THE CONTRIBUTION OF THE NATURALLY OCCURRING POLYMORPHISMS IN ALTERING THE BIOCHEMICAL AND STRUCTURAL CHARACTERISTICS OF HIV-1 SUBTYPE C PROTEASE

By
ROXANA MARIA COMAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2007
To my mother, Victoria Coman, and my father, Florin Coman, for their unwavering love and support
ACKNOWLEDGMENTS

I thank my mentor, Dr. Ben M. Dunn, for allowing me to pursue my doctorate training under his guidance. In his lab I have learned how to design a project, how to plan all the experimental steps, how to think like a scientist. I have been given the opportunity to present my work at several national and international meetings, to interact and to exchange ideas with scientists from around the world. I can wholeheartedly say that the training in his lab eased my transition from a graduate student to a postdoctoral fellow to a professor managing my own lab. I would like to express my appreciation to the members of my supervisory committee: Dr. Linda Bloom, Dr. Brian Cain, Dr. Richard Condit, and Dr. Robert McKenna for their patience and for their willingness to help and guide me through my graduate career.

I would like to extend special thanks to Dr. Maureen Goodenow for being a true adviser and a true friend who always was available to answer my questions or give me guidance.

I want to express my thanks to all my friends I have made while pursuing my doctorate, especially my colleagues, both graduate and undergraduate students in the Dunn lab and the McKennas’ lab.

I thank my parents for their never-ending love and unconditional support. They have been behind me every step of the way, encouraging me to follow my aspirations. I would also like to thank my sister, Anca Cristina, for being a friend and for taking pride in all my accomplishments. I would like to thank my grandparents, Veta and Gheorghe Albulescu for their love and support.

Most of all, I thank my husband, Altin Gjymishka, for his endless love and friendship. He has always encouraged me to pursue my dreams. He has been my inspiration to accomplish what I have accomplished today and, without him, I could not have made it through it as well as I have.
Finally, my compassion goes out to all who are living with HIV/AIDS and fight it with so much courage and dignity. I extend my sincere thanks and admiration to all involved in fighting this disease and in bringing hope to the millions afflicted by HIV/AIDS around the world.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>9</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>10</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>13</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 HUMAN IMMUNODEFICIENCY VIRUS TYPE 1</td>
<td></td>
</tr>
<tr>
<td>Introduction on World HIV Epidemic Status</td>
<td>15</td>
</tr>
<tr>
<td>HIV-1 Genome and Structure</td>
<td>15</td>
</tr>
<tr>
<td>Viral Life and Replication Cycles</td>
<td>16</td>
</tr>
<tr>
<td>Antiretroviral Therapy and Resistance</td>
<td>19</td>
</tr>
<tr>
<td>Reverse Transcriptase Inhibitors (RTIs)</td>
<td>21</td>
</tr>
<tr>
<td>Protease Inhibitors (PIs)</td>
<td>22</td>
</tr>
<tr>
<td>Entry and Fusion Inhibitors</td>
<td>23</td>
</tr>
<tr>
<td>Highly Active Antiretroviral Therapy (HAART)</td>
<td>23</td>
</tr>
<tr>
<td>HIV-1 Diversity</td>
<td>24</td>
</tr>
<tr>
<td>Groups and Subtypes</td>
<td>24</td>
</tr>
<tr>
<td>HIV-1 non-B Subtypes</td>
<td>25</td>
</tr>
<tr>
<td>HIV-1 Subtype C</td>
<td>28</td>
</tr>
<tr>
<td>2 HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 PROTEASE</td>
<td>40</td>
</tr>
<tr>
<td>HIV-1 Protease Structure</td>
<td>40</td>
</tr>
<tr>
<td>Function and Substrate Specificity</td>
<td>41</td>
</tr>
<tr>
<td>Gag/pol Processing</td>
<td>42</td>
</tr>
<tr>
<td>Matrix (MA)</td>
<td>44</td>
</tr>
<tr>
<td>Capsid (CA)</td>
<td>45</td>
</tr>
<tr>
<td>p2</td>
<td>46</td>
</tr>
<tr>
<td>Nucleocapsid (NC)</td>
<td>46</td>
</tr>
<tr>
<td>p1</td>
<td>46</td>
</tr>
<tr>
<td>p6(^{GAG})</td>
<td>47</td>
</tr>
<tr>
<td>Protease Inhibitors and Drug Resistance</td>
<td>47</td>
</tr>
<tr>
<td>Drug Resistance</td>
<td>49</td>
</tr>
<tr>
<td>Active site mutations</td>
<td>51</td>
</tr>
<tr>
<td>Non-active site mutations</td>
<td>51</td>
</tr>
<tr>
<td>Naturally Occurring Polymorphisms in HIV-1 Protease</td>
<td>52</td>
</tr>
</tbody>
</table>
3 MATERIALS AND METHODS

Sub-Cloning of HIV-1 Protease.................................63
Directional Cloning – Digestion and Ligation.........................64
Mutagenesis.............................................................65
Transformation............................................................66
Protein Expression and Inclusion Body Extraction..................66
Protein Refolding..........................................................67
Ammonium Sulfate Precipitation.........................................68
Size - Exclusion Chromatography.......................................68
Protease Kinetic Studies..................................................69
  Determination of Michaelis-Menten Constants..................69
  Dissociation Constant (K_i) Determination.........................70
  Enzyme Active Site Titration........................................70
Crystallization Studies....................................................71
  Protein Sample Preparation........................................71
  Crystal Preparation..................................................72
  Data Collection and Processing...................................72
In Vitro Gag/pol Processing...............................................73
  Amplification and Cloning of the Gag/Pol DNA Sequence into the TNT Vector ....73
  Site-Directed Mutagenesis to Introduce the Desired Mutations....74
In Vitro Transcription-Translation Experiment........................75
  Separation of the Translation Products on an SDS-PAGE........75
  Autoradiography and Densitometric Analysis....................75

4 KINETIC STUDIES ON THE CONTRIBUTION OF NATURALLY OCCURRING
POLYMORPHISMS IN ALTERING THE BIOCHEMICAL CHARACTERISTICS OF
HIV-1 SUBTYPE C PROTEASE .............................................85

Introduction........................................................................85
Results.............................................................................87
  Analysis of HIV-1 Subtype C Protease Sequence................87
  Kinetic Analysis of HIV-1 Subtype B and C Proteases ..........88
  Kinetic Analysis of HIV-1 Subtype B and C Proteases Harboring Drug-Resistant
    Mutations.................................................................89
    Kinetic analysis of the single mutants of subtype B and C proteases ....90
    Kinetic analysis of the double mutants of subtype B and C proteases ......91
    Kinetic analysis of the triple mutants of subtype B and C proteases ....92
Discussion........................................................................92

5 STRUCTURAL ANALYSIS OF HIV-1 SUBTYPE C PROTEASE ..........................113

Introduction.....................................................................113
Results..........................................................................114
  Crystallization................................................................114
    Unbound HIV-1 subtype C protease..............................115
    Drug-bound HIV-1 subtype C protease.........................115
6 ANALYSIS OF THE PROCESSING EVENTS OF HIV-1 SUBTYPE A, B, AND C GAG POLYPROTEINS ......................................................... 155

Introduction ........................................................................................................................... 155
Results ................................................................................................................................... 157
Polymorphic Sites within HIV-1 Subtype A, B, and C Gag Polyproteins ................................. 157
Cleavage Site Analysis .................................................................................................... 158
In Vitro Processing Studies of HIV-1 Subtype B, A, and C Gag Polyproteins ....................... 159
  In trans processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype B protease ................................................................. 160
  In trans processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype A protease .......................................................................................... 161
  In trans processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype C protease .......................................................................................... 162
Analysis of MA/CA Cleavage Site. ...................................................................................... 162
In Vitro Processing Studies of HIV-1 Subtype A and B Gag Polyproteins – 124 and QV Variants ................................................................. 162
  In trans processing of HIV-1 subtype A gag polyprotein – S124V and QV variants – by HIV-1 subtype B protease ................................................................. 162
  In trans processing of HIV-1 subtype B gag polyprotein – V124L, V124S and ΔQV variants – by HIV-1 subtype B protease ................................................... 163
Discussion .......................................................................................................................... 164

LIST OF REFERENCES ............................................................................................................. 185

BIOGRAPHICAL SKETCH ....................................................................................................... 208
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>FDA approved individual anti-HIV drugs</td>
<td>37</td>
</tr>
<tr>
<td>1-2</td>
<td>Summary of class and agent-specific side effects of antiretroviral agents.</td>
<td>39</td>
</tr>
<tr>
<td>4-1</td>
<td>The Michaelis-Menten constants for the HIV-1 subtype B and C proteases.</td>
<td>111</td>
</tr>
<tr>
<td>4-2</td>
<td>The $K_i$ values for HIV-1 subtype B and C proteases.</td>
<td>112</td>
</tr>
<tr>
<td>5-1</td>
<td>Data collection statistics</td>
<td>152</td>
</tr>
<tr>
<td>5-2</td>
<td>Refinement steps in SHELX for the unbound form of HIV-1 subtype C protease</td>
<td>153</td>
</tr>
<tr>
<td>5-3</td>
<td>Refinement statistics</td>
<td>154</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Adults and children estimated to be living with AIDS in 2006</td>
<td>31</td>
</tr>
<tr>
<td>1-2</td>
<td>The structure of the mature human immunodeficiency virus</td>
<td>32</td>
</tr>
<tr>
<td>1-3</td>
<td>The HIV life cycle</td>
<td>33</td>
</tr>
<tr>
<td>1-4</td>
<td>Binding events and potential sites of action for various viral-entry inhibitors</td>
<td>34</td>
</tr>
<tr>
<td>1-5</td>
<td>HIV life cycle and drug targets</td>
<td>35</td>
</tr>
<tr>
<td>1-6</td>
<td>Genetic epidemiology of HIV</td>
<td>36</td>
</tr>
<tr>
<td>2-1</td>
<td>HIV-1 protease</td>
<td>56</td>
</tr>
<tr>
<td>2-2</td>
<td>Cryo-electron micrographs and schematic representations of HIV-1 particles</td>
<td>57</td>
</tr>
<tr>
<td>2-3</td>
<td>Hydrogen bonding between HIV-1 protease and a modeled substrate</td>
<td>58</td>
</tr>
<tr>
<td>2-4</td>
<td>HIV-1 gag and gag/pol polyproteins</td>
<td>59</td>
</tr>
<tr>
<td>2-5</td>
<td>Structure of HIV-1 capsid monomer</td>
<td>60</td>
</tr>
<tr>
<td>2-6</td>
<td>Structures of nine individual protease inhibitors, approved for clinical use in HIV-infected patients.</td>
<td>61</td>
</tr>
<tr>
<td>2-7</td>
<td>HIV-1 protease inhibitors resistance chart</td>
<td>62</td>
</tr>
<tr>
<td>3-1</td>
<td>The expression vector pET23a</td>
<td>77</td>
</tr>
<tr>
<td>3-2</td>
<td>Cloning the protease gene into the expression vector pET23a</td>
<td>78</td>
</tr>
<tr>
<td>3-3</td>
<td>HIV-1 protease expression in BL21 DE3(Star) pLysS cells and inclusion bodies (IBs) extraction</td>
<td>79</td>
</tr>
<tr>
<td>3-4</td>
<td>HIV-1 protease purification through size-exclusion column Superdex 75</td>
<td>80</td>
</tr>
<tr>
<td>3-5</td>
<td>Kinetic constants determination</td>
<td>81</td>
</tr>
<tr>
<td>3-6</td>
<td>Inhibitor dissociation constant (K_i) determination</td>
<td>82</td>
</tr>
<tr>
<td>3-7</td>
<td>Enzyme titration curve</td>
<td>83</td>
</tr>
<tr>
<td>3-8</td>
<td>Cloning of gag/pol sequence into the TNT vector</td>
<td>84</td>
</tr>
<tr>
<td>4-1</td>
<td>Polymorphic sites within the subtype B and C protease sequences</td>
<td>105</td>
</tr>
</tbody>
</table>
6-3 Alignment of the cleavage sites within gag polyproteins of HIV-1 subtypes, B, C, and A................................................................................................................................177
6-4 The in vitro processing without the frameshift and without the addition in trans of the HIV protease .................................................................................................................................................178
6-5 In trans processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype B protease ..............................................................................................................................................179
6-6 In trans processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype A protease ......................................................................................................................................................180
6-7 In trans processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype C protease ......................................................................................................................................................181
6-8 In trans processing of HIV-1 subtype A gag polyprotein – S124V and QV variants – by HIV-1 subtype B protease ......................................................................................................................................................182
6-9 In trans processing of HIV-1 subtype B gag polyprotein – V124S and ΔQV variants – by HIV-1 subtype B protease ......................................................................................................................................................183
6-10 HIV-1 matrix protein ..................................................................................................................................................................................................................................................184
THE CONTRIBUTION OF THE NATURALLY OCCURRING POLYMORPHISMS IN ALTERING THE BIOCHEMICAL AND STRUCTURAL CHARACTERISTICS OF HIV-1 SUBTYPE C PROTEASE

By

Roxana Maria Coman

August, 2007

Chair: Ben M. Dunn
Major: Medical Sciences–Biochemistry and Molecular Biology

The discovery of a very diverse panel of HIV-1 subtypes poses the question of whether the currently used therapeutic regimens designed against HIV-1 subtype B are as effective against non-B subtypes, including subtype C viruses.

In our research, we focus on exploring the role of naturally occurring polymorphisms in subtype C protease. Our goal is to understand how these sequence variations affect the overall characteristics of subtype C protease and ultimately how the response to treatment is modulated by the preexistence of the baseline polymorphisms.

We obtained a subtype C protease clone through the NIH-AIDS Reagent and Reference Program and we determined its kinetic parameters. We then added several combinations of major drug resistance mutations: D30N, N88D, and L90M. Our data show that the naturally occurring polymorphisms by themselves, do not provide for a greater level of resistance against the clinically used protease inhibitors. However, we found that the baseline polymorphisms appear to have a role in preserving the catalytic efficiency of the protease upon acquisition of major drug resistance mutations.
We were also able to obtain crystals of subtype C protease and to perform structural studies of this enzyme in the unbound form and complexed with two protease inhibitors: indinavir and nelfinavir. In our crystal structures we observed that the protease inhibitors bind in the same orientation and make the same contacts within the active site as in subtype B protease. The results also showed that the naturally occurring polymorphisms within subtype C protease might have a role in modulating the structural stability and flexibility of this enzyme.

We also conducted in vitro gag processing studies, in which we compared the rates and order of processing among subtypes B, C, and A. The experiments were performed by expressing the gag polyprotein in a cell-free system and adding the active HIV protease in trans. Our data reveal that the gag amino acid sequence has a dominant effect on the order and the rate of cleavage. We also observed that the rate of gag cleavage appears to be modulated by both the cleavage site amino acid sequence and other determinants within gag polyprotein.

It is widely known that clinically used protease inhibitors have been designed using kinetic and structural information about HIV protease. The compilation and subsequent analysis of our data provide information about the biochemical and structural characteristics of subtype C protease, and its susceptibility to currently used protease inhibitors. This type of knowledge will hopefully aid in finding new drugs to combat HIV infection worldwide.
CHAPTER 1
HUMAN IMMUNODEFICIENCY VIRUS TYPE 1

Introduction on World HIV Epidemic Status

Acquired immunodeficiency syndrome (AIDS) continues to spread largely unchecked since its first documentation in 1983 (Gottlieb et al. 1983; Groopman and Gottlieb 1983; Karpas 2004). Even though promising developments have been seen in the last years in the global effort to control the AIDS epidemic, the number of people living with HIV continues to grow as does the number of deaths due to AIDS. A total of 39.5 million people were living with HIV in 2006, but this number can go as high as 47.1 million, according to the figures released by the Joint United Nations Program on HIV/AIDS (UNAIDS) and World Health Organization (WHO) (UNAIDS/WHO 2006) (Figure 1-1). One of the developing regions most devastated is Sub-Saharan Africa. Two thirds (63%) of all adults and children with HIV globally live in sub-Saharan Africa and almost three quarters (72%) of all adult and child deaths due to AIDS in 2006 occurred there. Overall, sub-Saharan Africa is home to an estimated 25 million adults and children infected with HIV.

The most striking increases in the number of HIV-positive people have occurred in East Asia, Eastern Europe and Central Asia, where the number of people living with AIDS in 2006 was 21% higher than in 2004.

Another important shift in the trend of 2006 AIDS epidemic is that, globally, and in every region, more adult women (15 years or older) than ever before are living now with HIV (UNAIDS/WHO 2006).

HIV-1 Genome and Structure

HIV is part of a family or group of viruses called lentiviruses and it is now generally accepted that HIV is a descendant of simian (monkey) immunodeficiency virus (SIV) (Hirsch et
A further classification categorizes HIV as a retrovirus that packages two copies of positive-sense RNA strands in its genome. HIV has a diameter of 100-120 nm, its genome is about 9 Kb in length and is flanked by two long terminal repeats (LTR) that are involved in integration and regulation of the viral genome. The genome can be read in three frames and there are several overlaps of viral genes in different reading frames, allowing for the encoding of many proteins in a small genome. The viral genes encode for structural (gag and env), enzymatic (pol), accessory (vif, vpr, vpu, nef) and regulatory (rev, tat) proteins (Saksena 1998).

The outer shell of the virus, known as the viral envelope, consists of a lipid bilayer that is acquired as the virus buds from the cell surface. Embedded in the viral envelope is a complex protein known as env, which consists of an outer protruding cap that is a glycoprotein named gp120 and a stem glycoprotein called gp41. Inside the viral envelope is a protein called p17 (matrix) and within this is the viral core or capsid, which is made of another viral protein p24 (core antigen). The major elements contained within the viral core, besides the viral RNA, are protein p7 (nucleocapsid) that associates with the RNA molecules, and three enzyme proteins, p51 (reverse transcriptase), p11 (protease) and p32 (integrase). Some other regulatory proteins, such as nef, vpr and vif, are also packaged in the virion (Figure 1-2).

**Viral Life and Replication Cycles**

The HIV-1 life cycle is a complex multistage process involving interactions between HIV-1 proteins and host macromolecules. The Early Phase of the virus life cycle comprises the infection of host cell and integration of viral genome, and the Late Phase comprises the regulation of the expression of the viral gene products and the production of viral particles (Turner and Summers 1999). Infection begins when an HIV particle encounters a host cell with a surface receptor called CD4 (Figure 1-3). The cells mainly targeted by HIV are T-Helper lymphocytes, macrophages, and dendritic cells. The virus particle uses gp120 to attach itself to
the CD4 receptor and this is sufficient for binding, but co-receptors are necessary for viral entry (Dragic et al. 1996; Alkhatib et al. 1997; Deng et al. 1997). Several proteins have been identified as possible co-receptors but HIV, generally, uses mainly two co-receptors to enter a target cell: either CCR5 or CXCR4, depending on the strain of virus (Doranz et al. 1996; Samson et al. 1996; Clapham and Weiss 1997; Moore 1997). The strains of HIV most commonly seen early in HIV disease, known as macrophage-tropic (M-tropic) viruses, use CD4 and CCR5 for cell entry. The importance of CCR5 has been demonstrated by the identification of polymorphisms within the CCR5 gene which affect transmission and/or disease progression (Samson et al. 1996).

The binding of viral gp120 env protein to the CD4 receptor causes a structural change in gp120 exposing the binding site for the co-receptors. Once the co-receptor is bound, further structural rearrangements occur, mostly in gp41 transmembrane protein, which lead to virus entry (Figure 1-4). Once within the cell, the virus particle releases its RNA, and the enzyme reverse transcriptase (RT) then makes a DNA copy of the viral RNA. Of importance in the process of retroviral DNA synthesis, for the purpose of developing adequate therapies, are the high degree of genome recombination and the high error rate of RT which provides means for a high level of viral genome variability. Recombination is facilitated by the packaging into the virus particle of two copies of viral RNA and the ability of RT to jump from one copy of the genome to another. The high error rate is due to lack of proofreading ability of RT. It is estimated that HIV-1 RT has an average error rate of $3 \times 10^{-5}$ errors per base per replication cycle (Mansky and Temin 1995). This value is an average of all types of point mutations, but deletions, insertions, and frameshift mutations are also commonly observed.

The template viral genomic RNA is degraded by RNase H and the new HIV dsDNA then moves into the nucleus of the cell where, with the help of the enzyme integrase (IN), it is inserted
into the host cell’s DNA. This is the last step of the early phase of the HIV life cycle. Formation of the pre-integration complex must occur before the integration can take place. The pre-integration complex carries sequences that interact with the cellular system and signal the transport of the viral proteins and nucleic acids into the nucleus. Proteins such as matrix, vpr and IN have been postulated to be involved in the integration of the viral DNA into the host genomic DNA. This process involves not only the viral IN enzyme but also the host repair system. Initially it was believed that the site of integration was random. However, there are several reports that sites of HIV integration in the human genome are not randomly distributed but instead are enriched in active genes and regional hotspots (Schroder et al. 2002). Global analysis of cellular transcription indicated that active genes were preferential integration targets, particularly genes that were activated in cells after infection by HIV-1. Regional hotspots for integration were also found, including a 2.4 kb region containing 1% of sites. These data document unexpectedly strong biases in integration site selection and suggest how selective targeting promotes aggressive HIV replication (Schroder et al. 2002).

Once located in the genome of the cell, HIV DNA is called a provirus. The HIV provirus is replicated by transcription into viral RNA, some of which become new viral genomic material and some of which is needed to direct the synthesis of viral polyproteins env, gag and gag/pol. After transcription, the viral mRNA, as any other cellular RNA, is modified by addition of a polyA tail. Also, in order to produce proteins such as Rev and Tat, the mRNA must be properly spliced. However, genomic viral RNA has to be transported out of the nucleus without further splicing. To circumvent the strong aversion of the cellular machinery for transporting “improperly” spliced RNA molecules, the retroviruses code for a constitutive transport element. This nucleotide sequence allows for the transport of the un-spliced mRNA, and in the case of
HIV, this is aided by the Rev protein (Bray et al. 1994). This creates a point of regulation where regulatory genes that are coded as nascent sequences and arise due to RNA splicing are transcribed and exported out of the nucleus for translation first. One of these genes is rev that encodes the regulatory protein Rev. Once Rev reaches high enough concentrations it promotes the export of the intact mRNA for translation of gag, pol, and env genes (Pollard and Malim 1998).

The final step in virus replication cycle is budding and maturation. The association of viral RNA, gag and gag/pol polyproteins just underneath the cell membrane precedes budding. Aggregation of gag and gag/pol polyproteins is mediated by both protein-protein interactions and protein-RNA interactions. Specifically, gag-gag interaction is mediated by nucleocapsid, matrix, and capsid interactions (Borsetti et al. 1998; Burniston et al. 1999; Gross et al. 2000; Ehrlich et al. 2001). The viral genomic RNA is recruited through interactions with the nucleocapsid protein within gag.

The accumulation of env, gag and gag/pol proteins within the lipid rafts in the cell membrane induces changes that promote budding (Scarlata et al. 1998; Murakami and Freed 2000; Nguyen and Hildreth 2000; Liao et al. 2001).

**Antiretroviral Therapy and Resistance**

Access to treatment and care has greatly increased in the recent years. Even if the coverage is still low in many HIV/AIDS-stricken regions, the benefits are dramatic. Through the expanded provision of antiretroviral (ARV) treatment, an estimated two million life-years were gained from 2002 in low- and middle- income countries (UNAIDS/WHO 2006).

The evolution of ARV therapy is an ongoing process aimed at discovering potent and tolerable drug regimens. The first ARV drug approved by the US-FDA in 1987 was zidovudine (AZT), a previously known potential anticancer agent. Shortly after, rapid progress in the
understanding of the structure and life cycle of the virus led to unprecedented development of other drugs targeted to a variety of viral proteins. The retroviral enzymes – RT, IN, and protease (PR) – were the obvious targets for drug discovery (Figure 1-5). The first drugs to be identified were inhibitors of RT (DeVita et al. 1987), which were discovered and developed long before the structure of RT itself was solved (Jacobo-Molina et al. 1993). Newer RT-targeted drugs, nonnucleoside inhibitors (NNRTI), and HIV protease inhibitors (PIs) have been developed bearing the enzyme structure in mind. The structure-assisted drug design and discovery process utilizes structural biochemical methods, such as protein crystallography, NMR, and computational biochemistry, to guide the synthesis of potential drugs. This information can, in turn, be used to help explain the basis of their activity and to improve the potency and specificity of new lead compounds. Crystallography plays a particularly important role in this process. The past years have seen a virtual explosion of crystallographic studies aimed at the characterization of the structures of HIV enzymes and of HIV enzyme/inhibitor complexes on an atomic level.

Initially, the ARV drugs were given as mono- or dual-therapy. However, this approach often led to treatment failures due to the development of resistant virus. Patients who were receiving monotherapy or dual therapy initially experienced decreases in viral load, increases in CD4 count, and improvement in quality of life; then they experienced viral rebound and decrease CD4 counts further into therapy. The concept of highly active antiretroviral therapy, HAART, was introduced once the PIs were developed. The first PI approved in 1995 by the FDA to be administered to HIV-positive patients was saquinavir (SQV). To date, twenty-one individual ARV compounds within four classes have been approved for the treatment of HIV-1 infection: seven nucleoside (NRTIs) and one nucleotide analog reverse transcriptase inhibitors (NtRTI),
three nonnucleoside RT inhibitors (NNRTIs), nine protease inhibitors (PIs), and one fusion inhibitor (Table 1-1).

**Reverse Transcriptase Inhibitors (RTIs)**

The mode of action of NRTIs and NtRTIs is essentially the same; they are analogues of the naturally occurring deoxynucleotides needed to synthesize the viral DNA and they compete with the natural deoxynucleotides for incorporation into the growing viral DNA chain. The difference between these two subclasses of drugs is that NRTIs, in order to be incorporated into the viral DNA, must be activated in the cell by the addition of three phosphate groups to their deoxyribose moiety. This phosphorylation step, carried out by cellular kinase enzymes, converts NRTIs into NtRTIs. Taking NtRTIs directly allows conversion steps to be skipped, causing less toxicity. Unlike the natural deoxynucleotides substrates, NRTIs and NtRTIs lack a 3’-hydroxyl group on the deoxyribose moiety. As a result, following incorporation of an NRTI or an NtRTI, the next incoming deoxynucleotide cannot form the next 5’-3’ phosphodiester bond needed to extend the DNA chain. Thus, when an NRTI or NtRTI is incorporated, viral DNA synthesis is halted, a process known as chain termination. All NRTIs and NtRTIs are classified as competitive substrate inhibitors.

In contrast, NNRTIs have a completely different mode of action. NNRTIs block RT by binding at a different site on the enzyme when compared to NRTIs and NtRTIs. NNRTIs are not incorporated into the viral DNA but instead inhibit the movement of protein domains of RT that are needed to carry out the process of DNA synthesis. NNRTIs are therefore classified as non-competitive inhibitors of RT.

The RTIs are also available as combination pills with the advantage of a reduced burden pill: zidovudine/lamivudine (Combivir), zidovudine/lamivudine/abacavir (Trizivir),
lamivudine/abacavir (Epzicom), tenofovir/emtricitabine (Truvada), and efavirenz/tenofovir/emtricitabine (Atripla).

The most common side effects of RTIs as a class are digestive problems (diarrhea, nausea, vomiting, abdominal pain) and constitutional problems (fatigue, fever) (Table 1-2).

**Protease Inhibitors (PIs)**

PIs work by inhibiting the proteolytic cleavage of the structural and enzymatic proteins, which prevents the virus from maturing into an infectious virion. The addition of PIs to the ARV therapeutic regimens significantly improved the life expectancy of HIV-positive patients.

Unfortunately, the older PIs came with multiple scheduling requirements and the need to take 10-16 capsules/day. Ritonavir (RTV) boosting is a relatively new concept, and today is one of the mainstays of therapy. Boosting reduces the frequency of dosing and the number of required forms/day (Kilby et al. 2000; Veldkamp et al. 2001). The concept of boosting involves pharmacokinetic drug interactions between a PI, RTV as the boosting PI, and inhibition of the cytochrome P450 (CYP) 3A4 enzyme. Pharmacokinetic enhancement or “boosting” of PI serum levels and prolongation of their half-life can be achieved with subtherapeutic doses of RTV. Metabolized mostly through the 2C19 enzyme of the cytochrome P450, nelfinavir (NFV) is the only PI not markedly boosted by RTV (Kurowski et al. 2002).

The adverse effects of PIs are listed in Table 1-2. Important to mention is that atazanavir (ATV), one of the newest and most potent PIs, is more tolerable than other PIs. However, it has some notable drug interactions: (1) its levels are decreased by 25% when combined with tenofovir, (2) it has significant interactions with H2 blockers and proton pump inhibitors (Le Tiec et al. 2005).
Entry and Fusion Inhibitors

This new class of ARV targets prevention of the fusion of HIV and CD4 cell, and prevents passing the viral genome to the CD4 cell. The only FDA approved fusion inhibitor, Enfuvirtide, is a structural analog of HR2 domain of gp41 and binds to the HR1 region preventing the change in conformation that allows the viral entry into the cell. Enfuvirtide is a synthetic peptide recommended for use in patients who are treatment experienced or who have had multiple treatment failure regimens and in those with ongoing viral replication or viral load greater than 400 copies/mL (Piacenti 2006).

