The Role of the Protease Cleavage Sites in Viral Fitness and Drug Resistance in HIV-1 Subtype C

by

Jennifer Giandhari

Submitted in fulfilment of the requirements for the degree of
Master of Medical Science in Molecular Virology
In the
Faculty of Health Sciences
University of Kwa-Zulu Natal
2010

PREFACE

The experimental work described in this dissertation was carried out in the Hasso Plattner Research Laboratory, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, from September 2008 to July 2010 under the supervision of Dr Michelle Lucille Gordon.

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any other University. Where use has been made of the work of others, it is duly acknowledged in the text.

Signed: _____ Date: _____

Jennifer Giandhari (Candidate)

Signed: _____ Date: _____

Dr Michelle Lucille Gordon (Supervisor)

Plagiarism:

Declaration

I …….declare that

(i) The research reported in this dissertation, except where otherwise indicated, is my original work.

(ii) This dissertation has not been submitted for any degree or examination at any other university.

(iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

(a) their words have been re-written but the general information attributed to them has been referenced;

(b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(v) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

(vi) This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the Reference sections.

Signed: _____ Date: _____

Jennifer Giandhari (Candidate)

Signed: _____ Date: _____

Dr Michelle Lucille Gordon (Supervisor)

Ethical Approval

Full ethical approval, from the Biomedical Research Ethics Committee of the Nelson R. Mandela School of Medicine, University of KwaZulu-Natal (ref: BF 068/08), and informed patient consent was obtained for this study.

Presentations

Part of this work was presented at the 10th Annual Symposium on Antiviral Drug Resistance, held in Viginia, USA on November 15-18, 2009. A poster was presented titled*: Characterisation of Protease Cleavage Sites in Protease Inhibitor-Naïve Patients Infected with HIV-1 Subtype C.*

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Michelle Gordon. With her understanding, inspiration, and great efforts to explain things clearly and simply, she provided me with a priceless foundation for my career. I am also deeply grateful for the encouragement, sound advice, great ideas and good company which she has provided me with during my thesis-writing period.

I would like to express my deep and sincere gratitude to Professor Thumbi Ndung'u. His wide knowledge and important support throughout this work have been of great value to me.

My sincere thanks are due to Mrs Taryn Green for her assistance in the lab and advice throughout my degree.

I would like to thank Mrs Keshni Hiramen, Ms Jaclyn Wright, Ms Vanessa Naidoo and Mr Ravesh Singh for their research assistance.

I wish to thank my friends at university, for their loving support and for providing a stimulating and fun environment in which to learn and grow. A special thanks to Dshanta and Kavidha for their advice, friendship and support which they provided.

My special thanks are due to my brother Ashley, sister-in-law Prayashnee and grandmother for their loving support.

My loving thanks are due to my husband Shane for his love, continuous support and for always believing in me.

This thesis is dedicated to my parents, without whom none of this would have been possible. I will always be grateful for their support, unconditional love and encouragement. This thesis is also in remembrance of Romeo and my late grandmother Yvonne.

I thank God for the strength to persevere and for making all things possible.

Abstract

There is an increasing number of patients failing second line highly active antiretroviral therapy (AZT, DDI and LPV/$_r$) in South Africa, where HIV-1 subtype C predominates. Mutations at *gag* cleavage sites (CS) have been found to correlate with resistance mutations in protease (PR). Therefore, it is important to collect data on subtype C *protease* and *gag* sequences from patients as these mutations may affect the efficacy of protease inhibitor (PI) containing drug regimens.

In this study, 30 subtype-C infected second-line failures were genotyped using the ViroSeq^{TM} resistance genotyping kit and the *gag* region from these isolates were then characterised. These sequences were then compared to 30 HIV-1 subtype C infected first-line failures (PI-naïve) and subtype B, C and group M naïve sequences that were downloaded from the Los Alamos Sequence Database. Amino acid diversity at the CS was measured using Mega version 4.0. To investigate the effect of CS mutations on replication capacity, a mutation was introduced by site-directed mutagenesis (Stratagene's QuikChange Site-Directed Mutagenesis kit).

Of the 30 second-line failures that we genotyped, only 16 had resistance mutations in PR and 23 in *gag*. The most frequent major PI mutations were: I54V/L, M46I, V82A, and I84V and in *gag* CS were V390L/I and A431V. Interestingly the A431V mutation significantly correlated with protease mutations M46I/L, I54V and V82A. The virus carrying the A431V mutation *in vitro* was found to have a lower replication capacity compared to the wild type.

These findings emphasize the need for further investigation of *gag* mutations and their contribution to the evolution of HIV resistance to PIs.

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

ABC	-	Abacavir
AIDS	-	Acquired Immunodeficiency Syndrome
ARV	-	Antiretroviral
ATV	-	Atazanavir
AZT	-	Zidovudine
CS	-	Cleavage site
D4T	-	Stavudine
DEPC	-	Diethylpyrocarbonate
DDI	-	Didanosine
DLV	-	Delavirdine
DNA	-	Deoxyribnucleic acid
dNTP	-	Deoxynucleotide Triphosphate
DRV	-	Darunavir
DTT	-	Dithiothreitol
EFV	-	Efavirenz
ETV	-	Etravirine
FPV	-	Fosamprenavir
FTC	-	Emtricitabine
gp41	-	Glycoprotein 41
gp120	-	Glycoprotein 120
HAART	-	Highly Active Antiretroviral Therapy
HIV	-	Human Immunodeficiency Virus
HIV-1	-	Human Immunodeficiency Virus Type-1
IDV	-	Indinavir

IPTG	-	Isopropyl β-D-1-thiogalactopyranoside
KCl	-	Potassium Chloride
LPV	-	Lopinavir
LPV/r	-	Kaletra
MRCA	-	Most Recent Common Ancestor
mRNA	-	Messenger Ribonucleic Acid
MuLV	-	Moloney Murine Leukaemia Virus
NFV	-	Nelfinavir
NNRTI	-	Non-Nucleoside Reverse Transcriptase Inhibitor
NRTI	-	Nucleoside Reverse Transcriptase Inhibitor
NVP	-	Nevirapine
PCR	-	Polymerase Chain Reaction
PI	-	Protease Inhibitor
PR	-	Protease
RNA	-	Ribonucleic Acid
RNAse H	-	Ribonuclease H
RT	-	Reverse Transcriptase
RTI	-	Reverse Transcriptase Inhibitor
RTV	-	Ritonavir
SQV	-	Saquinavir
TAM	-	Thymidine Analog Mutation
TDF	-	Tenofovir
TPV	-	Tipranavir
UV	-	Ultraviolet
3TC	-	Lamivudine

CHAPTER 1

Introduction

1.1 HIV-1

1.1.1 The Epidemiology of HIV-1

The HIV/AIDS pandemic has an unequal distribution of HIV-1 infections worldwide, resulting in some regions being affected far more than others. There is also a disparate distribution of HIV-1 subtypes worldwide, with subtype C predominating in Southern Africa, as shown in Figure 1.1 below. In 2009, the number of people living with HIV-1 worldwide was estimated to be 33.3 million by the UNAIDS, while Sub-Saharan Africa accounted for 22.5 million of these infections. Sub-Saharan Africa accounts for 68% of persons living with HIV/AIDS worldwide and for 72% of all deaths due to AIDS (UNAIDS, 2010).

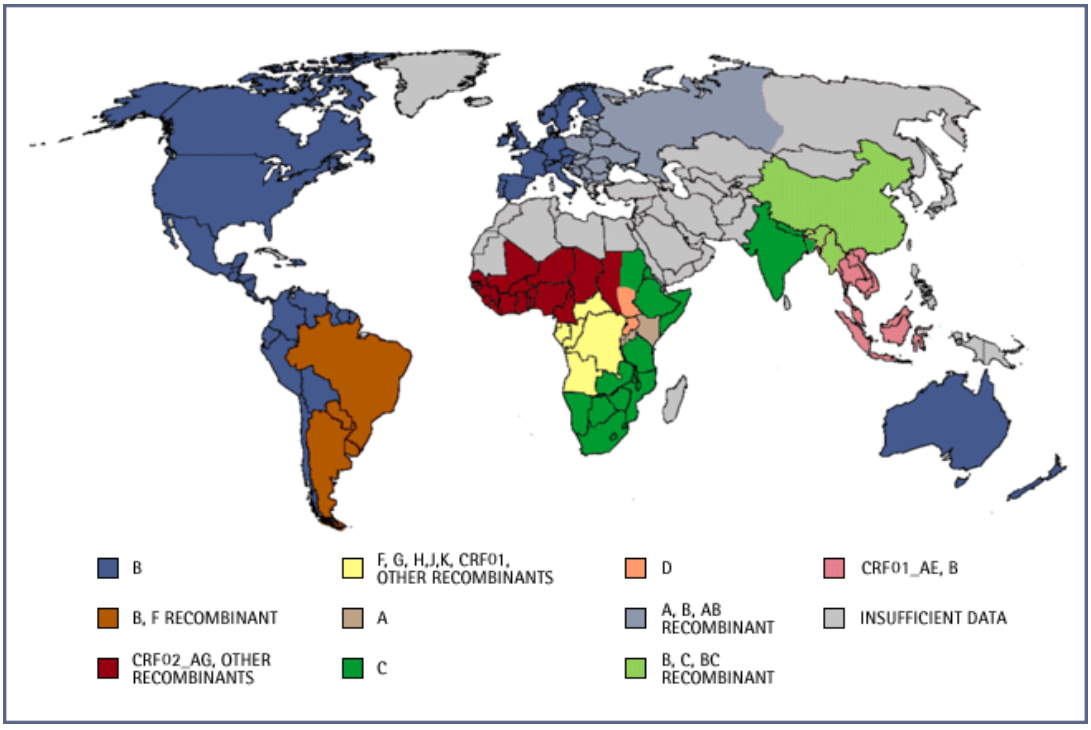

Fig 1.1 Global distribution of HIV-1. An overview of the geographical distribution of HIV-1 subtypes. This distribution shows the predominance of HIV-1 subtype C in Sub-Saharan Africa. Taken from (McCutchan, 2003).

1.1.2 The Classification of HIV-1

Human immunodeficiency virus (HIV) can be sub-divided into two groups i.e. HIV-1 and HIV-2, both which cause infection and disease in humans (Cohen et al., 2008). HIV-1 is thought to

have originated from cross-species transmission of a chimpanzee virus to humans (Hahn et al., 2000, Sharp et al., 2005), while HIV-2 is thought to have arisen from cross-species transmission of a Sooty mangabey virus to humans (Lemey et al., 2003). In comparison with HIV-1, HIV-2 is not as common and it is associated with a much slower progression to immunodeficiency (Cohen et al., 2008).

HIV-1 is characterised by extensive genetic diversity (Hemelaar et al., 2006). It is phylogenetically classified based on nucleotide sequences derived from multiple subgenomic regions (*gag, pol* or *nef*) from the same isolate or on full-length genome sequence analysis (Peeters, 2001). Phylogenetic analysis of the different HIV-1 strains have revealed that they can be subdivided into groups (M, N, O and P), subtypes, sub-subtypes and circulating recombinant forms (CRFs) (Santos and Soares, 2010).

Group M (main) accounts for the vast majority of strains found worldwide while group O (outlier) is endemic to Cameroon and neighbouring countries in West Africa, and group N (non-M/non-O) is relatively new and has only been identified in a few patients from the Cameroon (Santos and Soares, 2010). The recently discovered HIV-1 group P was first described in a Cameroonian woman and has not shown evidence of recombination with other HIV-1 subtypes (Santos and Soares, 2010). There are nine phylogenetic subtypes of group M (A-D, F-H, J and K). The DNA sequence of viruses in the distinct subtypes can differ by 15-20% (McCutchan, 2000). Certain subtypes contain groups of viral isolates that are genetically related sister clades (sub-subtypes), which are phylogenetically more closely related to each other than to other subtypes (Robertson et al., 2000). These include the A (sub-subtypes A1 and A2) and F (sub-subtypes F1 and F2) subtypes. The circulating recombinant forms are a result of recombination between subtypes within a dually infected individual, from whom the recombinant forms are

then passed to other people (Taylor and Hammer, 2008). Recombinant HIV-1 such as A/G recombinant strains are prevalent in West Africa while B/C recombinant strains are prevalent in China (McCutchan, 2000). Subtype B is known to predominate in the Americas, Europe and Australia while subtype C is most common in the most heavily affected part of the world, Southern Africa, as shown in Figure 1.1 (McCutchan, 2000, Buonaguro et al, 2007).

1.1.3 The Structure of the HIV-1

HIV-1 is a retrovirus belonging to the Lentivirus genus (Zhao et al, 2005). Similar to other retroviruses, HIV-1 is reverse transcribed to viral DNA by the enzyme reverse transcriptase. All lentiruses are enveloped by a lipid bilayer as shown in red in Figure 1.2. The exposed surface glycoproteins i.e. gp120, are anchored to the lipid membrane by the transmembrane protein gp41. A matrix shell containing copies of the matrix protein (Gag p17) lines the inner surface if the viral membrane, while a capsid core containing copies of capsid proteins (Gag p24) is located in the centre of the virus (Sierra et al., 2005).

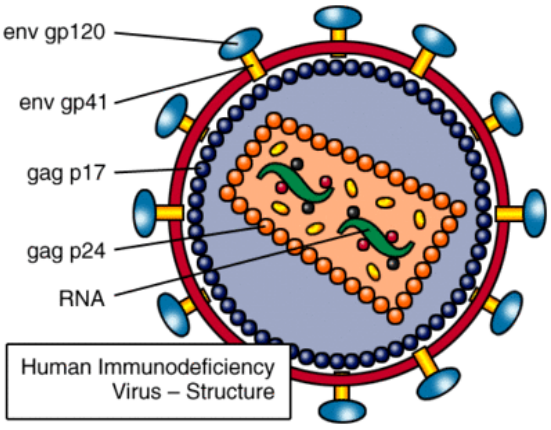

Fig 1.2 General features of the HIV virion. Taken from (Nielsen et al., 2005).

1.1.4 The Life Cycle of HIV-1

HIV-1 particles bind specifically to cells expressing CD4 receptors as illustrated in Figure 1.3. Gp120 which is located on the viral envelope interacts with the CD4 receptors and chemokine

co-receptors such as CCR5 and CXCR4 (Sierra et al., 2005). The HIV-1 life cycle can be divided into the early and late phase.

In the early phase, HIV-1 particles bind specifically to CD4 expressing cells via interaction between gp120 and the chemokine receptors (Taylor and Hammer, 2008) thereby triggering a conformational change in gp120 which uncovers the co-receptor binding site. After binding to the co-receptor the viral and cellular membranes fuse thereby allowing for uncoating and release of the viral core into the host cell. The single stranded viral RNA is then reverse transcribed to double strand DNA by the enzyme reverse transcriptase which is then transported to the nucleus (Pomerantz and Horn, 2003). After transportation to the nucleus, the viral DNA is integrated with the host cell genome by the enzyme integrase.

In the late phase of the viral life cycle, spliced and unspliced mRNA transcripts are synthesised and transported out of the nucleus for translation into polypeptides such as *gag*, *gag-pol* and *nef* (Sierra et al., 2005. *Gag* and *gag-pol* polyproteins assemble near the cell membrane and once they are budded off the cell, they are cleaved by the enzyme protease to produce the independent enzymes as well as structural proteins: matrix (MA), capsid (CA) and nucleocapsid (NC) (de Oliveira et al., 2003). These structural proteins mature to form the infectious viral particle.

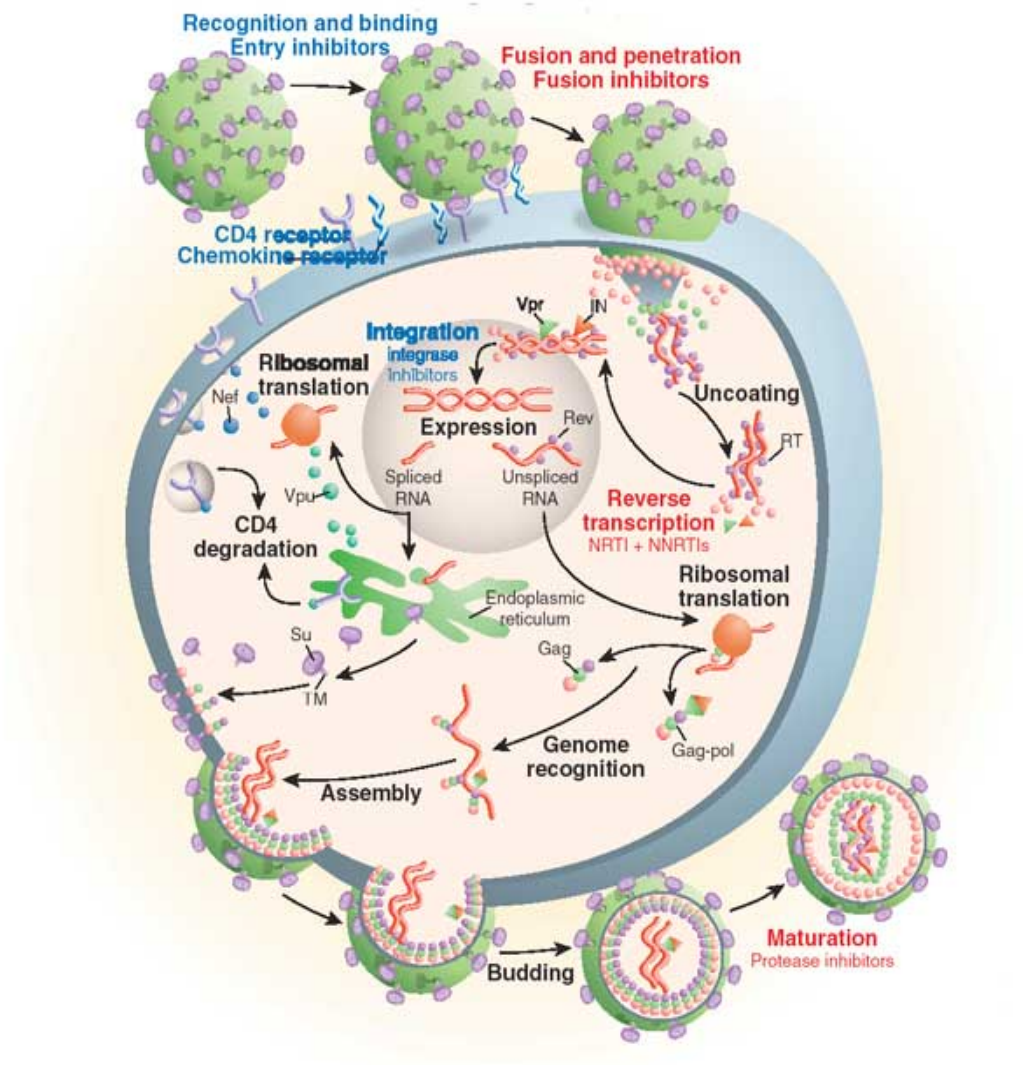

Fig 1.3 Diagram illustrating the features of the HIV-1 replication cycle. Once a virus binds and fuses with a CD4 cell, it releases its contents into the cell. The single stranded RNA is then reverse transcribed to double stranded DNA which is integrated with the host's cells DNA. The DNA is transcribed to form proteins which are assembled close to the cell's membrane prior to budding. Once the immature virus buds off the cell, the proteins are cleaved to render the virus mature and infectious. Taken from (Pomerantz and Horn, 2003).

1.2 Antiretroviral Therapy

Antiretroviral (ARV) therapy has helped transform HIV-1 infection from a fatal infection to a chronic disease (Cohen et al., 2008). Initial targets for ARV therapy were protease and reverse transcriptase due to their essential role in the HIV-1 viral replication cycle. Shortly after the first successful ARV agent zidovudine (AZT) became available, it was clear that HIV-1 had the

ability to generate drug-resistant mutants and that a combination of ARV agents were required for effective therapy (Larder, 2001).

Treatment with highly active antiretroviral therapy (HAART) has helped reduce the morbidity and mortality associated with HIV-1 infection (Wood et al., 2003). HAART is a combination of at least three drugs belonging to two different classes of ARV agents i.e two nucleoside reverse transcriptase inhibitors (NRTIs) and either a protease inhibitor (PI) or a non-nucleoside reverse transcriptase inhibitor (NNRTI).

1.3 HIV-1 Drug Resistance

The emergence of drug resistance complicates therapy. Drug resistant viruses emerge when therapies fail to achieve a complete and sustained suppression of virus replication (Hirsch et al., 2000). Even though antiretroviral drugs have been successfully used to treat HIV-1 infected individuals, the emergence of viruses with reduced susceptibility to these agents are inevitable (Clavel and Hance, 2004). Genetically distinct viral variants (viral quasispecies) are generated in the absence of effective treatment by continuous high levels of viral replication, lack of proofreading by the viral reverse transcriptase (RT) and recombination (Domingo et al., 1996).

Resistance mutations can be classified as primary and secondary mutations. Primary resistance mutations affect the enzyme's susceptibility to drug binding (Molla et al., 1996, Cabana et al., 2002) .These mutations are the first to appear and significantly affect drug resistance. Secondary resistance mutations develop later and serve to improve enzyme function. They contribute to drug resistance when they occur in combination with other primary mutations but have less of a direct effect on inhibitor binding (Molla et al., 1996, Cabana et al., 2002).

1.4 HIV-1 Reverse Transcriptase

Reverse transcriptase is a multifunctional enzyme that possesses RNA- and DNA-dependant DNA polymerase, RNase H, strand transfer and strand displacement activities (Menendez-Arias, 2010). The process of converting single stranded RNA to double stranded DNA is catalysed by the RNA- and DNA-dependant polymerase and ribonuclease H (RNase H) activities of the enzyme (Figueiredo et al., 2006). HIV-1 RT consists of a 66 kDa (p66) and 51 kDa (p51) subunit (di Marzo Veronese et al., 1986). The p51 subunit is composed of the first 440 amino acids of the RT gene while the p66 subunit is made up of all 560 amino acids of the RT gene (Shafer, 2002). Both p66 and p51 have an identical primary structure, however p51 lacks the 120 amino acid RNase H domain which is found in the C-terminal region of p66 (Menendez-Arias, 2010). The p66 subunit comprises of the DNA-binding groove and the active site while the p51 subunit has no enzymatic activity (Shafer, 2002). The crystal structure of HIV-1 reverse transcriptase shown in Figure 1.4, illustrates the four common subdomains between p66 and p51: 'fingers' (residues 1-85 and 118-155), 'palm' (residues 86-117 and 156-236), 'thumb' (237-318) and 'connection' (319-426) (Kohlstaedt et al., 1992, Jacobo-Molina et al., 1993, Huang et al., 1998). The DNA polymerase domain consists of the 'fingers', 'palm' and 'thumb' subdomains, as well as the active site carboxylates which are required for catalysis (Shafer, 2002).

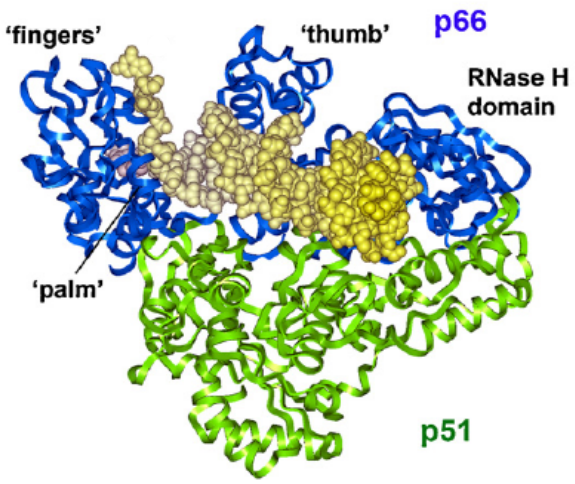

Fig 1.4 Crystal structure of HIV-1 reverse transcriptase. The p66 and p55 subunits are shown as ribbon structures in blue and green respectively. The 'fingers', 'palm' and 'thumb' is common to both subunits while RNase H belongs to the p66 domain. Taken from Menendez-Arias, (2010).

1.4.1 HIV-1 Reverse Transcriptase Inhibitors

Drugs targeting the HIV-1 reverse transcriptase are included in all regimens of HAART in South Africa. The reverse transcriptase inhibitors (RTIs) comprise of nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). The NRTIs are similar in structure to the normal nucleotides which make up DNA, therefore they are able to incorporate themselves into the growing viral DNA chain and prevent further addition of nucleotides thereby causing chain termination as shown in Figure 1.5a (Richman, 2001, Clavel and Hance, 2004). The NNRTIs act by binding to the hydrophobic pocket of reverse transcriptase thereby blocking polymerisation of the viral DNA as shown in Figure 1.5b (Clavel and Hance, 2004).

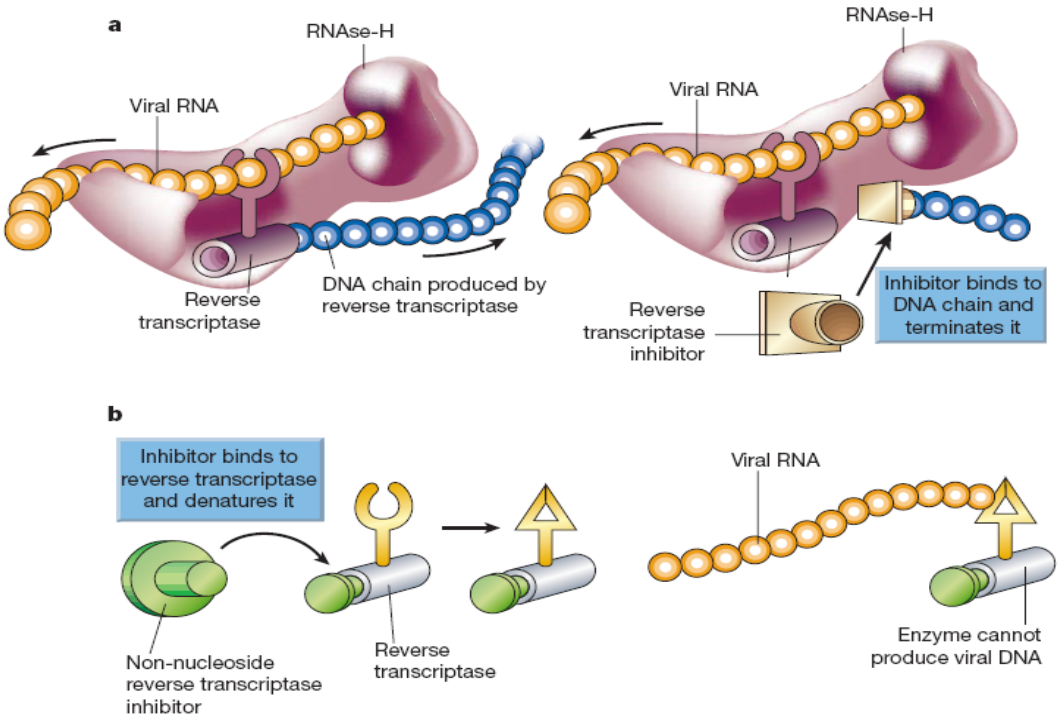

Fig 1.5 Mechanism of action of nucleoside (a) and non-nucleoside reverse transcriptase inhibitors (b). (a)The NRTIs are similar to the human nucleotides and therefore incorporate themselves into the growing DNA chain thereby blocking further production of the DNA chain. (b)The NNRTIs attach directly to the reverse transcriptase enzyme thereby blocking the functioning of the enzyme. The virus is therefore unable to convert RNA to DNA and infect a cell to produce new virions. Taken from (Richman, 2001).

1.4.1.1 Resistance to Nucleoside Reverse Transcriptase Inhibitors

One nucleotide and six nucleoside -RT inhibitors have been approved by the US Food and Drug Administration (FDA) to date. The nucleoside RT inhibitors include lamivudine (3TC), emtricitabine (FTC), abacavir (ABC), didanosine (DDI), stavudine (D4T) and zidovudine (AZT) while the nucleotide RT inhibitor is tenofovir (TDF) (Shafer, 2006). Since both nucleoside and nucleotide RT inhibitors have a similar mechanism of action, they are commonly referred to as NRTIs.

