5), 0.1-0.4M ammonium sulphate, and 20-31% PEG MME 5000 3) 0.1M citrate (pH 5.5), 0.5-3.0M ammonium sulphate. X-ray diffraction data for $PR_{D30N/N88D}$ —p1-$p6_{S451N}$ was collected on the BioCARS beamline 14-BMC at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) at a wavelength of 0.9 Å with a Quantum 315 CCD X-ray detector (Area Detector Systems Corporation, Poway, CA). Diffraction data for the other 5 complexes were collected on the BioCARS beamline 14-IDB at a wavelength of 1.033 Å with a Mar 165 CCD X-ray detector (Rayonix, LLC, Evanston, Illinois). All data were collected under cryo-cooled conditions.

Structure Solution and Crystallographic Refinement. The data were indexed and scaled using HKL-2000 software (HKL Research, Charlottesville, VA). Structures were solved and refined using the CCP4 program suite (Collaborative-Computational-Project, 1994). The Matthews co-efficient was calculated for all structures to determine the number of molecules in the asymmetric unit. For complexes with two molecules in the asymmetric unit, structure solution was carried out using the molecular replacement package Phaser (McCoy et al., 2007), whereas for complexes with one molecule in the asymmetric unit, AMoRe was used (Navaza, 1994). The structure of darunavir (inhibitor) in complex with active protease was used as the starting model (1T3R; (Surleraux et al., 2005). ARP/wARP was used to build solvent molecules into the electron density (Langer et al., 2008). Refinement was carried out using a combination

of TLS (Winn, Isupov, and Murshudov, 2001) and restrained refinement using Refmac5 (Murshudov, Vagin, and Dodson, 1997). Each cycle of refinement was followed by model building and real space refinement performed using COOT, a molecular-graphics software (Emsley and Cowtan, 2004). For all complexes, the peptides were built into the F_o-F_c electron density after two rounds of refinement cycles. The working R (R_{Work}) and its cross validation (R_{Free}) were monitored throughout refinement. The quality of the structures was assessed using MolProbity (Davis et al., 2007).

Structure Analyses. Comparisons of the crystal structures were made by superposing Cα atoms of the terminal regions (residues 1 to 9 and 86 to 99) from both monomers onto a previously solved structure of WT protease in complex with WT p1-p6 (PR_{WT}—p1-$p6_{WT}$) (PDB 1KJF; (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002b). Structures were visualized using PyMol molecular-graphics software (DeLano, 2002).

For structures with two molecules in the asymmetric unit, RMSDs were calculated and double-difference plots generated to visualize structural differences, if any, between the two molecules as previously described (Prabu-Jeyabalan et al., 2006). Briefly, distances between all the Cα atoms within the dimer were calculated for each complex. A distance-difference matrix was then computed for each atom for a given pair of complexes. The distance-difference matrix was then plotted as a contour plot using GNUPLOT plotting software.

Hydrogen Bonds and van der Waals Contacts

Hydrogen bond interactions were determined using the default settings of the HBPLUS program (McDonald, 1994). van der Waals interaction energies between protease and the p1-p6 peptide were computed using a simplified Lennard-Jones potential, *V(r)* using the relation $4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^{6}]$, where r is the protease-p1-p6 interatomic distance and ε and σ are *the depth of the potential well* and collision diameter, respectively for each protease substrate atom pair. *V(r)* is computed for all possible protease-p1-p6 atom pairs within 5.0 Å and potentials for nonbonded pairs separated less than a distance at the minimum of the potential were equated to - ε. Using this simplified potential for each non-bonded protease-p1-p6 atom pair, total van der Waals contact energy, Σ*V(r)* was then computed for each protease and p1-p6 residue.

RESULTS

Overall Structure of the Complexes

The four decameric p1-p6 substrate variants were crystallized in complex with inactive variants of WT and NFV-resistant D30N/N88D HIV-1 protease. Crystals were successfully grown and data obtained for all complexes. Structures were solved for all complexes, except $PR_{D30N/N88D}$ – $p1\text{-}p6_{L449F}$, to a resolution between 1.6 – 1.9 Å (Table IV.1). Four complexes were solved in the $P2_1$ space group with two molecules in the asymmetric unit, and two were solved in the $P2_12_12_1$ with one molecule in the asymmetric unit. In the structures where there are two molecules in the asymmetric unit,

Table IV.1

Variable	PR_{WT} - p1-$p6_{WT}$	$PR_{D30N/N88D}$ - p1-$p6_{WT}$	PR_{WT} - p1-$p6_{L449F}$	PR_{WT} - p1-	$PR_{D30N/N88D}$ - p1-	PR_{WT} - p1-p6	$PR_{D30N/N88D}$ - p1-
Data collection							
Space group	P212121	P21	P21	P212121	P21	P212121	P21
a (Å)	51.29	51.67	51.14	51.19	51.55	50.94	51.47
b (Å)	59.07	60.11	62.79	58.3	60.07	58.41	59.54
c (Å)	61.81	60.2	61.46	61.36	60.23	61.46	59.36
β-angle	90	99.09	98.03	90	99.25	90	98.6
Z	4	8	8	4	8	4	8
Temperature (°C)	-80	-80	-80	-80	-80		-80
Resolution (Å)	2.0	1.78	1.85	1.75	1.68	1.65	1.9
Total Reflections	41786	133292	113050	109140	167489	155872	126355
Unique Reflections	12376	34288	28931	14456	41265	22296	32879
R merge (%)	6.7	6.8	7.3	4.5	6.6	8.7	6.4
Completeness (%)	93.4	99.2	99.7	96	99.8	98.1	98.8
Crystallographic							
R value (%)	20.3	20.04	16.30	14.13	17.97	17.30	20.01
R free (%)	25.1	25.46	21.28	17.48	22.89	20.45	24.44
RMSD in:							
Bond length (Å)	0.006	0.009	0.009	0.009	0.009	0.009	0.009
Bond angles	1.3	1.404	1.423	1.461	1.460	1.457	1.441

Table IV.1: Crystallographic Statistics for WT and D30N/N88D Protease Complexes with the Four p1-p6 Peptide Variants. The statistics for the PR_{WT}—$p1\text{-}p6_{WT}$ structure are also included for reference.

the p1-p6 peptide could be modelled unambigously in at least one of the complexes. Additionally, because previously solved structures of substrate complexes from earlier studies have shown that Gag 444 and 453 have little contact with the protease and are often disordered, differences between the molecules observed at Gag 444 and 453 of the peptide were not taken into account. The structural differences between the two molecules are not significant as observed by the the double difference plots (Fig IV.1 a-d) and RMSDs which range between 0.14-0.19 (for the entire protease) and 0.11-0.16 for the p1-p6 peptide. Therefore, in the subsequent structural analyses, only one molecule from the asymmetric units was chosen based on the unambiguous location of sidechains for residues Gag 447-452 within the density. The complete crystallographic statistics are listed in Table IV.1.

All structures are similar to the previously solved PR_{WT}—p1-$p6_{WT}$ structure, with RMSDs ranging from 0.36 to 0.64. The three WT protease complexes show backbone shifts in the flap-hinge region compared to the PR_{WT}—p1-$p6_{WT}$ structure (RMSDs 0.36-0.4) (Fig IV.2a). Backbone shifts are also observed at residues 11-22, which form part of the protease core. This region is surface exposed and is known to be flexible. Structures of the three D30N/N88D protease complexes are also very similar to each other (RMSDs 0.13-0.3) (Fig IV.2b). However, they all differ from the PR_{WT}—p1-$p6_{WT}$ in the flap-hinge region and the β-strand 6 (AA 69-78) especially in monomer a, with shifts upto 1.5 Å. The differences in the flap-hinge region are likely due to the effect of crystal packing. The change in β-strand 6 is likely due to the N88D protease mutation and will be described in the following sections.

Figure IV.1

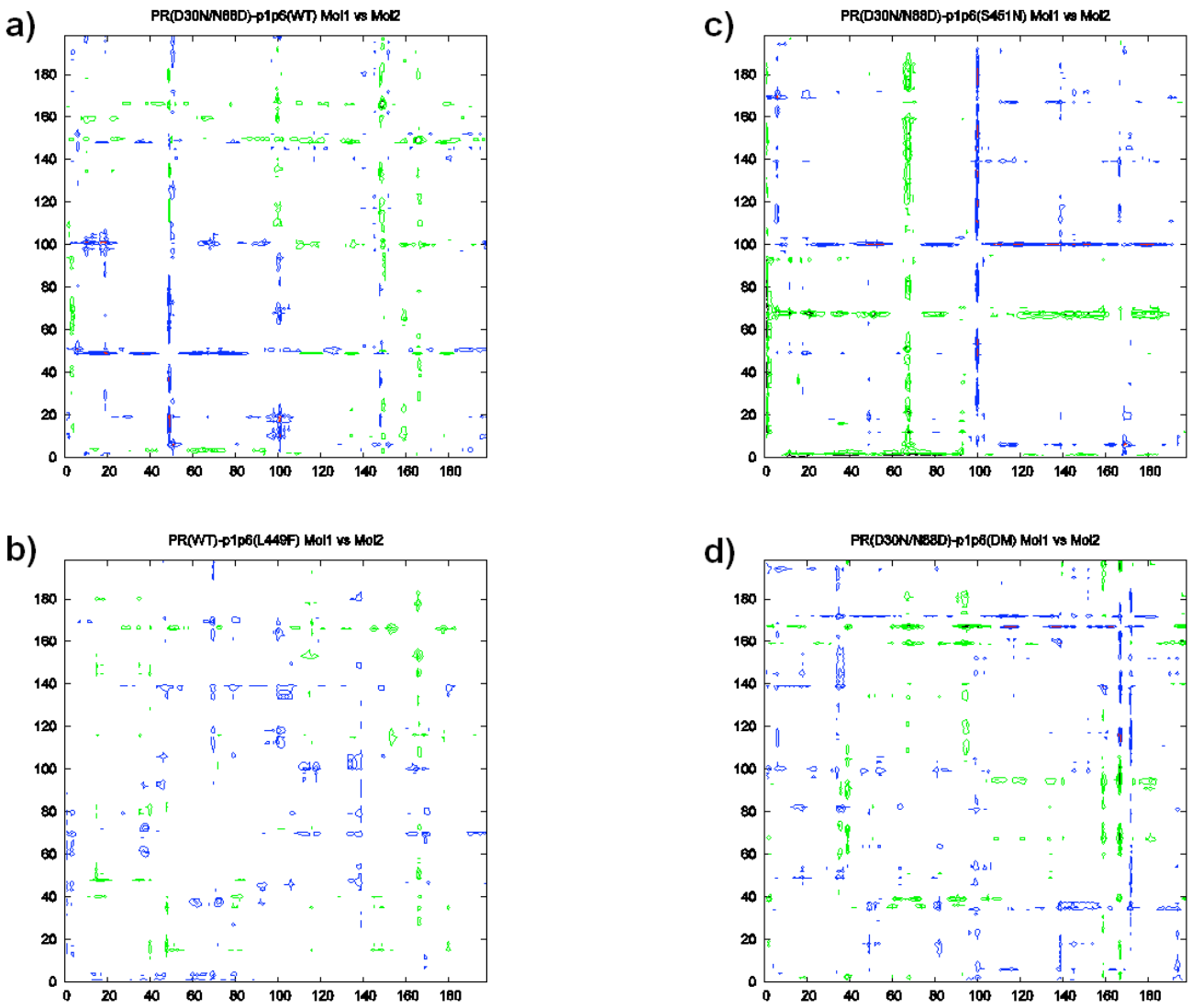

a)
PR(D30N/N88D)-p1p6(WT) Mol1 vs Mol2
b)
PR(WT)-p1p6(L449F) Mol1 vs Mol2
c)
PR(D30N/N88D)-p1p6(S451N) Mol1 vs Mol2
d)
PR(D30N/N88D)-p1p6(DM) Mol1 vs Mol2

Figure IV.1: Double difference plots comparing the two molecules in the asymmetric unit. (a) $PR_{D30N/N88D}$—p1-$p6_{WT}$ (b) PR_{WT}—p1-$p6_{L449F}$ (c) $PR_{D30N/N88D}$ —p1-$p6_{S451N}$ and (d) $PR_{D30N/N88D}$ —p1-$p6_{L449F/S451N}$. The color contours indicate distance differences of < 1.0 Å (black), -1.0 to -0.5 Å (green), 0.5 to 1.0 Å (Vivés, Brodin, and Leblue) and > 1.0 Å (red).

Figure IV.2

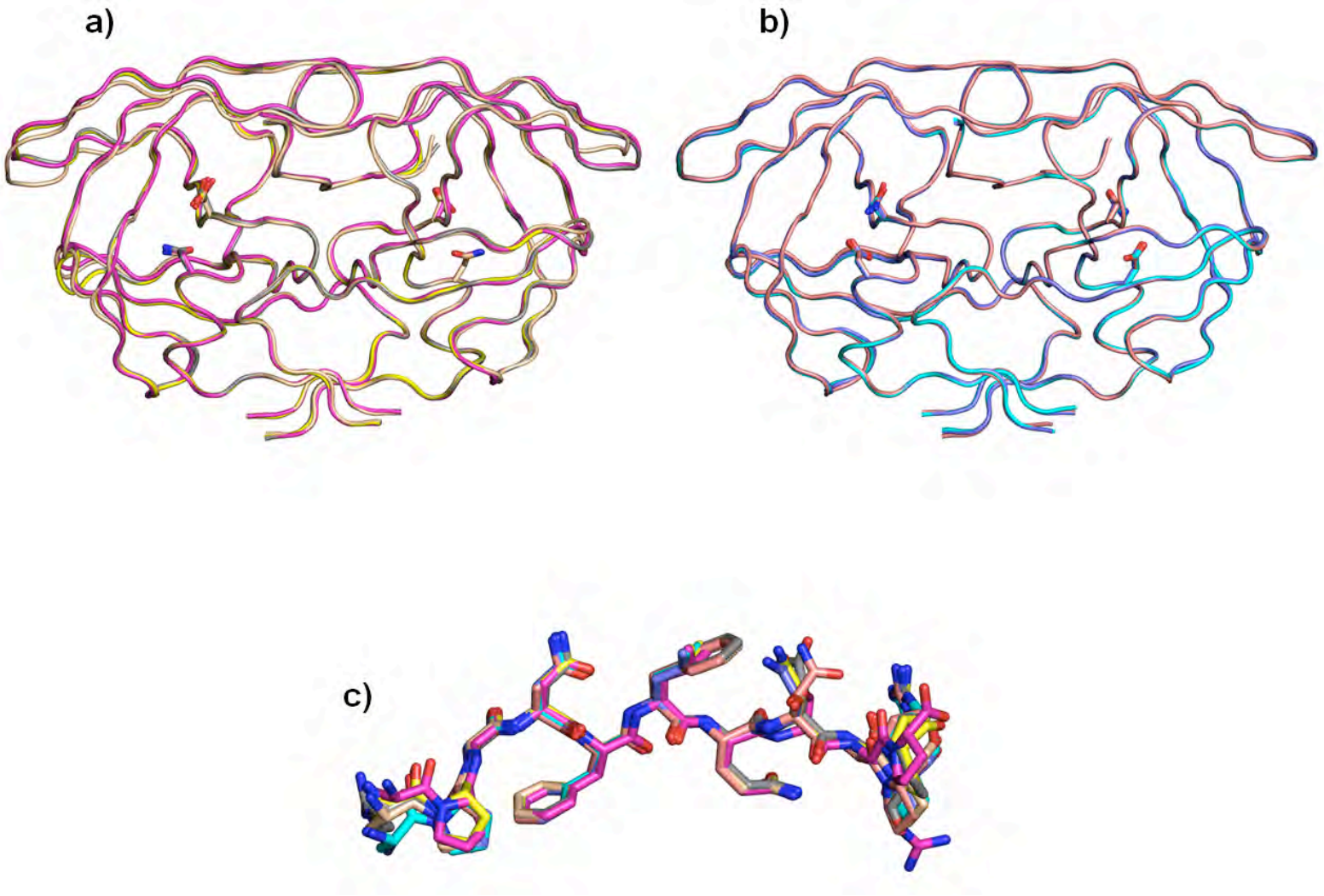


Figure IV.2: Cartoon Diagram for Superposition of the Protease-p1-p6 Complexes. (a) Superposition of WT proteases in complex with the p1-p6 peptide variants. (b) Superposition of D30N/N88D proteases in complex with the p1-p6 eptide variants. (c) Superpositions of all the p1-p6 peptides.