Highly Active Antiretroviral Therapy (HAART)

HAART is a combination of three or more drugs from two drug classes and has improved significantly the prognosis of HIV-infected individuals (Collier 1996; Carpenter et al. 1997). Since the advent of HAART, the following improvements have been noted: better patient quality of life, increased survival, slowed disease progression, decreased opportunistic infections, decreased viral loads, and increased CD4 counts. Many studies have shown that HAART is the most powerful and essential combination to treat HIV infection effectively and decrease the likelihood of emergence of resistant viruses. However, even with this complex therapeutic approach, the development of drug resistance poses the greatest challenge for treating HIV infection and the margin of success for achieving and maintaining virus suppression is narrow. The evolution of HIV-1 drug resistance within an individual depends on three main variables: different immunologic status and response to therapy, HIV-1 genetic variability that allows for rapid development of drug resistance mutations, and the adherence to therapy.

Adherence to a complex regimen is often a significant barrier to treatment success. It is known that HAART is expensive, requires the patient to ingest a large number of pills under a
complex dosing schedule and specific food requirements. It is also associated with various and severe side-effects (Schieferstein and Buhk 2005).

The occurrence of the drug-induced resistance mutations results from the inability of HAART to totally eradicate viral replication. With the extremely fast replication rate of HIV, which can reach $10^{10}$ viral particles/day, the high error rate of RT that incorporates, on average, one error per ten thousand bases, and the capacity for genomic recombination, the virus can quickly develop genomic variations that translate into protein structural changes. These mutations tend to decrease the affinity of the drug for the target enzyme promoting therapeutic failure (Erickson and Burt 1996). As of May 2006, the HIV guidelines recommend resistance testing in patients with acute or chronic HIV infection before therapy is started (U.S. Department of Health and Human Services 2006). This may help detect a virus that is resistant to initial treatment regimens, and regimens can be adjusted based on sensitivity to ARV agents.

**HIV-1 Diversity**

**Groups and Subtypes**

The large genomic diversity of viral subtypes in different geographical regions is the consequence of the high mismatch error rate of the HIV RT enzyme (between $1 \times 10^{-4}$ and $5 \times 10^{-5}$) (Mansky and Temin 1995; Wainberg et al. 1996) coupled with the absence of an exonuclease proofreading activity. Other factors that contribute to the rapid pace of genetic diversification include the replicative rate of each viral subtype, the number of mutations arising in each replicative cycle, the viral propensity for genomic recombination, and viral fitness. In addition, high rates of genomic evolution may result from host, environment, and/or therapeutic selection pressure (Simon et al. 1998).

HIV is characterized by significant genetic diversity among distinct types, groups and subtypes, (Brodine et al. 1995; Fleury et al. 2003; Kantor and Katzenstein 2004) and this
variability has implications in prevention, diagnostic tests, therapy response, and vaccine development (Peeters and Sharp 2000b; Romano et al. 2000; Holguin et al. 2002; Kantor and Katzenstein 2004).

This variability of HIV has led to the development of various distinct groups and subtypes of HIV virus (Figure 1-6) (Peeters and Sharp 2000a; Peeters 2001). HIV-1 group M, which accounts for most infections worldwide (Osmanov et al. 2002), is classified into 9 distinct subtypes (A-D, F-H, J, and K) and many circulating and unique recombinant forms. These subtypes and intersubtype recombinants differ from one another by 25-30% in the env gene (Robertson et al. 2000; Gaschen et al. 2002) and by 10-15% in the pol gene (Shafer et al. 1999; Robertson et al. 2000).

**HIV-1 non-B Subtypes**

While HIV-1 subtype B has been the most widely studied, subtypes A, C, and D predominate worldwide. Among the approximately 40 million people living with HIV/AIDS in 2003, more than 80% are infected with HIV-1 non-B subtypes (Quinn 2002; Kantor and Katzenstein 2003; Kantor and Katzenstein 2004).

In the absence of any drug exposure, RT and PR sequences from B and non-B HIV-1 are polymorphic among about 40% of the first 240 RT amino acids and 30% of the 99 PR amino acids. Based on the observation of differences between sequences from untreated and treated persons, mutations at PR positions 10, 20, 36, 63, 71, 77, 93 are characterized as secondary resistance mutations in subtype B PR (Hirsch et al. 2000; Shafer et al. 2000). These mutations are known to contribute to drug resistance when present together with certain primary protease mutations, or have been shown to compensate for the decrease in catalytic efficiency caused by primary protease mutations selected by protease inhibitors. Substitutions at many of the above-mentioned positions occur at high rates in certain non-subtype B viruses, and are designated as
naturally occurring or baseline polymorphisms. Whereas in treatment-naïve patients, many of these baseline polymorphisms do not confer resistance to drugs *per se* among different subtypes, they may facilitate the development of drug resistance (Spira et al. 2003). Recent studies have shown that non-B isolates were statistically associated with a more rapid progression to resistance after antiretroviral therapy and they had different mutational patterns when compared to those of B isolates (Pillay et al. 2000; Loemba et al. 2002). The combined effects of naturally existing polymorphisms and drug resistant mutations might have important consequences on the feasibility of continuing to use current HIV-1 PIs for non-subtype B infections. A study presented in 2006 found that Ugandans infected with subtype D or recombinant strains incorporating subtype D developed AIDS sooner than those infected with subtype A, and also died sooner. The study's authors suggested that subtype D is more virulent because it is more effective at binding to immune cells (Laeyendecker et al. 2006). This result was supported by another study presented in 2007, which found that Kenyan women infected with subtype D had more than twice the risk of death over six years compared with those infected with subtype A (Baeten 2007). An earlier study of sex workers in Senegal, published in 1999, found that women infected with subtype C, D or G were more likely to develop AIDS within five years of infection than those infected with subtype A (Kanki et al. 1999).

It has also been hypothesized that the HIV genetic variability might modulate the transmissibility rates and modes of different viral subtypes or CRFs. It has been observed that certain subtypes/CRFs are predominantly associated with specific modes of transmission. In particular, subtype B is spread mostly by homosexual contact and intravenous drug use (essentially via blood), while subtype C and CRF A/E tend to fuel heterosexual epidemics (via a mucosal route) (Bhoopat et al. 2001). However, whether there are biological causes for the
observed differences in transmission routes remain the subject of debate. Some scientists, such as Dr. Max Essex of Harvard, believe such causes do exist. However, this theory has not been conclusively proven (Pope et al. 1997). More recent studies have looked for variation within subtypes in rates of mother-to-child transmission. One of these found that such transmission is more common with subtype D than subtype A (Yang et al. 2003). Another reached the opposite conclusion (A worse than D), and also found that subtype C was more often transmitted than subtype D (Blackard et al. 2001). A third study concluded that subtype C is more transmissible than either D or A (Renjifo et al. 2004). Other researchers have found no association between subtype and rates of mother-to-child transmission (Murray et al. 2000; Tapia et al. 2003).

The main concern, however, remains how these various subtypes will respond to the current therapeutic strategies. Most current HIV-1 ARV drug regimens were designed for use against subtype B, and so hypothetically might not be equally effective in Africa or Asia where other strains are more common. At present, there is no compelling evidence that subtypes differ in their sensitivity to ARV drugs. However, some subtypes may occasionally be more likely to develop resistance to certain drugs. In some situations, the types of mutations associated with resistance may vary and the drug resistance threshold might be lower due to preexistent polymorphisms that could act as secondary resistant mutations, increasing the drug resistance while conserving the catalytic activity and fitness of the protease. Up to date there are no long term studies analysing the speed of acquiring drug resistance in HIV-1 non-B subtypes versus B subtype. This is an important subject for future research.

The effectiveness of HIV-1 treatment is monitored using viral load tests. It has been demonstrated that some such tests are sensitive only to subtype B and can produce a significant underestimate of viral load if used to process other strains. The latest tests do claim to produce
accurate results for most Group M subtypes, though not necessarily for Group O. It is important
that health workers and patients are aware of the subtype/CRF they are testing for and of the
limitations of the test they are applying.

Important to mention is that HIV-2, even if it accounts for a lower percentage of HIV-
infected individuals, has several specific traits. For example, not all of the drugs used to treat
HIV-1 infection are as effective against HIV-2. In particular, HIV-2 has a natural resistance to
NNRTI antiretroviral drugs and they are therefore not recommended to treat individuals
harboring this HIV type. As yet there is no FDA-licensed viral load test for HIV-2 and those
designed for HIV-1 are not reliable for monitoring the other type. Instead, response to treatment
may be monitored by following CD4+ T-cell counts and indicators of immune system
deterioration. More research and clinical experience is needed to determine the most effective
treatment for HIV-2.

**HIV-1 Subtype C**

HIV-1 subtype C is one of the nine HIV-1 subtypes but noteworthy are the facts that
subtype C has the highest global prevalence, currently estimated at 42% of all HIV infections
worldwide, and is now the dominant subtype in sub-Saharan Africa (McCormack et al. 2002;
Osmanov et al. 2002). Recent studies (Walker et al. 2005) demonstrated that the growth rate of
subtype C infections in southern Africa shows no evidence of declining. In addition, the subtype
C epidemic has also spread to South and Central China, India and Brazil (Yu et al. 1998; Soares
et al. 2003).

The sequence of HIV-1 subtype C viruses shows the same genetic organization as the other
HIV-1 subtypes. Characteristics that are shared only by subtype C viruses include three copies of
the NF-kB binding site in the LTR promoter enhancer region, an insertion in the vpu
transmembrane domain as well as a truncation in the rev gene of the virus. From a biological
point of view, subtype C viruses differ from other subtypes in the choice of receptor used to enter the host. HIV-1 subtype C viruses have preferential use of CCR5 over CXCR4 (Zhang et al. 2002). Centlivre et al. demonstrated that at the peak of primary infection, preferential replication of HIV-1 subtype C was supported by the gut-associated lymphoid tissue (GALT), an IL-7 rich microenvironment. This was shown by the correlation of the RNA viral genotype in blood and stools, compartments directly draining virions from the GALT (Centlivre et al. 2006). These data show that the GALT cytokine network may well favor HIV-1 subtypes C replication during primary infection, and this could result in enhanced transmission.

Another difference between B and C subtypes is that the main transmission route for subtype C viruses appears to be heterosexual contact (Buve et al. 2001), unlike subtype B that was primarily transmitted through homosexual sex and injecting drug use (UNAIDS/WHO 2006) but this could merely be due to the specific sexual practices of the different population groups.

The documentation of these dissimilarities advances the hypothesis that differences in response to antiretroviral therapy that was developed based on studies made with HIV-1 subtype B might occur as well. Initial reports suggest that HIV-1 subtype C viruses respond equally well to therapy (Shafer et al. 1999). However, it was found that some mutations (e.g. D30N against NFV) appear in very low rates in drug treated subtype C isolates whereas the same mutations appear more often in subtype B (Grossman et al. 2004). Another example of differences is the RT V106M resistance mutation in HIV-1 subtype C patients treated with efavirenz as compared to a V106A mutation that occurs in HIV-1 subtype B patients treated with nevirapine (Brenner et al. 2003). With the possibility of complicated mutation patterns arising in ARV treated isolates,
the impact of such and other signature mutations in HIV-1 subtype C needs to be investigated further to be able to design any salvage therapy as required.
Figure 1-1. Adults and children estimated to be living with AIDS in 2006. Total: 39.5 (34.1 – 47.1) million. Accessed and adapted on June 2007 from www.unaids.org - Joint United Nations Programme on HIV/AIDS.
Figure 1-2. The structure of the mature human immunodeficiency virus. Accessed and adapted on June 2007 from www.avert.org – AVERTing HIV and AIDS.
Figure 1-3. The HIV life cycle. (a) HIV (yellow) attaches to two cell-surface receptors (the CD4 antigen and a specific chemokine receptor). (b) The virus and cell membrane fuse, and the virion core enters the cell. (c) The viral RNA and core proteins are released from the virion core. (d) The viral RNA genome is converted into double-stranded DNA through an enzyme unique to viruses, reverse transcriptase (red dot). (e) The double-stranded viral DNA moves into the cell nucleus. (f) Using a unique viral enzyme called integrase, the viral DNA is integrated into the cellular DNA. (g) Viral RNA is synthesized by the cellular enzyme RNA polymerase II using integrated viral DNA as a template. Two types of RNA transcripts, shorter spliced RNA (h) and full-length genomic RNA (j) are produced. (h) Shorter spliced RNAs are transported to the cytoplasm and used for the production of several viral proteins that are then modified in the Golgi apparatus of the cell (i). (j) Full-length genomic RNAs are transported to the cytoplasm (k). (l) New virion is assembled and then buds off. (m) Mature virus is released. Accessed and adapted on June 2007 from www.hhmi.com - HHMI, Immunology and Infectious Diseases.
Figure 1-4. Binding events and potential sites of action for various viral-entry inhibitors. The viral-entry process consists of a series of coordinated interactions — binding to two different receptors (Panel A) and membrane fusion (Panel B). The viral envelope glycoproteins are synthesized as a single polyprotein that assembles into a trimer and then is broken down by host protease into surface glycoprotein subunits (gp120) and transmembrane glycoprotein subunits (gp41). Each gp120 monomer is a complex, folded structure, consisting of a series of variable loops formed by disulfide bonds, with noncontiguous segments brought together to form three-dimensional binding sites for the CD4 receptor and a chemokine receptor (either CCR5 or CXCR4). After CD4 binding, each gp120 undergoes a conformational change exposing the region that will bind to a seven-transmembrane chemokine receptor. Viral isolates have varying affinities for CCR5 or CXCR4 receptors. Gp41 contains a repeat domain which folds upon itself, thus bringing the two membranes in close proximity resulting in fusion. Reprinted and adapted from Kilby and Eron 2003.
NRTI, NtRTI, NNRTI interrupt transcription of viral RNA into viral DNA

Entry and Fusion Inhibitors prevent passing the viral genome

PIs inhibit HIV protease-mediate cleavage of the structural and enzymatic proteins

Figure 1-5. HIV life cycle and drug targets. Accessed and adapted on June 2007 from www.roche.com – Roche.
Figure 1-6. Genetic epidemiology of HIV. Classification of HIV in types, groups, subtypes, sub-subtypes. HIV-1 recombinants are categorized in two classes: circulating recombinant forms (CRFs) and unique recombinant forms (URFs). Reprinted and adapted from Takeb et al., 2004 (Takeb et al. 2004).
Table 1-1. FDA approved individual anti-HIV drugs.

<table>
<thead>
<tr>
<th>Generic name</th>
<th>Alternative name(s)</th>
<th>Brand name(s)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse Transcriptase Inhibitors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleoside Analog Reverse Transcriptase Inhibitors</td>
<td>Zidovudine</td>
<td>AZT, ZDV, azidothymidine</td>
<td>Retrovir</td>
</tr>
<tr>
<td>Didanosine</td>
<td>ddI</td>
<td>Videx</td>
<td>Second FDA-approved ARV drug Due to lower potency and serious side effects, it is now rarely used for the treatment of HIV</td>
</tr>
<tr>
<td>Zalcitabine</td>
<td>ddC, dideoxycytidine</td>
<td>Hivid</td>
<td></td>
</tr>
<tr>
<td>Stavudine</td>
<td>d4T</td>
<td>Zerit</td>
<td>It is also used for the treatment of HIV</td>
</tr>
<tr>
<td>Lamivudine</td>
<td>3TC</td>
<td>Epivir</td>
<td></td>
</tr>
<tr>
<td>Abacavir</td>
<td>ABC</td>
<td>Ziagen</td>
<td>The most powerful NRTI to treat HIV</td>
</tr>
<tr>
<td>Emtricitabine</td>
<td>FTC</td>
<td>Emtriva</td>
<td>It is the newest NRTI and is very similar to 3TC</td>
</tr>
<tr>
<td>Nucleotide Analog Reverse Transcriptase Inhibitors</td>
<td>Tenofovir</td>
<td>tenofovir disoproxil fumarate, PMPA, TDF</td>
<td>Viread</td>
</tr>
<tr>
<td>Non-nucleoside Reverse Transcriptase Inhibitors</td>
<td>Nevirapine</td>
<td>NVP</td>
<td>Viramune</td>
</tr>
<tr>
<td>Delavirdine</td>
<td>DVD</td>
<td>Rescriptor</td>
<td>It is now rarely used to its lower potency and complex drug interactions</td>
</tr>
<tr>
<td>Efavirenz</td>
<td>EFV</td>
<td>Sustiva/Stocrin</td>
<td>It is always given in combination with other ARV drugs</td>
</tr>
<tr>
<td>Generic name</td>
<td>Alternative name(s)</td>
<td>Brand name(s)</td>
<td>Comments</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Protease Inhibitors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saquinavir</td>
<td>SQV</td>
<td>Invirase</td>
<td>The first PI approved by the FDA</td>
</tr>
<tr>
<td>Ritonavir</td>
<td>RTV</td>
<td>Norvir</td>
<td>It is widely used as a booster for other PIs</td>
</tr>
<tr>
<td>Indinavir</td>
<td>IDV</td>
<td>Crixivan</td>
<td>It requires very precise dosing schedule</td>
</tr>
<tr>
<td>Nelfinavir</td>
<td>Nelfinavir mesylate, AG1343, NFV</td>
<td>Viracept</td>
<td>The only PI approved to treat HIV in pregnant women</td>
</tr>
<tr>
<td>Amprenavir</td>
<td>APV</td>
<td></td>
<td>Marketed as the prodrug fosamprenavir (Lexiva),</td>
</tr>
<tr>
<td>Lopinavir</td>
<td>ABT-378, LPV</td>
<td></td>
<td>Due to its insufficient bioavailability, it is marketed only as a co-formulation with ritonavir (Kaletra)</td>
</tr>
<tr>
<td>Atazanavir</td>
<td>ATV</td>
<td>Reyataz</td>
<td>The first PI approved for once-daily dosing</td>
</tr>
<tr>
<td>Tipranavir</td>
<td>TPV</td>
<td>Aptivus</td>
<td>The first non-peptidic PI FDA-approved for HIV treatment</td>
</tr>
<tr>
<td>Darunavir</td>
<td>DNV</td>
<td>Prezista</td>
<td>The latest ARV drug on the market</td>
</tr>
<tr>
<td><strong>Entry and Fusion Inhibitors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enfuvirtide</td>
<td>INN</td>
<td>Fuzeon</td>
<td>Due to its high cost ($25,000/year/patient) and inconvenient dosing schedule, it is used for “salvage” therapy only</td>
</tr>
</tbody>
</table>
Table 1-2. Summary of class and agent-specific side effects of antiretroviral agents.

<table>
<thead>
<tr>
<th>Antiretroviral Agent</th>
<th>Adverse Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reverse Transcriptase Inhibitors</strong></td>
<td></td>
</tr>
<tr>
<td>Nucleoside analog reverse transcriptase inhibitors</td>
<td>Class effects: lactic acidosis, hepatic steatosis, pancreatitis, bone marrow toxicity, rash</td>
</tr>
<tr>
<td>Zidovudine</td>
<td>Severe headache, nausea, hepatotoxicity</td>
</tr>
<tr>
<td>Didanosine</td>
<td>Pancreatitis, peripheral neuropathy</td>
</tr>
<tr>
<td>Zalcitabine</td>
<td>Peripheral neuropathy, stomatitis</td>
</tr>
<tr>
<td>Stavudine</td>
<td>Peripheral neuropathy, lipodystrophy, pancreatitis, hyperlipidemia</td>
</tr>
<tr>
<td>Lamivudine</td>
<td>Minimal toxicity</td>
</tr>
<tr>
<td>Abacavir</td>
<td>Severe hypersensitivity, nausea, diarrhea</td>
</tr>
<tr>
<td>Emtricitabine</td>
<td>Minimal toxicity, palmar discoloration</td>
</tr>
<tr>
<td>Nucleotide analog reverse transcriptase inhibitors</td>
<td>Headache, diarrhea, nausea, vomiting, renal insufficiency</td>
</tr>
<tr>
<td><strong>Non-nucleoside reverse transcriptase inhibitors</strong></td>
<td>Class effects: rash, elevated transaminase levels, nausea, abdominal pain, fatigue</td>
</tr>
<tr>
<td>Nevirapine</td>
<td>Steven-Johnson syndrome, hepatitis</td>
</tr>
<tr>
<td>Delavirdine</td>
<td></td>
</tr>
<tr>
<td>Efavirenz</td>
<td>Insomnia, abnormal dreams, confusion, impaired concentration</td>
</tr>
<tr>
<td><strong>Protease Inhibitors</strong></td>
<td></td>
</tr>
<tr>
<td>Class effects: nausea, vomiting, diarrhea, hyperlipidemia (except atazanavir), fat maldistribution, hyperglycemia, possible increased bleeding in patients with hemophilia, elevated transaminase levels</td>
<td></td>
</tr>
<tr>
<td>Saquinavir</td>
<td>Headache</td>
</tr>
<tr>
<td>Ritonavir</td>
<td>Abdominal pain, peripheral and peri-oral parasthesias</td>
</tr>
<tr>
<td>Indinavir</td>
<td>Nephrolithiasis, indirect hyperbilirubinemia, metallic taste, alopecia</td>
</tr>
<tr>
<td>Nelfinavir</td>
<td>Severe diarrhea</td>
</tr>
<tr>
<td>Amprenavir</td>
<td>Skin rash</td>
</tr>
<tr>
<td>Lopinavir</td>
<td>Pancreatic toxicity</td>
</tr>
<tr>
<td>Atazanavir</td>
<td>Indirect hyperbilirubinemia, can cause the prolongation of the PR interval on ECG</td>
</tr>
<tr>
<td>Tipranavir</td>
<td>Increased risk of intracranial hemorrhage</td>
</tr>
<tr>
<td>Darunavir</td>
<td>Cold-like symptoms</td>
</tr>
<tr>
<td>Entry and Fusion Inhibitors</td>
<td></td>
</tr>
<tr>
<td>Enfuvirtide</td>
<td>Skin reactions at the injection site, severe allergic reaction, renal toxicity, paralysis,</td>
</tr>
</tbody>
</table>
CHAPTER 2
HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 PROTEASE

HIV-1 Protease Structure

The HIV protease (PR) is a member of the class of aspartic proteases. The classification is based on a number of criteria: (1) homology alignment of HIV protease with cellular aspartic PRs reveals the Asp-Thr-Gly conserved sequence common to all aspartic PRs; (2) several studies have shown that substituting the aspartic residue in the Asp-Thr-Gly sequence removes all catalytic activity; (3) the HIV PR is inhibited by pepstatin, a natural inhibitor that is selective for aspartic PRs; (4) crystal structures of HIV PR show a homodimer with each monomer providing a copy of the Asp-Thr-Gly sequence.

Structural studies on HIV-1 PR were initially hampered by the fact that it constituted a minor component of mature virions (Henderson et al. 1988b), which therefore necessitated the use of recombinant and synthetic technologies to produce the milligram quantities needed for structural investigations.

The HIV-1 PR is an obligatory homodimer that consists of two identical 99-residue subunits (Wlodawer et al. 1989). The dimer is stabilized by a β-sheet formed between the N- and C- termini of each subunit (Weber et al. 1989; Gustchina and Weber 1990). The active site of the enzyme is located directly above the dimer interface and contains two aspartic residues, one Asp residue is provided by each monomer (Navia et al. 1989). The carboxylate groups of Asp25 from both chains are nearly coplanar and show close contacts. The network is quite rigid due to the interaction (called “fireman’s grip”) in which each side-chain oxygen of the Thr26 accepts a hydrogen bond from the Thr26 main-chain NH of the opposing loop. Thr26 also donates a hydrogen bond to the carbonyl oxygen atom of residue 24 on the opposite loop. These aspartic residues are responsible for the catalytic function of the enzyme. As oriented in the figure 2-1, at
the bottom of the active site are the catalytic aspartic acids and at the top are two β-hairpins, one from each monomer, called flaps. The flaps are thought to undergo a large conformational change to open up and allow access to the active site (Rick et al. 1998). It is believed that the flaps need to open as much as 20 Å to allow the substrate to enter the active site (Rick et al. 1998). The crystal structures of HIV-1 PR submitted to the protein data bank reveal a 7 Å change in the flap orientation in the unbound form of the enzyme (Navia et al. 1989; Wlodawer et al. 1989).

**Function and Substrate Specificity**

Several studies (Kohl et al. 1988; McQuade et al. 1990) confirmed the necessity of an active PR in order to produce mature infectious virions (Figure 2-2). The function of the PR is to cleave gag and gag/pol polyproteins into the independent viral structure components and enzymes during maturation. This process is absolutely essential for viral maturation and rendering the PR inactive completely abolishes the viral infectivity.

The cleavage of gag and gag/pol follows a sequential order, most likely due to both substrate specificity and structural accessibility of the cleavage sites (Wiegers et al. 1998). Unexpectedly, HIV-1 PR has a broad specificity. Studies done using non-viral proteins as substrate showed that the PR has a broad preference for cleavage site amino acids (Tomasselli et al. 1993; Tomasselli and Heinrikson 1994). Despite this broad substrate preference, it has been possible to assign some general features to PR substrate specificity. The substrate must be at least 7 residues in length, and must be able to bind in an extended conformation (Figure 2-3). Hydrophobic-hydrophobic or aromatic-proline (Pro) residues seem to be preferred at position P1-P1’ flanking the scissile bond. Hydrophobic or Gln/Glu residues appear to be the most common residues at position P2’. Aromatic residues are almost never seen at position P3’. Small residues are preferred at position P2 (Dunn et al. 1994; Tomasselli and Heinrikson 1994). Also, several
studies showed that a residue in one subsite can influence the specificity for the residue found in the adjoining subsites (Dauber et al. 2002).

Binding of a substrate or an inhibitor introduces substantial conformational changes to the enzyme. The overall movement of the subunits can be described as a rotation of up to about 2° around a hinge axis located in the subunit β-sheet interface. This motion, which slightly tightens the cavity of the active site, seems to be also accompanied by a large motion of the flap region (Wlodawer et al. 1989).

**Gag/pol Processing**

The precursors gag (p55) and gag/pol (p160) are translated from the gag open reading frame and, together with the regulatory protein nef, serve as the substrate for PR activity.

An essential step in the viral life cycle is the processing of these two polyproteins: gag and gag/pol. The gag precursor contains the structural proteins of the viral core: matrix (MA), capsid (CA) and nucleocapsid (NC) (Henderson et al. 1990; Oroszlan and Luftig 1990) and as well as regulatory proteins (p1, p2, p6) (Pettit et al. 1994). The gag/pol precursor itself is a fusion between gag and pol polyproteins and it is translated as a result of a −1 translational frameshift (Jacks et al. 1988) at the NC/p1 junction, occurring with a frequency of about 5%. While the gag domain of gag/pol encodes for approximately the same protein as gag precursors, the pol domain of gag/pol contains PR, reverse transcriptase (RT) and integrase (IN) (Figure 2-4). The role of the frameshift mechanism is to maintain the correct ratio of structural proteins to enzymatic proteins (20 : 1). The frameshift site contains the “slippery sequence” (UUUUUUA) followed by a stem loop. During translation, the ribosome reaches the stem loop and stalls on the “slippery sequence” approximately 5% of the time. The ribosome than slips back one nucleotide and translates the pol domain of gag/pol polyprotein (Jacks et al. 1988; Somogyi et al. 1993). Any change to the frameshifting site reduces the efficiency of frameshifting, thereby altering the gag
to gag/pol ratio. Numerous studies have shown that the correct gag to gag/pol ratio is necessary for both viral assembly and replication (Karacostas et al. 1993; Hung et al. 1998; Shehu-Xhilaga et al. 2001).

The gag polyprotein is synthesized from the viral mRNA on the cytosolic ribosomes. During this step a myristic acid moiety, which increases gag’s affinity for the cell membrane, is added through a cotranslational modification (Gottlinger et al. 1989; Bryant and Ratner 1990). There are conflicting reports as to whether gag multimerizes prior to or following membrane targeting. However, the process of viral maturation involves a series of ordered cleavages and the processing of gag and gag/pol precursors is accomplished by the viral PR encoded within the pol domain of gag/pol fusion protein, without assistance from cellular PRs. Both the order and kinetics of cleavage as well as the extent of precursor processing appear to be critical steps in the generation of fully infectious, appropriately assembled viral particles (Mervis et al. 1988; Pettit et al. 1994; Wiegers et al. 1998). It has been shown that the initial cleavage takes place at the p2/NC cleavage site for both gag and gag/pol polyprotein through trans and cis mechanisms respectively (Pettit et al. 1994; Wiegers et al. 1998). This step releases NC and allows the condensation of the viral core. Cleavage between MA and CA then allows CA to separate from the membrane, and the cleavage at CA/p2 site releases the p2 spacer peptide and allows for CA condensation (Wiegers et al. 1998).

It is obvious that at least the first cleavage occurs when the PR is embedded within gag/pol fusion protein. Recent studies demonstrate that the location of the embedded PR limits the number of cleavages that can be made (Pettit et al. 2005). In fact, the Pro located at the first position in the PR is responsible for preventing PR from cleaving downstream sites (Pettit et al.
Thus, these data raise the possibility that the structure of gag/pol helps to determine the order of the processing by the PR.