One of the mechanisms of NRTI resistance is via mutations which allow the RT enzyme to distinguish between the NRTIs and the natural substrate, thereby preventing their addition to the DNA chain (Larder et al., 1998, Sarafianos et al., 1999). Several NRTI resistance mutations

have been reported to confer cross resistance to other NRTIs such as M184V, L74V (Dykes and Demeter, 2007) as well as thymidine analogue mutations (TAMs). M184V/I mutation develops soon after the initiation of therapy with 3TC due to the drug's low genetic barrier, and confers high level resistance to both 3TC and FTC. The mutation K65R has been observed in patients treated with TDF, ABC and DDI, and was found to confer low- to high-level resistance to all NRTIs except AZT (Martinez-Picado and Martinez, 2008). The mutation K65R, which is negatively associated with TAMs, has been found to reduce replication capacities and RT processing *in vitro*, which is consistent with the low prevalence of the K65R mutation in isolates from drug-experienced patients (Miller et al., 1998, White et al., 2002, Deval et al., 2004).

Two distinct pathways of TAMs have been reported; TAM-1 and TAM-2. The TAM-1 pathway comprises of the 215Y mutation along with 41L and 210W while the TAM-2 pathway consists of the 215F, 70R, 67N and 219Q/E mutations (Armstrong et al., 2009, Miller et al., 2004, Marcelin et al., 2004). In subtype C, there is evidence that these two pathways are not distinctive. In a recent study by Armstrong et al. (2009), it was found that 67N, 70R and 215Y were the most commonly found mutations in a DDI containing regimen (along with AZT and EFV or NVP) in subtype C while the 41L mutation was rare. (Soares et al., 2007), found that the 41L and 210W mutations occurred more frequently in subtype B than C viruses. Further analysis on NRTI mutations may lead to a better understanding of the TAM pathways occurring in subtype C viruses.

1.4.1.2 Resistance to Non-Nucleoside Reverse Transcriptase Inhibitors

Currently there are three FDA-approved NNRTI agents used in therapy i.e. efavirenz (EFV), nevirapine (NVP) and delavirdine (DLV), however all three of these drugs have a low genetic barrier (Richman, 2001, Sardana et al., 1992). Etravirine (ETV) however is an NNRTI that is

promising and active against a variety of NNRTI-resistant viruses (Young et al., 1995, Andries et al., 2004, Ludovici et al., 2001).

The presence of single amino acid changes such as K103N/S or Y188C/H/L limit the use of currently approved NNRTIs (Antinori et al., 2002). K103N is the most predominant mutation found in patients receiving NNRTIs and has been found to persist for long periods of time even after NNRTI therapy has been discontinued (Gianotti et al., 2005). In the presence of NVP, the mutation Y188L confers the greatest selective advantage followed by Y188C and Y188H (Martinez-Picado and Martinez, 2008). Other substitutions that have also been involved in NNRTI resistance include: L100I, K101P, V106M/A, V108I, Y181C/I/V, G190A/S/C/E/Q, P225H, M230L P236L and Y318F (Martinez-Picado and Martinez, 2008). It has been observed that at all clinically relevant levels of drug exposure, NNRTI-resistant HIV-1 display higher replication capacity than wild type virus (Martinez-Picado and Martinez, 2008).

1.5 HIV-1 Protease

HIV-1 protease is a homodimeric enzyme consisting of two polypeptide chains, each containing 99 residues (Menendez-Arias, 2010). Each monomer contains the conserved sequence Asp-Thr-Gly which provides the aspartyl group that is necessary for catalysis. The HIV-1 protease is located at the 5' end of the *pol* gene and is expressed as part of the *gag-pol* polyprotein.

When the protease enzyme binds to its substrate, it recognises at least seven substrate residues i.e. four at the N-terminal and three at the C-terminal sides of the cleavage position (Menendez-Arias, 2010). It targets amino acids in the *gag* and *gag-pol* polyproteins which are necessary to cleave in order for mature virions to be formed (Flexner, 2008). The *gag* and *gag-pol* polyproteins are cleaved by HIV-1 protease to form mature functional proteins as shown in

Figure 1.6. Cleavage of the *gag* polyprotein releases the structural proteins matrix (MA), capsid (CA), nucleocapsid (NC) and p1, p2 and $p6^{gag}$ proteins while cleavage of the *gag-pol* polyprotein release essential enzymes such as protease (PR), reverse transcriptase (RT), RT-RNase and integrase (IN).

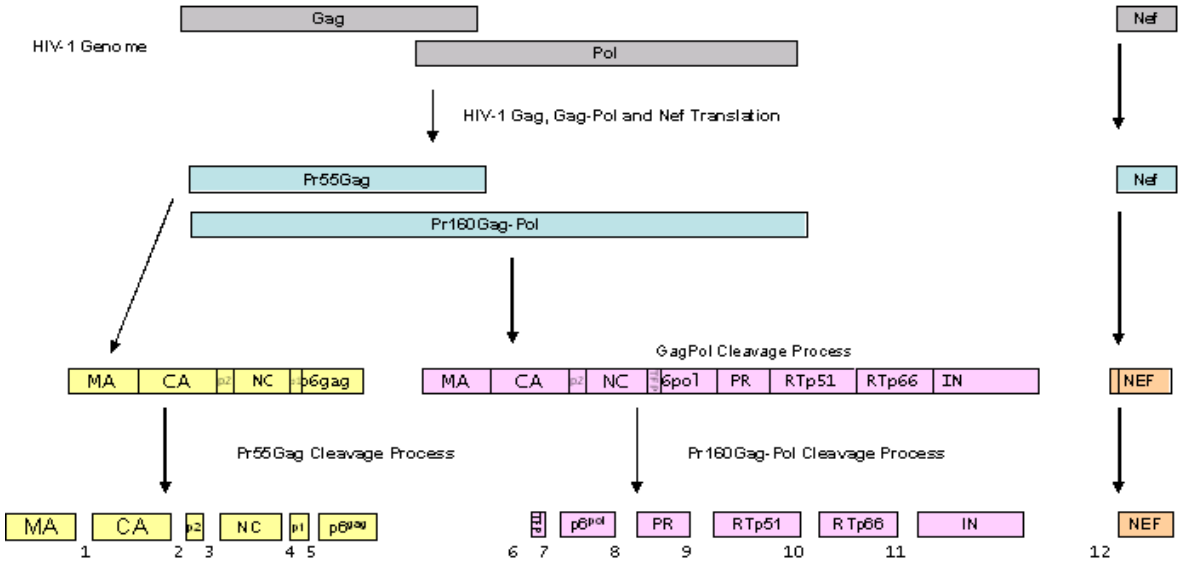

Fig 1.6 Schematic diagram of the *gag* and *gag-pol* processing sites. This diagram shows the 5 cleavage sites in *gag* (MA/CA, CA/p2, p2/NC, NC/p1 and p1/$p6^{gag}$) and 4 cleavage sites in *gag-pol* (NC/TFP, TFP/$p6^{pol}$, $p6^{pol}$/PR, PR/RT). Taken from (de Oliveira et al., 2003).

1.5.1 HIV-1 Protease Inhibitors

HIV-1 protease inhibitors prevent cleavage of the *gag* and *gag-pol* polyproteins thereby arresting maturation and blocking the infectivity of budding virions as shown in Figure 1.7. There are currently eight FDA-approved PIs: atazanavir (ATV), darunavir (DRV), fosamprenavir (FPV), indinavir (IDV), lopinavir (LPV), nelfinavir (NFV), saquinavir (SQV) and tipranavir (TPV). All of these drugs are manufactured in combination with ritonavir (RTV) except for NFV. It was found that RTV reduces the metabolism of PIs via hepatic and intestinal cytochrome P450 3A4 inhibition thereby allowing for improved bioavailability and half-life of these PIs (Wensing et al., 2010).

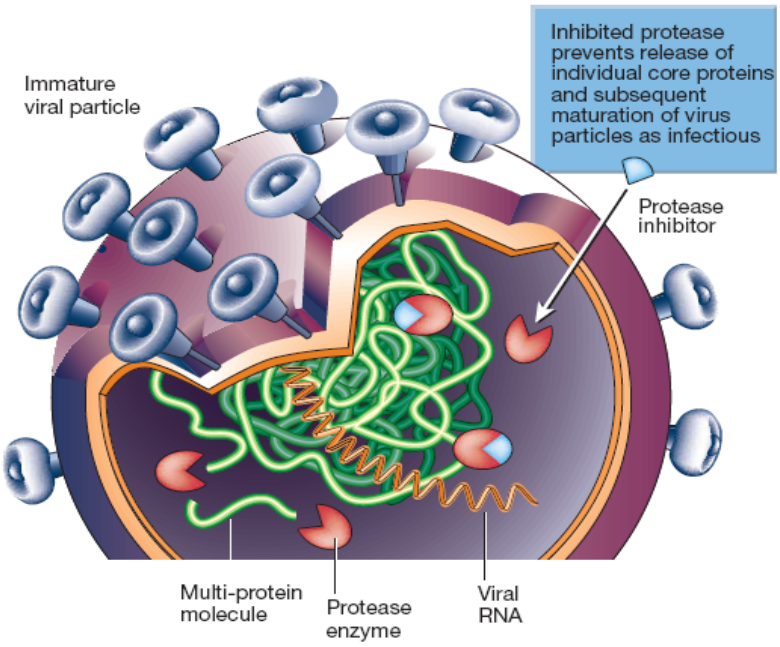

Fig 1.7 Mechanism of action of protease inhibitors (PIs). Protease inhibitors act by targeting viral assembly by inhibiting the activity of protease, thereby preventing cleavage of nascent proteins required for viral assembly. Taken from (Richman, 2001).

1.5.2 Resistance to Protease Inhibitors

Resistance can occur by mutations arising in the substrate cleft of protease which reduce the binding affinity of the inhibitor to the mutant protease, as well by mutations arising which compensate for the decreased activity of the mutant protease (Erickson et al., 1999). Many of these mutations can lead to reduced replication capacity. The HIV protease can tolerate a great amount of mutations since almost a third of the protease gene can deviate from the wild type sequence without altering the protein's function (Molla et al., 1996). Prolonged treatment with one PI can lead to the emergence of viruses with primary and secondary mutations, as well cross-resistance to other PIs (Flexner et al., 2008).

The side effects of drug resistance mutations on the protease enzyme function include the loss of replication capacity or viral fitness (Berkhout, 1999). Viral fitness refers to the ability of the virus to replicate in a given environment (Barbour and Grant, 2004). Mutations causing a loss of

viral fitness include D30N, G48V, I50V/L, V82A, I84V and L90M (Patick et al., 1998, Mammano et al., 2000, Colonno et al., 2004). A specific virus population is considered fitter than another when it is more able to multiply and spread (Martinez-Picado and Martinez, 2008). Some mutations have the ability to partially or completely improve the replication deficit conferred by certain resistance mutations. These include mutations such as M46I, L63P and A71V (Martinez-Picado et al., 1999, Perrin and Mammano, 2003, Schock et al., 1996). Sequence analysis of PI drug resistant viruses has revealed that mutations have not only been observed in the viral protease, but also in the protease CS in Gag (Verheyen et al., 2006, Nijhuis et al., 2007, Dam et al., 2009, Verheyen et al., 2009). Mutations occurring at the CS are believed to adapt the virus to the altered substrate-binding cleft of the mutant viral protease (Doyon et al., 1996a, Mammano et al., 1998, Maguire et al., 2002).

1.5.3 The Contribution of *Gag* Mutations to Protease Inhibitor Resistance

From *in vitro* studies the order of cleavage was found to be: firstly the p2/NC site (Erickson-Viitanen et al., 1989, Krausslich et al., 1989, Pettit et al., 1994), followed by CA/MA and p1/p6^{gag}, CA/p2 and finally NC/p1 (Pettit et al., 2002). Resistance mutations that occur in protease reduce the replicative capacity of the virus due to decreased binding to the natural substrate, HIV-1 *gag* (Parry et al., 2009).

Mutations occurring at the protease active site can result in altered specificity of the enzyme which could lead to co-evolution of the target CS sequences (Swanstrom and Erona, 2000). Mutations arising in *gag* were found to partially restore the replication capacity and are thus selected for, both at CS and non-CS (Doyon et al., 1996b, Gatanaga et al., 2002, Myint et al., 2004).

Several amino acid substitutions at the CS of *gag* have been shown to improve the catalytic efficiency protease in HIV-1 viruses resistant to PIs (Doyon et al., 1996b, Zhang et al., 1997, Mammano et al., 1998, Cote et al., 2001). Nijhuis et al. (2007) found that mutations (A431V, K436E and I437V/T) at the *gag* NC/p1 CS, alone without any change in the viral protease was selected during *in vitro* protease inhibitor exposure. When these mutations were introduced into a reference strain they were found to cause PI resistance which were directly related to increased *gag* processing (Wensing et al., 2010). In a recent study by Dam et al. (2009), it was found that mutations in the NC/p1 CS strongly contributed to HIV-1 resistance to PIs in highly drug-experienced patients besides compensating for viral fitness.

Even though several studies have focused on CS mutations and its effects on viral fitness and drug resistance, non-CS mutations in *gag* were found to recover the reduced viral replication capacity of HIV-1 caused by mutations in the protease (Gatanaga et al., 2002). Gatanaga et al. (2002) found that viruses carrying non-CS mutations H219Q and R409K contributed to the development of resistance to PIs, as well as being indispensable for viral replication in the presence of PIs, since upon reversion to the wild-type amino acid these viruses refused to replicate.

Studies have revealed that amino acids flanking the protease CS are generally hydrophobic (Pettit et al, 1994). The P4 to P4' residues shown in Figure 1.8, share an overall hydrophobicity especially in the P1 and P1' positions (Debouck, 1992). Most studies have focused on the P4-P3' positions which are in direct contact with the viral protease. However it was demonstrated by Nijhuis et al. in 2007 that the more distally located P4' and P5' positions also affect the efficiency of cleavage.

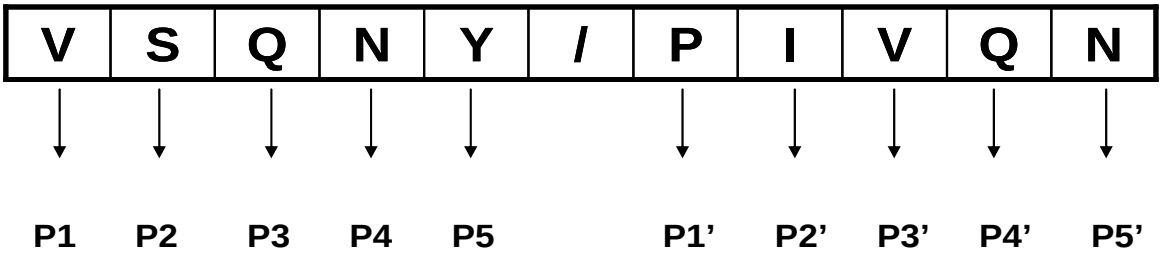

Fig 1.8 Illustration of a the MA/CA protease cleavage site and the nomenclature used for each amino acid position at the cleavage site.

1.6 Drug Resistance Testing

Drug resistance testing has been accepted as an important process to the management of patients with detectable plasma viraemia who are receiving ARV therapy (Hirsch et al., 2008). There are two general types of resistance assays i.e. genotypic assays and phenotypic assays. Genotypic assays involve sequencing of the HIV-1 gene to detect mutations that confer drug resistance, while phenotypic assays involve cell culture-based viral replication assays in the presence of drugs (Hirsch et al., 2008).

Commercially available genotyping methods include ViroSeq (Applied Biosystems, California), Trugene (Bayer, Pittsburgh, PA) and GeneSeq (Virologic, South San Francisco, CA). These tests involve viral RNA extraction from plasma, reverse transcription and amplification of the cDNA by PCR, and sequencing of the amplicons on an automated DNA sequencer. Genotypic testing is most commonly used in clinical settings due to their wider availability and quicker turnaround time. Genotypic tests have the ability to detect mutations present as mixtures which cannot be done using phenotypic tests (Hirsch et al., 2008). Results of genotypic tests use lists of predefined resistance mutations or classifications by computerised rule-based algorithms to characterise viral susceptibility to each antiretroviral drug (MacArthur, 2004, Shafer et al., 2007).

Standard phenotypic testing involves using recombinant virus assays, which includes amplifying specific genes from the plasma virus and generating a recombinant virus with other HIV-1 genes derived from a laboratory construct (Walter et al., 1999, Petropoulos et al., 2000, Whitcomb et al., 2007). The results of phenotypic testing are usually presented as a fold change in susceptibility of the test sample compared with a laboratory control isolate (Hirsch et al., 2008). However phenotypic assays are more expensive, time consuming and complex compared to genotypic assays.

With all resistance assays, the results are weighted in favour of the majority species present and this can obscure the detection of minority species (Petropoulos et al., 2000). This should always be kept in mind when interpreting test results.

1.7 Antiretroviral Treatment in South Africa

The main goal of ARV therapy is to decrease HIV-related morbidity and mortality. The selection criteria for ARV therapy initiation in adults and adolescents in South Africa as stated in the latest guidelines includes a CD4 count of less than 300 cells/mm^3 or WHO Stage IV AIDS-defining illness irrespective of CD4 count (Department of Health, 2010). For ARV therapy to be initiated in children, specific medical criteria have to be met. This includes recurrent hospitalisation for HIV-related disease, modified WHO stage II disease or CD4 percentage <20 % in children under 18 months old and <15 % in children over 18 months. Table 1.1 and 1.2 show the recommended regimen for adults and paediatrics before April 2010. Table 1.3 and 1.4 show the recommended regimen after April 2010.

Patients who are on regimen 1a were switched onto regimen 1b if they were unable to guarantee reliable contraception since EFV is teratogenic and can be harmful to the unborn child.

Table 1.1 : Recommended regimen for adults and adolescents prior to April 2010

Regimen	Drugs
1a	D4T/ 3TC/ EFV
1b	D4T/ 3TC/ NVP
2	AZT/ ddI/ LPV/r

Table 1.2: Recommended regimen for paediatrics prior to April 2010

1st Line	
6 months – 3 years	>3 years old and >10 kg
D4T	D4T
3TC	3TC
LPV/r	EFV
2nd Line	
6 months – 3 years	>3 years old >10 kg
AZT	AZT
DDI	DDI
NVP	LPV/r

Table 1.3: Recommended regimen for adults and adolescents after April 2010

1st Line	
All new patients requiring treatment, including pregnant women:	TDF + 3TC/FTC + EFV/NVP
Currently on a D4T based regimen with no side-effects:	D4T + 3TC + EFV/NVP
Contraindication to TDF: renal diseases:	AZT + 3TC + EFV/NVP
2nd Line	
Failing on a D4T or AZT-based 1st line regimen:	TDF + 3TC/FTC + LPV/r
Failing on a TDF-based 1st line regimen:	AZT + 3TC + LPV/r

Table 1.4: Recommended regimen for paediatrics after April 2010

1st Line	
All infants and children under 3 years:	ABC + 3TC + LPV/r
Children 3 years and over:	ABC + 3TC + EFV
2nd Line	
Children above 3 years Failed ABC + 3TC + EFV:	AZT + DDI + LPV/r
Failed on AZT or DDI-based regimen:	ABC + 3TC + LPV/r

When patients are failing their ARV regimen, they are assessed for virologic, immunologic and clinical failure before having their regimen changed. Primary virologic failure is defined as less

than 10-fold decrease in viral load after 6-8 weeks of therapy while secondary virologic failure is a 10-fold increase in the lowest recorded level. Immunologic failure is defined as a 30 % decrease in CD4 count from peak value or a return to pre-ART baseline. Clinical failure is defined as progression of disease with the development of opportunistic infections or malignancy occurring three months or more after the initiation of therapy. According to South African antiretroviral treatment guidelines patients who have experienced virologic failure with good adherence are changed onto second-line therapy. Those patients showing immunologic and clinical failure remain on first-line therapy. Unfortunately, since there is no official third-line regimen in South Africa, patients failing second-line therapy remain on that regimen until there is no further benefit from the treatment.

1.8 Project Aim and Objectives

The aim of this project was to determine how mutations in protease and *gag* impact on resistance and protease cleavage.

The specific objectives of this study included:

- To perform resistance genotyping assays on 30 patients failing a PI inclusive regimen.
- To sequence the *gag* region from these patients and compare their protease cleavage sites to a subtype B reference strain and matched sequences from other HIV-1 subtypes.
- To correlate polymorphisms at the protease cleavage sites and non-cleavage sites in *gag* with known drug resistance mutations in these patients.
- To compare cleavage rates of subtype C naïve vs subtype C treated isolates.

CHAPTER 2

Resistance Genotyping of HIV-1 *Pol*

2.1 Introduction

When a patient's viral load increases while on therapy, it is recommended that the antiretroviral therapy is changed (Dybul et al., 2002, Hirsch et al., 2003). Due to cross resistance between ARV drugs within a class, treatment cannot be changed based on the assumption that HIV will remain susceptible to alternative drugs within a class that the patient has previously received (Clavel and Hance, 2004). Therefore resistance tests which evaluate the susceptibility of HIV to particular drugs have been developed to facilitate the selection of alternative regimens.

The ViroSeq and In-house assay were used for this study. The ViroSeq HIV-1 Genotyping System detects mutations in the protease and reverse transcriptase regions of the *pol* gene that confer resistance to specific types of antiretroviral drugs. The analysis yields a report indicating evidence for viral resistance such as, resistance mutations to different classes of drugs and levels of drug resistance to particular drugs. The procedure involves viral RNA isolation from plasma, amplification of a 1.8kb product (complete protease gene and two-thirds of the RT gene) by RT-PCR. This amplicon is used as a sequencing template for primers A, B, C, D, F, G, H to generate a 1.3 kb consensus sequence. The ViroSeq HIV-1 genotyping system software assembles, edits and identifies mutations within the sequence. The software also compares the consensus sequence to a subtype B reference, HXB2, in order to determine mutations in the sample being analysed. The edited sequence is then submitted onto the Stanford Drug Resistance Database (Liu and Shafer, 2006) and a report is generated with details of protease and reverse transcriptase mutations present, and the susceptibility of the virus to particular drugs.

2.2 Materials and Methods

2.2.1 Study sites

This study was conducted at the Sinikithemba Outpatient HIV/AIDS Clinic at McCord Hospital (MCH) and the ARV clinic at King Edward Hospital (KEH) in Durban, South Africa. This study was approved by the respective hospital management at both hospitals as well as by the UKZN Biomedical Research Ethics Committee (ref: BF 068/08).

2.2.2 Specimen collection and processing

A total of 30 subtype C patients who were failing a PI-inclusive regimen were recruited for this study. For the purpose of this study virologic failure was defined as two consecutive viral loads >1000 copies/mL.

After obtaining informed consent, patients' blood samples were collected in heparin free EDTA tubes (to prevent coagulation of the blood). The tubes were centrifuged in a refrigerated Jouan MR22i centrifuge (Scientific Group, South Africa), at 1000 x *g* for 10 minutes to separate the plasma and buffy coat which were collected and stored at -80 °C until required.

2.2.2.1 Controls

A group of 30 subtype C 1st line failures were also used as controls since these patients had no previous exposure to PI therapy. These samples were used from stored plasma. The subtype C 1st line failures were previously genotyped in the SARC's (South African Resistance Cohort) study. All available subtype B naïve and subtype B treated full length sequences were downloaded from the Los Alamos Sequence Database (LANL), (Los Alamos National Laboratory, 2005) and used as control groups. Since subtype B predominates in the Americas and in Europe, and most countries use a PI in the first and second line regimen, it was difficult

to compare them to our regimens in South Africa. Therefore the subtype B treated sequences were grouped according to those sequences with protease mutations (B+PR) and those without protease mutations (B-PR). Thirty subtype C drug naïve sequences from KwaZulu-Natal were downloaded from the LANL database (Rousseau et al., 2006). Thirty group M sequences were also downloaded from the LANL database and used to compare our sequences to, since this group includes different subtypes (i.e subtypes A-D, F-H, J and K). The group M sequences were previously used by (de Oliveira et al., 2003).

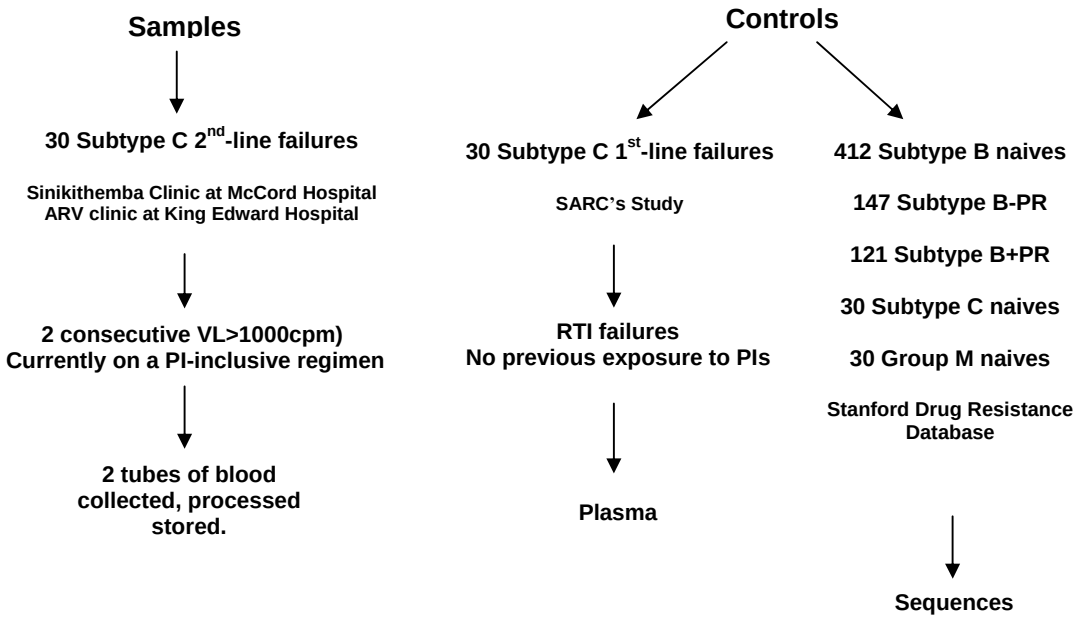

Fig 2.1 Diagram illustrating the samples and controls used for the study. The accession numbers for the sequences downloaded and used as controls are listed in Appendix I.

2.2.3 Patient Characteristics at Study Enrolment

The characteristics of patients recruited from the two centres are shown in Table 2.1. The patients recruited were aged between 12 and 52, having a mean age of 34. The mean CD4 cell count was 146 and the mean viral load was 56785 cpm. All 26 patients, for whom clinical data was available, were receiving Kaletra. The NRTI component varied for the patients: seven were receiving AZT and DDI, 15 were on AZT and 3TC, 2 were on D4T and DDI and 2 were on D4T and 3TC. Twenty three patients were on ARV therapy for an average of 49 months prior to enrolment.

Table 2.1: Patient characteristics at study enrolment. n= number of patients used for the analysis of each characteristic.

Characteristic	[a]**n=**	**Average (Range)**
Age (years)	25	34 (12-52)
CD4 cell count (cells/mm^3)	19	146 (2-352)
Plasma HIV-1 RNA level (cpm)	24	56785 (49-430000)
Duration of ART prior to enrolment (months)	23	49
Treatment regimen at time of enrolment	**n=**	**No. of patients on regimen**
AZT, DDI, and LPV/r	26	7
AZT, 3TC and LPV/r	26	15
D4T, DDI and LPV/r	26	2
D4T, 3TC and LPV/r	26	2

[a]Clinical data for each of the 30 patients was not available, therefore the data that is shown is only for those patients for whom information was available.