For all structures used in the analyses, including the previously solved PR_{WT}—p1-$p6_{WT}$, nine substrate residues from Gag 447 to 453 are unambiguously located within the electron density map. All the p1-p6 substrate peptides bound to the protease in the same extended conformation as shown in previous studies (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2000; Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002b), and the side chains of residues Gag 447 through 451 are similarly oriented (Fig IV.2c). However, they all differ significantly from the PR_{WT}—p1-$p6_{WT}$ structure in that the sidechain for Arg 452 is oriented in a different direction. The details of the resulting structural alterations are discussed in the following sections.

Hydrogen Bonds

All the p1-p6 complexes with WT and D30N/N88D protease variants show an extensive network of conserved hydrogen bonds, similar to previously solved substrate complexes (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002b). The five water molecules that bridge the peptide to the protease are also conserved across all six structures (Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002b). In all the complexes, including the PR_{WT}—p1-$p6_{WT}$, 30a does not make any sidechain-mediated hydrogen bonds with the p1-p6 peptide. Though residue 88 is not in direct contact with either p1-p6 or 30, alterations in the hydrogen bonding network around this residue are observed. Significant differences in hydrogen bonds observed around protease residues 30b and 88 are described below.

Hydrogen bonds around 30b. The most significant changes in hydrogen bonds are observed due to the D30N protease mutation (Fig IV.3a-g). The Gln 450 NE1 to Asp 30b OD1 hydrogen bond is conserved across all structures. Asp 30b also forms a hydrogen bond with a water molecule that is conserved across the WT complexes with all the p1-p6 substrate variants. In the case of PR_{WT}—p1-$p6_{WT}$, strong electrostatic interactions are observed between NE and NH1 of Arg 452 and OE2 of Asp 30b (Fig IV.3a). In all the other complexes Arg 452 is flipped away from residue 30b and these electrostatic interactions are lost (Fig IV.3b-g). In all complexes except the PR_{WT}—p1-$p6_{WT}$ structure, Arg 452 makes at least one hydrogen bond with water or other solvent molecules including glycerol and ethylene glycol. Additionally, Arg 452 NE makes a water mediated hydrogen bond with 30b OD1/2 in all except the PR_{WT}—p1-$p6_{WT}$ structure. In $PR_{D30N/N88D}$ —p1-$p6_{S451N}$ and PR_{WT}—p1-$p6_{S451N}$ complexes the Asn 451 forms hydrogen bonds with Arg 8a (Fig IV.3d and e) as well as water molecules, thus resulting in increased interactions as compared to Ser 451 in the PR_{WT}—p1-$p6_{WT}$ and PR_{WT}—p1-$p6_{L449F}$ and $PR_{D30N/N88D}$—p1-$p6_{WT}$ structure (Fig IV.3a-c). Thus the loss of hydrogen bond interactions with 30b as a result of the flipped Arg 452 appears to be compensated by several other hydrogen bonds.

Hydrogen bonds around 88. There is a network of hydrogen bonds that extend from 88 that help orient this residue in this region, both in monomer a and b of the protease (Fig IV.4 a-g) . In all the WT complexes with the p1-p6 variants, Asn 88 ND1 forms a hydrogen bond with Thr74 O. As a result of the N88D mutation, residue 88 is unable to maintain this hydrogen bond with Thr74. Instead, 88 flips slightly inwards and the OE1

Figure IV.3

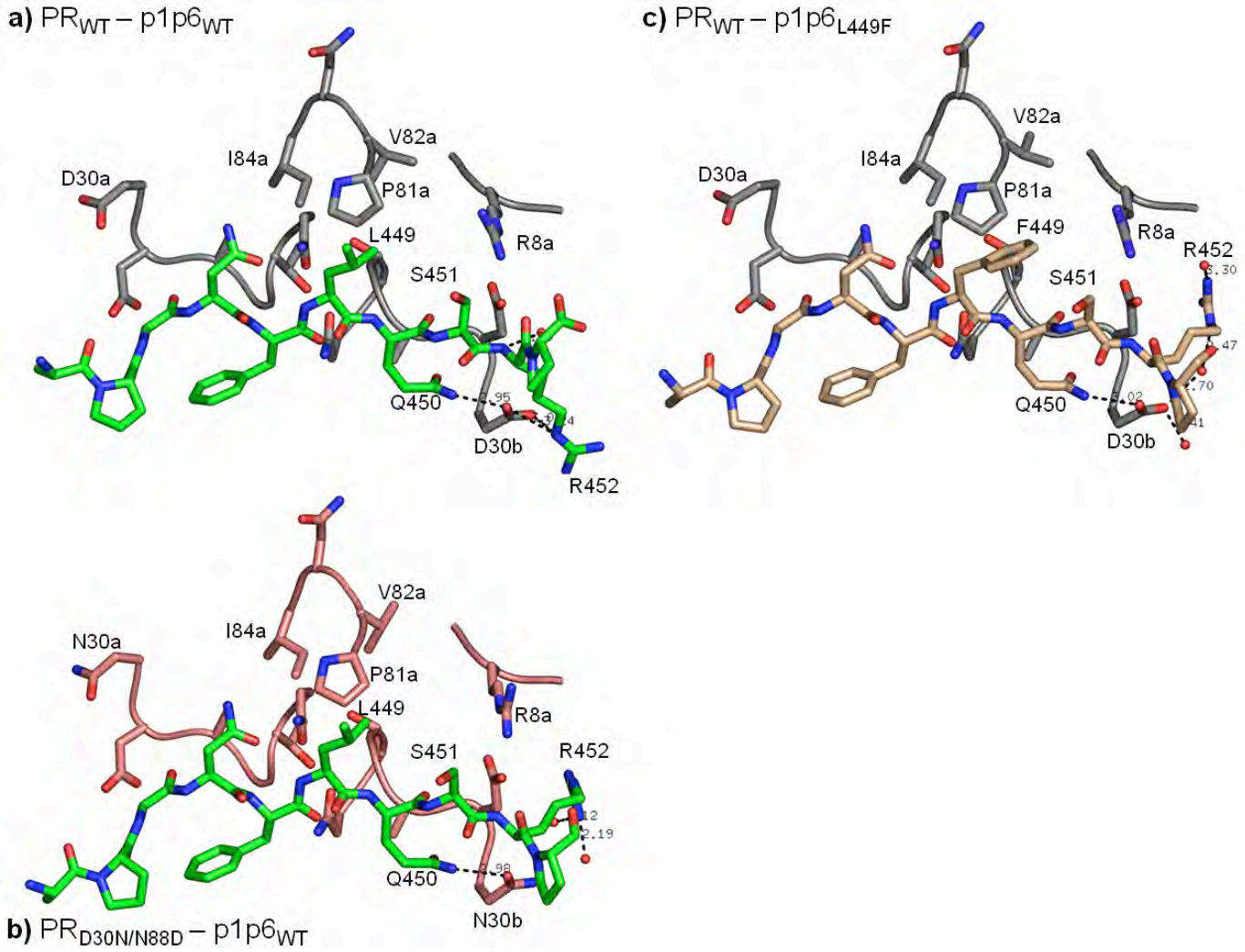

a) PR$_{WT}$ – p1p6$_{WT}$
V82a
I84a
D30a
P81a
L449
R8a
S451
Q450
D30b
R452
c) PR$_{WT}$ – p1p6$_{L449F}$
V82a
I84a
D30a
P81a
F449
R8a
R452
S451
Q450
D30b
b) PR$_{D30N/N88D}$ – p1p6$_{WT}$
V82a
I84a
N30a
P81a
L449
R8a
S451
R452
Q450
N30b

Figure IV.3 *(cont.)*

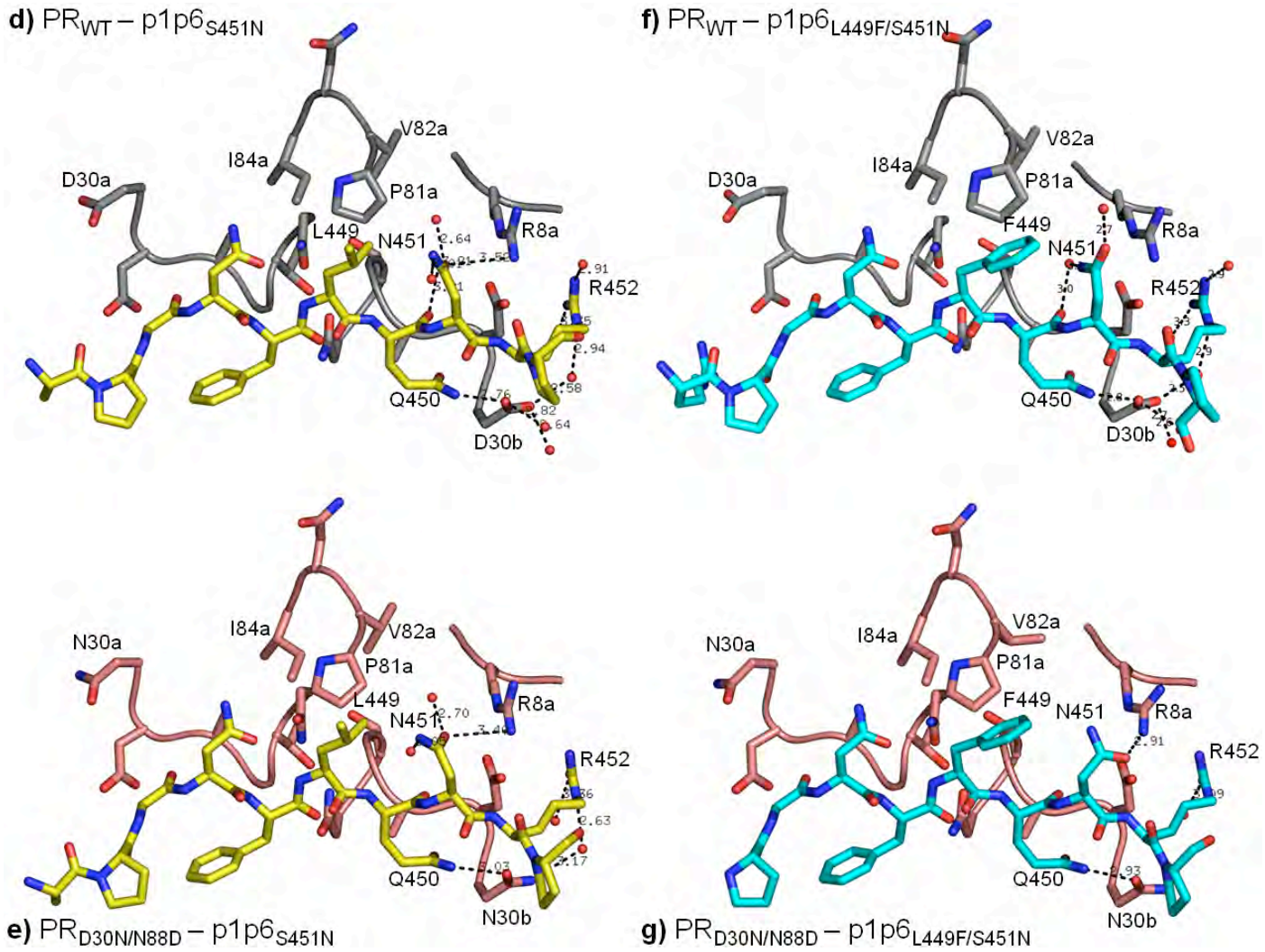


Figure IV.3: Hydrogen Bond Interactions of the p1-p6 Peptide. (a) PR_{WT}—p1-$p6_{WT}$ (b) $PR_{D30N/N88D}$—p1-$p6_{WT}$ (c) PR_{WT}—p1-$p6_{L449F}$, (d) PR_{WT}—p1-$p6_{S451N}$, (e) $PR_{D30N/N88D}$—p1-$p6_{S451N}$, (f) PR_{WT}—p1-$p6_{L449F/S451N}$, (g) $PR_{D30N/N88D}$—p1-$p6_{L449F/S451N}$. The orientation of Arg 452 and the hydrogen bonds around 30b in all structures are shown. Hydrogen bonds around Gag 451 are also significantly altered.

Figure IV.4

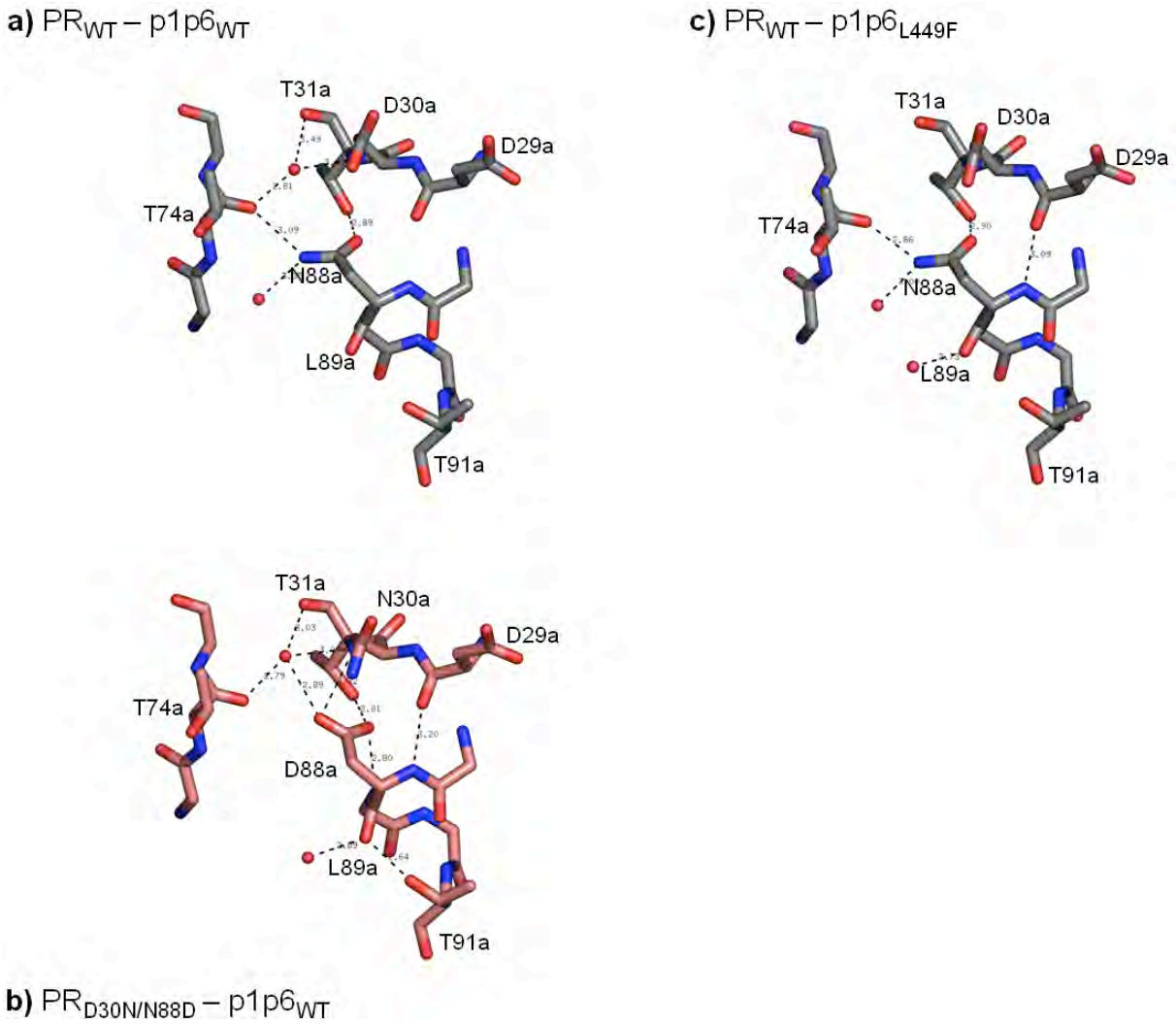
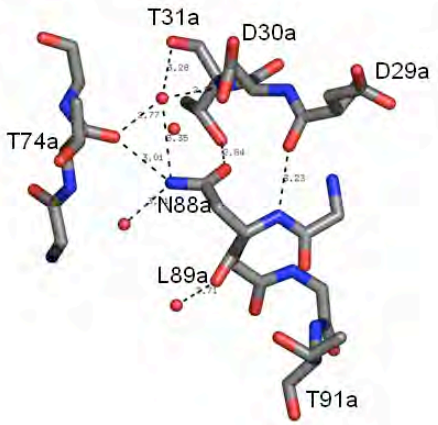
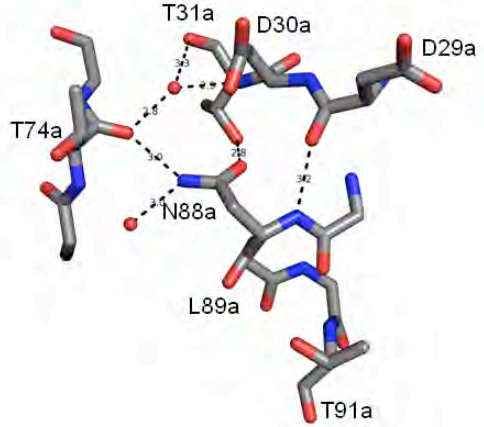

a) PR$_{WT}$ – p1p6$_{WT}$
T31a
D30a
D29a
T74a
N88a
L89a
T91a
c) PR$_{WT}$ – p1p6$_{L449F}$
T31a
D30a
D29a
T74a
N88a
L89a
T91a
T31a
N30a
D29a
T74a
D88a
L89a
T91a
b) PR$_{D30N/N88D}$ – p1p6$_{WT}$

