The HIV-1 Gag and Protease: Exploring the coevolving nature and structural implications of complex drug resistance mutational patterns in subtype C

By

Veronna Marie

Submitted in fulfilment of the academic requirements for the degree of Doctor of Philosophy in Virology

Nelson R Mandel School of Medicine, School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal

2019

Preface

The work described in this thesis was carried out at the HIV Pathogenesis Programme and KwaZulu-Natal Research Innovation and Sequencing Platform, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa, from July 2015 to March 2019, under the supervision of Dr. Michelle Gordon.

This work has not been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others, it is duly acknowledged in the text.

Signed

…….. ……..

Veronna Marie Date

Signed

…….. ……..

Dr. Michelle Gordon Date

Declaration

I, **Miss Veronna Marie** declare that:

1. The research reported in this dissertation, except where otherwise indicated, is my original research.

2. This dissertation has not been submitted for any degree or examination at any other university.

3. This dissertation does not contain other scientists' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other scientists.

4. This dissertation does not contain other scientists' writing, unless specifically acknowledged as being sourced from other scientists. Where other written sources have been quoted, then their words have been re-written but the general information attributed to them has been referenced.

5. This dissertation does not contain text, graphics or tables taken from the internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the reference section.

Signed

……..
Veronna Marie

Presentations

Part of the findings observed in this study was presented at the following conferences or symposia:

1. Oral presentation at the **College of Health Sciences (CHS) Research Symposium**, 8th–9th September 2016, Nelson R Mandela School of Medicine (UKZN), Durban, South Africa.

2. Oral presentation at the **7th International Bioinformatics Congress**, 24th–25th October 2016, Rome, Italy.

3. Poster presentation at the **WGC Advanced Course: Computational Molecular Evolution**, 8–19th May 2017, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom.

To my parents, this title is yours.

I also dedicate this thesis to all the nerds out there that continue to search for answers to questions we have not yet asked.

And to all those that conspire against me,
I will destroy you.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Michelle Gordon. Through this sometimes-daunting process, you have inspired me to always remain positive and to complete any task that I undertake with the greatest of efforts I can afford.

A warm thanks to the Protease Cleavage Site study participants for their noble contribution to the workings of this study.

I would like to acknowledge the National Research Foundation, Poliomyelitis Research Foundation, South African Medical Research Council, HIV Pathogenesis Programme and University of KwaZulu-Natal's College of Health Sciences for the financial support.

I thank my friends Khe Maphumulo and Avashna Singh for their assistance and for kindly listening to me complain during the entirety of this project.

To my parents, Alex and Jane Marie, I would have not been able to complete this degree without your guidance and profound understanding. Everything I do, I do for you.

To my sister, Savanna and brother in-law, Melvin, thank you for listening to me when I needed to vent and helped me when I needed the support.

To my beautiful niece and handsome nephew, you are the reason for my random smiles and bursts of happiness. You calm my mind when nothing else can.

God, for his guidance and allowing me to bring to fruition my most treasured thoughts.

Abstract

Due to the high prevalence of HIV-1 subtype C infection coupled with increasing antiretroviral (ARV) drug treatment failure, the elucidation of complex resistance mutational patterns arsing through protein coevolution is required. Despite the inclusion of LPV and DRV in second- and third-line, many patients still fail treatment. In this study, protease (PR) inhibitor resistance mutations were identified by comparing treatment versus naïve sequences datasets in Gag and PR. Thereafter, to investigate Gag-PR coevolution and pathways to LPV resistance, phylogenetic analyses and Bayesian networks were constructed. Following this, structural analyses combining homology modelling, molecular docking and molecular dynamic simulations were carried out on specific patterns of protease resistance mutations (PRMs). To complement these analyses, the structural impact of a mutated Gag cleavage site on PR resistance dynamics was also evaluated. Accordingly, this study identified 12 major PRMs and several resistance combinations. Of these, the M46I+I54V+V82A pattern frequently occurred. The second most frequently recurring pattern included L76V as a fourth mutation to the above triplet. Coevolution analyses revealed correlations between positions 10, 46, 54 and 82 in PR. Of these, minor PRM L10F occurred in 6.4% of the dataset and was involved in pathways to LPV resistance. Additionally, Gag cleavage site (CS) mutation A431V was also correlated with L10F and the major PRMs. Distinct changes in PR's active site, flap and elbow regions due to the PRMs (L10F, M46I, I54V, L76V, V82A) were found to alter LPV and DRV drug binding. When the PRMs were combined with the mutant Gag CS binding was greatly exacerbated. While the A431V Gag CS mutation coordinated several amino acid residues in PR, the L76V mutation was found to have a significant role in substrate recognition rather than directly inhibiting the drugs. These data show that the co-selection of mutations in Gag-PR greatly contributes to resistance outcomes and that our understanding on drug resistance is largely lacking, particularly where structure is concerned.

Table of Content

List of Figures

List of Tables

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AA	Amino acid
AIC	Akaike information criterion
AIDS	Acquired immune deficiency syndrome
AMBER	Assisted model building with energy refinement
APV	Amprenavir
ARV	Antiretroviral
BI	Bayesian inference
BNL	Bayesian network learning
CA	Capsid; p24
CAPS	Coevolution analysis for protein sequences
CDC	Centre for Disease Control and Prevention
CRF	Circulating recombinant form
CS	Cleavage site
CTD	C-terminal domain
DRV	Darunavir
DRV/r	Darunavir boosted with ritonavir
FDA	Food and Drug Administration
FI	Fusion inhibitor
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HIV-1	Human immunodeficiency virus type I
HIV-2	Human immunodeficiency virus type II
HTLV-III	Human T-lymphotropic virus type III
INSTI	Integrase stand transfer inhibitors
KZN	KwaZulu-Natal
LAV	Lymphadenopathy-associated virus
LPV	Lopinavir
LPV/r	Lopinavir boosted with ritonavir
LRT	Likelihood ratio test
LTR	Long terminal repeat

MA	Matrix; p17
MD	Molecular dynamics
MDR	Multi-drug resistant
MHR	Major homology region
ML	Maximum-likelihood
MM-GBSA	Molecular mechanics-generalized born surface area
MM-PBSA	Molecular mechanics-poisson boltzmann surface area
NC	Nucleocapsid; p7
NFV	Nelfinavir
NNRTI	Non-nucleoside reverse transcriptase inhibitors
NPT	Constant pressure and normal temperature
NRTI	Nucleoside reverse transcriptase inhibitors
ns	Nanoseconds
NTD	N-terminal domain
NVT	Constant volume and normal temperature
p1	Spacer peptide two
p2	Spacer peptide one
PAI	Post-attachment inhibitor
PCS	Protease cleavage site
PDB	Protein Databank
PI	Protease inhibitor
PIC	Pre-integration complex
PK	Pharmacokinetic enhancers
ps	Picoseconds
PR	Protease
PRMs	Protease resistance mutations
RTV	Ritonavir
SCF	Self consistent field
SIV	Simian immunodeficiency virus
United States	US
vdW	van der Waals
WT	Wild-type

Symbols

α	Alpha
β	Beta
Å	Angstroms
Cα	Alpha carbon
K	Kelvin
kcal	Kilocalories
mol	per mole
π	pi

Chapter one

Literature review

1.1 History of the human immunodeficiency virus

In June 1981, a group of clinical experts based at the Centre for Disease Control and Prevention (CDC) documented that five relatively healthy patients between the ages of 29 and 36 all contracted a rare respiratory infection called *Pneumocystis carinii* pneumonia[1] (PCP) in their Morbidity and Mortality Weekly Report. Additionally, whilst all five men confessed to having sexual relations with other men and used inhalant drugs, three patients had defective lymphocyte proliferation coupled with considerably low T cell counts (CDC, 1981). Following the CDC's initial report, several cases all leading into the year of 1982 described larger groups of homosexual men showing the same symptoms, with mortality steadily increasing. By September 1982, the CDC had concluded that given the clinical manifestations and nature of the emerging infectious disease, this new illness would be known as acquired immune deficiency syndrome (AIDS). Later AIDS would be identified in various groups, including patients receiving blood transfusions, injection drug users, Haitian individuals residing in the United States (US) and in children of parents with AIDS (Goedert and Gallo, 1985). This implied that this unidentified etiological agent was transmitted via blood-borne, vertical and sexual means and was not limited to homosexuals.

In 1983, Françoise Barré-Sinoussi and Luc Montagnier isolated a retrovirus from the lymph node of a 33-year-old French homosexual patient. Since the virus was only able to infect T cells, they concluded that the patient was infected with a T cell-tropic retrovirus and was subsequently named lymphadenopathy-associated virus (LAV) (Barré-Sinoussi et al., 1983). Although the link between AIDS and LAV remained uncertain at this time, Luc Montagnier and Barré-Sinoussi would go on to win the Nobel Prize for initially isolating and characterizing HIV in 2008. Elsewhere, Robert Gallo and colleagues had isolated a virus from 48 patients in 1984. These patients included those with AIDS, a healthy homosexual male, mothers with AIDS and patients with symptoms of pre-AIDS[2]. After multiple tests, Gallo's team coined the virus human T-lymphotropic virus type III (HTLV-III) and suggested its role as the causative agent of AIDS (Gallo et al., 1984).

It was eventually revealed that LAV and HTLV-III were in fact the same virus and was subsequently renamed the human immunodeficiency virus (HIV) in 1986 (Coffin et al., 1986).

[1] PCP is a rare, serious infection of the lungs that are usually seen in severely immunocompromised individuals.

[2] Pre-AIDS is referred to as symptoms that precede the CDC's official definition of AIDS, e.g. lymphadenopathy.

1.2 Classification and origin of HIV

HIV belongs to the Lentivirus genus within the *Retroviridae* family (Williams and Burdo, 2009) and is classified into two main types, namely, HIV type I (HIV-1) and HIV type II (HIV-2). Although both types share structural similarities, HIV-1 is responsible for the global pandemic seen today (Xiao et al., 2017). However, while HIV-2 was initially localized to West Africa, it has since traversed to the United States, Europe, India and other parts of Africa (Campbell-Yesufu and Gandhi, 2011).

HIV-1 is divided into groups M (main), N (non-M; non-O), O (outlier) and P. Group M is further subdivided into nine subtypes (A–D, F–H, J and K) (Lau and Wong, 2013) and about 98 circulating recombinant forms[3] (CRFs). Of these, group M subtype C is currently the most prevalent form of HIV-1, globally (Castley et al., 2017). HIV-2 is divided into eight groups, namely A–H, where groups A and B are the most frequent and likely the only pathogenic ones (Eberle and Gürtler, 2012). Both HIV-1 and HIV-2 originated from independent zoonotic events when a simian immunodeficiency virus (SIV) jumped species from non-human primates to humans (Peeters et al., 2010). Two distinct SIVs were involved in cross-species transmission for the main groups of HIV-1 to arise. As such, the natural viruses found in chimpanzees (SIVcpz) led to the genetic development of HIV-1 groups M and N whilst those found in gorillas (SIVgor) led to the HIV-1 group O variants. Similarly, the natural sooty mangabey[4] virus (SIVsmm) is the closest genetic relative of HIV-2 (Pandrea et al., 2008).

Interestingly, Kalish et al. (2004) showed that people hunting primates for bushmeat in Central Africa were often exposed to SIV infection. Nevertheless, the weak nature of SIV allows for the human immune system to fully suppress the virus within weeks of acquisition. Therefore, several successive infections from person-to-person and the presence of high-risk transmission networks were required for SIV to mutate into HIV. Consequently, the turn of the 20th century which led to the collapse of colonialism, increased trade with global proprietors, sociopolitical reform and an increase in prostitutes provided the perfect channel for HIV to develop and spread. A study conducted by Faria et al. (2014) on the transmission network of HIV-1 concluded that group M utilized the behavioural and socioeconomic changes to establish and maintain future transmissions (Parrish et al., 2008) in what would become one of the "most devastating infectious diseases" to emerge in human history (Sharp and Hahn, 2011).

[3] CRFs are combinations of different HIV-1 subtypes, e.g. CRF08_BC is a combination of subtypes B and C.

[4] The sooty mangabey (*Cercocebus atys*) are old world monkeys indigenous to West Africa.

1.3 Geographical distribution of the HIV-1 subtypes

According to a report published in 2018 by the Joint United Nations Programme on HIV/AIDS, there were 36.9 million people living with HIV, globally. In addition, there were 1.8 million new infections and 940,000 AIDS-related deaths (UNAIDS, 2018). An estimated 20% of the population of Lesotho, South Africa, Zimbabwe and Botswana are HIV-positive (Chin, 2017). In South Africa alone, there is an estimated 7.52 million individuals currently living with HIV (STATS SA, 2018). One of the most challenging issues faced currently is preventing the transmission of HIV-1 (Kurth et al., 2011). Finding ways to control the epidemiological spread of HIV-1 can not only prevent further increases in the number of new HIV infections but can also reduce the economic burden faced by many developing countries.

Thus far, universal efforts in generalizing the pandemic is hindered by the worldwide geographical heterogeneity of the HIV-1 subtypes (Buonaguro et al., 2007) as seen in Figure 1.1. Due to the diversity of HIV-1 in Southern Africa, several HIV-1 subtypes and CRFs are found in Cameroon and in the Democratic Republic of Congo. Accordingly, subtype C is endemic to Eastern and sub-Saharan Africa, the south of Brazil and the Indian Pacific whereas subtype B is mostly prevalent in the US, Japan, Europe and Northern Africa (Chin, 2017). CRF01_AE (recombinant between HIV-1 subtypes A and E) and CRF02_AG (recombinant between subtypes A and G) accounts for 5% and 8% of global HIV infections and are prevalent in Southeast Asia and West Africa (Daw et al., 2017). Interestingly, the idea that a single virus preferentially introduced into a dense population which rapidly transmitted, or otherwise known as the founder-effect was suggested to play a role in the geographical distribution of the HIV-1 subtypes seen today (Buonaguro et al., 2007; Chin, 2017).

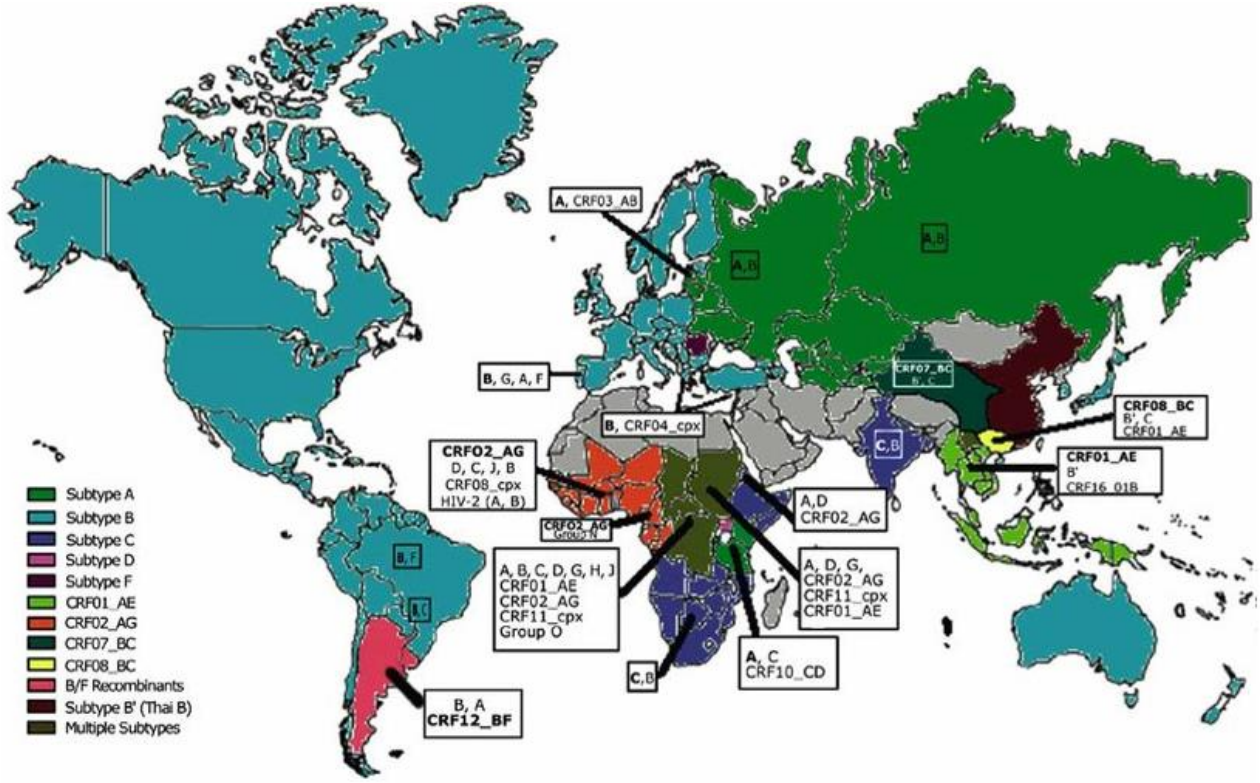


Figure 1.1 Worldwide geographical distribution of the HIV-1 subtypes and CRFs. Illustration taken from Butler et al. (2007).

1.4 Clinical manifestations of HIV infection

Over the years several mathematical models have been used to breakdown the clinical progression from HIV infection to AIDS. Though differences exist between patients, a general trend was outlined as seen in Figure 1.2. Accordingly, this trend encompassed three distinct stages known as the acute infection, chronic infection and AIDS phase (Alizon and Magnus, 2012). During acute infection, flu-like symptoms including headache, cough and fever occurs within 2–4 weeks (Robb et al., 2016). Additionally, weight loss, oral and genital ulcers, skin rash as well as fatigue have also been associated with the early stage of HIV infection. Importantly, individuals in acute stage have high viral turnover rates and can easily transmit to non-infected people quickly. Consequently, 38–50% of all new infections occur at the acute HIV infection stage (Henn et al., 2017).