Two methods were used for resistance genotyping the samples i.e. the HIV-1 ViroSeq Genotyping system and the In-house genotyping system. Samples were first tested using the HIV-1 ViroSeq Genotyping system and if they failed to amplify using this kit, the assay was repeated using the In-house genotyping method.

2.2.4 ViroSeq Resistance Genotyping Assay

The ViroSeq HIV-1 Genotyping System detects mutations in the protease and reverse trancriptase region of HIV-1 thereby providing the physician with a report indicating genetic evidence of viral resistance (ViroSeq HIV-1 Genotyping System Manual v2.0).

2.2.4.1 RNA Extraction

RNA was extracted using the extraction module contained in the ViroSeq HIV-1 genotyping system (Celera Diagnostics, California) according to the manufacturer's protocol. Briefly, 500ul of plasma was centrifuged in a refrigerated Jouan MR22i centrifuge (Scientific Group, South Africa) at 4°C for 1h, at 25,000 x *g*. The resulting pellet was purified using the ViroSeq extraction module as per the manufacturer's protocol. The RNA was resuspended in 55 µL cold DEPC treated water instead of the recommended elution buffer. RNA was stored at -80 °C until required.

2.2.4.2 Reverse Transcription and Amplification of HIV-1 *Pol*

Reverse transcription was performed in a pre-amplification room. The HIV RT mix and DTT were thawed and vortexed for 5 seconds to mix. The enzymes RNase inhibitor and MuLV reverse transcriptase were not vortexed. All reagent tubes were briefly centrifuged to collect the contents at the bottom. The cDNA master mix was prepared as follows:

Reagent	Volume for 1 reaction(µL)	Final Concentration
HIV RT Mix	8	-
RNase Inhibitor (20U/ul)	1	20U
MuLV Reverse Transcriptase (50U/ul)	1	50U
DTT (100mM)	0.4	2mM
Final volume	**10.4**	

Ten microliters of RNA was added to individual tubes and placed in the thermocycler.

The reaction was run under the following conditions:

Temperature (°C)	Time	Process
65	30 seconds	Relaxes the RNA secondary structure
42	5 minutes	Cools to the optimum enzyme activity temperature
Manually pause thermocycler, add RT mix and resume reaction		
42	60 minutes	Reverse transcription
99	5 minutes	Inactivates MuLV reverse transcriptase
4	Hold (>10 minutes)	Holds until ready to proceed

After completion of the RT program the samples were held in the thermocycler at 4 °C for atleast 10 minutes. If PCR was not performed on the same day, the cDNA was stored at -20 °C.

Amplification was performed in a pre-amplification area. The HIV PCR mix was thawed and vortexed for 5 seconds to mix and briefly centrifuged. The AmpliTaq Gold and AmpErase UNG reagents were only centrifuged for 5 seconds at 2000 x *g* to collect the contents at the bottom of the tubes. The PCR master mix was prepared in a 1.5 mL microcentrifuge tube as follows:

Reagent	Volume for 1 Reaction (µL)	Final Concentration
HIV PCR Mix	29.5	-
AmpliTaq Gold polymerase (5U/ul)	0.5	2U
AmpErase UNG (1U/ul)	1	1U
Final volume	**31**	

Care was taken to only have one tube open at a time while adding 30 µL of PCR mastermix to each RT reaction tube containing HIV-1 cDNA resulting in a final volume of 50 µL in each reaction tube. The PCR was run on the thermocycler under the following conditions:

Temperature (°C)	Time	Cycles
50	10 min	1
93	12 min	1
93	20 sec	
64	45 sec	40
66	3 min	
72	10 min	1
4	Hold	-

Following amplification the PCR products were immediately purified or stored at -20 °C.

2.2.4.3 Amplicon Purification and Quantification

This procedure was performed in a post-amplification area using the ViroSeq RT-PCR Purification module. Briefly, for each sample a microcon-100 spin column was inserted into a microcon 1.5 mL collection tube and 300 µL of KCL was added to the top of each column without touching the membrane. The entire 50 µL of the PCR product was added to the centre of the KCL-filled column and centrifuged at 900 x *g* for 15 minutes in a Jouan A14 centrifuge (Scientific Group, France). The columns were washed with 300 µL of sterile distilled. To elute the DNA, 35 µL of sterile distilled water was added to the centre of each column, the columns were inverted and placed into new 1.5 mL microcentrifuge tube centrifuged at 900 x *g* for 5 minutes in the centrifuge. The columns were removed and discarded leaving approximately 40-50 µL of purified PCR product. In order to verify presence of DNA and to quantify it, an agarose gel was run.

To make a 1 % agarose gel, 0.5 g of agarose per 50 mL of 1 X TBE buffer was heated in a microwave until all agarose was completely dissolved and the resulting solution was clear. When cool to the touch 2 µL of ethidium bromide (0.5 µg/mL) was added and the gel was poured into perspex casting trays. Combs were inserted appropriately and the gel was left to set at room temperature for 30 minutes. Once the gel was set, the combs were carefully removed and the gel was placed in the electrophoresis tank. Electrophoresis running buffer was added to the electrophoresis tank to a level just covering the gel. Two µL of the DNA mass ladder was mixed with 1 µL of gel loading dye (made up of 0.1% bromophenol blue, 40% sucrose, 3.7% EDTA and 0.5% lauryl sulphate sodium) and added to lane 1 of the gel. Five µL of the purified PCR product was mixed with 1 µL of gel loading dye and this was carefully loaded into the

sample wells. The gel was run at 100 V for 45 minutes on an Electrophoresis Power Supply-EPS 301 (Amersham Biosciences, Sweden). The gels were viewed under UV light using the GelVue UV Transilluminator (SynGene, London), and the images were captured and analysed.

Using the picture obtained from the gel the concentration of the PCR products are calculated. The concentration of the bands of the mass ladder are as follows:

Table 2.2: The amount of each DNA fragment using 2 µL of ladder.

Fragment Size (bp)	Amount of DNA (ng)
2000	100
1200	60
800	40
400	20
200	10
100	5

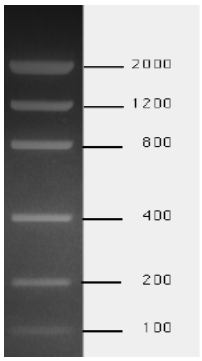

Fig 2.2 An agarose gel loaded with 2 µL of Low DNA mass ladder stained with ethidium bromide.

The PCR products were diluted to an optimal concentration as follows:

If the intensity of the PCR product is:	Then…
> 100 ng	Make a 1:10 dilution with sterile filtered H_2O
60 – 100 ng	Make a 1: 4 dilution with sterile filtered H_2O
40 – 60 ng	Make a 1: 2 dilution with sterile filtered H_2O
20 – 40 ng	Adjust the volume of the sample to 60 ul (minimum volume for sequencing)
<20 ng	Sample not suitable for sequencing, but can try with increasing the number of cycles to 35.

2.2.4.4 Sequencing of HIV-1 *Pol*

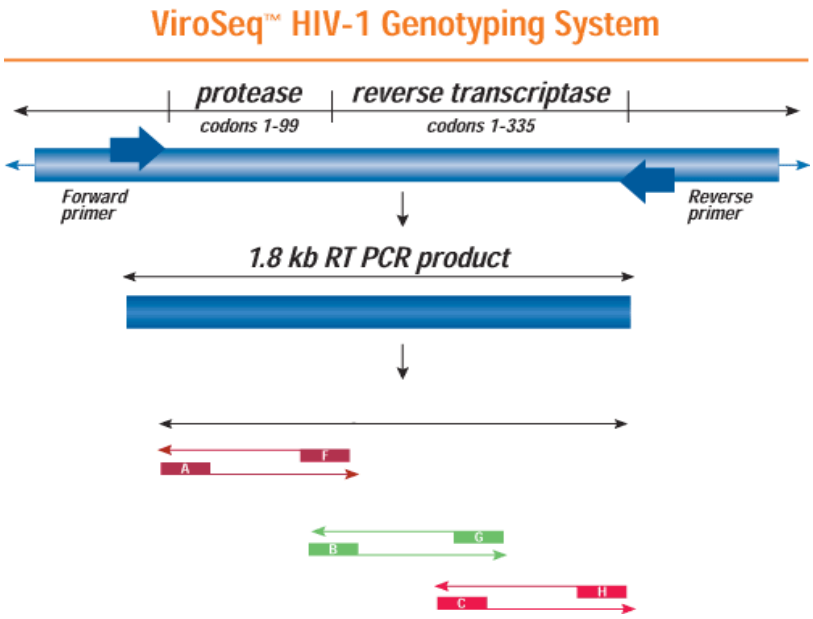

Fig 2.3 Layout of primers used for sequencing the HIV-1 *Pol* gene. Primer D was excluded for this study. The primers amplified amino acids 1-99 of protease and 1-335 of RT. Taken from Applied Biosystems, 2000, <http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/>.

Six primers were used to sequence the *pol* region as shown in Figure 2.3. The sequencing primer mixes were thawed, vortexed and briefly centrifuged. Twelve µL of each primer mix was added to an empty well in a MicroAmp optical 96-well reaction plate (Applied Biosystems, Singapore). Eight µL of the diluted PCR product was added to each primer well. The sequencing reaction was run under the following conditions:

No. of Cycles	Temperature	Time	Process
25	96°C	10 sec	DNA denaturation
	50°C	5 sec	Primer annealing
	60°C	4 min	Primer extension
-	4°C	HOLD	-

The sequencing products were purified on the same day.

To purify the sequenced products a fresh stock of 3.0 M Sodium acetate and 100 % ETOH was prepared to precipitate the DNA as follows in a buffer reservoir.

Reagent	1 well (ul)
3M Sodium acetate pH4.6	2
100% ETOH	50
Final Volume	**52**

Since the stock was prepared in a reservoir and a multi-channel pipette was to be used, sufficient volume for at least 16 extra samples was prepared.

Using a multi-channel pipette 55 µL of the stock solution prepared above, was added to each well. The plate was sealed using adhesive plate covers and it was vortexed and centrifuged at 3000 x *g* for 20 minutes on an Eppendorf Plate Centrifuge 5810R (Merck, Germany). The plate covers were carefully removed and the plate was inverted on folded kimwipe and centrifuged at 150 x *g* for 5 minutes. One hundred and fifty µL of 70 % ethanol was added immediately to each well to prevent drying out of the well which could result in excess of unincorporated dye terminators during the run of the sequencer. The plate was centrifuged at 3000 x *g* for 5 minutes. The plate was inverted one again on a folded kimwipe and centrifuged at 150 x *g* for 1 minute before being left at room temperature for 5 minutes in the dark to allow for the ethanol to dry completely. The plate was sealed with an adhesive foil cover and stored at -20 °C in the dark till sequenced. Prior to sequencing, the dried pellet was resuspended with 10 µL formamide and place in a thermocycler to denature at 95 °C for 3 minutes. The samples were then loaded onto the ABI 3130 XL Genetic Analyzer (Applied Biosystems, California).

If the ViroSeq Resistance assay failed to work on a sample, an In-house assay was used which also amplifies the HIV-1 *pol* region.

2.2.5 In-house Resistance Genotyping Assay

The protocol for the In-house resistance genotyping assay was obtained from the National Institute of Communicable Diseases, AIDS Virus Research Unit. This was further optimized and validated in our laboratory.

This assay is a cost-effective test for drug resistance-related mutation detection on genetically diversified HIV-1. It also amplifies the *pol* region of HIV-1 and the sequence generated is submitted to the Stanford Drug Resistance Database where a report indicating the presence of drug resistance was generated.

2.2.5.1 RNA Extraction

RNA was extracted from the plasma of the samples using the QIAamp Viral RNA Mini Kit (QIAGEN, Germany) according to the manufacturer's protocol.

However, before commencing with the manufacturer's protocol a 1 hr spin at 25000 x *g* at 4 °C was performed on 500 µL of the plasma to concentrate the RNA. Another amendment to the manufacturer's protocol was the use of DEPC water to elute the RNA in instead of the provided elution buffer. The viral RNA was stored at -80 °C until required.

2.2.5.2 Reverse Transcription and Amplification of HIV-1 *Pol*

The cDNA master mix was prepared in a pre-amplification area. The 5 X First strand buffer, DTT and primer IN3 were thawed and vortexed for 5 seconds to mix. The enzymes RNase OUT and Superscript III were not vortexed. The cDNA master mix was prepared in a 1.5 mL RNase/DNase free microcentrifuge tube as follows:

Reagent	Volume for 1 reaction(μL)	Final concentration
5 X First Strand Buffer	4	1X
0.1 M DTT	1	5mM
RNase OUT (40U/μL)	1	40U
Superscript III (200U/ μL)	1	200U
Final volume	**7**	

The RT mix was kept at room temperature until it was added to the reaction tubes. Care was taken to be quick at the following steps since the RT mix should not stand at room temperature for longer than 30 minutes.

One μL of primer IN3, 1 μL of dNTP mix and 1 μL of DEPC water were combined in a 0.2 mL MicroAmp reaction tube. To each reaction tube, 10 μL of RNA was added and pipette mixed. The reaction tubes were placed in a GeneAmp PCR System 9700 thermocycler (Applied Biosystems, Singapore) and run on the following program:

Temperature (°C)	Time
65	5 minutes
4	1 minute
Pause manually, add RT mix and resume reaction	
50	60 minutes
70	15 minutes
4	Hold (>10 minutes)

After the 45 °C step for 1 minute, 7 μL of the master mix was added to each reaction tube. Once the program had finished, the samples were left in the thermocycler for >10 minutes at 4 °C before 1 μL of RNase H was added to each reaction tube and run for 20 minutes at 37 °C on the thermocycler. A 1.7 kb fragment of *pol* was then amplified by nested PCR. The HXB2 positions and sequence of the primers used for the PCR are shown below in Table 2.3.

Table 2.3: Primers targeting the *pol* region during the in-house PCR

Primer Name	HXB2 Positions	Primer Sequence (5' – 3')
G25REV (10 pmol/µL)	1867→1892	GCAAGAGTTTTGGCTGAAGCAATGAG
IN3 (10 pmol/µL)	4246←4212	TCTATCCCATCTAAAAATAGTACTTTCCTGATTCC
AV150 (10 pmol/µL)	2036→2062	GTGGAAAGGAAGGACACCAAATGAAAG
Pol M4 (10 pmol/µL)	3892←3870	CTATTAGCTGCCCCATCTACATA

The layout of the primers on the HIV-1 genome is shown below on Figure 2.4.

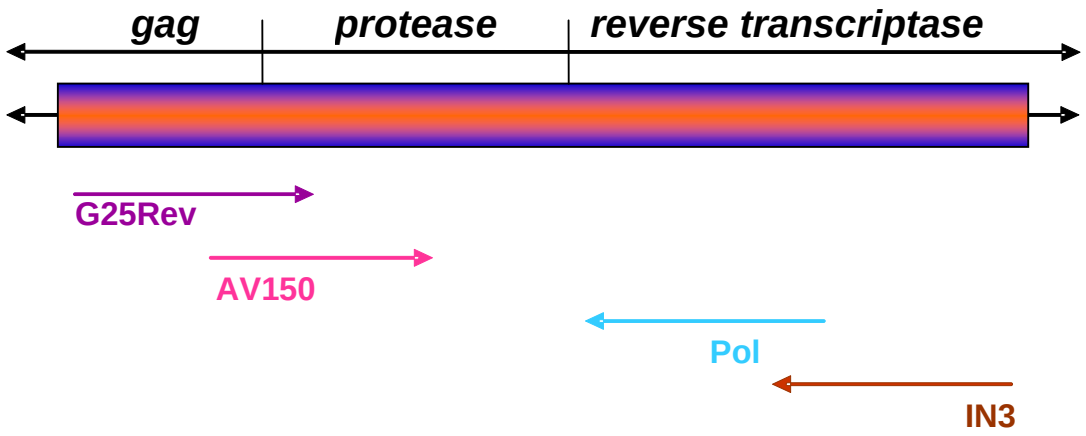

Fig 2.4 Layout of the primers used for the In-house PCR. Four primers were used to amplify the 1.7 kb *pol* region of HIV-1.

The master mix for the first round PCR was prepared in a 1.5 mL microcentrifuge tube as follows:

Reagent	Volume for 1 Reaction (µL)	Final concentration
10 X Buffer + $MgCl_2$	2.5	1X
10mM dNTPs	0.88	0.35mM
10pmol/ul G25REV	0.25	0.125µM
10pmol/ul IN3	0.25	0.125µM
Expand Long Template (5U/ul)	0.38	1.9U
DEPC treated water	16.3	-
Final volume	**20.5**	

A volume of 20.5 µL of PCR mastermix was added to each appropriately labelled 0.2 mL reaction tube. Care was taken to have only one reaction tube open at a time while 4.5 µL of cDNA was added. The conditions for the first and second PCR were as follows:

Temperature (°C)	Time	Cycles
94	2 min	1
94	10 sec	
52	30 sec	10
68	2 min	
94	15 sec	
52	30 sec	52
68	*2 min	*plus 20 sec/cycle
68	7 min	1
4	>10 min	

The master mix for the second round PCR was made up as follows:

Reagent	Volume for 1 Reaction (µL)	Final concentration
10 X Buffer + $MgCl_2$	5.0	1X
10mM dNTPs	1.75	0.35mM
10pmol/ul AV150	0.5	0.125µM
10pmol/ul Pol M4	0.5	0.125µM
Expand Long Template (5U/ul)	0.75	1.9U
DEPC treated water	39.0	-
Final volume	**47.5**	

A volume of 47.5 µL of PCR mastermix was added to each 0.2 mL reaction tube. Care was taken to have only one reaction tube open at a time while 2.5 µL of first round PCR product was added. The PCR was run on the thermocycler following the same PCR conditions as that of the first round PCR.

2.2.5.3 Amplicon Puification and Quantification

The PCR products were detected by agarose gel electrophoresis as described in Chapter 2.2.4.3, except that a 100 bp DNA ladder (New England Biolabs, Massachusetts) was used instead of a DNA mass ladder as shown in Figure 2.5.

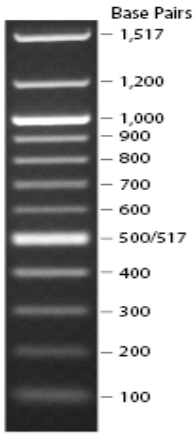

Fig 2.5 100 bp DNA ladder illustrating the sizes of different products on an ethidium bromide stained agarose gel. The figure was taken from the product insert for the 100 bp DNA ladder (New England Biolabs, Massachusetts).

Amplicons were purified using the QIAGEN PCR Purification kit as per manufacturer's protocol. Sterile water was used to elute the DNA instead of the provided elution buffer.

An agarose gel was run to quantify the amount of DNA obtained following purification of the PCR product as described in Chapter 2.2.4.3.

2.2.5.4 Sequencing of HIV-1 *Pol*

The HXB2 positions and sequence of the primers used for the sequencing of the HIV-1 *pol* are shown in Table 2.4.

Table 2.4: Primers used for sequencing the *Pol* region

Primer Name	HXB2 Positions	Primer Sequence (5' – 3')
Pol M0 (1.6 pmol/µL)	2251→2272	TCCCTCAGATCACTATTTGGCA
Pol M1 (1.6 pmol/µL)	2610→2632	GTTAAACAATGGCCATTGACAGA
Pol M4 (1.6 pmol/µL)	3892←3870	CTATTAGCTGCCCCATCTACATA
Pol M8 (1.6 pmol/µL)	3323←3302	CTGTATATCATTGACAGTCCAG
Pol MG (1.6 pmol/µL)	2823←2798	ATTGAACTTCCCAGAAGTCTTGAGTT
AV150 (1.6 pmol/µL)	2036→2062	GTGGAAAGGAAGGACACCAAATGAAAG

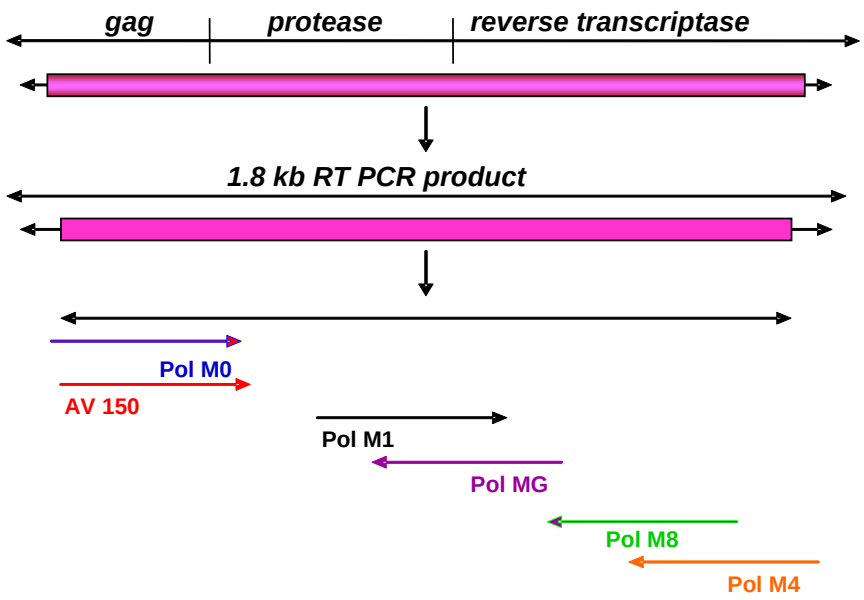

Fig 2.6 Layout of primers used to sequence the HIV-1 *Pol* gene. Six primers were used to sequence the protease and reverse transcriptase regions of HIV-1.

Six primers were used to sequence the *pol* gene as shown in Figure 2.6. The sequencing primers, BigDye and 5 X sequencing buffer were thawed, vortexed and centrifuged. A master mix for each primer was prepared as follows:

Reagent	Volume for 1 Reaction (µL)	Final concentration
BigDye Ready reaction mix	0.4	-
5 X Sequencing Buffer	2	1X
Primer (1.6 pmol/µL)	2	0.32µM
Water	3.1	-
Final volume	**7.5**	

The sequencing reaction was set up in a MicroAmp optical 96-well reaction plate (Applied Biosystems, California), where 7.5 µL of each primer mix and 2.5 µL of diluted PCR product (4ng/ul) was added to a designated well and the plate was sealed with adhesive foil cover strips. The sequencing reaction was run under the following conditions:

No. of Cycles	Temperature	Time
1	96 ^{o}C	1 min
35	96 ^{o}C	10 sec
	50 ^{o}C	5 sec
	60 ^{o}C	4 min
-	4^{o}C	HOLD

The sequencing products were purified on the same day.

To purify the sequenced products a fresh stock of 3.0 M Sodium acetate and 100 % ETOH was prepared in a buffer reservoir as follows:

Reagent	1 well (ul)
3M Sodium acetate pH4.6	1
100% ETOH	25
Final Volume	**26**

Using a multi-channel pipette, 1 µL of EDTA was added to each well and pipette mixed to ensure complete mixing of the EDTA. Twenty six µL of the sodium acetate and ethanol stock solution prepared above, was added to each well. The plate was sealed using adhesive foil covers and it was vortexed and centrifuged at 3000 x *g* for 20 minutes. The plate covers were then carefully removed and the plate was inverted on folded kimwipe and centrifuged at 150 x *g* for 5 minutes to dry. Thirty five µL of 70 % ethanol was immediately added to each well. The plate was centrifuged at 3000 x *g* for 5 minutes. The plate was inverted once again on a folded kimwipe and centrifuged at 150 x *g* for 1 minute before being left at room temperature for 5 minutes in the dark to allow the sample to dry completely. The plate was sealed with an adhesive foil cover and stored at -20 °C in the dark until sequenced. Prior to sequencing, the dried pellet was resuspended with 10 µL formamide and denatured in a thermocycler at 95 °C for 3 minutes. The samples were then loaded onto the ABI 3130 Genetic Analyzer (Applied Biosystems, California).

2.2.6 DNA Sequencing Analysis

Analysis of the DNA sequences was done using the ViroSeq® HIV-1 Genotyping System Software for both the sequences generated from the ViroSeq assay as well as the In-house assay. This software processes the 6 primer sequence files to generate a project. A project is an assembly of the sample files containing all the sequence information required to produce a genotype. The project format allows you to edit the electropherogram data while comparing it to a reference sequence (HXB2) to generate a final consensus sequence for the HIV-1 protease and reverse transcriptase (RT) genes. A project was created for each patient and the consensus sequence was submitted to the Stanford Drug Resistance Database (Liu and Shafer, 2006). A resistance report was generated for each patient's sequence. Each reverse transcriptase and protease sequence was compared to that of a subtype B reference strain, HXB2, in the database. Mutations associated with reduced sensitivity to antiretroviral drugs were assigned a drug penalty score based on genotypic-phenotypic correlative data. The report included all the polymorphisms and drug resistance mutations selected by reverse transcriptase inhibitors (RTIs) and protease inhibitors (PIs).

2.2.7 Statistical Analysis

The Fischer's exact test was used to compare the frequency of mutations between the different groups analysed. The GEE (Generalized Estimating Equation) binary test was used to analyse longitudinal samples. All analysis was performed using GraphPad Prism version 5 (GraphPad Software, California).

2.3 Results

2.3.1 Resistance Genotyping of the Protease Gene

2.3.1.1 Analysis of the Naïve Sequences from Subtype B, Subtype C and Group M

The downloaded protease amino acid sequences of group M, subtype B and C naïve were compared to each other as shown in Figure 2.7. No primary resistance mutations to protease inhibitors were detected in any of the sequences. However secondary mutations were found at the following positions: T74S (10% in C naïve and 0.24% in B naive), K20R (6.67% in C naive) and L10F/I/V (5.58% in B naive). The frequency of mutations T74S and K20R in the C naïve viruses were significantly (P <0.05) higher than in B viruses.

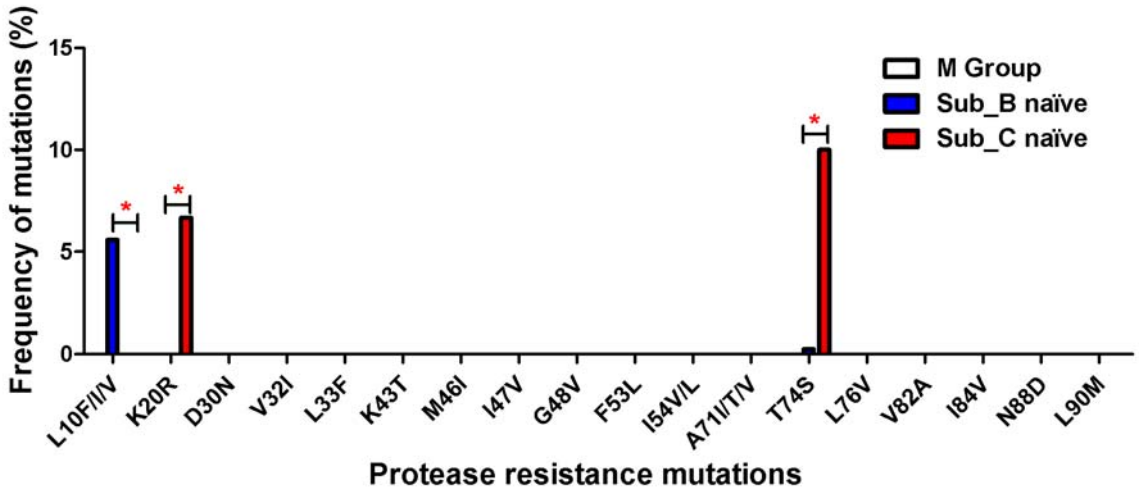

Fig 2.7 Frequency of protease resistance mutations in subtype B, subtype C and M group naïve patients. *P* values <0.05 were considered significant and are indicated by an asterix (*).