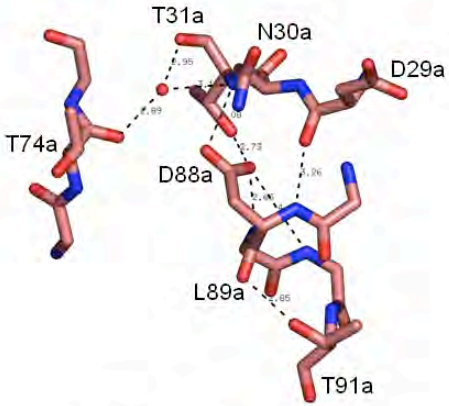
Figure IV.4 *(cont.)*

d) PR_{WT} – $p1p6_{S451N}$

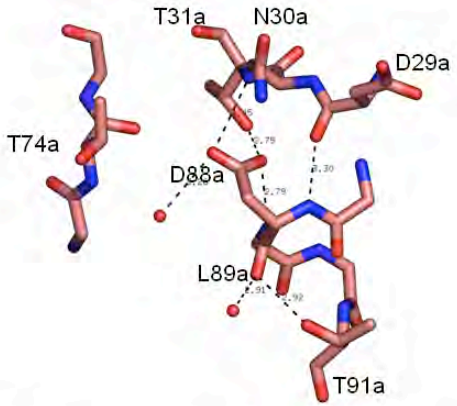


f) PR_{WT} – $p1p6_{L449F/S451N}$





e) $PR_{D30N/N88D}$ – $p1p6_{S451N}$



g) $PR_{D30N/N88D}$ – $p1p6_{L449F/S451N}$

Figure IV.4: Hydrogen Bond Network Around Residue 88. (a) PR_{WT}—p1-$p6_{WT}$ (b) $PR_{D30N/N88D}$—p1-$p6_{WT}$ (c) PR_{WT}—p1-$p6_{L449F}$, (d) PR_{WT}—p1-$p6_{S451N}$, (e) $PR_{D30N/N88D}$—p1-$p6_{S451N}$, (f) PR_{WT}—p1-$p6_{L449F/S451N}$, (g) $PR_{D30N/N88D}$—p1-$p6_{L449F/S451N}$. Loss of interactions with Thr 74 as a result of the N88D mutation are shown. Asp 88 shows significant deviation from the WT Asn.

makes a hydrogen bond with Thr31 N. Additionally, N88D O forms a hydrogen bond with Thr91 OG1 and N88D OE2 hydrogen bonds with Leu89 N, further stabilizing this residue. The hydrogen bond between 88 OE1 and Thr31 OG1 are conserved across all complexes, including the previously solved PR_{WT}—p1-$p6_{WT}$ structure.

Van der Waals Interactions

Significant but similar changes in the overall van der Waals interactions are observed in all the structures compared to the PR_{WT}—p1-$p6_{WT}$ structure as revealed by the interaction energies. Analysis of the van der Waals interaction energies of the active site residues show significant interactions between active site residues of monomer b and the p1-p6 peptide, and large but similar changes in interaction energies are observed in all structures compared to the PR_{WT}—p1-$p6_{WT}$ (Fig IV.5). In contrast, the few contacts between monomer a and the p1-p6 peptide remain relatively constant across all seven structures (Fig IV.5). A more detailed examination reveals reduced interactions between 30b and the p1-p6 peptide for all structures compared to the PR_{WT}—p1-$p6_{WT}$ structure (Fig IV.5). In contrast, for all structures except PR_{WT}—p1-$p6_{WT}$, the van der Waals interaction energies of D29b increase significantly and in a similar manner, indicating increased interactions with the p1-p6 peptide (Fig IV.5). A similar increase in favorable van der Waals interactions are also observed between residues 47 48 and 50 of monomer b and the p1-p6 peptide of all structures compared to PR_{WT}—p1-$p6_{WT}$, but to a smaller extent (Fig IV.5).

Figure IV.5

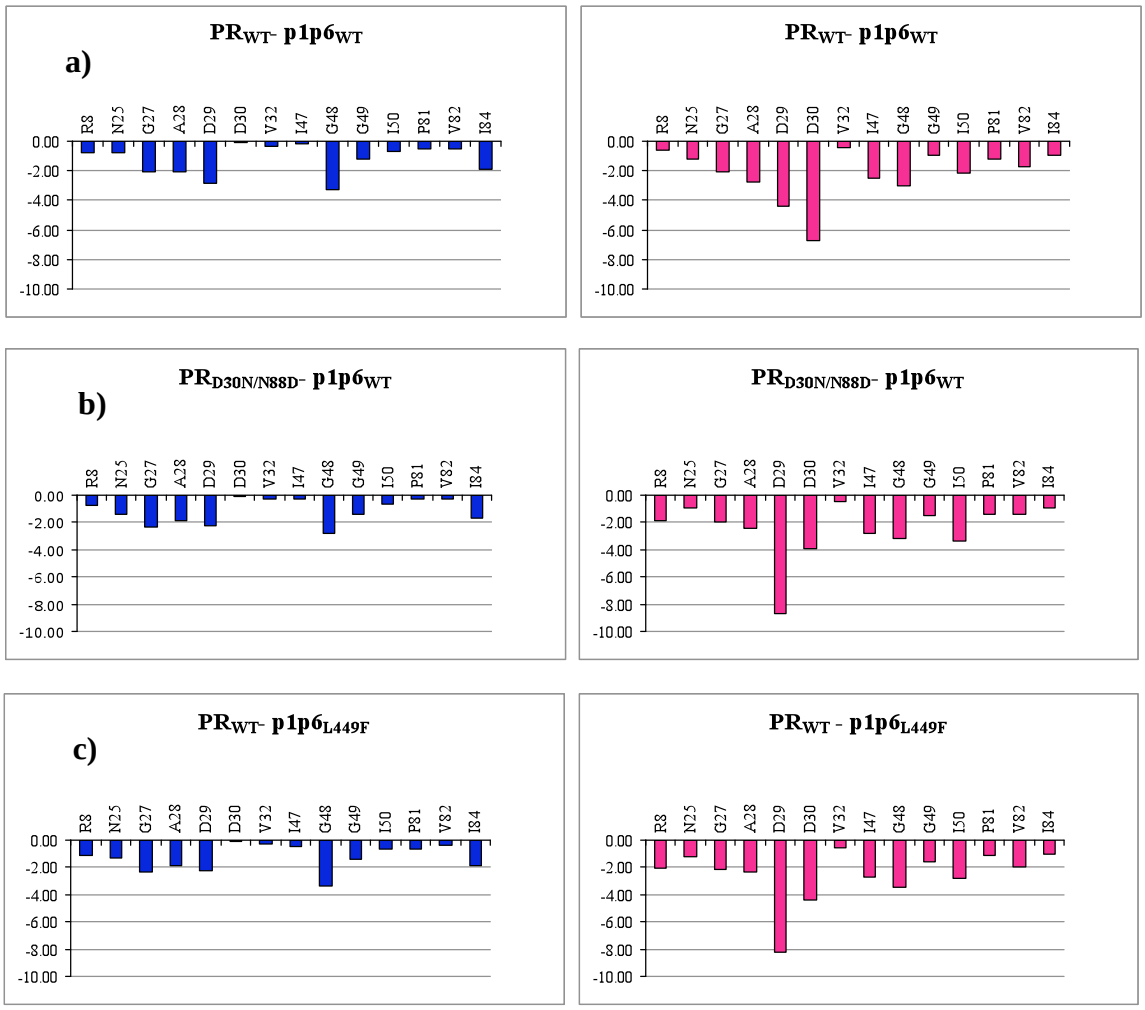

a)
PRWT- p1p6WT
R8 N25 G27 A28 D29 D30 V32 I47 G48 G49 I50 P81 V82 I84
0.00 -2.00 -4.00 -6.00 -8.00 -10.00
PRWT- p1p6WT
R8 N25 G27 A28 D29 D30 V32 I47 G48 G49 I50 P81 V82 I84
0.00 -2.00 -4.00 -6.00 -8.00 -10.00
b)
PRD30N/N88D- p1p6WT
R8 N25 G27 A28 D29 D30 V32 I47 G48 G49 I50 P81 V82 I84
0.00 -2.00 -4.00 -6.00 -8.00 -10.00
PRD30N/N88D- p1p6WT
R8 N25 G27 A28 D29 D30 V32 I47 G48 G49 I50 P81 V82 I84
0.00 -2.00 -4.00 -6.00 -8.00 -10.00
c)
PRWT- p1p6L449F
R8 N25 G27 A28 D29 D30 V32 I47 G48 G49 I50 P81 V82 I84
0.00 -2.00 -4.00 -6.00 -8.00 -10.00
PRWT - p1p6L449F
R8 N25 G27 A28 D29 D30 V32 I47 G48 G49 I50 P81 V82 I84
0.00 -2.00 -4.00 -6.00 -8.00 -10.00

Figure IV.5 *(cont.)*

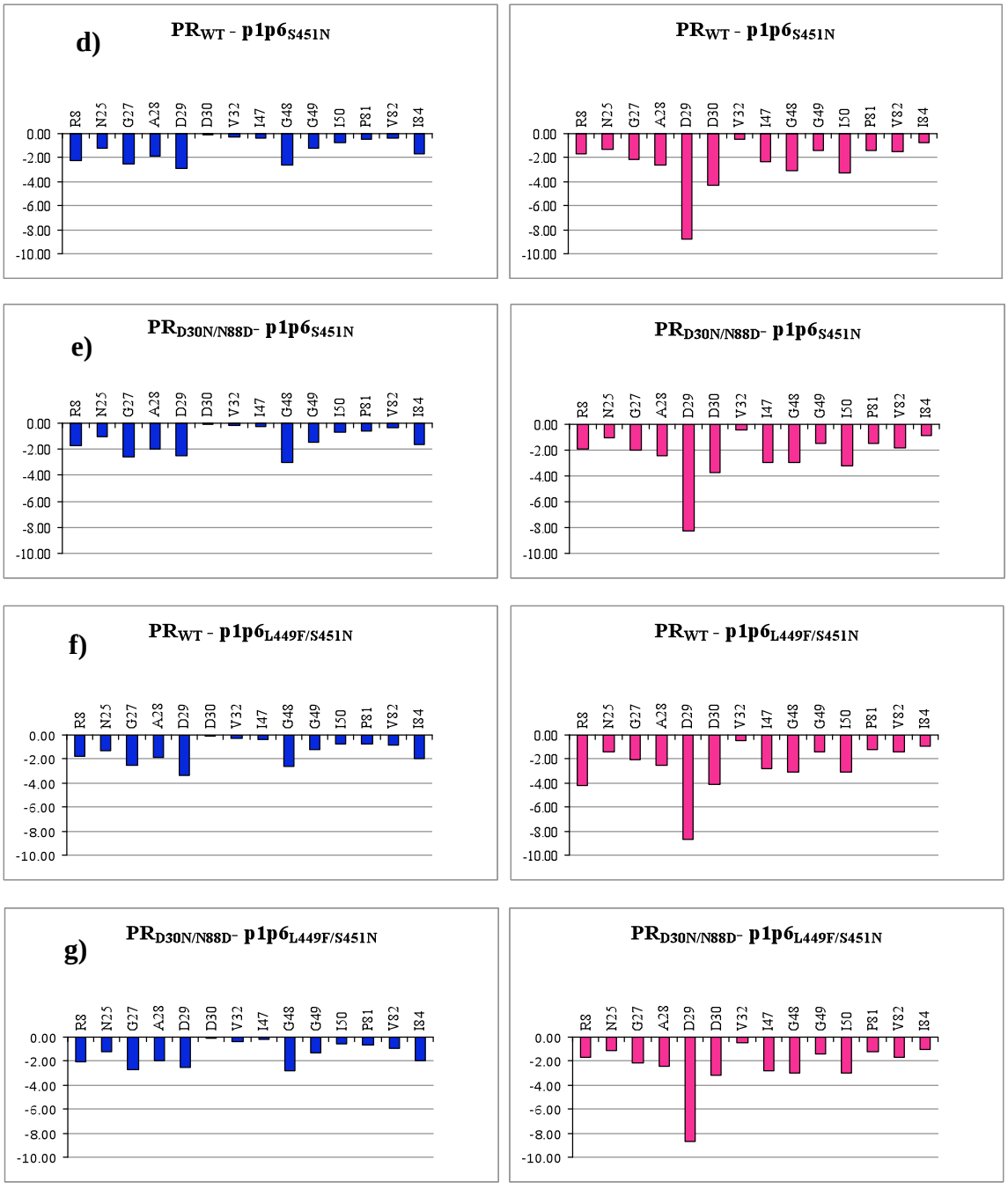


Figure IV.5: The Total van der Waals Interaction Energies for Active Site Residues in Protease. The interaction energies for residues in monomer a are shown in blue and those for residues in monomer b are shown in magenta. (a) PR_{WT}—$p1\text{-}p6_{WT}$ (b) $PR_{D30N/N88D}$—$p1\text{-}p6_{WT}$ (c) PR_{WT}—$p1\text{-}p6_{L449F}$, (d) PR_{WT}—$p1\text{-}p6_{S451N}$, (e) $PR_{D30N/N88D}$—$p1\text{-}p6_{S451N}$, (f) PR_{WT}—$p1\text{-}p6_{L449F/S451N}$, (g) $PR_{D30N/N88D}$—$p1\text{-}p6_{L449F/S451N}$. No significant changes in interaction energies are seen in monomer a. In the case of monomer b, mutations in protease or p1-p6 result in decrease van der Waals interaction energy at 30, which are compensated by interactions made by 29b.

Analysis of the van der Waals interaction energies of the p1-p6 peptide with the protease shows that the overall contacts made by each residue of the peptide follows a common trend across all structures (Fig IV.6 a-g). However, the most significant change is observed in the PR_{WT}—p1-$p6_{L449F}$ (Fig IV.6c), PR_{WT}—p1-$p6_{L449F/S451N}$ (Fig IV.6d) and $PR_{D30N/N88D}$—p1-$p6_{L449F/S451N}$ (Fig IV.6g) structures as a result of the larger Phe at Gag 449, leading to increased interactions with the surrounding protease residues, as compared to the PR_{WT}—p1-$p6_{WT}$ structure (Fig IV.6a). A similar increase in favorable interactions is observed in PR_{WT}—p1-$p6_{S451N}$ (Fig IV.6d), $PR_{D30N/N88D}$—p1-$p6_{WT}$ (Fig IV.6b) and $PR_{D30N/N88D}$—p1-$p6_{S451}$ (Fig IV.6e), though not to the same extent. At least in the case of the PR_{WT}—p1-$p6_{S451N}$ (Fig IV.6d and Fig IV.7d) and $PR_{D30N/N88D}$—p1-$p6_{S451}$ (Fig IV.6e and Fig IV.7e) structures this could be due to the larger Asn in the same pocket of the protease resulting in more favorable contacts with Leu 449. To investigate this, the van der Waals interactions made by the p1-p6 peptide from residues 449 to 452 were examined in detail. Lastly, van der Waals interactions between 30b and the peptide were also examined.

Interactions at Gag 449. Parallel to the increased overall van der Waals energies at Gag 449 observed in the previous section, a detailed examination of interactions reveals that more favorable and more number of interactions are made with the protease in the PR_{WT}—p1-$p6_{L449F}$ and $PR_{D30N/N88D}$ —p1-$p6_{L449F/S451N}$ structures (Table IV.2). However, a similar increase is not observed in the PR_{WT}—p1-$p6_{L449F/S451N}$ and the interactions are comparable to those seen in the $PR_{D30N/N88D}$ —p1-$p6_{S451N}$, $PR_{D30N/N88D}$—p1-$p6_{WT}$, and PR_{WT}—p1-$p6_{S451N}$ structures as indicated by the van der Waals interaction energies.