With the onset of flu-like symptoms during acute HIV infection, the body's immune response kicks in. The cellular immunity or cell-mediated immune response[5] which is mediated by lymphocytes known as T-cells is activated after 1–2 weeks of infection. This is followed by the humoral response[5] (or antibody-mediated humoral response which sequesters specific B-cells to form plasma cells to secrete antibodies into the lymph and blood) 4–8 weeks post-infection (Alizon and Magnus, 2012). Through the activation of immune responses, viral copy numbers decrease with lower levels of replication keeping the viral load constant. This is referred to as the viral load set point and is the chronic infection stage (Dykes and Demeter, 2007). Since host CD4 T-cells are actively targeted by the virus (Douek et al., 2002), its numbers are greatly reduced, which leads to the AIDS stage (Dykes and Demeter, 2007). This stage is the most severe and is characterized by extremely low CD4 T-cell count (usually <200 per μL), high viral copies and damaged immune systems. Furthermore, the frailty of the immune system results in the acquisition of several opportunistic infections (Friedman-Kien et al., 1981).

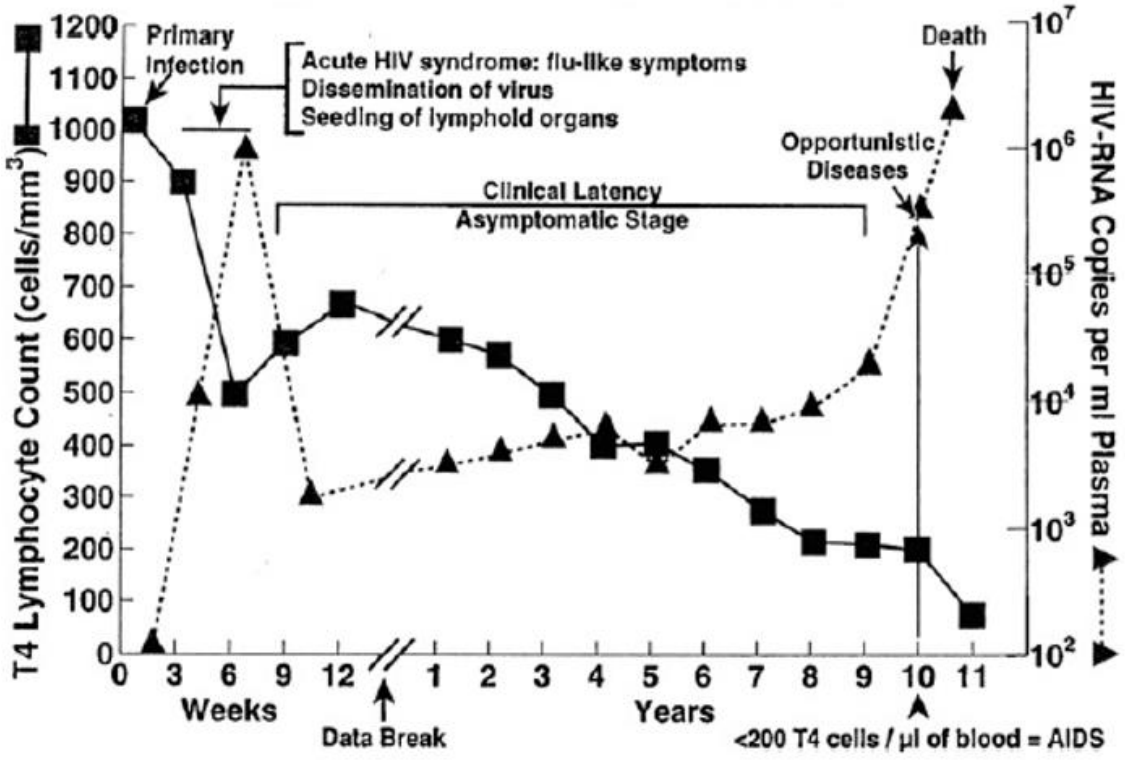


Figure 1.2 The typical three clinical stages observed in untreated HIV-1 infected individuals with axes showing the viral load and CD4 T-cell count. Illustration taken from Landi et al. (2008).

[5] Information on immune responses adapted from Nauta J. 2011. Humoral and cellular immunity. In Nauta J. (eds), Statistics in Clinical Vaccine Trials. Springer-Verlag Berlin Heidelberg, pp. 13–17.

1.5 Genetic and physical characteristics of HIV

In general, all retroviruses possess similar physical and genetic characteristics. One of the main features that make retroviruses unique is that the virus comprises two positive-sense RNA strands (Kieken et al., 2002). Each strand comprises nine genes over three reading frames as shown in Figure 1.3 below. Of these, the *gag*, *pol* and *env* genes encodes the virus's main structural and enzymatic components. The last six include regulatory genes *tat* and *rev* as well as the *vif* (viral infectivity factor), *vpu* (viral protein U), *vpr* (viral protein R) and *nef* (negative regulation factor) accessory proteins. The *gag* gene encodes the virus's main structural components, matrix, capsid and nucleocapsid while *pol* encodes for the viral protease, reverse transcriptase and integrase enzymes. The surface envelope glycoprotein is encoded with the *env* gene (Watts et al., 2009). Importantly, these genes are flanked by identical 5' and 3' long terminal repeats (LTRs). Transcription regions within the LTRs can activate and repress factors that influence the rates of gene expression (Burnett et al., 2009; Shah et al., 2014).

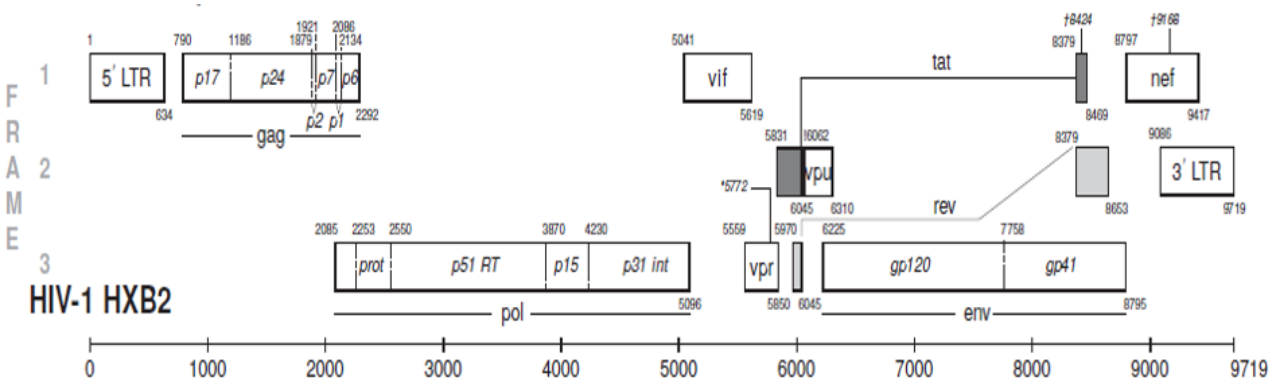


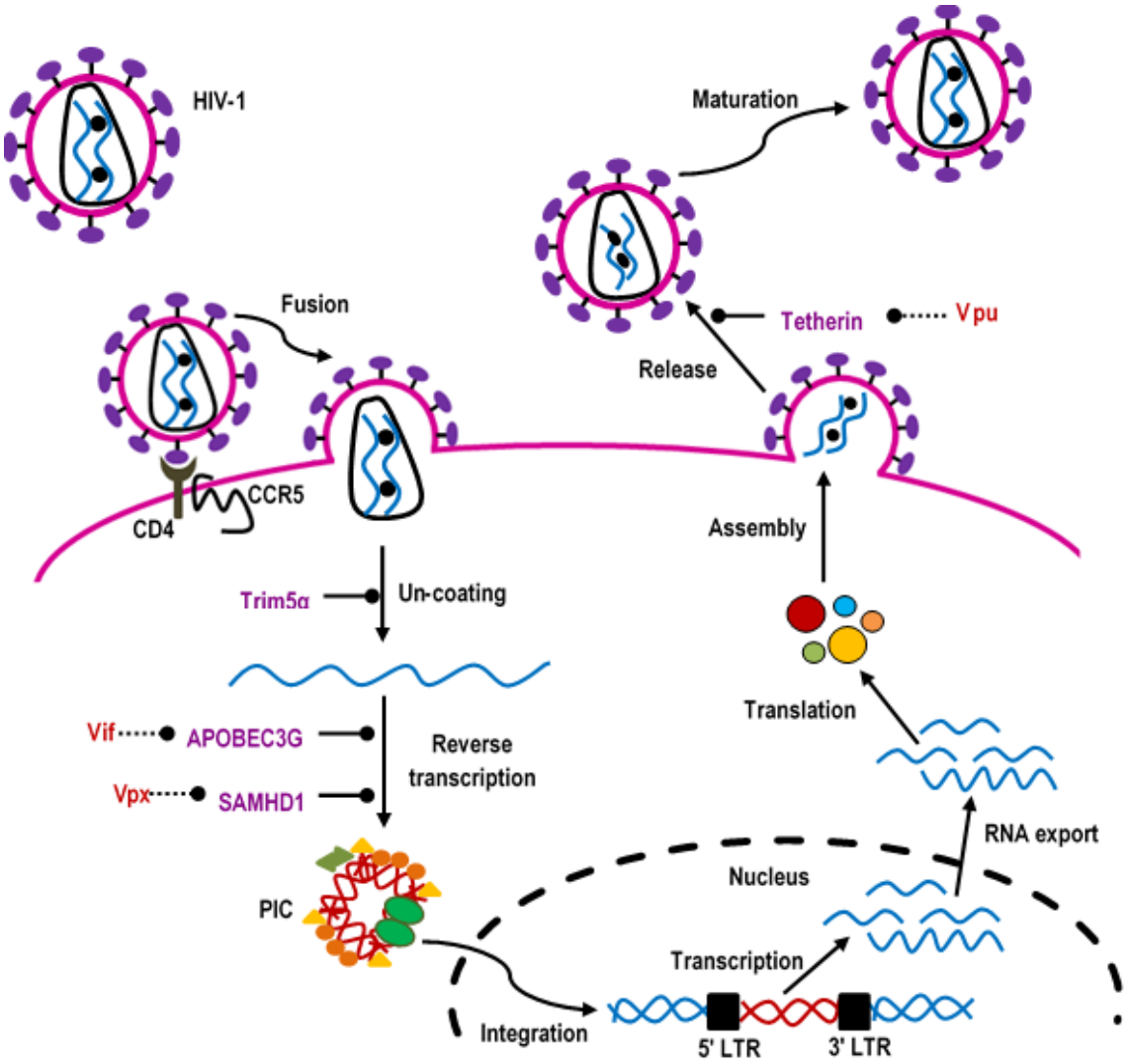
Figure 1.3 Genetic map of HIV-1 showing the three reading frames, the virus's nine genes and the 5' to 3' long terminal repeat (LTR) regions. Image taken from Foley et al. (2017).

Physically, the HIV-1 virion[6] is approximately 100 nm in size and spherical in shape. The viral envelope contains 72 spikes which are composed of trimers of envelope's (Env) gp120 and gp41 surface and transmembrane glycoproteins, respectively (Gelderblom et al., 1987). The HIV-1 matrix (MA; p17) protein which is a thin layer lining the inner surface of the virion lipid bilayer helps hold the Env proteins together after maturation has occurred (Gelderblom et al., 1987). Furthermore, in a mature virion, the HIV capsid (CA, p24) protein oligomerizes to form a shell around the virus's RNA and core-associated proteins (Freed, 1998).

[6] A structurally and genetically complete infectious virus particle as it exists outside a host.

1.6 The HIV-1 lifecycle

The replication cycle of HIV-1 comprises two distinct phases: (a) the early phase where virions bind to the host cells followed by cDNA integration and (b) the late phase in which viral gene expression occurs to allow for maturation and release of the now infectious viral particles (Nisole and Saïb, 2004). Several steps occur (Figure 1.4) to produce mature, infectious virus: (i) binding of HIV to its receptor (CD4) and preferred co-receptor[7] (ii) fusion with the host cell membrane, (iii) un-coating of the viral capsid and subsequent nucleic acid release, (iv) reverse transcription of viral RNA, (v) integration into host cell's DNA, (vi) transcription and translation, (vii) assembly of the immature virions and budding from cells and (viii) final maturation of immature viral particles to fully infectious virions (reviewed in Barré-Sinoussi et al., 2013).



[7] Co-receptors are a group of chemokine proteins. In the HIV-1 lifecycle, R5 and X4 viruses interact with CCR5 or CXCR4 co-receptors, respectively.

Figure 1.4 Diagram of the HIV-1 lifecycle. Replication begins when HIV-1 binds the CD4 receptor and its co-receptors following which fusion of the virion to the host membrane occurs. After HIV's capsid is uncoated the viral proteins and RNA are released into the cytoplasm. Thereafter, the virus's RNA is reverse-transcribed into DNA and the pre-integration complex (PIC) is formed and translocated into the nucleus. At this stage, the viral DNA is integrated with the host's genetic material and encoded into new viral proteins and RNA. These components are then transported to the cell surface where they are assembled into immature virus particles. These progeny viruses are subsequently released. Finally, the protease cleaves the virus's structural polyprotein, Gag to form fully, infectious HIV virions. **Note:** restriction factors (tetherin, APOBEC3G, tripartite motif-containing 5α (TRIM5α) and SAMHD1; purple) and the virus's regulatory proteins (Vpu, Vif and Vpx; red) are also shown. **Abbreviations:** CCR5 = CC-chemokine receptor 5 (Illustration adapted from Barré-Sinoussi et al., 2013).

1.7 HIV-1 antiretroviral therapy

The elucidation of the HIV-1 lifecycle provided invaluable information for the development of several drugs designed to actively target specific steps in the virus's replicative process (Gu et al., 2014). Zidovudine, a reverse-transcriptase inhibitor was the first antiretroviral (ARV) drug approved for HIV treatment by the Food and Drug Administration (FDA) (Warnke et al., 2007). Since then, it has become apparent that a combination of ARV drugs is more effective against the virus that a single inhibitor. This approach is often referred to as highly active antiretroviral therapy or HAART (Cohen, 2006; Arts and Hazuda, 2012). Since its inception, HAART is generally able to reduce the viral load to undetectable numbers (Fang et al., 2007) whilst extending patient life expectancy by several years (Smith et al., 2013; Meintjes et al., 2017).

There are over 26 FDA approved ARV drugs that belong to eight drug classes (NIH, 2018). These are the nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), CCR5 antagonist, fusion inhibitor (FI), post-attachment inhibitor (PAI), integrase strand transfer inhibitors (INSTIs), pharmacokinetic enhancers (PKs) and the protease inhibitors (PIs) as seen in Table 1.1 below. Briefly, the NRTIs are considered chain terminators as they block reverse transcriptase from producing DNA via its incorporation into the process. Contrastingly, NNRTIs binds to reverse transcriptase thereby changing the flexibility of the enzyme (Tang and Shafer, 2012). The CCR5 antagonist blocks the CCR5 co-receptor thus preventing virus entry into the cell (Lopalco, 2010) whilst the FIs prevent HIV-1 from entering the CD4 cells (Falkenhagen and Joshi, 2018). Contrastingly, the PAIs are monoclonal antibodies that bind the receptive CD4 cells to block HIV-1 entry (Henrich and Kuritzkes, 2013). The INSTIs are a class of drugs that inhibits the HIV-1 integrase enzyme thus preventing it from making more copies of itself. The PKs are used in HIV-1 therapy to pharmacokinetically enhance or boost the

effectiveness of ARVs (Arts and Hazuda, 2012). Lastly, the PIs are a class of competitive inhibitors that target the HIV-1 protease enzyme during maturation (Yang et al., 2012) and will be further discussed in section 1.9.3 as they are the focus of the current study.

Table 1.1 List of commonly used ARV drugs approved by the FDA.

DRUG CLASS	GENERIC NAME	TRADE NAME
Nucleoside reverse transcriptase inhibitors	Abacavir	ZIAGEN
	Tenofovir alafenamide	VEMLIDY
	Emtricitabine	EMTRIVA
	Tenofovir disoproxil fumarate	VIREAD
	Lamivudine	EPIVIR
	Zidovudine	RETROVIR
Non-Nucleoside Reverse Transcriptase Inhibitors	Doravirine	PIFELTRO
	Efavirenz	SUSTIVA
	Etravirine	INTELENCE
	Nevirapine	VIRAMUNE VIRAMUNE XR
	Rilpivirine	EDURANT
Pharmacokinetic Enhancers	Cobicistat	TYBOST
	Ritonavir*	NORVIR
Post-Attachment Inhibitor	Ibalizumab-uiyk	TROGARZO
Fusion Inhibitor	Enfuvirtide	FUZEON
Protease Inhibitors	Atazanavir	REYATAZ
	Lopinavir/Ritonavir	KALETRA
	Nelfinavir	VIRACEPT
	Darunavir	PREZISTA
	Saquinavir	INVIRASE
	Tipranavir	APTIVUS
	Fosamprenavir	LEXIVA
	Amprenavir	AGENERASE
	Indinavir	CRIXIVAN
Integrase Strand Transfer Inhibitors	Bictegravir	BIKTARVY#
	Dolutegravir	TIVICAY
	Elvitegravir	STRIBILD#
	Raltegravir	ISENTRESS
CCR5 Antagonists	Maraviroc	SELZENTRY

* Ritonavir is a protease inhibitor, however, it serves as a pharmacokinetic enhancer or booster.
# Available as part of a fixed dose combination pill only.