2.3.1.2 Analysis of Subtype C Naïve vs Treated Sequences

The frequency of subtype C naïve and treated patients were then compared as shown in Figure 2.8. The C naïve and first-line failures did not harbour any primary resistance mutations, however they did have secondary mutations. The K20R and T74S were found in both naives and the first-line failures, while L10F was only found in the first-line failures. The second-line

failures harboured primary as well as secondary resistance mutations at a variety of sites. Primary mutations found included: M46I, I54V/L, L76V, V82A and I84V. Mutations at M46I, I54V/L, L76V and V82A in the second-line failures were significantly ($P < 0.05$) higher than in the C naives and first-line failures. Four of the six patients who harboured L76V from the second-line failures also had the M46I mutation. When correlations tests were performed using the Spearman rank order correlation test, these mutations were significantly (P <0.0038) correlated with each other (R =0.5123).

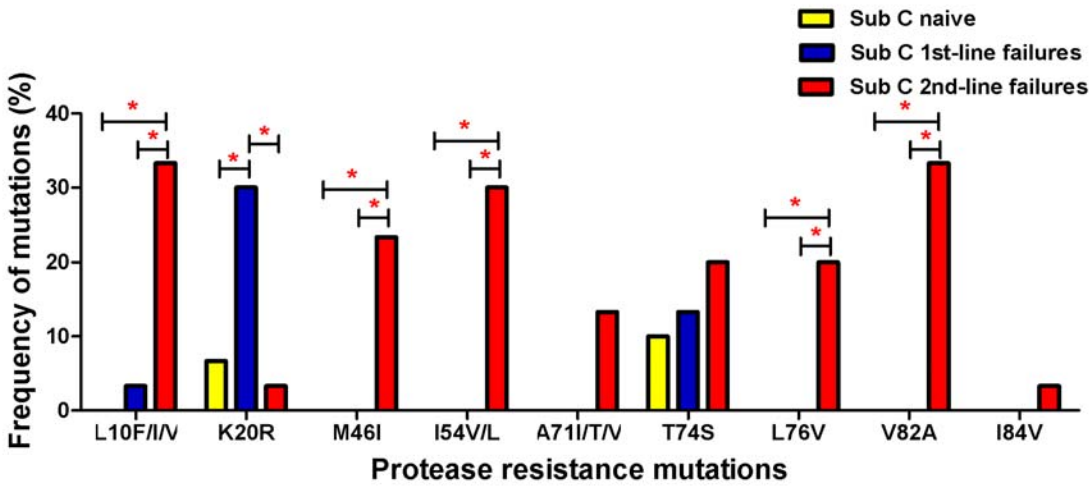

Fig 2.8 Frequency of protease resistance mutations in subtype C naïve and treated patients. P values <0.05 were considered significant and are indicated by an asterix (*).

2.3.1.3 Analysis of Subtype B Naïve vs Treated Sequences

Since most of the studies on resistance have focused on subtype B response to ARVs, it was important for us to compare our subtype C sequences to the subtype B sequences.

The frequency of protease resistance mutations in the downloaded sequences from subtype B naïve and treated patients were analysed as previously described and shown in Figure 2.9. The B naïve patients only harboured the secondary mutation L10F. Primary mutations present in the treated group were significantly ($P < 0.0001$) higher in frequency compared to the B naïve

group. These mutations included: D30N, M46I, I54V/L, V82A, I84V, N88D and L90M. Other primary mutations present in the treated group were V32I, L33F, K43T, G48V and F53L, which were significantly ($P < 0.05$) more frequent than the B naïve group. Similar to the subtype C second-line failures, secondary mutations which were present in the B+PR mutations group were L10F/I/V, A71V/T/I and T74S.

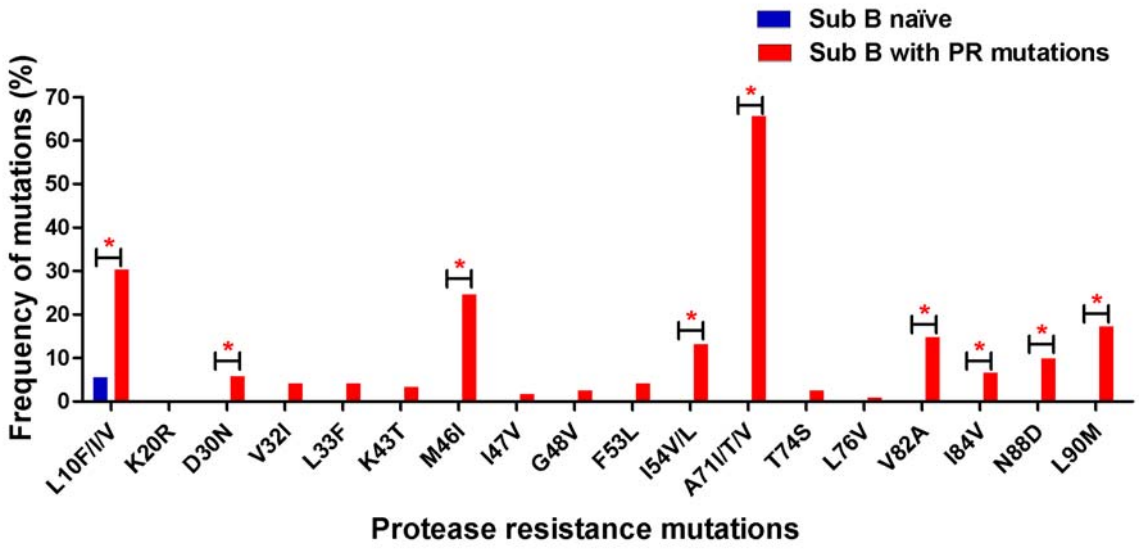

Fig 2.9 Frequency of protease resistance mutations in subtype B naïve and treated patients. *P* values <0.05 were considered significant. Those having a *P* value <0.0001 are indicated by an asterix (*).

2.3.2 Resistance Genotyping of the Reverse Transcriptase Gene

For this section, only subtype C patients were analysed. The amino acid sequences of subtype C first- and second-line failures were investigated to identify mutations associated with reverse transcriptase inhibitors.

2.3.2.1 Analysis of the NRTI Resistance Mutations in Reverse Transcriptase

The frequency of NRTI mutations in subtype C first- and second-line failures was analysed as shown in Figure 2.10. The most commonly detected mutation was M184V in both the first (50%) and second-line failures (23%), however this was not significantly different ($P >0.05$).

TAM 1 (M41L and T215Y) mutations were found both in the first (6.7%) as well as second-line failures (16.7%). The TAM 2 (D67N, K70R, T215F and K219Q) mutations were also present in the first (30%) and second-line failures (36.7%). The presence of any TAM was seen in 33% of the first-line failures and 37% of the second-line failures. Two patients from the first-line failures had three TAMs while three of the second-line failures had three TAMs. Only two first-line and two second-line failures had > 3 TAMs.

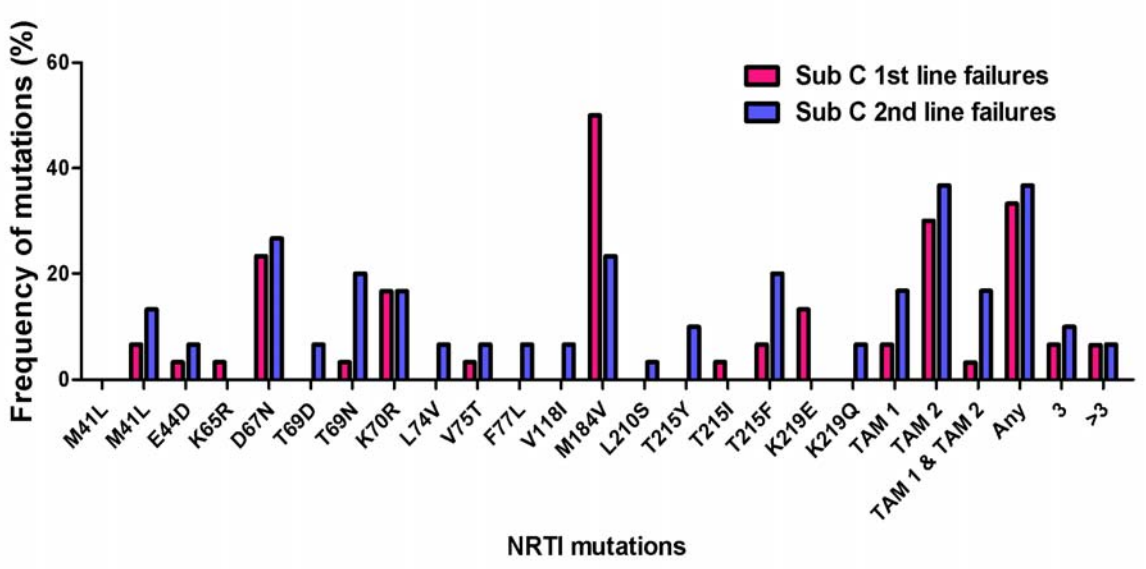

Fig 2.10 Frequency of NRTI resistance mutations in subtype C first- and second-line failures. *P* values <0.05 were considered significant.

2.3.2.2 Analysis of the NNRTI Resistance Mutations in Reverse Transcriptase

The frequency of NNRTI mutations in subtype C first- and second line failures was analysed as shown in Figure 2.11. The most frequent mutation detected in both groups was the K103N mutation, however this was not statistically significant (*P* >0.05) which shows that the K103N mutation persists long after therapy has stopped. The Y188L mutation was only present in 3/30 (10%) of the first-line failures. The K103N, V106M, V179D, G190A and P225H mutations were more frequent in the first-line failures than the second-line failures, 46.7% vs 30%, 30% vs 6.7%, 13% vs 3%, 20% vs 3% and 6.7% vs 3.3% respectively.

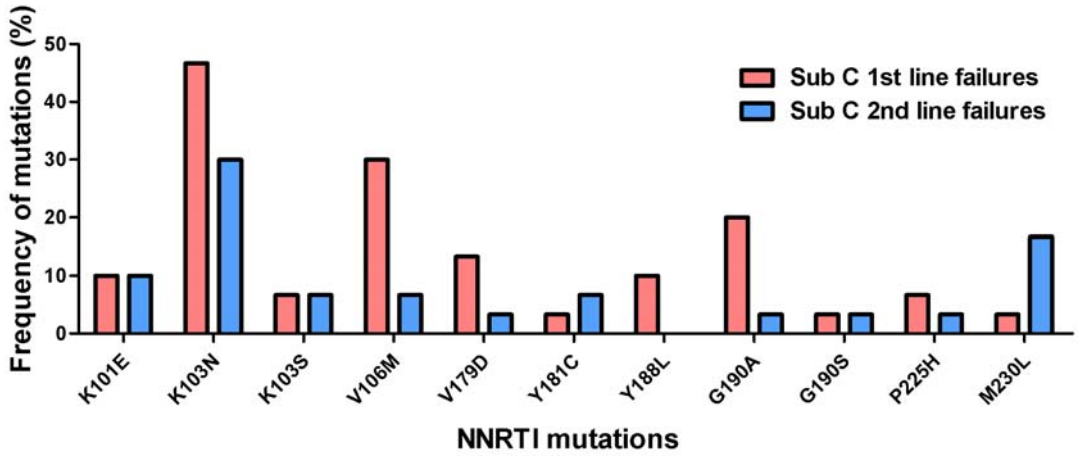

Fig 2.11 Frequency of NNRTI resistance mutations in subtype C first- and second-line failures. *P* values <0.05 were considered significant.

2.4 Discussion

Great efforts have been made in the past few years to control and manage HIV-1 infection in resource-limited settings by programs providing ARV treatment to those in need. However this advancement is threatened by the development of drug resistance which may be due to poor adherence and/or suboptimal regimens. Most studies on drug resistance have focused on HIV-1 subtype B viruses, however C viruses account for the majority of infections globally (Kantor et al., 2002). Therefore it is important to investigate the responsiveness of subtype C viruses to ARV therapy relative to that which has been observed in subtype B viruses.

We compared sequences from treated and untreated patients infected with different HIV-1 subtypes to help us investigate the prevalence of mutations with or without drug selection pressure as well as subtype differences. Mutations which occur at a higher rate among treated patients in both subtype B and C may have implications for the response to therapy. Only the second-line failures (PI-experienced) harboured major protease mutations. Protease mutation L76V was significantly ($P < 0.05$) more frequent in subtype C patients who have been exposed

to PIs in the second-line regimen compared to the first-line regimen. Mutations M46I and L76V were found to be significantly correlated with each other which is consistent with current literature since they were recently reported to be a novel resistance pathway for patients receiving LPV boosted with RTV (Nijhuis et al., 2009). In this study by Nijhuis et al. it was found that the L76V mutation in combination with M46I increased the maximum inhibitory concentration of LPV by 11 fold.

A comparison of the subtype B and C groups revealed the absence of protease mutations D30N, V32I, I47V, G48V, F53L, N88D and L90M in the subtype C groups. Mutations D30N, N88D and L90M are mutations which have been found to be associated with NFV resistance (Patrick et al., 1998, Zhang et al., 1997). The absence of NFV from the drug regimen for the subtype C second-line failures is probably the reason for the lack of these mutations. The absence of the minor mutation K20R, and the low frequency of L76V in the subtype B+PR group compared to the subtype C groups, is also a reflection of the treatment differences in the patients from the different regions. Mutation T74S was found at a high frequency in the naïve and treated subtype C patients, but only at a low frequency in the subtype B+PR group. This may suggest implications for treatment since T74S was found to restore fitness of multi-drug resistant viruses in both subtype B and C (Soares et al., 2009).

The most common mutations present in reverse transcriptase (M184V and K103N) were associated with resistance to NRTI and NNRTI resistance, respectively. The K103N mutation has been found to persist for 1 to 3 years (Delaugerre et al., 2004, Brenner et al., 2002). None of the subtype C second-line failures were on NNRTIs, however the high frequency of K103N (26.7%) could be due to the persistence of this mutation from their first-line therapy. The TAM 2 pathway (30% in first-line failures and 27% in second-line failures) was found to be more

common than the TAM 1 pathway which contrasts with that which is known in subtype B, where the TAM 1 pathway is twice as frequent as the TAM 2 pathway (Armstrong et al., 2009). The majority of the TAMs from both pathways (M41L, T215Y, D67N, K70R, T215F and K219Q) were found to be more frequent in the subtype C second-line failures.

The low frequency of protease mutations could be attributed to the high genetic barrier of LPV i.e. numerous mutations are required to cause resistance to the ARV agent (Kempf et al., 2001), or poor adherence. As adherence is an ongoing problem, patients who were enrolled into our study had participated in intensified adherence counselling. However many still presented with no resistance mutations. This was noticed at both centers, the Sinikithemba Clinic at McCord Hospital as well as the ARV clinic at King Edward Hospital. Therefore, the results were not related to samples collected from a particular clinic. The viral loads did not appear to be related to whether or not mutations were present or not. Viral load is often used as a surrogate marker for adherence. There were patients with viral loads >100000 cpm (n=4) and some with viral loads as low as 1400 cpm that did not harbour any resistance mutations in protease or reverse transcriptase. This suggests an alternate mechanism of protease inhibitor drug resistance via mutations occurring at the protease CS which might be in play. This has been previously described (Nijhuis et al., 2007). Previous studies have shown that mutations occurring in *gag* not only act as compensatory mutations but can directly affect PI drug resistance (Dam et al., 2009, Malet et al., 2007, Nijhuis et al., 2007).

CHAPTER 3

Molecular Characterisation of HIV-1 Protease Cleavage Sites

3.1 Introduction

HIV-1 protease is responsible for cleavage of its *gag* and *gag-pol* substrates in order for mature and infectious viral particles to be formed. Resistance to PIs occurs by the stepwise accumulation of mutations in protease itself which consequently lead to a reduced replication capacity of the virus (Condra et al., 1996, Molla et al., 1996, Croteau et al., 1997). Therefore mutations which arise in *gag* restore the replicative capacity of the virus and are selected for both at CS and non-CS (Gatanaga et al., 2002, Myint et al., 2004, Tamiya et al., 2004).

Since substitutions at the CS can block processing, this can lead to the production of abberant and non-infectious viruses (Lambert-Niclot et al., 2008). Many studies have investigated the characteristics of sequences at the protease CS to gain a better understanding of the protease-substrate interaction. It was found that the amino acids which flank the target scissile bond at the cleavage sites are generally hydrophobic (Henderson et al., 1990). Non-CS substitutions as well as inserts in the proximity of the *gag* CS of PI-resistant HIV-1 variants have been shown to restore the otherwise compromised activity of the mutant protease (Tamiya et al., 2004). *Gag* CS mutations have also been shown to confer resistance in the absence of protease mutations (Nijhuis et al., 2007, Dam et al., 2009).

To further investigate the characteristics of amino acids at the CS, HIV-1 *gag* was sequenced and analysed. Sequencing of *gag* was performed directly on a 700- and 896-bp PCR product (nucleotides 1157 to 2549, relative to HXB2).

3.2 Materials and Methods

3.2.1 RNA Extraction

RNA was extracted using the ViroSeq extraction module Chapter 2.2.4.1. RNA was extracted from the stored plasma of first-line failures using the QIAamp Viral RNA Mini Kit (QIAGEN, Germany) as described in Chapter 2.2.5.1.

3.2.2 Reverse Transcription and Amplification of HIV-1 Protease Cleavage Sites

The cDNA master mix was prepared in a pre-amplification area. The 5 X First strand buffer, DTT and random hexamers was thawed and vortexed for 5 seconds to mix. The enzymes RNase OUT and Superscript III were not vortexed.

The cDNA master mix was prepared in a 1.5 mL RNase/DNase free microcentrifuge tube as follows:

Reagent	Final concentration
5 X First Strand Buffer	X1
0.1 M DTT	0.005M
RNase OUT (40U/µL)	0.005M
Superscript III (200U/µL)	10U/µL
Final volume	**3.5**

Two µL of random hexamer primer and 1 µL of dNTPs were added to 0.2 mL MicroAmp reaction tubes. To each reaction tube, 4.5 µL of RNA was added and the tubes were placed in a PCR System 9700 thermocycler (Applied Biosystems, California) and run on the following program:

Temperature (°C)	Time
65	5 minutes
45	1 minute
Pause manually, add RT mix and resume reaction	
45	60 minutes
70	5 minutes
4	Hold (>10 minutes)

After the 45 °C step for 1 minute, 3.5 µL of the master mix was added to each reaction tube. Once the program had completed 0.5 µL of RNase H was added to each reaction tube and run for 20 minutes at 37 °C on the thermocycler.

A 1.4 kb fragment of *gag* and protease was amplified by nested PCR. The HXB2 positions and sequence of the primers used are shown below in Table 3.1. For the first-round PCR, the outer most primers i.e. Gas 1F and 3' InProt were used. The *gag* region was amplified to yield two regions of *gag*, the 5'-end which is 800bp and the 3'-end which is 896bp. The second-round PCR used the Gas 1F and Gas 6R to amplify the 5' end, while primers Gas 5F and 3'InProt amplified the 3' end as shown in Figure 3.1.

Table 3.1: Primers used to amplify the *gag* region

Primer Name	HXB2 Positions	Primer Sequence (5' – 3')
Gas 1F (10 pmol/µL)	1090 to 1109	TTAGACAAGATAGAGGAAGA
Gas 6R (10 pmol/µL)	1771 to 1790	AAAATGGTCTTACAATCTGG
Gas 5F (10 pmol/µL)	1697 to 1716	CTTTAAGAGCTGAACAAGCT
3' InProt (10 pmol/µL)	2569 to 2593	CCTGGCTTTAATTTTACTGGTACAG

The primer coverage is seen below in Figure 3.1.

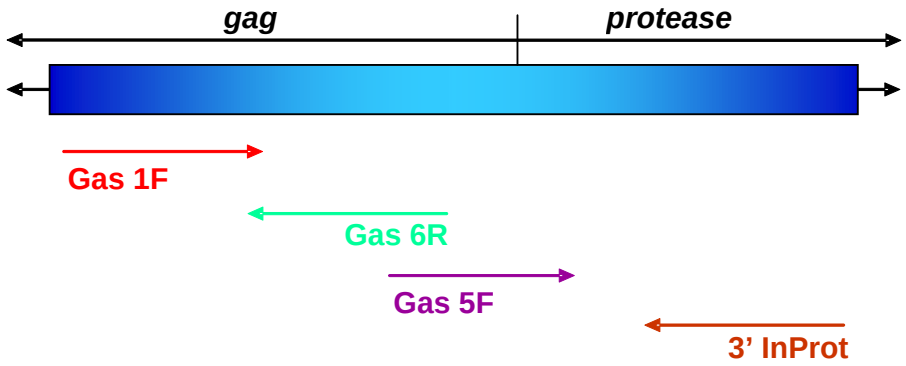

Fig 3.1 Diagram showing the layout of the primers used for the amplification of the HIV-1 protease cleavage sites. Gas 1F and 5F (forward primers) and Gas 6R and 3'InProt (reverse primers) amplify a 1.4kb fragment of *gag* and protease.

The 10 X Buffer, $MgCl_2$, dNTPs, primers and water was thawed and vortexed prior to use. The master mix for the first round PCR was prepared in a 1.5 mL microcentrifuge tube as follows:

Reagent	Final Concentration
10 X Buffer	x1
$MgCl_2$ (25 mM)	2 mM
dNTPs (10 mM)	0.2 µM
Gas 1F	0.2 µM
3' InProt	0.2 µM
Amplitaq Gold (5 U/µl)	1.25 U
Water	
Final volume	

Twenty-three µL of PCR mastermix was added to each appropriately labelled 0.2 mL reaction tube. Care was taken to have only one reaction tube open at a time while 2 µL of cDNA was added. The conditions for the first and second PCR were as follows:

Temperature (°C)	Time	Cycles
94	13 min	1
94	30 sec	
52	30 sec	40
72	1 min	
72	7 min	1
4	Hold	-

Following amplification, the PCR products were stored at -20 °C if the second PCR step was not done immediately. The master mix for the second round PCR was made up as follows:

Reagent	Final Concentration	Reagent	Final Concentration
10 X Buffer	X1	10 X Buffer	X1
$MgCl_2$ (25 mM)	2mM	$MgCl_2$ (25 mM)	2mM
dNTPs (10 mM)	1mM	dNTPs (10 mM)	1mM
Gas 1F (10 pmol/µl)	0.2 pmol/µl	Gas 5F (10 pmol/µl)	0.2 pmol/µl
Gas 6R (10 pmol/µl)	0.2 pmol/µl	3' InProt (10 pmol/µl)	0.2 pmol/µl
Amplitaq Gold (5 U/µl)	0.05 U/µl	Amplitaq Gold (5 U/µl)	0.05 U/µl
Water		Water	
Final volume	**48**	Final volume	**48**

Forty eight µL of PCR mastermix was added to each appropriately labelled 0.2 mL reaction tube. Care was taken to have only one reaction tube open at a time while 2 µL of first round PCR product was added. The PCR was run on the thermocycler following the same PCR conditions as that of the first round PCR.

3.2.3 Amplicon Purification and Quantification

Detection of amplicons was done by running a 1 % agarose gel as described in Chapter 2.2.5.3. The PCR products were purified using the QIAquick PCR Purification kit from QIAGEN and quantified as described in Chapter 2.2.5.3.

3.2.4 Sequencing of the HIV-1 Protease Cleavage Sites

The sequencing primers, BigDye and 5 X sequencing buffer were thawed and vortexed prior to use. A master mix for each primer was prepared as follows:

Reagent	Final Concentration
BigDye	Ready reaction mix
5 X Sequencing Buffer	X1
Primer (1.6 pmol/µL)	0.32 pmol/µL
Water	
Final volume	**7.5**

The sequencing reaction was set up in a MicroAmp optical 96-well reaction plate (Applied Biosystems, California), where 7.5 µL of each primer mix and 2.5 µL of the diluted PCR product was added to a designated well. The plate was sealed with adhesive foil cover strips and vortexed to allow the master mix and diluted PCR product to mix. The plate was briefly centrifuged before being transferred to a GeneAmp PCR System 9700 thermocycler. The following sequencing program was followed on the thermocycler:

No. of Cycles	Temperature	Time
1	96 ^{o}C	1 min
35	96 ^{o}C	10 sec
	50 ^{o}C	5 sec
	60 ^{o}C	4 min
-	4^{o}C	HOLD

The sequencing products were purified on the same day, as described in Chapter 2.2.5.4.

3.2.5 DNA Sequencing Analysis

In addition to the sequences generated, those downloaded from the Los Alamos Sequence Database of subtype C naïve, subtype B naïve and treated, and group M were analysed. Nucleotide sequences were aligned by CLUSTAL W (Thompson et al., 1994) and manually edited with the codon alignment of the Genetic Data Environment (GDE version 2.2) program (Smith et al., 1994).

Cleavage sites were trimmed from the original sequences and the amino diversity at each CS was measured using a Poisson distribution method implemented in the MEGA program version 4.0 (Arizona State Univerity, Tempe). The most recent common ancestor (MRCA) was downloaded from the Los Alamos Sequence Database (Shafer, 2006) and used to compare the amino sequence at the CS with the different groups. The mutations analysed in *gag* were based on literature and those which have been associated with resistance (Knops et al., 2010, Dam et al., 2009, Verheyen et al., 2009, Lambert-Niclot et al., 2008, Nijhuis et al., 2007).

3.2.5.1 Statistical Analysis

P values for the difference in frequencies between the groups were measured using the Fischer's exact test. *P* values for diversity measurements were calculated by applying a *t* test to the distance matrix of each data set. In order to account for multiple comparisons the Bonferroni's test was used to reduce the threshold of significance. The associations between the protease mutations and *gag* mutations were analysed and associations were tested using the Spearman rank order correlation in GraphPad Prism 5 (GraphPad Software, California).

3.3 Results

3.3.1 Mutational Patterns Occurring in *Gag*

3.3.1.1 Analysis of the Naïve Sequences from Subtype B, Subtype C and Group M

The mutations that were included in the analysis were found at CS (A431V, K436R, I437V/L, L449P/F and P453L/T) and non-CS (H219Q, V390L/I and E468G) in *gag*. These mutations were chosen since they are at sites related to drug resistance in subtype B.

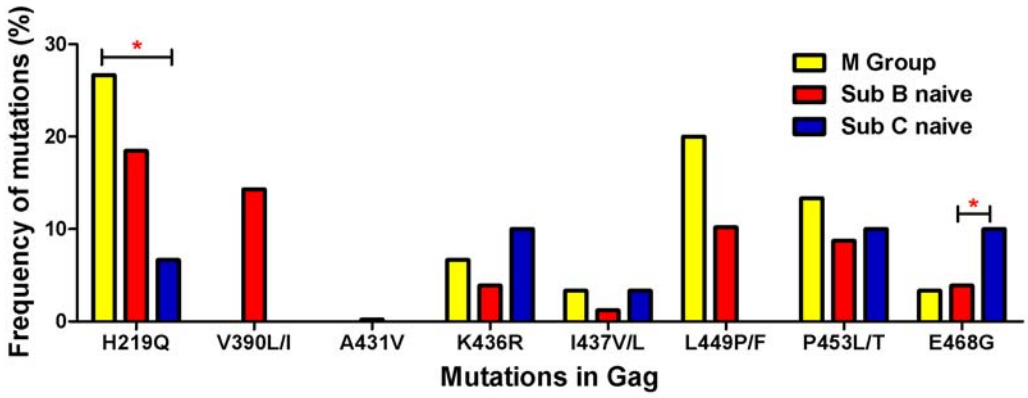

Fig 3.2 Frequency of mutations occurring in *gag* in subtype B, subtype C and M naïve group. A *P* value <0.05 was considered significant and is indicated by an asterix (*).

The *gag* amino acid sequences of M group, and subtype B and C naïves were compared to each other as shown in Figure 3.2. The M group sequences harboured mutations: H219Q, K436R, I437V/L, L449P/F, P453L/T and E468G in *gag*. The H219Q mutation was significantly ($P < 0.05$) more frequent in the M group (26.7%) compared to the C naïve group (6.7%). The subtype B naïve group harboured mutations: H219Q, V390L/I, K436R, I437V/L, L449P/F, P453L/T and E468G. The C naïves harboured mutations H219Q, K436R, I437V/L, P453L/T and E468G. The frequency of the mutation E468G in the C naives (10%) was significantly ($P < 0.05$) higher compared to the B naives (3.8%). Interestingly in the subtype C naïve group, mutations V390L/I, A431V and L449P/F were absent.