Figure IV.6

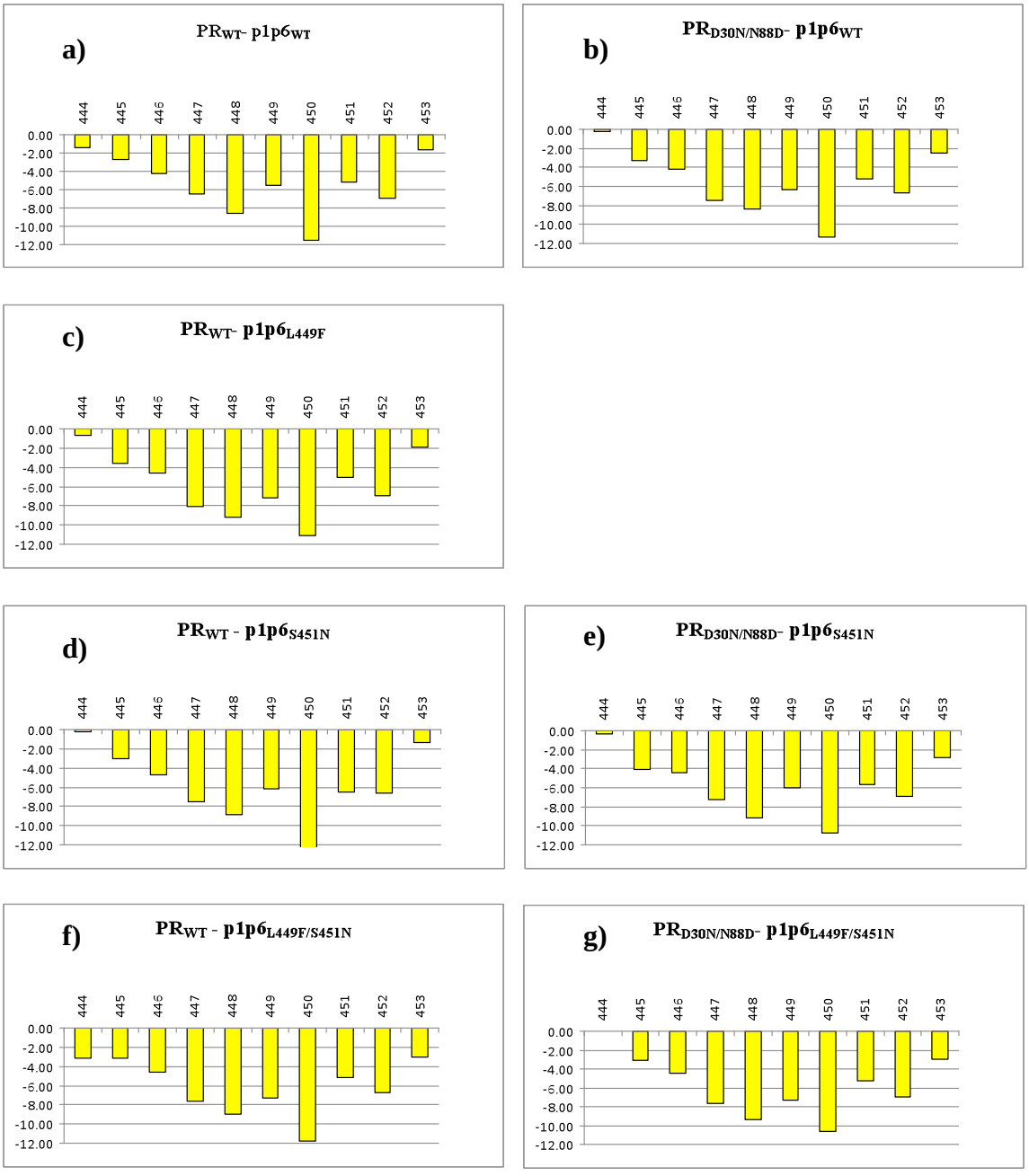

a)
PRWT- p1p6WT
b)
PRD30N/N88D- p1p6WT
c)
PRWT- p1p6L449F
d)
PRWT - p1p6S451N
e)
PRD30N/N88D- p1p6S451N
f)
PRWT - p1p6L449F/S451N
g)
PRD30N/N88D- p1p6L449F/S451N
444 445 446 447 448 449 450 451 452 453
0.00 -2.00 -4.00 -6.00 -8.00 -10.00 -12.00

Figure IV.6: Total van der Waals Interaction energies of the p1-p6 Peptide. (a) PR_{WT}—p1-$p6_{WT}$ (b) $PR_{D30N/N88D}$—p1-$p6_{WT}$ (c) PR_{WT}—p1-$p6_{L449F}$, (d) PR_{WT}—p1-$p6_{S451N}$, (e) $PR_{D30N/N88D}$—p1-$p6_{S451N}$, (f) PR_{WT}—p1-$p6_{L449F/S451N}$, (g) $PR_{D30N/N88D}$—p1-$p6_{L449F/S451N}$. No significant differences are observed in the overall trend in van der Waals interaction energies across all the structures. Increased van der Waals energies are observed at Gag 449 and 451 in all structures as compared to the PR_{WT}—p1-$p6_{WT}$ structure.

Figure IV.7

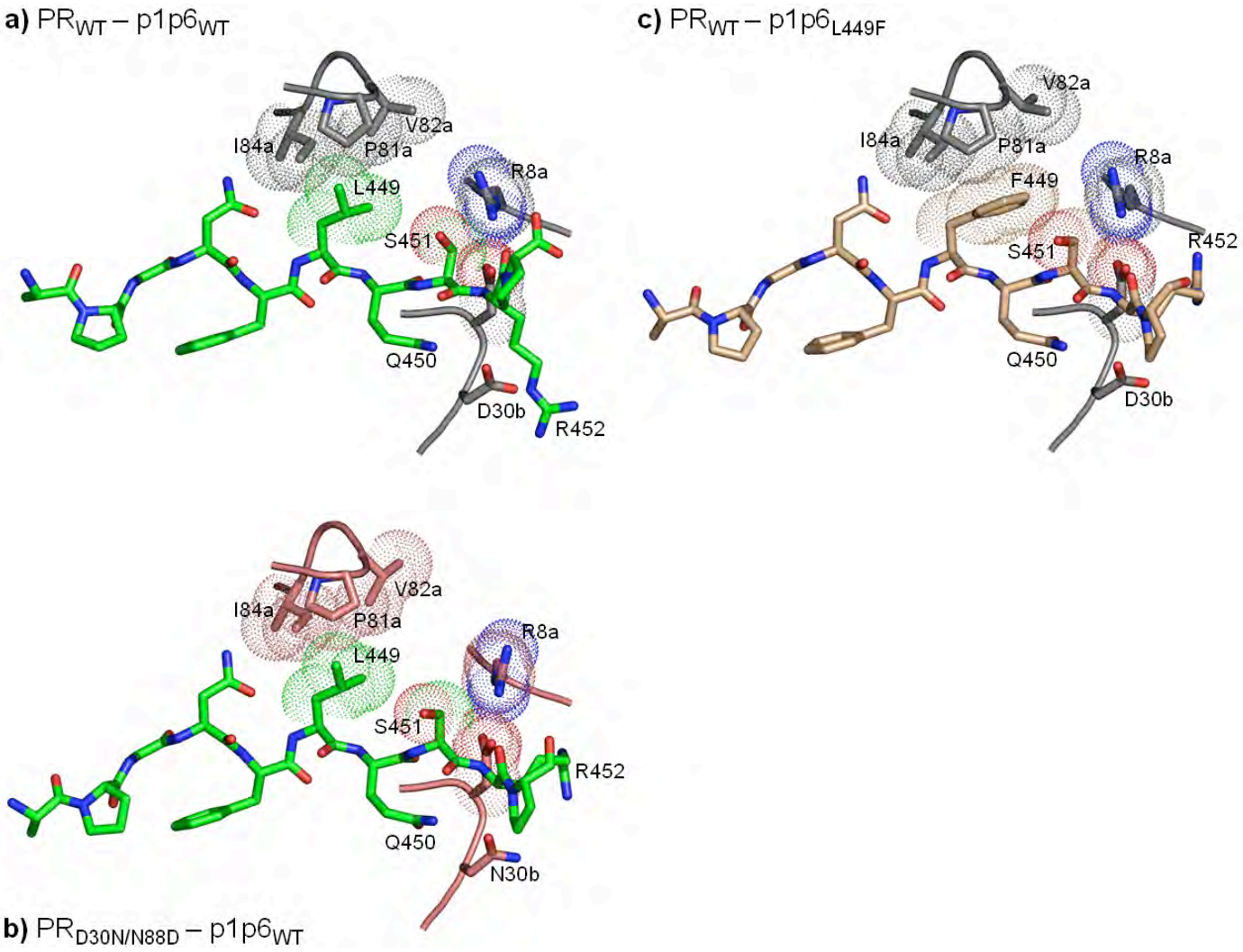

a) PR$_{WT}$ – p1p6$_{WT}$
V82a
I84a
P81a
R8a
L449
S451
Q450
D30b
R452
c) PR$_{WT}$ – p1p6$_{L449F}$
V82a
I84a
P81a
R8a
F449
S451
R452
Q450
D30b
V82a
I84a
P81a
R8a
L449
S451
R452
Q450
N30b
b) PR$_{D30N/N88D}$ – p1p6$_{WT}$

Figure IV.7 *(cont.)*

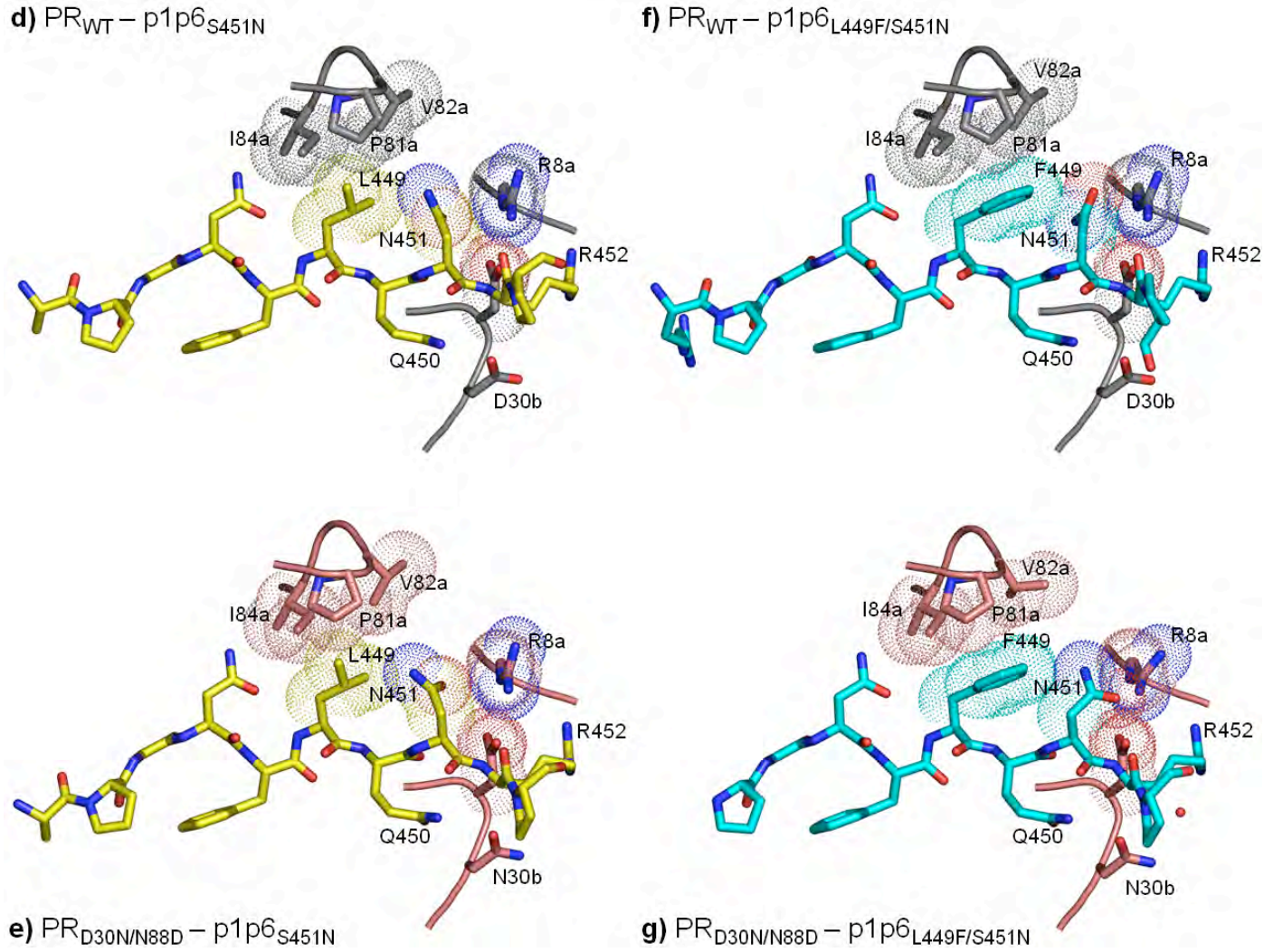


Figure IV.7: van der Waals Interactions around p1-p6 Peptide Residues Gag 449 and Gag 451. (a) PR_{WT}—p1-$p6_{WT}$ (b) $PR_{D30N/N88D}$—p1-$p6_{WT}$ (c) PR_{WT}—p1-$p6_{L449F}$, (d) PR_{WT}—p1-$p6_{S451N}$, (e) $PR_{D30N/N88D}$—p1-$p6_{S451N}$, (f) PR_{WT}—p1-$p6_{L449F/S451N}$, (g) $PR_{D30N/N88D}$—p1-$p6_{L449F/S451N}$. Increased van der Wals interactions are observed at Gag 449 and Gag 451 upon mutation.

Table IV.2		van der Waals Interaction Energy (kcal/mol)						
		PR_{WT}	$PR_{D30N/N88D}$	PR_{WT}	PR_{WT}	$PR_{D30N/N88D}$	PR_{WT}	$PR_{D30N/N88D}$
p1p6 residue	protease residue	$p1p6_{WT}$	$p1p6_{WT}$	$p1p6_{L449F}$	$p1p6_{S451N}$	$p1p6_{S451N}$	$p1p6_{L449F/S451N}$	$p1p6_{L449F/S451N}$
	8A			-0.35				
	23A	-0.12	0.82	-0.03	-0.18	0.31	0.22	0.24
	25A	-0.11	-0.31	-0.30	-0.24	-0.19	-0.28	-0.28
	27B	-0.44	-0.44	-0.53	-0.44	-0.43	-0.53	-0.52
	48B	-	-	-0.25			-0.09	-0.28
	49B	-0.46	-0.73	-0.79	-0.70	-0.74	-0.67	-0.81
Gag449	50B	-0.73	-0.97	-0.45	-0.95	-1.00	-0.79	-0.62
	80A	-0.09	-0.07		-0.08	-0.07		
	81A	0.74	-0.30	-0.65	0.45	-0.13	0.39	-0.61
	82A	0.02	-0.31	-0.39	-0.18	-0.35	-0.22	-0.94
	84A	-0.41	-0.67	-0.47	-0.59	-0.56	-0.62	-0.33
	Total vdW energy	-1.60	-2.97	-4.21	-2.92	-3.16	-2.60	-4.16
	27B	-0.08	-0.08	-0.07	-0.08	-0.08	-0.08	-0.07
	28B	0.44	-0.89	-0.78	-0.72	-0.78	-0.69	-0.74
	29B	-1.07	-1.28	-1.10	-1.19	-1.19	-1.26	-1.14
	30B	-2.15	-1.82	-2.34	-1.29	-1.82	-1.91	-1.95
	32B	-0.45	-0.46	-0.55	-0.46	-0.44	-0.46	-0.47
Gag450	47B	-1.19	-1.12	-1.12	-1.05	-1.17	-1.10	-1.10
	48B	-0.42	-0.25	-0.44	-0.45	-0.23	-0.47	-0.34
	50A	-0.15	-0.18	-0.18	-0.20	-0.15	-0.16	-0.12
	76B	-0.11	-0.14	-0.16	-0.12	-0.13	-0.10	-0.14
	84B				-0.26			
	Total vdW energy	-5.17	-6.22	-6.74	-5.81	-5.98	-6.23	-6.08
	8A	0.85	-0.28	-0.27	-1.26	-0.87	-1.26	-1.00
	29B	-0.45	-0.36	-0.32	-0.82	-0.56	-1.09	-0.66
Gag451	48B	-0.29	-0.92	-0.86	-0.43	-0.67	-0.80	-0.47
	81A				-0.16	-0.08		
	Total vdW energy	0.11	-1.56	-1.45	-2.68	-2.18	-3.15	-2.13
	8A		-0.25	-0.21	-0.77	-0.64	-0.95	-0.89
	29B		-3.25	-3.50	-3.38	-3.05	-3.50	-3.13
	30B	-0.27	-0.37	-0.46	-0.54	0.12	-0.53	-0.30
	45B	-0.04						
	46B	-0.05						
Gag452	47B	-0.14						
	58B	-0.98						
	74B	-0.40						
	76B	-0.83						
	87B		-0.45	-0.51	-0.35	-0.50	-0.33	-0.50
	Total vdW energy	-2.71	-4.31	-4.68	-5.04	-4.07	-5.31	-4.81
protease residue	p1p6 residue							
	7	-1.46	-1.14	-1.49	-0.69	-1.09	-1.20	-1.10
	8	-0.22	-0.13	-0.23	-0.21	-0.13	-0.20	-0.09
30B	9	-0.27	-0.35	-0.62	-1.04	-0.11	-0.96	-0.37
	Total vdW energy	-1.96	-1.62	-2.34	-1.94	-1.33	-2.36	-1.57

Table IV.2: Side Chain van der Waals Interaction Energies for Gag 449-452 of p1-p6 Peptide and 30b of Protease. These van der Waals interaction energies were calculated for only the side chains.