In South Africa, the classes of drugs currently available include the NRTIs, NNRTIs, PIs and INSTIs (NDOH, 2019). Drugs from these classes are administered in three varying regimens depending on the patient's age, weight, secondary infections and even gender. In the first-line treatment regimen, one NNRTI and two NRTIs are administered as a fixed dose combination

tablet. However, recent guidelines have also stipulated the use of an INSTI drug, namely Dolutegravir, in first-line treatment regimens (NDOH, 2019). Currently, Dolutegravir is administered as a fixed dose combination with Tenofovir and Lamivudine and is prescribed to clients > 10 years of age and > 35 kg, and as a standard dose in adolescents, adults and children > 20 kg. In second-line, the regimen consists of two NRTIs and a PI whilst third line regimens consists of a PI, NNRTI and an INSTI (NDOH, 2019). An overview of the available ARVs used in the three treatment regimens are described in Table 1.2.

Table 1.2 Overview of the ARV drugs employed in the three drug regimens in South Africa (adapted from the NDOH, 2019 guidelines).

First-line regimen: Two NRTIs + one NNRTI	Emitricitabine (FTC) Tenofovir (TDF) Lamivudine (3TC) Abacavir (ABC)	NRTI
	Efavirenz (EFV) Nevirapine (NVP)	NNRTI
Second-line regimen: Two NRTIs + one PI	Tenofovir (TDF) Lamivudine (3TC) Zidovudine (AZT) Emitricitabine (FTC) Abacavir (ABC)	NRTI
	Lopinavir (LPV)* Atazanvir (ATV)	PI
Third-line regimen: One NNRTI + one PI + one INSTI	Etravirine (ETR)	NNRTI
	Darunavir (DRV)*	PI
	Raltegravir (RAL)	INSTI

* LPV and DRV are boosted with ritonavir (r): LPV/r and DRV/r

1.8 ARV drug resistance

Due to the low proof-reading capability of the HIV-1 reverse transcriptase enzyme, the virus can accumulate numerous mutations in the genome (Hu and Hughes, 2012). Although most of these mutations adversely affect enzymatic function, the trade-off between reduced fitness and the virus's ability to evade ARVs allow some of these mutations to gain an evolutionary advantage (Richman et al., 2004). Hence, these mutant strains continue to replicate and will eventually predominate thereby rendering treatment ineffective as seen in Figure 1.5 below. Although the success of ARV therapy cannot be disputed, the development of ARV drug resistance remains a serious concern (Dudley et al., 2014; Silver et al., 2018). Even though these drug resistance mutations exhibit different trends for different drugs, certain mutations are often enough to resist

multiple inhibitors belonging to the same family or class. This is known as cross-resistance and has been observed in several studies (Condra et al., 1995; Hertogs et al., 2000; Race, 2001; Rhee et al., 2010; Diphoko et al., 2018; Penrose et al., 2018).

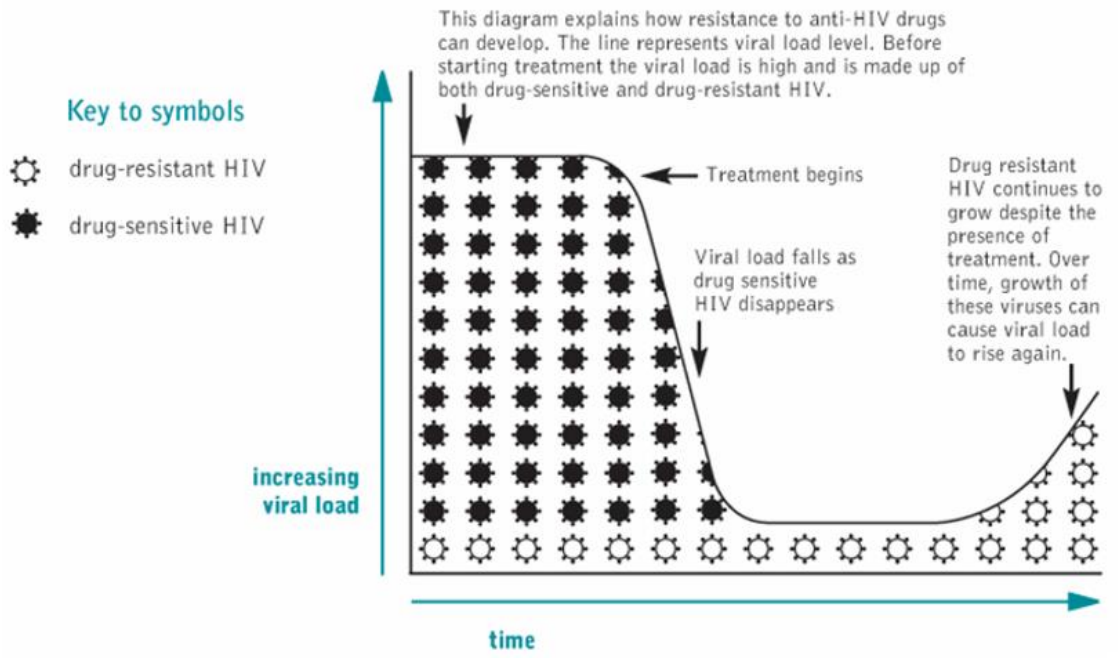


Figure 1.5 Diagram showing the development of drug resistant viruses over time (Image taken from AIDSMAP, 2014).

Usually, to ensure the drug regimens are effective, patients are carefully monitored (Iacob et al., 2017). Patient monitoring includes a sequence of steps from initial diagnosis to treatment and are assessed by stages. These stages are diagnosis, linkage to medical care, initiating ART, adherence to treatment regimens and viral suppression to undetectable levels in the blood (HIV. gov, 2019) through the monitoring of CD4 cell counts and viral load data (WHO, 2006). Although genotyping the viral sequences can provide invaluable information on which resistance mutations are present, the relationship between resistance mutations are complicated (Zhang et al., 2010). Currently, clinical decisions on the selection of drug regimens are based on predictive algorithms such as the HIVdb algorithm on the Stanford HIV Drug Resistance Database (Sing and Däumer, 2006; Shen et al., 2016). Since the focus area of this study is to evaluate the mechanism by which mutations in protease (PR) and its natural substrate Gag work to evade the PIs, the enzyme, substrate and its associated resistance will be discussed in detail in the following sections.

1.9 Protease structure, function, inhibitory drugs and resistance

1.9.1 Structure

The HIV-1 PR belongs to a family of aspartic proteases that is symmetrically structured to form a homodimer consisting of two identical subunits each comprising 99 amino acids (AAs) (Katoh et al., 1987; Yang et al., 2012). Each PR subunit is composed of a single α-helix and nine β-sheets. Of these, four β-sheets that run anti-parallel relative to each other form the conserved dimer interface (Todd et al., 1998; Velazquez-Campoy et al., 2003) as shown in Figure 1.6. Structural studies indicate that residues 1–4 and 96–99 in each subunit of the four anti-parallel β-sheets forms the dimer interface of PR (Choudhury et al., 2003). The dimer interface results in the formation of the hydrophobic active site cavity consisting of PR's catalytic triad D25-T26-G27 from both monomers (Wlodawer et al., 1989; Zhang et al., 2008). More so, the anti-parallel β-sheets restrict access to the substrate cavity by forming two flexible flaps to cover the active site region. In a ligand-free state, PR assumes a semi-open conformation whereas a PR-ligand complex induces a closed conformation. Two models have been proposed for flap opening and closing states (Yang et al., 2012). The first theorizes that a collision complex between the ligand and PR occurs when the flaps are fully opened which then induces a closed conformation as the substrate approaches the active site (Scott and Schiffer, 2000). The second model proposes that the ligand first approaches a semi-open state which then prompts the flaps to fully open as it enters the substrate cavity. The flaps then extend over the substrate to adopt a closed conformation for proteolysis (Collins et al., 1995; Toth and Borics, 2006). While the flaps are indispensable for catalysis of the viral PR, their exact conformational states during cleavage is debatable (Soares et al., 2016).

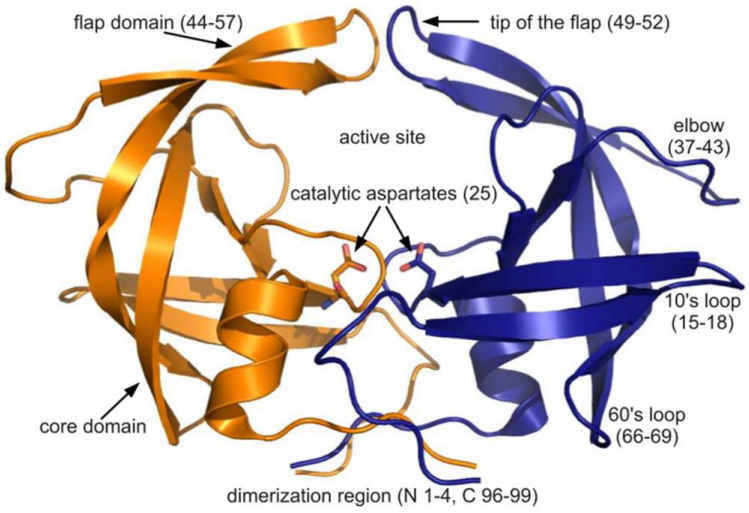


Figure 1.6 Structural representation of the HIV PR showing its active site, flap and identical subunits. Image take from Venkatakrishnan et al. (2012).

1.9.2 Function

In nature, the role of the HIV-1 PR is to cleave the nascent polyproteins and is thus essential for viral replication (Adachi et al., 2009). Consistent with studies evaluating other non-viral PR's, it is recognized that the acid-base mechanism involving two important active site aspartic residues are essential for catalysis to occur (Brik and Wong, 2003). The widely accepted mechanism of catalysis was initially described by Suguna et al. (1987). The authors suggest that of the two aspartic acids, only one is unprotonated. Therefore, the negatively charged aspartic group activates the nucleophilic water located between the aspartic acids which then proceeds to attack the substrate's carbonyl group to generate an oxyanion tetrahedral intermediate. Finally, the substrate's amide nitrogen atom is protonated and rearranged to breakdown the intermediate leading to the formation of the hydrolysis products (Figure 1.7).

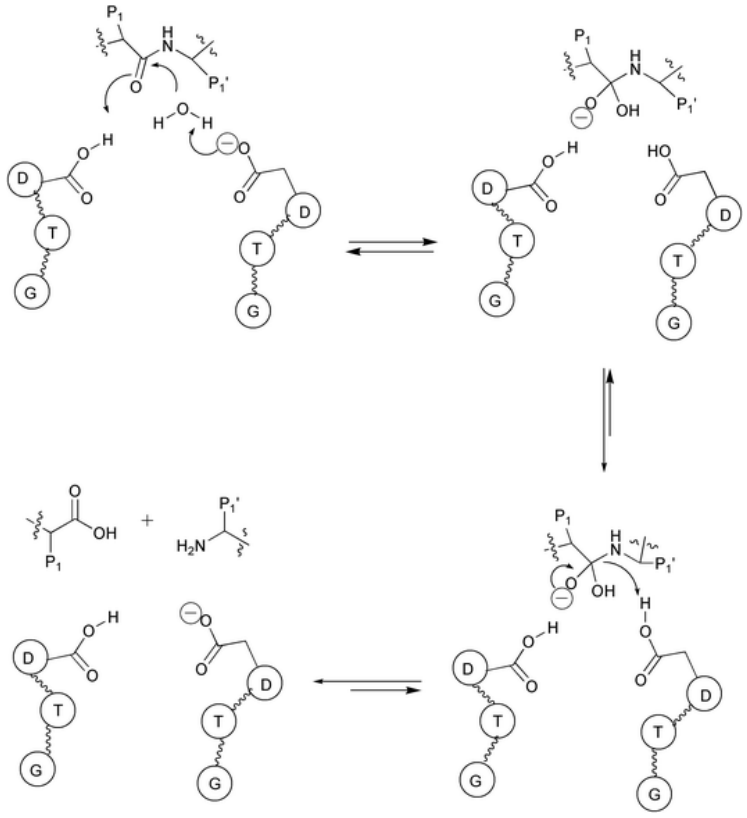


Figure 1.7 Chemical mechanism showing the proteolytic cleavage of the hydrolysis products. Image taken from Brik and Wong (2003).

In Gag, five cleavage sites (CSs) are recognized by PR during hydrolysis as depicted in Table 1.3 below. Although these regions share limited sequence similarity they are similarly structured. As such, it was proposed that the mechanism by which PR cleaves Gag also depends on the structure of the substrate. Moreover, while PR exhibits symmetry within its subunits, the Gag substrates are asymmetrical in size and charge of the AA residues (Prabu-Jeyabalan et al., 2002).

Table 1.3 WT sequences for the HIV-1 subtype C Gag CSs (adapted from De Oliveira et al., 2003).

CLEAVAGE SITES	SEQUENCE
Matrix-Capsid	VSQNY -- PIVQN
Capsid-p2	KARVL -- AEAMS
p2-Nucleocapsid	NTNIM -- MQKSN
Nucleocapsid-p1	ERQAN -- FLGKI
p1-p6	RPGNF -- LQSRP

1.9.3 Inhibitory drugs

As previously mentioned, the analogous structure of the PIs to the Gag CSs make them ideal inhibitors of HIV-1 maturation (Mudgal et al., 2018). As shown in Table 1.1 above, there are 10 FDA approved PIs currently available on market, nine of which are shown in Figure 1.8 below. In general, the mechanism of inhibition begins when the hydroxyl group of the PI interacts with the carboxyl group of the aspartic active site residues via hydrogen bonding. Consequently, binding to the active site AAs prevents the HIV-1 PR from successfully cleaving Gag. While several attempts have been made to modify the current PIs, none have been as successful as second-generation PIs Lopinavir (LPV) and Darunavir (DRV) which were modified from Ritonavir (RTV) and Amprenavir (APV), respectively (Lv et al., 2015).

In LPV, the 5-thiazolyl and the 2-isopropylthiazolyl of RTV's P2 and P2' groups are replaced with a phenoxyacetyl and six-membered cyclic urea, respectively. These substitutions have been shown to improve the potency against drug resistant variants (Sham et al., 1998). On the other hand, having been FDA-approved in 2006, DRV is the newest PI on market (Lv et al., 2015). The only difference between DRV and APV is the substitution of tetrahydrofuran with bis-tetrahydrofuran in the P2 group. Consequently, replacing APV's chemical moiety allows DRV to form more hydrogen bonds with PR. Due to these modifications LPV boosted ritonavir (LPV/r) and DRV boosted ritonavir (DRV/r) have high genetic barriers against the resistant PR variants in comparison to the other PIs (Doherty et al., 2011; Aoki et al., 2018).

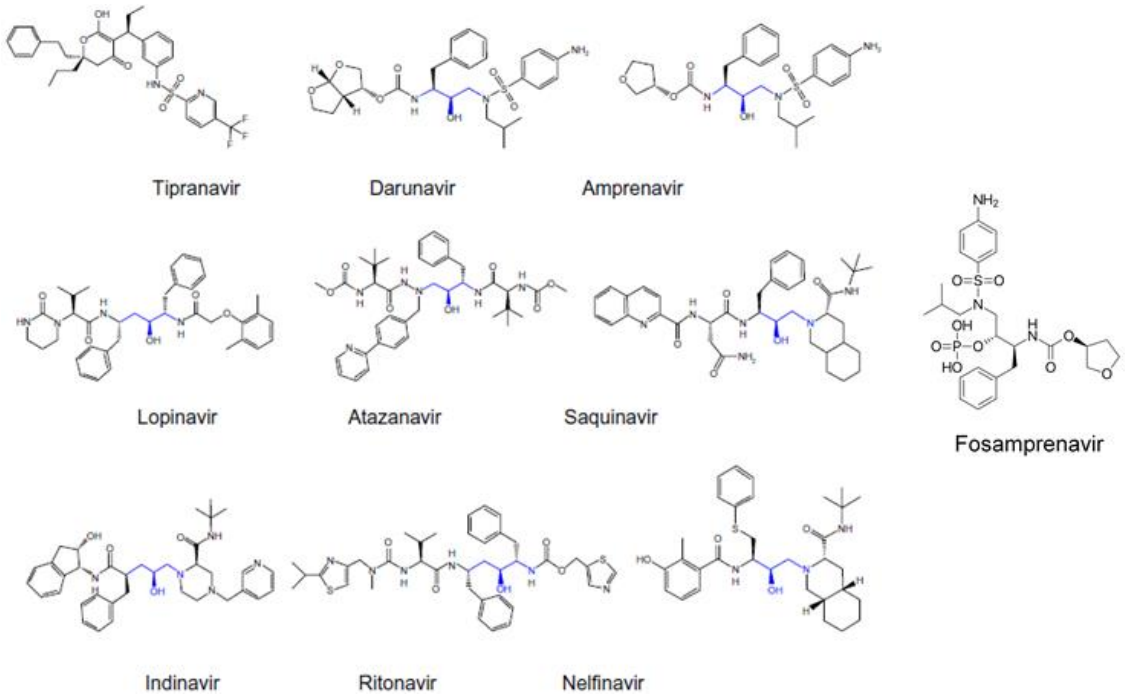


Figure 1.8 Chemical structures of the PIs. Image adapted from Lv et al. (2015).