Several studies demonstrated that the PR, flanked by short sequences that correspond the native sequences in the gag/pol polyprotein, has a decreased conformational stability (Wondrak and Louis 1996). These results suggest that the subunit interactions and, hence the dimer stability of the PR domain within gag/pol polyprotein, differ from those of the mature PR. This seems to have an effect on both the enzymatic activity of the enzyme (Louis et al. 1994; Louis et al. 1999) and the drug sensitivity of the HIV PR (Pettit et al. 2004). In fact, the mature, free PR is significantly more sensitive to the protease inhibitor (PI) ritonavir (RTV) than the immature, embedded PR (Pettit et al. 2004). The first cleavage to release PR from gag/pol polyprotein occurs at the N-terminus of the PR (Wondrak and Louis 1996), precedes the cleavage at the C-terminus, is accompanied by a large increase in the enzymatic catalytic activity and seems to lead to the stabilization of the dimer structure (Louis et al. 1999). Overall, PR release from the polyprotein appears to be a relatively late event in virion assembly, most likely occurring after the virion has been released from the membrane of the host cell.

**Matrix (MA)**

MA is located at the N-terminus of gag polyprotein and has two main roles: (1) targeting of gag to the plasma membrane, and (2) incorporation of the envelope glycoproteins into the new virus particles. Targeting of MA, thus of the gag or gag/pol polyproteins, to the cell membrane is mediated by both a cluster of conserved basic residues located near the N-terminal region (Yuan et al. 1993; Zhou et al. 1994) and a myristyl group that is added to the N-terminus (Lee and Linial 1994). In fact, Lee et al. showed that efficient particle formation can occur if the MA domain of HIV-1 gag is substituted by a myristylation signal. Evidence suggests that the myristyl moiety is exposed when MA is still part of the gag precursor, but is sequestered and hidden after
gag processing (Spearman et al. 1997; Hermida-Matsumoto and Resh 1999; Paillart and Gottlinger 1999). This myristyl switch model would account for the association of the gag precursor with the membrane during budding, but, at the same time, allow MA in the mature viral particles to dissociate from the membrane after infection and enable the viral core to enter the cytoplasm of the newly infected cells. MA is necessary for the incorporation of the envelope glycoproteins during virus assembly, and a stable interaction between MA and the cytoplasmic domain of the envelope suggests that gag and envelope interact prior to transport to the plasma membrane (Dorfman et al. 1994b; Wyma et al. 2000).

Capsid (CA)

HIV-1 CA protein is located between MA and p2 and consists of two domains separated by a flexible linker region (Figure 2-5) (Gamble et al. 1996; Gitti et al. 1996). The C-terminal domain is necessary for particle assembly as well as core formation, whereas the N-terminal domain is only required for core formation (Dorfman et al. 1994a; McDermott et al. 1996; Reicin et al. 1996; Borsetti et al. 1998). During virion maturation, CA assembles into a closed conical shell that surrounds the NC/RNA complex.

The C-terminal domain of CA mediates dimerization of gag/pol. Evidence suggests that the dimerization domain extends from CA into the spacer peptide p2, forming an $\alpha$-helix, which appears to be necessary for assembly (Gamble et al. 1997). The dimerization process is important in activating the PR and initiating the processing events. Studies showed that small molecules targeting the CA/p2 junction can slow PR-mediated cleavage at this site, thereby delaying virion maturation and reducing the viral infectivity (Zhou et al. 2004). The major homology region (MHR), which is a stretch of 20 amino acids in the C-terminal domain (Figure 2-5), does not seem to be involved in gag and gag/pol dimerization but is required for both viral particle assembly and the correct assembly of the viral core (Mammano et al. 1994).
p2

The p2 domain is located between CA and NC and its location is well conserved among primate lentiviruses, but the sequence and length are not well conserved (Henderson et al. 1988a). It has been suggested that p2 plays a role in the regulation of gag processing. Petit et al. demonstrated that cleavage at the CA/p2 site was accelerated significantly when cleavage at p2/NC site was blocked, suggesting that p2 may function to slow the cleavage rate at CA/p2 site during the processing (Pettit et al. 1994). Further evidence by this group also indicates that p2 is necessary for the correct formation of the core of the virus, as virions produced from viruses with p2 deleted were not infectious and contained an improperly formed core (Pettit et al. 1994).

Nucleocapsid (NC)

NC has highly conserved zinc binding motifs whose alteration results in decreased efficiency of reverse transcription and initial integration process, decreased packaging of genomic RNA, as well as decreased viral infectivity (Poon et al. 1998; Buckman et al. 2003). Because of the important role NC plays in the viral life cycle, there were numerous attempts to develop drugs that would remove the zinc atom, thus destabilizing the binding motif. These trials have been proven unsuccessful because these drugs also removed zinc atoms from the cellular proteins. Current studies are focusing on the identification of compounds that bind reversibly and are non-chelating (Stephen et al. 2002).

p1

p1 is located between NC and p6. It is 16 amino acids long in HIV-1 subtype B and has a remarkable high sequence identity between all the HIV-1 subtypes. Research centered on p1 has been limited and as yet no function has been ascribed to this spacer peptide. Because of its location, p1 overlaps with the “slippery site” and the stem loop of the frameshift site.
Several studies (Hill et al. 2002; Hill et al. 2005) showed that the two highly conserved Pro residues within p1 are important for viral infectivity. Also multiple other changes in the amino acid sequence of p1 resulted in altered protein processing, reduced viral genomic RNA dimer stability, and abolished viral infectivity (Hill et al. 2002).

**p6\textsuperscript{GAG}**

\textsuperscript{p6}\textsuperscript{GAG} is located at the C-terminus of gag polyprotein and three functional domains have been mapped to it. (1) The P(T/S)AP motif located at the N-terminus binds to the host cellular factor TSG101 and is required for efficient virus release (Gottlinger et al. 1991; Huang et al. 1995; Martin-Serrano et al. 2001). (2) AN LXSLFG motif in the C-terminus binds vpr, and mutations in this region diminish vpr incorporation into virions (Kondo and Gottlinger 1996). (3) LYP motif is situated upstream to LXSLFG, and both are necessary for the binding to the cellular factor AIP1 and modulate virion budding (Strack et al. 2003). AIP1 directly binds to TSG101 and von Schwedler et al. suggested a model in which gag forms a ternary complex with both TSG101 and AIP1 during budding (von Schwedler et al. 2003). In uninfected cell, these factors are involved in a pathway that sorts cellular proteins into vesicle that bud into multivesicular bodies, and the interaction of HIV-1 p6 with these proteins suggests that in infected cells, HIV-1 uses this cellular network to promote budding of new virions.

**Protease Inhibitors and Drug Resistance**

As the proteolytic activity of the PR is absolutely required for the production of mature, infectious virions, this enzyme was soon identified as an attractive target for therapeutic interventions, and the discovery and development of PIs are a great success of modern pharmacology and structural biology.

There are several criteria that have to be taken in consideration when designing a PI. The basic design criterion relied on the observation that HIV-1 PR, unlike other PRs, is able to cleave
Tyr-Pro or Phe-Pro sequences in the viral polyprotein. Because the amide bonds of Pro residues are not susceptible to cleavage by mammalian endopeptidases, the design of HIV-1 PR inhibitors based on this criterion was expected to bring potential advantages of higher selectivity: the PIs should be able to bind the PR with high affinity but should also be able to avoid binding with high affinity structurally similar human enzymes, such as pepsin or cathepsin D, that can lead to various adverse effects. Due to the high polymorphic nature of HIV PR, the PIs must be able to bind tightly to as many variants as possibly. This will also aid in evading the occurrence of resistance. Because the fight against HIV requires the use of multiple drugs targeted at both HIV and opportunistic diseases, the PIs must be able to function in the presence of other drugs circulating in the blood at high concentration. If possible, it is preferable that the PIs be able to access body compartments that tend to function as viral reservoirs such as the CNS. Also, the costs of producing the drug should be low enough to allow open access for those in need of treatment. This is very important because HIV/AIDS is a chronic disease requiring a lifetime of treatment with many drugs.

PIs (Figure 2-6) are small peptidic and nonpeptidic molecules that mimic the natural substrate in the gag/pol polyprotein and therefore compete for active-site binding. Once bound, the inhibitor cannot be cleaved and consequently inactivates the enzyme. Most of the inhibitors co-crystallized with HIV PR, including all peptidomimetic inhibitors, are bound in the enzyme active site in an extended conformation and the hydrogen bonds are made mostly between the main-chain atoms of both the enzyme and the inhibitor. The hydroxyl group at the non-scissile junction present in inhibitors is positioned between the D25/D25 carboxyls of the PR, within hydrogen-bonding distance to at least one carboxylate oxygen of each aspartate. A feature common to almost all complexes of HIV-1 PR is a buried water molecule that bridges the P2 and
P1’ carboxyl groups of the inhibitor and Ile50 and Ile150 amino groups of the flaps. This water is approximately tetrahedrally coordinated and is completely separated from the bulk solvent.

A number of distinct subsites or pockets that accommodate side chains of the inhibitors can be identified in HIV PR. Three subsites on each side of the non-scissile bond (S1 – S3 and S1’ – S3’) are very well defined, whereas more distant subsites are not as clear. The amino acid composition of these pockets defines the specificity and affinity of the enzyme for a particular inhibitor or substrate, as was described previously in this chapter.

**Drug Resistance**

PI resistance, defined as the loss of virus susceptibility to a PI, involves multiple gradual changes that accumulate within the PR sequence (Erickson et al. 1999b). This happens during continued viral replication in the presence of PIs that allows mutations to accumulate either at the active site or within the core of the PR.

At the molecular level, resistance to PIs predominantly takes the form of mutations within the HIV PR molecule that preferentially lower the affinity of PIs with respect to PR substrates, while still maintaining a viable catalytic activity. Basically, resistant mutations arise because they provide the virus with an advantage to survive in the presence of the drug.

There are numerous ways to assess drug resistance. There are clinical studies, genotypic and phenotypic assays, and biochemical studies. Biochemical resistance is a measure of the reduction in drug binding affinity to a variant enzyme compared to the wild type enzyme. This measure is reported as a measure of the dissociation constant for a particular inhibitor with a particular enzyme variant. These are measurements conducted in our studies. It is important to distinguish the differences in classifying resistance because clinical resistance can be due to a number of factors, including viral and host genetics, biochemical, cellular, and pharmacological drug levels. Though most resistance can be traced to genomic mutations, there have been
instances where resistance has occurred without detectable mutations within the PR sequences. It is now understood that factors that influence drug absorption, distribution, metabolism, and excretion have a strong role in the development of resistance and can vary between individuals (Burger et al. 1997; Hugen et al. 1999).

A number of resistance mutations have been characterized for all clinically used PIs, including tipranavir (TPV), one of the most recent and most potent PI on the market (Figure 2-7). In fact, over 57 mutations have been observed in at least 27 positions in the 99-residue HIV PR monomer in response to drug selection pressure (Hammond et al. 1997).

PI resistance mutations can be classified as shown in Figure 2-7, retrieved from HIV Drug Resistance Database. The mutations are categorized into five groups: (I) Major mutations that, with the exception of L90M, are located in the substrate cleft. Each by itself is capable of reducing susceptibility to one or more PIs. With the exception of V82I, which does not cause drug resistance, mutations at these positions are nonpolymorphic in that they do not occur in PI-naïve persons. (II) Flap mutations which occur in the PR flap and are second in importance only to the major mutations. Some may cause resistance by themselves, but these mutations are more often accessory. They are also nonpolymorphic. (III) Other nonpolymorphic PI-resistance mutations that usually indicate past PI exposure. Several cause resistance by themselves but more often they are accessory. (IV) With the exception of L10F/R, these mutations are polymorphic. They contribute to resistance only when present in combination with one or more category I-III mutations. (V) There are recently reported mutations that have been associated with resistance to ATV, TPV, or DRV (TMC-114). Some are non-polymorphic and likely to be significant. Others are polymorphic and less likely to be significant.
PI resistance mutations can also be classified as active site versus non-active mutations according to whether they occur inside or outside the inhibitor binding subsites.

**Active site mutations**

The active site mutations are the first mutations observed in response to therapy and have received a great deal of attention. They occur in the substrate-binding cleft of the enzyme and usually directly reduce inhibitor binding to PR. Most notable are V82A/F/T/S, I84V, and D30N. Mutations at position 82 accumulate in response to APV, IDV, RTV, and SQV (Condra et al. 1995; Molla et al. 1996; Patick et al. 1998). Mutations at residue 84 are seen in response to APV, IDV, and SQV therapy. The D30N mutation is very specific, only seen in response to NFV therapy. Other residues commonly mutated during drug therapy are found in the flaps. These mutations, unlike the previous mutations discussed, are more selective to specific PIs. The mutation I47V and I50V are selected against APV, the mutation G48V is selected in response to SQV, and I50L arises in response to TPV treatment.

Even if the active site mutations have varying degree of specificity for different PIs, substantial cross-resistance has been reported for most of them (Hirsch et al. 2000; Kantor et al. 2001). In general, mutations conferring significant resistance are absent in untreated populations; this is consistent with the idea that these substitutions are themselves deleterious to virus replication and are maintained only in the presence of the inhibitor.

**Non-active site mutations**

It was argued that the main role of the non-active site mutation is to compensate for reduced catalytic efficiency of the enzyme. However, not too long ago, non-active site mutations have been biochemically characterized as having an active role in decreasing inhibitor binding affinity (Olsen et al. 1999; Perno et al. 2002; Muzammil et al. 2003). Most of the non-active site residues, which mutate in response to drug therapy, are buried in the PR core; and changes in the
PR core interaction could, through a cooperative effect and via long range interactions, translate to structural changes in the active site. Recent studies confirm the fact that many of these substitutions that arise in response to drug therapy in subtype B PR, occur as baseline polymorphisms in non-B subtypes as discussed below (Tanuri et al. 1999; Shafer et al. 2000; Vergne et al. 2000; Pires et al. 2004).

One of the non-active site mutations analyzed in our work is L90M. It is considered a major mutation because its occurrence decreases the binding affinity for almost all clinically used PIs without a significant effect on the catalytic efficiency of the enzyme. The L90M mutation appears to exert its effect by affecting the structural conformation of the active site loop containing the catalytic aspartic acids (Prabu-Jeyabalan et al. 2003). Mutations in the flaps at positions 46 and 54 could also affect the interactions of the flaps with the PI or change the dynamics of the opening and closing of the flaps (Clemente et al. 2004). It is worth noting that most non-active site mutations, similar to most of the active site mutations, are not specific to any single PI.

**Naturally Occurring Polymorphisms in HIV-1 Protease**

These mutations are defined as position specific differences from consensus sequence within the subtype being evaluated, occurring in more than 1% of the sequences isolated from drug-naïve patients. Usually these polymorphisms are expected to appear at sites that are less critical for enzyme activity and where variation within the protein may be driven by immune selection and functional adaptations (Kantor and Katzenstein 2003). These polymorphisms are far more frequent in non-B subtype than in subtype B PRs and seem to be located at positions known to be associated with PI resistance in subtype B PR (Tanuri et al. 1999; Caride et al. 2001; Perez et al. 2001; Holguin et al. 2002; Gordon et al. 2003). Accessory/secondary mutations are defined as residue changes that arise in response to drug therapy. For certain positions,
depending on the subtype analyzed, these two terms can be used interchangeably. Some other residues are known only as baseline polymorphisms, while other residues are considered to be only secondary mutations and others are not categorized in one of the previously mentioned groups.

The precise role of baseline polymorphisms in non-B subtypes is not well understood but a number of studies postulate that they can have various functions:

- **Increase the catalytic efficiency of HIV protease.**

  Recent *in vitro* studies performed on subtype A and C PR mutants (Velazquez-Campoy et al. 2001) showed that in the presence of PIs such as IDV, SQV, RTV, and NFV, PRs from subtypes A and C consistently performed their catalytic functions better than PRs from subtype B. These results point to a greater biochemical fitness of the subtypes A and C PRs in the presence of the existing inhibitors.


  D30N is a primary mutation that renders the HIV-1 PR less susceptible to NFV, and does not exhibit cross-resistance with other PIs (Patick et al. 1998). Grossman et al. showed that there is a significant difference in the occurrence of the D30N mutation between subtype B and subtype C infected patients (P = 0.03) when treated with NFV (Grossman et al. 2004). The authors conclude that the rates at which these mutational pathways develop differ in subtype C and subtype B infected patients failing therapy, possibly due to differential impact of baseline polymorphisms. This has important implications because NFV was often the first-line drug of choice for subtype B infected patients, as the frequent emergence of the non-cross-resistant D30N mutant would not bar the use of alternative drug combinations. This strategy has no such advantage over other PI in the treatment of HIV-1 subtype C infected patients.
• **Influence the speed of acquiring PI-related resistance mutations.**

   It has been argued that PI resistance-associated accessory substitutions in subtype B that correspond to baseline polymorphisms in non-B subtypes, may not result in a significant decrease in sensitivity to drugs but are associated with an increase in viral fitness (Erickson et al. 1999a; Hirsch et al. 2000). Thus, the appearance of a major mutation in a genome already containing accessory mutations could influence the speed with which highly resistance viruses are selected during therapy (Vergne et al. 2000).

• **Contribute to resistance and/or maintenance of viral fitness once primary resistance mutations occur.**

   Rose et al. (Rose et al. 1996) showed that baseline polymorphisms or secondary mutations such as L10I are necessary to accommodate a combination of other three mutations responsible for drug resistance.

   Also, these naturally occurring polymorphisms could introduce differences in structural stability of the PR that may influence binding affinity due to the required conformational change associated with substrate and inhibitor binding (Velazquez-Campoy et al. 2003). This is important because conformationally constrained ligands, such as the PIs in clinical use, have little capacity to adapt to changes to the target site and they lose affinity significantly, even when confronted with conservative changes (Velazquez-Campoy et al. 2002).

• **Promote a poorer response to therapy.**

   Servais et al. provided evidence that response to triple-drug therapy including a PI is closely related with the overall number of baseline polymorphisms. The main effects of the natural polymorphisms are most likely to manifest themselves after the onset of drug resistant mutations associated with ARV therapy. If this is the case, the amplification of drug resistance
effects might have serious consequences on the long-term viability of PI therapy in non-subtype B patients (Velazquez-Campoy et al. 2003).

In our work, we aim to explore the role of naturally occurring polymorphisms in subtype C PR. Our goal is to understand how these sequence variations affect the overall characteristics of subtype C PR and ultimately how the response to treatment is modulated by the preexistence of the baseline polymorphisms. Combining kinetic and structural data will give us a broad understanding of the interactions that are being affected by residue changes and how different inhibitors or substrates interact with the same enzyme. This study will bring more information about the contribution the baseline polymorphisms have in modulating the subtype C PR affinity for the substrates and inhibitors. If our study supports the theory that the high rate of natural polymorphisms in the protease gene in subtype C viruses compromises the protease-inhibitor interactions, this might have important consequences on the feasibility of continuing to use current HIV-1 PIs for subtype C infections. Furthermore, our data will allow for prediction about how a certain viral variant, depending on the polymorphisms preexistent within the PR sequence, will respond to drug therapy and what mutational pathway is more likely to evolve. This sort of knowledge will provide clinicians with information that will allow for optimization of therapeutic regimens. This will also provide researchers with clues to designing inhibitors that will retard the evolution of resistant protease variants and exhibit less cross-resistance with other classes of PIs.
Figure 2-1. HIV-1 protease (PBD code 1HVP). The catalytic aspartic residues (D25/D25‘) are shown in salmon-colored sticks. In the active site, the protease inhibitor indinavir (IDV) is shown in gray. Figure rendered with Pymol (DeLano Scientific).
Figure 2-2. Cryo-electron micrographs and schematic representations of HIV-1 particles. A) Immature HIV-1 virions containing the gag and gag/pol polyproteins assembled in spherical layer underneath the viral envelope. B) Once active, the viral PR cleaves gag and gag/pol into their constitutive components: matrix (MA), capsid (CA), nucleocapsid (NC) and p6 proteins thus causing the formation of mature, infection virion. Characteristic for this stage is the cone-like shape of the viral core. The color code is as follows: blue – viral envelope proteins, yellow – lipid bilayer, black – viral proteins, green – viral RNA. Reprinted and adapted from Briggs et. al., 2004.
Figure 2-3. Hydrogen bonding between HIV-1 protease and a modeled substrate. Substrate and inhibitor residue side chains are designated as $P_n$ for those at the N-terminus of the scissile bond and $P_n'$ for those at the C-terminus of the scissile bond as defined by Schechter and Berger, 1967. Reprinted and adapted from Wlodawer and Vondrasek, 1998.
Figure 2-4. HIV-1 gag and gag/pol polyproteins. The gag precursor contains the structural proteins of the viral core: matrix (MA), capsid (CA) and nucleocapsid (NC) and as well as regulatory proteins (p1, p2, p6$_{GAG}$). The gag domain of gag/pol encodes for approximately the same protein as gag precursors and the pol domain of gag/pol contains PR, reverse transcriptase (RT) and integrase (IN). The HIV PR is represented as blue ribbon.
Figure 2-5. Structure of HIV-1 capsid monomer (PDB code 1E6J). The color code and abbreviations are as follows: green – NTD – amino-terminal domain, blue – CTD – carboxyl-terminal domain, red – FLR – flexible linker region, yellow – MHR – major homology region. Figure rendered with Pymol (DeLano Scientific).
Figure 2-6. Structures of nine individual protease inhibitors, approved for clinical use in HIV-infected patients.
Figure 2-7. HIV-1 protease inhibitors resistance chart. Color code: dark blue represents high level of resistance, lighter tones match lower levels of resistance, (*) – hypersusceptibility, (?) – unknown. Also see text. Accessed and adapted on June 2007 from http://hivdb.stanford.edu – HIVDatabase, Stanford University.
CHAPTER 3
MATERIALS AND METHODS

Sub-Cloning of HIV-1 Protease

The HIV-1 subtype B protease LAI clone (wild type) was acquired from Dr. Maureen Goodenow, Department of Pathology, University of Florida. The HIV-1 subtype C protease (PR) clone p94IN476.104 and subtype A PR clone p92UG037.1 were obtained through the AIDS Research and Reference Reagent program, Division of AIDS, NIAID, from Drs. Rodenburg, Gao, and Hahn (Rodenburg et al. 2001). The HIV-1 PR was sub-cloned from the gag/pol gene sequence using primers to either end of the PR gene sequence. The primers were engineered to insert the restriction site NdeI at the 5’ end and BamHI at the 3’ end of the PR for directional cloning into expression vector pET23a shown in figure 3-1. The PCR reaction was initiated by mixing together gag/pol DNA, forward and reverse primers, Taq-polymerase, 10X Taq-polymerase buffer, MgCl₂, dNTPs and water to a final volume of 50 μl. The PCR reaction protocol consists of 4 steps: (1) the reaction was started by heating the PCR mixture to 95°C for 1 min; (2) the reaction was then cycled 18 times through a melting step at 95°C for 30 sec, an annealing step at 45, 51 or 55 °C for 30 sec, and an elongation step at 72 °C for 1 min; (3) the temperature of the reaction remained to 72 °C for 10 min; (4) the temperature of the reaction was dropped to 4 °C until the PCR cycler was turned off. The results were checked on a 1% agarose gel (Figure 3-2A). The PCR product was cloned into TOPO-cloning vector, pCR2.1, using the TA Cloning Kit from Invitrogen. TA-cloning reaction entails combining 1 μL PCR product, 1 μL T-4 Ligase (4 Weiss Units), 1 μL 10X Ligation buffer (Invitrogen), 2 μL pCR 2.1 vector (25 ng/μL), and 5 μL of H₂O with an incubation step at room temperature (25 °C) for 1 h. One microliter of TA-cloning reaction was transformed into chemically competent Top10 cells (Invitrogen), using the transformation procedure described below. The Top10 cells harboring the
cloned PR gene into the pCR2.1 vector, 750 μL of cell suspension, were stored at –80 °C in LB media (1 Liter preparation contains 10 gm of Tryptone, 5 gm of yeast extract, and 10 g NaCl, pH 7) supplemented with 10% glycerol (120 μL of 75% glycerol).

**Directional Cloning – Digestion and Ligation**

The constructs containing the PR gene cloned in pCR2.1 vector, in Top10 cells from glycerol stocks, were grown over night (16 hours) in 7 mL of LB with 50 μg/ml of ampicillin. The cells were harvested using a Beckman GS-15R centrifuge with Beckman rotor S4180 at 4800 rpm for 10 minutes. The plasmid containing the gene of interest was purified from the Top10 cells using a Qiagen Spin Mini-Prep Kit (Qiagen). In brief, the procedure involves the resuspension of pelleted bacterial cell in prechilled (4 °C) P1 buffer (Qiagen) in the presence of RNase A. The lysis solution, solution P2, is a mixture of NaOH and SDS (Qiagen). The lysate solution is neutralized and brought to a high salt concentration in one step using solution N3 (Qiagen). The high salt concentration causes cellular components, chromosomal DNA and SDS to precipitate leaving the plasmid DNA in the supernatant. The plasmid DNA is bound to a DNA binding column. The DNA on the column is washed with 75% ethanol to remove any salt. The DNA is eluted off the column using 35 - 50 μl of water. The concentration of DNA prep was measured at 260 nm with a UV spectrophotometer (CaryUV Bio50), using the formula: OD_{260DNA} \times 50 \text{ ng/μL} \times \text{ dilution factor}, where the dilution factor is the DNA : final solution ratio in microliters. The plasmid was then double digested with the restriction enzymes NdeI and BamHI. Plasmid pET23a was subjected to the same procedure. Digestions were done using 3-10 μg of DNA, 1 μL of NdeI (10 U/μL) and/or 1 μL of BamHI (10 U/μL), 1 μL of 10X BamHI Buffer, 0.5 μL of BSA, and dH2O to total volume of 30 μL, 37 °C for 1 h. The bands of the double digested samples at 300 base pairs (HIV-1 protease) and 3600 base pairs (pET23a) were
excised and gel purified using Spin Gel Purification Kit (Qiagen). The insert gene and plasmid DNA (3:1 or 10:1 weight ratio) were ligated together in a reaction mixture containing 1 μL of T-4 ligase (4 Weiss units), 1 mM dATP, and 1.5 μL ligase buffer (10X Buffer), in a total volume of 15 μL. One microliter of ligation product was transformed into Top10 cells. DNA was purified from these cells using the Qiagen Spin Mini-Prep procedure (Qiagen) and it was used to transform (see page 63) expression cells BL21 DE3(Star) pLysS (Invitrogen). The presence of the insert in the expression vectors was verified by double digestion with NdeI and BamHI (Figure 3-2B). Two micrograms of DNA were also sent for DNA sequencing. DNA sequencing was performed by the UF-ICBR sequencing core.

**Mutagenesis**

Mutations D30N, N88D, and L90M were added to HIV-1 subtype B and subtype C proteases using the Quick-Change Site Directed Mutagenesis Kit (Stratagene). Mutations are generated using two primers that are complementary to both the coding and the non-coding strands surrounding the base or bases to be mutated. The reactions were initiated by mixing together 50-100 ng of plasmid DNA, 5 μL of Pfu polymerase buffer, 100 ng of dNTPs, 125 ng of upper and lower primer, 1 μL of Pfu polymerase (2.5 U), and brought to a final volume of 50 μL with ddH2O in a 250 μL PCR tube. The PCR reaction protocol consists of 4 steps: (1) the reaction was started by heating the reaction to 95°C for 30 sec; (2) the reaction was then cycled 18 times through a melting step at 95°C for 30 sec, an annealing step at 51 or 55 °C for one minute, and an extension step at 68°C for 4 minutes; (3) the temperature of the reaction was dropped to 4°C and held for 7 minutes; (4) the temperature of the reaction was held at 4 °C until the the PCR cycler was turned off. To remove the template, 1 μL of restriction enzyme DpnI (10
U/μL) was added to all PCR reactions at the end of the cycle and placed at 37°C for two hours. 1.5 μL of the PCR reaction was used to transform competent Top10 cells (Invitrogen).

Transformation

All transformations were done using chemically competent cells, which require heating to 42°C to uptake the DNA vector. The vector containing the DNA of interest was mixed with 15 – 20 μL of cell stock in a microcentrifuge tube and placed on ice for 30 min. The reaction tube was then placed into a 42 °C water bath for 45 – 60 sec. Immediately afterwards it is placed on ice for 10 min. One hundred microliters of SOC media (2.0% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, and 20 mM glucose, pH 7) were then added to the reaction tube and placed in an incubator rotating at 250 rpm at 37°C for 45 – 60 min. Following incubation 50-100 μL of cell culture were spread over plates containing LB-ampicillin (50 μg/mL) for Top10 cells and LB-ampicillin and chloramphenicol (34 μg/mL) for BL21 Star DE3 pLysS cells. The plates were then incubated overnight in a 37 °C incubator to promote colony growth. The next day a colony was randomly picked and grown overnight (16 hr) in 7 mL of LB-ampicillin (50 μg/mL) or LB-ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL) to make 10% glycerol stocks and stored at -80°C.