Thus, there is increased van der Waals interactions with the protease as result of the mutation from Leu to a larger Phe.

Interactions at Gag 450. Overall, Gln 450 makes the most number of favorable contacts with the protease as indicated by the large van der Waals interaction energies (Fig IV.6a-g). These are comparable to the PR_{WT}—p1-$p6_{WT}$ and did not vary significantly irrespective of the protease or the p1-p6 variant. The energies of the specific interactions with the protease also follow the same trend (Table IV.2).

Interactions at Gag 451. Interactions made by Gag 451 with the protease are observed in the PR_{WT}—p1-$p6_{S451N}$ (Fig IV.6d and Fig IV.7d), $PR_{D30N/N88D}$ —p1-$p6_{S451N}$ (Fig IV.6e and Fig IV.7e), PR_{WT}—p1-$p6_{L449F/S451N}$ (Fig IV.6g and Fig IV.7f) and $PR_{D30N/N88D}$—p1-$p6_{L449F/S451N}$ (Fig IV.6g and Fig IV.7g) complexes with the Ser to Asn mutation within the p1-p6, and were the most favorable as indicated by the overall van der Waals interaction energies. Analysis of the the specific interactions reveals a large increase in favorable contacts with Arg 8a as a result of the S451N mutation in the p1-p6 peptide in these complexes. The S451N mutation also leads to increased contacts with residues 29b and 48b of the protease. Thus, Gag 451 makes more favorable van der Waals interactions with the protease as a result of the S451N mutation.

Interactions at Gag 452. The overall van der Waals energies show very little differences in the total interactions between Arg 452 and the protease (Fig IV.6). However, a detailed examination of the individual interactions reveals that though the Arg 452 in the PR_{WT}—p1-$p6_{WT}$ interacts with more protease residues as compared to all

the other structures, the sum total energies of the interactions are much less favorable (Table IV.2). Although Arg 452 is oriented a different direction in all the structures compared to the PR_{WT}—p1-$p6_{WT}$ structure, the change in interactions with protease 30b is not notable as indicated by the energies. Most significant are the interactions with protease 8a and 29b, which are not seen in the PR_{WT}—p1-$p6_{WT}$ structure.

Packing around 30b. The D30N mutation does not involve changes in the size of the side chain. However, significant decreases in the total van der Waals interactions between 30b and the p1-p6 peptide are observed in all structures when compared with the PR_{WT}—p1-$p6_{WT}$ structure (Fig IV.5). The specific interactions between 30b and Gag 450 are similar across all structures except PR_{WT}—p1-$p6_{S451N}$ (Table IV.2). Interactions with Gag 451 are very similar for all the structures. 30b makes more interactions with Gag 452 PR_{WT}—p1-$p6_{S451N}$ and PR_{WT}—p1-$p6_{L449F/S451N}$ compared to the remaining complexes (Table IV.2). Though the interactions made by 30b with Gag 450, 451 and 452 residues of the p1-p6 peptide vary slightly between structures, the overall van der Waals interactions are very similar across all structures as indicated by the interaction energies (Table IV.2).

DISCUSSION

Cleavage sites within Gag co-evolve with primary PI-resistance mutations in the HIV-1 protease (Bally et al., 2000; Doyon et al., 1996; Mammano, Petit, and Clavel, 1998; Zhang et al., 1997). Previous studies have shown that the p1-p6 cleavage site co-evolves with the D30N/N88D protease mutations that arise specifically in response to NFV therapy (Chapter II and III). This study reports novel data on structural changes in protease and p1-p6 cleavage site as a result of p1-p6 coevolution with the D30N/N88D protease mutations.

As shown in Chapter II, p1-p6 cleavage site mutations at Gag 449 and 451 mutate in a correlated manner with the NFV-resistant D30N/N88D protease mutations. Specifically, Leu at Gag 449 mutates frequently to Phe and Ser at Gag 451 mutates to an Asn. In order to study the structural changes as a result of these mutations, x-ray crystallographic studies were performed with the WT and D30N/N88D protease variants in complex with the different p1-p6 substrate variants. Crystals and diffraction data were obtained for all seven complexes after considerable effort, however, the structure of the PR_{WT}—$p1\text{-}p6_{L449F}$ could not be solved. Though several crystallization conditions were used, this complex crystallized in only one set of buffer conditions and were fragile and plate-like, probably resulting in the poor diffraction data.

Overall, the structures of both WT and D30N/N88D protease complexed with the different p1-p6 peptides were very similiar to the structure of WT protease complexed with WT (Fig IV.2a,b). Though some backbone shifts were observed in the flap-hinge

region and the surface-exposed region of the core of the protease, these differences are likely due to changes in the crystal packing and flexibility of the solvent exposed part of the core region. The most striking difference is observed in the orientation of the Arg at Gag 452. In all the structures where, either the protease or the substrate is mutated, Arg 452 is oriented differently when compared to the PR_{WT}—p1-$p6_{WT}$ structure (Fig IV.2c). This difference, could not be attributed to differences in crystal packing or buffer conditions, as there are no crystal contacts or symmetry related contacts in that region, nor were there any consistent differences in buffer conditions. In the PR_{WT}-p1-$p6_{WT}$ structure, Arg 452 is part of a hydrogen bonding network involving Asp 30b and Gln 450. The altered conformation of Arg 452 in all the other structures leads to the loss of its hydrogen bond with Asp 30b.

Previous studies have shown that Asp 30b of the protease is important for recognition of the p1-p6 cleavage site (Kolli, Lastere, and Schiffer, 2006; Prabu-Jeyabalan, Nalivaika, and Schiffer, 2002a). In Chapter II, the modeling studies suggested that the D30N mutation would disrupt the network of hydrogen bonds between Gln 450, Asp 30b and Arg 452. Structural studies in this chapter revealed that the D30N mutation serves to disrupt the hydrogen bonds with Arg 452 only (IV.3a-g). A similar disruption of hydrogen bond network was also observed in structures of the WT protease complexes with the various p1-p6 peptides due to the different orientation of Arg 452. However, a water mediated hydrogen bond between Asp/Asn 30b OD1/ND2 and Arg8a NE compensates for the loss of the hydrogen bond between Asp 30b and Arg 452.

Additionally, Arg 452 forms several new hydrogen bonds with water or other solvent molecules.

A detailed analysis of the van der Waals energies for interactions between only the sidechain of Arg 452 and the protease revealed significantly higher values for all structures compared to the PR_{WT}—p1-$p6_{WT}$ structure (Table IV.2). However, the overall van der waals energies for Arg 452 interactions with the protease were in the same range for all structures including the PR_{WT}—p1-$p6_{WT}$ structure (Fig IV.6a-b).

A significant decrease was observed in the overall van der Waals interactions made by 30b with the p1-p6 peptide as compared to the PR_{WT}—p1-$p6_{WT}$ structure, which are compensated by large increases in interactions of Asp 29b with the the p1-p6 peptide (IV.5a-g). In addition, Ile 47b, Gly 48b and Ile 50b protease residues in all structures except PR_{WT}—p1-$p6_{WT}$ structure made more favorable van der Waals interactions with the p1-p6 peptide, further compensating for the loss of interactions.

In chapter II and other studies (Feher et al., 2002; Maguire et al., 2002), it was hypothesized that the L449F mutation would result in increased interactions by this residue with the protease due to the larger size of the Phe residue. This is demonstrated by the van der Waals interaction energies listed in Table IV.2 and Fig IV.6c. A similar increase in interactions at Gag 449 is expected in the $PR_{D30N/N88D}$—p1-$p6_{L449F}$ structure. However, a smaller increase in van der Waals interaction energies is observed in the PR_{WT}—p1-$p6_{L449F/S451N}$ (Fig IV.6f) structure. Thus, it appears that a mutation at Gag 449

leads to increased interactions at 449 leading to the more efficient processing of the p1-p6 cleavage site as shown in previous studies (Feher et al., 2002; Maguire et al., 2002).

Similarly, the larger Asn residue in the peptide with the Gag S451N mutation, causes improved van der Waals interactions with the protease as indicated by the lower van der Waal interaction energies (Table IV.2). This increase in interaction energies is mainly due to a significant increase in contacts with Arg 8a and, to a lesser extent with Asp 29b and 48b, of the protease. Also, several new hydrogen bonds are observed between Asn 451 and Arg 8a or water molecules (Fig IV.3d-g). Together, these changes resulting from the S451N mutation are likely to improve p1-p6 recognition by the WT and D30N/N88D protease.

Finally, asparagine 88 to aspartate is a compensatory mutation that commonly occurs with the D30N, NFV-resistant protease mutation. As described in the results, changes in the hydrogen bonding network around residue 88 were observed as a result of this mutation (Fig IV.4a-g). Most significantly, in all the WT protease complexes Asn 88 ND2 forms a hydrogen bond with Thr74 O that is disrupted upon mutation to Asp. This disruption, along with a few other changes in the hydrogen bonding network, serve to reorient this residue towards the interior of the molecule. This change in hydrogen bonding was also reported in previously solved structures of the NFV-resistant D30N/N88D HIV-1 protease in complex with NFV (PDB 2PYM; (Kozisek et al., 2007). This reorientation of Asp 88 and rearrangement of hydrogen bonds were also observed in the absence of the D30N mutation (Mahalingam et al., 2001). Thus, the mechanism by

which the N88D mutation improves viral fitness/replicative capacity in the context of the D30N mutation is not well undertood. It is also very likely that this protease mutation does not directly play a role in the co-evolution of the p1-p6 cleavage site.

In summary, the p1-p6 cleavage site co-evolves with the D30N/N88D protease mutations as shown in Chapters II and III. Structural studies described here show that the L449F and S451N mutations in the p1-p6 cleavage site significantly increase the overall interactions, both by means of hydrogen bonds and van der Waals interactions with the protease. Previous studies have demonstrated that the L449F mutation improves Gag processing at this site and it is possible that similar effects on Gag processing maybe also be observed with the S451N mutation. Co-occurance of L449F and S451N within the p1-p6 peptide has not been observed previously. In chapter II, it is hypothesised that van der Waals clashes between the two large amino acid residues may prevent this co-occurance. However, no such van der Waals clashes were observed in the PR_{WT}-p1-$p6_{L449F/S451N}$ and the $PR_{D30N/N88D}$-p1-$p6_{L449F/S451N}$ structures. Also, the overall van der Waals interaction energies did not vary significantly from the structures with single mutant peptides. While it appears that the occurrence of either one of the mutations may be sufficient to compensate for the loss of binding efficiency of the WT peptide to the mutant protease, whether the compensation is sufficient to prevent co-occurence is not clear.

Interestingly, the differences observed in the interactions between mutant peptide and mutant protease were also observed in all the structures WT protease and mutant peptides

as well. Earlier have shown that p1-p6 peptide with the L449F mutation was cleaved more efficiently even by the WT protease studies (Feher et al., 2002; Maguire et al., 2002), suggesting that this mutant peptide is a more efficient substrate than the WT p1-p6 cleavage site. This suggests that the co-evolution of this cleavage is not specific for particular protease mutations. This observation is consistent with results from chapter III and other studies (Maguire et al., 2002) that this site co-evolves with other primary resistance mutations as well. Earlier kinetic studies and amino acid preference studies have suggested that the WT p1-p6 cleavage site may not be the most optimum sequence (Feher et al., 2002). This sub-optimal amino acid sequence results in the later cleavage of the p1-p6 site during the processing of Gag. The coevolution of the p1-p6 site resulting in an improved substrate may reestablish the correct order of Gag processing.

References

Bally, F., Martinez, R., Peters, S., Sudre, P., and Telenti, A. (2000). Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: Clinical significance and implications for resistance to protease. *AIDS Res Hum Retrovir* **16**(13)**,** 1209-1213.

Clemente, J. C., Hemrajani, R., Blum, L. E., Goodenow, M. M., and Dunn, B. M. (2003). Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N. *Biochemistry* **42**(51)**,** 15029-35.

Collaborative-Computational-Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50,** 760-763.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**(Web Server issue)**,** W375-83.

DeLano, W. L. (2002). The PyMol Molecular Graphics System. Delano Scientific, San arlos, CA, USA.

Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L., and Lamarre, D. (1996). Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol* **70**(6)**,** 3763-9.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D: Biological Crystallography* **60**(Pt 12 Pt 1)**,** 2126-32.

Feher, A., Weber, I. T., Bagossi, P., Baross, P., Mahalingam, B., Louis, J. M., Copeland, T. D., Yorshin, I. Y., Harrison, R. W., and Tozser, J. (2002). Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. *J Biochem* **269,** 4114-4120.

Johnson, V. A., Brun-Vezinet, F., Clotet, B., Kuritzkes, D. R., Pillay, D., Schapiro, J. M., and Richman, D. D. (2006). Update of the drug resistance mutations in HIV-1: Fall 2006. *Top HIV Med* **14**(3)**,** 125-30.

Kaldor, S. W., Kalish, V. J., Davies, J. n., Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., Dressman, B. A., Hatch, S. D., Khalil, D. A., Kosa, M. B., Lubbehusen, P. P., Muesing, M. A., Patick, A. K., Reich, S. H., Su, K. S., and Tatlock, J. H. (1997). Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J Med Chem* **40**(24)**,** 3979-3985.

Kempf, D. J., Marsh, K. C., Denissen, J. F., McDonald, E., Vasavanonda, S., Flentge, C. A., Green, B. E., Fino, L., Park, C. H., Kong, X. P., and et al. (1995). ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proceedings of the National Academy of Sciences of the United States of America* **92**(7)**,** 2484-8.

Kolli, M., Lastere, S., and Schiffer, C. A. (2006). Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virology* **347**(2)**,** 405-9.

Kolli, M., Stawiski, E., Chappey, C., and Schiffer, C. A. (2009). Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. *Journal of Virology* **83**(21)**,** 11027-42.

Kozisek, M., Bray, J., Rezacova, P., Saskova, K., Brynda, J., Pokorna, J., Mammano, F., Rulisek, L., and Konvalinka, J. (2007). Molecular analysis of the HIV-1 resistance development: enzymatic activities, crystal structures, and thermodynamics of nelfinavir-resistant HIV protease mutants. *Journal of Molecular Biology* **374**(4)**,** 1005-16.

Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**(7)**,** 1171-9.

Maguire, M. F., Guinea, R., Griffin, P., Macmanus, S., Elston, R. C., Wolfram, J., Richards, N., Hanlon, M. H., Porter, D. J., Wrin, T., Parkin, N., Tisdale, M., Furfine, E., Petropoulos, C., Snowden, B. W., and Kleim, J. P. (2002). Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J Virol* **76**(15)**,** 7398-406.

Mahalingam, B., Louis, J. M., Hung, J., Harrision, R. W., and Weber, I. T. (2001). Structural implications of drug-resistant mutants of HIV-1 protease: high-resolution crystal structures of the mutant protease/substrate analogue complexes. *Proteins* **43,** 455-464.