1.9.4 Drug resistance

Despite their success, mutations arising in the viral PR have been associated with LPV and DRV drug resistance (Weber et al., 2015). In LPV, the most common resistance mutations include V32I, M46I/L, G48V/M, I54V/T/A/L/M, L76V, V82A/T/F/S, I84V and L90M (Lv et al., 2015) whilst V32I, L33F, I47V/A, I50V, I54L/M, L76V and I84V have been clinically associated with DRV resistance (Tremblay, 2008). Generally, resistance to the PIs occur when the viral PR accumulates several primary or major mutations in various regions of the enzyme (Wensing et al., 2010). Major mutations directly affect resistance by altering the structure of the catalytic site. This results in reduced contact between PR and the inhibitors. In addition, several secondary or minor resistance mutations emerge at a later stage to compensate for the changes induced by the major protease resistance mutations (PRMs) (Budambula et al., 2015). Therefore, this suggests that the way in which PR resists the PIs is dependent on complex evolutionary and resistance dynamics.

1.10 Gag structure, function and its role in PI resistance

1.10.1 Structure and function

The autocatalytic cleavage of the aspartyl protease enzyme from Gag-Pol (PR160) prompts the cleavage of both Gag-Pol and Gag (PR55) at predefined sites in the polyproteins (Weber and Agniswamy, 2009). Gag or group specific antigen is essential for viral infectivity and maintaining the structural integrity of HIV (Ning et al. 2016). The protein comprises four main domains, i.e. matrix, capsid, nucleocapsid and p6 (Bharat et al., 2014). In addition to these domains, two small spacer peptides (p2 and p1) that assist in the regulation of structural changes accompanying maturation are also found within Gag (Sundquist and Kräusslich, 2012). For successful maturation of the virions, the proteolytic cleavage of Gag occurs in a very precise and ordered manner as seen in Figure 1.9 below.

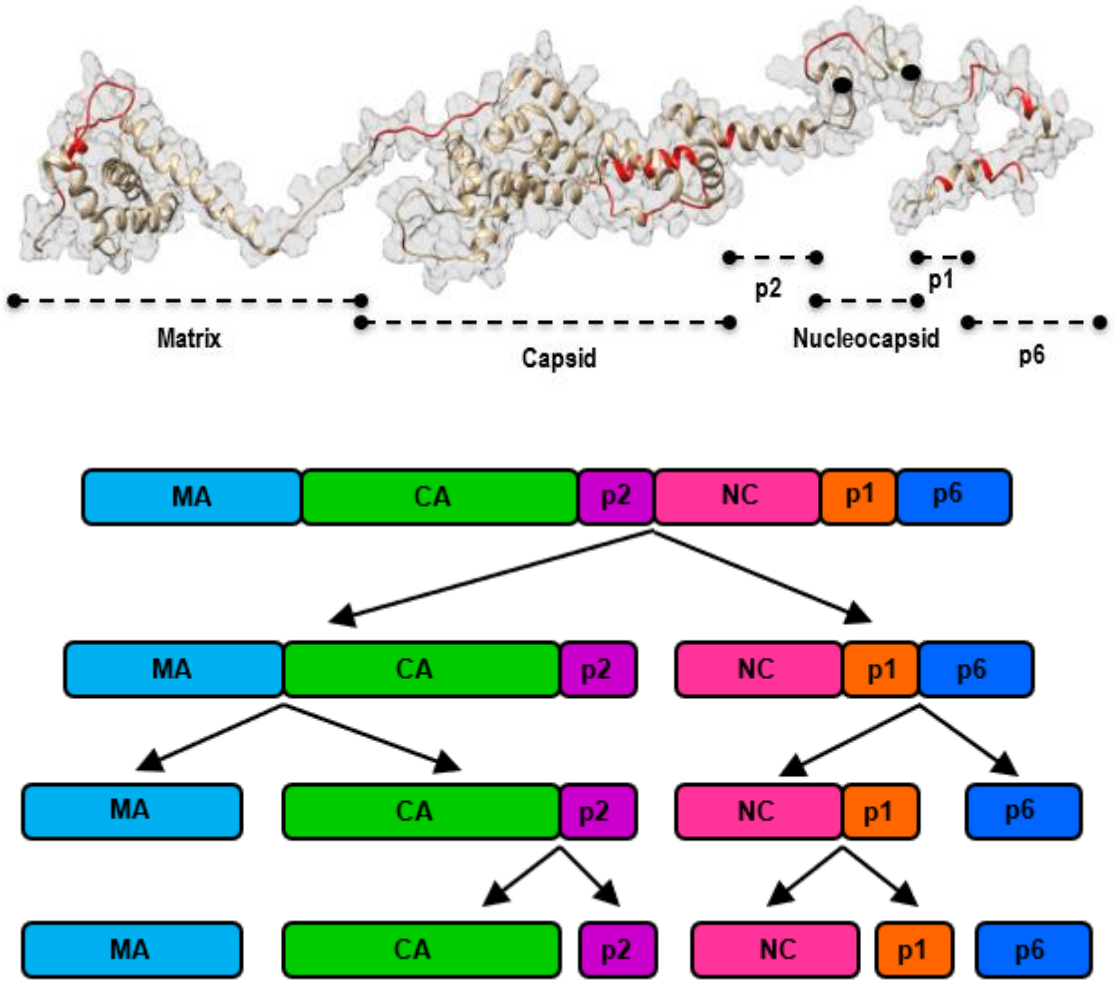


Figure 1.9 Diagram showing the proteolytic cleavage of Gag by PR. The first site of cleavage in between p2 and nucleocapsid. This is followed by cleavage at the matrix-capsid and p1|p6 sites. Finally, the small spacer peptides are cleaved from capsid and nucleocapsid. **Abbreviations:** MA

= matrix, CA = capsid and NC = nucleocapsid. **PDB codes:** 2H3I (Matrix), 1L6N (N-terminal Capsid), 1A43 (C-terminal Capsid), 1U57 (p2), 1A1T (Nucleocapsid) and 2C55 (p6).

1.10.1.1 Matrix

The HIV-1 matrix (MA; p17) protein is a thin layer lining the inner surface of the virion lipid bilayer after maturation (Doherty et al., 2005). Structurally, the 132 amino acid protein (Hill et al., 1996) consists of an amino-terminal (N-terminal) globular head and a flexible carboxyl-terminal (C-terminal) tail. Globular structures are typically spherical or "globe-shaped" water soluble proteins. In matrix, the head is composed of a 3_{10} helix (Freed, 2015) and four alpha helices that is capped by a 3-stranded beta sheet. A 5^{th} α-helix connects the capsid and MA domains (Freed, 2015), projecting away from the β-sheet to expose the C-terminal residues and forms the tail (Fiorentini et al., 2006). Based on the ionic strength and concentration of the protein buffer, MA may exist as either monomers, dimers, timers or oligomers of monomers, dimers and trimers (Forster et al., 2000).

Functionally, MA facilitates the targeting of Gag-Pol and Gag to the plasma membrane via myristoylation and PIP2 signals[8]. Additionally, MA is also involved in the nuclear import of the pre-integration complex (PIC) (Freed, 1998). The PIC is a large nucleoprotein composed of several viral and cellular proteins that allows integration during replication (Thierry et al., 2015). Interestingly, MA also has cytokine-like functions that act on pre-activated human T cells encouraging proliferation, viral replication after p17R host cellular receptor binding and pro-inflammatory cytokine release. Therefore, MA may have an important role in the virus- and host-derived factors that contribute to favourable HIV-1 infection and replication (Fiorentini et al., 2006).

1.10.1.2 Capsid

In a mature virion, the HIV capsid (CA, p24) protein oligomerizes to form a shell around the virus's RNA and core-associated proteins (Freed, 1998). All morphological variances of the CA domain are derived from a collection of hexameric (rings) CA monomers (Maillard et al., 2011) that has the same tertiary structure (Ganser-Pornillos et al., 2007). CA comprises two separate domains: the N-terminal domain (NTD 1–145) and the C-terminal domain (CTD 146–231) (Gitti et al., 1996), connected by a flexible linker (Lingappa et al., 2014). Structural studies revealed

[8] Plasma membrane-specific phosphatidylinositol-4,5-bisphosphate (PIP2) and myristoylation facilitates Gag membrane targeting to the plasma membrane.

that CA is predominantly composed of α-helices, with seven and four α-helices in the NTD and CTD, respectively (Bharat et al., 2012). Moreover, the NTD of CA contains an amino-terminal β-hairpin and resembles an arrowhead whereas the globular CTD comprises one 3_{10} helix (Berthet-Colominas et al., 1999). Extensive intermolecular NTD-NTD interactions and to a lesser extent NTD-CTD interactions stabilize the hexameric rings. This stabilization occurs predominantly between helices 8–11 on the C-terminal side and 4–7 on the N-terminal side. Flexible CTD-CTD interactions link the adjacent hexameric rings, allowing for the formation of the curved lattice (Fassati, 2012) while inter-hexamer angle variation between the longitudinal or lateral positions or at the broad and narrow ends of CA has been suggested. Furthermore, the asymmetrical structure infers that no two CAs are identical (Zhao et al., 2013). Therefore, the flexibility and complex molecular interactions supports a variety of conformations that contribute to structural variation in CA (Gres et al., 2015).

Additionally, each CA domain has a different function in HIV-1 morphogenesis. Although the NTD of CA is not necessary for immature virion assembly, it is indispensable for mature core formation (Borsetti et al., 1998). Contastingly, the CTD of CA is important for both the formation of the core and in virion assembly (McDermott et al., 1996). Additionally, all retroviruses comprise a conserved region of 20 AAs known as the major homology region (MHR) which is located in the CTD of CA (Wills and Craven, 1991). While this region is not clearly understood (Bell and Lever, 2013), its presence has been suggested to have a role at post-assembly and during assembly stages (Cairns and Craven, 2001). The entire MHR region forms a compacted strand turned helix stabilized by hydrogen bonds (Gamble et al., 1997). The MHR has been suggested to have a role CTD dimerization contributing to the stability of the viral shell (Ivanov et al., 2005). As such, mutations in this region may affect viral assembly.

1.10.1.3 Nucleocapsid

The 55 amino acid nucleocapsid (NC; p7) protein (Bell and Lever, 2013) packages two copies of the viral RNA genome into the assembling virions (Berkowitz et al., 1996). A prominent feature of NC is the presence of two highly conserved Cys-X2-Cys-X4-His-X4-Cys (CCHC) signatures that resemble zinc-finger motifs. The zinc-fingers, each containing an aromatic residue (F16 in the N-terminal zinc finger and W37 in the C-terminal zinc finger) (Darlix et al., 2011) coordinate a zinc ion (Lingappa et al., 2014) and are divided by a RAPRKKG basic domain linker (Godet et al., 2012). These separately folded zinc-fingers resemble beads attached to a string (Summers et al., 1992).

Functionally, in addition to genome packaging, NC can renature nucleic acids with catalytic rates of about four orders of magnitude. In reverse transcription, NC can also stimulate $tRNA^{Lys}$ binding to the primer binding site found at the N-terminal region of the genome, initiate reverse transcription from the bound $tRNA^{Lys}$ and partake in strand transfer (Rodriguez-Rodriguez et al., 1995; Guo et al., 1997; Freed, 1998). Furthermore, annealing by NC enhances the formation of the converted loop-loop interaction into a stable duplex and the formation of RNA dimers (Lu et al., 2011).

1.10.1.4 p6

The C-terminal region of Gag comprises a proline-rich (Freed, 1998), 52 amino acid p6 domain (Votteler et al., 2011) where two amino acid sequences are translated, the -1 frameshift Gag-Pol p6 and the in-frame Gag p6 domain (Bell and Lever, 2013). Although the structure of the p6 domain of Gag remains unclear (Freed, 2015), it is known that the domain displays little, if any secondary structure (Stys et al., 1993), instead serving as a flexible docking site for cellular host factors (Gottlinger et al., 1991). The p6 domain is also necessary for the incorporation of viral accessory protein, Vpr into the virions (Solbak et al., 2013). Vpr is associated with trans-activation of the LTR, import of the pre-integration complex (PIC) in non-dividing cells, the induction of apoptosis[9] and cell halt at the G2/M transition pathway[10].

1.10.1.5 p2 and p1 spacer peptides

The 14 amino acid p2 peptide is found wedged between the N-terminal MA and C-terminal NC proteins (Bell and Lever, 2013). The C- and N-termini of the CA and p2 domains respectively, were shown to form two parts of an α-helix. Electron cryotomography analyses show that p2 forms a six α-helical bundle that stabilizes Gag hexamers in immature virus particles (Wright et al., 2007). As a result, cleavage of p2 from NC is necessary for the formation of ribonucleoprotein within RNA and condensation of the CA domain. The late cleavage of CA|p2 allows morphogenesis through the cadence of CA-CA interactions. The importance of the integrity of the CA|p2 segment in formation of immature CA was shown in an electron microscopic study (Gross et al., 2000).

The 16 amino acid p1 peptide lies between the CTD of NC and the NTD of p6. This peptide contains two highly conserved proline residues, namely P445 and P439. Hill et al. (2007)

[9] Programmed cell death

[10] The G2/M transition is a point in the cell's lifecycle where after the second growth phase (G2) and DNA replication (S phase), mitosis occurs (M phase) to separate the cell into identical daughter cells.

implicated the importance of p1 for Pol and Gag incorporation into the virus particles. The "slippery site" involved in the Gag-Pol ribosomal frameshift overlaps with the end of p1 leading to potentially complex effects of mutations at this site (Bell and Lever, 2013). The substitution of either proline (P445 and P439) by leucine results in lower stability of the NC-RNA complex and rescinds infectivity (Hill et al., 2002). Moreover, mutations at the cleavage sites at either end of p1 have discrepant effects. For example, mutations in NC|p1 have no effect on proviral integration, whereas mutations at the C-terminal that produces p15 (NC|p1|p6) or p8 (p1|p6) peptides decreases the amount of integrated provirus on subsequent infection. This differential effect may suggest the role of NC|p1 in proviral integration as opposed to p15 and p8 (Coren et al., 2007).

1.10.2 Resistance

The collection of resolved structures on individual Gag domains contributes to the intrinsic and functional behaviour of Gag at the molecular level (Novikova et al., 2018). As a result, it is now evident that Gag and its individual domains have significant roles in the lifecycle and perseverance of HIV (Freed, 1998; Mailler et al., 2016; Tomasini et al., 2018). Even so, an extensive review by Fun et al. (2012) highlights several Gag mutations within and outside the CSs that are associated with drug resistance or exposure. Mutations in Gag, particularly at NC|p1 and p1|p6 CSs have been extensively reported (Borman et al., 1996; Nijhuis et al., 1999; Maquire et al., 2002; Prado et al., 2002; Dam et al., 2009). Additionally, while mutations in Gag can restore substrate processing (Myint et al., 2004; Tamiya et al., 2004; Ho et al., 2008; Parry et al., 2009), it has also been shown that Gag can confer resistance in the absence of mutations in the viral PR (Nijhuis et al., 2007; Gupta et al., 2010).

1.11 Molecular mechanisms of drug resistance

The molecular mechanisms of drug resistance are commonly elucidated through structural comparisons of the wild type and mutant proteins. In the HIV-1 PR, several structural changes associated with single amino acid substitutions have been identified: i) direct active site mutations that alter the interactions of PR with the substrate or inhibitor, ii) mutations occurring at the dimer interface affecting PR stability and iii) distal mutations showing an assortment of structural affects, such as mutations occurring at the flaps and those without direct contacts with inhibitors (Weber and Agniswamy, 2009). Whilst several mutations at the Gag CSs can be co-selected along with resistance mutations in the viral PR, the molecular mechanism by which these mutations contribute to resistance is unclear. However, it has been shown that Gag CS mutations can enhance the van der Waals interactions between the mutant PR and substrate (Özen et al., 2014).

As Gag is a large protein in nature, the mechanisms behind the structural synergistic effects of Gag and PR mutations requires further investigation (Su et al., 2019).

1.12 Variability in HIV-1 subtype C

Genomic variability differs across the subtypes. These discrepancies are based not only on the subtype but also the genomic region. Therefore, specific, functionally important regions in HIV are conserved across the subtypes, such as the CD4 binding sites (Alexandre et al., 2011). HIV-1 subtype C exhibits high sequence variability (Lynch et al., 2009) which has been attributed to several distinct genetic features including three NF-κB binding sites at the long terminal repeats (Rodenburg et al., 2001), as well as a truncated Rev (Ndung'u et al., 2001, Gordon et al., 2003) and a five AA insertion in the Vpu protein. While the truncated Rev protein and the additional NF-κB binding sites may alter gene expression by influencing viral replication (Iordanskiy et al., 2010), the enlarged Vpu protein may impact subtype C virulence via the regulation of Vpu functions including enhancing virion release and/or CD4 degradation (McCormick-Davis et al., 2000). Interestingly, some mutations associated with drug resistance in other subtypes have been identified as natural polymorphisms in subtype C (Lynch et al., 2009). Examples of these signature mutations include PR's M36I in non-B subtypes and I93L in subtype C (Santos and Soares, 2010). Contrastingly, some studies have reported the contribution of polymorphic residues in drug resistance (Bessong, 2008; Wainberg and Brenner, 2012). Furthermore, a higher level of diversity exists in subtype C Gag CSs in comparison to other group M subtypes (Li et al., 2013). This diversity may arise through ARV drug selection pressure, mutations or as natural polymorphisms (Verheyen et al., 2009).