3.3.1.2 Analysis of Subtype C Naïve and Treated Sequences

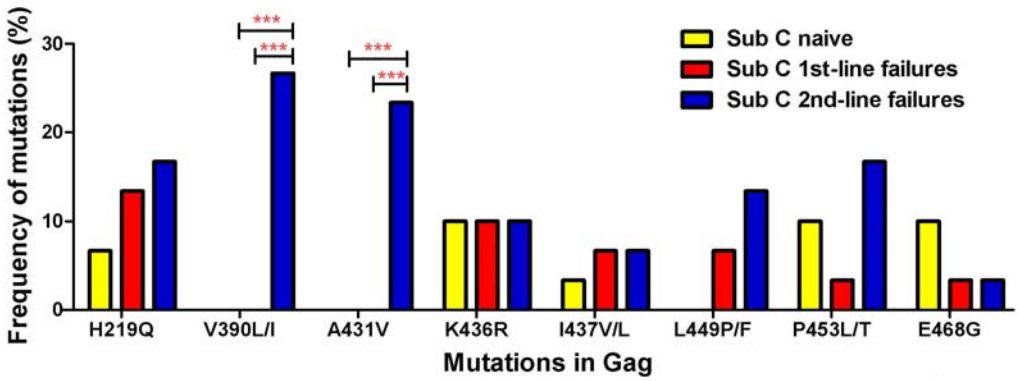

Fig 3.3 Frequency of mutations occurring in *gag* in subtype C naïve and treated patients. A *P* value <0.05 (*), <0.001 (**) and <0.0001 (***) was considered significant and are indicated by asterisks. The only two mutations which were present at significantly higher frequencies were V390L/I and A431V in the second-line failure group.

The frequency of *gag* mutations in subtype C naïve and treated patients were then compared as shown in Figure 3.3. The second-line failures harboured mutations in *gag* at all of the positions analysed (219, 390, 431, 436, 437, 449, 453 and 468). Of note, mutations V390L/I and A431V were only found in the second-line failures and at a significantly higher frequency compared to the naives and the first-line failures ($P < 0.0001$). In addition, the second-line failures had a higher frequency of H219Q, L449P/F and P453L/T, however this was not statistically significant. Surprisingly the naïve sequences had a higher frequency of the E468G (10%) mutation compared to the first-line failures (3.3%) and second-line failures (3.3%) but this was not significant.

From our findings it appears that mutation L449P/F only occurred in patients who were on first and second-line therapy, while mutations V390L/I and A431V were selected for during PI therapy since it was only present in the second-line failures.

3.3.1.3 Analysis of Subtype B Naïve and Treated Sequences

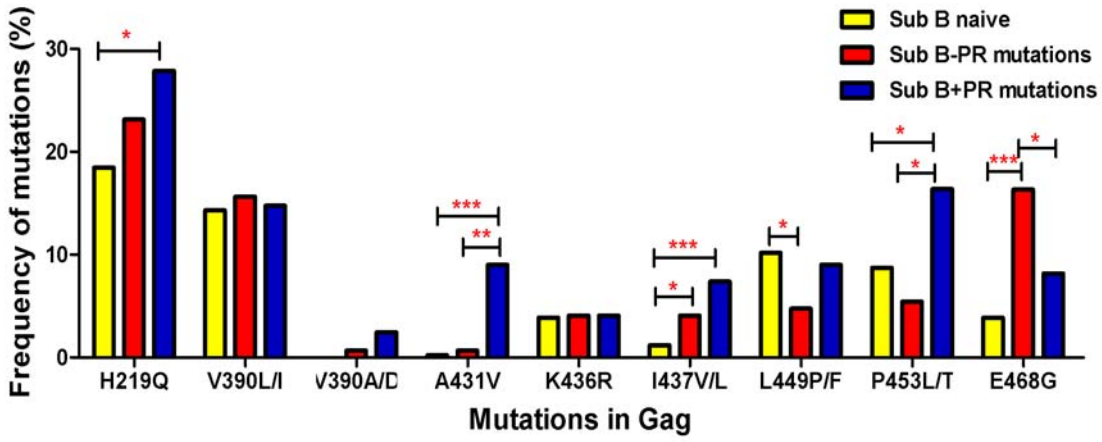

Fig 3.4 Frequency of mutations occurring in *gag* in subtype B naïve and treated patients. A *P* value <0.05 (*), <0.001 (**) and <0.0001 (***) was considered significant and are indicated by asterisks. Sequences harbouring protease mutations had a significantly higher frequency of *gag* mutations at positions 219, 431, 437, 453.

The frequency of *gag* mutations in subtype B naïve and treated patients were also analysed as shown in Figure 3.4. Mutations H219Q, A431V, I437V/L and P453L/T were found at a significantly ($P < 0.05$) higher frequency in the subtype B+PR group compared to the naïve sequences but only A431V and P453L/T remained significant when compared to the B+PR group. In contrast to the subtype C groups, the V390L/I mutation was present in all subtype B sequences at similar frequencies. This was interesting since the V390A/D mutations have been more commonly reported in subtype B. Mutations were present in the subtype B naives at all positions. Interestingly, the B naïve sequences had a higher frequency of the L449P/F mutation compared to the subtype B-PR ($P < 0.05$), but was not significantly different from the B+PR group ($P > 0.05$). The B-PR group had a significantly higher frequency of the I437V/L ($P < 0.05$) and E468G ($P < 0.0001$) mutation compared to the B naïve sequences suggesting that these mutations emerge before protease mutations emerge.

From our findings it is evident that mutation V390A/D is selected only during therapy in the subtype B patients. However, the V390L/I mutations seem to be polymorphic in subtype B since

it occurs in all subtype B groups at similar frequencies. The A431V mutation seems to be selected during PI therapy since it was at a significantly higher frequency in the B+PR group.

3.3.2 Correlation of Mutations in *Gag* with Protease Resistance Mutations

An analysis of the association between mutations in subtype C *gag* and protease was done as shown in Table 3.2. The only *gag* CS mutations which correlated significantly with protease mutations were A431V, L449F/P and P453L/T. The A431V mutation significantly correlated with protease mutations L10F/I/V, M46I, I54V/L, L76V and V82A while L449F/P and P453L/T were significantly correlated with I84V in protease.

Table 3.2 Association of mutations in *gag* with protease resistance mutations in subtype C second-line failures

		L10F/I	**M46I**	**I54V/L**	**A71T/V**	**T74S**	**L76V**	**V82A**	**I84V**
A431V	Spearman R	0.4458	0.6273	0.6707	0.2473	0.1182	0.5123	0.613	0.3366
	P value	0.0135	0.0002	<0.0001	0.1877	0.5338	0.0038	0.0003	0.0689
		*	***	***	ns	ns	**	***	ns
L449F/P	Spearman R	-0.0693	0.0154	0.1712	0.1346	0.0490	0.2942	0.1387	0.4734
	P value	0.7158	0.9354	0.3657	0.4782	0.797	0.1146	0.4649	0.0082
		ns	Ns	ns	ns	ns	ns	ns	**
P453L/T	Spearman R	0.0632	0.1762	0.2928	0.0877	0.2236	0.2236	0.0632	0.4152
	P value	0.7399	0.3516	0.1164	0.6449	0.2349	0.2349	0.7399	0.0225
		ns	Ns	ns	ns	ns	ns	ns	*

Table 3.3 Association of mutations in *gag* with protease resistance mutations in subtype C second-line failures

		L10F/I	M46I	I54V/L	A71T/V	T74S	L76V	V82A	I84V
H219Q	Spearman R	0.1438	0.152	0.1902	-0.1235	-0.0996	0.146	-0.0810	0.2776
	P value	0.1156	0.096	0.0367	0.177	0.2767	0.11	0.3769	0.002
		ns	Ns	*	ns	ns	ns	ns	**
A431V	Spearman R	0.1645	0.0847	-0.0385	0.1098	0.3193	-0.0288	0.41	0.1472
	P value	0.0714	0.355	0.6744	0.2306	0.0004	0.7533	<0.0001	0.107
		ns	Ns	ns	ns	***	ns	***	ns
I437V/L	Spearman R	0.2904	0.2019	0.0753	0.00820	0.3599	0.322	0.4593	0.0513
	P value	0.0012	0.0263	0.4117	0.9289	<0.0001	0.0003	<0.0001	0.5761
		**	*	ns	ns	***	***	***	ns
L449P/F	Spearman R	0.1021	-0.0484	-0.0385	0.0494	0.3193	-0.0288	0.41	0.0315
	P value	0.2651	0.5979	0.6744	0.5904	0.0004	0.7533	<0.0001	0.7312
		ns	Ns	ns	ns	***	ns	***	ns

An analysis of the association between mutations in subtype B *gag* and protease was done as shown in Table 3.3. The *gag* mutations H219Q, A431V, I437V/L and L449F/P were the only mutations that were found to be associated with mutations in protease. It was found that the non-CS *gag* mutation H219Q significantly correlated with I54V/L and I84V. The CS mutations A431V, I437V/L and L449F/P were found to correlate significantly with T74S and V82A while I437V/L also correlated significantly with L10F/I, M46I and L76V.

3.3.3 Longitudinal Analysis of Patients Failing Second-line Therapy

Seven of the patients in the study had genotypic data from earlier time points. Their drug histories and resistance profiles were analysed to determine the effects of ARV therapy over time. The patient history profiles were drawn on the ART-AIDe program at the Stanford Drug Resistance Database. Patients were separated based on those who were not previously exposed to PIs and those who were exposed.

3.3.3.1 Longitudinal analysis of patients previously not exposed to a protease inhibitor

Four of the seven patients were not previously exposed to a protease inhibitor. These patients were studied to analyse their resistance profiles before and after commencing a PI-inclusive regimen.

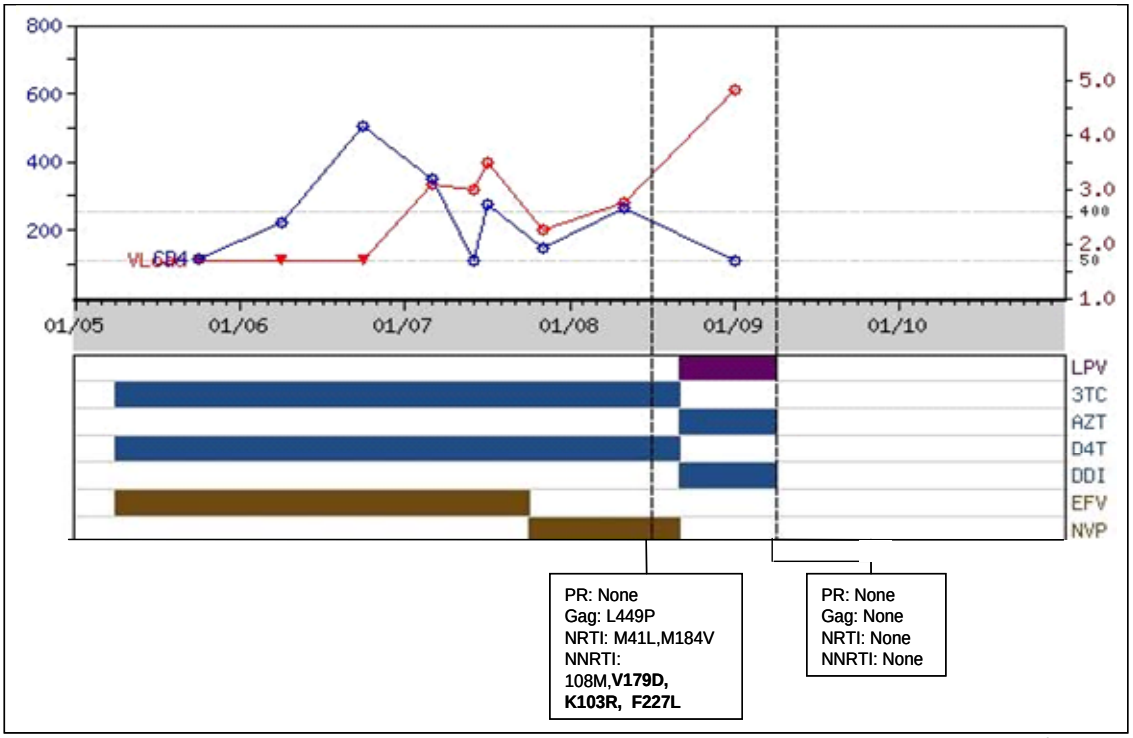

Fig 3.5 Drug history and resistance profile of patient PCSM004. The CD4 cells counts (cells/mm^3) (blue) are indicated on the left-sided y-axis and the log viral loads (red) are indicated on the right-sided y-axis. The dates on which the regimens were changed are shown on the x-axis. PIs are illustrated as purple bars, NRTIs as blue bars and NNRTIs as brown bars. The list of mutations is shown in open squares. For this patient, no mutations in the protease, *gag* and reverse transcriptase regions were present after commencing a PI-inclusive regimen.

Patient PCSM004 had been on D4T, 3TC and EFV prior to October 2007 when he was switched to NVP as shown in Figure 3.5. In July 2008, a genotype was performed which showed resistance to both NRTIs and NNRTIs and no mutations in protease. At this time point the L449P mutation was present in *gag*. The patient was then switched onto second-line therapy (AZT, DDI and Kaletra). In January 2009, the viral load had increased to 72000 cpm and the CD4 cell count dropped to 112 cells/ml. A second genotype was performed in April 2009, surprisingly all mutations that were previously present were lost after being exposed to the PI.

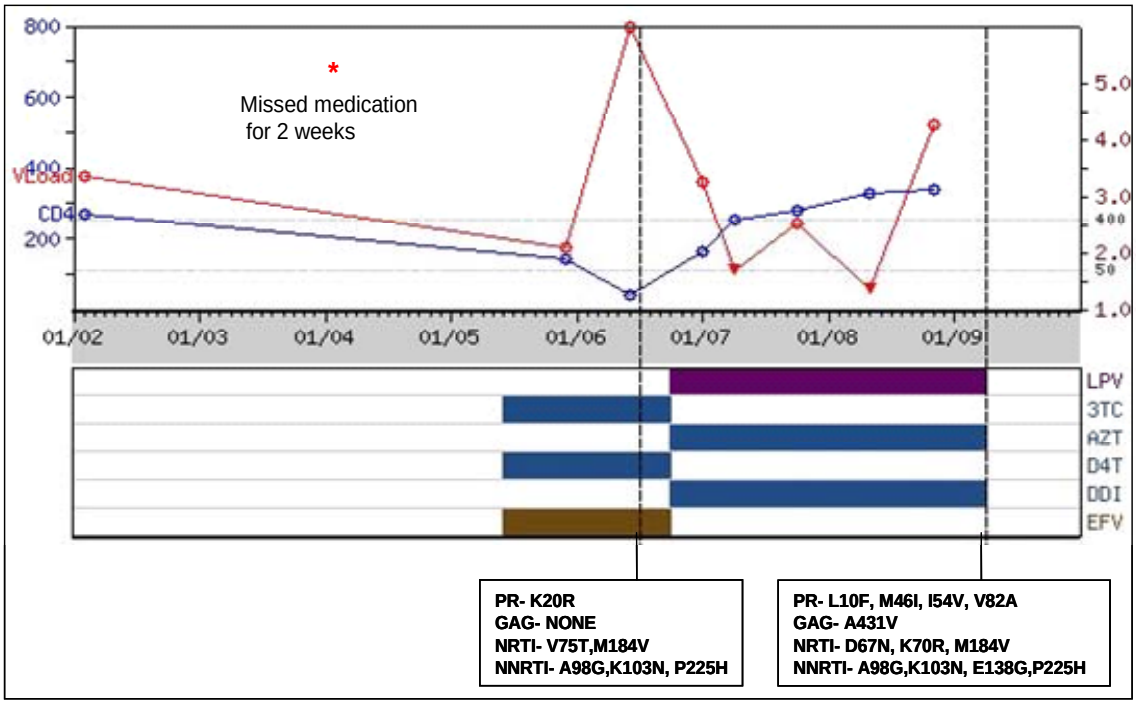

Fig 3.6 Drug history and resistance profile of patient PCSM005. This patient had a dramatic decrease in viral load after commencing a PI-inclusive regimen, however additional *gag* and reverse transcriptase mutations were acquired. Annotation of this figure is as described in Fig 3.5.

Patient PCSM005 had been on D4T, 3TC and EFV prior to July 2006 as shown in Figure 3.6. In July 2006, a genotype was performed which showed resistance to both NRTIs and NNRTIs. The resistance may have developed due to the patient missing two weeks of medication. During this time protease only harboured the accessory mutation K20R while no mutations were present in *gag*. The patient was then switched onto second-line therapy (AZT, DDI and Kaletra). In April 2009 the viral load had decreased to 2400 cpm while the CD4 cell count had increased to 344 cells/ml. During April 2009 a second genotype was performed which showed resistance to PIs, NRTIs and NNRTIs. During this time, the A431V *gag* CS mutation emerged after the patient was exposed to a PI.

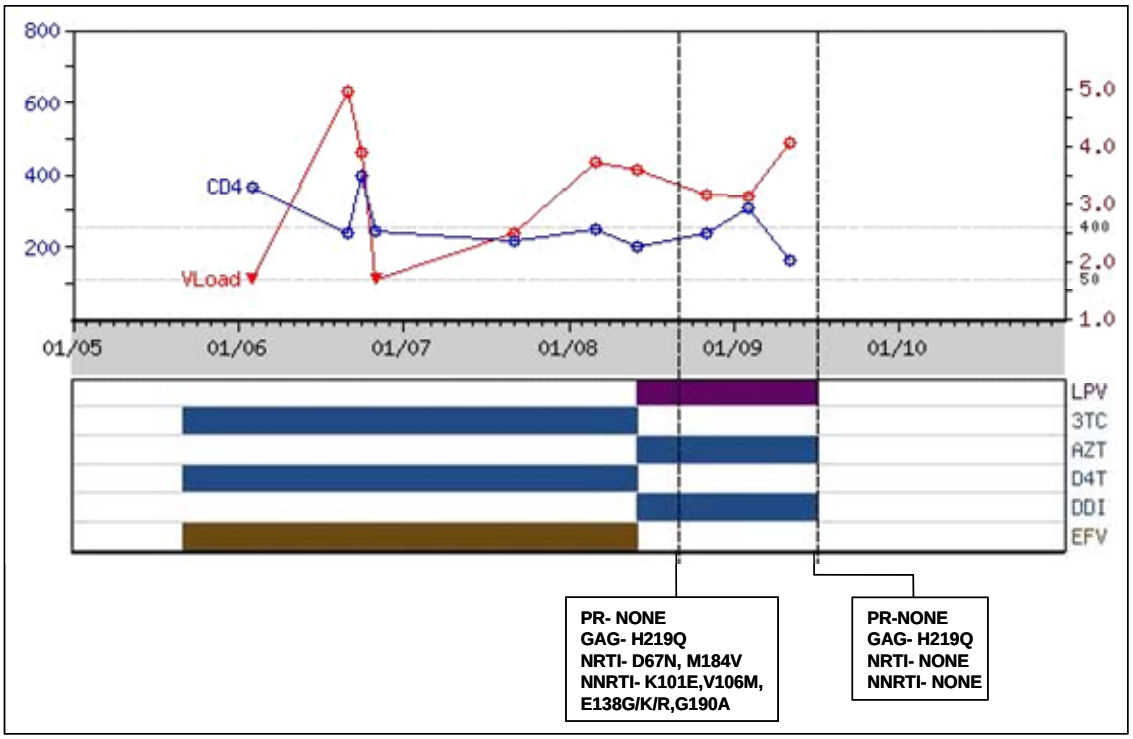

Fig 3.7 Drug history and resistance profile of patient PCSM009. This patient had lost mutations in reverse transcriptase following therapy with a protease inhibitor. Annotation of this figure is as described in Fig 3.5.

Patient PCSM009 had been on D4T, 3TC and EFV prior to September 2008 as shown in Figure 3.7. In September 2008, a genotype was performed which showed resistance to both NRTIs and NNRTIs and no mutations in protease. During this time the H219Q mutation was present in *gag*. The patient was then switched onto second-line therapy (AZT, DDI and Kaletra). In July 2009 the viral load had increased to 8300 cpm while the CD4 cell count decreased to 167 cells/ml. During this time a second genotype was performed. Surprisingly, all the mutations in RT which were previously present were lost after being exposed to a PI. However, the H219Q mutation in *gag* still persisted during this time.

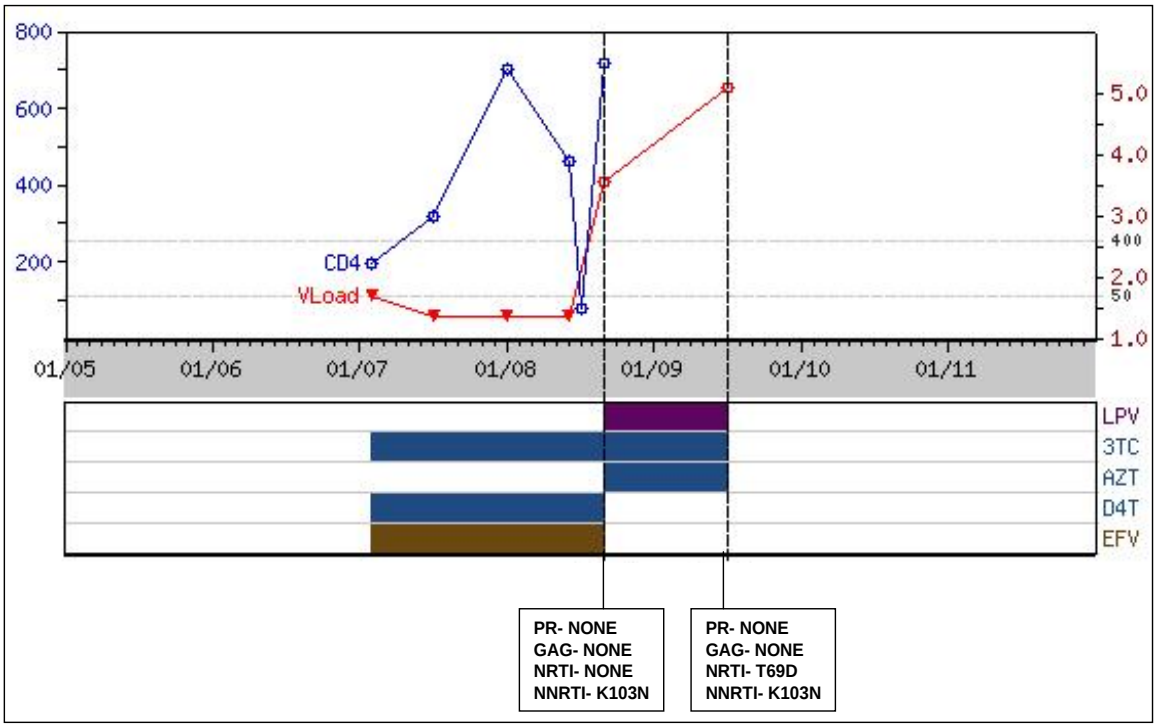

Fig 3.8 Drug history and resistance profile of patient PCSM012. This patient had acquired mutations in reverse transcriptase as well as an increase in viral load over five months. Annotation of this figure is as described in Fig 3.5.

Patient PCSM012 had been on D4T, 3TC and EFV prior to August 2008 as shown in Figure 3.8. In August 2008, a genotype was performed which showed resistance to NNRTIs while no mutations were present in protease and *gag*. The patient was then switched onto second-line therapy (AZT, 3TC and Kaletra). In July 2009 the viral load had increased to 13000 cpm and a second genotype was performed which showed resistance to both NRTIs and NNRTIs. No mutations had developed in protease or *gag*.

From these results it was found that patients who were on PI therapy for less than year did not develop resistance mutations in protease. This may be due to the high genetic barrier of LPV, therefore a longer time was required to acquire mutations. For the patient who had been on PI therapy for long (PCSM005), there was a gain of numerous protease mutations and one *gag* CS mutation. These findings imply that the development of PI resistance depends on the length of LPV treatment.

3.3.3.2 Longitudinal analysis of patients previously exposed to a protease inhibitor

Three patients who were previously exposed to a PI were studied to determine the effects of PI therapy on protease, *gag* and reverse transcriptase region of HIV-1.

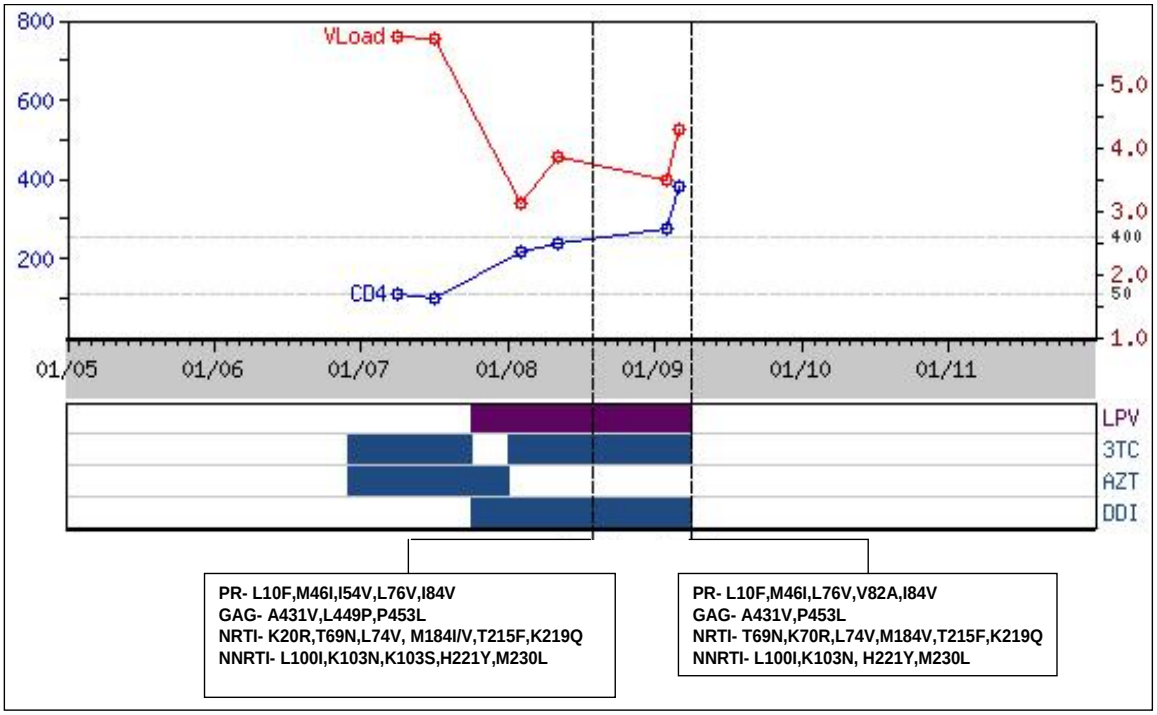

Fig 3.9 Drug history and resistance profile of patient PCSM001. This patient had numerous *gag*, protease and reverse transcriptase mutations while on a PI. Annotation of this figure is as described in Fig 3.5.

Patient PCSM001 had been on DDI, AZT and Kaletra before replacing AZT with 3TC in August 2008 as shown in Figure 3.9. In August 2008, a genotype was performed which showed resistance to PIs, NRTIs and NNRTIs. During this time, three CS mutations (A431V, L449P and P453L) were present in *gag*. In April 2009, the viral load had increased to 21000 cpm while the CD4 cell count also increased to 344 cells/ml. A second genotype was performed in April 2009 which still showed resistance to PIs, NRTIs and NNRTIs. However, some mutations in PR and RT were lost and replaced by others. During this time point the L449P *gag* mutation was lost.