Protein Expression and Inclusion Body Extraction

The recombinant PR variants were purified from the over-expression of the insert gene in BL21 Star DE3 pLysS cells pET23a vector. Four 1 L expressions in M9 minimal media (in 1L: 6.8 g Na₂HPO₄, 3 g KH₂PO₄, 0.5 g NaCl, 1 gm NH₄SO₄, 5 g Casamino Acids autoclaved together in 987 mL of H₂O, then 1 mL of 0.1 M CaCl₂, 2 mL of 1.0 M MgSO₄, and 10 mL of 20% glucose are added) were initiated with a 4% inoculation from overnight cultures of cells grown in LB media supplemented with 50 μg/mL of ampicillin and 34 μg/mL of
chloramphenicol for BL21 Star DE3 pLysS cells. The expression cultures were grown to an OD$_{600}$ of 0.8 in a 37 °C incubator rotating at 250 rpm. Expression of the recombinant protein was induced with 1 mM IPTG. Two 1 mL samples were taken at zero, one, two, and three hours post induction to monitor expression. One milliliter was used to determine the OD$_{600}$ of the culture. The second milliliter was centrifuged at 14,000x g for 1 minute to harvest the cells. The supernatant was discarded and the cell pellet was resuspended in 200 μL of IB buffer (50 mM Tris, pH 7.5, 3 mM DTT, 2 mM EDTA). 30 μL of the resuspension and 6 μL of LSB (Laemmli sample buffer) were mixed together and boiled for 3 – 5 minutes to lyse the cells and denature the proteins. 30 μL sample for the zero hour and a standardized volume based on the OD$_{600}$ for each sample at one, two, and three hours were loaded into a 15% or 18% Tris-tricine SDS PAGE gel (Figure 3-3). After three hours after induction with IPTG the cells were harvested by centrifugation at 14,300x g for 10 min and resuspended in 30 – 50 mL IB buffer. Cells were lysed using an SLM-Aminco French Pressure cell at a 1000 psi. The inclusion body pellet was retrieved by centrifuging the cell lysate over 10 mL of 27% sucrose in 30 mL Corex tubes in a JS-13.1 swinging bucket rotor at 12,700x g for 50 min. The procedure was repeated two times. After the last spin the wet weight of the inclusion bodies was recorded and a sample was taken to run on a gel shown in Figure 3-3.

**Protein Refolding**

The inclusion bodies were dissolved at 1 mg/mL in 8 M urea supplemented with 300 mM β-mercaptoethanol, 50 mM CAPS buffer pH 10, and 1 mM EDTA, and stirred at room temperature for 1 h. The urea solution was prepared in 1 L total volume and mixed with 5 g of Dowex ion exchange resin and stirred at room temperature for 1 h. The ion exchange resin was removed using a 2 μm filter. The suspension of denatured protein was loaded into Spectra-Por
MWCO 6000-8000 dialysis membrane (Spectrum). The suspension was dialyzed against 4 L of sodium phosphate buffer (5 mM Na₂HPO₄, pH 7.0, 5 mM EDTA, 300 mM NaCl, and 1 mM DTT) in a 4-L graduated cylinder for 2 h at 4 °C. The buffer was then exchanged for fresh buffer dialyzed over night at 4 °C. The protein suspension was never more than 6% of the total volume of dialysis buffer.

Ammonium Sulfate Precipitation

To concentrate the HIV-1 protease, the post-dialysate was subjected to ammonium sulfate ((NH₄)₂SO₄) precipitation. The post-dialysate was stirred on ice and (NH₄)₂SO₄ was slowly added to 30% saturation and allowed to stir for 1 h. The solution was centrifuged at 14,300x g for 60 min and the supernatant was returned to ice and stirred. The precipitate was dissolved in 5 – 10 mL of FPLC buffer (50 mM K₂PO₄, pH 6.5, 2 mM EDTA, 100 mM NaCl, 2 mM DTT, 5% glycerol) and transferred to a 15 mL conical tube. The above procedure was repeated for 50% ammonium sulfate saturation. The resuspended precipitates from the 30% and 50% ammonium sulfate fractions in the 15 mL conical tubes were centrifuged in a Beckman GS-15R centrifuge with Beckman rotor S4180 at 4800 rpm to remove any precipitate that did not re-dissolve. The supernatants were transferred to a clean 15 mL conical tube and checked for activity.

Size - Exclusion Chromatography

The ammonium sulfate fractions that contain activity were loaded onto a Hi-Load 16/60 Superdex 75 (Amersham Pharmacia) gel filtration column that was prewashed with 0.1 M NaOH and water and equilibrated with the FPLC buffer. The column was connected to an FPLC system (Amersham Pharmacia) driven by an LCC-500 Plus controller. Two mL fractions were collected using a Frac 200 (Amersham Pharmacia). Fractions showing protein peaks by absorbance at 280 nm were tested for activity (Figure 3-4A) and fractions containing active enzyme were verified.
to be HIV-1 PR by SDS-PAGE, which also allowed us to determine the presence of any contaminants (Figure 3-4B).

**Protease Kinetic Studies**

**Determination of Michaelis-Menten Constants**

All HIV-1 subtype B and C variants were assayed kinetically in 50 mM sodium phosphate buffer, pH 4.7 containing 150 mM sodium chloride, 2 mM EDTA, at 37 °C. Reactions were initiated when 230 μL containing enzyme, buffer, and filtered water were added to 20 μL of a chromogenic substrate giving a total volume of 250 μL. Both the enzyme mixture and the substrate were pre-incubated at 37 °C for 3 min. Variable chromogenic substrate concentrations were used ranging from 5 – 120 μM. The substrates were synthesized by the ICBR Protein Chemistry Core using the solid phase method with an Applied Biosystems Model 432A automated peptide synthesizer. A stock of the substrate was made at 10 mg/mL in 5% formic acid. The substrate stock was analyzed by the ICBR Protein Chemistry Core to determine an accurate substrate concentration. At least six different substrate concentrations were used to determine the Michaelis-Menten constants (Michaelis and Menten 1913). The cleavage of the substrate was monitored using a Cary 50 Bio Varian spectrophotometer equipped with a 18-cell multitransport. Constant temperature of the reaction was maintained at 37°C in quartz cuvetts using a water pump. A plot of the initial rate data for each substrate concentration versus substrate concentration gives the characteristic Michaelis-Menten curve. Fit of the curve to the following equation:

\[
v = \frac{V_{\text{max}} \cdot S}{(K_m + S)}
\]
using non-linear Marquardt analysis was used to determine the $K_m$ and $V_{max}$ for the substrate (Figure 3-5) (Marquardt 1963). The $k_{cat}$ values were determined using the following equation:

$$k_{cat} = \frac{V_{max}}{E_t}$$

**Dissociation Constant ($K_i$) Determination**

To assess the effect of various mutations on inhibitor binding the $K_i$ (dissociation constant) values for various clinically used inhibitors of the HIV-1 PR were determined for each variant. Inhibition is measured as the decrease of the rate of substrate cleavage in the presence of inhibitor. After acquiring the Michaelis-Menten curve in the absence of inhibitor the assay is repeated twice in the presence of two different concentrations of inhibitor. The curves are then simultaneously fit to the following equation:

$$v = \frac{V_{max}}{(1 + \frac{K_m}{S}) \cdot (1 + \frac{I}{K_i})}$$

to determine $K_i$ values of classical competitive inhibitors (Figure 3-6). The following equation was used to determine the $K_i$ values of tight binding inhibitors (Williams and Morrison 1979).

$$\frac{v}{v_o} = \frac{E_t - I_t - (K_i \cdot ((S/K_m) + 1)) + \sqrt{(E_t - I_t - (K_i \cdot ((S/K_m) + 1)))^2 + 4 \cdot E_t \cdot (K_i \cdot ((S/K_m) + 1))}}{2 \cdot E_t}$$

**Enzyme Active Site Titration**

The active enzyme concentration in our assays is determined through active site titration of the enzyme using a tight binding inhibitor ($K_i \leq 1$ nM). The assay consisted of monitoring the rate of substrate cleavage with a single substrate concentration in the presence of various inhibitor concentrations in a total volume of 250 μL. Reactions are initiated when a series (12-18
different reactions) of 230 μL containing buffer (same as used in $K_m$ and $K_i$ determinations), fixed enzyme concentration, zero and increasing concentration of inhibitor (until the rate of the reaction was inhibited totally) and water is mixed with 20 μL of a fixed substrate concentration. Both the enzyme mixture and the substrate were pre-incubated at 37 °C for 3 min before mixing. Cleavage of the substrate in the presence of inhibitor is monitored as described before. The assay is set up to have the condition where $[E] \gg K_i$ for the inhibitor used. A plot of initial rate versus inhibitor concentration gives an exponential decay curve that when fitted to the equation above gives the active enzyme concentration (Figure 3-7).

**Crystallization Studies**

**Protein Sample Preparation**

The clone used to express HIV-1 subtype C PR was ordered through DNA2.0. The DNA composition of this clone was codon-optimized: modified to harbor codons that are found with high frequency in the E. coli strains. Such changes allow for expression of higher quantities of protein, necessary for crystallization studies. However, the amino acid sequence is similar to that used in our kinetic studies except three differences: Q7K, L33I, L63I. The HIV-1 subtype C PR was expressed, refolded, purified and concentrated as described previously.

To set up crystallization trays, the HIV-1 subtype C PR was concentrated to 3 mg/mL (140 μM) using a 5 kD VivaSpin 15R MWCO 3000 spin concentrator (VivaScience). Then the buffer was exchanged for 30 mM sodium acetate, pH 4.7 containing 2 mM DTT. To obtain the drug-bound forms, the HIV-1 subtype C PR was incubated with either indinavir (IDV) or nelfinavir (NFV) (obtained from the NIH Research and Reference Reagent Resources Program), two of the clinically used PR inhibitors, in a 3-fold molar excess, for 60 min at 4 °C in 20 mM sodium acetate pH 4.5, 2 mM DDT prior to crystallization. The enzyme-inhibitor mixture was then spun
at 14,000 rpm for 30 min to pellet any precipitation that might have occurred during incubation step.

**Crystal Preparation**

Initial crystallization trials were conducted using the hanging-drop vapor-diffusion method (McPherson 1982), screening various conditions from Crystal Screen 1, Light and Cryo kits (Hampton Research). Two microliters of the unbound or the drug-complexed forms were mixed in 1:1 ratio with the various reservoir solution on siliconized glass circle cover slides (Hampton). The droplet was suspended over 1 mL of the crystallization buffer. Based on the results of the crystallization screening, useful X-ray diffraction-quality crystals of HIV-1 subtype C PR were obtained by mixing 2 µL of enzyme with 2 µL of reservoir solution consisting of 30 mM citric acid, pH 5.0 and 1 M sodium chloride. The crystals formed overnight or within several days and grew at their full size in 7 – 10 days.

**Data Collection and Processing**

Data were collected using A MAR CCD 225 detector at the SER-CAT beam line BM22 at the Advanced Photon Source, Argonne National Laboratory, with the kind help of the scientists team, especially Dr. Dauter Zbigniew. The crystal-to-detector distance was 200 mm. The crystals were soaked in 35% glycerol solution and flash cooled at 100 K.

All diffraction data frames were collected using a 0.5° oscillation angle with an exposure time of 5 sec per frame. The data sets was indexed and scaled with DENZO and HKL-2000 software (Otwinowski and Minor 1997).

**Refinement and Structure Analysis**

Cross-rotation and translational searches were performed using the program AmoRe (Navaza 2001) as implemented in the CCP4 suite (Collaborative Computational Project 1994), and rigid body refinement using the CNS package (Brünger et al. 1998). Positional and
temperature factors refinement steps were performed in CNS up to 1.6 Å resolution and in SHELX (Sheldrick 1997) when data were above 1.6 Å resolution. Electron density maps with 2Fo-Fc and Fo-Fc coefficients were used to guide manual fitting of the PR and the bound inhibitor, when applicable. Interactive manual model building, using the molecular graphics program O, version 7 (Jones et al. 1991) allowed for a further improvement of the structures. The quality of the final refined structure was validated with the PROCHECK (Laskowski et al. 1993).

In Vitro Gag/pol Processing

Amplification and Cloning of the Gag/Pol DNA Sequence into the TNT Vector

The subtype C gag/pol sequence was amplified using as template the subtype C p94IN476.104 clone ordered from NIH using primers to either ends of the gag/pol gene sequence: forward primer: 5’-AGGCTAGAAGGCTCGAGATGGGTGCGA-3’, and reverse primer: 5’-TGGTGTTGTACTACGCGTTTACTAGTTCTGATCCTC-3’

The primers were engineered to insert the restriction site XhoI at the 5’ end and MluI at the 3’ end of the sequence for directional cloning into the expression vector TNT. The PCR amplification step was performed as described above.

The PCR product was resolved on a 1% agarose gel, cut out and purified using a QIAquick Gel Extraction Kit (Qiagen) (Figure 3-8A). The PCR product was then double digested with the restriction enzymes XhoI and MluI. The TNT vector was subjected to the same procedure. Digestions were done using 5-10 μg of DNA, 0.3 μL of XhoI (20 U/μL) and/or 0.3 μL of MluI (10 U/μL), 1 μL of NEB 3, 0.3 μL of BSA, and ddH₂O to total volume of 40 μL at 37 °C for 1 h. The vector was further treated with Intestinal Calf Alkaline Phosphatase (1 U/μL) (Novagen), 10X buffer and ddH₂O to total volume of 20 μL and incubated at 37 °C for 1.5 hours. The insert gene and plasmid DNA (3:1 weight ratio) were ligated together in a reaction mixture containing
1 μL of T-4 ligase (4 Weiss units) (Novagen), 1 μL ligase buffer (10X Buffer) in a total volume of 10 μL, at room temperature, for 2 h. Two microliters of the ligation product were transformed into Top10 cells, as it has been described previously. The DNA extracted from these cells was checked by digestion with KpnI or XbaI and yielded the expected digestion products: 3 fragments for KpnI (6200 bp, 900 bp and 200 bp), and 3 fragments for XbaI (5600 bp, 1100 bp, 300 bp) (Figure 3-8B). The presence of an insert in the TNT vector was also verified by double digestion with XhoI and MluI. Two micrograms of DNA were sent for DNA sequencing, performed by the UF-ICBR sequencing core. The identity of the gag/pol sequence was verified using T7-promoter and T7-terminator primers as well as primers complementary to the gag/pol DNA sequence. The purified DNA plasmid was concentrated down to 1 μg/μL and directly used for in vitro transcription/translation reactions.

**Site-Directed Mutagenesis to Introduce the Desired Mutations**

A continuous gag/pol open reading frame was created by site-directed mutagenesis to reproduce exactly the amino acid sequences of the gag/pol product proteins found in virions. For the same purpose, another frameshift was introduced in the HIV-1 subtype C gag/pol fusion protein within RT gene. These mutations were engineered using Quick-Change Site Directed Mutagenesis Kit (Stratagene). Mutations are generated using two primers that are complementary to both the coding and non-coding strand surrounding the base, or bases, to be mutated. The reactions were initiated by mixing together 50-100 ng of plasmid DNA, 5 μL of 10X Pfu polymerase buffer, 800 ng of dNTPs, 250 ng of upper and lower primer, 1 μL of Pfu polymerase (2.5 U), and ddH2O to a final volume of 50 μL. The PCR amplification cycle was started by heating the reaction to 95 °C for 30 sec. The reaction was then cycled 18 times through a melting step at 95 °C for 30 sec, an annealing step at 55 °C for 1 min, and an extension
step at 68 °C for 8 min. After the last cycle the temperature of the reaction was dropped to 4 °C. To remove the template, 1 μL of restriction enzyme DpnI (10 U/μL) was added to all PCR reactions at the end of the cycle and placed at 37°C for 1 hour. Two microliters of the PCR reaction were used to transform chemically competent Top10 cells (Invitrogen).

**In vitro Transcription-Translation Experiment**

The gag/pol DNA sequences used for *in trans* transcription-translation experiments did not harbor the continuous open reading frame. Two microliters of DNA sample (1 μg/μL) were mixed with a solution containing 40 μL TNT T7 Quick Master Mix (contains rabbit reticulocite lysate and a mixture of all the amino acids except methionine), 2 μL 35S-Met (1000 Ci/mmol at 10 mCi/mL) and nuclease-free H2O up to 50 μL final volume, and the reaction mixture was incubated at 37 °C. After 2 h, 10 – 50 nM of the active HIV-1 PR was added in the reaction mixture and two to five microliter samples were taken at different time points and quenched with 2x Laemmlil buffer at 1:1 ratio and heated to 70 °C for 2 min.

**Separation of the Translation Products on an SDS-PAGE**

The samples were spun down for 30 – 60 sec and loaded onto a 10-20% Tris-HCl SDS-PAGE gel (BioRad) that was used for separation of the gag/pol products with different MW. The samples were run at 30 mA until the front of the loading dye reached the bottom of the gel.

**Autoradiography and Densitometric Analysis**

The gels were fixed in a solution containing 10% acetic acid and 5% hydrochloric acid for 30 min, and then soaked in 10% glycerol solution for 5 min. The gels were then placed on a sheet of absorbant filter paper (Biorad), dried, and exposed to either a XAR-5 film (Kodak) or a phospho-screen, at room temperature, for 12 – 24 h. The amount of labeled proteins was
quantified by scanning the screens using a Molecular Dynamics PhosphorImager, Storm 860 model.
Figure 3-2. Cloning the protease gene into the expression vector pET23a. A) HIV-1 subtype C PR gene amplified using a PCR protocol with three different annealing temperatures: 45 °C (lane 2), 51 °C (lane 3), and 55 °C (lane 4). The marker in lane 1 is 1 kb DNA ladder (BioRad). B) Double digestion with NdeI and BamHI of the pET23a vector containing the 300 bp insert (HIV-1 subtype C PR gene).
Figure 3-3. HIV-1 protease expression in BL21 DE3(Star) pLysS cells and inclusion bodies (IBs) extraction. The gel is Tris-glycine 15% polyacrylamide. The lanes are labeled as follows: 1-ladder (Broad, BioRad), 2-empty, 3-immediately before induction with IPTG, 4-one hour after induction, 5-two hours after induction, 6-three hours after induction, 7-inclusion body sample.
Figure 3-4. HIV-1 protease purification through size-exclusion column Superdex 75 (Amersham). A) Plot of the enzymatic rate for the fractions 37 – 48. B) The gel is Tris-glycine 18% polyacrylamide. The lanes are labeled as follows: 1-ladder (Precision Plus Protein Kaleidoscope, BioRad), 2 – 9 lanes: fractions 37 to 44 eluted from the column.
Figure 3-5. Kinetic constants determination. Michaelis-Menten curve fit of rate (µmol/min/mg) versus substrate concentration (µM) to determine $K_m$ and $V_{max}$. 

![Kinetic constants determination graph](image-url)
Figure 3-6. Inhibitor dissociation constant (Kᵢ) determination. Michaelis-Menten curve fit of rate (µmol/min/mg) versus substrate (µM) with increasing inhibitor concentration [I] in nM (● = 0 nM; ○ = 20 nM; ▼ = 40 nM).
Figure 3-7. Enzyme titration curve. A) Linear extrapolation to the x-axis gives \([I] = [E]\); B) Dixon plot.
Figure 3-8. Cloning of $gag/pol$ sequence into the TNT vector. A) Gag/pol sequence amplified by PCR and purified by gel extraction procedure (Qiagen). B) TNT vector containing the gag/pol insert cut with KpnI (lane 1), cut by XbaI (lane 2) and uncut (lane 3).
CHAPTER 4
KINETIC STUDIES ON THE CONTRIBUTION OF NATURALLY OCCURRING POLYMORPHISMS IN ALTERING THE BIOCHEMICAL CHARACTERISTICS OF HIV-1 SUBTYPE C PROTEASE

Introduction

There is extensive and growing literature on sequence data from untreated and treated persons infected with HIV-1 subtype B virus (Kantor and Katzenstein 2004; Shafer et al. 2007). This has led to increasingly accurate, though complex, interpretations of HIV-1 subtype B drug resistance (Hirsch et al. 2000; Vergne et al. 2006; Shafer et al. 2007). Such data are generally not available for non-B subtypes. More information needs to be gathered and analyzed in order to fully understand the important role that differences in protease (PR) sequence of non-B subtypes, including subtype C play in the interaction of enzyme with the substrates and the inhibitors (Figure 4-1).

In this chapter, we investigate the role that the eight polymorphic residues have in altering the enzymatic and structural features of subtype C PR as compared to subtype B PR. These residues are located in the elbow of the flaps (M36I, and S37A), at the base of the enzyme (H69K, and I93L), and within the PR core (T12S, I15V, L19I, and L89M). Our goal is to analyze whether these baseline polymorphisms engender a decrease in the affinity of the PR for the clinically used inhibitors, or confer an increased catalytic activity and thus possibly a greater biochemical fitness in the presence of the inhibitors.

We also engineered combinations of three major mutations (L90M, D30N, and N88D) that arise mainly in response to nelfinavir (NFV) treatment. NFV is a PI preferably used, especially in developing countries because of relatively lower cost and fewer side effects. Also, its relatively safe biochemical and pharmacodynamic profiles make NFV the first choice to treat HIV-positive pregnant women. Recent studies have shown that the most common primary
mutation observed in treated patients infected with HIV-1 subtype C was L90M (Cane et al. 2001), a major mutation that confers multi-PI resistance especially to saquinavir (SQV), ritonavir (RTV) and NFV (Condra et al. 1996; Hertogs et al. 2000; Kempf et al. 2001). The leucine (Leu) at the position 90 is not located in the active site but the side chains lie on either side of the catalytic Asp25, so the catalytic activity may be more sensitive to substitutions at Leu90 (Mahalingam et al. 2001). Its position in the core of the PR might affect the stability of the enzyme as shown by Xie et al. The L90M mutant seems to be more sensitive to urea denaturation than other mutants. Sedimentation equilibrium studies have also shown that L90M has reduced dimer stability at pH 7.0 relative to the wild-type PR (Xie et al. 1999).

D30N is a primary mutation that renders the HIV-1 PR less susceptible to NFV, and does not exhibit cross-resistance with other PIs (Patick et al. 1998). The crystal structure of the HIV-1 PR complexed with a peptide containing the wild type sequence of CA/p2 cleavage site from P5 to P5’, Lys-Ala-Arg-Val-Leu*Ala-Glu-Ala-Met-Ser, shows that the Asp30 residue side chain is involved in hydrogen bonding to the GluP2’ side chain (Prabu-Jeyabalan et al. 2000). This is the only direct hydrogen bonding interaction seen between the enzyme and a side chain of the substrate. The D30N mutation of Asp to an Asn in the PR likely results in a weaker hydrogen bond that destabilizes NFV binding (Kolli et al. 2006). Mutations at position 88 (N88D and N88S) commonly occur in patients receiving NFV and occasionally in patients receiving indinavir (IDV). By itself, a mutation at this position causes low-level resistance to NFV. However, when associated with D30N or M46I, the D88N mutation causes high-level NFV resistance (Petropoulos et al. 2000; Ziermann et al. 2000). These three mutations are considered major mutations against NFV and three patterns of mutational associations were identified. First, D30N was positively associated with N88D but negatively associated with N88S. Second, D30N
and L90M were negatively associated except in the presence of N88D, which facilitated the co-occurrence of D30N and L90M. Third, D30N + N88D + L90M formed a stable genetic backbone for the accumulation of additional PI resistance mutations. In 16 patients having isolates with more than one combination of mutations at positions 30, 88, and 90, all exhibited one of the steps in the following progression: D30N → D30N + N88D → D30N + N88D + L90M → D30N + N88D + L90M + (L33F +/- I84V or M46I/L +/- I54V) (Mitsuya et al. 2006).

In this chapter, we analyze whether the preexistent baseline polymorphisms, by themselves or in combination with drug resistance mutations, differentially alter the biochemical features of the subtype C PR compared with that of subtype B PR.

Our goal is to understand the effects of secondary mutations in conjunction with published data and to evaluate mutations/polymorphisms as a predictive measure for therapy development and effectiveness. We performed kinetic studies for each variant. Michaelis-Menten constants and \( K_i \) values provide biochemical information about the differences in the binding affinity of different mutants to the substrates and clinically used inhibitors, and could represent a guide for further \textit{in vitro} and \textit{in vivo} studies, pinpointing which inhibitor works best for a certain mutation or combination of mutations.

\textbf{Results}

\textbf{Analysis of HIV-1 Subtype C Protease Sequence}

The HIV-1 subtype C PR sequence analyzed in our study comes from an HIV-positive patient from India, and the clone was provided by the NIH Research and Reference Reagent Resources Program (Rodenburg et al. 2001). This clone contains eight amino acid differences (T12S, I15V, L19I, M36I, S37A, H69K, L89M, and I93L) from the subtype B PR (LAI sequence) (Figure 4-2A). Two of these polymorphisms are located in the elbow of the flaps, three are located in the 10’s loop, and three are situated within the loops at the base of the PR.
All these regions harboring the subtype C PR polymorphic differences are located in the outside regions of the protease avoiding the active site, the flaps and the dimerization region (Figure 4-2B).

The initial clone contained one amino acid difference, N88D, considered a major mutation for NFV resistance. In our study, this enzyme is designated HIV-1 subtype C N88D variant. Subtype C PR analyzed here was obtained by back-mutating the aspartate (D) at the position 88 to the wild-type residue, asparagine (N).

**Kinetic Analysis of HIV-1 Subtype B and C Proteases**

The kinetic constants were measured for the HIV-1 subtype B and C PRs using a synthetic substrate: Lys-Ala-Arg-Val-Nle-nPhe-Glu-Ala-Nle-NH$_2$. The P1’ position was replaced with nitrophenylalanine to permit spectrophotometric monitoring of the substrate cleavage. The P4’ position was replaced with norleucine for stability.

This peptide resembles the cleavage sites between the capsid protein and p2 (KARVL/AEAM). This site is well conserved among different HIV-1 subtypes and it is essentially identical for subtype B and C gag polyproteins. Previously published reports noted that the sequence of the capsid/p2 cleavage site is one of the best HIV-1 processing sites when assayed as a peptide at low pH (Richards et al. 1990; Tozser et al. 1991). Several studies showed that CA/p2 site appears to be important in regulating the formation of the viral core being the rate-limiting step during gag and gag/pol processing (Pettit et al. 1994; Wiegers et al. 1998).

- **Michaelis-Menten constants ($K_m$, $k_{cat}$)**

  As seen in Table 4-1, $K_m$ for the subtype C PR is similar to that of the B subtype PR, with values of 19.5 μM and 17 μM, respectively. However, the situation is not the same with $k_{cat}$ values. Subtype C PR showed a 2-fold decrease in $k_{cat}$ (5.6 sec$^{-1}$) when compared to subtype B.
PR (10 sec⁻¹) (Figure 4-3A). Accordingly, the catalytic efficiency, defined as $k_{\text{cat}}/K_m$, of the C subtype PR is about 2 times lower than that of the B subtype PR.

- **Inhibition constants (Kᵢ)**

  The Kᵢ values for subtype B and C PRs were calculated for eight protease inhibitors (PIs) used in clinical settings: SQV, RTV, IDV, NFV, amprenavir (APV), lopinavir (LPV), atazanavir (ATV), tipranavir (TPV). The results obtained for all of the inhibitors are summarized in Table 4-2. There are three PIs: SQV, IDV and NFV that have slightly higher Kᵢ values than the other five inhibitors (Figure 4-3B). Even if the Kᵢ values are comparable between subtype B and subtype C PRs, the naturally occurring polymorphisms within subtype C PR seem to have differential effect for different inhibitors. They increase the Kᵢ value in subtype C protease for IDV by 2-fold when compared to that of subtype B, and they have similar effect on NFV binding but to a lesser extent, with only 1.6-fold increase in Kᵢ value. On the other hand, the influence on the change in the Kᵢ values is in the opposite direction for RTV and TPV with a decrease in Kᵢ values for subtype C PR by 2.6- and 3.6- fold, respectively, when compared to those of subtype B PR (Figure 4-3B).

  However, even with the differences described above, the Kᵢ values for these enzymes with all of the inhibitors tested were within sub-nanomolar to low nanomolar range, in agreement with other studies analyzing both subtype B and C PRs (Muzammil et al. 2003; Velazquez-Campoy et al. 2003).

**Kinetic Analysis of HIV-1 Subtype B and C Proteases Harboring Drug-Resistant Mutations**

The mutations we analyzed in this study were introduced within both subtype B and subtype C PR sequences through site-directed mutagenesis technique. We engineered single,
double and triple mutants, harboring various combinations of D30N, N88D, and L90M. There were a total of 14 mutants we constructed and analyzed.

**Kinetic analysis of the single mutants of subtype B and C proteases**

- **Michaelis-Menten constants (Kₘ, kₜₐₜ)**

  The addition of D30N mutation increased the Kₘ by to 2- and 3-fold for subtype B and subtype C PRs, respectively. However, the effect on the kₜₐₜ value was divergent with a 2-fold decrease for subtype B PR and 2-fold increase for subtype C PR (Figure 4-4A). However, despite these varied effects on Kₘ and kₜₐₜ values, the enzymatic catalytic efficiency (kₜₐₜ/Kₘ) showed similar calculated values for both D30N subtype B and C mutants: 0.13 and 0.18 sec⁻¹ μM⁻¹, respectively (Table 4-1).

  The addition of the N88D mutation resulted in relatively unchanged Kₘ and kₜₐₜ values for subtype B PR with a 1.4-fold increase in the catalytic efficiency of the mutant. The HIV-1 subtype C N88D PR variant showed a slight increase in Kₘ and an approximate 2-fold decrease in the enzymatic catalytic efficiency. The variants harboring L90M mutation exhibited similar changes in their kinetic parameters when compared with the N88D variants.