1.13 Proteins

Proteins are the macromolecular machines of life as they perform complex molecular processes with rapid efficiency and accuracy. These processes are usually dependent on the protein's ability to bind other small molecules (Chica, 2018). Specifically, proteins form the abundance of cellular dry mass and is imperative for regulating biological function and molecular structure. Some of these functions include cellular messaging, maintenance of cell shape, catalytic reactions and the transport of small molecules. Although, these macromolecules exhibit great diversity in shape and size, all proteins are comprised from 20 common AAs (Janke et al., 2019). Together with these AAs, proteins display primary, secondary, tertiary structures that will be discussed in the subsequent sections of this chapter.

1.13.1 Primary structure

Simply, the primary structure of a protein is a linear chain of AAs (Farhadi, 2018). Large polypeptide chains are formed when the carboxyl and amino groups of two different AAs form a peptide bond thus linking them together (Janke et al., 2019). Importantly, while the amino and carboxyl groups are conserved across all the AAs, the individual side chains are different. These side chains are what distinguishes AAs from each other (Farhadi, 2018).

In nature, the amino acids that form specific proteins is dependent on an RNA template which is transcribed from a DNA sequence. The template codes for the AAs as triplet of bases, also referred to as a codon, from adenine, cytosine, guanine or uracil. As such, AAs can be identified based on either a three letter or single code (Wright, 2011). Since proteins form complex structures in nature, which is known as folding, the sequence of AAs plays a role in the conformation of the final structure. Furthermore, the AA side chains can also adopt numerous conformations referred to as rotamers. Specific conformational rotamers adopt the lowest energy. This energy is determined by the conformation of the backbone, the side chain the rotamer is interacting with and the environment (Wright, 2011).

1.13.2 Secondary structure

Secondary structures are formed due to repeated regular conformations of a polypeptide chain (Anfinsen et al., 1961; Zhang et al., 2018). The arrangement of peptide bonds results in the formation of specific structural elements in various regions of the proteins. The most common being α-helices and β-sheets (Eisenberg, 2003; DeBenedictis and Keten, 2019). Briefly, α-helices are formed when the phi and psi angles exhibit constant displacement between the nitrogen and α-carbon of N-Cα, Cα and carbonyl carbon of Cα-C and C-N. Furthermore, there are 3.7 AA residues per turn (Haimov and Srebnik, 2016). Contrastingly, β-sheets are formed from consecutive strands separated by turns. Varied hydrogen bond networks can form depending on the orientation of the strands. In essence, β-sheets can assume parallel, anti-parallel and mixed orientations (Craveur et al., 2013).

1.13.3 Tertiary structure

Tertiary structures are formed when specific regions known as loops and the secondary structural elements adopt a satisfactory three-dimensional (3D) state during protein folding. Since loops rarely have hydrogen bonds between residues, they exhibit high flexibility and are therefore less ordered (Subramani and Floudas, 2012). Moreover, folding is dependent on the hydrophobic packing of residues, dispersion of hydrophobic and hydrophilic residues in water and the non-

covalent interactions between AA side chains. Additionally, regions of a protein that can independently fold into a stable 3D structure are referred to as domains (Wright, 2011), for example the individual Gag domains discussed in section 1.10.

1.13.4 Quaternary structure

Simply put, the quaternary structure of proteins refers to the 3D structure of several protein subunits such as dimers, trimers and tetramers (Godbey, 2014). These structures are bound by non-covalent bonds, salt linkages and disulfide bonds. Since these bonds are considered weak, dissociation of the individual protein subunits can occur (Pelley, 2007). As such, quaternary structures allow proteins to provide substrate binding sites, perform specialized functions and create concise spatial arrangements for the catalysis of chemical reactions (Veenhoff et al., 2002).

1.13.5 Relationship between protein sequence and structure

In 1973, Christian Anfinsen postulated that the 3D conformation, corresponding to the native structure of a protein depends on the AA sequence and has since become known as Anfinsen's dogma. In essence, this meant that given optimal conditions, the specificity of an AA sequence results in a stable native structure. Contrastingly, Levinthal (1968) hypothesized that the tertiary structure of proteins relies on complex pathways based on their environment and chemical properties. This hypothesis has since been coined Levinthal's paradox (Wright, 2011).

Furthermore, proteins display great structural dynamics as they are not static even during equilibrium (Argudo et al., 2017; Chen and Makhatadze, 2017). Still this inherent flexibility is essential for proteins to accurately and efficiently perform their regular functions (Teilum et al., 2009). Coupled with this, it is simpler to generate sequence data experimentally than to predict its native structure. This is especially highlighted by Levinthal's paradox. Therefore, many software accommodate this complexity in their predictions of 3D protein structures (Wright, 2011). An overview of the method used to predict and assess protein dynamics is discussed in the subsequent sections.

1.14 Computational methods to predict protein interactions

1.14.1 Homology modelling

While the collection of protein structures is slowly gaining momentum, the structure of numerous proteins remains unsolved. Though it is impossible to solve all protein structures in existence, innovation into the computational prediction of protein tertiary structures is slowly filling this

gap. One such method includes homology modelling. Homology or comparative modelling is a method used to predict the tertiary structure of proteins from its AA sequence. The notion is that the structures exhibits greater stability and evolves more slowly over time. As such, homology modelling assumes that while similar sequences will adopt an almost identical structure, distant sequences will also retain a high level of similarity. Consequently, the method relies on a template of a known structure to compare against and predict the unknown structure (Muhammed and Aki-Yalcin, 2018). Several steps are involved in homology modelling, namely, template selection, alignment of the template and target sequences, model building, model optimization and validation. Currently, several homology modelling software are available, however, the two most commonly used software include MODELLER and Swiss-Model. Unfortunately, a drawback of homology modelling is that these theoretical tertiary structures cannot be predicted if the corresponding template's structure of a protein in the same family has not been solved (e.g. NMR or X-ray crystallography) (Chenug and Yu, 2018). Despite this, homology modelling is considered highly accurate, fast and low cost with clearly defined steps (Muhammed and Aki-Yalcin, 2018).

1.14.2 Molecular docking

Proteins do not act in an independent manner in complex biological processes. In fact, proteins often interact with small molecules or ligands to complete important functional tasks (Salmaso, 2018). This phenomenon has since attracted a great many scientists in the drug industry. The ability to accurately predict binding modes and interpret these data can provide invaluable information for the development of novel drug binding ligands (Salmaso and Moro, 2018). These methods have since been referred to as molecular docking. Since molecular docking estimates the best binding pose of a ligand and its associated protein, it requires an experimental or theoretical predicted tertiary structure (Salmaso, 2018).

Molecular docking occurs in two stages, i.e. the search for conformation or binding poses and the scoring function that correlates a score to each binding pose generated (Kitchen et al., 2004; Huang and Zou, 2010). During conformational searches, the algorithm should accurately evaluate the specified conformation space outlined by the free energy landscape whereas the binding scores should be associated with the global minimum energy of the hypersurface[11] (Salmaso and Moro, 2018). In general, two approaches can be considered during docking, i.e. rigid and flexible docking. In the first approach, the independent protein and ligand are bound based on shape and

[11] A physical space that considers the N^{th} dimension instead of only 2D and 3D models.

volume. The second approach assumes flexible docking in which binding is accomplished based on reciprocal effects of both the ligand and protein (Prieto-Martínez et al., 2018). Based on these approaches, there are currently several molecular docking software currently available for commercial and research purposes (Ciemny et al., 2018). Importantly, while molecular docking is fervently employed in the drug industries, this approach has been used to evaluate the interactions of existing protein-ligand systems, such as in HIV (Tong et al., 2017; Tarasova et al., 2018; Vora et al., 2019).

1.14.3 Molecular dynamics

Since protein-ligand interactions play vital roles in key biological processes, studying these interactions is an important approach in understanding the basis of protein functionality (Fu et al., 2018). A popular approach to evaluate these interactions is molecular dynamics (MD). Simply put, a MD simulation is a method used to compute the behavior of atoms and molecules over time (Lipkowitz, 1990). Briefly, the atomic particles are modelled with a specified charge and mass. The electrostatic force field, calculated from the atomic charge is used to estimate the force on each atom in the system. The force is then used to update the positions of the atoms and evolve the system over time. Consequently, MD simulations provide useful information on protein conformation and dynamics (Frenkel and Smit, 2002). This method has gained enormous popularity over the years due to software upgrades that has vastly improved the speed of calculations (Perricone et al., 2018).

MD software include Chemistry of HARvard Macromolecular Mechanics (CHARMM), GROningen Machine for Chemical Simulations (GROMACS) and Assisted Model Building with Energy Refinement (AMBER). Apart from classical MD simulations, these packages also provide useful tools for protein assessment, including predicting the binding-free energy scores of protein-ligand systems as discussed below.

1.14.4 Predicting protein-ligand binding-free energies

As previously discussed, the recognition of ligands by proteins drive many biological processes. The strength of this recognition is characterized by binding affinities (Jiao et al., 2008). The two most popular methods used to determine binding-free energies are the Molecular Mechanics-Poisson Boltzmann and Generalized Born Surface Area calculations (MM-PBSA and MM-GBSA). The MM-P(G)BSA approach to estimate binding-free energies are based on thermodynamics as shown in equation (1) (Case, 2014). The equation shows that the binding-free energy is the difference between the solvated complex, ligand and protein.

$$\Delta G_{bind,solv} = \Delta G_{com,solv} - (\Delta G_{rec,solv} + \Delta G_{lig,solv}) \quad (1)$$

Each of these components are calculated based on the trajectory of the complex as well as the atomic interactions. Thus, equation (1) can also be written as equation (2):

$$G = E_{bnd} + E_{el} + E_{vdW} + G_{pol} + G_{np} - TS \quad (2)$$

In equation (2) the first three terms (E_{bnd}, E_{el}, E_{vdW}) are standard MM terms corresponding to bond, electrostatic and van der Waals (vdW) interactions. The polar and non-polar contributions are represented by G_{pol} and G_{np}. In the MM-P(G)BSA calculations, the polar contributions are estimated either using the GB model or PB equation. Contrastingly, the non-polar contribution is estimated by the solvent accessible surface area (SASA). The TS term refers to the absolute temperature (T) multiplied by the entropy (S). Entropy calculations are estimated using the normal mode analysis (Genheden and Ryde, 2015). Normal mode analyses have a large margin of error therefore introducing significant uncertainty. Additionally, these calculations are computationally expensive (Graham et al., 2013). As a result, entropy calculations can be omitted if comparisons are being made between ligands binding to the same protein.

1.15 Sequence-based methods in predicting protein interactions

In addition to the structural approach, sequence-based methods have also been employed to evaluate protein interactions (Rao et al., 2014). These methods are usually applicable to larger protein datasets since sequence information is available for most proteins as opposed to their structure (Sael et al., 2012) as discussed in 1.13.1. Additionally, a vast majority of protein functional information is maintained in the sequence databases (Sael et al., 2012). Examples of sequence-based methods include gene cluster methods, phylogenetic reconstruction, coevolution analyses and network-related methods (Liu et al., 2012).

In this study, we employed the use of phylogenic reconstruction, coevolution analyses and Bayesian networks to evaluate various co-evolutionary resistance pathways in Gag-PR (chapter three).

1.16 Study significance, hypothesis and research aims

South Africa has one of the largest ARV programs, globally that is unfortunately coupled with limited access to third-line treatment regimens (NDOH, 2019). An important challenge of this

consequence is the development of ARV drug resistance. Consequently, the emergence of acquired drug resistance in South Africa is on the rise (Chimukangara et al., 2019). With two PIs, i.e. LPV and DRV on the forefront of second- and third-line treatment regimens, respectively, the emergence of PI resistance mutations is concerning as is the evidence supporting the accumulation of resistance mutations in Gag (Fun et al., 2012). More so, information regarding the way in which Gag-PR mutational patterns are evolutionarily selected and interact on a structural level is limited (Zhang et al., 2010). Additionally, in spite of the large amount of structural information available on subtype B proteins, limited studies have focused on subtype C itself (Costa et al., 2014).

Thus, the focus of this study was to elucidate the mechanisms by which Gag-PR coevolution can affect LPV and DRV resistance in HIV-1 subtype C. In doing so, we did not only evaluate the pathways to resistance but also the structural implications of complex resistance patterns in evading drug selection pressure.

1.16.1 Hypothesis

It is hypothesized that Gag-PR coevolution drives the accumulation of complex drug resistance mutational patterns. It is further hypothesized that interactions of these patterns of mutations can alter the molecular structure of Gag-PR to evade drug binding and maintain enzymatic function.

1.16.2 Research aims

Based on the hypothesis, two main aims were evaluated in this study:

i. To determine Gag-PR coevolution under drug selection pressure.

ii. To evaluate the structural mechanisms by which complex, coevolving resistance mutational patterns affect PI drug binding and Gag cleavage.

1.17 Thesis outline

This thesis is presented in six chapters as follows:

- **Chapter one** provides a literature review of the HIV discovery, diversity and ARV treatment with focus on PR and its natural substrate, Gag. This section also highlights several concepts surrounding protein structure and function.
- **Chapter two** highlights the drug resistance mutations in Gag and PR using sequences from a South Africa cohort and public databases. This chapter also pinpoints patterns of PRMs.

- **Chapter three** investigated Gag-PR coevolution and describes the probable pathways to LPV/r resistance using combination of phylogenetic methods and Bayesian network learning.
- **Chapter four** describes the structural implications of a pattern of PRMs on LPV and DRV binding in two South Africa viral PR sequences.
- **Chapter five** investigated the coevolutionary impact between the mutant A431V NC|p1 Gag CS and PRMs described in chapter four. The focus of this chapter was to elucidate the mechanism by which the CSs are favoured over the drugs.
- **Chapter six** provides a general discussion and conclusion highlighting the main findings observed in chapters two to five. It also provides a list of study limitations and future recommendations.

Chapter two

Detecting drug resistance mutations in Gag and protease

2.1 Introduction

The PIs are amongst some of the most potent ARVs currently used to treat HIV-1 infected patients (Midde et al., 2016). Although the success of PIs cannot be disputed, the development of drug resistance mutations can increase the risk of therapy failure (Kožíšek et al., 2012; Baxter et al., 2016; Tsai et al., 2017). Therefore, the selection of protease resistance mutations (PRMs) can limit the options available for treatment while possibly allowing for resistance to develop in the other drug classes within the treatment regimen (Barber et al., 2012). Hence, the identification and effective characterization of PRMs can aid in pre-treatment screening processes that can possibly lead to the reduced risk of treatment failure.

Whilst mutations in the viral PR are widely known to inhibit the effects of PIs (Shafer, 2017), mutations in Gag have been documented to do the same (Tamiya et al., 2004; Dam et al., 2009; Clavel and Mammano, 2010; Li et al., 2014). Though mutations in Gag are noted, their presence has not been considered in the selection of drug regimens for the current PIs. The ability of Gag and PR to behave as "partners in resistance" (Fun et al., 2012) emphasizes our disregard on the selection of Gag mutations in resistance outcomes. Moreover, as with PR, certain known PI-resistance or exposure associated mutations in Gag have been linked to certain PIs (Cote et al., 2001; Prado et al., 2002; Mo et al., 2007; Larrouy et al., 2010). This suggest that Gag follows a similar trend to PR when selecting mutations under drug selection pressure.

Although information for patient sequences, drug regimens with their predictive efficaciousness and genotyped data indicating PI-resistance is publicly available, this information is primarily based on subtype B (Li et al., 2013). As previously mentioned in section 1.2 of the literature review, subtype C is currently the most predominate subtype (Günthard and Scherrer, 2016). Although subtype associated differences in the development of drug resistance is still debatable (Lessells et al., 2012), subtype genetic diversity (Wainberg and Brenner, 2012; Santoro and Perno, 2013) and discrepancies in the selection of mutations between subtypes (Grossman et al., 2004; Sui et al., 2014) is still important to consider.

Here we assessed the prevalence of drug resistance mutations in Gag and PR in HIV-1 subtype C PI-treatment associated datasets from publicly available repositories combined with sequences from a KZN cohort. It should be noted that some of these data was previously analyzed by Singh (2015) and is being reassessed here to confirm former observations made in our lab as well as to contextualize later chapters.

2.2 Methods

2.2.1 Sequence dataset

The dataset consisted of HIV-1 viral sequences genotyped from South African patients failing a PI-inclusive treatment regimen. To increase the sample size of the dataset, Gag and PR sequences were also retrieved from two HIV public databases, as described below.

2.2.1.1 PR Cleavage Site (PCS) cohort

The PCS cohort consisted of patients recruited between 2009–2013 at McCords and King Edward VIII hospitals in Durban, South Africa. All PCS study participants were on a LPV/r-inclusive treatment regimen for at least six months at the time of enrolment. These participants had a viral load of >1,000 copies/mL. The median CD4 cell count was 135 cells/µL prior to treatment. CD4 data following drug therapy was not available for these patients. The number of PR and Gag sequences included in the dataset was dependent on whether the samples were successfully sequenced by previous students (Pillay, 2015; Singh, 2015). Consequently, 86 Gag and 130 treatment exposed PR sequences were included in the final dataset.

This retrospective study was approved by the Biomedical Research Ethics Committee (BE446/15) and can be viewed in Appendix E.

2.2.1.2 Sequences obtained from public databases

A total of 44 Gag and 2,316 treatment exposed PR sequences from HIV-1 subtype C infected patients were obtained from the HIV Stanford Drug Resistance (https://hivdb.stanford.edu/) and Los Alamos Sequence (http://www.hiv.lanl.gov) databases. To ensure confidence in the study observations, the patients had to be on at least one PI to be included in the dataset.

The control groups, obtained from the Los Alamos database comprised 2,610 and 7,303 HIV-1 subtype C Gag and PR sequences, respectively. The dataset obtained for the control groups followed a stringent search criterion where all sequences had to be ARV naïve.