Mammano, F., Petit, C., and Clavel, F. (1998). Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol* **72**(9)**,** 7632-7.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography* **40**(Pt 4)**,** 658-674.

McDonald, I. K., Thornton, J.M. (1994). Satisfying hydrogen bonding potentialin proteins. *J Mol Biol* **238,** 777-793.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **D53,** 240-255.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr D Biol Crystallogr* **A50,** 157-163.

Pai, V. B., and Nahata, M. C. A. P., 33: 325-339 (1999). Nelfinavir mesylate: a protease inhibitor. *Ann Pharmacotherapy* **33**(3)**,** 325-39.

Patick, A., Duran, M., Cao, Y., Shugarts, D., Keller, M., Mazabel, E., Knowles, M., Chapman, S., Kuritzkes, D., and Markowitz, M. (1998). Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob Agents Chemother* **42**(10)**,** 2637-44.

Patick, A., Mo, H., Markowitz, M., Appelt, K., Wu, B., Musick, L., Kalish, V., Kaldor, S., Reich, S., Ho, D., and Webber, S. (1996). Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrob Agents Chemother* **40**(2)**,** 292-7.

Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2000). How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J Mol Biol* **301**(5)**,** 1207-20.

Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2002a). Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure* **10**(3)**,** 369-81.

Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., and Schiffer, C. A. (2004). Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *Journal of Virology* **78**(22)**,** 12446-54.

Prabu-Jeyabalan, M., Nalivaika, E. A., Romano, K., and Schiffer, C. A. (2006). Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *Journal of Virology* **80**(7)**,** 3607-16.

Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2002b). Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes. *Structure* **10**(3)**,** 369-381.

Rose, J. R., Salto, R., and Craik, C. S. (1993). Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *J Biol Chem* **268,** 11939-11945.

Shafer, R. W., and Schapiro, J. M. (2008). HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev* **10**(2)**,** 67-84.

Surleraux, D. L., Tahri, A., Verschueren, W. G., Pille, G. M., de Kock, H. A., Jonckers, T. H., Peeters, A., De Meyer, S., Azijn, H., Pauwels, R., de Bethune, M. P., King, N. M., Prabu-Jeyabalan, M., Schiffer, C. A., and Wigerinck, P. B. (2005). Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *Journal of Medicinal Chemistry* **48**(6)**,** 1813-22.

Vivés, E., Brodin, P., and Leblue, B. (1997). A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *Lett Pept Sci* **4,** 429-436.

Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallographica. Section D: Biological Crystallography* **57**(Pt 1)**,** 122-33.

Zhang, Y. M., Imamichi, H., Imamichi, T., Lane, H. C., Falloon, J., Vasudevachari, M. B., and Salzman, N. P. (1997). Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol* **71,** 6662-6670.

CHAPTER V

DISCUSSION

Despite tremendous advances in scientific research on AIDS and its causative organism, HIV, there are still millions of infected people worldwide with no hope for a permanent cure nor is there a preventive vaccine that could curb the growing epidemic. Two decades of research have led to the development of several antiviral drugs including inhibitors to the viral protease, which have significantly reduced HIV-related morbidity and mortality. The success of these inhibitors is severely compromised due to the development of resistance-associated mutations within the protease. The virus also evolves compensatory mutations, not only in the protease but also in the protease cleavage sites within Gag, which contribute to increased viral fitness and resistance.

Drug resistance is a change in molecular recognition where the protease mutates to decrease inhibitor binding even as it continues to recognize and cleave its substrates. A better understanding of the underlying mechanisms of drug resistance to protease inhibitors (PIs), including substrate co-evolution, will contribute to the design of more robust inhibitors that are effective against the drug-resistant protease variants and prevent further development of drug resistance. My dissertation research focused on investigating substrate co-evolution and examining the effect this has on viral fitness and protease inhibitor susceptibility. In addition, I performed X-ray crystallographic studies to investigate structural changes as a result of this substrate co-evolution.

In Chapters II and III, I examined viral sequences derived from infected individuals treated with protease inhibitors, for mutations in the cleavage sites that correlated with primary drug resistance mutations. I investigated the effect of these correlated mutations

on viral fitness and phenotypic susceptibilities to various PIs. Finally, I studied protease-p1-p6 complexes, in Chapter II and IV, to probe the underlying structural changes occurring as a result of this co-evolution.

Co-evolution of HIV-1 Protease and the Viral Substrates

Associations between specific mutations in the protease and its cleavage sites have been reported previously, and were demonstrated to act as compensatory mutations (Bally et al., 2000; Doyon et al., 1998; Maguire et al., 2002; Mammano, Petit, and Clavel, 1998; Zhang et al., 1997). These associated mutations are likely to become more common as the number of infected individuals receiving protease inhibitor therapy increases. Therefore I performed statistical analyses to investigate co-evolution of substrate cleavage site and protease mutations that are the primary cause of PI resistance (Chapters II and III). My initial study involving a small set of viral sequences from infected individuals that are resistant to protease inhibitor therapy (Chapter II) revealed that the p1-p6 cleavage site mutations correlate with the NFV-resistant D30N/N88D protease mutations. In addition to this, I identified several other correlations between cleavage site mutations and primary drug resistant mutations in Chapter III. Of the four cleavage sites analyzed, mutations are most frequent in the NC-p1 and p1-p6 cleavage sites and correlate with PI resistance protease mutations, in different combinations.

Effect of Substrate Co-evolution on Viral Fitness and PI Susceptibility

Primary PI resistance mutations, especially in the active site, reduce both protease catalytic efficiency and viral replicative capacity (RC) (Bleiber et al., 2001; Croteau et al., 1997; Martinez-Picado et al., 2000; Martinez-Picado et al., 1999). Several studies have demonstrated that the evolution of compensatory mutations within cleavage sites leads to improved viral fitness compensating for the loss in fitness resulting from the protease resistance mutations (Doyon et al., 1996; Mammano, Petit, and Clavel, 1998; Robinson et al., 2000; Zhang et al., 1997). However, in my studies, no significant differences in viral fitness were observed in viruses with protease resistance mutations in the presence and absence of mutations within the Gag cleavage sites (Chapter III). This reflects studies by Dam et al. (Dam et al., 2009), where, although RC was fully restored by cleavage site mutations in the context of WT pNL43 Gag-patient derived protease-RT, the reversion of the same cleavage site mutations to WT in the context of patient derived Gag had little impact on RC. Likely there are fitness determinants in the down-stream region of Gag, especially the p1-p6/p6* region, although mechanism for this is poorly understood (Myint et al., 2004; Whitehurst et al., 2003).

Additionally, in Chapter III, I statistically analyzed the effect of the observed correlations on phenotypic susceptibilities to various PIs, and these correlations were observed to significantly affect PI susceptibilities. In most instances, a significant decrease in phenotypic susceptibility to particular protease inhibitors was observed. Although mutations at either Gag 431 or Gag 437 were not associated with D30N/N88D protease

mutations, significantly lower PI susceptibilities were observed. A similar trend was also observed with Gag A431V and the L90M protease mutation. Mutations at either of these residues within the NC-p1 cleavage site likely directly enhance resistance to PIs as was observed and demonstrated previously (Dam et al., 2009; Parkin, 2005). At least in the case of the Gag A431V mutation this is likely due to enhanced Gag processing at this site as demonstrated by Nijhuis et al., (Nijhuis et al., 2007). Thus, I observed that Gag cleavage site mutations enhance resistance to PIs in combination with primary drug resistance mutations in the protease.

Structural Insights into Substrate Co-evolution

As observed in Chapter III, the NC-p1 and p1-p6 cleavage sites mutated very frequently and were associated with primary drug resistance mutations in the protease. Although several studies demonstrated the effect of both the Gag A431V and Gag L449F mutations on Gag processing (Feher et al., 2002; Maguire et al., 2002), only a few studies examined the structural changes resulting from substrate co-evolution (Prabu-Jeyabalan et al., 2004). To this end, I performed structural studies using the co-evolution of the p1-p6 cleavage site with the D30N/N88D protease mutation as an example. Within the p1-p6 cleavage site, mutations occur specifically at Gag 449, Gag 451 or Gag 453 (Chapters II and III). Leu 449 mutated frequently to a Phe whereas Ser 451 mutated to an Asn. However, the simultaneous occurrence of these two mutations was infrequent and I hypothesized to be due to unfavorable van der Waals contacts (Chapter II). Since Pro 453 is located outside the substrate-binding pocket and is frequently disordered in crystal

structures of protease-p1-p6 complexes no meaningful structural conclusions can be made. Therefore this mutation was not included in this structural analysis. Thus, WT and D30N/N88D protease variants in complex with p1-p6 variants including either L449F or S451N or both mutations, were analyzed for structural changes resulting from this co-evolution.

Structural studies in Chapter IV show that both L449F and S451N mutations in the p1-p6 cleavage site significantly increase the overall interactions with the protease active site, both by means of hydrogen bonds and van der Waals interactions, and is in agreement with previous studies in the case of the L449F mutation (Feher et al., 2002). The L449F mutation has been demonstrated to improve Gag processing at this site (Maguire et al., 2002). Further studies are warranted to investigate if similar effects on Gag processing are observed with the S451N mutation. The two larger residues at Gag 449 and 451, Phe and Asn, respectively are able to pack well in both PR_{WT}-p1-$p6_{L449F/S451N}$ and the $PR_{D30N/N88D}$-p1-$p6_{L449F/S451N}$ structures. The occurrence of either one of the mutations may be sufficient to compensate for the loss of binding efficiency of the WT peptide to the mutant protease, but there is no obvious selective advantage for having both changes simultaneously.

The overall structures of the complexes are very similar to each other and the PR_{WT}—p1-$p6_{WT}$ structure though significant differences are observed within the p1-p6 peptide, which affect the interactions with protease 30b. Modeling studies in Chapter II suggest that the D30N mutation likely affects p1-p6 recognition as a result of disruption of strong

hydrogen bond interactions between 30b and Arg 452 of the p1-p6 peptide. This was confirmed in the crystal structures of D30N/N88D protease complexes but unexpectedly also the WT protease complexes with variants of p1-p6 (Chapter III). The disruption of the interaction between 30 and 452 is due to a large conformational change involving both the backbone and side chain of R452. This conformational change cannot be attributed to differences in crystal packing or buffer conditions as there are no crystal contacts or symmetry related contacts in that region, nor are there any consistent differences in buffer conditions. Rather, the rearrangements of the packing of the substrate either due to changes of the protease or the substrate result in a suboptimal packing and thus change in orientation of R452.

Contrary to the predictions made in Chapter II, my structural analyses reveal that the differences observed in the interactions between mutant protease and mutant peptide compared to WT protease with WT p1-p6 peptide were also observed in structures of WT protease and mutant peptides. Earlier studies demonstrating efficient cleavage of the L449F p1-p6 peptide even by the WT protease studies (Feher et al., 2002; Maguire et al., 2002), show that this mutant peptide is a more efficient substrate than the WT p1-p6 cleavage site. This, and my results from Chapter III suggest that the co-evolution of the p1-p6 cleavage may not be specific for particular protease mutations, however may play a role in modulating Gag processing.

Future Directions

Structural Changes due to Substrate Co-evolution

I've shown, in Chapter III, that different cleavage sites mutate depending on the mutation at residue 50. NC-p1 mutations correlate with the I50L protease mutation whereas with I50V, we observe correlated mutations within the p1-p6 cleavage site. Structural studies of wild-type and mutant NC-p1 and p1-p6 in complex with I50L and I50V proteases respectively, will give information regarding structural changes as a result of the correlated mutations. Additionally, these studies will give insights into this differential pattern of co-evolution of NC-p1 and p1-p6 substrates with the I50L and I50V protease mutations respectively. Similar studies can be performed with other observed co-evolving pairs of mutations.

Effect of Substrate Co-evolution on Gag Processing

The WT p1-p6 cleavage site, like the NC-p1 cleavage site, may not be the most optimum sequence (Feher et al., 2002) and this sub-optimal amino acid sequence likely influences the sequential processing of Gag. Both A431V and L449F mutations within NC-p1 and p1-p6 cleavage sites, respectively, alter Gag processing by the wild-type protease. Previous studies have shown that both Gag A431V and Gag L449F mutations enhance Gag processing and influence viral fitness (Feher et al., 2002; Maguire et al., 2002; Maguire et al., 2001; Myint et al., 2004). The enhanced processing of Gag A431V is likely the result of additional contacts with the protease (Prabu-Jeyabalan et al., 2004).

Peptides corresponding to the p1-p6 cleavage site with the L449F mutation have previously been shown to be a better substrate for the protease (Feher et al., 2002). In addition, this mutation has been shown to enhance Gag processing (Maguire et al., 2002). Thus, to study the effect of mutations in the protease and/or the co-evolving cleavage sites, one can measure the relative cleavage rates at different sites during Gag processing by the protease as shown by Pettit et al (Pettit et al., 2002). The relative rates of cleavage of wild-type and mutant Gag, containing the p1-p6 mutations, by the D30N/N88D protease could be assayed and compared with processing by the wild-type protease. Similar studies should be performed with the I50L and I50V proteases to determine the effect of NC-p1 and p1-p6 co-evolution on Gag processing. I predict that the above mentioned protease mutations will likely affect Gag processing, either by slowing down the relative rates or by changing the order of cleavage between the cleavage sites. If the order of cleavage is altered, possibly the co-evolved substrate sites may restore the correct order of Gag processing. Thus these experiments to study processing of full length Gag variants by WT and mutant proteases will give insights into how co-evolving mutations influence Gag processing, and in turn influence viral fitness and/or resistance to various inhibitors.

Influence of Gag Non-Cleavage Site Mutations on Resistance

The p1-p6^{Gag}-p6^{pol} regions of Gag/Gag-Pol is one of the least understood regions of the viral genome. The p6^{Gag} is known to facilitate viral budding (Gottlinger et al., 1991; Huang et al., 1995; Yu et al., 1995), Pol and Env packaging (Ott et al., 1999; Yu et al.,

1998), and has also been implicated in controlling particle size (Garnier et al., 1999; Garnier et al., 1998), and $p6^{Pol}$ acts as a regulator of protease activation (Chatterjee et al., 2005; Chiu et al., 2006; Dautin, Karimova, and Ladant, 2003; Louis et al., 1998; Partin et al., 1991). The $p6^{Gag}$ protein has a highly-conserved N-terminal P(T/S)AP motif which mediates binding to the cellular protein Tsg101 (Demirov, Orenstein, and Freed, 2002; Garrus et al., 2001; Martin-Serrano, Zang, and Bieniasz, 2001; Pornillos et al., 2002; VerPlank et al., 2001). There have been several reports that suggest that changes in this region of Gag, including duplications and/or insertions, in viruses from patients undergoing anti-retroviral therapy lead to increased infectivity and increased resistance to RT and protease inhibitors (Doyon et al., 1998; Gallego et al., 2002). In Chapter III I showed that mutations within the PTAP region were more frequent in combination with protease resistance mutations than in the absence. These mutations/ duplications/ insertions may improve viral fitness and/or enhanced resistance to protease inhibitors, either in the presence or absence of protease resistance mutations. A statistical analysis can be performed to look for association between these changes in $p6^{Gag}$ and replicative capacity and resistance to protease inhibitors. One can also investigate the association of changes in the $p6^{Gag}$ region with protease resistance mutations. In depth viral studies incorporating these changes within the $p6^{Gag}$ could give insights into the various mechanisms leading to alterations in viral fitness and resistance.

More recently, Parry et al., have demonstrated that mutations in the matrix and partial capsid in the N-terminal regions of Gag fully restore RC to WT levels and thus play a key role in fitness (Parry et al., 2009). More specifically, their studies have narrowed down

the mutations to the following: R76K, Y79F, S81A and a double insertion Q116TQ. In addition to improving fitness, these mutations significantly enhanced resistance to PIs even in the absence of PI resistance mutations in the protease (Parry et al., 2009). Although this study was based on a single clinical isolate, other mutations in the N-terminal region of Gag outside of the cleavage sites may influence viral fitness and/or protease inhibitor resistance. The Monogram Biosciences database could be examined and statistical analyses performed to investigate resistance mutations in the protease and their associations with mutations in the N-terminal region of Gag. Biochemical and viral studies can be performed to investigate the effect of these mutations on Gag processing, viral fitness and resistance to protease inhibitors. Additionally, modeling studies and/or structural studies could provide insights into how mutations in these distant regions of Gag are likely to influence viral fitness and resistance to protease inhibitors. These studies will also increase our understanding of the structure of Gag and tertiary interactions between the protease and Gag.