- **Inhibition constants (Kᵢ)**

  As expected, the most significant effect of D30N mutation was on NFV binding for both subtype B and C PRs, with a fold-increase of Kᵢ values of 26 and 89, respectively (Table 4-2). The other PI significantly affected by D30N is ATV with a 12-fold increase of Kᵢ value for subtype C PR and only 3-fold increase for subtype B variant. However, both Kᵢ values remain in subnanomolar range, indicating tight binding inhibition. HIV-1 subtype C D30N variant also exhibited a 9-fold increase in Kᵢ value against RTV, while the same mutant of subtype B PR showed only 2-fold increase.
The subtype B and C N88D mutants exhibited a 2-fold and 4-fold increase of \( K_i \) values when tested with NFV (Figure 4-4B). The N88D mutation is not selected by other clinically used inhibitors but decreases the binding affinity for SQV (3-fold) and IDV (6-fold and 13-fold) for subtype B and C PRs, respectively. The L90M mutation alone introduces, in both subtypes, slight increases (between 1.5- and 3.5-fold) in \( K_i \) values for almost all clinically used inhibitors tested in this study, except RTV where there is a 6-fold and 11-fold increase in \( K_i \) for subtype B and C PRs, respectively.

The newer PIs APV, LPV, and TPV seem unaffected by any single mutations introduced within subtype B and C PRs backbone.

**Kinetic analysis of the double mutants of subtype B and C proteases**

- **Michaelis-Menten constants \((K_m, k_{cat})\)**

  The subtype C PR double mutants D30N/N88D and D30N/L90M exhibited an approximate 2-fold decrease in their catalytic activity. On the other hand, the same double mutants of subtype B PR showed more significant changes, retaining only about 10% of the enzymatic activity of the wild type enzyme.

  The double mutant N88D/L90M of subtype B PR showed slight changes in the \( K_m \) and \( k_{cat} \) values: 24 \( \mu \text{M} \) and 13 sec\(^{-1}\), with a catalytic activity similar to that of the wild type subtype B PR. The introduction of the same combination of mutations in HIV-1 subtype B PR engendered a 2-fold and 4-fold increase in \( K_m \) and \( k_{cat} \) values respectively, resulting in doubling the variant catalytic activity when compared to the wild type subtype C PR (Figure 4-5A).

- **Inhibition constants \((K_i)\)**

  The addition of D30N and N88D of both subtype B and C PRs resulted, as expected, in a comparable increase in \( K_i \) values for IDV, NFV, and SQV. RTV showed 11-fold increase of \( K_i \)
value for subtype B PR and almost no change for subtype C variant. APV showed a 5- and 2.5-fold decrease in $K_i$ values for subtype B and C variants, respectively (Figure 4-5B).

The subtype B PR D30N/L90M double mutant exhibited an increased in $K_i$ values for all PIs tested in this study when compared with the same variant of subtype C PR. The most significant differences were for SQV, IDV, and NFV with an increase of $K_i$ values by 7-, 6-, and 52- fold and 3-, 3-, 24- fold for subtype B and C PRs, respectively.

Upon the acquisition of N88D/L90M, there was noted an increase in $K_i$ values for all inhibitors analyzed except TPV, which showed a decrease in $K_i$ values for both subtypes, thus indicating an increased affinity for the enzymes harboring the N8D/L90M combination. The highest fold-changes were observed for SQV, RTV, IDV with a 22-, 10-, 10- fold increase for subtype B PR and a 50-, 14-, 28- fold increase for subtype C PR.

**Kinetic analysis of the triple mutants of subtype B and C proteases**

- **Michaelis-Menten constants ($K_m$, $k_{cat}$)**

  Both variants of subtype B and C PRs harboring the triple combination D30N/N88D/L90M showed a 2-fold increase in their catalytic efficiencies (Figure 4-6A).

- **Inhibition constants ($K_i$)**

  The increase in $K_i$ values is comparable between the triple mutants with the most significant differences for RTV and NFV with 14- and 52-fold increase and 31- and 83-fold increase for HIV-1 subtype B and C variants respectively (Figure 4-6B).

**Discussion**

The discovery of such a diverse panel of non-B subtypes of HIV-1 has posed the question if the currently used methods of prevention, diagnostic, and treatment are as effective as they are for HIV-1 subtype B, the subtype these methods have been originally developed to combat. A large number of laboratories and researchers and numerous resources have been allocated to
investigate the matter. One of the most researched topics is the efficacy of the treatment and the development of drug resistance to currently available PIs, which have been designed based on biochemical and structural information on subtype B PR. The main concern is that the preexisting naturally occurring polymorphisms within non-B subtypes PRs might increase the speed of which highly resistant viruses are selected, promoting a poorer response to the clinically used PIs.

The aim of this chapter is to analyze the contributions to the catalytic efficiency and inhibitor resistance by naturally occurring polymorphisms and therapy-selected, active and non-active site, residue changes in subtype C compared to subtype B PRs. As mentioned before, even though the baseline polymorphisms found in subtype C PR have arisen in the absence of PI therapy, residue changes in C PR at positions 36, 69, and 89 have been associated with PI resistance in vivo and in vitro in B PRs (Gong et al. 2000; Rusconi et al. 2000; Clemente et al. 2003; Holguin et al. 2004). Thus, these polymorphisms can potentially influence substrate processing and the binding of currently available PIs and/or facilitate the development of resistance.

In the PR sequence analyzed in this study all naturally occurring polymorphism changes represent conservative amino acid changes, except at position 37 where a hydrophilic amino acid, serine, is replaced by a small hydrophobic residue, alanine. Two polymorphic residues (M36I and S37A) are located in the flap making contact with the solvent molecules. More precisely, M36I lies in the interface between the enzyme core and the flap and develops as a secondary drug resistance mutation in subtype B PR, decreasing the enzyme affinity for several clinically-used PIs (Clemente et al. 2003; de Mendoza et al. 2007). The S37A is rarely encountered in clinical isolates (Stanford 2007) and has unknown role if any in the development
of drug resistance but in our *in vitro* experiments this mutation seems to decrease the enzyme solubility in polar solvents and to promote precipitation. Three other polymorphisms are located in the 10’s loop: T12S, I15V, and L19I. This combination is highly specific for subtype C PR (Stanford 2007). The role of these polymorphisms was not fully explored but the I15V mutation has been described to arise in association with TPV treatment. However, there is no clear link between the occurrence of this polymorphism and treatment failure to TPV (Rusconi et al. 2000).

H69K is located in the loop at the base of the PR, and together with M36I was documented to predict decreased phenotypic susceptibility or diminished antiviral responses to TPV (Baxter et al. 2006; de Mendoza et al. 2007). L89M is located at the base of the PR and Sanches et al. hypothesized that this polymorphism may lead to early development of drug resistance in patients infected with non-B HIV subtypes (Sanches et al. 2007). Also, recent studies inferred that the L89M mutation in subtype F viruses is a high genetic barrier to the accumulation of the L90M resistance mutation and can function as a resistance mutation, depending on the presence of other polymorphisms in the subtype F PR backbone (Calazans et al. 2005).

The I93L polymorphism is located within a hydrogen-bonded turn immediately upstream from the PR/RT cleavage site, in close proximity L89M and H69K and the dimerization domain. It seems to play a dual role, depending on the ARV regimen administered to the HIV-positive patients. A higher relative risk for developing treatment failure was observed with the presence of I93L in a subgroup of patients treated with IDV, RTV or SQV in several studies (Harrigan et al. 1999; Servais et al. 2001). Also, although the I93L substitution has been described as a naturally occurring polymorphism in subtype B PR as well, the frequency of this substitution appears to be substantially higher in the treated population than in protease inhibitor–naive patients (40% vs. 20%) (Kozal et al. 1996). On the other hand, I93L seems to increase the
susceptibility of subtype C PR to LPV as observed by Gonzalez et al. (Gonzalez et al. 2003). In this study, the authors noticed that the subtype C PRs harboring I93L polymorphism presented significant hypersusceptibility to LPV.

The results of all these studies, and of many others not mentioned here, demonstrate the need for further investigating the role of naturally occurring polymorphisms in the development of drug resistance and their contribution in evading drug inhibition and in preservation of the catalytic activity of the PR upon accumulation of major drug resistance mutations.

Our kinetic data on the wild type subtype B and C PRs showed that there are no significant differences in the substrate affinity constant between these two enzymes, both having $K_m$ values within 17 – 19 μM range. This lack of a significant effect in $K_m$ probably reflects the fact that the sequence variations found in subtype C PR are located outside the active site and that the chemical environment surrounding the aspartyl diad remains unchanged. Also, even if the naturally occurring polymorphism, such as L89M and I93L, located in the hydrophobic core of the enzyme would change the shape of the substrate-binding cleft, the flexibility of the substrate seems to overcome this effect.

From our data it is also apparent that the naturally occurring polymorphisms in subtype C PR by themselves do not provide an advantage for substrate catalysis when compared to subtype B PR, with the catalytic enzymatic efficiency of subtype C PR being 60% of that of subtype B PR. These results come in contrast to the data reported earlier for subtype C PR. Velazquez-Campoy et al. showed that the subtype C and A PRs exhibit catalytic advantage over B PR, with the C subtype PR displaying lower $K_m$ values against two different substrates and resulting in a higher (2.4-fold) catalytic efficiency than the subtype B PR (Velazquez-Campoy et al. 2001). These different results can be explained by the differences in the amino acid sequences between
the subtype C PRs analyzed in these studies. The enzyme analyzed in our studies harbors five additional polymorphisms: T12S, I15V, L19I, S37A, and I93L, whose overall effects on the biochemical characteristic of the HIV PR have not been studied in detail.

The inhibition kinetic studies conducted with eight clinically used PIs showed that the $K_i$ values for all the inhibitors tested on subtype B and C wild type PRs are comparable between these two enzymes and have low nanomolar and sub-nanomolar ranges, indicating high affinity of currently used PIs for these enzymes. These data are in agreement with previous clinical studies that report that, at least at the initiation at ARV treatment, drug naïve patients infected with subtype C viruses respond to ARV treatment as well as HIV-1 subtype B infected patients (Alexander et al. 2002; Frater et al. 2002; Pillay et al. 2002; Weidle et al. 2002; Bocket et al. 2005; Wester et al. 2005). However, several studies showed that differences in response to ARV treatment, measured as viral load or CD4 cells count, seem to arise after 1-2 years of treatment, with non-B subtypes infected patients performing worse than subtype B infected patients (Caride et al. 2001; Loveday et al. 2001; De Wit et al. 2004; Atlas et al. 2005). But other factors, such as adherence to treatment, ethnicity, psychosocial conditions, should be taken into consideration as well when analyzing these differences. However, up to date there are no long-term clinical studies (more than 5 years) to exploring the efficacy of the existing PIs in patients infected with HIV-1 non-B subtypes, including subtype C (Frater et al. 2002). Due to the severity of this disease and its life-long implications, together with the use of antiretroviral therapy becoming more widespread across Africa, it is imperative to characterize baseline molecular variability and subtype-specific peculiarities of drug targets in non-subtype B HIV-1 infection.

We hypothesize that the effects of naturally occurring polymorphisms on the enzymatic catalytic activity and inhibitor binding affinity of subtype C PR are augmented by the acquisition
of major drug resistance mutations. As a first step to explore this possibility we engineered a series of single mutants containing one PI-induced mutation: D30N, N88D or L90M. For most PIs analyzed in this study, when comparing the effects of various combinations of mutations between subtype B and C PRs, we observed that the trend in the changes in $K_i$ values are similar, exhibiting either concurrent increases or decreases against the clinically used PIs analyzed in this study. The difference consists in the magnitude of the change that varies between subtype B and C PRs. The most significant fold-increase in $K_i$ values was noted for NFV when tested against PRs harboring D30N mutation. Subtype B variants showed a 26-fold increase when compared to the wild type, while D30N subtype C PR exhibited 89-fold increase, and thus a significant decrease in binding affinity towards NFV. The catalytic efficiencies for both variants were similar when compared as absolute values. However, when compared to the wild type enzyme, D30N subtype B PR preserved only 25% of the catalytic efficiency of the wild type, while D30N subtype C PR maintained 50% catalytic efficiency. These results do not bring a biochemical explanation to the previous observation that HIV-1 subtype C viruses preferentially select L90M over D30N when exposed to NFV treatment (Grossman et al. 2004). However, other clinical studies also could not corroborate with Grossman et al. findings, as no differences were observed in the frequency of development of resistance mutations L90M and D30N in B and non-B viruses, including subtype C, upon NFV treatment (Pillay et al. 2002; Doualla-Bell et al. 2006; Stanford 2007).

Subtype C PR harboring D30N also exhibited larger decrease in the susceptibility to PIs, such as RTV and ATV inhibition when compared to similar mutant of subtype B PR. Although this might predict a predisposition to developing resistance to these two inhibitors by the subtype
C variants, it does not seem to be a clear marker of resistance, because both PIs maintained a binding strength sufficient for effective inhibition of the PR.

Both single mutants, N88D and L90M, of subtype B PR showed a slight decrease in the $K_m$ values and a preservation of the enzymatic catalytic efficiency when compared to the wild type subtype B PR. The similar mutants of the subtype C PR exhibited a slight decrease in substrate affinity and in enzymatic catalytic efficiency. Both mutations are located on each side of the 89 position occupied by a Leu in subtype B PR and a Met in subtype C PR. The differences in enzyme activity could be explained by slight distortions that a larger 89M could induce in the subtype C PR core. The same hypothesis was recently advanced by Sanches et al. after structural characterization of subtype F PR which also contains the L89M polymorphisms (Sanches et al. 2007).

Upon acquisition of N88D, the only significant change in binding affinity was observed for IDV, with a higher increase in $K_i$ values for subtype C PR (13-fold increase in $K_i$ value, compared to 6-fold increase for subtype B PR mutant). Our study demonstrates that N88D by itself does not provide higher level of resistance for NFV in both subtype B and C PRs. L90M mutants exhibited slight decrease in binding affinity for SQV, IDV and RTV, for both subtype B and C PRs.

So far, our data on single mutants for subtype B and C PRs confirmed some of the previous studies on subtype B PR and highlighted several differences in the kinetic profile between these two enzymes. These differences are minor for newer PIs such as APV, LPV, ATV, and TPV and of greater amplitude for SQV, RTV, IDV, and NFV.

The variants harboring D30N/N88D combination of mutations exhibited similar kinetic behavior for both subtype B and C PRs. The subtype B mutant maintained 70% of the catalytic
efficiency when compared with the wild type, much better than the mutant harboring D30N alone. Previous studies showed that there is a strong association between D30N and N88D in subtype B PR, with the latter probable having the role to stabilize the enzyme by compensating for the negative charge lost with D30N substitution (Roge et al. 2003; Mitsuya et al. 2006). The D30N/N88D subtype C PR preserved 55% of the wild type enzymatic activity. The $K_i$ value profile is comparable between these two mutants with, as expected, a significant decrease in binding affinity for NFV.

For the D30N/L90M double mutants, we noticed a more notable increase in $K_i$ value for NFV for subtype B PR (52-fold), compared to subtype C mutant (24-fold). However, we also observed a significant decrease of almost 10-fold in the catalytic efficiency of the D30N/L90M subtype B mutant. On the other hand, upon acquisition of this combination of mutation, subtype C PR exhibited only a 50% decrease in its enzymatic activity. Our kinetic results appear to be in accordance with previous studies that showed that subtype B mutants carrying D30N/L90M combination are only rarely found in vivo, and generally have additional, potentially compensatory, mutations. Interestingly, a major loss of replicative capacity for a mutant clone harboring both D30N and L90M was also described (Martinez-Picado et al. 1999; Sugiura et al. 2002; Perrin and Mammano 2003). It can be inferred that in the context of this combination of major resistance mutations, the naturally occurring polymorphisms within subtype C PR could act as compensatory mutations having role in preserving the catalytic efficiency of the enzyme.

N88D/L90M combines two non-active site mutants: N88D, which has high stability with L90M, which shows poor structural stability (Xie et al. 1999; Mahalingam et al. 2002). Both N88D and L90M are not part of the substrate-binding site, but they can influence the catalytic process through indirect interactions. This combination of mutations appeared to be very
favorable in the context of the naturally occurring polymorphisms of subtype C PR when compared to subtype B PR. The subtype C mutant exhibited increased catalytic efficiency when compared with both the wild type subtype C PR and the similar variant of subtype B PR. The inhibition profile showed higher $K_i$ values for subtype C PR when compared to subtype B PR mutant for all PIs tested in this study. Not surprisingly the fold-increase in $K_i$ value for NFV was modest, but the $K_i$ values for SQV and IDV showed 50- and 28- fold increase, respectively for N88D/L90M subtype C PR and 22- and 10- fold increase, respectively for subtype B variant.

Surprisingly, the mutants harboring the triple combination D30N/N88D/L90M showed improved catalytic efficiencies for both subtypes and increased $K_i$ values for NFV, SQV, and RTV. The fold-change was higher for NFV in subtype C PR (83-fold for NFV versus 52-fold increase in subtype B PR).

Our data also demonstrated that the D30N mutation in combination with D88N and L90M provides a higher level of cross-resistance for subtype C PR, specifically to RTV. This is of interest as no such combination was described in response to RTV in subtype B PR. This result could argue for a new mutational pathway in developing resistance to RTV in subtype C PR. However, despite the high fold-increase, the absolute $K_i$ value remains low, indicating that RTV maintains sufficient binding strength to effectively inhibit the D30N/N88D/L90M mutant of subtype C PR. For these reasons the use of pretherapy and therapy resistance testing, especially in cases of early failure, is crucial for providing optimum treatment potency.

An interesting observation is that the kinetic parameters of subtype C PR harboring the D30N mutation do not improve in the same way as happens for subtype B PR upon acquisition of other drug resistance mutations. Specifically, D30N subtype C PR does not improve its catalytic efficiency upon acquisition of N88D, as happens for subtype B PR. Also, the $K_i$ values
for NFV for D30N subtype C PR do not change after combination with N88D and L90M. A possible interpretation of these results is that the naturally occurring polymorphisms in subtype C PR have already set a favorable context in which the acquisition of D30N by itself is sufficient enough to attain a higher level of resistance, while in subtype B PR other major mutation are required to develop a similar degree of resistance. We have to mention that this effect can be limited to the combination of baseline polymorphisms in subtype C PR and the type of major mutations induced by the ARV treatment. We do not attempt to make the same assumption for other subtype PRs, due to the fact that they harbor different combinations of polymorphisms that might overall not have the same effect on the biochemical and structural characteristics of the PR. This is another reason why the exploration of each subtype PR is important in the context of designing the best ARV treatment strategy, especially for patients who have failed multiple drug regimens.

In our study the binding affinities of the newer PIs, APV, LPV, ATV, TPV, seem less affected by the introduction of three mutations studied here. We also have noticed that most of the subtype C PR mutants analyzed in this study appear to exhibit a modest increase in the susceptibility to TPV, as shown by the lower $K_i$ values for this PI. It might be due to the fact the newly designed PIs are able to evade drug resistance mutations better that the older PIs. Also, the mutations we studied are quite specific for these PIs. On the other hand, we have not studied mutations that specifically arise upon treatment with these inhibitors, such as I50L for ATV or I84 V for LPV but such studies are under way in our laboratory (data not shown). Also, the viruses isolated from patients that failed ARV treatment usually have PRs harboring a much larger number of mutations. So, in the long term, the minor effects introduced by the baseline polymorphisms, exemplified in our studies by modest changes in $K_i$ values, could be even
further accentuated by the multiplicative effect of addition of multiple drug resistance mutations (Wu et al. 2003; Clemente et al. 2006). We expect to see even larger differences in $K_i$ values between subtype B and C PRs upon acquisition of a larger number of drug resistance mutations, as usually happens during ARV therapy.

Gonzalez et al. showed that the addition of three or four IDV drug resistance mutations has led to a highly resistant subtype C virus compared with its subtype B counterpart (Gonzalez et al., 2003). The same phenomenon was observed with NFV. This means that the IDV drug resistance mutations impact was much more noticeable in subtype C than in subtype B viruses. Differences in behaviors between the subtype C and subtype B viruses observed in vitro in this study and in our study may have a clinical relevance, suggesting a lower genetic barrier in subtype C viruses than in subtype B isolates.

The PR clone used in our studies does not harbor a frequent polymorphism found in subtype C PRs: R41K. It is one of the residues that appears to be involved in preserving or augmenting the catalytic efficiency of the subtype C PR and enhancing the viral fitness when compared to subtype B viruses (Ziermann et al. 2000; Velazquez-Campoy et al. 2001). It also appears to decrease viral susceptibility to RTV and APV treatment. Our intention was not to study the effects of each individual polymorphic residue but their influence as a group within the subtype C PR context.

In conclusion, the differences observed in our study between various mutants of subtype B and C PRs are due to the presence of the preexisting polymorphisms. Our study showed that the presence at the start of therapy of the naturally occurring polymorphisms could give the virus an advantage in the rapid development of drug resistance while preserving the virus viability in the presence of a specific inhibitor. We also believe that the effects of the naturally occurring
polymorphisms in subtype C PR are influenced by the PI choice for treatment and by the type of major mutations acquired upon drug treatment. The same is probably true for all other HIV-1 subtypes. These mutation patterns are complex and frequently overlapping. It is necessary to test the major groups of naturally occurring polymorphisms for each subtype in combination with several major drug resistant mutations to further be able to predict the response to treatment based on the combination of baseline polymorphisms. Also, determining the biochemical and biophysical properties of enzymes with these patterns of mutations will be important for designing new PIs that are less likely to trigger resistance or are effective against already drug-resistant isolates.

It is important to mention that studying the addition of PI-resistance mutations within the PR is limited only to their effect on the biochemical characteristics of the enzyme. The set up of our experiments prevents us from studying the effects of the naturally occurring polymorphisms within the gag or the viral context. For example, several studies showed that there is an interrelation between acquisition of D30N and N88D and changes occurring within the rest of the gag and gag/pol polyproteins. It has been shown that mutations in the HIV-1 protease substrate cleavage-site p1–p6 covary with the D30N/N88D protease mutations. Aspartyl at position 30 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 (Kolli et al. 2006).
All these studies highlight the importance of further characterization of the development of resistance in non-subtype B viruses. For some drugs, the differences seen in baseline polymorphisms between subtypes may influence which mutational patterns develop. Subtle effects of such polymorphisms on drug susceptibility and replicative capacity may underlie such changes.

Despite their limitations, our studies verify and complement data obtained through phenotypic studies. While experiments involving virus cultures might require more time and are more expensive, these biochemical studies can be done quickly and at lower cost. Also biochemical studies are performed in a controlled environment, in which differences in response can be tracked down to a single variable. The results of our studies should prove useful in the design of new and continuing therapy.
Figure 4-1. Polymorphic sites within the subtype B and C protease sequences. Accessed and adapted on June, 2007 from http://hivdb.stanford.edu/pages/barChart_MutPrevBySubtype/PR_untreated.html - Stanford Database
Figure 4-2. HIV-1 subtype C protease. A) Pair-wise sequence alignment of HIV-1 subtype B PR LAI strain (designated as B) and subtype C PR (designated as C). B) Cartoon representation of HIV-1 subtype C PR. The naturally occurring polymorphisms in subtype C PR, represented as red spheres, are superimposed onto the crystal structure of HIV-1 subtype B PR (black ribbon). Amino acid positions are as numbered.
Figure 4-3. Kinetic analysis of HIV-1 subtype B and C proteases. A) The Michaelis-Menten constants: $k_{\text{cat}}/K_m$. The dark gray and light gray bars designate subtype B PR and C PR, respectively. B) The $K_i$ values for subtype B and C wild type PRs. The blue and red bars designate subtype B PR and C PR, respectively.
Figure 4-4. Kinetic analysis of the single mutants of the subtype B and C proteases. A) The Michaelis-Menten constants: $k_{cat}/K_m$. B) The $K_i$ values for subtype B and C PRs.
Figure 4-5. Kinetic analysis of the double mutants of subtype B and C proteases. A) The Michaelis-Menten constants: $\frac{k_{cat}}{K_m}$ (sec$^{-1}$ μM$^{-1}$). B) The $K_i$ values for subtype B and C PRs.
Figure 4-6. Kinetic analysis of the triple mutants of subtype B and C proteases. A) The Michaelis-Menten constants: $k_{\text{cat}}/K_m$. B) The $K_i$ values for subtype B and C PRs.
<table>
<thead>
<tr>
<th>Enzyme/Subtype</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}$ (sec$^{-1}$)</th>
<th>$k_{cat}/K_m$ (sec$^{-1}$mM$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR/B</td>
<td>19.5 ± 2</td>
<td>10 ± 1</td>
<td>0.55 ± 0.06</td>
</tr>
<tr>
<td>PR/B-D30N</td>
<td>38.6 ± 2</td>
<td>5 ± 0.5</td>
<td>0.13 ± 0.02</td>
</tr>
<tr>
<td>PR/B-N88D</td>
<td>16 ± 1.5</td>
<td>12 ± 1.1</td>
<td>0.75 ± 0.10</td>
</tr>
<tr>
<td>PR/B-L90M</td>
<td>10 ± 1</td>
<td>5.8 ± 0.5</td>
<td>0.58 ± 0.06</td>
</tr>
<tr>
<td>PR/B-D30N/N88D</td>
<td>35 ± 3.5</td>
<td>12.4 ± 1</td>
<td>0.4 ± 0.04</td>
</tr>
<tr>
<td>PR/B-D30N/L90M</td>
<td>61 ± 6</td>
<td>3.5 ± 0.4</td>
<td>0.06 ± 0.008</td>
</tr>
<tr>
<td>PR/B-N88D/L90M</td>
<td>30 ± 2</td>
<td>10.2 ± 1</td>
<td>0.34 ± 0.04</td>
</tr>
<tr>
<td>PR/B-30/88/90</td>
<td>30 ± 2.3</td>
<td>30.6 ± 7.5</td>
<td>1.0 ± 0.3</td>
</tr>
<tr>
<td>PR/C</td>
<td>17 ± 1</td>
<td>5.6 ± 0.2</td>
<td>0.32 ± 0.02</td>
</tr>
<tr>
<td>PR/C-D30N</td>
<td>64 ± 6</td>
<td>12 ± 1</td>
<td>0.18 ± 0.02</td>
</tr>
<tr>
<td>PR/C-N88D</td>
<td>22 ± 3</td>
<td>4.3 ± 0.4</td>
<td>0.19 ± 0.03</td>
</tr>
<tr>
<td>PR/C-L90M</td>
<td>23 ± 3</td>
<td>4.8 ± 0.3</td>
<td>0.2 ± 0.04</td>
</tr>
<tr>
<td>PR/C-D30N/N88D</td>
<td>19 ± 1.6</td>
<td>3.2 ± 0.3</td>
<td>0.17 ± 0.02</td>
</tr>
<tr>
<td>PR/C-D30N/L90M</td>
<td>23 ± 3</td>
<td>4 ± 0.2</td>
<td>0.17 ± 0.02</td>
</tr>
<tr>
<td>PR/C-N88D/L90M</td>
<td>24 ± 3</td>
<td>13 ± 1.7</td>
<td>0.51 ± 0.09</td>
</tr>
<tr>
<td>PR/C-30/88/90</td>
<td>35.5 ± 3</td>
<td>22 ± 2.4</td>
<td>0.6 ± 0.1</td>
</tr>
</tbody>
</table>
Table 4-2. The $K_i$ values (µM) for HIV-1 subtype B and C proteases. The $K_i$ values are in bold font and the errors are in parenthesis.