2.2.2 Sequence analysis

Each Gag and PR treatment associated dataset was grouped separately for sequence alignment and downstream analyses. For quality control purposes, duplicates and sequences with hypermutations were detected using the Los Alamos ElimDupes and Hypermut tools (http://www.hiv.lanl.gov) and subsequently removed (Li et al., 2013) as shown in Figure 2.1 below. The Gag and PR sequences were then aligned against a drug naïve patient-derived subtype C reference sequence (accession number: AY772699) and the subtype B HXB2 reference strain (accession number: K03455) in MAFFT v.7.402 (Katoh and Standley, 2013) and manually edited in BioEdit v.7.2.5 (Hall, 1999).

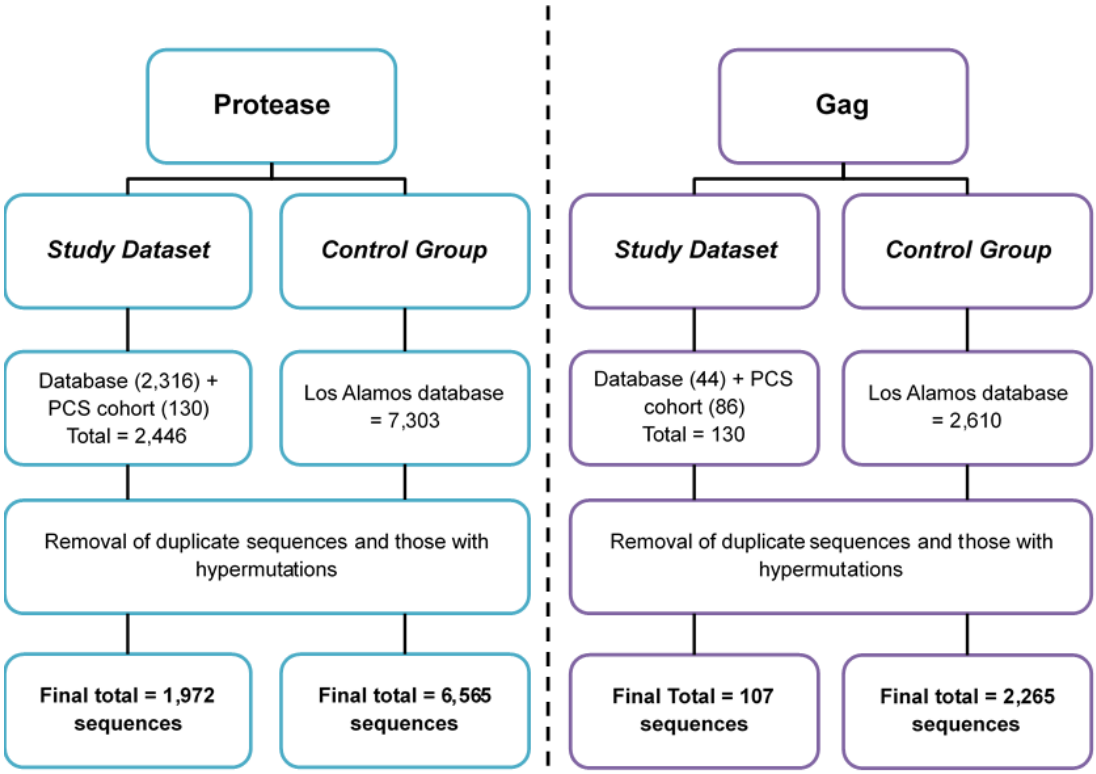


Figure 2.1 Overview of the study dataset and control groups utilized in this study for PR and Gag.

The RegaDB sequence analysis tool (www.rega.Kuleuven.be/cev/regadb) was used to perform a codon-by-codon analysis on the sequence datasets. The sequence analysis tool compared each sequence to the reference sequences and calculated the presence or absence of a mutation at each amino acid (AA) position. The frequency of mutations at each AA position was then calculated per the following formula:

$$F\ (\%) = \frac{S}{N}\ x\ 100$$

F : Frequency of each mutation
S : Sum of mutation
N : Total number of sequences at each position

Thereafter, the Stanford Resistance HIVdb algorithm (http://hivdb.stanford.edu/) was used to identify and interpret the PR drug resistance mutations. Mutations in Gag found associated with PI drug resistance or exposure was identified based on Fun et al. (2012).

2.2.3 Statistical analysis

The Fisher's exact test was performed to assess proportional differences of mutations between the treatment exposed and naïve Gag and PR sequence datasets. A p-value of <0.01 was considered statistically significant. The statistical tests were performed in SPSS v.25 (SPSS Inc., Illinois).

2.3 Results

2.3.1 Drug resistance mutations in PR

From the data analysis, 12 major PRMs and 20 amino acid variants were identified as seen in Figure 2.2. Overall, the highest percentage of mutations were identified at position 82 (25%). This was followed by positions 54 (23%) and 46 (19%). Interestingly, variation at these positions indicate that certain AAs are selected at a greater frequency than others. For example, mutation V82A (n=221; 23%) occurred at a higher frequency than V82C/M/S/T (n=23; 2%). This could also be seen for I54V (n=211; 22%) and M46I (n=157; 16%). Contrastingly, all other variants occurred at much lower frequencies (<3%), excluding L76V (n=70; 7%) and L90M (n=138; 14%). Of note, D30N, V32I, G48V, L76V, I84V, N88S and L90M were the only variants observed at these positions.

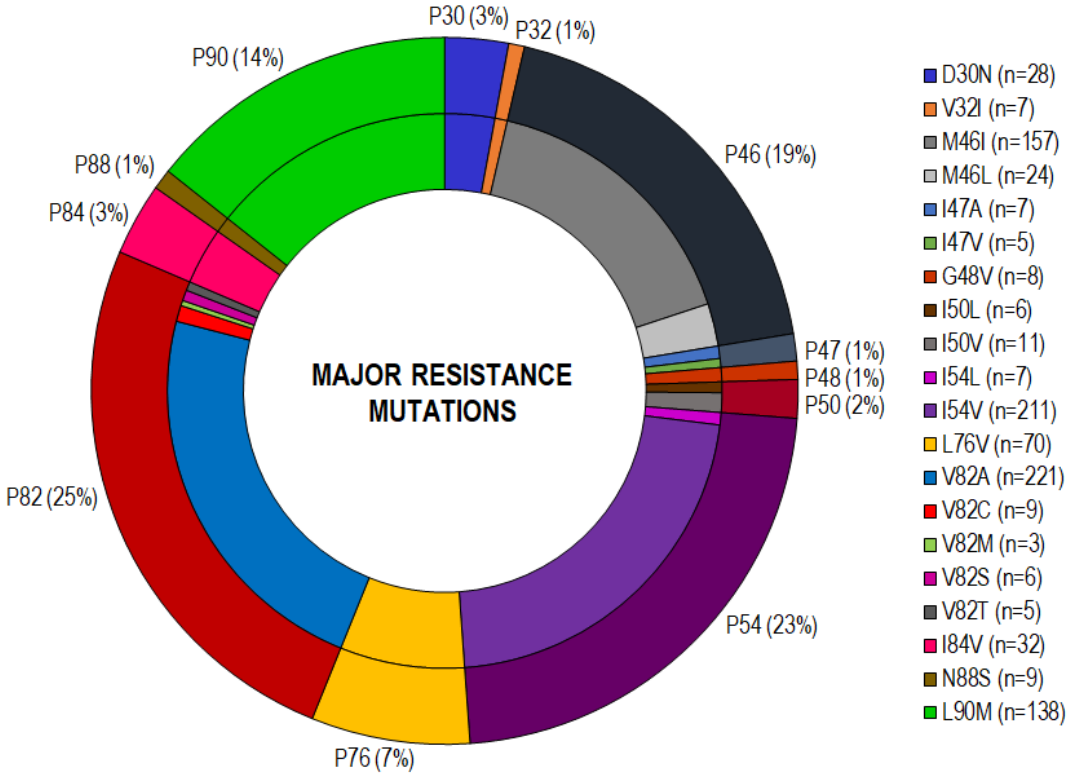


Figure 2.2 Proportion of position and specific major resistance mutations identified in the PR sequence dataset (only 964 out of the 1,972 PR sequences had major PRMs). **Outer ring:** Percentage (%) of overall mutations at a specific position. **Inner ring:** Proportion of specific amino acid variants. **Abbreviations:** P = position and n = total number of mutations at specific amino acid positions. **Note 1:** All amino acid variants were found significantly ($p<0.01$) higher in the treated vs naïve groups. Only I47V ($p=0.026$) and V82M ($p=0.011$) were non-significant ($p>0.01$). **Note 2:** Data from the naïve sequence dataset not shown.

Interestingly, several combinations of mutations ranging between one and six were observed amongst the major PRMs (Table 2.1). Apart from the mutations that occurred alone, one of the most commonly occurring combinations was M46I+I54V+V82A (red). In addition, this recurring pattern was also observed in many of the other combinations (i.e. combinations of 4–5) as highlighted in red in Table 2.1, including the most frequent combination of four M46I+I54V+L76V+V82A (purple). All other combinations varied and depicted no observable mutational pattern.

Table 2.1 Combinations of major resistance mutations identified in the 964 protease sequences. **Note:** Combinations ranged from one to six.

COMBINATIONS	n	COMBINATIONS	n	COMBINATIONS	n
ONE					
L90M	43	I50L	5	I47V	1
V82A	35	M46L	4	V82T	1
D30N	23	I47A	2	I84V	1
M46I	8	L76V	2	V82T	1
I54V	6	N88S	2		
TWO					
I54V+V82A	35	N88S+L90M	2	M46I+I84V	1
M46I+L90M	10	M46I+L76V	2	M46L+I47A	1
V82A+L90M	4	M46I+N88S	2	I47A+N88S	1
D30N+L90M	4	L76V+V82A	2	M46L+V82A	1
I54V+L90M	3	V32I+V82A	1	I54L+L90M	1
I54V+I84V	3	M46I+I50L	1	V32I+I47A	1
I84V+L90M	3	M46I+I50V	1	I54V+V82M	1
M46I+V82A	2	M46I+I54V	1	V82M+L90M	1
G48V+V82A	2	V32I+I47V	1	L76V+V82C	1
THREE					
M46I+I54V+V82A	43	I54V+V82S+L90M	2	M46I+V82A+L90M	1
I54V+V82A+L90M	13	M46I+I54V+L90M	2	M46I+L76V+L90M	1
M46L+I54V+V82A	7	I54V+V82A+I84V	2	G48V+I54V+V82T	1
I54V+L76V+V82A	5	M46I+I47A+I84V	1	I54L+I84V+L90M	1

M46L+V82A+L90M	4	V32I+I47A+V82A	1	V82A+I84V+L90M	1
M46I+I84V+L90M	3	M46I+L76V+V82A	1	D30N+I54V+V82A	1
I54V+I84V+L90M	3	M46I+N88S+L90M	1	I54V+V82C+L90M	1
M46I+I54L+L90M	2	M46L+I47A+N88S	1		
FOUR					
M46I+I54V+L76V+V82A	35	M46I+I54V+L76V+I84V	2	M46I+I54V+L76V+V82M	1
M46I+I54V+V82A+L90M	8	M46L+G48V+I54V+V82S	1	M46I+I50V+I54V+L90M	1
M46I+I50V+I54V+V82A	6	V32I+**M46I+I54V+V82A**	1	M46L+I54V+V82C+L90M	1
M46I+I54V+I84V+L90M	2	G48V+I54V+V82T+L90M	1	I54V+V82T+I84V+L90M	1
M46I+I54L+L76V+I84V	2	M46I+I54V+V82C+I84V	1	V32I+I54L+L76V+V82A	1
G48V+I54V+V82S+L90M	2	M46I+I54V+V82T+L90M	1		
M46L+I54V+L76V+V82A	2	I54V+V82A+I84V+L90M	1		
FIVE					
M46I+I54V+L76V+V82A+L90M	8	M46I+I50V+I54V+L76V+V82A	1	V32I+M46I+I54V+L76V+I84V	1
M46I+I50V+I54V+V82A+L90M	2	M46I+I47V+I54V+L76V+V82A	1	M46L+G48V+I54V+V82S+L90M	1
M46I+I54V+L76V+V82C+I84V	2	M46I+I47V+I54V+V82C+L90M	1		
M46I+I54V+V82A+I84V+L90M	1	M46I+I47V+I54V+L76V+V82C	1		
SIX					
M46I+I54V+L76V+V82C+I84V+L90M	2				

Note: n – frequency of occurrences observed within the study group.

Additionally, several resistance-associated mutations occurred at varying frequencies as seen in Figure 2.3. Of these, 12 were classified as non-polymorphic, minor resistance mutations and four as other, polymorphic PI accessory mutations by the HIVdb algorithm. Excluding positions 10 (16%), 20 (5.9%), 71 (6.1%) and 74 (16.8%), variation at the remaining positions were infrequent, occurring at 2.1% or less. The most common variants at these positions included L10F (n=126; 6.4%), L10I (n=107; 5.4%), A71V (n=111, 5.6%) and T74S (n=311; 15.8%).

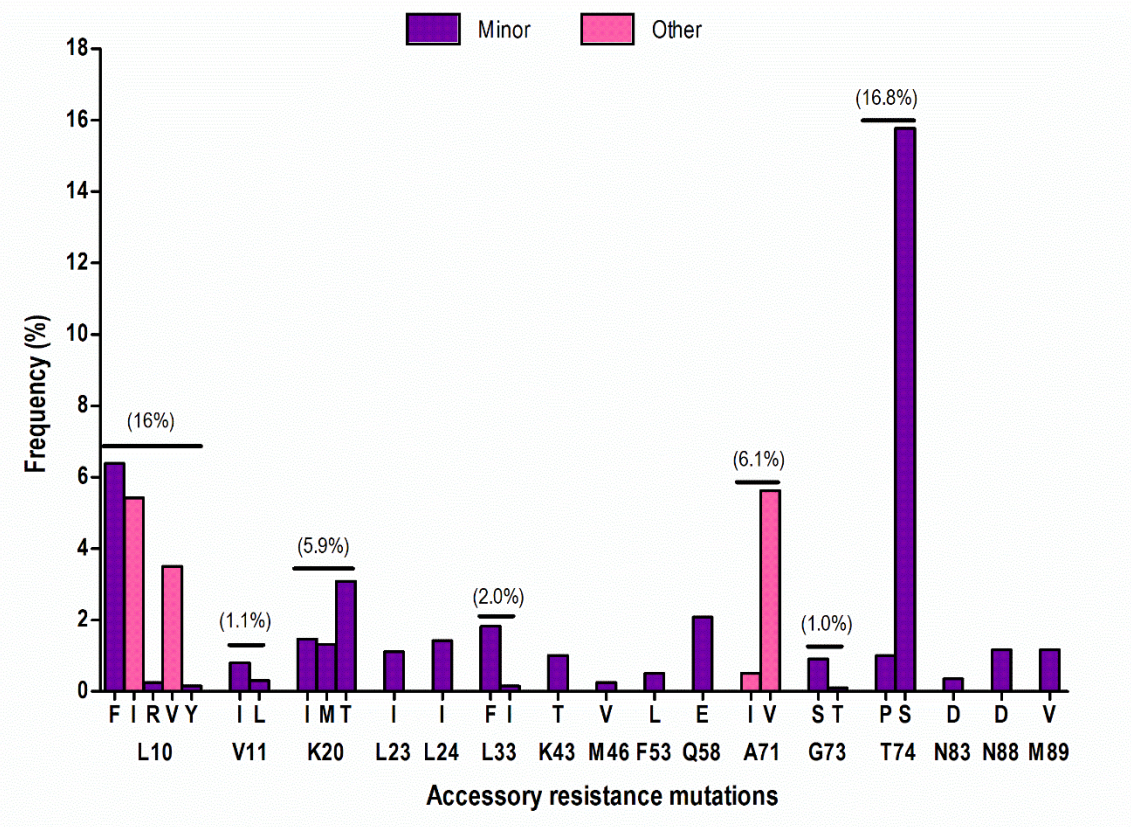


Figure 2.3 Frequency of accessory resistance mutations[12] identified in the viral PR with known, non-polymorphic, minor mutations represented by the purple bars and other, polymorphic mutations identified during drug resistance represented by the pink bars. **Note 1:** For those with multiple amino acid variants, the cumulative percentage is displayed above the coloured bars. **Note 2:** All amino acid variants were found significantly ($p<0.01$) higher in the treated (n=1,972) vs naïve (n=6,565) groups except for L10V (p=0.000), V11I (p=0.270), K20M (p=0.000) and T74S (p=0.000) which occurred more frequently in the control groups. **Note 3:** Data from the naïve control dataset not shown.

[12] The Stanford HIVdb algorithm classifies PR's accessory resistance mutations in two ways: i) non-polymorphic, accessory/minor resistance mutations that are associated with reduced susceptibility to PI drugs and as ii) other, polymorphic, accessory mutations that can increase viral replication with PI-resistance mutations.

2.3.2 Gag mutations associated with PI resistance/exposure

As the concept of PI drug resistance associated with mutations in Gag is still gaining momentum, there are currently no software-based tools that can identify resistance mutations in Gag based on the sequence dataset. Consequently, manual methods of identification are required based on previously published works. The review conducted by Fun et al. (2012) provides a comprehensive list of resistance associated mutations in Gag that were determined based on *in vitro* studies. The paper also goes on to classify these mutations as either resistance associated, i.e. they were found to directly impact PI resistance, or exposure associated (those that arose as a result of PI exposure but may not have had a significant impact of drug resistance). Therefore, based on Fun et al. (2012), a total of 30 Gag mutations were identified in the Gag treated sequence dataset. These mutations were further grouped into two categories, comprising 11 resistance and 19 exposure associated mutations.