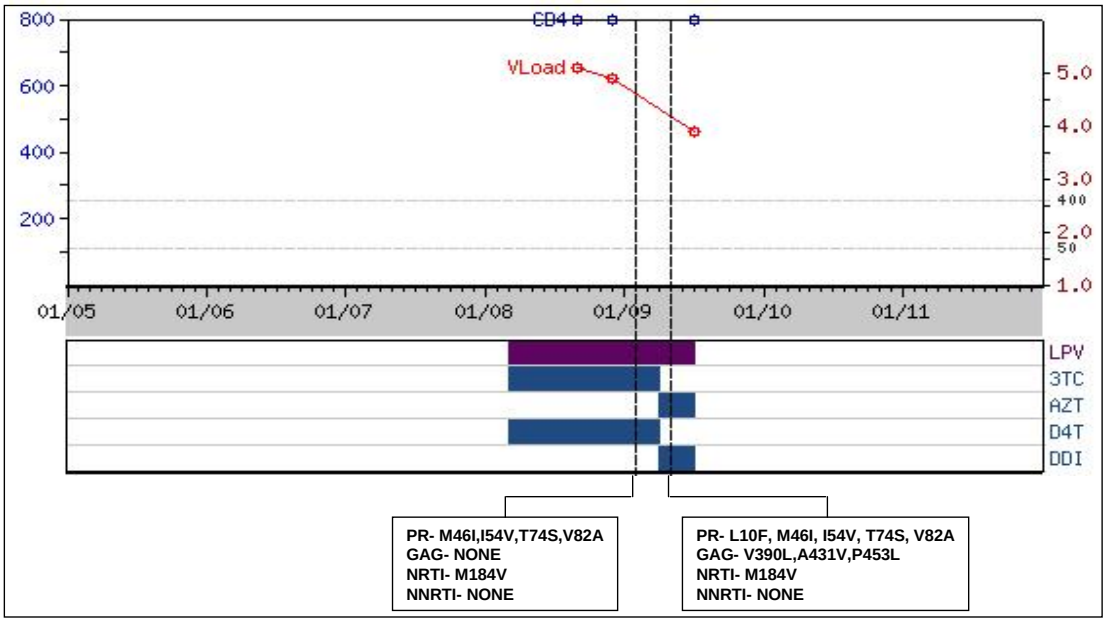

Fig 3.10 Drug history and resistance profile of patient PCSM002. Within a year patient PCSM002 had acquired additional mutations in protease and *gag*. Only the M184V mutation persisted in reverse transcriptase. Annotation of this figure is as described in Fig 3.5.

Patient PCSM002 had been on D4T, 3TC and Kaletra prior to March 2009 as shown in Figure 3.10. In February 2009, a genotype was performed which showed resistance to PIs and NRTIs. During this time no mutations were present in *gag*. In May 2009, there was a decrease in viral load to 84000 cpm and CD4 cell count to 1562 cells/ml. A second genotype was performed in May 2009 which showed resistance once again to PIs and NRTIs. However, the minor L10F mutation in protease had developed while the M184V mutation still persisted. During this time point, mutations at *gag* CS (A431V and P453L) and non-CS V390L had also developed with time.

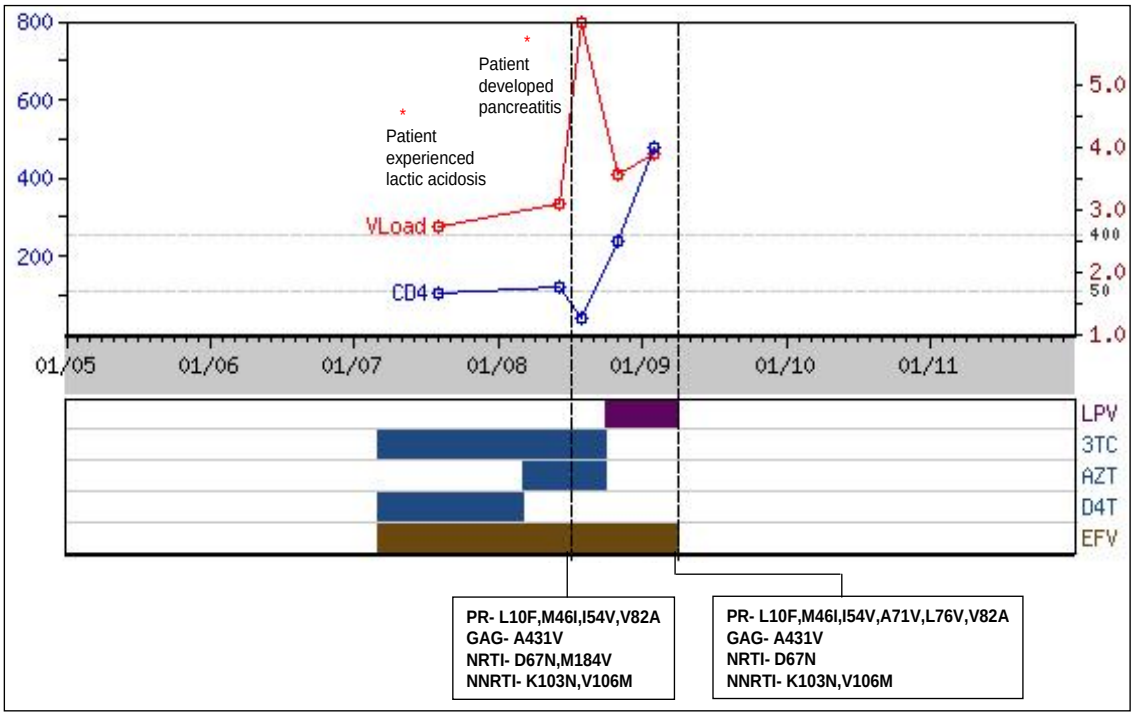

Fig 3.11 Drug history and resistance profile of patient PCSM007. Over time patient PCSM007 experienced an increase in viral load. There was an increase in protease mutations over time as well as the acquisition of a CS mutation in *gag*. Annotation of this figure is as described in Fig 3.5.

Patient PCSM007 had been on D4T, 3TC and EFV before replacing D4T with AZT in March 2008 since he experienced lactic acidosis while on D4T as shown in Figure 3.11. However, while on AZT, 3TC and EFV he developed pancreatitis. In October 2008 the patient was switched to second-line therapy (EFV and Kaletra). In July 2008, the viral load had decreased to 3800 cpm the CD4 cell count increased to 241 cells/ml. A genotype was performed in July 2008, which showed resistance to PIs, NRTIs and NNRTIs, while *gag* harboured CS mutation A431V. In April 2009 the viral load increased to 7900 cpm while the CD4 cell count had also increased to 480 cells/ml. A second genotype was performed in April 2009 which showed resistance once again to PIs, NRTIs and NNRTIs. During this time PR gained additional mutations (A71V and L76V) while the M184V mutation in RT was lost. However the A431V mutation in *gag* still persisted.

These findings justify what is already known in the literature that with prolonged PI therapy there is an accumulation of drug resistance mutations which was seen in two of the patients (PCSM002 and PCSM007). Patients who harboured major mutations in protease also presented with mutations in *gag*.

3.3.4 Genetic Diversity and Patterns of Amino Acid Variability at Individual Cleavage Sites

Variation at 9 CS of M group, subtype B naïve and subtype C naïve and treated is shown in Table 3.4. The genetic diversity was measured using the Poisson distribution method implemented in the MEGA program version 2.0. To determine if the sequences had evolved over time, the amino acid profiles for the *gag* and *gag-pol* CS were compared to the inferred MRCA for that site.

Table 3.4 Inter- and intrasubtype diversity at 9 cleavage sites, expressed as amino acid distances between sequences.

Protease Site	Mean % distance					*P* value					
	Sub B Naive	M Group	Sub C Naïve	Sub C 1st-Line Failures	Sub C 2nd-Line Failures	B naive vs C naive	B naive vs M group	C naive vs M group	C naive vs 1st-line failures	C naïve vs 2nd-line failures	1st-line failures vs 2nd-line failures
MA/CA	0.033	0.027	0.034	0.017	0.071	<0.0001*	0.285	<0.0001*	0.582	<0.0001*	<0.0001*
CA/p2	0.026	0.024	0.014	0.028	0.000	0.4363	0.7977	0.4013	<0.0001*	0.0059	<0.0001*
p2/NC	0.313	0.458	0.387	0.328	0.556	<0.0001*	<0.0001*	0.2234	<0.0001*	0.1141	<0.0001*
NC/p1	0.015	0.047	0.047	0.028	0.076	0.002	<0.0001*	0.6569	0.9747	0.0809	0.0967
p1/p6	0.092	0.114	0.105	0.019	0.15	<0.0001*	<0.0001*	0.0059	<0.0001*	<0.0001*	0.0057*
NC/TFP	0.015	0.15	0.038	0.024	0.076	<0.0001*	<0.0001*	<0.0001*	0.5944	<0.0001*	<0.0001*
TFP/P6pol	0.184	0.255	0.089	0.076	0.069	0.002	0.0222	0.0028*	0.0065*	0.1271	<0.0001*
P6pol/PR	0.046	0.299	0.075	0.103	0.023	0.0492	<0.0001*	<0.0001*	<0.0001*	0.0566	<0.0001*
PR/RT	0.002	0.027	0.031	n/a	n/a	<0.0001*	<0.0001*	0.7289	n/a	n/a	n/a

P values which remained significant following correction using the Bonferroni multiple comparisons test are indicated by an asterisk (*).

Variation at the 9 CS of subtypes B and C and group M is shown in Table 3.4 and Table 3.5. The *P* values were corrected for using the Bonferroni multiple comparisons test and those which were significant following correction are indicated by an asterisk (*). For the purpose of this study, we have referred to the patterns of diversity at the CS as conserved, moderately variable, and variable. Four (44.4%) sites (MA/CA, CA/p2, NC/p1 and PR/RT) were found to be relatively well conserved with an intersubtype distance ranging from 0.0% to 7.6% while the remaining five sites exhibited moderate (p1/p6^{gag}, NC/TFP, TFP/p6^{pol} and p6^{pol}/PR) to extensive (p2/NC) variation with intersubtype diversities reaching levels as high as 55.6 %.

An investigation of the naïve groups i.e. subtype B, C and group M revealed that overall, three CS (MA/CA, NC/TFP, p6^{pol}/PR) had significantly higher levels of diversity among subtype C naïve viruses ($P < 0.002$); while TFP/p6^{pol} was highly variable among B and group M viruses (mean distances, 18.4 and 25.5%, respectively) but relatively conserved in C viruses (mean distance, 8.9%). In all three groups the most variable CS was p2/NC with a mean divergence ranging from 31.3% to 45.8%. Group M was the most variable (mean distance, 45.8%) at this site compared to subtype B naïve ($P < 0.0001$) and C naive.

An investigation of the subtype C naïve and first-line failures revealed that two CS (CA/p2 and p6^{pol}/PR) were significantly more diverse ($P<0.0001$) in the first-line failures compared to the naives. Interestingly, four CS (MA/CA, NC/TFP, p2/NC and p1/p6^{gag}) had significantly higher levels of diversity among subtype C second-line failures ($P < 0.05$) compared to the first-line failures. Once again, in all three groups the most variable CS was p2/NC while the subtype C 2nd-line failures group was the most variable (mean distance, 55.6%) site compared to subtype C naïve ($P < 0.0001$) and the first-line failures.

3.4 Discussion

The occurrence of mutations in HIV-1 protease would be expected to result in compensatory changes occurring in the substrate CS which are recognised and cleaved by the protease enzyme (de Oliveira et al., 2003). To investigate this, the prevalence and patterns of subtype C CS mutations in *gag* and *gag-pol* polyproteins were compared to those of non-C viruses.

Most of the mutations found in *gag* which have been previously associated with PI resistance, were found to occur at the NC/p1 (A431V, K436R and I437V/L) and p1/p6^{gag} (L449P/F and P453L/T) CS. Numerous studies investigating PI resistance have focused on these two CS and have found them to significantly recover viral replicative capacity in the presence of a mutant protease (Doyon et al., 1996a, Mammano et al., 1998, Zhang et al., 1997, Coren et al., 2007). The fact that the A431V CS mutation was only seen in the subtype C second-line failures, strongly confirms its association with PI therapy. This mutation has been previously reported to occur with protease mutations M46I, I54V/L and V82A in subtype C (Dam et al., 2009, Verheyen et al., 2009, Verheyen et al., 2006, Lambert-Niclot et al., 2008).

Using the Spearman rank order correlation, significant association was found between the A431V *gag* mutation and protease mutations L10F/I/V, M46I, I54V/L, L76V and V82A. The L76V is a major protease mutation which causes high-level resistance to LPV. This mutation has been suggested to occur after patients have failed other PIs or accumulated other protease resistance mutations (de Mendoza et al., 2008). Four of the seven patients with L76V also harboured the A431V mutation in *gag*. These results suggest that A431V in *gag* may influence the selection of the protease mutation L76V. In the subtype B, we

observed a similar trend of the *gag* CS mutations being more prevalent in the group harbouring protease mutations. In addition, the non-CS mutation V390L/I was significantly more prevalent in the subtype C second-line failure group. Substitutions of an alanine or aspartic acid at the V390 position in *gag* were found to recover viral fitness in the presence of a mutant protease in subtype B (Gatanaga et al., 2002). The role of the leucine and isoleucine warrants further investigation.

From the small group of patients who were followed up, it was found that those who were not previously exposed to PI therapy did not harbour any major mutations in protease after switching to second-line therapy. However in two of these patients *gag* mutations A431V and V390L did appear after starting second-line therapy. The high genetic barrier of LPV is probably the reason why mutations did not develop within a short period of time. Patient's who were previously exposed to PI therapy still harboured numerous mutations in protease over time. All three of the patients who were previously exposed to a PI gained *gag* mutations within a year. A similar analysis using a larger group of patients would provide greater insights into the genetic profile of patients before and after commencing a PI-inclusive regimen.

Our study also revealed important differences between HIV-1 subtypes. Subtype C naïve viruses were significantly ($P < 0.0001$) more variable than B naïve viruses at MA/CA, p2/NC, NC/p1, p1/p6^{gag}, NC/TFP and PR/RT. One CS (p2/NC) was found to be the most variable across all subtypes, with the subtype C second-line failures being the most variable. Previous studies have shown that p2/NC is the first site that is cleaved and controls both the rate and order of *gag* and *gag-pol* polyprocessing (Shehu-Xhilaga et al., 2001, Pettit et al., 2005a, Pettit et al., 2005b). Three of the CS (MA/CA, CA/p2 and

NC/p1) were well conserved in all HIV-1 subtypes which suggests that mutations are not well tolerated in these regions. Cleavage of the MA/CA site plays an important role in virion maturation while cleavage of NC/p1 is needed for *gag-pol* expression (Erickson-Viitanen et al., 1989, Kaplan et al., 1994, Pettit et al., 1994). It has been suggested that slow, regulated cleavage of these structural proteins may be a common approach employed by the virus to ensure that the assembled virions have the full complement of proteins required for a new round of replication (Tessmer and Krausslich, 1998). Mutations which occur in the protease under drug selection pressure result in the development of compensatory mutations in the protease CS (Verheyen et al., 2006). Therefore the increased variability seen at the different CS may be an indication of improved viral fitness.

It was found that the second-line failures had an increased variability at two sites (MA/CA and NC/TFP) compared to C naïve viruses and four sites (MA/CA, p2/NC, p1/p6^{gag} and NC/TFP) compared to the first-line failures. The increased diversity at the MA/CA and p2/NC sites could be due to the increased compensatory changes which may have occurred in response to PI therapy. However, since these are structural proteins, their increased variability may affect the virus's replication capacity. Cleavage of the p1/p6^{gag} site releases the phosphoprotein p6^{gag}, which is an important protein required for the release of mature, infectious virions (Muller et al., 2002). Increased variability at this site in the second-line failures may lead to increased phosphorylation of the *gag* and *gag-pol* polyproteins thereby resulting in increased cleavage of these polyproteins.

The differences in variability and frequency of mutations in *gag* seen between the subtype B and C viruses may suggest that there are important differences in the way that C viruses

regulate viral processing and respond to antiretroviral therapy relative to that which has been observed in B viruses.

CHAPTER 4

The Effect of the Cleavage Site Mutation A431V on Viral Replication

4.1 Introduction

The protease enzyme is essential for viral replication. It is responsible for the cleavage of the *gag* and *gag-pol* precursors thereby releasing the functional proteins (Myint et al., 2004). In order for mature, infectious viruses to be formed, the enzyme has to recognise and bind to its substrate efficiently so that cleavage can occur. However, in the presence of drug selection pressure mutations arise in protease and this leads to the selection of compensatory mutations arising in *gag* (Nijhuis et al., 2007).

Certain mutations which arise in *gag* have also been shown to effect viral fitness. Two CS mutations, L449F and P453L, were shown to partially improve the replication fitness of viruses containing the I50V protease mutation and increased the degree of PI resistance (Doyon et al., 1996b). Mutation L449F has also been shown to improve fitness of viruses with the D30N, L90M and N88D mutations (Myint et al., 2004). In the absence of these CS mutations, the I50V mutant virus accumulated uncleaved p1/p6 proteins and this was partially corrected by the addition of one of these CS mutations. Mutations at CS which were found to recover viral fitness in PI-resistant viruses were most commonly found at the p2/NC, NC/p1 and p1/p6^{gag} CS.

In this chapter we concentrate on the effect of the A431V CS mutation on viral fitness in a subtype C virus. Previously this mutation has shown that it compensates for reduced viral fitness by a mutated protease. However, even in the absence of protease mutations the A431V mutation has shown that it can increase viral fitness in the presence of drug selection pressure (Dam et al., 2009). It has been found to recover viral fitness in drug experienced patients carrying mutations M46I/L, I54V, and V82A (Malet et al., 2007).

Based on the data from the previous chapter, where it was shown that the prevalence of the A431V mutation was highest in the second-line failures we decided that this mutation could be a good candidate for a primary PI resistance mutation in subtype C and we investigated its effect on replication capacity by site-directed mutagenesis.

4.2 Materials and Methods

4.2.1 Amplification of *Gag-Protease* Region from the C Naïve Virus

RNA was extracted from a subtype C virus (patient 254) using the QIAamp Viral RNA Mini Kit as previously described in Chapter 2.2.5.1, and an 837 bp fragment was amplified. The RNA was reversed transcribed using random hexamers as described in Chapter 2.2.5.2. The 837 bp fragment was amplified in a one-step RT-PCR.

The HXB2 positions and sequence of the primers used are shown below in Table 4.1.

Table 4.1: Primers used to amplify the *gag-protease* region

Primer Name	HXB2 Positions	Primer Sequence (5' – 3')
ApaI (10 pmol/µL)	1199→1226	ACATAGCCAGAAATTGCAGGGCCCCTAG
Sse8387 (10 pmol/µL)	2865→2835	CTGATTTTTTTCTGTTTTAACCCTGCGGATG

The 10 X Buffer, $MgCl_2$, dNTPs, primers and water was thawed and vortexed prior to use. The AmpliTaq Gold was briefly centrifuged to collect the contents at the bottom of the tubes. The master mix for the first round PCR was prepared in a 1.5 mL microcentrifuge tube as follows:

Reagent	Volume for 1 Reaction (µL)	Final Concentration
10 X Buffer	2.5	1X
$MgCl_2$ (25 mM)	2	1.5mM
dNTPs (10 mM)	0.5	0.2mM
ApaI (10pmol/µl)	0.5	0.2µM
Sse8387 (10pmol/µl)	0.5	0.2µM
Amplitaq Gold (5 U/µl)	0.25	1.25U
Water	16.75	
Final volume	**23**	

Twenty three µL of PCR mastermix was added to each appropriately labelled 0.2 mL reaction tube. Care was taken to have only one reaction tube open at a time while 2 µL of cDNA was added. The conditions for the PCR were as follows:

Temperature (°C)	Time	Cycles
94	13 min	1
94	30 sec	40
52	30 sec	
72	1 min	
72	7 min	1
4	Hold	-

The amplicon generated here was used as a template for the site-directed mutagenesis.

4.2.2 Site-Directed Mutagenesis

In vitro site-directed mutagenesis is a useful technique for studying protein structure-function relationships and gene expression. Stratagene's QuikChange Site-Directed Mutagenesis kit (Stratagene, Texas) allows site-specific mutations to be generated in almost any double-stranded plasmid. The QuikChange Site-Directed Mutagenesis kit is used to make point mutations, switch amino acids, and delete or insert single or multiple amino acids. The basic procedure uses a supercoiled double-stranded DNA vector with an insert of interest and two synthetic oligonucleotide primers containing the desired mutation. The primers, each complimentary to opposite strands of the vector, are extended during temperature cycling by *PfuTurbo* DNA polymerase. Incorporation of the primers

generates a mutated plasmid containing staggered nicks. The product is then treated with Dpn I following temperature cycling. The Dpn I endonuclease is used to digest the parental DNA template and to select for mutation-containing synthesised DNA. The nicked vector DNA containing the desired mutation is then transformed into competent cells.

4.2.2.1 Mutagenic Primer Design

To generate the A431V mutation mutagenic primers were designed using Stratagene's web-based QuikChange primer design program available online at http://www.stratagene.com/qcprimerdesign.

The following considerations were made when designing the mutatgenic primers:

- Both the mutagenic primers had to have the desired mutation and anneal to the same sequence on opposite strands of the plasmid.
- The primers had to be between 25 and 45 bases in length with a melting temperature T_m of ≥78 °C.
- The desired mutation was in the middle of the primer with ~10-15 bases of correct sequence on both sides.
- The primers should have a minimum GC content of 40 % and should end on one or more C or G bases.

The primers were designed on a subtype C naïve sequence lacking any resistance associated mutation in *gag* and protease. The following segment containing the appropriate segment was chosen from the C naïve sequence and uploaded onto the above mentioned website:

CACCAAATGAAAGACTGTACTGAGAGGCAG***GCT***AATTTTTTAGGGAAAATT TGGCCTTCCCACTA

In order to generate an A431V mutation, a 'C' nucleotide (underlined and italicised in the above sequence) had to be changed to a 'T' nucleotide. This would produce an amino acid change of alanine (GCT) to a valine (GTT). The output from the website is shown below.

Primer Name	HXB2 Position	Primer Sequence (5' to 3')
c32t	2060→2103	aagactgtactgagaggcaggttaatttttttagggaaaatttgg
c32t_antisense	2060←2103	ccaaattttccctaaaaaattaacctgcctctcagtacagtctt

Oligonucleotide information:

Primer Name	Length (nt.)	Tm
c32t	44	78.80°C
c32t_antisense	44	78.80°C

The primers designed met all the requirements for mutagenic primers. The primers were then ordered online from Roche Diagnostics, South Africa, where it was manufactured and HPLC purified.

4.2.2.2 The Mutagenesis Reaction

The mutagenesis PCR was the first step of the site-directed mutagenesis. The positive control plasmid and sample reactions were run in parallel to generate the desired mutation. Two sample reactions were set up individually using 10 and 30 ng of dsDNA while keeping the primer concentration constant. The PCR master mix preparation for the control and sample reactions are shown below:

Control Reaction - Final Concentrations	Sample Reaction - Final Concentrations
10 X reaction buffer - **1X**	10 X reaction buffer - **1X**
(10 ng) of pWhitescript - **0.4ng**	(10 and 30 ng) of dsDNA template -**0.4ng and 1.5ng**
(125 ng)of of oligonucleotide control primer #1 - **3.1ng**	(125 ng)of of oligonucleotide primer #1 - **3.1ng**
(125 ng)of of oligonucleotide control primer #2 - **3.1ng**	(125 ng)of of oligonucleotide primer #2 - **3.1ng**
10mM dNTP mix - **0.2mM**	10mM dNTP mix - **0.2mM**
39.5 µl of sterile water to a final volume of 50 µl	39 µl of sterile water to a final volume of 50 µl

Each master mix was made up in individual 0.2 mL microcentrifuge tubes. One µl of *PfuTurbo* DNA polymerase was added to each tube containing the master mix. The thermocycler conditions for the QuikChange Site-Directed Mutagenesis method was as follows:

No. of Cycles	Temperature	Time
1	95 ^{o}C	30 sec
16	95 ^{o}C	30 sec
	55 ^{o}C	1 sec
	68 ^{o}C	6 min
-	4^{o}C	2 min

Following temperature cycling, 1 µl of the Dpn I restriction enzyme was added directly to each amplification reaction. Each reaction mixture was microcentrifuged for 1 minute and immediately incubated at 37 °C for 1 hour in a thermocycler to digest the parental dsDNA.

4.2.2.3 Transformation of XL10-Gold Ultracompetent Cells

The XL10-Gold ultracompetent cells were treated very carefully. The cells were removed from the -80 °C freezer and immediately placed on ice to thaw. For each control and sample reaction to be transformed, 50 µl of the ultracompetent cells were aliquoted to a prechilled 14 mL BD falcon polypropylene round bottom tube. As an optional control to verify the transformation efficiency of the XL10-Gold ultracompetent cells, a pUC18 control plasmid was used. One µl of the pUC18 control plasmid was added to a 14 mL round bottom tube containing an aliquot of the cells. One µl of the Dpn I-treated DNA from each control and sample reaction was transferred to separate aliquots of the ultracompetent cells. The transformation reactions were swirled gently to mix and incubated on ice for 30 minutes. Each transformation reaction was then heat pulsed for 45 seconds at 42 °C and then immediately placed on ice for 2 minutes. Five hundred µl of S.O.C medium was added to the transformation reactions which were then incubated at 37

°C for 1 hour in a shaking incubator. Fifty µl of the X-gal and 20 µl of IPTG was plated onto each agar plate. The appropriate volume of each transformation was then plated on agar plates as indicated in the table below.

Table 4.2: Transformation reaction plating volume

Reaction Type	Volume to Plate
pWhitescript mutagenesis control	250 µl
pUC18 transformation control	5 µl in 200 µl of S.O.C medium
Sample mutagenesis	250 µl on each of two plates

The plates were incubated at 37 °C for 16hrs. Master plates for each dsDNA concentration was then prepared. Colonies from each plate was picked and touched onto the master plate before being transferred to a 0.2 mL microcentrifuge tube containing 10 µl of sterile water. The master plates were incubated at 37 °C for 48 hrs before being stored at 4 °C. Each 0.2 mL microcentrifuge tube containing the colonies were then boiled in a thermocycler at 90 °C for 10 mins and microcentrifuged at 12000 X g for 10 minutes. Following centrifugation, the supernatant from each tube was collected and used as the template DNA in the screening PCR.

4.2.2.4 Screening of Mutants

A master mix for the PCR was prepared as follows:

Reagent	Volume for 1 Reaction (µL)	Final Concentration
10 X Buffer	2.5	1X
25mM $MgCl_2$	2	1.5mM
10mM dNTPs	0.5	0.2mM
10pmol/ul Gas 5F	0.5	0.2µM
10pmol/ul 3' InProt	0.5	0.2µM
AmpliTaq Gold(5U/ul)	0.25	1.25U
Water	8.75	-
Template DNA	10	-
Final Volume	**25**	

The PCR was run under the following cycling conditions:

Temperature (°C)	Time	Cycles
94	13 min	1
94	30 sec	
52	30 sec	40
72	1 min	
72	7 min	1
4	Hold	-

A 1 % agarose gel was then run as described in Chapter 2.2.5.3 to confirm the presence of the desired clone and to check if the clone is of the correct size as that amplified. Samples sets were randomly chosen from each plate. The primers used produce a fragment that is 896 bp long. The PCR products were sequenced as previously described in Chapter 2.2.5.4 to check for the presence of the desired mutation, A431V. The PCR primers were used for sequencing at a concentration of 1.6 pmol/µl.