<table>
<thead>
<tr>
<th>Subtype-Mutant</th>
<th>RTV</th>
<th>IDV</th>
<th>NFV</th>
<th>SQV</th>
<th>APV</th>
<th>LPV</th>
<th>ATV</th>
<th>TPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.07</td>
<td>1.8</td>
<td>1.7</td>
<td>2.2</td>
<td>0.4</td>
<td>0.11</td>
<td>0.07</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.1)</td>
<td>(0.4)</td>
<td>(0.3)</td>
<td>(0.1)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>B-D30N</td>
<td>0.15</td>
<td>9.4</td>
<td>44.0</td>
<td>3.8</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(1.1)</td>
<td>(5.5)</td>
<td>(0.3)</td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>B-N88D</td>
<td>0.24</td>
<td>9.9</td>
<td>3.4</td>
<td>5.7</td>
<td>0.39</td>
<td>0.08</td>
<td>0.21</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.8)</td>
<td>(0.3)</td>
<td>(0.7)</td>
<td>(0.05)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.1)</td>
</tr>
<tr>
<td>B-L90M</td>
<td>0.42</td>
<td>5.0</td>
<td>5.4</td>
<td>7.5</td>
<td>0.36</td>
<td>0.10</td>
<td>0.17</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.4)</td>
<td>(0.5)</td>
<td>(0.9)</td>
<td>(0.05)</td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>B-D30N/L90M</td>
<td>0.32</td>
<td>10 (1)</td>
<td>88 (10)</td>
<td>15 (2)</td>
<td>0.47</td>
<td>0.44</td>
<td>0.28</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.04)</td>
<td>(0.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-D30N/N88D</td>
<td>0.87</td>
<td>6.9</td>
<td>68 (6)</td>
<td>20 (2)</td>
<td>0.08</td>
<td>ND</td>
<td>0.09</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.6)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-N88D/L90M</td>
<td>0.71</td>
<td>17.7</td>
<td>12 (1)</td>
<td>49 (4)</td>
<td>0.89</td>
<td>0.15</td>
<td>0.35</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(1.3)</td>
<td>(0.11)</td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-30/88/90</td>
<td>1.0</td>
<td>5 (1)</td>
<td>88 (7)</td>
<td>47 (6)</td>
<td>0.61</td>
<td>1.0</td>
<td>0.37</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(0.07)</td>
<td>(0.2)</td>
<td>(0.05)</td>
<td>(0.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.27</td>
<td>3.3</td>
<td>2.7</td>
<td>2.6</td>
<td>0.29</td>
<td>0.19</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.3)</td>
<td>(0.3)</td>
<td>(0.2)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>C-D30N</td>
<td>0.6</td>
<td>10.4</td>
<td>151 (4)</td>
<td>9.0</td>
<td>0.6</td>
<td>0.3</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>(0.1)</td>
<td>(0.3)</td>
<td>(0.1)</td>
<td>(0.1)</td>
<td>(0.03)</td>
<td>(0.1)</td>
<td>(0.1)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>C-N88D</td>
<td>0.1</td>
<td>24 (2)</td>
<td>6.2</td>
<td>5.7</td>
<td>0.5</td>
<td>0.15</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.7)</td>
<td>(0.1)</td>
<td>(0.2)</td>
<td>(0.01)</td>
<td>(0.1)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>C-L90M</td>
<td>0.78</td>
<td>4.6</td>
<td>5.2</td>
<td>7.1</td>
<td>0.36</td>
<td>0.48</td>
<td>0.27</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.4)</td>
<td>(0.5)</td>
<td>(0.8)</td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>C-D30N/L90M</td>
<td>0.27</td>
<td>4.8</td>
<td>41.6</td>
<td>6.5</td>
<td>0.18</td>
<td>0.19</td>
<td>0.08</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.1)</td>
<td>(4.3)</td>
<td>(0.2)</td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.01)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>C-D30N/N88D</td>
<td>0.04</td>
<td>5.8</td>
<td>40.1</td>
<td>10.9</td>
<td>0.16</td>
<td>0.18</td>
<td>0.06</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.4)</td>
<td>(3.4)</td>
<td>(0.8)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>C-N88D/L90M</td>
<td>1.0</td>
<td>50 (4)</td>
<td>14 (2)</td>
<td>109</td>
<td>1.5</td>
<td>0.43</td>
<td>0.54</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(13)</td>
<td>(0.2)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>C-30/88/90</td>
<td>2.2</td>
<td>8.1</td>
<td>141</td>
<td>47 (4)</td>
<td>0.39</td>
<td>0.8</td>
<td>0.29</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>(0.3)</td>
<td>(0.1)</td>
<td>(17)</td>
<td>(0.08)</td>
<td>(0.2)</td>
<td>(0.03)</td>
<td>(0.06)</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5
STRUCTURAL ANALYSIS OF HIV-1 SUBTYPE C PROTEASE

Introduction

Extensive structural studies have been done with HIV-1 protease (PR) in an attempt to better understand the molecular mechanisms that govern the interactions between this enzyme and substrates or inhibitors. HIV-1 subtype B PR three-dimensional structure has been determined both alone and complexed with different inhibitors (Clemente et al., 2006; Clemente et al., 2004; Logsdon et al., 2004; Ringhofer et al., 1999; Rutenber et al., 1993). Crystal structures of HIV-1 PR were first reported in 1989 (Miller et al., 1989; Navia et al., 1989; Wlodawer et al., 1989) and their availability had a major role in the process of drug development (Vondrasek and Wlodawer, 2002). Crystal structures show that HIV-1 PR forms a binding site that consists of subsites S4-S4’, which span about eight residues (P4-P4’) of a peptide substrate (Figure 2-3). Many HIV-1 PR mutants have also been crystallized alone or complexed with peptido-mimetic or non-peptido-mimetic inhibitors (Clemente et al., 2003; Mahalingam et al., 2001; Mahalingam et al., 2004). Some of these mutants show structural changes consistent with the differences seen in their enzymatic activity (Mahalingam et al., 2002; Mahalingam et al., 2001; Mahalingam et al., 2004). In a recent study, Prabu-Jeyabalan et al. (Prabu-Jeyabalan et al., 2003) solved the crystal structures of an inactive PR bearing the mutations D25N and V82A in complex with three substrates and two inhibitors, (saquinavir (SQV) and ritonavir (RTV)). The study analyzes the mechanisms through which mutations in the active site can drastically reduce the affinity of the PR for the inhibitors, but at the same time, exert a smaller effect on the enzyme ability to cleave the substrate. These findings suggest that future inhibitor design may reduce the probability of the appearance of drug-resistant mutations by targeting residues that are essential
for substrate recognition. Perhaps this type of directed inhibitor-design will make it more
difficult for further drug resistance to evolve.

This wealth of information regarding interactions between HIV-1 PR and substrates or
inhibitors is available only for subtype B PR. The first crystal structure of a non-B subtype PR
(subtype F PR) was reported recently (Sanches et al., 2007; Sanches et al., 2004). In addition,
Coman et al. reported the crystallization of the HIV-1 subtype C PR complexed with indinavir
(IDV) and nelfinavir (NFV) (Coman et al., 2007).

In this chapter we analyze and compare the differences and similarities between subtype B
and C PRs. We believe that the successful crystallization of the protein-inhibitor complex is of
great significance. It will provide us with insight on how this PR interacts with clinically used
inhibitors at each individual subsite at the atomic level, especially for those residues presumably
critical for substrate and inhibitor binding. Thus, the structural information obtained from the
crystallographic analysis can not only verify the data observed in kinetic and functional studies
but also provide structural reference for substrate and inhibitor design. Furthermore, it will reveal
instructive information about the effect of baseline polymorphisms on the overall structure of
PR, and will help identifying the long-range interactions through which these non-active site
mutations can affect the binding of substrates/inhibitors in the active site.

Results

Crystallization

X-ray crystallography is one of the few methods available to study the atomic three-
dimensional structures of proteins and macromolecular complexes. The first step in
macromolecular structure determination is the production of highly purified sample (95-99%
homogenous) in high enough amounts. This is a crucial process, and often can mean the
difference between diffraction-quality crystals and amorphous precipitate. The next most
important, and the most time-consuming and patience-demanding step is the finding of the
condition(s) for crystallization of the protein under study. The overall approach is an iterative
one. An initial set of crystallization conditions are screened, the outcome is observed, and, if
crystals do not immediately form, the conditions are then modified taking into account what was
learned during the previous attempt. The most common approach is to use commercially
available crystallization kits. These sample kits contain a range of previously successful
crystallization conditions with varying precipitant concentrations as well as pH ranges.

**Unbound HIV-1 subtype C protease**

Crystals of the unbound subtype C PR appeared under the following conditions: 1 M NaCl,
30 mM citric acid, pH 5.0 with Triton-X 100 as additive (Hampton). The initial inspection of the
crystallization drop, after 12 hours following the equilibration against the precipitant solution at
room temperature, revealed fine precipitation that increased in amount over the next 48 hours. At
the beginning of the third day several diamond-shape crystals appeared and they increased their
size over the next five days (Figure 5-1A).

**Drug-bound HIV-1 subtype C protease**

We obtained crystals of the subtype C PR bound to Indinavir (IDV) or Nelfinavir (NFV). The crystallization conditions for both were: 1 M NaCl, 30 mM citric acid, pH 5.0. Crystals
suitable for X-ray diffraction studies of the drug-bound subtype C PR appeared approximately 24
hours after setting up the drop. These crystals grew as thin, rectangular plates, stacked in top of
each other, with approximate dimensions of 0.4 x 0.3 x 0.1 mm (Figure 5-1B).

**Diffraction Data Collection, Processing and Scaling**

The degree of oscillation data that should be collected from a crystal depends on the space
group of the crystal. According to Friedel’s law, 180 ° of data is the most, theoretically, that need
to be collected if no other symmetry exists within the crystal system (triclinic system). Thus,
higher symmetry space groups require less data to be collected, this being due to the occurrence of more symmetry equivalent planes due to the lattice symmetry.

Once a complete data set has been collected, the initial diffraction images can be used for space group determination, followed by processing of the complete set of diffraction images, reduction of data to only unique reflections based on the assigned space group, and scaling of the data for intensity normalization and calculation of $R_{\text{sym}}$. The $R_{\text{sym}}$ value measures how well the collected data fits the imposed space group symmetry elements. The collection of high quality data is of utmost importance for structure determination as well as refinement of the macromolecular model. The calculation of the $R_{\text{work}}$ and $R_{\text{free}}$ values (two important determinants of crystal structure quality) will all be dependent upon the initial measurements of the intensities.

**Unbound HIV-1 subtype C protease**

A total of 180° of data were collected using a 0.5° oscillation angle. This resulted in a total 360 images and a total of 279351 reflections measured. A complete data set was collected to 1.2 Å resolution.

Initially, the unbound form of the subtype C PR crystals were believed to belong to the tetragonal symmetry group P422, with unit-cell parameters $a = b = 46.7$ Å, $c = 100.8$ Å. The data were scaled in the tetragonal space group P422 and were merged and reduced to a set of 35611 independent reflections (99.3% completeness, 95.7% in the outer resolution shell) resulting in a scaling $R_{\text{sym}}$ of 0.076 (0.276 in the outer resolution shell). But, due to the fact that a good electron density map could not be obtained, as explained later in this chapter, during the refinement the data were rescaled in the space group P41 and were merged and reduced to a set of 66386 independent reflections (98.9% completeness, 91.2% in the outer resolution shell) resulting in a scaling $R_{\text{sym}}$ of 0.073 (0.246 in the outer resolution shell) (Table 5-1). The $R_{\text{sym}}$ values are very similar between the two scaled data sets processed in two different space groups.
From the unit-cell volume and the molecular weight of subtype C PR, $V_m$ values (Matthews, 1968) of $\sim 2.56 \, \text{Å}^3/\text{Da}$ (52% solvent content) were calculated assuming four homodimers per asymmetric unit using CNS v.1.1 (Brunger et al., 1998).

**Drug-bound HIV-1 subtype C protease**

A total of 180° of data were collected (360 images) from each of two single crystals of subtype C PRs complexed with IDV or NFV and a total of 65956 and 61466 reflections were measured, respectively. Complete data sets were collected to 2.3 Å resolution for both structures.

The data for the subtype C PR complexed with IDV were initially processed in the monoclinic space group P2. The diffraction data were subsequently processed and reduced with an overall $R_{\text{sym}}$ of 0.084 (0.332 in the outer resolution shell), and completeness of 96.7% (96.4% in the outer resolution shell). A similar data set for the subtype C PR complexed with NFV was processed as well in P2 space group and resulted in an $R_{\text{sym}}$ of 0.113 (0.373 in the outer resolution shell) and completeness of 97.3 (93.7% in the outer resolution shell) Table 5-1 gives a full summary of the data collection statistics.

On inspection of the intensities of the 0k0 reflections, the existence of a twofold screw axis along the $b$ direction could be inferred, therefore implying that the space group was P2$_1$ for both crystals. From the unit-cell volume and the molecular weight of subtype C PR complexed with IDV and NFV, $V_m$ values (Matthews, 1968) of $\sim 2.9 \, \text{Å}^3/\text{Da}$ (56% solvent content) for both crystals were calculated assuming two homodimers per asymmetric unit using CNS v.1.1 (Brunger et al., 1998).

**Molecular Replacement: Particle Orientation and Position**

Molecular replacement exploits the similarity of the tertiary fold of the protein or macromolecular complexes when structures are being determined. A clue to the degree of sequence and, thus, structural homology that exists between two protein molecules is an
alignment analysis or pair-wise BLAST search. It is generally agreed that, if two proteins share more than 30% sequence identity their three-dimensional structures will be superimposable to within few Angstroms. Structure determination using molecular replacement consists of two separate operations: a rotation search and a translation search. Both operations rely upon the three-dimensional Patterson function to derive a structure solution (Rossmann, 1990). There are many programs with slightly different algorithms that have been developed to implement this process, such as those in CNS and CCP4, and molecular replacement has become the fastest and the most direct method of macromolecular structure determination (Brunger et al., 1998).

**Unbound HIV-1 subtype C protease**

We decided to use 1HHP (Spinelli et al., 1991) as model to solve the unbound subtype C PR molecule orientation and position. The cross-rotation function in the point group P422 searches provided one clear solution with the correlation factor of 0.1503, with the next highest peak having a correlation coefficient of 0.0682.

Using this orientation matrix, a translation function search in the Laue group P4₁2₁2 for the non-crystallographic dimer rotation solution provided one peak with correlation coefficient of 0.654 and R-values of 0.587. We also performed translation function searches for P4₂2₁2 and P4₃₂₁₂ but no peaks were found.

**Drug-bound HIV-1 subtype C protease**

Cross-rotation function searches with the subtype C PR – IDV complex data in the point group P2, using the 1W5Y (Lindberg et al., 2004) polypeptide model provided four solutions, two of which reflected the molecular two-fold rotation axis of the model. The four unique peaks were approximately twice the values of the next highest peaks in the rotation function.

Translation function search using the subtype C PR – IDV complex data in the space group P₂₁ for the two non-crystallographic dimer rotation solutions provided peaks with correlation...
coefficients 0.380 and 0.287, and R-values of 0.473 and 0.518, respectively. A second translation search was required to define the relative orientation of the two dimers along the y direction, by constraining the position in the A dimer in a second translation search. This solution had a correlation coefficient and R-value of 0.675 and 0.358, respectively. The fractional translations were $T_x=0.4238$, $T_y=0.0000$, and $T_z=0.0486$ for site A, and $T_x=-0.4222$, $T_y=0.0245$, and $T_z=-0.4449$ for site B. The solution for the subtype C PR – NFV complex data used the orientations and positions as defined from the subtype C PR – IDV complex data.

**Structure Refinement**

Once an initial model has been constructed, iterative refinement can be carried out either with CNS package or with the CCP4 program Refmac5. These algorithms will refine the model based upon accepted bond lengths and angles that have been found in well-refined proteins and peptides solved to date. The goal of the refinement is to improve the agreement between the observed amplitudes obtained from the diffraction data and calculated values for the model.

To refine the HIV-1 subtype C PR structures, both bound and unbound forms, we employed the CNS refinement algorithm that consists of several steps: rigid-body, simulated annealing, B-factor refinement and conjugate-gradient energy minimization.

For all three structures a rigid body refinement was undertaken in order to optimize the positions of the molecules in the asymmetric unit. This is a preliminary step in refinement that does not take in the account the conformation of the amino acid side-chains, but facilitates a finer treatment of the model in the subsequent refinement steps. Individual B factor refinement allows for a better measurement of the movements and dynamics of the molecules. This, combined with conjugate-gradient energy minimization, will apply restraints on bond lengths and angles to bring the molecule to the most energetically favorable conformation. Once these steps have been finished, electron density maps are calculated using these improved phases and the degree of
agreement with the model is examined manually and adjusted accordingly. These refinement steps are part of an iterative refinement loop and are repeated several times to attain the best model.

**Unbound HIV-1 subtype C protease**

The solution obtained through rotation and translation searches was further refined with cycles of rigid body, individual B-factor, energy minimization, and water-pick procedures in CNS interspersed with rounds of manual modeling using the program O (Jones, 1978).

Initial Fo-Fc and 2Fo-Fc electron density maps were calculated using the phases from the model obtained by molecular replacement and were contoured at 3.0 \( \sigma \) and 1.5 \( \sigma \), respectively. The maps were of good quality, the main chain electron density was continuous and the side chains of the amino acids similar between the two protein molecules were well fitted into the density. The maps also clearly showed that the flap regions of the model were not fitting within the electron map, and further extensive rearrangements on the backbone and side chains were required.

However, despite extensive cycles of refinement in CNS, the R\text{work} and R\text{free} values did not converge to less than 20.6% and 21.5%. Also, at resolutions higher than 2.3 Å, the F\text{o}-F\text{c} maps showed patches of positive electron density along the main chain of the protein molecule. The results called for a re-examination of the assigned space group and data processing. The data were rescaled in P4\text{1} space group. Upon this reassignment, the R values and the quality of the electron density map were substantially improved and further refinement was performed in CNS out to 1.6 Å resolution. Figure 5-2 shows the packing arrangement for the 8 monomers for the unbound HIV-1 subtype C PR in the space group P4\text{1}.

The PDB coordinate file obtained from the final cycle of CNS refinement was used to generate the fractional coordinates and equivalent isotropic thermal parameters for further
refinement using \textit{SHELX} and \textit{SHLEXPRO} (Sheldrick, 1997; Sheldrick and Schneider, 1997). The water and solvent molecules placed in the model during the \textit{CNS} refinement steps were not removed prior to input into \textit{SHELX}. The first cycle of refinement resulted in an initial $R_{\text{work}}$ of 23.8\% and $R_{\text{free}}$ of 26.4\% (Table 5-2). Anisotropic refinement of the protein atoms only during the second cycle in \textit{SHELX} and removal of 50 weaker water molecules increased significantly the number of parameters ($N_{\text{par}}$) and brought the $R_{\text{work}}$ and $R_{\text{free}}$ down to 14.9\% and 18.3\% respectively. In the next two steps, the addition of more solvent molecules in the model did not significantly improved the R-values. The next stage of refinement involved the generation of 1618 hydrogen atoms (according to the riding H-atom model) resulting in a significant decrease in $R_{\text{work}}$ and $R_{\text{free}}$ to 13.9\% and 16.9\%, respectively. Full anisotropic refinement proceeded from this point given the high resolution (1.20 Å) and yielded an $R_{\text{work}}$ of 12.4\% and $R_{\text{free}}$ of 15.3\%. The last step of refinement in \textit{SHELX} was performed using 100\% of the data, including 5\% of the data put aside for $R_{\text{free}}$ calculation and resulted a final $R_{\text{work}}$ of 12.3\% (Table 5-3).

\textbf{Drug-bound HIV-1 subtype C protease}

Rigid body refinement of the solution for HIV-1 subtype C PR – IDV complex, using data between 15 and 4.0 Å resolution, converged at $R_{\text{free}} = 30.3$ \% and $R_{\text{work}} = 32.1$ \%. Figure 5-3 shows the packing arrangement for the two homodimers in the P2$_1$ space group for the subtype C PR – IDV complex.

Initial $F_o-F_c$ electron density maps for both complexes, contoured at 3.0 $\sigma$, showed clear interpretable density for the respective inhibitors IDV and NFV present in both dimers. Topology and parameter files for IDV and NFV were obtained from the HIC-UP website. Iterative cycles of refinement were performed in \textit{CNS} including data between 15 and 2.3 Å resolution. The final R values converged to $R_{\text{work}}$ of 25.3 \% and $R_{\text{free}}$ of 28.3 \% and $R_{\text{work}}$ of 21.1 \% and $R_{\text{free}}$ of 23.7 \% for NFV and IDV complexes, respectively (Table 5-3).
Structure Validation

Structure validation of the refined model is an important final step in the structure determination and it can be performed using programs such as PROCHECK (Laskowski et al., 1993). Validation is used to ensure that bond lengths and angles are in accordance with acceptable values. Once validated, the model can be reliably interpreted to elucidate the biological implications the structure contains. Once structure refinement and validation is complete, final models are submitted to the Protein Data Bank (PDB) for later release to the public.

Unbound HIV-1 subtype C protease

The quality of the final refined structured of HIV-1 subtype C PR was verified with the PROCHECK program (Laskowski et al., 1993). 94.9 % of the dihedral angles were found to be located in the most favored regions, with all the other in the additional allowed regions (Figure 5-4A).

Drug-bound HIV-1 subtype C protease

The final models for IDV- and NFV- bound subtype C PR showed 96.2 % and 95.9 % of the dihedral angles in the most favorable regions, respectively, with all the other in the additional allowed regions (Figure 5-4B, C).

Structure Analysis

As expected, given the high amino acid sequence identity, all three crystal structures solved in this study, the unbound form of subtype C PR and the enzyme complexed with IDV and NFV, had the overall shape and fold of HIV-1 subtype B PR. The PRs crystallized as homodimers with the substrate-binding cleft in the middle of the molecule, covered by two flexible arms known as the flaps.
**Unbound HIV-1 subtype C protease**

This high-resolution structure showed excellent electron density for all the PR atoms and solvent molecules. The electron density map of the naturally occurring polymorphisms at positions 36 and 37 is shown in the Figure 5-5.

An unassigned density was located between the open flaps of the unbound subtype C PR. The F_o-F_c electron density has a “C-shaped” appearance and was initially interpreted and built as a string of disordered water molecules. Later, during the refinement, the density was interpreted and built as a fragment of Triton-X 100, the additive added in the crystallization drop (Figure 5-6).

The average B values for the main-chain and the side-chain atoms were 13.92 and 16.1 Å², respectively (Table5-3). The distribution of the B factors (atom thermal motion) for the main chain atoms showed similar peaks at the C- and N- termini, at the elbow (34-38 residues), at the tip of 60’s loop (65-70 residues) and in the active site (78-84) for both subunits for subtype C and subtype B multi-drug resistant PR (Figure 5-7). The B factors profile for the wild type subtype B PR was different from the outer two structures in that it had a much lower peak for the active site residues 78-84. Interestingly, the B factors for subtype C PR showed a peak in the 12-20 β-sheet region, where three of the naturally occurring polymorphisms are located. This peak does not appear in the multi-drug resistant B PR (PDB code 1RPI). Also, the B values for the flap regions in subtype B PR are below the average indicating that there is limited motion of the atoms within these structures. The B values of the same regions in subtype C PR show values slightly above the average and higher that those for subtype B PR, while the wild type subtype B PR showed higher B values for the flap region and a slight shift of the peak B values for the residues located in the elbow of the flap.
The solvent for the final model included 341 water and 4 glycerol molecules, with average B values of 26.7 and 31.6 Å², respectively (Table 5-3).

Analysis of the anisotropy of the structure was performed using the final output file from SHELX. Anisotropy is defined as the ratio between the minimum and maximum eigenvalues of the matrix of anisotropic displacement parameters (Merritt, 1999a). An analysis of the anisotropically refined structures in the Protein Data Bank performed by Merritt, 1999 indicates that the standard mean anisotropy is 0.45 with a standard deviation of 0.150 (Merritt, 1999b); the current model of the unbound subtype C PR conforms to these values. We were particularly interested in the distribution of the anisotropic parameters of the atoms in the flaps as this would help in understanding the direction of the motion of these highly flexible structures (Figure 5-8). A careful analysis of the ellipsoidal atomic motions indicates that a possible direction of the motion of the flaps is perpendicular on the active site. This motion is accompanied, at a lesser extent, by the similar movements in the elbow of the flap.

High-resolution structural information allows for better interpretation of the structural disorder, including amino acid side chains that exhibit alternate conformations (Esposito et al., 2000). In our model, several residues showed disordered density for the side chain and had alternate conformations. The alternate conformations for Glu21, Glu34, Glu35, Pro44, Arg57, Lys69, and Val82 were generated in the graphics program O version 10.0.1 (Jones, 1978). All of these residues are located in the outer loops of the enzyme, in contact with the solvent molecules.

As expected, the HIV-1 subtype C PR structure is highly similar in shape and fold with subtype B PR. A least-square superimposition of our model of subtype C PR in the unbound form and the previously reported wild type (PDB code 1HHP) and multi-drug resistant form (PDB code 1RPI) of subtype B PR was conducted using the graphic software Pymol (DeLano...
Scientific) (Figure 5-9A). The wild-type structure of subtype B PR (1HHP) known as the “semiopen” form, crystallized in the space group P4_12_12 with cell constants similar to those of the subtype C PR and multi-drug resistant mutant of subtype B PR (1RPI), which crystallized in P4_1 space group.

The mean r.m.s. deviations for C^α atoms between 1HHP, 1RPI and our model were 1.09 and 0.65 Å, respectively (Figure 5-9B). These values are above the ~0.5 Å that represents the crystallographic error margin when comparing to crystal structures. The catalytic triplet residues 25-27 located in the active site showed values of 0.14 and 0.35 Å, consistent with the highly conserved core structure. However, an interesting observation was that the main chain atoms of the active site residue Val82 showed an r.m.s. deviation of 1.07 to 1.40 Å. The highest difference was observed, as detailed below, between the flaps of the PR.

The relatively high mean r.m.s.d between subtype B and C PRs is the result of several differences of more than 1.0 Å that were observed for the main chain atoms of the residues: 35-42 (the elbow of the flaps), 49-53 (the tip of the flaps), 63-70 (the 60s’ loop), and 80-81. The highest r.m.s. deviations were located in the outer loops, especially in the loops harboring the naturally occurring polymorphisms such as the elbow of the flap and the 60s’ loop. The most striking difference between the wild-type subtype C and B PRs is a very large conformational change in the flap region, with the distance between the tips of the flaps of the subtype B PR being 4.4 Å while the distance for subtype C PR is 12.2 Å (Figure 5-10A). This accounts for a movement of more than 8 Å of the flaps of the subtype C PR relative to the subtype B PR structure. Interestingly, our structure showed a similar distance between the flaps as that of a subtype B multi-drug resistant PR (1RPI) that shows an opening of the flaps of 12.53 Å (Figure 5-10B).
M36I is one of the baseline polymorphisms occurring in subtype C PR that is considered a secondary drug resistance mutation in subtype B PR. Analyzing both PRs, we observed that the side chain of the smaller Ile residue make half of the van der Waals interactions observed in subtype B PR harboring the Met residue. We also noted a downshift in the subtype C PR 36 – 41 loop with an average displacement of 1.6 and 1.3 Å, when compared with the wild-type and multi-drug resistant subtype B PR, respectively (Figure 5-11A).

Another polymorphism, known to occur in more than 95% of the subtype C strains, is L89M. This residue is located in the core of the PR and participates in an extensive network of hydrophobic interactions, about twice the number of interactions in subtype C PR when compared with subtype B PR. H69K is located at the base of the PR, within a loop that exhibited an average r.m.s.d of 1.7 and 1.0 Å when compared subtype C PR to the wild-type and multi-drug resistant subtype B PR, respectively. The I93L polymorphism is located in close spatial vicinity of position 69 and, in both subtype B and C PRs, makes numerous interactions with the surrounding residues (Figure 5-11B).

**Drug-bound HIV-1 subtype C protease**

Both inhibitor-bound structures of subtype C PR were solved in the P2₁ space group with two homodimers in the unit cell. As for the unbound PR, the crystallographic asymmetric unit had a dimer of subtype C PR. In both inhibitor-bound structures the PR exhibited a closed conformation with the inhibitor buried in the active site, covered by the two flaps. Both structures, solved at a resolution of 2.3 Å, revealed a good quality density map. The solvent for the final model included 169 water molecules and 2 sodium ions for IDV-bound PR and 136 water molecules for NFV-bound PR. The average B factors for the main-chain of the IDV- and NFV- bound structures were 28.5 and 31.5 Å², respectively (Table 5-3). In each structure the inhibitor, either IDV or NFV, was observed to have a single orientation.
The least square superimposition of the IDV-bound structures of our model and subtype B PR (PDB code 1SDT) was performed in the program \textit{O} (Jones, 1978) and revealed an r.m.s. deviation of 0.30 Å for the C\textsubscript{α} atoms, with the highest values measured in the outer loops 65-69 of 1.3 Å (Figure 5-12A,B).

In the IDV-bound subtype C PR, the profile of mean main-chain B factors is similar to that of bound subtype B PR indicating a similar behavior of these two proteins upon inhibitor binding. However, there were several differences in the B value profiles of these two enzymes (Figure 5-13). We observed that monomer A in subtype C PR has significantly higher B factors for the elbow of the flap when compared with the B monomer. Also, the 60’s loop in the A monomer of the bound subtype B PR showed higher B factors when compared to the same region in B monomer. We also noticed that the flap regions in the bound subtype B PR showed a ~2-fold increase in the B values when compared with the bound subtype C PR. This is a perplexing finding, as we would expect that the closed flaps, due to decrease flexibility and atomic motion, would exhibit low B values. The observed difference between the B factors of these two structures might be due to crystallization conditions, crystal quality and/or crystal packing environment. The bound subtype B PR structure was solved at 1.3 Å in the \textit{P2}_1\textit{2}_1\textit{2}_1 space group while our structure was solved at 2.3 Å in the \textit{P2}_1 group.

The IDV atoms were in well-defined electron density, and the superimposition of the bound subtype B and C PRs showed that the IDV structures superposed very well (Figure 5-14) and the number of subtype C PR – IDV van der Waals contacts showed only a small variation among the subtype C and B PR crystal structures (Figure 5-15). Also these crystal structures showed a very similar arrangement of PR-IDV hydrogen bond interactions (H-bonds), including the same water-mediated interactions. Six water molecules that mediate interactions between
IDV and the subtype C PR were observed in our structure forming 14 H-bonds, while there are 7 water molecules that form 16 H-bonds in the subtype B PR. The number of H-bonds between the inhibitor and the active site residues is 3 for both structures compared in our study.

In our model, the NFV atoms are in well-defined electron density (Figure 5-16). We attempted to perform the same comparison of the NFV-bound subtype C PR with a similar subtype B PR. To our surprise, despite the fact that NFV was one of the protease inhibitors frequently used to treat HIV-positive patients and NFV-induced drug resistance mutations, such as D30N, L90M, N88D, have been extensively studied, only one NFV-bound subtype B PR was deposited in the Protein Data Bank (PDB code 1OHR). This structure was deposited in 1997, crystallized in the P2₁2₁2₁ space group and was solved at 2.1 Å resolution. However, when superimposed, the monomer B of the subtype B PR structure showed a spatial shift of ~0.5 Å relative to our structure, apparently due to a rotational movement at the dimerization region at the base of the PR. Consequently a correct alignment could not be performed.

**Discussion**

We were able to obtain crystals of subtype C PR and have solved the first crystal structure of this enzyme, in both unbound and inhibitor-bound forms. The unbound form of the HIV-1 subtype C PR was solved at 1.20 Å resolution, representing the first structure of subtype C PR and the highest resolution solved structure of non-B subtype PR reported to date.