The resistance associated mutations included R76K, Y79F, T81A, V128I, Y132F, A431V, K436R, I437V, L449F/P/V, R452K and P453L/T as shown in Figure 2.4. Amongst these, CS mutations A431V (NC|p1), L449F (p1|p6) and R452K (p1|p6) revealed significantly ($p<0.01$) higher frequencies in the treated versus naïve groups whilst the remaining mutations occurred at comparable, non-significant ($p>0.05$) frequencies in both sequence datasets.

This suggests that unlike the CS mutations which occurs to actively play a role in drug resistance (i.e. genetic selection), the other resistance-associated mutations possibly occur more randomly but aids in resistance when present under drug selection pressure.

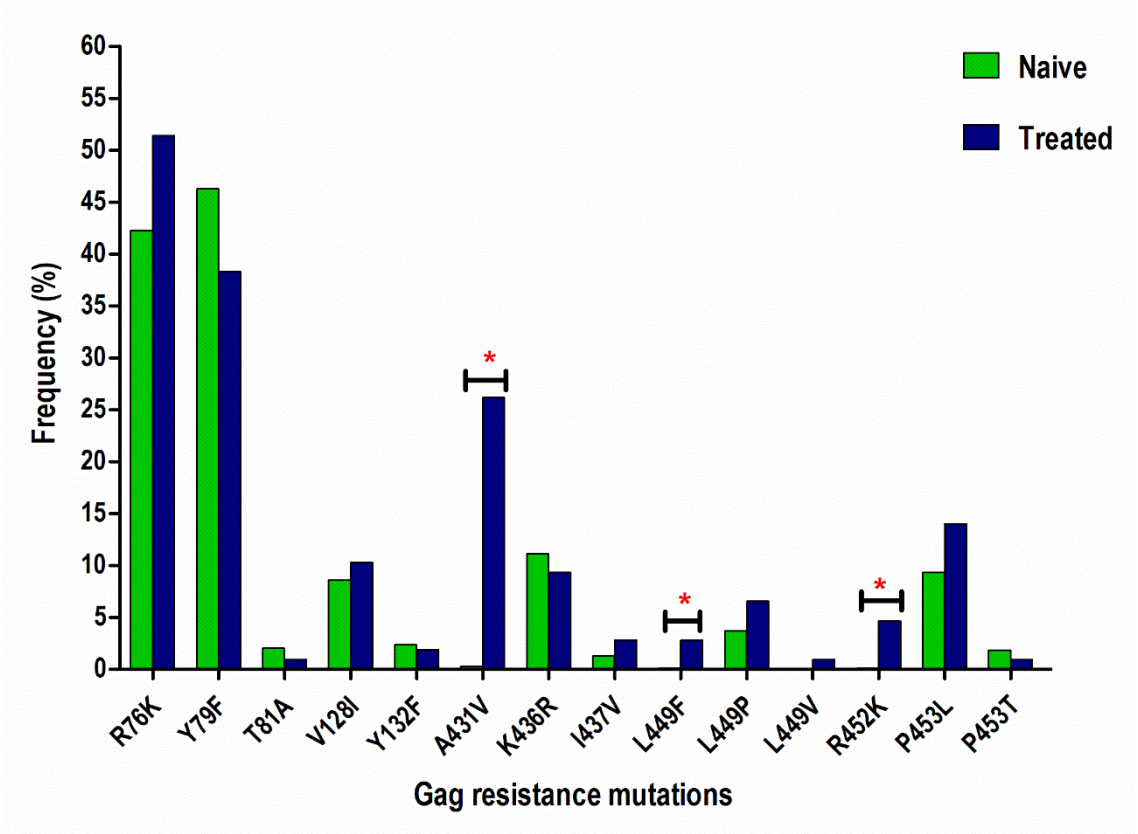


Figure 2.4 Frequency of Gag resistance-associated mutations in the treated (n=107) versus naïve (n=2265) sequence datasets. Significant differences ($p<0.01$) are denoted by the red asterisks (*).

Figure 2.5 provides an overview of the known PI exposure associated Gag mutations identified in the sequence dataset. Accordingly, mutations K62R, N373P, I401V, R409K and S451T were significantly ($p<0.01$) higher in the treated versus naïve group. Additionally, apart from N373P and S451T which are p2|NC and p1|p6 CS mutations, respectively, all other mutations were detected at non-CS positions.

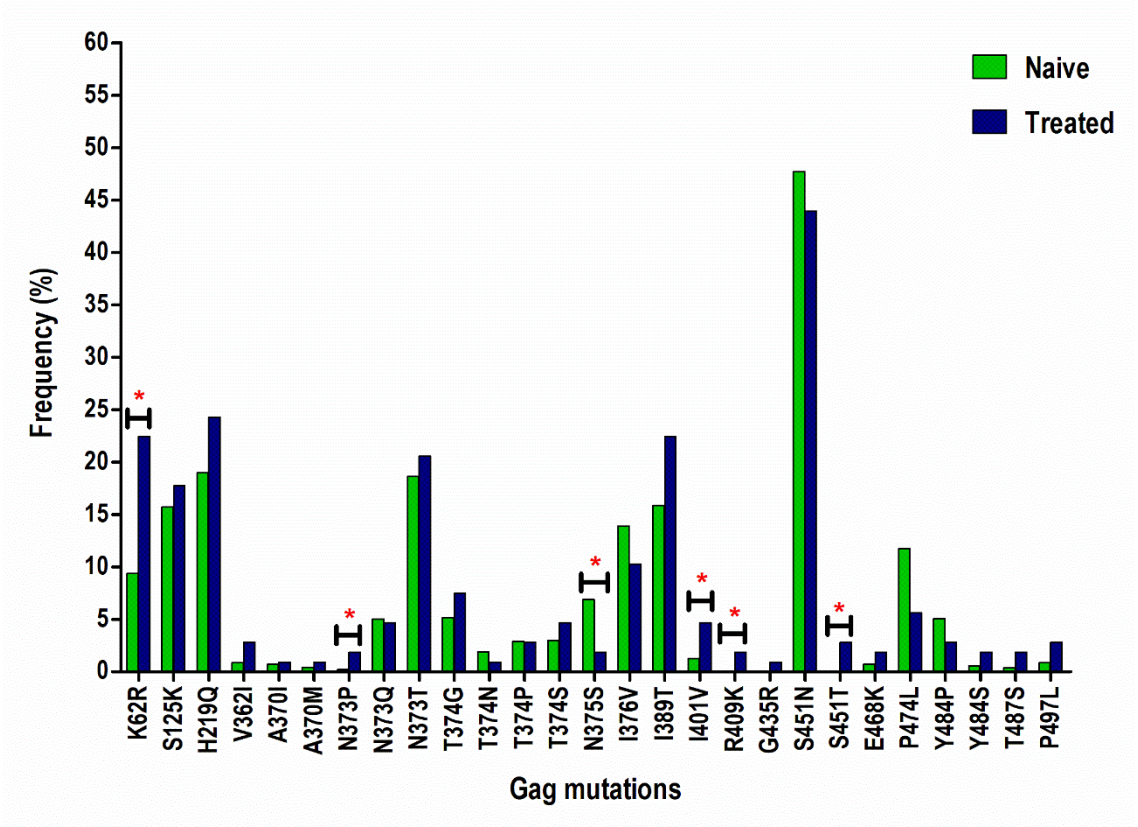


Figure 2.5 Frequency of Gag exposure-associated mutations in the treated (n=107) versus naïve (n=2265) sequence datasets. Significant differences ($p<0.01$) are denoted by the red asterisks (*).

2.3.3 Identification of presumptive Gag resistance associated mutations

To identify presumptive Gag resistance-associated mutations, the frequencies of mutations in the treated versus naïve groups were compared. Accordingly, mutations that occurred significantly ($p<0.01$) greater than 10% in the treated versus naïve group were considered resistance associated mutations.

Twelve presumptive Gag mutations were identified in the sequence dataset, as depicted in Figure 2.6. Interestingly, these mutations were only found at non-CSs. Of the 12 mutations, six (T53N, S54A, Q69K, D122T, S125D and N126E) were identified in matrix (p17), four (D260E, I256V, D319E and R335K) in capsid (p24) and two (P485A and L498S) in p6. No mutations were identified in the NC domain.

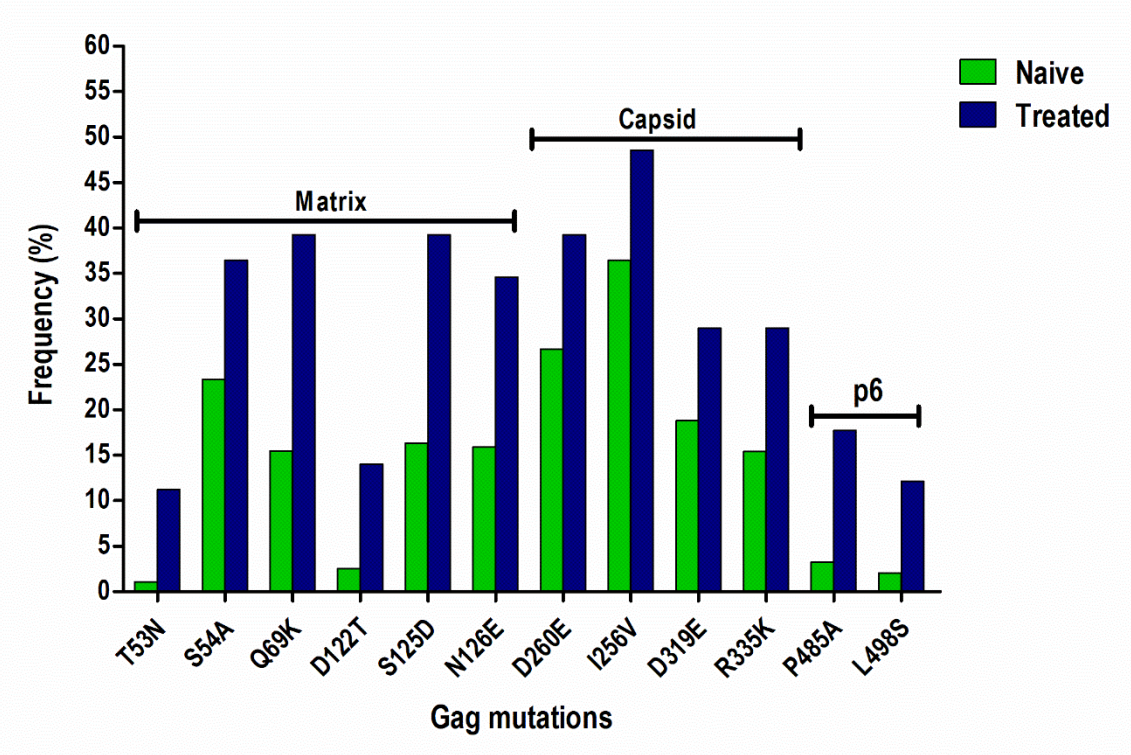


Figure 2.6 Frequency of presumptive Gag resistance-associated mutations identified in the treated (n=107) versus treatment naïve (n=2265) sequence datasets.

2.4 Discussion

Using a large HIV-1 subtype C dataset comprised of publicly available sequences and those of patients recruited in KZN, this study identified various PI-resistance mutations previously linked to both PR and Gag. Additionally, several presumptive Gag mutations associated with PI-resistance/exposure were also identified.

These data revealed that out of the 1,972 PR sequences, only 964 (49%) harboured major PRMs. As such, more than half (51%) of the PR sequence dataset did not present with drug resistance mutations, suggesting that non-adherence is a prominent factor in the development of resistance mutations, as reported by other studies (Levison et al., 2011; Wallis et al., 2011; Court et al., 2016; Kyaw et al., 2017).

In this study positions 82, 54 and 46, having the highest percentage of variation, are in the active site (82) and flap regions (46, 54) of the viral PR and are therefore integral for substrate cleavage (Saleh et al., 2017). As such, mutations at these positions can aid the virus in outcompeting the inhibitor for more efficient cleavage and subsequently escape drug selection pressure (Liu et al.,

2008; Weber and Agniswamy, 2009; Yu et al., 2015). Additionally, the selection of specific mutational variants (M46I, I47A, G50V, I54V, V82A) over others (M46L, I47V, G50L, I54L, V82CMST) at these positions suggests that AA selection also has an important role in maintaining protein functionality when resistance is induced. Dwyer (2001) showed that the electronic characteristics[13] of an AA could be comparable to the effect of hydrophobicity on the conformational preferences during protein folding. In this study, the frequently occurring triplet of M46I+I54V+V82A are described as common PI-resistance mutations by the Stanford Drug Resistance Database in ATV/r[14], FPV/r, IDV/r, LPV/r[2] and NFV/r (Appendix D; Figures 1–2). Moreover, the selection of L76V in conjunction with the above mutations, as observed in this study, has been shown to not only enhance LPV/r resistance (Young et al., 2010) but also allows for low-level DRV/r resistance to occur (HIVdb algorithm; Appendix D; Figure 3).

Whilst several accessory PI mutations were observed in this study, the most prevalent included L10F/I, A71V and T74S. The L10F/I mutations have been linked to reduced susceptibility and/or increased replication capacity to numerous PIs when found associated with other resistance mutations (Prado et al., 2002; Rhee et al., 2010; De Luca et al., 2016; Aoki et al., 2018). The polymorphic A71V mutation has been found in association with other PI-resistance mutations in numerous studies (Shafer et al., 2007; Lopes et al., 2015; Huang et al., 2017; Pessôa and Sanabani, 2017). Although the role of A71V is largely unknown (Gonzalez et al., 2008), a study conducted by Chang and Torbett (2011) revealed that in combination with I84V, A71V restored the structural stability of the PR to within 1°C of the WT[15]. The high prevalence of T74S observed in this study was consistent with other studies investigating the selection of drug resistance mutations in patients receiving a PI-inclusive treatment regimen (Taylor et al., 2011; Etta et al., 2017; Lu et al., 2017). Although T74S is polymorphic in most non-B subtypes (Deforche et al., 2006), its presence in treatment-associated outcomes suggests that it provides a selective advantage and can either improve or maintain viral replication capacity in the presence or absence of PR resistance mutations, respectively.

[13] Electronic characteristics refers to the physico-chemical nature of the amino acid side chains such as the hydrophobicity, pKa, steric effects (bulk of the side chain), etc. Thus, these characteristics were used to provide a scale: O (strong electron donor), U (weak donor), Z (ambivalent), B (weak electron acceptor) and X (strong acceptor) (Dwyer, 2001).

[14] High-level drug resistance.

[15] In other words, there was only a 1°C difference in the melting temperature between the WT and mutant when measured using Differential Scanning Calorimetry. Therefore, the mutant is comparatively stable.

Of the 11 resistance-associated Gag mutations, only the A431V (NC|p1), L449F (p1|p6) and R452K (p1|p6) mutations were found significantly higher in the treatment exposed versus naïve sequence dataset (Figure 2.4). The A431V mutation has been linked to several PIs, including LPV/r (Mammano et al., 1998; Cote et al., 2001; Mo et al., 2007; Nijhuis et al., 2009). In this study, the A431V mutation was found in 26% (n=107) of the viral Gag sequences. It has been shown to confer resistance in the absence of major PRMs (Dam et al., 2009; Nijhuis et al., 2009) and act in a compensatory manner when PRMs are present by improving viral fitness (Mammano et al., 2000; Cote et al., 2001). This mutation often occurs with major PRMs, V82A (Bally et al., 2000; Malet et al., 2007) and M46I/L (Myint et al., 2004). A study conducted by Nijhuis et al. (2009) on subtype B infected patients receiving a LPV/r-inclusive treatment regimen revealed that A431V was also selected in viruses comprising an M46I+L76V PR double mutant, increasing LPV/r resistance by 10.6-fold and improving viral replication capacity by 10%. Contrastingly, Myint et al. (2004) revealed that A431V had deleterious effects on the replication capacity of D30N, N88D and L90M PR mutants suggesting that A431V is synergistically co-selected with specific PRMs.

Unlike A431V, the L449F mutation increased viral replication capacity in the PR mutants observed in the Myint et al. (2004) study. This mutation has been shown to reduce the susceptibility of various PIs when combined with major PRMs D30N/N88D and with V82A or L90M in subtype B (Fun et al., 2012). However, unlike A431V, L449F cannot confer resistance alone (Fun et al., 2012), and instead acts in a solely compensatory role (Yates et al., 2006; Girnary et al., 2007; Kolli et al., 2009). Recently, in a structural study investigating the effects of p1|p6 CS mutations on an I50V+A71V subtype B PR mutant, Özen et al. (2014) demonstrated that L449F increased the van der Waals forces in the enzyme-substrate complex, thereby restoring PR's active site dynamics. Similar to L449F, the R452K Gag mutation, frequently occurring with D30N/N88D, I84V and I50V, has a compensatory role during PI-selection pressure (Kolli et al., 2009).

Additionally, several Gag mutations associated with PI exposure were identified in this study, five (G62R, S373P, I401V, R409K and S451T) of which occurred at significantly higher frequencies in the treated dataset (Figure 2.5). Although the G62R mutation occurred in 22% of the Gag sequence dataset in this study, its role in resistance is not yet understood (Koh et al., 2009). In a study investigating a novel PI (GRL-02031) against a multi-drug resistant PR, Koh et al. (2009) observed that G62R along with three other Gag mutations developed upon a 37th passage after exposure to GRL-02031 on MT-4 cell lines. Interestingly, one of the other mutations

found in the Koh study included NC mutation, R409K. Despite evidence linking R409K to APV/r (Gatanaga et al., 2002; Aoki et al., 2009), these studies did not report changes to APV susceptibility *in vitro*. As such, its role in PI-resistance is unclear. However, Stray et al. (2013) reported that R409K together with five other Gag mutations was able to reduce the efficacy of a novel PI (GS-8374) by 14-fold after >300 days of passage on MT-2 cell lines. This study suggests that since only one polymorphism was selected in PR, the Gag mutations were selected to confer resistance, playing a role as primary resistance mutations. In a similar manner to R409K, NC mutation I401V has been linked to treatment failure (Myint et al., 2004), however the mechanism by which it acts is not understood.