The 1.7 kb product containing the A431V mutation was then amplified to remove the product from the TOPO vector under the following conditions:

Reagent	Volume for 1 Reaction (µL)	Final Concentration
10 X Buffer	2.5	1X
25mM $MgCl_2$	2	1.5mM
10mM dNTPs	0.5	0.2mM
10pmol/ul Gag+1	0.5	0.2µM
10pmol/ul 3'RVP	0.5	0.2µM
AmpliTaq Gold (5U/ul)	0.25	1.25U
Water	16.75	-
Template DNA	2	-
Final Volume	**25**	

The PCR was run under the following cycling conditions:

Temperature (°C)	Time	Cycles
94	2 min	1
94	15 sec	
55	30 sec	35
72	1 min 30 sec	
72	7 min	1
4	Hold	-

The resulting amplicon comprised of *Gag* and Protease.

4.2.3 Transfection of CEM-GXR 25 Cells

For site-directed mutagenesis a pNL4-3ΔGag-Pro plasmid was used that was a gift from Ms Jaclyn Wright and Dr Mark Brockman. This plasmid was digested in a 1.5 ml tube under the following conditions:

Reagent	Volume
Plasmid	10 µg/sample
10 X BSTEII Buffer	1/10 reaction volume
100 X BSA	1/100 reaction volume
BSTEII Enzyme	2 U/µg plasmid
Water	To make up final volume

The plasmid was digested for 2 hrs at 60 °C in a waterbath. Following digestion CEM-GXR 25 cells were cotransfected by electroporation with the digested pNL4-3ΔGag-Pro plasmid and 1.8kb PCR product to generate a chimeric virus. 80 µl of the PCR product, 10 µg of plasmid per sample and 800 µl of CEM-GXR 25 cells (containing ~ 2 million cells) were pipette-mixed in a matrix cluster-tube before being transferred to an electroporation cuvette. The samples were electroporated for: 300 V, 500 µF, R= ∞. The samples were then incubated in the cuvettes for 1 hour at room temperature before adding the contents of the tube to 25 ml flasks containing 4 ml media. The flasks were incubated at 37 °C and after 5 days, 5 ml of R-10^{+} media was added to the flasks. From day 12 the samples were run daily on the flow cytometer to determine the percentage of infected cells and harvested when they reached ~30% infectivity.

The viruses were titered to achieve a 0.6% infectivity on day 2 of the replication assay. The titer was set up in a 24-well plate where 1 million cells in 100 µl were plated per well and 400 µl of stock virus was added. A negative control was also set up which included media

instead of virus. The plate was incubated overnight and 1 ml of media was added to each well the following day. On day 2 the percent infectivity was measured by flow cytometry.

4.2.4 Replication Assay

The replication assay was set up in a 6-well plate in duplicates as shown in Figure 4.1. The total volume in each well was 9 mL.

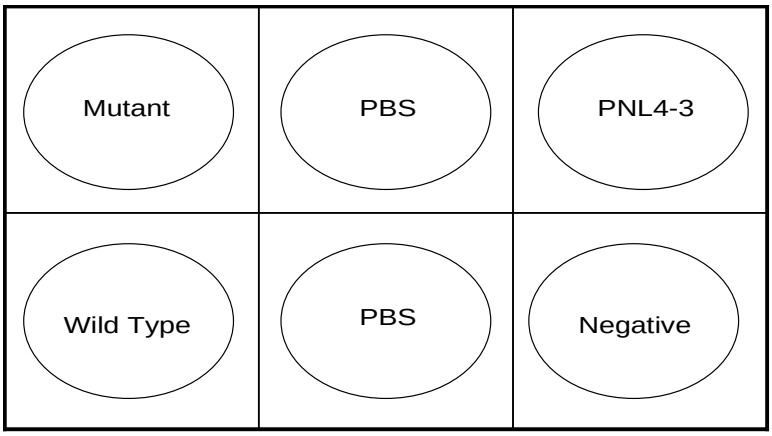

Fig 4.1. Layout of plate for the replication assay. The experiment was performed in duplicate. A positive (PNL4-3) and a negative control (media) was also included.

The cells were plated at a concentration of 1 million cells/100µl and 600µl per well. So 8 wells= 48 million cells/4800ul media. Enough volume from flasks was removed for 48 million cells. The cells were spun down 1800 x g for 10 mins at room temperature. The supernatant was discarded and 4800 µL of warmed media was added to the cells and mixed well. Six hundred microliters of cells were added to each well.

Based on results from the titer, all viruses were added in different amounts to achieve a 0.6% infection on day 2.

Amount of virus added (μL) = [0.6/ (% total cells fluorescing on day 2 of titre experiment)] x 2400 μL

This amount was subtracted 2400 in order to obtain the number of μL of media to add.

	Titer (%)	Amount of Virus (μL)	Amount of Media (μL)
Mutant	1.2	1200	1200
Wild Type	0.9	1600	800
PNL4-3	1.1	1310	1090
Negative	-	-	2400

The calculated virus and media were added to 15 ml tubes and vortexed to mix followed by 2400 μL of the virus/media mix being added to the designated wells. Phosphate buffered saline (PBS) were added to the two remaining wells. The plates were incubated overnight at 37 °C. The following day (day 1), 6 mL of warm media was added to each well. On day 2 till day 6, 2500 μL of virus was stored while 500 μL was used to measure the percent of cells infected by flow cytometry. Each day 3 mL of warm media was added to each well till the end of the six day assay.

4.3 Results

4.3.1 Replication Capacity of Mutant HIV-1

The growth and infectivity rates of the viral variants were calculated based on the enhanced green fluorescent protein (eGFP) expression. The mean slope was determined by best-fit analysis to an exponential curve. The *P* values were calculated by Students *t* test.

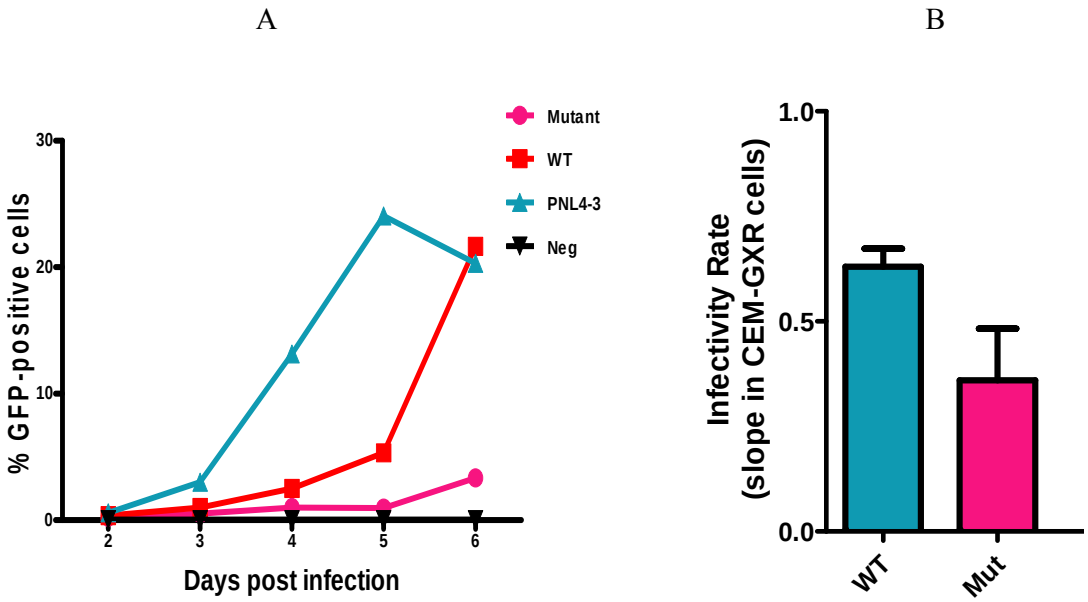

Fig 4.2. Replication kinetics of mutant HIV-1. CEM-GXR25 cells were infected in duplicate with viruses carrying the A431V mutation and compared to those infected with the wild-type. The replication kinetics of the viruses were followed daily after infection. (a) Viral infectivity was measured by the percentage of $eGFP^+$ cells by FACS analysis. (b) The infectivity rate was calculated as the slope of log $eGFP^+$ cells measured in the kinetics experiments.

The replication assay was run over six days and virus was sampled each day. The percent infectivity was measured by flow cytometry. There was a significant decrease in the growth rate however this was not statistically significant (P =0.0987) and infectivity rate for the A431V mutant compared with the wild type as shown in Figure 4.2. The growth rate of the mutant virus was 3.3% compared to the wild type which was 21.7% as shown in Figure 4.2a. On day six the infectivity of the positive control decreased, this may be due to the initial high infection rate of the virus as well as a depletion of healthy cells to infect by day six. The mutant virus had an infection rate almost half of that of the wild type (0.36 vs 0.63). However, throughout the six days of the assay both the mutant and the wild type grew slower than the positive control and only started to grow exponentially on day six and day five respectively.

4.4 Discussion

Many studies have reported on the improved viral fitness seen in drug-resistant subtype B viruses harbouring the A431V mutation in patients with mutations such as M46I/L, I54V and V82A (Dam et al., 2009, Malet et al., 2007, Knops et al., 2010). We decided to look at this mutation independent of drug selection pressure and changes in protease since it will add to our knowledge of mutations at *gag* CS.

The A431V mutation in *gag* was studied since it was the most prevalent mutation which only occurred in the second-line failures and was found to correlate positively with other mutations in protease.

From our study it was found that the mutant virus had a slower growth rate compared to the wild type as well as a viral infectivity half of that of the wild type. This may be due to the protease binding less efficiently to its cleavage site thereby infecting fewer cells and having less replication. The subtype C viruses (both mutant and wild type) grew slower than the subtype B positive control. This maybe an indication of differences in the way that subtype B and C viruses regulate *gag-pol* processing. However, to confirm this, a larger set of B and C viruses should be compared to each other. The A431V mutation is located at the generally conserved NC/p1 cleavage site. Cleavage at this site is essential for virus assembly (Freed, 2001). In order for efficient cleavage to occur protease has to recognise and cleave *gag*.

Thus far most studies have described the CS as five amino acids on either side of the scissile bond. The amino acid positions at the CS are named as P1-P5 (left to the scissile bond) and P1'-P5' (right to the scissile bond). Studies have described that the amino acids flanking the scissile bond are generally hydrophobic (Pettit et al., 2002). For the A431V

mutation, valine replaces alanine. However, both amino acids are polar. Valine is has a molar mass of 117.15 $gmol^{-1}$ compared to alanine which is 89.09 $gmol^{-1}$ (Mathews et al., 2000). The substitution of a larger amino acid could lead to a change in conformation of the CS therefore affecting the binding and cleavage capacity. If this theory is correct then substitutions within the P4-P4' region of the different cleavage sites should be further considered for its role in viral fitness.

Future studies on CS mutations in subtype C should look at the different combinations of mutations such as A431V-M46I/L, A431V-I54V and A431V-V82A in the presence and absence of drug selection pressure. Also of interest would be the modelling of the mutated *gag* substrate with the wild type protease, as well as the mutated protease. This would allow us to see the differences in which the enzyme and substrate bind.

This study provided an understanding of the effect of the A431V CS mutation on viral fitness. A further investigation into the drug susceptibility of the mutant virus compared to the wild type virus would provide us with a better understanding of the effect of this mutant on drug resistance in subtype C viruses.

CHAPTER 5

Conclusion

While great efforts have been made to reduce morbidity and mortality by providing ARV treatment to those in need, a huge problem which we are faced with is the widespread development of drug resistance. Thus far most studies have focused on the response of subtype B viruses to ARVs compared to subtype C. Therefore, it is essential to determine the patterns of drug resistance among subtype C viruses which are predominant in Southern Africa.

Subtype specific differences in variability and frequency of mutations seen in the study may suggest that there are important differences in the way that C viruses regulate viral processing and respond to antiretroviral therapy relative to that which has been observed in B viruses. Most algorithms used to detect drug resistance are designed on subtype B therefore when analysing subtype C viruses caution should be taken when interpreting patterns of drug resistance.

The low number of patients presenting with resistance mutations in protease may be due to the high genetic barrier of LPV which is supported by the findings of the seven patients which were followed up and required a longer period of time to develop resistance mutations. The potency of LPV can be highly beneficial in controlling HIV progression in patients if they adhere to the regimen and are constantly monitored for viral load and CD4 deviations. Most studies thus far have shown that CS mutations improved the fitness of the virus with a mutated protease. We were able to demonstrate that the A431V mutation decreases viral fitness in the absence mutations in protease and drug selection pressure.

Despite study limitations including the small sample size, possible patient adherence issues and the impact of different/suboptimal therapeutic regimens on the resistance patterns, our findings expose the importance of *gag* CS mutations in drug resistance and viral fitness. It is expected that with an increasing number of patients accessing treatment in future and an increased number of subtype C patients being exposed to PI therapy, future studies will reveal fundamental aspects of the interactions between *gag* CS and protease to help us better understand viral pathogenesis.

References

ANDRIES, K., AZIJN, H., THIELEMANS, T., LUDOVICI, D., KUKLA, M., HEERES, J., JANSSEN, P., DE CORTE, B., VINGERHOETS, J., PAUWELS, R. & DE BETHUNE, M. P. 2004. TMC125, a novel next-generation nonnucleoside reverse transcriptase inhibitor active against nonnucleoside reverse transcriptase inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob Agents Ch,* 48**,** 4680-6.

ANTINORI, A., ZACCARELLI, M., CINGOLANI, A., FORBICI, F., RIZZO, M. G., TROTTA, M. P., DI GIAMBENEDETTO, S., NARCISO, P., AMMASSARI, A., GIRARDI, E., DE LUCA, A. & PERNO, C. F. 2002. Cross-resistance among nonnucleoside reverse transcriptase inhibitors limits recycling efavirenz after nevirapine failure. *AIDS Res Hum Retroviruses,* 18**,** 835-8.

APPLIED BIOSYSTEMS 2000, *ViroSeq HIV-1 Genotyping System for the Analysis of HIV-1 Drug Resistance Mutations*, accessed 3 November 2009, <http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/>

ARMSTRONG, K. L., LEE, T. H. & ESSEX, M. 2009. Replicative capacity differences of thymidine analog resistance mutations in subtype B and C human immunodeficiency virus type 1. *J Virol,* 83**,** 4051-9.

BARBOUR, J. D. & GRANT, R. M. 2004. The Clinical Implications of Reduced Viral Fitness. *Curr Infect Dis Rep,* 6**,** 151-158.

BERKHOUT, B. 1999. HIV-1 evolution under pressure of protease inhibitors: climbing the stairs of viral fitness. *J Biomed Sci,* 6**,** 298-305.

BRENNER, B. G., ROUTY, J. P., PETRELLA, M., MOISI, D., OLIVEIRA, M., DETORIO, M., SPIRA, B., ESSABAG, V., CONWAY, B., LALONDE, R., SEKALY, R. P. & WAINBERG, M. A. 2002. Persistence and fitness of multidrug-resistant human immunodeficiency virus type 1 acquired in primary infection. *J Virol,* 76**,** 1753-61.

BUONAGURO, L., M., TORNESELLO, M. L. & BUONAGURO, F. M. 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J Virol,* 91, 10209-10219.

CABANA, M., FERNANDEZ, G., PARERA, M., CLOTET, B. & MARTINEZ, M. A. 2002. Catalytic efficiency and phenotype of HIV-1 proteases encoding single critical resistance substitutions. *Virology,* 300**,** 71-78.

CLAVEL, F. & HANCE, A. J. 2004. HIV drug resistance. *N Engl J Med,* 350**,** 1023-35.

COHEN, M. S., HELLMANN, N., LEVY, J. A., DECOCK, K. & LANGE, J. 2008. The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. *J Clin Invest,* 118**,** 1244-54.

COLONNO, R., ROSE, R., MCLAREN, C., THIRY, A., PARKIN, N. & FRIBORG, J. 2004. Identification of I50L as the signature atazanavir (ATV)-resistance mutation

in treatment-naive HIV-1-infected patients receiving ATV-containing regimens. *J Infect Dis,* 189**,** 1802-10.

CONDRA, J. H., HOLDER, D. J., SCHLEIF, W. A., BLAHY, O. M., DANOVICH, R. M., GABRYELSKI, L. J., GRAHAM, D. J., LAIRD, D., QUINTERO, J. C., RHODES, A., ROBBINS, H. L., ROTH, E., SHIVAPRAKASH, M., YANG, T., CHODAKEWITZ, J. A., DEUTSCH, P. J., LEAVITT, R. Y., MASSARI, F. E., MELLORS, J. W., SQUIRES, K. E., STEIGBIGEL, R. T., TEPPLER, H. & EMINI, E. A. 1996. Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor. *J Virol,* 70**,** 8270-6.

COREN, L. V., THOMAS, J. A., CHERTOVA, E., SOWDER, R. C., GAGLIARDI, T. D., GORELICK, R. J. & OTT, D. E. 2007. Mutational analysis of the C-terminal Gag cleavage sites in human immunodeficiency virus type 1. *J Virol,* 81**,** 10047-10054.

COTE, H. C. F., BRUMME, Z. L. & HARRIGAN, P. R. 2001. Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir. *J Virol,* 75**,** 589-594.

CROTEAU, G., DOYON, L., THIBEAULT, D., MCKERCHER, G., PILOTE, L. & LAMARRE, D. 1997. Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. *J Virol,* 71**,** 1089-1096.

DAM, E., QUERCIA, R., GLASS, B., DESCAMPS, D., LAUNAY, O., DUVAL, X., KRAEUSSLICH, H.-G., HANCE, A. J. & CLAVEL, F. 2009. Gag Mutations Strongly Contribute to HIV-1 Resistance to Protease Inhibitors in Highly Drug-Experienced Patients besides Compensating for Fitness Loss. *PLoS Pathogens,* 5, 1-11.

DE MENDOZA, C., GARRIDO, C., CORRAL, A., ZAHONERO, N. & SORIANO, V. 2008. Prevalence and impact of HIV-1 protease mutation L76V on lopinavir resistance. *AIDS,* 22**,** 311-3.

DE OLIVEIRA, T., ENGELBRECHT, S., VAN RENSBURG, E. J., GORDON, M., BISHOP, K., ZUR MEGEDE, J., BARNETT, S. W. & CASSOL, S. 2003. Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? *J Virol,* 77**,** 9422-9430.

DEBOUCK, C. 1992. The HIV-1 protease as a therapeutic target for AIDS. *AIDS Res Hum Retroviruses,* 8**,** 153-64.

DELAUGERRE, C., MORAND-JOUBERT, L., CHAIX, M. L., PICARD, O., MARCELIN, A. G., SCHNEIDER, V., KRIVINE, A., COMPAGNUCCI, A., KATLAMA, C., GIRARD, P. M. & CALVEZ, V. 2004. Persistence of multidrug-resistant HIV-1 without antiretroviral treatment 2 years after sexual transmission. *Antivir Ther,* 9**,** 415-21.

DEPARTMENT OF HEALTH 2010, *The South African Antiretroviral Treatment Guidelines 2010*, accessed 22 May 2010, <http://www.doh.gov.za/docs/factsheets/guidelines/art/>

DEVAL, J., WHITE, K. L., MILLER, M. D., PARKIN, N. T., COURCAMBECK, J., HALFON, P., SELMI, B., BORETTO, J. & CANARD, B. 2004. Mechanistic basis for reduced viral and enzymatic fitness of HIV-1 reverse transcriptase containing both K65R and M184V mutations. *J Biol Chem,* 279**,** 509-16.

DI MARZO VERONESE, F., COPELAND, T. D., DEVICO, A. L., RAHMAN, R., OROSZLAN, S., GALLO, R. C. & SARNGADHARAN, M. G. 1986. Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. *Science,* 231**,** 1289-91.

DOMINGO, E., ESCARMIS, C., SEVILLA, N., MOYA, A., ELENA, S. F., QUER, J., NOVELLA, I. S. & HOLLAND, J. J. 1996. Basic concepts in RNA virus evolution. *FASEB J,* 10**,** 859-64.

DOYON, L., CROTEAU, G., THIBEAULT, D., POULIN, F., PILOTE, L. & LAMARRE, D. 1996a. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol,* 70**,** 3763-3769.

DOYON, L., CROTEAU, G., THIBEAULT, D., POULIN, F., PILOTE, L. & LAMARRE, D. 1996b. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol,* 70**,** 3763-9.

DYBUL, M., FAUCI, A. S., BARTLETT, J. G., KAPLAN, J. E. & PAU, A. K. 2002. Guidelines for using antiretroviral agents among HIV-infected adults and adolescents. *Ann Intern Med,* 137**,** 381-433.

DYKES, C. & DEMETER, L. M. 2007. Clinical significance of human immunodeficiency virus type 1 replication fitness. *Clin Microbiol Rev,* 20**,** 550-78.

ERICKSON-VIITANEN, S., MANFREDI, J., VIITANEN, P., TRIBE, D. E., TRITCH, R., HUTCHISON, C. A., 3RD, LOEB, D. D. & SWANSTROM, R. 1989. Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease. *AIDS Res Hum Retroviruses,* 5**,** 577-91.

ERICKSON, J. W., GULNIK, S. V. & MARKOWITZ, M. 1999. Protease inhibitors: resistance, cross-resistance, fitness and the choice of initial and salvage therapies. *AIDS,* 13 Suppl A**,** S189-204.

FIGUEIREDO, A., MOORE, K. L., MAK, J., SLUIS-CREMER, N., DE BETHUNE, M. P. & TACHEDJIAN, G. 2006. Potent nonnucleoside reverse transcriptase inhibitors target HIV-1 Gag-Pol. *PLoS Pathog,* 2**,** e119.

FLEXNER, C. 2008. Pharmacoecology: a new name for an old science. *Clin Pharmacol Ther,* 83**,** 375-9.

FREED, E. O. 2001. HIV-1 replication. *Somat Cell Mol Genet,* 26**,** 13-33.

GATANAGA, H., SUZUKI, Y., TSANG, H., YOSHIMURA, K., KAVLICK, M. F., NAGASHIMA, K., GORELICK, R. J., MARDY, S., TANG, C., SUMMERS, M. F. & MITSUYA, H. 2002. Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J Biol Chem,* 277**,** 5952-5961.

GIANOTTI, N., GALLI, L., BOERI, E., MAILLARD, M., SERRA, G., RATTI, D., GALLOTTA, G., VACCHINI, D., TREMOLADA, Y., LAZZARIN, A., CLEMENTI, M. & CASTAGNA, A. 2005. In vivo dynamics of the K103N mutation following the withdrawal of non-nucleoside reverse transcriptase inhibitors in Human Immunodeficiency Virus-infected patients. *New Microbiol,* 28**,** 319-26.

HAHN, B. H., SHAW, G. M., DE COCK, K. M. & SHARP, P. M. 2000. AIDS as a zoonosis: scientific and public health implications. *Science,* 287**,** 607-14.

HEMELAAR, J., GOUWS, E., GHYS, P. D. & OSMANOV, S. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS,* 20**,** W13-23.

HENDERSON, L. E., SOWDER, R. C., COPELAND, T. D., OROSZLAN, S. & BENVENISTE, R. E. 1990. GAG PRECURSORS OF HIV AND SIV ARE CLEAVED INTO 6 PROTEINS FOUND IN THE MATURE VIRIONS. *J Med Primatol,* 19**,** 411-419.

HIRSCH, M. S., BRUN-VEZINET, F., CLOTET, B., CONWAY, B., KURITZKES, D. R., D'AQUILA, R. T., DEMETER, L. M., HAMMER, S. M., JOHNSON, V. A., LOVEDAY, C., MELLORS, J. W., JACOBSEN, D. M. & RICHMAN, D. D. 2003. Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type 1: 2003 recommendations of an International AIDS Society-USA Panel. *Clin Infect Dis,* 37**,** 113-28.

HIRSCH, M. S., BRUN-VEZINET, F., D'AQUILA, R. T., HAMMER, S. M., JOHNSON, V. A., KURITZKES, D. R., LOVEDAY, C., MELLORS, J. W., CLOTET, B., CONWAY, B., DEMETER, L. M., VELLA, S., JACOBSEN, D. M. & RICHMAN, D. D. 2000. Antiretroviral drug resistance testing in adult HIV-1 infection: recommendations of an International AIDS Society-USA Panel. *JAMA,* 283**,** 2417-26.

HIRSCH, M. S., GUNTHARD, H. F., SCHAPIRO, J. M., BRUN-VEZINET, F., CLOTET, B., HAMMER, S. M., JOHNSON, V. A., KURITZKES, D. R., MELLORS, J. W., PILLAY, D., YENI, P. G., JACOBSEN, D. M. & RICHMAN, D. D. 2008. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. *Clin Infect Dis,* 47**,** 266-285.

HUANG, H., CHOPRA, R., VERDINE, G. L. & HARRISON, S. C. 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science,* 282**,** 1669-75.

JACOBO-MOLINA, A., DING, J., NANNI, R. G., CLARK, A. D., JR., LU, X., TANTILLO, C., WILLIAMS, R. L., KAMER, G., FERRIS, A. L., CLARK, P. & ET AL. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc Natl Acad Sci U S A,* 90**,** 6320-4.

KANTOR, R., ZIJENAH, L. S., SHAFER, R. W., MUTETWA, S., JOHNSTON, E., LLOYD, R., VON LIEVEN, A., ISRAELSKI, D. & KATZENSTEIN, D. A. 2002. HIV-1 subtype C reverse transcriptase and protease genotypes in Zimbabwean patients failing antiretroviral therapy. *AIDS Res Hum Retroviruses,* 18**,** 1407-13.

KAPLAN, A. H., MANCHESTER, M. & SWANSTROM, R. 1994. The activity of the protease of human immunodeficiency virus type 1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency. *J Virol,* 68**,** 6782-6.

KEMPF, D. J., ISAACSON, J. D., KING, M. S., BRUN, S. C., XU, Y., REAL, K., BERNSTEIN, B. M., JAPOUR, A. J., SUN, E. & RODE, R. A. 2001. Identification of genotypic changes in human immunodeficiency virus protease that correlate with reduced susceptibility to the protease inhibitor lopinavir among viral isolates from protease inhibitor-experienced patients. *J Virol,* 75**,** 7462-9.

KNOPS, E., DAUMER, M., AWERKIEW, S., KARTASHEV, V., SCHULTER, E., KUTSEV, S., BRAKIER-GINGRAS, L., KAISER, R., PFISTER, H. & VERHEYEN, J. 2010. Evolution of protease inhibitor resistance in the gag and pol genes of HIV subtype G isolates. *J Antimicrob Chemother,* 65**,** 1472-6.

KOHLSTAEDT, L. A., WANG, J., FRIEDMAN, J. M., RICE, P. A. & STEITZ, T. A. 1992. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science,* 256**,** 1783-90.

KRAUSSLICH, H. G., INGRAHAM, R. H., SKOOG, M. T., WIMMER, E., PALLAI, P. V. & CARTER, C. A. 1989. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. *Proc Natl Acad Sci U S A,* 86**,** 807-11.

LAMBERT-NICLOT, S., FLANDRE, P., MALET, I., CANESTRI, A., SOULIE, C., TUBIANA, R., BRUNET, C., WIRDEN, M., KATLAMA, C., CALVEZ, V. & MARCELIN, A. G. 2008. Impact of gag mutations on selection of darunavir resistance mutations in HIV-1 protease. *J Antimicrob Chemother,* 62**,** 905-8.

LARDER, B. 2001. Mechanisms of HIV-1 drug resistance. *AIDS,* 15 Suppl 5**,** S27-34.

LARDER, B., KEMP, S., BLOOR, S., MILLER, V., STASZEWSKI, S., VAN DEN EYNDE, C., PAUWELS, R. & HERTOGS, K. 1998. A complete survey, in over 1,500 clinical HIV-1 isolates, of phenotypic and genotypic protease inhibitor

resistance profiles (including gag cleavage site sequences) and their relation to therapy history. *AIDS,* 12**,** OP35.