Our research is of interest because many *in vivo* and *in vitro* studies (Clemente et al., 2006; Gonzalez et al., 2006; Kantor and Katzenstein, 2003; Peeters, 2001; Sanches et al., 2007; Tanuri et al., 1999; Velazquez-Campoy et al., 2003) advanced the hypothesis that the naturally occurring polymorphisms play a role in modulating antiretroviral drug susceptibility with the possibility of faster development of drug resistance during therapy. In this study we analyzed the structural parameters of the subtype C PR, unbound and complexed with IDV and NFV, and we
also compared them to available structures of subtype B PR in an attempt to understand the structural differences due to the baseline polymorphisms and their implications in antiretroviral drug resistance/susceptibility.

One of the PR regions believed to be involved in modulating the affinity of the PR for inhibitors is the flap domain. Understanding the factors underlining the HIV PR flap mobility has profound implications in elucidating the detailed mechanism of substrate/inhibitor binding of this enzyme and in the design of new therapeutic agents such as allosteric inhibitors intended to interfere with the flap opening and thereby with enzymatic function. The mechanisms and the factors involved in coordinating and modulating the motion of the flaps have been the focus of study for many researchers. Several studies showed that that flaps open upward and laterally (Ishima et al., 1999; Nicholson et al., 1995; Toth and Borics, 2006; Wlodawer and Erickson, 1993) while others argued that the tip of the flaps curl inside, making hydrophobic contacts with the several residues located in the active site (Scott and Schiffer, 2000). It is generally agreed that the large motion of the tip of the flap is accompanied by changes in the hinge and the elbow of the flap as well (Clemente et al., 2004; Perryman et al., 2006). Several NMR and molecular dynamics studies investigated the conversion between closed, semi-open, and fully open forms of HIV PR. These conformations appear to be in dynamic equilibrium, with the semi-open form being the most prevalent (Freedberg et al., 2002; Hamelberg and McCammon, 2005; Hornak et al., 2006b; Nicholson et al., 1995). The main focus of our study was to analyze the structural differences between subtype B and C PRs in an attempt to better understand the role of the naturally occurring polymorphisms. However, our 1.2 Å resolution structure allowed us to study the thermal motion and anisotropic parameters within subtype C PR and make inferences about the possible mechanism of opening/closing of the flaps. In our structure the amplitude of the
atomic motion in the flaps does not appear to be significantly higher than the core of the enzyme. These data, correlated with relatively low B values for the flaps (Figure 5-7), argue for a limitation of the flap movements, probably due to crystal contacts as has been proposed by Hornak et al. (Hornak et al., 2006a). Among the crystal contacts involved in holding the flap open are: H-bonding of the carbonyl of Gly49 with side chain amino group of Arg41’ and hydrophobic interaction Ile50 with Pro81’. In interpreting the factors involved in keeping the flaps open, we have also to consider the role of the C-shaped density observed in our structure. The multi-drug resistant form of subtype B PR contains about 100 water molecules in the active site cavity. Martin et al. proposed that these water molecules form a scaffold in the active site cavity, preventing the PR from collapsing in the absence of a ligand (Martin et al., 2005).

Also, an unbound structure of the wild-type subtype B PR has been recently released in the Protein Data Bank (code 2PC0) by Heaslet et al. (Heaslet et al., to be published). The structure crystallized in the P4_1212 space group and has widely open flaps. This new finding argues for crystal contacts having a prominent role in propping open the flaps, when HIV PR crystallizes in P4 or P422 space groups. However, it could be that the unbound PR prefers the open conformation in solution and the prevalence of this form induces the enzyme to crystallize in the P4 or P422 space groups and, consequently, the crystal contacts are formed due to the open form of the PR and are not alone the cause of the flaps staying open.

Other regions of interest that could further elucidate the changes in the flaps during binding/releasing the substrate/inhibitor are the hinge and the elbow of the flaps. When superimposing the unbound subtype C PR with either wild type or drug resistant subtype B PR, we observed that there are several interesting differences. Position 36 occupies a region in the PR that is highly mobile during flap opening and closing in the course of ligand binding. It has been
argued that the M36I mutation may promote long-range structural changes in the active site or changes in the flexibility of the PR which may lead to either the closed or the open conformation of the PR being dominant (Clemente et al., 2004). As seen in our structure and other several previous studies (Clemente et al., 2004; Martin et al., 2005), Met36 makes extensive interactions with residues located in the 10s’ and 60s’ loops. In subtype C PR the bulkier Met is exchanged for a smaller Ile and consequently there is a decrease in the number of van der Waals interactions between the 10s’ and the 60s’ loops. This effect is augmented by the I15V polymorphisms where a smaller Val replaces the Ile residue. The overall effect is a decreased number of interactions between these two loops in subtype C PR, allowing for an increased flexibility of the elbow of the flap. Previous studies of subtype B and F PRs have also argued that this polymorphic change causes a collapse of the elbow of the flap, resulting in displacement of the main chain of this loop toward the loop 76 – 83, stabilizing the catalytic S1/S1’ pockets (Sanches et al., 2007). We observed a similar effect in our structure, where the catalytic residues Pro81 and Val82 are shifted towards the active site.

The amino acid residue at position 89 is located in the hydrophobic core of the PR and, when mutated to a Met, it makes extensive hydrophobic contacts with neighboring residues, more than in subtype B PR which harbors a Leu at this position. Variation in the number of hydrophobic residues appears to be important for both maintaining the structural stability of the enzyme and allowing conformational changes. It has been hypothesized that the hydrophobic core residues slide by each other, exchanging one hydrophobic van der Waals contact for another, with little energy penalty, while maintaining many structurally important hydrogen bonds. Such hydrophobic sliding may represent a general mechanism by which proteins undergo conformational changes. Consequently, mutation of these residues in HIV-1 PR would alter the
packing of the hydrophobic core, affecting the conformational flexibility of the PR. It has been proposed that these residues impact the dynamic balance between processing substrates and binding inhibitors, and thus any change in this region could contribute to drug resistance/susceptibility (Foulkes-Murzycki et al., 2007). The increased number of van der Waals interactions with the presence of L89M polymorphism might increase the stability of the subtype C PR and affect the dynamic properties of the PR and potentially affect its ability to bind inhibitors and substrates. Furthermore, a previous study hypothesized that Met89 was assumed to mimic the role of the L90M mutation, by displacing Asp25 and thus constraining the S1/S1’ pockets (Sanches et al., 2007).

The subtype C PR harbors three signature residues: 12, 15, 19 which are located in a β-sheet that forms what is called the 10s’ loop. The influence of the polymorphisms in this region was not widely studied, but there are two interesting observations in our study. First, as mentioned above, the polymorphic change from a larger Ile15 to a smaller Val15 in subtype C PR further reduces the number of interactions between the 10’s loop and the elbow of the flap, changing in this way the dynamics of the elbow of the flap. Second, the analysis of the B factors (Figure 5-7) showed that there is a significant difference between the main-chain B values of the 10-22 residues between subtype B and C PRs. In the unbound form of subtype C PR the 10s’ loop had B factors just below the average B value, while the unbound subtype B PR exhibited a ~2.5-fold increase in the B values in this region. This finding leads to the conclusion that the 10s’ loop is more ordered and probably less flexible in subtype C PR. A similar effect, but of a lesser magnitude, happens for 60s’ loop. All these data taken together, the increased hydrophobic contacts due to L89M polymorphisms and decreased stability of the 10’s and 60’s loops, could suggest that, in subtype C, there is an increased stability at the base of the PR. The large number
of van der Waals interactions forms a scaffold on which the flaps can swing open easier, with fewer energetic requirements. Also, this arrangement could change the size of the active site to an extent where, upon addition of major drug resistance mutations, the inhibitor binding is hindered, while at the same time maintaining a reasonable affinity for the more flexible substrate.

The role of polymorphisms might be that they stabilize the core of the HIV PR while maintaining the flexibility of the flaps, promoting the open, flexible conformation of subtype C PR. Since inhibitors are rigid and are designed to bind the closed conformation, they would preferentially bind to enzymes that carry mutations that favor the closed conformation (Clemente et al., 2004). Consequently, this open conformation of the HIV subtype C PR would be less favorable for inhibitor binding. These results correlate with our kinetic data as well as with other recent structural and kinetic studies (Sanches et al., 2007) that showed that the naturally occurring polymorphisms in subtype F PR might amplify the effect of drug resistance mutations.

The overall superimposition of IDV-bound structures of subtype B and C PRs showed that the most significant differences are located in the 60’s loop, at the base of the PR, with slight differences in the elbow of the flap. The flaps in both enzymes are firmly closed over the active site enclosing the inhibitor within. The LIGPLOT (Wallace et al., 1995) analysis of the IDV bound in the active site of subtype B and C PRs revealed that there are no significant differences in the number of H-bonds and hydrophobic interactions between the inhibitor and the active site of the PR. These results correlate well with previous kinetic data that showed that the $K_i$ values for IDV for subtype B and C PRs are comparable, indicating that both enzymes bind the inhibitor with similar affinity.

Our structural study revealed several structural differences between subtype B and C PRs. Even though crystallography offers a static exploration of a structure, we were also able to make
several inferences about the dynamics of the flaps. These results could add to the general effort in explaining if and how the naturally occurring polymorphisms contribute to the mechanism through which subtype C PR could gain resistance to protease inhibitors.

These data and subsequent studies with other inhibitors will greatly aid in our efforts to understand the influence of baseline polymorphisms in modulating the enzyme sensitivity and resistance to current drug therapy and hopefully provide new insight into designing novel inhibitors less likely to promote development of drug-resistance mutations.
Figure 5-1. Optical photographs of HIV-1 subtype C protease crystals. A) Crystal of unbound HIV-1 subtype C PR. Approximate dimensions are 0.5 x 0.4 x 0.3. B) Crystal of the IDV-bound subtype C PR. Approximate dimensions are 0.4 x 0.3 x 0.1 mm. NFV-PR complexes formed crystals similar shape and dimensions.
Figure 5-2. Packing diagram for HIV-1 subtype C protease – unbound form. Ribbon Packing Diagram in P4₁ Crystal Lattice. There are four subtype C PR homodimers in the unit cell (shown in red and light red, and green and light green). The box depicts the unit cell. Figure rendered with Pymol (Delano Scientific).
Figure 5-3. Packing diagram of HIV-1 subtype C protease – IDV-bound form. There are two subtype C PR homodimers in the unit cell (shown in red and light red, and green and light green). The box depicts the unit cell. Figure rendered with Pymol (Delano Scientific).
Figure 5-4. Ramachandran diagrams. A) Unbound subtype C PR. B) IDV-bound subtype C PR. C) NFV-bound subtype C PR. The color code is as follows: pink – most favored regions; yellow – additional allowed regions, gray – disallowed regions. Plots created using Coot (Emsley and Cowtan, 2004).
Figure 5-5. The 2Fo-Fc electron density map of the unbound subtype C protease. The map is contoured at 3 $\sigma$ using data to 1.2 Å resolution. Figure rendered with Pymol (Delano Scientific).
Figure 5-6. The C-shaped electron density between the flaps. A) Ribbon representation of the subtype C PR with the naturally occurring polymorphisms showed as red spheres. The alternative structures of the Triton X100 fragments are shown in light and dark yellow between the flaps of the PR. B) Close-up view from above the PR molecule. The electron density map is contoured at 2σ level, using data to 1.2 Å resolution. Figure rendered with Pymol (Delano Scientific).
Figure 5-7. The normalized mean B values for the main chain atoms of the wild type subtype C and multi-drug resistant subtype B proteases. The normalized B values are plotted for the residues of subtype C PR (red), multi-drug resistant B PR (green) and wild type subtype B PR (blue). Normalization was done by dividing the B values for the main chain of each residue by the average B values for the entire PR molecule. The residues in the two subunits are labeled 1-99 and $1'-99'$. 
Figure 5-8. Thermal ellipsoid diagram for unbound HIV-1 subtype C protease. A) Thermal diagram representing the anisotropy for the 42-59 residues within the flaps, including the side chains. B) Thermal diagram representing the anisotropy for the main-chain atoms of the 42-59 residues within the flaps. C, O, N and S atoms are colored gray, red, blue and yellow, respectively. The arrows indicate the direction of the flap motion. Figure created using RASTEP (Merritt, 1999a; Merritt and Bacon, 1997).
Figure 5-9. Superimposition of the wild type subtype C protease with the wild type and multidrug resistant mutant of subtype B protease. A) Ribbon representations of wild type subtype C PR (red) superimposed over wild type subtype B PR (blue) or multidrug resistant mutant B PR (green). The naturally occurring polymorphisms in subtype C PR are represented as red spheres. B) The r.m.s differences (Å) per residues are plotted for the Cα atoms of wild type subtype B PR (blue) and mutant B PR (green) compared with the wild type subtype C PR. The residues in the two subunits are labeled 1-99 and 1’-99’.
Figure 5-10. The comparison between the flaps of subtype C and B proteases. A) Ribbon representation of the flap regions of the wild type subtype C PR (red) superimposed over the wild type subtype B PR (blue). B) Ribbon representation of the flap regions of the wild type subtype C PR (red) superimposed over the mutant subtype B PR (green). The gray surface represents the active site. Figure rendered with Pymol (Delano Scientific).
Figure 5-11. The naturally occurring polymorphisms in subtype C protease. A) Ribbon representation of the flap and elbow regions of the wild type subtype C PR (red) superimposed over the wild type B PR (blue). B) Ribbon representation of the 60’s loop and residues 89 and 93 of the wild type subtype C PR (red) superimposed over the wild type B PR (blue). The amino acid residues are represented as sticks. Figure rendered with Pymol (Delano Scientific).
Figure 5-12. Superimposition of the IDV-bound subtype C protease with the IDV-bound subtype B protease. A) Ribbon representations of IDV-bound subtype C PR (red) superimposed over IDV-bound subtype B PR (gray). The naturally occurring polymorphisms in subtype C PR are represented as red spheres. B) The r.m.s differences (Å) per residues are plotted for the Cα atoms of bound B PR compared with the bound subtype C PR. The residues in the two subunits are labeled 1-99 and B1'-99'.
Figure 5-13. The normalized mean B values for the main chain atoms of the IDV-bound subtype C and B proteases. The normalized B values are plotted for the residues of subtype C PR (red) and B PR (gray). Normalization was done by dividing the B values for the main chain of each residue by the average B values for the entire PR molecule. The residues in the two subunits are labeled 1-99 and 1’-99’.
Figure 5-14. IDV in the active site of HIV protease. A) Electron Density Map for IDV in the Bound Subtype C PR Crystal Structure. The contour level is $2\sigma$. The pyridyl group of IDV is labeled. B) Stick representation of the IDV in the active site of the subtype C (red) and B (blue) PRs. Ball-and-stick representation of the amino acid residues in the subtype C (red) and subtype B (blue) surrounding the inhibitor. Figure rendered with Pymol (Delano Scientific).
A)

Figure 5-15. Ligplot analysis. A) Schematic representation of the H-bonds and hydrophobic interactions of the IDV bound in the active site of subtype C PR B) Schematic representation of the H-bonds and hydrophobic interactions of the IDV bound in the active site of subtype B PR. Figure made with LIGPLOT (Wallace et al., 1995).
Figure 5-15. Continued
Figure 5-16. Electron density map for NFV in the bound subtype C protease crystal structure. The contour level is 2 $\sigma$. Figure rendered with Pymol (Delano Scientific).
Table 5-1. Data collection statistics

<table>
<thead>
<tr>
<th></th>
<th>Subtype C protease with indinavir (IDV)</th>
<th>Subtype C protease with nelfinavir (NFV)</th>
<th>Unbound subtype C protease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P4₁</td>
</tr>
<tr>
<td>Unit-cell parameters (Å, °)</td>
<td>a=46.7, b=59.5, c=87.4, β=95.6</td>
<td>a=46.7, b=60.1, c=86.7, β=94.7</td>
<td>a=46.7, b=46.7, c=100.8, β=90.0</td>
</tr>
<tr>
<td>Unit cell volume (Å³)</td>
<td>241,915</td>
<td>242,491</td>
<td>219,737</td>
</tr>
<tr>
<td>Vₘ (Å³ Da⁻¹) / solvent fraction (%)</td>
<td>2.90 / 55</td>
<td>2.90 / 55</td>
<td>2.56 / 52</td>
</tr>
<tr>
<td>Total reflections</td>
<td>65956</td>
<td>61466</td>
<td>279351</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>21311</td>
<td>21028</td>
<td>35611</td>
</tr>
<tr>
<td>Crystal mosaicity (°)</td>
<td>0.9</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0 – 2.3 (2.38 – 2.3)</td>
<td>20.0 – 2.3 (2.38 – 2.3)</td>
<td>30.0 – 1.2 (1.24 – 1.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.7 (96.4)</td>
<td>97.3 (93.7)</td>
<td>99.3 (95.7)</td>
</tr>
<tr>
<td>Rₛₚₘᵦ (%)</td>
<td>8.4 (33.2)</td>
<td>11.3 (37.3)</td>
<td>7.6 (27.3)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.1 (3.0)</td>
<td>2.9 (2.6)</td>
<td>7.9 (4.7)</td>
</tr>
<tr>
<td>Average I/σ(I)</td>
<td>8.4</td>
<td>5.6</td>
<td>14.0</td>
</tr>
<tr>
<td>I/σ(I) &gt; 3 (%)</td>
<td>71.1 (36.8)</td>
<td>59.7 (21.0)</td>
<td>76.0 (41.7)</td>
</tr>
<tr>
<td>Job</td>
<td>Action taken</td>
<td>NP</td>
<td>NH</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>First refinement in SHELX</td>
<td>1508</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Protein atoms only anisotropic</td>
<td>1560</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>More solvent added</td>
<td>1560</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Minor adjustments</td>
<td>1560</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Riding hydrogens added</td>
<td>1560</td>
<td>1618</td>
</tr>
<tr>
<td>6</td>
<td>All atoms anisotropic</td>
<td>1560</td>
<td>1618</td>
</tr>
<tr>
<td>7</td>
<td>More disorder</td>
<td>1578</td>
<td>1618</td>
</tr>
<tr>
<td>8</td>
<td>Minor adjustments</td>
<td>1578</td>
<td>1618</td>
</tr>
<tr>
<td>9</td>
<td>Minor adjustments</td>
<td>1582</td>
<td>1618</td>
</tr>
<tr>
<td>10</td>
<td>Minor adjustments</td>
<td>1583</td>
<td>1618</td>
</tr>
<tr>
<td>11</td>
<td>Minor adjustments</td>
<td>1587</td>
<td>1618</td>
</tr>
<tr>
<td>12</td>
<td>Fragments of TX-100 added</td>
<td>1583</td>
<td>1618</td>
</tr>
<tr>
<td>13</td>
<td>SIMU changed from 0.1 to 0.05</td>
<td>1583</td>
<td>1618</td>
</tr>
<tr>
<td>14</td>
<td>Minor adjustments</td>
<td>1583</td>
<td>1618</td>
</tr>
<tr>
<td>15</td>
<td>WGHT changed from 0.1 to 0.2</td>
<td>1583</td>
<td>1618</td>
</tr>
<tr>
<td>16</td>
<td>Further refinement</td>
<td>1583</td>
<td>1618</td>
</tr>
<tr>
<td>17</td>
<td>100% of data refined</td>
<td>1583</td>
<td>1618</td>
</tr>
</tbody>
</table>
Table 5-3. Refinement statistics.

<table>
<thead>
<tr>
<th></th>
<th>Subtype C protease with IDV</th>
<th>Subtype C protease with NFV</th>
<th>Unbound subtype C protease</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R values (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{\text{work}}$</td>
<td>21.1</td>
<td>25.3</td>
<td>12.3</td>
</tr>
<tr>
<td>$R_{\text{free}}$</td>
<td>23.8</td>
<td>28.4</td>
<td>15.5</td>
</tr>
<tr>
<td><strong>No. of water molecules</strong></td>
<td>169</td>
<td>136</td>
<td>341</td>
</tr>
<tr>
<td><strong>No. of inhibitors per PR dimer</strong></td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Ramachandran statistics (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored regions</td>
<td>96.2</td>
<td>95.9</td>
<td>94.9</td>
</tr>
<tr>
<td>Allowed regions</td>
<td>3.8</td>
<td>4.1</td>
<td>5.1</td>
</tr>
<tr>
<td><strong>B factors (Å$^2$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average main chain atoms</td>
<td>28.5</td>
<td>31.5</td>
<td>13.3</td>
</tr>
<tr>
<td>Average, side-chain atoms</td>
<td>27.7</td>
<td>53.1</td>
<td>15.3</td>
</tr>
<tr>
<td>Solvent</td>
<td>35.9</td>
<td>32.8</td>
<td>29.1</td>
</tr>
<tr>
<td>Inhibitor</td>
<td>27.5</td>
<td>27.5</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: $R_{\text{work}} = \left[ \frac{\sum (F_o - F_c)}{\sum F_o} \right] \times 100$; $R_{\text{free}}$ is identical to $R_{\text{work}}$ for 5% of data omitted from refinement.
CHAPTER 6
ANALYSIS OF THE PROCESSING EVENTS OF HIV-1 SUBTYPE A, B, AND C GAG POLYPEPROTEINS

Introduction

Gag and gag/pol processing, an important step in the HIV life, is mediated by the viral protease (PR) and occurs at the plasma membrane of the infected cells during budding and release of progeny virions (Kaplan et al. 1994). It has been shown that this is a sequential process, regulated by the order of the cleavage and the rates of the proteolytic processing at individual cleavage sites (Wiegers et al. 1998). An effective approach to prevent HIV maturation and spread from cell to cell has proven to be the inhibition of gag and gag/pol cleavage.

Several studies (Henderson et al. 1990; Carter and Zybarth 1994; Pettit et al. 1994; Louis et al. 1999; Goodenow et al. 2002; Iga et al. 2002; Liao et al. 2004; Liao and Wang 2004; Pettit et al. 2004) conducted in a number of in vivo and in vitro systems, have examined the order of cleavage for different sites within the HIV-1 gag/pol precursors. There is a general agreement that in HIV-1 subtype B the first cleavage occurs between the p2 spacer peptide and the NC protein. Intermediate cleavages occur at the matrix/capsid (MA/CA) and transframe/p6POL (TF/p6POL) sites (Pettit et al. 2003) (Figure 6-1). Not only the order of cleavage but also the rates of cleavage at different sites have been evaluated and they are estimated to vary by as much as 400-fold between sites (Pettit et al. 1994).

Baseline polymorphisms or mutations arising in response to protease inhibitor (PI) therapy can affect the order and the rates of cleavage. Studies have shown that resistance mutations in the protease of subtype B virus are associated with impaired proteolytic processing and decreased enzymatic activity (Zennou et al. 1998) and that compensatory mutations within gag and gag/pol cleavage sites can partially overcome these defects (Mammano et al. 1998). Furthermore, recent in vitro studies showed not only that the HIV-1 PR embedded within the gag/pol polyprotein
seems to be less sensitive to the effect of clinically used PIs than is the processed PR, but also that the addition of PIs might promote a differential cleavage site affinity (Pettit et al. 2004). Detailed understanding of these events would make possible designing new drugs that would prevent gag/pol processing and thus the virus spreading from cell to cell.

While a relatively large body of literature describes and interprets data about subtype B gag and gag/pol processing, there are no data about the order or the rate of gag and gag/pol cleavage in HIV-1 subtype C or any other subtypes. Sequence analysis of the cleavage sites within the gag and gag/pol precursors showed that these sites are relatively well conserved for HIV-1 group M subtypes. In a recent study performed on 84 full-length nucleotide sequences (de Oliveira et al. 2003) it has been shown that overall, 58.3% of the 12 HIV-1 cleavage sites are significantly more diverse in C than in B viruses. The same study supports the theory that these cleavage sites have evolved in a subtype-specific manner, and that the variation of subtype C cleavage sites began early, prior to the diversification of HIV-1 subtypes.

There are sequence differences at the level of cleavage site among different subtype polyproteins studied. It has been shown that the sites situated upstream of the protease domain directly influence both the (auto) processing activity of PR and the order and efficiency of cleavage of gag and gag/pol precursors (Pettit et al. 1994; Zybarth et al. 1994; Zybarth and Carter 1995; Louis et al. 1999).

Natural variation at p2/nucleocapsid (p2/NC) cleavage sites may play an important role, not only in regulation of the viral cycle but also in disease progression and response to therapy (de Oliveira et al. 2003). Also, HIV-1 subtype C can be considered to be natural and viable variants of subtype B viruses. More data about processing events in subtype C would allow for evaluation of efficiency of current therapy and also would contribute to the effort of finding new
drugs to inhibit this process. The identification of common patterns may facilitate the
development of broad-based inhibitors with increased specificity and improved binding to the
mutated PR. These secondary inhibitors may preempt or delay the emergence of resistance.
These studies will add to our growing understanding of interaction within the gag/pol
polyprotein pre-processing structure that will be an excellent target for new drug discovery.

In this study we aim to analyze the processing events in HIV-1 subtype C and subtype A
gag polyproteins and to bring qualitative and quantitative information regarding the order and the
rate of processing in subtype C and subtype A gag polyproteins. We performed similar analysis
for HIV-1 subtype B gag polyprotein and the data obtained were compared and interpreted. Our
goals were to understand whether the naturally occurring polymorphisms within the PR or/and
the cleavage sites modulate the processing in HIV-1 subtype C and subtype A gag polyproteins.
Overall, the results of proposed experiments will aid in developing more effective therapeutic
approaches for patients infected with subtype C viruses.

Results

Polymorphic Sites within HIV-1 Subtype A, B, and C Gag Polyproteins

The alignment of the three gag polyproteins was performed with ClustalW software. The
HIV-1 subtype B gag polyprotein contains 512 amino acid residues, while HIV-1 subtype C and
A gag polyproteins have only 491 and 496 amino acid residues, respectively. The three gag
sequences had 74.7% identity, 9.9% strong similarity, 4.1% weak similarity and 11.3%
difference (Figure 6-2). The most conserved regions were capsid (CA) (with 95.2% identity and
strong similarity and only 1% difference), nucleocapsid (NC) (with 92.7% identity and strong
similarity and 5.4% difference), and p1 (with 94.7% identity and strong similarity and 5.3%
difference). The most polymorphic regions were p2 (with 53.3% identity and strong similarity
and 33.3% difference) and p6 (with 47% identity and strong similarity and 45.5% difference).
Analysis of the matrix (MA) protein yielded intermediate values: 84.1% identity and strong similarity and 13.6% difference. Several small gaps and inserts within A and C gag sequences were observed as well. We considered gaps to be amino acid residues missing within subtype A and C gag polyproteins when compared with subtype B gag, used as a reference sequence. Inserts are defined as extra amino acid residues within subtype A and C gags when compared with subtype B gag. The largest gaps (3 amino acid residues) were found in both subtype C and A gag polyproteins and they were located in the close vicinity of the MA/CA and p2/NC cleavage sites. We also noticed that there is a duplication of the FLQSRPEPTAPP sequence within p6 of subtype B gag polyprotein. It has been shown that the PTAP motif binds the host cellular factor TSG101 and is required for efficient virus release (Gottlinger et al. 1991; Huang et al. 1995; Yu et al. 1998; Garrus et al. 2001; Demirov et al. 2002).

It is worth mentioning that p1 protein in all three subtypes has all the Pro residues well conserved. Recent studies showed that these residues are important for the regulatory role of p1 in maintaining the viral infectivity (Hill et al. 2002).

**Cleavage Site Analysis**

There are 5 cleavage sites within gag polyprotein: MA/CA, capsid-p2 (CA/p2), p2/NC, nucleocapsid-p1 (NC/p1), and p1-p6 (p1/p6) (Figure 6-3). The cleavage sites were defined as being 10 amino acid residues long, with 5 amino acids on each side of the scissile bond. Two of the cleavage sites, CA/p2 and NC/p1, are very well conserved among subtype A, B, and C gag polyproteins, having 100% sequence identity. The MA/CA cleavage site is identical between subtype B and C gag polyproteins and differs in only one amino acid residue at position P5 (serine → valine) in subtype A gag. The p2/NC cleavage site differs by three residues, located at P3, P4, and P5, while the p1/p6 site differs by three residues as well, but these are situated on the carboxyl side of the cleavage site, at P1’, P3’, and P5’ positions.
In Vitro Processing Studies of HIV-1 Subtype B, A, and C Gag Polyproteins

To study the influence that the naturally occurring polymorphisms have on the rate and order of cleavage in subtype A, B, and C gag and gag/pol polyproteins processing, we employed an assay in which labeled gag precursors is processed sequentially in vitro after the addition of HIV-1 PR (Erickson-Viitanen et al. 1989; Krausslich et al. 1989; Partin et al. 1990; Tritch et al. 1991; Pettit et al. 1994).

The newly synthesized proteins are labeled with radioactive $^{35}$S which allowed visualization by autoradiography. The bands were recorded using a Molecular Dynamics PhosphorImager Storm 860, and the intensity of each band was measured with ImageQuant software (GE Healthcare). The calculations were made based on the known composition of the proteins and the number of labeled methionine (Met) residues of each protein. The total intensity of all the bands of interest in a given lane was considered 100 and the amount of each product was calculated as percentage of the total amount of labeled substrate in the lane.