Implications for Diagnosis and Treatment and Drug Design Strategies

Structural and biochemical insights into mechanisms by which substrate co-evolution alters viral fitness and resistance will enhance our understanding of substrate recognition and cleavage as well as the development of resistance. The detailed structural analyses of various protease-substrate complexes, including those I described in Chapter III, in combination with our knowledge of interactions between mutant protease variants and existing protease inhibitors will significantly contribute to design of better protease

inhibitors. Structural studies of the co-evolving mutations within protease and the NC-p1 and p1-p6 cleavage sites could lead to the calculation of a new substrate envelope, as has been done previously with wild type protease-substrate complexes. New inhibitors may be designed to fit within the new substrate envelope such that they are functional not only against mutant protease variants and but will also leave no room for co-evolution.

Resistance testing strategies involve phenotypic drug susceptibility assays that measure level of resistance to various protease and RT inhibitors. In addition, replicative capacity assays are also performed to measure infectivity of the virus, which reflects the overall viral fitness. Currently, these are single-cycle replicon assays that include the PR and RT regions amplified from patient plasma samples. However, my research in Chapter III showed that mutations within the NC-p1 and p1-p6 cleavage sites enhanced resistance to protease inhibitors, in combination with protease resistance mutations. Additionally, several other studies also demonstrated that mutations within the NC-p1 cleavage site enhanced resistance even in the absence of resistance mutations within the protease. Thus, these studies suggest that mutations within Gag must also be considered when assessing resistance to various protease inhibitors.

Summary

Drug resistance is a change in molecular recognition where, balance is maintained between loss of inhibitor binding even as substrate recognition is maintained. When this balance is disturbed, the substrates are likely to evolve so as to maintain viral fitness even as they contribute to drug resistance. My research showed that the NC-p1 and p1-p6

cleavage sites co-evolve even as the protease continues to accumulate primary resistance-associated mutations within the active site. Though NC-p1 and p1-p6 co-evolution are observed not to contribute significantly to viral fitness, there were extensive effects on phenotypic susceptibilities to various protease inhibitors. In most instances I observed a significant decrease in susceptibility to various protease inhibitors, whereas there were cases where the co-evolution actually increased susceptibility to particular inhibitors. My structural studies of the p1-p6 co-evolution with the D30N/N88D protease showed that loss of p1-p6-protease interactions as a result of the D30N mutation is likely to be compensated by improved van der Waals contacts as a result of correlated mutations in the p1-p6 cleavage site, in a manner similar to the A431V–V82A co-evolution. These findings only just begin to describe the interplay between protease and substrate co-evolution, details of which will be elucidated as more extensive viral sequencing is routinely performed on patient isolates. Thus, my research furthers our understanding of the evolution of resistance to protease inhibitors and also provides insights into the biology of HIV-1. Additionally, the results presented in my thesis dissertation, and by extension to other regions of Gag, could also be used in future strategies to design more potent protease inhibitors.

References:

Bally, F., Martinez, R., Peters, S., Sudre, P., and Telenti, A. (2000). Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: Clinical significance and implications for resistance to protease. *AIDS Res Hum Retrovir* **16**(13)**,** 1209-1213.

Bleiber, G., Munoz, M., Ciuffi, A., Meylan, P., and 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. *Journal of Virology* **75**(7)**,** 3291-300.

Chatterjee, A., Mridula, P., Mishra, R. K., Mittal, R., and Hosur, R. V. (2005). Folding regulates autoprocessing of HIV-1 protease precursor. *Journal of Biological Chemistry* **280**(12)**,** 11369-78.

Chiu, H. C., Wang, F. D., Chen, Y. M., and Wang, C. T. (2006). Effects of human immunodeficiency virus type 1 transframe protein p6* mutations on viral protease-mediated Gag processing. *Journal of General Virology* **87**(Pt 7)**,** 2041-6.

Croteau, G., Doyon, L., Thibeault, D., McKercher, G., Pilote, L., and Lamarre, D. (1997). Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. *J Virol* **71**(2)**,** 1089-96.

Dam, E., Quercia, R., Glass, B., Descamps, D., Launay, O., Duval, X., Krausslich, H. G., Hance, A. J., and Clavel, F. (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**(3)**,** e1000345.

Dautin, N., Karimova, G., and Ladant, D. (2003). Human immunodeficiency virus (HIV) type 1 transframe protein can restore activity to a dimerization-deficient HIV protease variant. *Journal of Virology* **77**(15)**,** 8216-26.

Demirov, D. G., Orenstein, J. M., and Freed, E. O. (2002). The late domain of human immunodeficiency virus type 1 p6 promotes virus release in a cell type-dependent manner. *Journal of Virology* **76**(1)**,** 105-17.

Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L., and Lamarre, D. (1996). Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol* **70**(6)**,** 3763-9.

Doyon, L., Payant, C., Brakier-Gingras, L., and Lamarre, D. (1998). Novel Gag-Pol frameshift site in human immunodeficiency virus type 1 variants resistant to protease inhibitors. *J Virol* **72**(7)**,** 6146-50.

Feher, A., Weber, I. T., Bagossi, P., Baross, P., Mahalingam, B., Louis, J. M., Copeland, T. D., Yorshin, I. Y., Harrison, R. W., and Tozser, J. (2002). Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. *J Biochem* **269,** 4114-4120.

Gallego, O., Corral, A., de Mendoza, C., and Soriano, V. (2002). Prevalence of the HIV protease mutation N88S causing hypersensitivity to amprenavir. *Clinical Infectious Diseases* **34**(9)**,** 1288-9.

Garnier, L., Parent, L. J., Rovinski, B., Cao, S. X., and Wills, J. W. (1999). Identification of retroviral late domains as determinants of particle size. *Journal of Virology* **73**(3)**,** 2309-20.

Garnier, L., Ratner, L., Rovinski, B., Cao, S. X., and Wills, J. W. (1998). Particle size determinants in the human immunodeficiency virus type 1 Gag protein. *Journal of Virology* **72**(6)**,** 4667-77.

Garrus, J. E., von Schwedler, U. K., Pornillos, O. W., Morham, S. G., Zavitz, K. H., Wang, H. E., Wettstein, D. A., Stray, K. M., Cote, M., Rich, R. L., Myszka, D. G., and Sundquist, W. I. (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107**(1)**,** 55-65.

Gottlinger, H. G., Dorfman, T., Sodroski, J. G., and Haseltine, W. A. (1991). Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proceedings of the National Academy of Sciences of the United States of America* **88**(8)**,** 3195-9.

Huang, M., Orenstein, J. M., Martin, M. A., and Freed, E. O. (1995). p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *Journal of Virology* **69**(11)**,** 6810-8.

Louis, J. M., Dyda, F., Nashed, N. T., Kimmel, A. R., and Davies, D. R. (1998). Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. *Biochemistry* **37**(8)**,** 2105-10.

Maguire, M. F., Guinea, R., Griffin, P., Macmanus, S., Elston, R. C., Wolfram, J., Richards, N., Hanlon, M. H., Porter, D. J., Wrin, T., Parkin, N., Tisdale, M., Furfine, E., Petropoulos, C., Snowden, B. W., and Kleim, J. P. (2002). Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J Virol* **76**(15)**,** 7398-406.

Maguire, M. F., MacManus, S., Griffin, P., Guinea, C., Harris, W., Richards, N., Wolfram, J., Tisdale, M., Snowden, W., and Kleim, J. P. (2001). Interaction of HIV-1 protease and *gag* gene mutations in response to amprenavir-selective pressure exerted in amprenavir-treated subjects - contribution of Gag p6 changes L449F and P453L. *Antivir Ther* **6 (Suppl 1),** 48.

Mammano, F., Petit, C., and Clavel, F. (1998). Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol* **72**(9)**,** 7632-7.

Martin-Serrano, J., Zang, T., and Bieniasz, P. D. (2001). HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nature Medicine* **7**(12)**,** 1313-9.

Martinez-Picado, J., Savara, A. V., Shi, L., Sutton, L., and D'Aquila, R. T. (2000). Fitness of human immunodeficiency virus type 1 protease inhibitor-selected single mutants. *Virology* **275**(2)**,** 318-22.

Martinez-Picado, J., Savara, A. V., Sutton, L., and D'Aquila, R. T. (1999). Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *Journal of Virology* **73**(5)**,** 3744-52.

Myint, L., Matsuda, M., Matsuda, Z., Yokomaku, Y., Chiba, T., Okano, A., Yamada, K., and Sugiura, W. (2004). Gag non-cleavage site mutations contribute to full recovery of viral fitness in protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob Agents Chemother* **48**(2)**,** 444-52.

Nijhuis, M., van Maarseveen, N. M., Lastere, S., Schipper, P., Coakley, E., Glass, B., Rovenska, M., de Jong, D., Chappey, C., Goedegebuure, I. W., Heilek-Snyder, G., Dulude, D., Cammack, N., Brakier-Gingras, L., Konvalinka, J., Parkin, N., Krausslich, H. G., Brun-Vezinet, F., and Boucher, C. A. (2007). A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med* **4**(1)**,** e36.

Ott, D. E., Chertova, E. N., Busch, L. K., Coren, L. V., Gagliardi, T. D., and Johnson, D. G. (1999). Mutational analysis of the hydrophobic tail of the human immunodeficiency virus type 1 p6(Gag) protein produces a mutant that fails to package its envelope protein. *Journal of Virology* **73**(1)**,** 19-28.

Parkin, N., Chappey,C.,Lam, E., Petropoulos, C. (2005). Reduced susceptibilityto protease inhibitorsin the ansence of primary protease inhibitor resistance-associated mutations. *Antivir Ther* **10,** S118.

Parry, C. M., Kohli, A., Boinett, C. J., Towers, G. J., McCormick, A. L., and Pillay, D. (2009). Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1. *Journal of Virology* **83**(18)**,** 9094-101.

Partin, K., Zybarth, G., Ehrlich, L., DeCrombrugghe, M., Wimmer, E., and Carter, C. (1991). Deletion of sequences upstream of the proteinase improves the proteolytic processing of human immunodeficiency virus type 1. *Proceedings of the National Academy of Sciences of the United States of America* **88**(11)**,** 4776-80.

Pettit, S. C., Henderson, G. J., Schiffer, C. A., and 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.

Pornillos, O., Alam, S. L., Rich, R. L., Myszka, D. G., Davis, D. R., and Sundquist, W. I. (2002). Structure and functional interactions of the Tsg101 UEV domain. *EMBO Journal* **21**(10)**,** 2397-406.

Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., and Schiffer, C. A. (2004). Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *Journal of Virology* **78**(22)**,** 12446-54.

Robinson, L. H., Myers, R. E., Snowden, B. W., Tisdale, M., and Blair, E. D. (2000). HIV type 1 protease cleavage site mutations and viral fitness: implications for drug susceptibility phenotyping assays. *AIDS Res Hum Retroviruses* **16,** 1149-1156.

VerPlank, L., Bouamr, F., LaGrassa, T. J., Agresta, B., Kikonyogo, A., Leis, J., and Carter, C. A. (2001). Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). *Proceedings of the National Academy of Sciences of the United States of America* **98**(14)**,** 7724-9.

Whitehurst, N., Chappey, C., Petropoulos, C., Parkin, N., and Gamarnik, A. (2003). Polymorphisms in p1-p6/p6* of HIV type 1 can delay protease autoprocessing and increase drug susceptibility. *AIDS Research and Human Retroviruses* **19**(9)**,** 779-84.

Yu, X. F., Dawson, L., Tian, C. J., Flexner, C., and Dettenhofer, M. (1998). Mutations of the human immunodeficiency virus type 1 p6Gag domain result in reduced retention of Pol proteins during virus assembly. *Journal of Virology* **72**(4)**,** 3412-7.

Yu, X. F., Matsuda, Z., Yu, Q. C., Lee, T. H., and Essex, M. (1995). Role of the C terminus Gag protein in human immunodeficiency virus type 1 virion assembly and maturation. *Journal of General Virology* **76 (Pt 12),** 3171-9.

Zhang, Y. M., Imamichi, H., Imamichi, T., Lane, H. C., Falloon, J., Vasudevachari, M. B., and Salzman, N. P. (1997). Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol* **71,** 6662-6670.

APPENDIX I

VIRAL ASSAYS – MEASUREMENTS OF PHENOTYPIC SUSCEPTIBILITIES AND REPLICATIVE CAPACITIES

The PhenoSense Assay™

The PhenoSense Assay™ is used by Monogram Biosciences for rapid and accurate measurement of susceptibility of clinical isolates to most currently available anti-retroviral drugs. In Chapter III, I used both sequence and drug susceptibility data generated to perform statistical analyses to examine the effect of substrate co-evolution on resistance to various protease inhibitors. The assay, adapted from Petropoulos et al.,(Petropoulos et al., 2000) is described below in detail.

Antiviral Drugs. The following protease inhibitors are used to determine the phenotypic susceptibility profile of the patients' HIV-1 isolates: amprenavir (APV), atazanavir (ATV), indinavir (IDV), lopinavir (LPV), nelfinavir (NFV), ritonavir (RTV), saquinavir (SQV), and tipranavir (TPV).

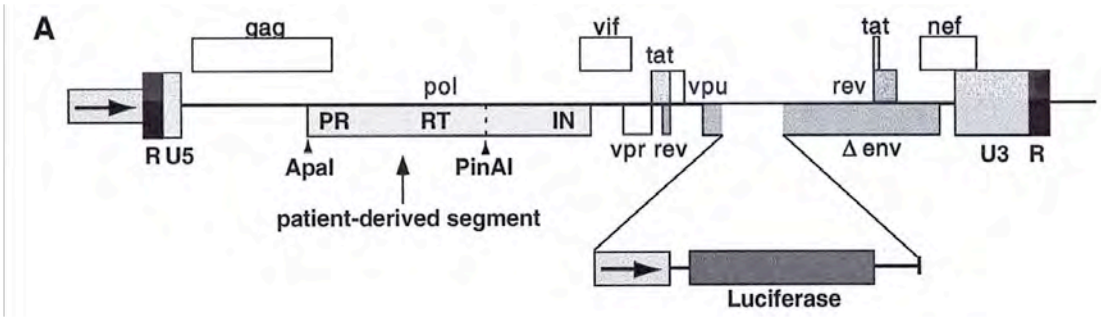
Preparation of sample and amplification. The patient viral particles are pelleted by centrifugation at 20,400 x g for 60 min for 1ml of plasma. The viral particles are disrupted and part of the viral genome is amplified with a forward primer containing the *Apa*I site and a reverse primer containing the *Pin*AI site. The 1.5kb amplification product includes the NC-p1-p6 cleavage sites in *gag* and the PR and RT coding regions within *pol*.

Resistance test vectors (RTV). The retroviral vector used to measure drug susceptibilities is constructed using a modified infectious vector derived from the NL4-3 HIV-1 clone. The vector, referred to as an indicator gene viral vector (IGVV), is

replication defective and lack the PR and RT genes. A region of the envelope (*env*) gene is deleted and in its place is a luciferase expression cassette. The amplified patient-derived viral PR and RT regions were digested using the ApaI and PinAI restriction sites, purified by agarose gel electrophoresis and then ligated to the IGVV DNA that was digested using the same restriction enzymes, thus forming the RTVs. Libraries of the RTVs are prepared in order to capture and preserve the heterogeneity of the patient viral populations. Ligation reactions are used to transform competent E Coli and plated onto agar. The number of colonies are used as an estimate of the number of viral segments in each RTV library. Libraries less than 100 clones are not considered representative of the patient virus.