LEMEY, P., PYBUS, O. G., WANG, B., SAKSENA, N. K., SALEMI, M. & VANDAMME, A. M. 2003. Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci U S A,* 100**,** 6588-92.

LIU, T. F. & SHAFER, R. W. 2006. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin Infect Dis,* 42**,** 1608-18.

LOS ALAMOS NATIONAL LABORATORY. 2005. *HIV Sequence Database* [Online]. Los Alamos: Los Alamos National Laboratory. Available: http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html [Accessed April 2010].

LUDOVICI, D. W., DE CORTE, B. L., KUKLA, M. J., YE, H., HO, C. Y., LICHTENSTEIN, M. A., KAVASH, R. W., ANDRIES, K., DE BETHUNE, M. P., AZIJN, H., PAUWELS, R., LEWI, P. J., HEERES, J., KOYMANS, L. M., DE JONGE, M. R., VAN AKEN, K. J., DAEYAERT, F. F., DAS, K., ARNOLD, E. & JANSSEN, P. A. 2001. Evolution of anti-HIV drug candidates. Part 3: Diarylpyrimidine (DAPY) analogues. *Bioorg Med Chem Lett,* 11**,** 2235-9.

MACARTHUR, R. D. 2004. An updated guide to genotype interpretation. *AIDS Read,* 14**,** 256-8, 261-4, 266.

MAGUIRE, M. F., GUINEA, R., GRIFFIN, P., MACMANUS, S., ELSTON, R. C., WOLFRAM, J., RICHARDS, N., HANLON, M. H., PORTER, D. J. T., WRIN, T., PARKIN, N., TISDALE, M., FURFINE, E., PETROPOULOS, C., SNOWDEN, B. W. & KLEIM, J. P. 2002. Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to 150V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J Virol,* 76**,** 7398-7406.

MALET, I., ROQUEBERT, B., DALBAN, C., WIRDEN, M., AMELLAL, B., AGHER, R., SIMON, A., KATLAMA, C., COSTAGLIOLA, D., CALVEZ, V. & MARCELIN, A. G. 2007. Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. *J Infection,* 54**,** 367-374.

MAMMANO, F., PETIT, C. & CLAVEL, F. 1998. Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: Phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol,* 72**,** 7632-7637.

MAMMANO, F., TROUPLIN, V., ZENNOU, V. & CLAVEL, F. 2000. Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. *J Viro,* 74**,** 8524-31.

MARCELIN, A. G., DELAUGERRE, C., WIRDEN, M., VIEGAS, P., SIMON, A., KATLAMA, C. & CALVEZ, V. 2004. Thymidine analogue reverse transcriptase inhibitors resistance mutations profiles and association to other nucleoside reverse

transcriptase inhibitors resistance mutations observed in the context of virological failure. *J Med Virology,* 72**,** 162-5.

MARTINEZ-PICADO, J. & MARTINEZ, M. A. 2008. HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and ex vivo. *Virus Res,* 134**,** 104-23.

MARTINEZ-PICADO, J., SAVARA, A. V., SUTTON, L. & D'AQUILA, R. T. 1999. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J Virol,* 73**,** 3744-52.

MATHEWS, C. K., VAN HOLDE, K. E. & AHERN, K. G. 2000. *Biochemistry,* San Francisco, Benjamin Cummings.

MCCUTCHAN, F. E. 2000. Understanding the genetic diversity of HIV-1. *AIDS,* 14 Suppl 3**,** S31-44.

MCCUTCHAN, F. E. 2003. *HIV-1 Global Distribution* [Online]. IAVI Report. Available: http://www.pbs.org/wgbh/pages/frontline/aids/atlas/clade.html [Accessed 8 April 2010].

MENENDEZ-ARIAS, L. 2010. Molecular basis of human immunodeficiency virus drug resistance: an update. *Antiviral Res,* 85**,** 210-31.

MILLER, M. D., LAMY, P. D., FULLER, M. D., MULATO, A. S., MARGOT, N. A., CIHLAR, T. & CHERRINGTON, J. M. 1998. Human immunodeficiency virus type 1 reverse transcriptase expressing the K70E mutation exhibits a decrease in specific activity and processivity. *Mol Pharmacol,* 54**,** 291-7.

MILLER, M. D., MARGOT, N., LU, B., ZHONG, L., CHEN, S. S., CHENG, A. & WULFSOHN, M. 2004. Genotypic and phenotypic predictors of the magnitude of response to tenofovir disoproxil fumarate treatment in antiretroviral-experienced patients. *J Infect Dis,* 189**,** 837-46.

MOLLA, A., KORNEYEVA, M., GAO, Q., VASAVANONDA, S., SCHIPPER, P. J., MO, H. M., MARKOWITZ, M., CHERNYAVSKIY, T., NIU, P., LYONS, N., HSU, A., GRANNEMAN, G. R., HO, D. D., BOUCHER, C. A., LEONARD, J. M., NORBECK, D. W. & KEMPF, D. J. 1996. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat Med,* 2**,** 760-6.

MULLER, B., PATSCHINSKY, T. & KRAUSSLICH, H. G. 2002. The late-domain-containing protein p6 is the predominant phosphoprotein of human immunodeficiency virus type 1 particles. *J Virol,* 76**,** 1015-24.

MYINT, L., MATSUDA, M., MATSUDA, Z., YOKOMAKU, Y., CHIBA, T., OKANO, A., YAMADA, K. & SUGIURA, W. 2004. Gag non-cleavage site mutations contribute to full recovery of viral fitness in protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob Agents Ch,* 48**,** 444-452.

NIELSEN, M. H., PEDERSEN, F. S. & KJEMS, J. 2005. Molecular strategies to inhibit HIV-1 replication. *Retrovirology,* 2**,** 10.

NIJHUIS, M., VAN MAARSEVEEN, N. M., LASTERE, S., SCHIPPER, P., COAKLEY, E., GLASS, B., ROVENSKA, M., DE JONG, D., CHAPPEY, C., GOEDEGEBUURE, I. W., HEILEK-SNYDER, G., DULUDE, D., CAMMACK, N., BRAKIER-GINGRAS, L., KONVALINKA, J., PARKIN, N., KRAUSSLICH, H. G., BRUN-VEZINET, F. & BOUCHER, C. A. B. 2007. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *Plos Medicine,* 4**,** 152-163.

NIJHUIS, M., WENSING, A. M., BIERMAN, W. F., DE JONG, D., KAGAN, R., FUN, A., JASPERS, C. A., SCHURINK, K. A., VAN AGTMAEL, M. A. & BOUCHER, C. A. 2009. Failure of treatment with first-line lopinavir boosted with ritonavir can be explained by novel resistance pathways with protease mutation 76V. *J Infect Dis,* 200**,** 698-709.

PARRY, C. M., KOHLI, A., BOINETT, C. J., TOWERS, G. J., MCCORMICK, A. L. & PILLAY, D. 2009. Gag Determinants of Fitness and Drug Susceptibility in Protease Inhibitor-Resistant Human Immunodeficiency Virus Type 1. *J Virol,* 83**,** 9094-9101.

PATICK, A. K., DURAN, M., CAO, Y., SHUGARTS, D., KELLER, M. R., MAZABEL, E., KNOWLES, M., CHAPMAN, S., KURITZKES, D. R. & MARKOWITZ, M. 1998. Genotypic and phenotypic characterization of human immunodeficiency

virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob Agents Ch,* 42**,** 2637-44.

PATRICK, D. M., MONEY, D. M., FORBES, J., DOBSON, S. R., REKART, M. L., COOK, D. A., MIDDLETON, P. J. & BURDGE, D. R. 1998. Routine prenatal screening for HIV in a low-prevalence setting. *CMAJ,* 159**,** 942-7.

PEETERS, M. 2001. The genetic variability of HIV-1 and its implications. *Transfus Clin Biol,* 8**,** 222-5.

PERRIN, V. & MAMMANO, F. 2003. Parameters driving the selection of nelfinavir-resistant human immunodeficiency virus type 1 variants. *J Virol,* 77**,** 10172-5.

PETROPOULOS, C. J., PARKIN, N. T., LIMOLI, K. L., LIE, Y. S., WRIN, T., HUANG, W., TIAN, H., SMITH, D., WINSLOW, G. A., CAPON, D. J. & WHITCOMB, J. M. 2000. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob Agents Chemother,* 44**,** 920-8.

PETTIT, S. C., CLEMENTE, J. C., JEUNG, J. A., DUNN, B. M. & KAPLAN, A. H. 2005a. Ordered processing of the human immunodeficiency virus type 1 GagPol precursor is influenced by the context of the embedded viral protease. *J Virol,* 79**,** 10601-10607.

PETTIT, S. C., HENDERSON, G. J., SCHIFFER, C. A. & SWANSTROM, R. 2002. Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag

processing sites can inhibit or enhance the rate of cleavage by the viral protease. *J Virol,* 76**,** 10226-10233.

PETTIT, S. C., LINDQUIST, J. N., KAPLAN, A. H. & SWANSTROM, R. 2005b. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology,* 2.

PETTIT, S. C., MOODY, M. D., WEHBIE, R. S., KAPLAN, A. H., NANTERMET, P. V., KLEIN, C. A. & SWANSTROM, R. 1994. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J Virol,* 68**,** 8017-27.

POMERANTZ, R. J. & HORN, D. L. 2003. Twenty years of therapy for HIV-1 infection. *Nat Med,* 9**,** 867-73.

RICHMAN, D. D. 2001. HIV chemotherapy. *Nature,* 410**,** 995-1001.

ROBERTSON, D. L., ANDERSON, J. P., BRADAC, J. A., CARR, J. K., FOLEY, B., FUNKHOUSER, R. K., GAO, F., HAHN, B. H., KALISH, M. L., KUIKEN, C., LEARN, G. H., LEITNER, T., MCCUTCHAN, F., OSMANOV, S., PEETERS, M., PIENIAZEK, D., SALMINEN, M., SHARP, P. M., WOLINSKY, S. & KORBER, B. 2000. HIV-1 nomenclature proposal. *Science,* 288**,** 55-6.

ROUSSEAU, C. M., BIRDITT, B. A., MCKAY, A. R., STODDARD, J. N., LEE, T. C., MCLAUGHLIN, S., MOORE, S. W., SHINDO, N., LEARN, G. H., KORBER, B. T., BRANDER, C., GOULDER, P. J., KIEPIELA, P., WALKER, B. D. &

MULLINS, J. I. 2006. Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *J Virol Methods,* 136**,** 118-25.

SANTOS, A. F. & SOARES, M. A. 2010. HIV genetic diversity and drug resistance. *Viruses*, 2, 503-531.

SARAFIANOS, S. G., DAS, K., CLARK, A. D., JR., DING, J., BOYER, P. L., HUGHES, S. H. & ARNOLD, E. 1999. Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. *Proc Natl Acad Sci U S A,* 96**,** 10027-32.

SARDANA, V. V., EMINI, E. A., GOTLIB, L., GRAHAM, D. J., LINEBERGER, D. W., LONG, W. J., SCHLABACH, A. J., WOLFGANG, J. A. & CONDRA, J. H. 1992. Functional analysis of HIV-1 reverse transcriptase amino acids involved in resistance to multiple nonnucleoside inhibitors. *J Biol Chem,* 267**,** 17526-30.

SCHOCK, H. B., GARSKY, V. M. & KUO, L. C. 1996. Mutational anatomy of an HIV-1 protease variant conferring cross-resistance to protease inhibitors in clinical trials. Compensatory modulations of binding and activity. *J Biol Chem,* 271**,** 31957-63.

SHAFER, R. W. 2002. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clin Microbiol Rev,* 15**,** 247-77.

SHAFER, R. W. 2006. Rationale and uses of a public HIV drug-resistance database. *J Infect Dis,* 194 Suppl 1**,** S51-8.

SHAFER, R. W., RHEE, S. Y., PILLAY, D., MILLER, V., SANDSTROM, P., SCHAPIRO, J. M., KURITZKES, D. R. & BENNETT, D. 2007. HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS,* 21**,** 215-23.

SIERRA, S., KUPFER, B. & KAISER, R. 2005. Basics of the virology of HIV-1 and its replication. *J Clin Virol*, 34, 233-244.

SHARP, P. M., SHAW, G. M. & HAHN, B. H. 2005. Simian immunodeficiency virus infection of chimpanzees. *J Virol,* 79**,** 3891-902.

SHEHU-XHILAGA, M., KRAEUSSLICH, H. G., PETTIT, S., SWANSTROM, R., LEE, J. Y., MARSHALL, J. A., CROWE, S. M. & MAK, J. 2001. Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J Virol,* 75**,** 9156-64.

SMITH, S. W., OVERBEEK, R., WOESE, C. R., GILBERT, W. & GILLEVET, P. M. 1994. The genetic data environment an expandable GUI for multiple sequence analysis. *Comput Appl Biosci,* 10**,** 671-5.

SOARES, E. A., SANTOS, A. F., GONZALEZ, L. M., LALONDE, M. S., TEBIT, D. M., TANURI, A., ARTS, E. J. & SOARES, M. A. 2009. Mutation T74S in HIV-1 subtype B and C proteases resensitizes them to ritonavir and indinavir and confers fitness advantage. *J Antimicrob Chemother,* 64**,** 938-44.

SOARES, E. A., SANTOS, A. F., SOUSA, T. M., SPRINZ, E., MARTINEZ, A. M., SILVEIRA, J., TANURI, A. & SOARES, M. A. 2007. Differential drug resistance acquisition in HIV-1 of subtypes B and C. *PLoS One,* 2**,** e730.

SWANSTROM, R. & ERONA, J. 2000. Human immunodeficiency virus type-1 protease inhibitors: therapeutic successes and failures, suppression and resistance. *Pharmacol Ther,* 86**,** 145-70.

TAMIYA, S., MARDY, S., KAVLICK, M. F., YOSHIMURA, K. & MISTUYA, H. 2004. Amino acid insertions near gag cleavage sites restore the otherwise compromised replication of human immunodeficiency virus type 1 variants resistant to protease inhibitors. *J Virol,* 78**,** 12030-12040.

TAYLOR, B. S. & HAMMER, S. M. 2008. The challenge of HIV-1 subtype diversity. *N Engl J Med,* 359**,** 1965-6.

TESSMER, U. & KRAUSSLICH, H. G. 1998. Cleavage of human immunodeficiency virus type 1 proteinase from the N-terminally adjacent p6* protein is essential for efficient Gag polyprotein processing and viral infectivity. *J Virol,* 72**,** 3459-63.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res,* 22**,** 4673-80.

TURNER, B. G. & SUMMERS, M. F. 1999. Structural biology of HIV. *J Mol Biol,* 285**,** 1-32.

UNAIDS 2010. UNAIDS report on the global AIDS epidemic: December 2010. Geneva.

VERHEYEN, J., KNOPS, E., KUPFER, B., HAMOUDA, O., SOMOGYI, S., SCHULDENZUCKER, U., HOFFMANN, D., KAISER, R., PFISTER, H. & KUCHERER, C. 2009. Prevalence of C-terminal gag cleavage site mutations in HIV from therapy-naive patients. *J Infection,* 58**,** 61-67.

VERHEYEN, J., LITAU, E., SING, T., DAUMER, M., BALDUIN, M., OETTE, M., FATKENHEUER, G., ROCKSTROH, J. K., SCHULDENZUCKER, U., HOFFMANN, D., PFISTER, H. & KAISER, R. 2006. Compensatory mutations at the HIV cleavage sites p7/p1 and p1/p6-gag in therapy-naive and therapy-experienced patients. *Antivir Ther,* 11**,** 879-887.

WALTER, H., SCHMIDT, B., KORN, K., VANDAMME, A. M., HARRER, T. & UBERLA, K. 1999. Rapid, phenotypic HIV-1 drug sensitivity assay for protease and reverse transcriptase inhibitors. *J Clin Virol,* 13**,** 71-80.

WENSING, A. M., VAN MAARSEVEEN, N. M. & NIJHUIS, M. 2010. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res,* 85**,** 59-74.

WHITCOMB, J. M., HUANG, W., FRANSEN, S., LIMOLI, K., TOMA, J., WRIN, T., CHAPPEY, C., KISS, L. D., PAXINOS, E. E. & PETROPOULOS, C. J. 2007.

Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. *Antimicrob Agents Chemother,* 51**,** 566-75.

WHITE, K. L., MARGOT, N. A., WRIN, T., PETROPOULOS, C. J., MILLER, M. D. & NAEGER, L. K. 2002. Molecular mechanisms of resistance to human immunodeficiency virus type 1 with reverse transcriptase mutations K65R and K65R+M184V and their effects on enzyme function and viral replication capacity. *Antimicrob Agents Ch,* 46**,** 3437-46.

WOOD, E., MONTANER, J. S., YIP, B., TYNDALL, M. W., SCHECHTER, M. T., O'SHAUGHNESSY, M. V. & HOGG, R. S. 2003. Adherence and plasma HIV RNA responses to highly active antiretroviral therapy among HIV-1 infected injection drug users. *CMAJ,* 169**,** 656-61.

YOUNG, S. D., BRITCHER, S. F., TRAN, L. O., PAYNE, L. S., LUMMA, W. C., LYLE, T. A., HUFF, J. R., ANDERSON, P. S., OLSEN, D. B., CARROLL, S. S. & ET AL. 1995. L-743, 726 (DMP-266): a novel, highly potent nonnucleoside inhibitor of the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob Agents Ch,* 39**,** 2602-5.

ZHAO, X., LI, M., XU, Y., LOU, Z., MENG, Z., LI, S., TIAN, B., GAO, G. F. & RAO, Z. 2005. Cloning, expression, purification, crystallization and preliminary crystallographic study of the protein module (BIV2-Helix) in the fusion core of

bovine immunodeficiency-like virus (BIV) gp40. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 61, 249–251.

ZHANG, Y. M., IMAMICHI, H., IMAMICHI, T., LANE, H. C., FALLOON, J., VASUDEVACHARI, M. B. & SALZMAN, N. P. 1997. Drug resistance during Indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol,* 71**,** 6662-6670.

Appendix I

Table 1.1: Genbank Accession Numbers of Controls

Sub B Naïve		Sub B without PR Mutations	Sub B with PR Mutations	Sub C Naïve	M Group
A04321	DQ853460	AB097870	AB287369	DQ396364	AF004885
A07867	DQ853462	AB428558	AB287370	DQ396369	AF005494
AB078005	DQ853463	AB428559	AB287371	DQ396371	AF061641
AB221005	DQ853464	AB428560	AB287372	DQ396373	AF077336
AB253432	DQ854714	AB428561	AB289587	DQ396374	AF082394
AB286955	DQ854716	AB428562	AB289588	DQ396376	AF082395
AB286956	DQ886032	AF049494	AB289589	DQ396374	AF084936
AB287363	DQ886033	AF069140	AB289590	DQ396378	AF110967
AB287364	DQ886034	AF070521	AB428551	DQ396379	AF190127
AB287365	DQ990880	AF538304	AB428552	DQ396381	AF190128
AB287366	EF175209	AF538305	AB428553	DQ396383	AJ249235
AB287367	EF175210	AF538306	AB428554	DQ396384	AJ249236
AB287368	EF175211	AY037282	AB428555	DQ396385	AJ249237
AB485638	EF175212	AY314044	AB428556	DQ396386	AJ249239
AB485639	EF178338	AY314045	AB428557	DQ396387	K03454
AB485640	EF178354	AY314046	AB480692	DQ396390	K03455
AB485641	EF363122	AY314047	AB480693	DQ011166	M17451
AB485642	EF363123	AY314048	AB480694	DQ011171	M27323
AF003887	EF363124	AY314049	AB480695	DQ011172	M62320
AF003888	EF363125	AY314050	AB480698	DQ011173	U21135
AF033819	EF363126	AY314051	AF004394	DQ011174	U46016
AF042100	EF363127	AY314052	AF042103	AY463217	U51190
AF042101	EF514697	AY314053	AJ271445	AY463218	U52953
AF042102	EF514698	AY331294	AY037268	AY463219	U63632
AF042104	EF514700	AY331295	AY314054	AY463220	U88822
AF042105	EF514701	AY423381	AY314055	AY463231	U88826
AF042106	EF514702	AY423382	AY314056	AY585264	AF005495
AF049495	EF514704	AY423383	AY314057	AY585266	AF005496
AF086817	EF514705	AY423384	AY314058	AY585267	U88823
AF146728	EF514706	AY423385	AY314059	AY703911	U88824
AF224507	EF514707	AY423386	AY314060		
AF256204	EF514708	AY423387	AY314061		
AF256205	EF514709	AY561236	AY314062		
AF256206	EF514710	AY586542	AY314063		
AF256207	EF514711	AY586543	AY779555		
AF256208	EF514712	AY779550	AY779557		
AF256209	EF637046	AY779553	AY779558		
AF256210	EF637049	AY779560	AY779562		
AF256211	EF637050	AY779563	AY839827		
AF286365	EF637051	AY835752	DQ322224		
AF538302	EF637053	AY835761	DQ322225		
AF538303	EF637054	AY835764	DQ322226		
AF538307	EF637056	AY835768	DQ358809		
AJ006287	EF637057	AY835772	DQ676875		
AX032749	EU541617	AY835775	DQ676883		
AX078307	EU547186	AY857144	DQ676884		

Sub B Naïve		Sub B without PR Mutations	Sub B with PR Mutations
AY037270	EU616640	DQ354116	EF637048
AY173951	EU616641	DQ358805	EF694037
AY173952	EU786678	DQ358810	FJ288890
AY173953	EU839596	DQ487189	FJ388895
AY173954	EU839598	DQ823363	FJ388947
AY173955	EU839600	DQ853436	FJ388957
AY173956	EU839605	DQ853437	FJ388958
AY308960	EU839606	DQ853439	FJ388962
AY308761	EU839607	DQ853440	FJ388963
AY308762	EU839608	DQ853441	FJ388965
AY331282	EU839609	DQ853442	FJ469682
AY331283	FB341548	DQ853443	FJ469686
AY331284	FB707281	DQ853444	FJ469687
AY331285	FJ195086	DQ853446	FJ469689
AY331286	FJ195090	DQ853447	FJ469703
AY331287	FJ195091	DQ853448	FJ493714
AY331288	FJ388898	DQ853449	FJ469717
AY331289	FJ388899	DQ853453	FJ469718
AY331290	FJ388904	DQ853459	FJ469720
AY331291	FJ388910	DQ853461	FJ469721
AY331292	FJ388911	DQ854715	FJ469724
AY331293	FJ388912	DQ886031	FJ469725
AY331296	FJ388915	DQ886037	FJ469726
AY331297	FJ388916	EF178358	FJ469737
AY332236	FJ388930	EF514703	FJ469740
AY332237	FJ388931	EU786674	FJ469747
AY352275	FJ388934	EU786675	FJ469757
AY560107	FJ388935	EU786676	FJ469760
AY560108	FJ388936	EU786680	FJ469762
AY560109	FJ388937	EU839597	FJ496081
AY560110	FJ388949	EU839601	FJ496082
AY561237	FJ388949	EU839603	FJ496083
AY561238	FJ388955	EU839604	FJ496084
AY561239	FJ388956	EU839610	FJ496085
AY561240	FJ388964	FJ195088	U26546
AY682547	FJ403482	FJ195089	Z11530
AY713408	FJ460499	FJ388905	AB221125
AY713409	FJ460500	FJ388918	AB221126
AY713410	FJ460501	FJ388933	AB480696
AY713411	FJ469683	FJ388939	AB480697
AY713412	FJ469690	FJ388940	AY180905
AY751406	FJ469692	FJ388941	DQ127537
AY751407	FJ469694	FJ388960	DQ127538
AY779551	FJ469696	FJ469684	DQ127539
AY779552	FJ469697	FJ469685	DQ354119
AY779554	FJ469700	FJ469688	DQ383750
AY779559	FJ469705	FJ469691	DQ676870
AY779561	FJ469706	FJ469693	DQ676871
AY779564	FJ469707	FJ469695	DQ886035

Sub B Naïve		Sub B without PR Mutations	Sub B with PR Mutations
AY781127	FJ469716	FJ469699	EF057102
AY795904	FJ469722	FJ469702	EF514699
AY795905	FJ469723	FJ469708	EF637047
AY818644	FJ469727	FJ469709	EU616642
AY819715	FJ469728	FJ469710	EU616643
AY835748	FJ469731	FJ469711	EU616644
AY835749	FJ469732	FJ469712	EU616645
AY837750	FJ469733	FJ469713	EU616649
AY837751	FJ469734	FJ469719	EU786672
AY835753	FJ469735	FJ469729	EU786679
AY835754	FJ469736	FJ469739	EU839602
AY835755	FJ469738	FJ469742	FJ388891
AY835756	FJ469741	FJ469744	FJ388914
AY835757	FJ469743	FJ469745	FJ388919
AY835758	FJ469749	FJ469746	FJ388920
AY835759	FJ469750	FJ469748	FJ388923
AY835760	FJ469753	FJ469755	FJ388924
AY835762	FJ469754	FJ469756	FJ388927
AY835763	FJ469758	FJ469766	FJ388959
AY835765	FJ469759	FJ469767	FJ469701
AY835766	FJ469761	FJ469769	FJ469704
AY835767	FJ469763	FJ469771	FJ469730
AY835769	FJ469764	FJ495818	FJ469751
AY835770	FJ469765	FJ495819	FJ469752
AY835771	FJ469770	FJ495820	FJ469768
AY835773	FJ469772	FJ495821	
AY835774	FJ496072	FJ495822	
AY835776	FJ496073	FJ495823	
AY835777	FJ496074	FJ495824	
AY835778	FJ496075	FJ495825	
AY835779	FJ496076	FJ495826	
AY835780	FJ496077	FJ495937	
AY835781	FJ496078	FJ495939	
AY857022	FJ496079	FJ495940	
AY857165	FJ496080	FJ495941	
AY945710	FJ496145	FJ495942	
AY945711	FJ496146	FJ495943	
AY970946	FJ496147	FJ496000	
AY970947	FJ496148	FJ496001	
AY970948	FJ496149	FJ496002	
AY970949	FJ496150	FJ496003	
AY970950	FJ496151	FJ496004	
BD238372	FJ496152	FJ496005	
BD298114	FJ496153	FJ496006	
CS272319	FJ496154	FJ496007	
D10112	FJ496155	U34604	
D86068	FJ496156	U63632	
D86069	FJ496157	U69584	
DD153311	FJ496158	U69588	
DD153312	FJ496159	U71182	

Sub B Naïve	
DM111087	FJ496161
DQ007901	FJ496162
DQ007902	FJ496163
DQ007903	FJ496164
DQ127534	FJ496165
DQ127535	FJ496166
DQ127536	FJ496167
DQ127548	FJ496168
DQ127549	FJ496169
DQ127550	FJ496170
DQ127551	FJ496171
DQ207942	FJ496172
DQ295192	FJ496174
DQ295193	FJ496175
DQ295194	FJ496176
DQ295195	FJ496177
DQ322223	FJ496178
DQ322227	FJ496179
DQ322228	FJ496180
DQ322229	FJ496181
DQ322230	FJ496182
DQ322239	FJ496184
DQ354112	FJ647145
DQ354114	FJ694790
DQ354118	GU177863
DQ358808	K02007
DQ383746	K02013
DQ383748	K02083
DQ383749	K03455
DQ383751	L02317
DQ383752	L31963
DQ396398	M15654
DQ487188	M17449
DQ487190	M17451
DQ487191	M19921
DQ676874	M26727
DQ676876	M38429
DQ676877	M38431
DQ676878	M93258
DQ676879	M93259
DQ676880	NC001802
DQ676881	U12055
DQ676882	U21135
DQ676886	U23487
DQ676887	U26942
DQ823362	U34603
DQ823364	U39362
DQ837381	U43096
DQ853438	U43141
DQ853445	U69585

Sub B Naïve	
DQ853451	U69587
DQ853452	U69589
DQ853454	U69590
DQ853455	U69591
DQ853456	U69592
DQ853457	U69593
DQ853458	X01762