Expression of the HIV-1 subtype B, A, and C gag polyproteins in an in vitro rabbit reticulocyte lysate transcription-translation system resulted in four major bands with sizes of 55, 41, 25, and 15-17 kD observed on the autoradiographs. The 55 kD protein corresponds to the gag precursor. The other 3 bands correspond to the intermediate cleavage products: MA-p2, CA-p2, and NC-p6$^{\text{GAG}}$. In the time frame of our experiments, we were not able to detect the synthesis of entire gag/pol polyprotein (p160) indicating that no active PR, able to carry out the cleavage of gag and gag/pol polyproteins, was synthesized. The transcription-translation reaction was followed for 5 hours. The p55 polyproteins continue to accumulate but no processing intermediate products were observed (Figure 6-4). The processing of gag polyproteins was accomplished by addition in trans of the active PR that was separately expressed in our bacterial system, purified and activated as described previously in Material and Methods chapter, page 63.
These experiments yielded proteins with a large range of sizes, from 55 kD to less than 2 kD. To be able to capture the entire picture we used a gradient (4-20%) gel that allows for a good separation. The top layers, with a lower percent of polyacrylamide, trapped and resolved the higher molecular weight proteins while the bottom layers with an increased concentration of polyacrylamide allowed for separation of smaller protein. The gels used for this study are denaturant gels so the proteins migrate according to their monomeric molecular weight.

*In trans* processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype B protease

The addition *in trans* of the active HIV-1 subtype B PR resulted in ordered processing of the subtype A, B, and C gag (p55) precursors (Figure 6-5). As expected, the initial cleavage was observed between the NC and p2 proteins and yielded two intermediates: MA-p2 (p41) and NC-p6\textsuperscript{GAG} (p15). The cleavage of this site proceeded at an increased rate with less than 5% subtype B p55 remaining after 5 min of processing. The cleavage at the corresponding site within subtype C gag polyprotein occurred at a slightly faster rate, with no p55 remaining after 5 min.

The second cleavage occurred at the site between the MA and CA (Pettit et al. 2003) and the cleavage product CA-p2 (p25) can be easily visualized as a intense band of about 25 kD size for all three subtypes. The production of p25 increased steadily during the *in vitro* transcription-translation reaction. However, the amount of p25 was markedly different between the three subtypes analyzed in this study. The p25 quantities for subtypes B and C were roughly similar, with about 10% more subtype C p25 than subtype B p25 after one hour of processing. Interestingly, the production of subtype A p25 occurred at a slower rate indicating a reduced cleavage rate at the MA/CA cleavage site. The densitometric analysis showed that, after one hour of processing, the amount of subtype A p25 was 15-fold less than that of subtype B or C p25.
The subtype A and B MA proteins do not contain Met residues so they cannot be identified on the gels. However, the subtype C MA protein contains one Met residue and it can been seen as a faint 17 kD band after 1 min and as a more intense band at the end of the experiment (60 min).

The subtype B NC-p6[^GAG] intermediate has a duplication of the LQSRPEPTAPP sequence within the p6[^GAG]. Consequently the subtype B NC-p6[^GAG] intermediate migrates on the gel at a slightly different position when compared to the intermediate in subtypes A and C. Worth mentioning is that the NC-p6[^GAG] intermediate products for all three subtypes steadily decreases over time, probably due to the HIV PR-mediated cleavage at the p1/p6[^GAG] cleavage site. However, the rates of cleavage at this site seemed to be different between the three subtypes studied here, with the rate of cleavage for subtype C intermediate slightly higher than for subtypes A and B. Also, the amounts of subtype C NC-p6[^GAG] intermediate synthesized at both the beginning and the end of the experiment are lower when compared with subtypes A and B. Due to their low molecular weight and lack of Met residues, we have not been able to follow the accumulation of the cleavage products at the p1/p6[^GAG] cleavage site: NC-p1 and p6[^GAG].

**In trans processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype A protease**

The HIV-1 subtype B, A, and C gag polyproteins were also cleaved *in trans* by addition of HIV-1 subtype A PR. The pattern of cleavage is similar to that observed after the addition of subtype B PR. However, the cleavage rates at the first cleavage site are slightly lower when compared to those performed by subtype B PR, shown by the retarded disappearance of p55, as seen on densitometric analysis (Figure 6-6). Despite this fact, the final amounts of p25, after one hour of processing, were 3-5-fold higher when compared with those produced by subtype B PR-mediated gag cleavage. Interestingly, the production of subtype A p25 was still lower when
compared with the amount of subtype B or C p25. However, the quantitative difference was only 4-fold less subtype A p25 than subtype B or C p25 compared with 15-fold difference observed in the subtype B-PR mediated cleavage.

**In trans processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype C protease**

The cleavage of all three different gag polyproteins by HIV-1 subtype C PR occurred in a similar pattern to cleavages mediated by the other enzymes, subtype A and B PRs, with a reduced production of subtype A p25 (12-fold lower) when compared with subtype B and C p25.

**Analysis of MA/CA Cleavage Site**

The MA/CA cleavage site appears to be among the sites cleaved immediately after the first cleavage event. The P5 position in this site harbors a valine (Val) residue for subtype B and C gag polyproteins and a serine (Ser) residue for subtype A gag. Furthermore, when an alignment of the cleavage site is performed, a gap of 2 and 3 residues for subtype A and C gag polyproteins, respectively is noted (Figure 6-3).

**In Vitro Processing Studies of HIV-1 Subtype A and B Gag Polyproteins – 124 and QV Variants**

The mutants for both subtype A and B gag polyproteins were engineered using Site Directed Mutagenesis technique. The order and the rates of gag processing were studied after the addition of HIV-1 subtype B PR in the reaction mixture, as described above.

**In trans processing of HIV-1 subtype A gag polyprotein – S124V and QV variants – by HIV-1 subtype B protease**

The first step was to mutate the MA/CA cleavage site within subtype A gag polyprotein to engineer a site identical to that in subtype B gag. Because only one residue was different between subtype A and B gag polyproteins, one change was introduced: the Ser found at position 124 in subtype A gag was mutated to the corresponding amino acid in subtype B – Val; the
mutant was named AgagS124V. We decided on another approach as well and engineered a two-
residue insert (glutamine -Val) that would cover the gap observed in the subtype A gag when
compared to subtype B gag, at the amino-terminus position of the MA/CA cleavage site; the
mutant was named Agag_QV.

The introduction of the S124V mutation at the P5 position of the MA/CA cleavage site
increased the rate of processing at this site by a factor of 2, with twice as much p25 produced
when compared to V124 subtype A gag (Figure 5-8). A similar effect was observed for
Agag_QV mutant. However, the amount of p25 generated by these two mutants is not equivalent
to that of subtype B gag, both being about 7-fold less.

In trans processing of HIV-1 subtype B gag polyprotein – V124L, V124S and ΔQV variants
– by HIV-1 subtype B protease

We also decided to mutate the MA/CA subtype B cleavage site to mimic the site in
subtype A gag polyprotein. We followed the same steps as when we engineered the subtype A
gag mutants but backwards. We constructed three variants: (1) the amino acid Val at position
124 was substituted by Ser (BgagV124S); (2) two residues, glutamine (Gln) and Val, located at
the amino-terminus position at the MA/CA cleavage site were removed (Bgag_ΔQV); (3) the
amino acid Val at position 124 was substituted with Leu (BgagV124L). The last mutant harbors
a conservative change and it would be useful to further analyze the importance of a hydrophobic
amino acid at the P5 position of the cleavage site.

The V124L mutations did not appear to affect the rate of cleavage at the MA/CA site,
while the V124S decreased the amount of p24 by 6-fold (Figure 6-9). The deletion of QV
dipeptide had similar results. However, the production of p24 is at least 2-fold higher than in
subtype A gag polyproteins harboring similar changes at the N-terminus of the MA/CA site.
Discussion

To our knowledge there are no other studies analyzing the gag and gag/pol processing events in the subtypes A and C. This work represents the first attempt in studying and characterizing the effect of naturally occurring polymorphisms located at the cleavage and non-cleavage sites in gag and gag/pol polyproteins. In our study we aim to evaluate the phenotypic impact of gag baseline polymorphisms on HIV-1 gag processing events.

The naturally occurring polymorphisms within the gag polyprotein can be classified as cleavage site and non-cleavage site polymorphisms. The former ones have received a great deal of attention and many studies have demonstrated that some of the amino acid variations within the cleavage site of gag are baseline polymorphisms while others develop in response to PI treatment and evolution of a catalytically impaired PR. It has been previously noted that some of the most frequently mutated positions under PI therapy also happen to be more polymorphic, such as positions P2 to P5 and P3′ of the p2/NC cleavage site or position P3 of the p66/INT site (Cote et al. 2001). One study suggested that, in subtype B gag polyprotein, the development of gag cleavage site mutations is associated with heavily mutated PRs (“dead end”) for which the concomitant evolution of additional mutations in the PR and in the gag substrate may be the only way for the virus to survive in an increasingly selective environment (Doyon et al. 1996; Maguire et al. 2002). More recently, the analysis of resistant viral isolates from IDV-treated patients indicated gag adaptation as a common evolutionary pathway, taking place as early as 6 weeks after the start of therapy and in the presence of as few as two PR mutations (Zhang et al. 1997). These findings emphasize the relevance of cleavage site mutations in the evolution of HIV resistance to PIs. Therefore, it was proposed that the therapy-associated cleavage site mutations should be considered in HIV resistance tests to estimate viral fitness in different clinical settings (Verheyen et al. 2006).
The non-cleavage site polymorphisms have also been indicated as possible mechanisms of adaptation for compensating for the reduced catalytic activity of mutant PRs. Moreover, certain amino acid substitutions in non-cleavage sites have been shown to contribute to the development of high levels of viral resistance to multiple PIs. These data strongly suggest that non-cleavage site amino acid substitutions in the gag protein recover the reduced replicative fitness of HIV-1 caused by mutations in the viral PR and may open a new avenue for designing PIs that resist the emergence of PI-resistant HIV-1 (Gatanaga et al. 2002). Such non-cleavage site gag mutations should render the polyprotein cleavage sites more accessible to the PR, make polymerization of viral proteins more efficacious, and/or make assembly and disassembly more efficient.

Alteration(s) of other unknown functions of gag proteins may also contribute to the HIV-1 acquisition of resistance to PIs, but it appears that HIV-1 resistance to PIs is acquired with multiple mechanisms (Gatanaga et al. 2002). Previous studies observed for the first time K416N and E399V substitutions in the NC protein, and E455A and E455V substitutions in the p6\textsuperscript{GAG} protein in resistant viruses suggesting that there are additional possibilities for gag adaptation besides the previously described substitutions in the cleavage sites surrounding the p2, NC, p1 and p6\textsuperscript{GAG} peptides (Mammano et al. 1998).

Analysis of consensus and individual sequences of the non-B subtype gag polyproteins showed an increased number of polymorphic residues when compared with subtype B gag (Los Alamos 2007). The subtype A and C clinical isolates analyzed in this study showed cleavage site sequences that sometimes differed considerably from the “consensus sequences” in subtype B gag after which HIV PR peptide substrates are often designed (Figure 6-3). None of the cleavage site polymorphisms present at the cleavage site that have been previously described as drug resistance-associated mutations were observed within our HIV-1 subtype A and C gag
sequences. Also these clones do not harbor any of the previously described non-cleavage site substitution, except E481K, a baseline polymorphism in subtype C gag polyprotein, that appears to enhance the replicative capacity of the virus (ref), and L496P, a polymorphism in subtype A gag, that arises as a post-PI treatment mutation in patients harboring HIV-1 subtype B (Ho et al. manuscript in preparation).

The subtype A and C gag polyproteins analyzed in this study also harbored several small gaps and inserts (Figure 6-2). Two small gaps are located at the amino-terminus of the MA/CA cleavage site, in the carboxyl-terminus of the MA. Several groups determined the structure of the MA protein through NMR analysis (Massiah et al. 1994; Matthews et al. 1994). They observed that the C-terminus 20 amino acid residues do not adopt a rigid conformation in the solution. It appears that the C-termini of the monomers form helical arms that fit into a cleft between the helices A and C of the adjacent monomer, a common feature in the viral architecture (Stuart et al. 1993). Consequently, the small gaps within this region do not interfere with a secondary structure but might affect the MA/MA interactions, that are important for both gag and gag/pol dimerization and for the formation of the viral MA shell. Another gap is located within the p2/NC cleavage site of subtype C gag, at the N-terminus of the p2 protein. However, as discussed later in this chapter, there appear to be no consequences of this gap on the rate of the order of gag processing.

The FLQSRPEPTAPP duplication was observed within subtype B p6\text{GAG} protein. The influence of this duplication on the results of our in vitro experiments is not clear. Previous studies showed that the complete or partial duplication of PTAP appears to provide an advantage for survival of the virus in the presence of NRTIs and PIs (Ibe et al. 2003). Insertions such as the PTAP repeat or SRPE in p6 reduce the enzymatic activity of the wild-type HIV-1 PR, while
restoring gag processing and subsequent replication of protease inhibitor-resistant variant (Tamiya et al. 2004). A truncated p6 did not inhibit processing of p160 gag/pol in COS-7 cells, while Demirov et al. showed that in certain T cell lines, mutation in the PTAP motif did cause processing defects (Yu et al. 1998; Demirov et al. 2002). All of these studies were performed in a cellular system where the budding and releasing of whole virions was analyzed. In the cell-free context of our experiment set-up, we can follow neither the assembly nor the budding. We believe the influence of the duplication over the rate and order of the processing is minimal, if any, in our system.

We analyzed the processing events in gag polyprotein cleavage by the PR from the same subtype, but also cleaved by the other two subtype PRs. This kind of “subtype interplay” would occur in the case of a recombinant form, where two different subtypes recombine and form a virus with a new genetic makeup (Takeb et al. 2004). This approach enabled us to evaluate the cleavage that occurs in the recombinant forms and would also bring information about compatibility between combinations of naturally occurring polymorphisms within the PR and within the gag polyproteins.

We observed that the order of gag cleavage is similar between all three HIV-1 subtypes analyzed in this study. As expected, upon addition of active HIV PR \textit{in trans}, the first cleavage occurred at the p2/NC cleavage site closely followed by MA/CA cleavage. When the same gag polyprotein was cleaved \textit{in trans} by PRs from different subtypes the pattern of cleavage was highly similar. However, the same PR rendered different rates of cleavage for different cleavage sites. It appears that subtype B PR cleaves the p2/NC site well and has a slightly lower MA/CA cleavage rate. Subtype A and C PRs exhibited a relatively slower cleavage rate for p2/NC when compared to subtype B PR, but they showed an increased MA/CA cleavage rate.
Independent of the PR added *in trans*, the p2/NC cleavage appears to occur at a reduced rate for subtype A gag when compared to B and C gag polyproteins. The p2/NC cleavage site is very polymorphic among these three gag polyproteins, each having a unique combination at the P3-P5 positions. In the same time, we observed a marked reduction in the amount of p25 produced by *in trans* processing of subtype A gag. Even when processed by the subtype A PR, the amount of subtype A p25 is 4-fold less than the amount of p25 that resulted from the cleavage of subtype B gag polyprotein by the subtype A PR. The slightly decreased p2/NC cleavage rate in subtype A gag polyprotein might be due to several factors: (1) the baseline polymorphisms at the p2/NC site, (2) the baseline non-cleavage site polymorphisms within p2, NC or the other structural proteins within subtype A gag, and/or (3) negative feed-back from the MA/CA site. With so many baseline variations it might be difficult to pinpoint the exact source of this difference in the p2/NC cleavage rate among subtypes, and probably would not be feasible to pursue this avenue due to the fact that such small difference might not have significant influence on the virus replicative capacity or PI treatment susceptibility. However, we believe that it is important to mention that even if small, these differences in cleavage rates might increase upon acquisition of other ARV treatment-induced mutations, and the overall effect on the virus could be significant.

As mentioned above, the cleavage of the MA/CA within subtype A gag proceeded at a much slower rate when compared with those for subtype B and C gag polyproteins. Upon a closer analysis of the amino acid sequence, we noticed that there is one residue difference between subtype A and the other two subtypes: the Val residue found at the P5 position within subtype B and C is mutated to a Ser in subtype A gag.
In order to understand the role of this cleavage site polymorphism on the cleavage rate at MA/CA site, we created a subtype B MA/CA cleavage site within subtype A gag. Our rationale was that if the cleavage site sequence has the dominant influence on the cleavage rate, the rate of cleavage of the mutated subtype A MA/CA site would be comparable with those the subtype B. If there are other determinants within gag polyprotein, most likely the rates of cleavage at the MA/CA site would be different between these two subtypes. We engineered two mutants designated AgagS124V and Agag_QV, where the Ser was mutated to a Val, and the Gln-Val dipeptide was inserted at the N-terminus of the cleavage site (Figure 6-3). The in vitro transcription-translation analysis was conducted as previously described, and the in trans cleavage was performed by the subtype B PR. The amount of p25 increased by 2-fold for the subtype A gag mutants but did not equal the p25 production upon cleavage of subtype B gag. We believe that this might be due to non-cleavage site polymorphisms present within gag polyprotein.

We also engineered a subtype A MA/CA cleavage site within subtype B gag. We obtained three mutants BgagV124L, BgagV124S and Bgag_ΔQV, as explained in the Results section of this chapter. The production of p25 was not affected by the V124L mutation (Figure 6-9), but the cleavage rate at the MA/CA cleavage site decreased significantly upon introducing V124S or deleting the Glu-Val dipeptide. After 90 min, the amount of p25 for BgagV124S was twice as much as for subtype A gag and 3-times less when compared to subtype B gag. The p25 amount in Bgag_ΔQV mutant was slightly greater than for BgagV124S.

These results point to the fact that the baseline polymorphisms within gag polyprotein play a dominant role in modulating the rate of cleavage. In this case, it appears that rate of cleavage is modulated by both the cleavage and the non-cleavage site polymorphisms. Our results indicate
that the presence of hydrophobic residue at position P5 of the MA/CA site appears to be a requirement for an enhanced cleavage rate. These results add to previous studies that suggested that the MA/CA site might be more sensitive to mutations than other processing sites (Margolin et al. 1990; Partin et al. 1990; Pettit et al. 1994). The importance of the non-cleavage site polymorphisms is underlined by the fact that even after we introduced a MA/CA cleavage site identical with that in subtype B, the amount of p25 produced by processing of subtype A gag is still much lower than for subtype B. Also, when a subtype A MA/CA cleavage site was engineered in subtype B gag the drop in the cleavage rate was not to the level encountered in subtype A gag. These results also point not only to the effect of cleavage and non-cleavage site polymorphisms but also to a probable cooperative effect among different regions in gag.

A better understand of the role of non-cleavage site polymorphisms could possible come from crystallographic or NMR studies of gag polyprotein which could provide us with a detailed map of the interactions and structural arrangements among different components of gag. Such information would be useful in designing novel therapeutic approaches to disrupt these interactions.

An interesting finding was that the sequence analysis through the Los Alamos Database showed that the consensus sequence of subtype A sub-subtype A2 contains a serine at the same position. Sequence analysis of the rest of the gag and gag/pol genes and proteins of the subtype A clone analyzed here revealed that it belongs to subtype A sub-subtype A2 group. This sub-subtype is confined to small region in the western part of Africa, and is not frequently found in any other part of the world. There might be a connection between a slower processing rate and the reduced number of HIV-positive persons infected with this sub-subtype but many other factors, such as ethnicity, geographically confined population groups, and access to care have to
be taken in consideration. Also, this in vitro assay of gag processing cannot replace phenotypic studies on the replicative capacity of the virus harboring this polymorphism. It is possible that, due to other compensatory factors not assayed here, the replicative capacity of the virus is not to be affected at the same extent as seen with the in vitro assay.

PR processing activity depends on (1) the primary amino acids sequence of PR, (2) the composition of the cleavage sites, (3) determinants in gag, and (4) the accessibility to the site of the PR (Pettit et al. 1994; Goodenow et al. 2002). We showed that determinants in gag could play a dominant role in modulating the rate of cleavage. This research and previous studies support the sequential, ordered substrate cleavage by the HIV-1 PR. The idea of strict requirements for gag processing is also supported by the fact that, despite many differences between the three subtypes studied here, the order and, largely, the rate of cleavage are comparable. Further mutational and structural studies would bring more insight and would pinpoint the key residues involved in coordinating the processing events. Also future studies where PI will be added in the reaction mixture should answer the question whether this naturally occurring polymorphism represents a mechanism through which the virus can acquire faster drug resistance.

The validity of these experiments is limited by the extent to which this in vitro processing system mimics the precursor cleavage that accompanies virus assemble in vivo (Pettit et al. 2004). Several studies support the theory that the spontaneous processing of in vitro-translated precursors is an appropriate model for processing in vivo (Carter and Zybarth 1994; Platt and Haffar 1994; Zybarth and Carter 1995). Several groups (Moody et al. 1995; Spearman and Ratner 1996; Lee et al. 1999) have observed that gag or pol precursors produced in rabbit reticulocite lysate (RRL) multimerize into higher-order components similar to those observed in
virions. These studies indicate that the concentration of gag/pol precursors in this system is high enough to support multimerization. Other lines of evidence suggest that the intermediates and final products that arise during sequential processing of gag polyprotein with this assay correlate well with the processing events observed in infected cells (Mervis et al. 1988; Gowda et al. 1989). Others (Platt and Haffar 1994; Hermida-Matsumoto and Resh 1999) have noted efficient membrane association of translated gag in RRL. Campbell et al. found that components within the reticulocyte lysate are required for proper multimerization of purified HIV gag into 110-nm particles in vitro and later linked that requirement to specific phospholipids (Campbell and Rein 1999; Campbell et al. 2001).

The correct interpretation of the bands seen on the autoradiographs was confirmed by Western blot analysis (data not shown). Despite the high similarity in the amino sequence of the CA protein among all HIV-1 subtypes, the monoclonal antibodies against subtype B that we initially used to identify the cleavage products containing p24 (NIH AIDS reagent program) did not bind to subtype C p24. However, a second polyclonal antibody did recognize the subtype C p24.

One of the questions we asked was why we were not able to observe processing in the absence of a forced frameshift mutation. It has been well documented that the frameshift occurs in 5% of translational events and is the result of the ribosomal slippage along UUUUUUA sequence (Jacks et al. 1988). Our initial expectations were that we would be able to document processing without addition of PR in trans or engineering a forced frameshift mutation. However, even after allowing the reaction to proceed for as long as 3 hours, no processing was observed either by visually inspecting the autoradiographs or by performing densitometric analysis (Figure 6-4). Several previous studies showed that processing in vivo occurs in a limited...
space (the interior of the virion) where the concentration of gag and gag/pol, and so of the HIV-1 PR embedded within gag/pol, is high. This allows for efficient dimerization and thus activation of the PR and for the initiation of gag/pol cleavage. This spatial limitation and the high number of gag and gag/pol polyproteins directed at the cell membrane are the key factors in promoting processing and maturation of the virions. Even if we assume that the frameshift occurs with the same frequency in our system, the concentration of gag/pol might be too low to promote the efficient dimerization of the protease monomers. This tentatively explains how the processing events happen upon addition of an increased concentration of PR.

Despite its limitations, this *in vitro* system has several advantages: (1) it is significantly faster than any other expression-purification systems using prokaryotic or eukaryotic cells; (2) it does not involve the generation of live HIV; therefore it does not require special biosafety facilities; (3) it is limited to the mechanistic relationship of PR and gag or gag/pol, thus eliminating variable factors that would complicate even more the interpretation of the results.

The present data, taken together, suggest that HIV-1 resistance to PIs is associated with primary and secondary mutations in the viral PR and is also associated with the cleavage site amino acid substitutions in gag together with substitutions at non-cleavage sites. We believe it is a necessity to examine regions within gag polyprotein for predictive sequences that, in combination with naturally occurring polymorphisms or ARV-induced mutations in PR, will enhance genotype and drug susceptibility predictions and allow customization of therapeutic regimens, especially needed in patients failing multiple anti-HIV drug regimens.
Figure 6-1. HIV-1 gag processing. *In cis* cleavage occurs when the HIV-1 PR cuts within the same gag/pol polyprotein in which it is embedded, and *in trans* cleavage occurs when the PR cleaves the gag polyprotein.
Figure 6-2. Alignment of the gag polyproteins of HIV-1 subtype B, C and A. The near full-length clones of gag/pol gene from subtypes A, B and C were obtained from NIH AIDS Reference and Reagent Program. The alignment was performed with ClustalW (Expasy Proteomics Server).
Figure 6-2. Continued
Figure 6-3. Alignment of the cleavage sites within gag polyproteins of HIV-1 subtypes, B, C, and A.
Figure 6-4. The *in vitro* processing without the frameshift and without the addition *in trans* of the HIV protease. Protein marker is shown in the left and the gag polyproteins in the right of the gel. The time points are indicated in the top of the gel.
Figure 6-5. *In trans* processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype B protease. A) Autoradiographs. The time points are indicated in the top of the gel. The protein marker is in the left and the processing intermediates are in the right of the gel. B) Densitometric Analysis. The time indicated on the x-axis is shown in minutes. The legend is: **p55**, **p41**, **p24**, **p17**, **p15**.
Figure 6-6. *In trans* processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype A protease. A) Autoradiographs. The time points are indicated in the top of the gel. The protein marker is in the left and the processing intermediates are in the right of the gel. B) Densitometric Analysis. The time indicated on the x-axis is shown in minutes. The legend is: p55, p41, p24, p15.
Figure 6-7. *In trans* processing of HIV-1 subtype B, A, and C gag polyproteins by HIV-1 subtype C protease. A) Autoradiographs. The time points are indicated in the top of the gel. The protein marker is in the left and the processing intermediates are in the right of the gel. B) Densitometric Analysis. The time indicated on the x-axis is shown in minutes. The legend is: p55, p41, p24, p15.
Figure 6-8. *In trans* processing of HIV-1 subtype A gag polyprotein – S124V and QV variants – by HIV-1 subtype B protease. A) Autoradiographs. The black triangles in the top of the gel indicate 7 time points from 0 to 90 min after subtype B PR addition. The protein marker is in the left of the gel. B) Densitometric Analysis. The time indicated on the x-axis is shown in minutes. The legend is: p55, p41, p24, p15.
Figure 6-9. In trans processing of HIV-1 subtype B gag polyprotein – V124S and ΔQV variants – by HIV-1 subtype B protease. A) Autoradiographs. The black triangles in the top of the gel indicate 5 time points from 0 to 90 min. The protein marker is in the left of the gel. B) Densitometric Analysis. The time indicated on the x-axis is shown in minutes. The legend is: p55, p41, p24, p15.
Figure 6-10. HIV-1 matrix protein (PDB code 1TAM). It has 5 $\alpha$-helices, A to E, (shown in red) connected through unstructured coils (shown in green). The last 20 amino acids at the C-terminus adopt a flexible coil-like conformation, important for MA-MA dimerization. Figure rendered with Pymol (DeLano Scientific).


Baeten D (2007) HIV-1 subtype D infection is associated with faster disease progression than subtype A, in spite of similar HIV-1 plasma viral loads. In: 14th Conference on Retroviruses and Opportunistic Infections, Los Angeles


Buckman JS, Bosche WJ, Gorelick RJ (2003) Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. J Virol. 77:1469-1480


Caride E, Hertogs K, Larder B et al. (2001) Genotypic and phenotypic evidence of different drug-resistance mutation patterns between B and non-B subtype isolates of human immunodeficiency virus type 1 found in Brazilian patients failing HAART. Virus Genes 23:193-202


Centlivre M, Sommer P, Michel M et al. (2006) The HIV-1 clade C promoter is particularly well adapted to replication in the gut in primary infection. Aids. 20:657-666


Dragic T, Litwin V, Allaway GP et al. (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature. 381:667-673


Heaslet H, Rosenfeld R, Giffin M et al. (to be published) Conformational flexibility in the flap domains of ligand free HIV protease. To be published


Kondo E, Gottlinger HG (1996) A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr. J Virol. 70:159-164


Loemba H, Brenner B, Parniak MA et al. (2002) Genetic divergence of human immunodeficiency virus type 1 Ethiopian clade C reverse transcriptase (RT) and rapid development of resistance against nonnucleoside inhibitors of RT. Antimicrob Agents Chemother 46:2087-2094


Renjifo B, Gilbert P, Chaplin B et al. (2004) Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. Aids. 18:1629-1636


Rose RE, Gong YF, Greytok JA et al. (1996) Human immunodeficiency virus type 1 viral background plays a major role in development of resistance to protease inhibitors. Proc Natl Acad Sci U S A 93:1648-1653


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


Soares MA, De Oliveira T, Brindeiro RM et al. (2003) A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. AIDS 17:11-21


Tomasselli AG, Sarcich JL, Barrett LJ et al. (1993) Human immunodeficiency virus type-1 reverse transcriptase and ribonuclease H as substrates of the viral protease. Protein Sci. 2:2167-2176


Xie D, Gulnik S, Gustchina E et al. (1999) Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. Protein Sci. 8:1702-1707


206


BIOGRAPHICAL SKETCH

Roxana Maria Coman was born in August 1976 to Victoria and Florin Coman, and grew up in Barlad, Vaslui, Romania, with her sister Anca Cristina. She received her MD from Carol Davila University of Medicine and Pharmacy, Bucharest, Romania, in September 2000. Roxana entered the Interdisciplinary Program in Biomedical Sciences at the College of Medicine, University of Florida in the summer of 2003. One year later she joined the laboratory of Dr. Ben M. Dunn and began training towards her doctorate. After her graduation, she plans to pursue a dual career as a clinician and researcher.