Protease Inhibitor Susceptibility Assay (Fig A I.1). Viral stocks for the assay are prepared by co-transfection of human embryonic kidney 293 cells (HEK 293, host cells) with RTV plasmid DNA and an expression vector encoding the ENV proteins of amphotrophic murine leukemia virus. Host cells are trypsinized 16 hours after transfection and distributed into 96 well plates containing serial dilutions of protease inhibitors. This was done for each of the protease inhibitors tested. Viral particles generated in the presence of the protease inhibitors are harvested 48 hours after transfection and used to infect fresh 293 cultures in 96 well plates in the absence of drugs. Measurement of luciferase activity 48 hours after infection indicates the amount of replication. The protease inhibitors affect luciferase activity in a dose-dependent manner, thus allowing for a quantitative measure of drug susceptibility. Low levels of

Figure A I.1


A
gag
vif
tat
nef
pol
vpu
rev
PR
RT
IN
R U5
ApaI
PinAI
vpr rev
Δ env
U3 R
patient-derived segment
Luciferase

B

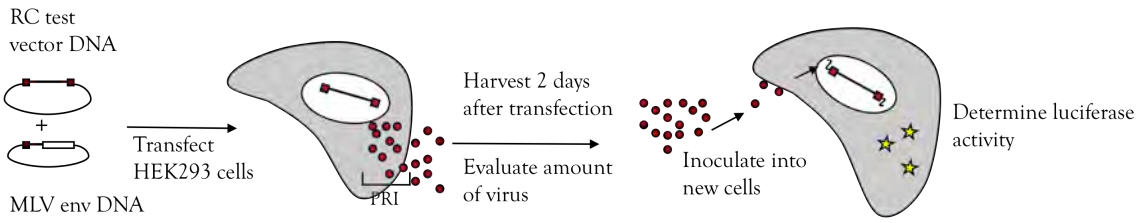

RC test
vector DNA
+
MLV env DNA
Transfect
HEK293 cells
PRI
Harvest 2 days
after transfection
Evaluate amount
of virus
Inoculate into
new cells
Determine luciferase
activity

Figure A I.1: Structure of the resistance test vector (RTV) and overview of the assay. (A) PR and RT sequences from patient-derived viral samples are inserted into a luciferase containing viral vector using the *Apa*I and *Pin*AI restriction sites. (B) RTV DNA and Amphotropic MLV DNA are transfected into cells and viral particles harvested. After harvesting, the viral particles are used to infect fresh target cells. Measurement of luciferase activity in the target cells quantitates the ability of the virus to complete one round of replication. When measuring replicative capacity, this assay is performed in the absence of any anti-retroviral drugs. Susceptibility to either protease inhibitors or RT inhibitors is measured by adding protease inhibitors to transfected cells and RT inhibitors to infected cells.

luciferase activity is indicative of a susceptible virus whereas high levels of luciferase activity indicate a resistant virus.

The percent inhibition is calculated as follows: [1 – (luciferase activity in the presence of drug/luciferase activity in the absence of drug)] X 100. For each drug concentration this is repeated several times and the mean percent inhibition determined by a bootstrapping procedure. This data is displayed by plotting the percent inhibition of luciferase activity versus $\log_{10}$ drug concentration. Inhibition curves defined by the four-parametric sigmoidal function $f(x) = a - [b/(1+(x/c)^d)]$ are fit to the data and used to calculate drug concentrations required to inhibit viral replication by 50% (IC_{50}). This calculated IC_{50} from the sample virus is then compared with the IC_{50} for a drug-sensitive reference NL4-3 strain of HIV-1, to calculate and determine the fold change in drug susceptibility.

Replicative Capacity™

Replicative Capacity™ of HIV measured by Monogram Biosciences gives a direct measure of infectivity i.e., the ability of the virus to infect target cells (PR activity) and to complete the reverse transcription and integrase steps of the life cycle (RT and Int activity). As described in Chapter III, I used both sequence and replicative capacity to examine the effect of substrate co-evolution on viral fitness. The replicative capacity of clinical isolates was measured as described in Deeks et al., (Deeks et al., 2001) using a modified version of the PhenoSense Assay™described in the previous section.

Assay description (Fig A I.1). Patient-derived viral PR and RT are prepared and amplified and resistance test vectors (RTV) are constructed as described in the previous section for the PhenoSense Assay™. The viruses are harvested 48 hours after transfection. The amount of virus produced is evaluated and then inoculated into fresh cell cultures and incubated for 48 hours. Luciferase activity in infected cells is determined after normalizing for the amount if input virus. This luciferase activity gives a measure of the ability of the virus to replicate in the absence of anti-retroviral drugs. The relative RC of the patient-derived virus is calculated as a ratio of the luciferase activity from patient-derived viruses to luciferase activity from reference virus (wild-type) derived from a molecular clone of HIV-1 (NL4-3). RC measurements are then expressed as a percent of the wild-type reference virus. Measurements less than 100 % indicate reduced replicative capacity of the virus when compared to that of the reference wild-type virus.

References:

Deeks, S. G., Wrin, T., Liegler, T., Hoh, R., Hayden, M., Barbour, J. D., Hellmann, N. S., Petropoulos, C. J., McCune, J. M., Hellerstein, M. K., and Grant, R. M. (2001). Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia. *New England Journal of Medicine* **344**(7)**,** 472-80.

Petropoulos, C. J., Parkin, N. T., Limoli, K. L., Lie, Y. S., Wrin, T., Huang, W., Tian, H., Smith, D., Winslow, G. A., Capon, D. J., and Whitcomb, J. M. (2000). A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob Agents Chemother* **44**(4)**,** 920-8.

APPENDIX II

STATISTICAL METHODS

Investigate co-evolution of Gag cleavage sites and primary PI resistance-associated mutations in the protease

1. Co-evolution of Gag cleavage sites with D30N/N88D protease mutations

I investigated the association of Gag cleavage site mutations with the D30N/N88D protease mutations. The analysis focused on qualitative changes in cleavage-site sequences, as compared to the HXB2 sequence, that covary with the D30N/N88D drug-resistant substitutions in the protease. For each protease sequence with the D30N/N88D mutations, the corresponding substrate cleavage sites were scored for the presence or absence of mutations. The simultaneous occurrence of cleavage-site mutations and the D30N/N88D protease mutations were compared for each of the seven substrates studied.

Method. In order to examine association between mutations in the Gag cleavage sites and the D30N/N88D protease mutations, I used the Chi-square test of independence. The null and alternate hypotheses tested are:

H_0: Association of mutations in the cleavage sites and the D30N/N88D protease mutations are independent.

H_A: Mutations in the cleavage sites and the D30N/N88D protease mutations are not independent.

The Chi-square test of independence is used to compare two nominal variables, each with two or more variables and tests the null hypothesis that the two variables are independent and is calculated as follows:

$$X^2 = \sum_{i=1}^{k}\left[\frac{(O_i - E_i)^2}{E_i}\right] \quad (1)$$

In this case there are two nominal variables: the presence and absence of particular cleavage site mutations each with two variables, the presence or absence of D30N/N88D. A data set like this is called a 2 X 2 contingency table and the Chi square statistic, X^2, was calculated by the following shortcut formula:

$$X^2 = \frac{n(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)} \quad (2)$$

where *a* , *b*, *c*, and *d* are the observed cell frequencies. The degrees of freedom for a 2 X 2 table is calculated to be 1 using the rule $(r - 1)(c - 1)$, where r is the number of rows and c, the number of columns. I also used the Yates correction for continuity when the total number of samples, N, is less than 20 or when the expected cell frequency was smaller than 5. The corrected Chi-square is then calculated as follows:

$$X^2_{corrected} = \frac{n((ad - bc) - 0.5n)^2}{(a + c)(b + d)(a + b)(c + d)} \quad (3)$$

I used an interactive web-based Pearson's Chi-square calculator to calculate the Chi-square statistic based on which the null hypothesis (H_0) is accepted or rejected. Associations where the *p* values are < 0.05 are considered significant.

In addition, I calculated the phi correlation coefficient to determine positive or negative correlations. The phi coefficient of correlation is used to measure the strength of association between two variables. This method is used when the two variables being studied are binary variables i.e. yes or no variables coded by 1 or 0 respectively. For this kind of binary data, this is modified and expressed in terms of p(for 1) and q(for 0) as

$$r_{\varphi}(\text{or simply } \varphi) = \frac{p_{xy} - p_x p_y}{\sqrt{p_x q_x p_y q_y}} \quad (4)$$

The phi coefficient of correlation is also related to the X^2 in the following manner:

$$\varphi = \sqrt{\frac{X^2}{N}} \quad (5)$$

I used the above equation to calculate the phi coefficient of correlation in this study. The phi coefficients vary from -1 (perfect negative correlation) to +1 (perfect positive correlation) through 0 (no correlation).

2. Co-evolution of Gag cleavage sites with primary PI resistance protease mutations

I wanted to further investigate substrate co-evolution and so extended this study, beyond the NFV-resistant mutations, to other primary drug resistance mutations. Anti-HIV drugs have become increasingly accessible (in the US at least) and resistance testing has become more common. As a result, we had access to a larger database of sequences.

This analysis also focused on qualitative changes in the NC-p1, p1-p6, autoproteolysis and the Pr-RT cleavage-site sequences, as compared to the HXB2 sequence, that covary with several primary resistance mutations within the protease. The following protease mutations, singly or in combination, were analyzed in this study: D30N, I50V/L, V82A, I84V, N88D and L90M.

Method. As in the previous analysis, we chose to use the Chi-square test of independence, as described in the previous section, to test for the association of cleavage site and protease mutations. Since this database consisted of more than 39,000 sequences, scripts were used to extract sets of data and the Chi-square test was computed using the software R (R function chisq.test). Since the same dataset was used to test for cleavage site mutations with several protease mutations, adjustments were made for multiple testing using the Bonferroni correction. Thus, if there are *n* number of hypotheses being tested on a single set of data, then the significance level is *1/n* times significance level if only one hypothesis is being tested. In other words, the p-values are multiplied by *n* and the corrected p-values < 0.05 indicate statistical significance. The Yates correction was not needed for this set of sequences as the dataset was very large.

Results and Discussion. The Pearson Chi-square is the simplest and one of the most commonly used tests for significance of association between two categorical variables. Covariation between cleavage site mutations and resistance mutations in the protease were observed in both studies. In the first study, I observed that D30N/N88D and p1-p6 substrate site mutations were significantly correlated and the phi values indicate a

positive correlation. Mutations at NC-p1 also covaried with the D30N/N88D mutations in the protease. However, in contrast with the p1-p6 site, I observed that mutations within the NC-p1 cleavage site were negatively correlated to the protease D30N/N88D mutations indicated by a negative phi value.

Investigate the effect of substrate co-evolution on viral fitness and phenotypic susceptibilities to various protease inhibitors.

In order to study the effect of substrate co-evolution on viral fitness and drug resistance, Replication Capacity™(RC) and phenotypic susceptibilities for viruses including the protease mutations of interest with and without Gag cleavage site mutations were extracted and analyzed. RC was determined and expressed as a percentage relative to that of a reference virus, NL4-3 (Deeks et al., 2001). Phenotypic susceptibilities are expressed as fold change (FC) in IC50 of the patient virus relative to that of the drug-sensitive reference, NL4-3.

Method. We used the Mann-Whitney test in order to examine if substrate co-evolution had an effect on PI susceptibility and RCs. The Mann-Whitney test can be used if following assumptions of the are met, as in this study: 1) the two samples being tested are independently and randomly drawn, 2) the variables are ordinal and continuous and 3) the difference in populations, if any, is with respect to their medians.

This test assesses whether a difference exists between the two samples of observation. The null hypothesis here is that the two samples are drawn from a single population and therefore, their probability distributions are equal. In this case, the null and alternate hypotheses tested are:

H_0: $M_X = M_Y$

H_A: $M_X \neq M_Y$

where M_X is the median of RCs of a set of viruses with mutations in both the particular

cleavage site and protease and M_Y is the median of a set of viruses with particular mutations in the protease alone in the absence of cleavage site mutations. This is a two-tailed test. The level of significance $\alpha = 0.05$. The expectation is that mutations in the cleavage site would affect RC or PI susceptibility when present in combination with primary resistance-associated protease mutations.

In order the compute the test statistic the two samples are combined and ranked in ascending order. If two or more observations are tied, their rank is equal to the mean of the rank positions for which they are tied. The test statistic, *U*, is calculated as follows:

$$U = S - \frac{n(n+1)}{2}$$

where *n* is the number of sample *X* observations (sequences with both the particular cleavage site mutation and protease mutation being analyzed) and *S* is the sum of ranks of sample *X*. If the computed values of *U* is either less than $\omega_{\alpha/2}$ or greater than $\omega_{1-(\alpha/2)}$ then the null hypothesis is rejected, where

$\omega_{\alpha/2}$ is the critical value of *U* (for a two-sided test as in this case), *n* is number of samples of *X* and *m*, number of samples of *Y*. However, if either n or m is greater than 20, then we calculate a new statistic *z*, as follows:

$$z = \frac{U - mn/2}{\sqrt{mn(n+m+1)/12}}$$

and compare it with critical values of the standard normal distribution to either reject or accept the null hypothesis.

The R software environment, version 2.4.1 (http://www.r-project.org) was used to perform the Mann-Whitney test to calculate the test statistic, as well as the p-values (R function "Wilcoxon.test"). Adjustments were made for multiple testing using the Bonferroni method. The corrected p-values < 0.05 were taken to be statistically significant.

Results and Discussion. In majority of cases, substrate co-evolution had no significant effect on replicative capacity. In contrast, in most instances where substrate co-evolution is seen, there was significant decrease in susceptibility to one or more protease inhibitors.

The Mann-Whitney test is used in same situations as the student's t-test. We chose the Mann-Whitney test for the following reasons: 1) Mann-Whitney is usually used when the data is ordinal, as in this case, 2) Mann-Whitney is much more robust as compared the t-test as it is less likely to give falsely significant results as a result of one or two outliers, 3) the t-test is always used when populations are normally distributed, whereas for distributions that are not normal and/or very large sample sizes, as in this case, the Mann-Whitney test is much more efficient, and 4) Mann-Whitney test can be used even if the shapes of distribution of the two samples are different whereas in the case of the t-test, the distributions are required to be the same. Though the Mann-Whitey test is commonly used to test for differences in medians, this is not strictly true and can also be used to test for differences in means of a population. The two-sample test, as used here, may be thought of as testing the null hypothesis that the probability of an observation from one population exceeding an observation from the second population is 0.5. Thus, the Mann-

Whitney test was used to test the difference in means in the case of RCs whereas in the case of phenotypic susceptibilities, it was used to test difference in median fold change in IC50. Though a formal test for determining the distribution was not performed, the Mann-Whitney test was used here, as it is the best-known non-parametric tests for significance.

Investigate whether there were significant differences in the average number of secondary protease mutations in viruses with or without particular Gag mutations

In order to rule out the possible effect of additional protease mutations we examined if there were differences in the average number of secondary mutations in viruses with particular protease mutations in the presence and absence of Gag mutations.

Method. The student's t-test was used to evaluate the difference in the average number of secondary protease mutations in viruses with and without particular Gag mutations. The student's t-test maybe used if the following assumptions are met, as in this study: 1) the two sample populations are normally distributed and, 2) the variances are equal and 3) data used should be sampled independently from the two populations being compared.

The t-test assesses whether a difference exists between the means of the two samples. The null hypothesis here is that the means of the two populations are equal. In this case, the null and alternate hypotheses tested are:

$H_0: m_x = m_y$

$H_A: m_x \neq m_y$

where m_X is the average number of secondary mutations in a set of viruses with mutations in both the particular cleavage site and protease and m_Y is the average number of secondary mutations in a set of viruses with particular mutations in the protease alone in the absence of cleavage site mutations. This is a two-tailed test. The level of significance $\alpha = 0.05$. We are testing to see if the mean number of secondary mutations is different

between the two populations which could somehow affect the median drug susceptibilities of these two populations to various protease inhibitors.

In order to examine if the means of the two populations are equal the following calculations are performed:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ where}$$

$$S_{X_1X_2} = \sqrt{\frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}}$$

$S_{X_1X_2}$ is an estimate of the common standard deviation of the two samples, where n is the number of samples, 1= group one and 2 = group 2. n –1 is the number of degrees of freedom for either group and $n_1 + n_2 - 2$ is the total number of degrees of freedom and is used in significance testing. Using the above equations we can compute the 95% confidence interval for the difference in population mean. If the interval includes zero, then we can conclude that the population means may be equal. Additionally, the test statistic can be used to calculate the p-value. p-values < 0.05 are considered significant.

The R software environment, version 2.4.1 (http://www.r-project.org) was used to perform the t-test test to calculate the test statistic, as well as the p-values (R function "t.test"). Adjustments were made for multiple testing using the Bonferroni method. The corrected p-values < 0.05 were taken to be statistically significant.

In a similar manner, we also used the t-test to investigate the effect of Gag cleavage site mutations on protease inhibitor susceptibility, in the absence of resistance mutations in the protease. Sequences with and without Gag mutations in the absence of protease mutations were extracted from the database. The difference in mean FC in IC50s to various protease inhibitors was evaluated using the t-test as described in the preceding section.

References:

1. Daniel, W.W. (2005). “Biostatistics: A foundation for analysis in the health sciences.” Eighth ed. John Wiley & Sons, Inc.
2. Miller, I., Miller, M. (1999). “ John. E. Freund’s mathematical statistics.” Sixth ed. Prentice-Hall, Inc.