This study identified two Gag CS PI-exposure associated mutations (S373P and S451T). Codon 373 of the p2|NC CS is highly variable and comprises three mutational variants (S373P/Q/T) that are associated with PI-exposure (Fun et al., 2012). The S373P mutation has been linked to LPV studies in subtype B (Mckinnon et al., 2011; Sutherland et al., 2014) and in CRF02_AG (Teto et al., 2017). The S451T mutation in the p1|p6 region has been associated with LPV in subtype B (Masse et al., 2007) and DRV in the POWER[16] trials (Dierynck et al., 2007).

Finally, 12 presumptive resistance associated mutations spanning the MA, CA and p6 regions of Gag were also observed (Figure 2.6). Previous data from our lab identified only two of the 12 presumptive Gag mutations, namely, Q69K and I256V (Singh, 2015). However, the dataset was considerably smaller (954 subtype C naïve and 54 acutely infected individuals vs. 80 PI-treated individuals) than what was used in the current study. While mutations outside the Gag CSs have been linked to drug resistance (Gatanaga et al., 2002; Callebaut et al., 2011; Parry et al., 2011), the numerous functional roles of MA, CA and p6 as discussed in section 1.10 of the literature review, suggests that mutations in these domains arise to maintain proper function at various stages in the virus's lifecycle rather than affecting resistance in a more direct manner, like the CS mutations (van Maarseveeen et al., 2012).

2.5 Conclusions

The aim of this chapter was to identify the selection of drug resistance mutations in Gag and PR in treatment vs naïve groups. In this study, several major PRMs, mostly associated with LPV/r were identified. Using a large dataset, these data supported previous findings observed in our lab

[16] The POWER trials were clinical trials used to evaluate the efficacy and safety of Darunavir with low dose Ritonavir in comparison to other PIs.

(Singh, 2015) on the selection of M46I+I54V+V82A as well as several studies reported elsewhere (Van Zyl et al., 2013; Grossman et al., 2014). Moreover, the inclusion of L76V to this triple combination was not only a prominent feature of the findings observed in this study but also of those reported by Singh (2015). Additionally, these data indicated that certain Gag mutations, specifically A431V, appears to play a role in treatment associated outcomes. Therefore, chapter three investigated Gag-PR coevolution to evaluate how mutations in PR and Gag interact in pathways leading to LPV/r resistance.

Chapter three

Evaluating the coevolving nature of the HIV-1 Gag-protease proteins under ARV drug selection pressure

--This chapter was adapted from our research publication entitled "*Gag-protease coevolution shapes the outcome of Lopinavir-inclusive treatment regimens in chronically infected HIV-1 subtype C patients*".

V. Marie and M. Gordon, *Bioinformatics*, Published Ahead of Print, doi: 10.1093/bioinformatics/btz076

3.1 Introduction

As mentioned previously (see section 1.9.4 of chapter one), LPV/r and DRV/r are two HIV-1 PIs that have the highest genetic barriers, requiring at least three mutations in LPV/r and more in DRV/r for high-level drug resistance to occur (Doherty et al., 2011). Consequently, the multi-step mechanism and high genetic barrier of these PIs results in a complex combination of mutational pathways leading to drug resistance (Rabi et al., 2013). Although many combinations of PI mutations occur, some are more frequent than others suggesting that these patterns of mutations are not occurring at random (Zhang et al., 2010).

However, recognizing Gag-PR coevolution and its role in facilitating alternative pathways to PI resistance is crucial. Despite our knowledge on the impact of PRMs on therapy failure, the introduction of Gag mutations with PR remains largely unclear (Li et al., 2014). Since protein coevolution can result in mutations at functionally important sites (Chakrabarti and Panchenko, 2010), the intimate interactions of Gag and PR during viral maturation can allow these proteins to coevolve and subsequently escape drug selection pressure (Giandhari et al., 2016). Therefore, it is important to understand these resistance dynamics to fully elucidate the pathways leading to drug resistance and ultimately therapy failure.

Here we used Bayesian network learning (BNL), which is a statistical model that describes the conditional independencies between a set of variables (Pearl, 1998), to elucidate pathways leading to LPV/r failure. To identify the sites at which coevolving Gag and PR residues occur, we utilized a coevolution sequence analysis tool and conducted a positive selection analysis. Finally, using these data we evaluated the multivariable effect of Gag and PR mutations. To the best of our knowledge, this is the first study used to explore the combined effects of Gag and PR mutations in pathways to LPV/r resistance.

3.2 Methods

3.2.1 Sequence dataset

Fifty-Eight complete Gag and PR sequences were retrieved from the PCS cohort as described in section 2.2.1.1 of chapter two. Noteworthy, unlike the other 57, one patient was on another PI (i.e. Ritonavir) during first-line before switching to LPV/r and failing second-line treatment, like the other 57 patients. Prior to concatenation, the Gag, PR and reference sequences were aligned

and manually edited as described in chapter two, section 2.2.2. The final concatenated sequence alignment length was 1,797 nucleotides.

3.2.2 Phylogenetic reconstruction

Phylogenetic trees were constructed in MrBayes v.3.2 (Ronquist et al., 2012) for Bayesian inference (BI) and RaxML v.8.1.17 (Stamatakis, 2014) for maximum-likelihood (ML) estimation. Prior to tree generation, the best-fitting nucleotide substitution model was determined according to the Akaike Information Criterion (AIC) using jModelTest. The GTR+I+G model selected by AIC, together with the likelihood calculations, was used to modify the following parameters in the MrBayes block: revmatpr = dirichlet (1.9444, 6.2812, 1.0534, 1.0472, 8.3907, 1.0000), statefreqpr = dirichlet (0.3750, 0.1868, 0.2374, 0.2008), shapepr = fixed (0.6600) and pinvarpr = fixed (0.3860). Tracer v.1.5 (http://beast.bio.ed.ac.uk/software/tracer) was used to assess the prior and posterior distributions, following which 25% of the samples were discarded as burn-in and a consensus tree was generated. The RaxML trees were generated via 1000 steps of rapid bootstrapping and a ML search with model parameters estimated up to 0.001. FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) was used to observe the topology of the phylogenetic trees and the Java-based application Compare2Trees (http://www.mas.ncl.ac.uk/~ntmwn/compare2trees/index.html) was used to compare tree topology between the two methods.

3.2.3 Identifying positive selection

To examine the selection pressure in Gag-PR, the rate ratio of non-synonymous to synonymous (ω) substitutions were estimated using the CodeML program in the PAML 4.0 package (Yang, 2007). Accordingly, ω=<1, ω=1, ω=>1 assumes purifying, neutral and positive selection, respectively. The ω value and likelihood estimates were calculated for three pairs of site models: M0 (one ω) vs M3 (discrete), M1a (nearly neutral) vs M2a (positive selection) and M7 (beta) vs M8 (beta&ω=>1). Using the likelihood ratio test (LRT), the statistical significance between pairs of site models were used to assess the model fit to the data and indicate positive selection. Briefly, the LRT is twice the log-likelihood difference between each pair of models [$2\Delta\ell = 2(\ell_1 - \ell_0)$]. Since the LRT follows an asymptotically chi ($\chi 2$) distribution with the number of degrees of freedom calculated in the difference of free parameters, it can be used to test if the null model is accepted against the alternative model (p-value). Thereafter, the Bayes Empirical Bayes method was used to identify specific sites under positive selection.

3.2.4 Evaluating the coevolving nature of Gag-PR

The Coevolution Analysis for Protein Sequences (CAPS) program (Fares and Travers, 2006) was used to evaluate Gag-PR coevolution. Briefly, the time-dependent evolutionary variation identifies the coevolving AA sites *e* and *k* by comparing the transition of AAs at sequences *i* and *j* ($\theta ek)ij$. The evolutionary variation is estimated using a blocks substitution matrix (BLOSUM) and the transitions of AAs are corrected by calculating the divergence time of sequences *i* and *j*. The comparison of 5000 randomly sampled estimates were used to identify the coevolving codons. Furthermore, CAPS corrects for multiple tests of non-independence data by a step-down permutation procedure.

3.2.5 Mapping resistance pathways with BNL

BNL was performed using the B-Course software (Myllymäki et al., 2002). The BNL dataset consisted of an equal amount of HIV-1 subtype C PI-naïve sequences of Gag-PR (obtained from the Los Alamos Database). Only mutations greater than a 10% prevalence in the treatment exposed and naïve sequence datasets were included in the network. Two networks were generated, one independently constructed for PR whilst the second incorporated variables from Gag-PR to observe a compounded effect. The robustness of the networks was assessed with a non-parametric bootstrap analysis of 5,000 replicates calculated in R. Additionally, for every resampled dataset, BNL was performed again in B-course to ensure the dependencies observed in the networks remained unchanged after multiple comparisons. In the networks, nodes are the colour coordinated variables while arcs are the arrows connecting them. Variables belonging to the same position are grouped together. While BNL implies bi-directionality wherein the presence of a mutation affects the prevalence of another mutation, arc direction does not indicate causality. Instead, arc direction allows a multivariable effect to be observed based on individual mutation conditional dependencies (Deforche et al., 2006).

3.3 Results

3.3.1 Tree reconstruction and the detection of positively selected Gag-PR sites

To evaluate the selection pressure and coevolution of Gag and PR, phylogenetic trees were inferred using ML and BI methods. Tree topologies for both the ML and BI methods were similar with an overall topological score of 93.67% as observed in Figure 3.1.

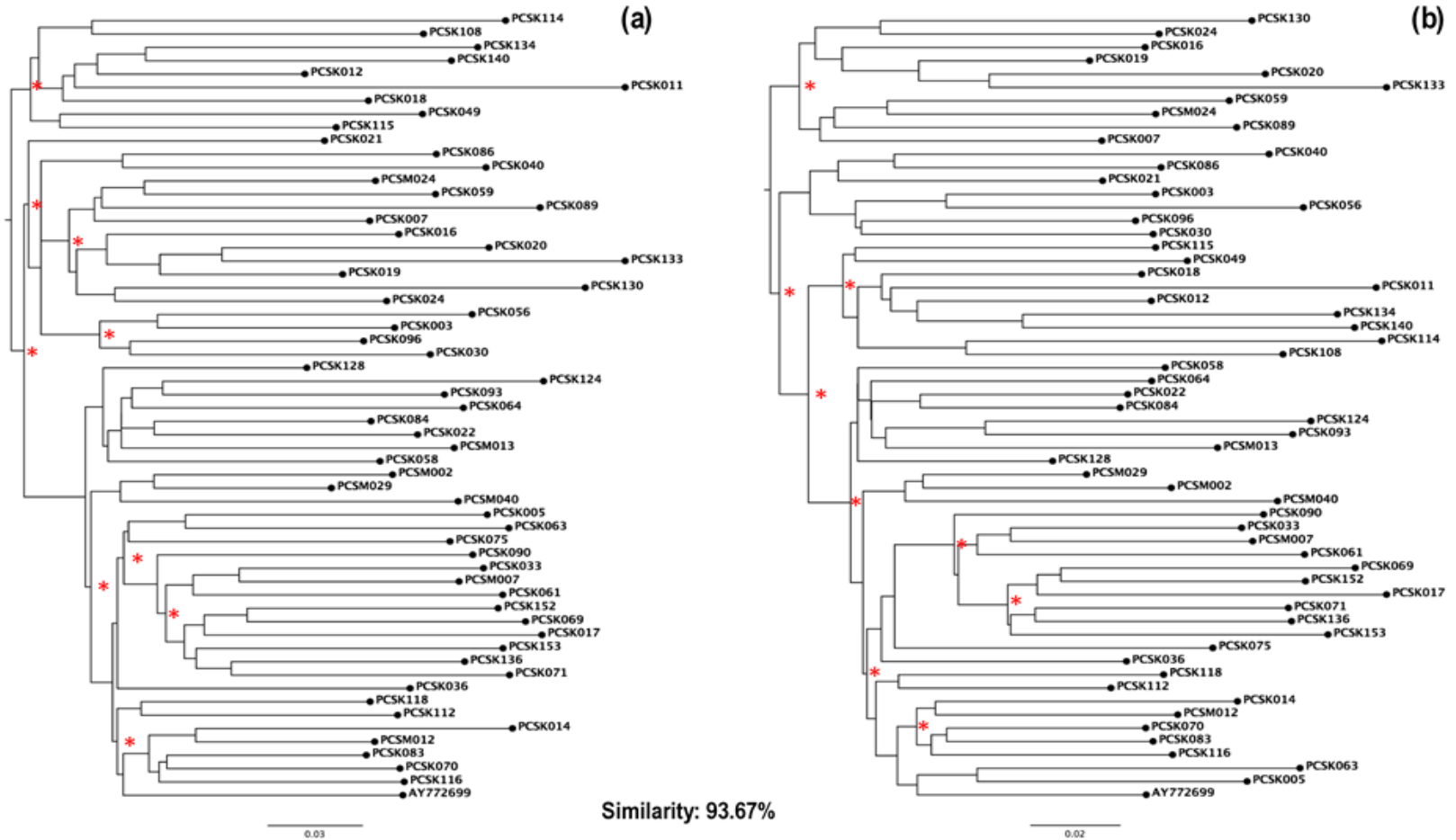


Figure 3.1 Phylogenetic reconstruction of the Gag-PR sequences showing **(a)** maximum-likelihood (ML) and **(b)** Bayesian inference (BI) trees. Bootstrap support from 1000 replicates (ML) and posterior probabilities (BI) >70% are shown by the red asterisks (*). Similarities between the trees as computed by the Java-based application Compare2trees is indicated below the figure.

The PAML package identified 24 positively selected sites at the posterior probability level of greater than 99% as seen in Figure 3.2. Out of the 24 positively selected sites, three were from PR, namely codons L10, L63 and V82. While codons 10 and 82 are sites at which drug resistance mutations occur, codon 63 is regarded as one of the most variable sites in PR. In Gag, 21 positively selected sites were identified including eight in MA, seven in CA, three between the p2 and NC regions as well as three in the p6 domain. Interestingly, four codons linked to PI exposure associated mutations, namely A62, N374, S381 and P474 were identified (Figure 3.2) while only one PI resistance associated codon (Y79) was selected by the CodeML analysis.

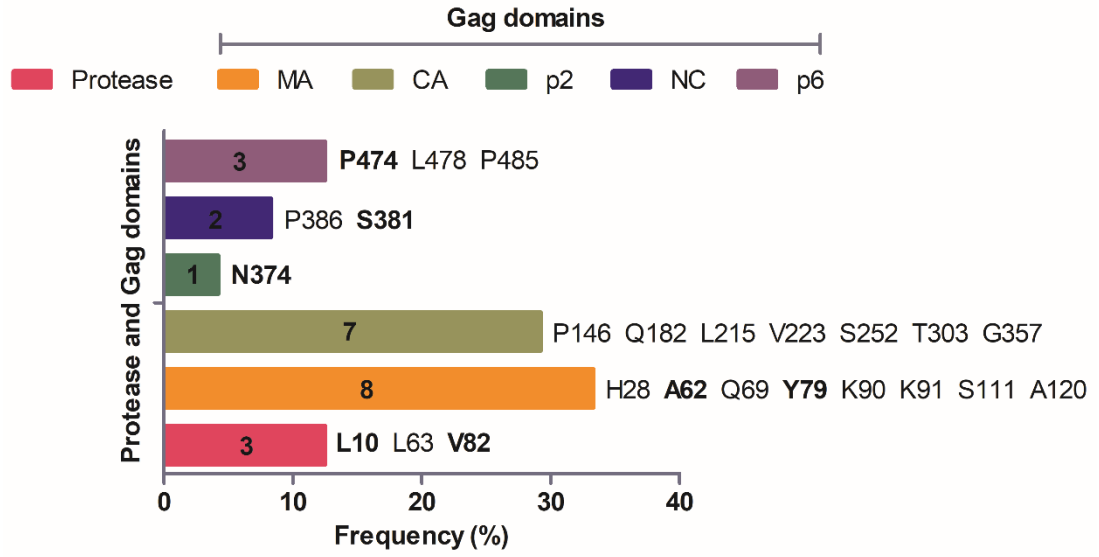


Figure 3.2 Frequency of the positively selected amino acid codons in Gag and PR. The numbers within the coloured bars indicate the number of codons selected at a specific region. **Note:** Bold font – sites associated with drug resistance/exposure mutations.

3.3.2 Gag and PR coevolution

Similarly, to previously reported studies (Travers et al., 2007; Codoñer et al., 2017), we used CAPS to identify several coevolving Gag-PR residues under drug selection pressure. Our analyses revealed significant correlations in PR between codons at which minor (10, 24 and 58) and major (46, 54 and 82) PRMs occur. Furthermore, an overlapping coevolving codon group of 10, 46, 54 and 82 was also observed. Interestingly, out of these codons, 10 and 82 was also positively selected by PAML (Figure 3.2). Finally, Table 3.1 describes the PRMs identified by the CAPS analysis and their structural location in PR.

Table 3.1 Protease amino acid codons identified in the coevolution analysis.

Residue in Protease	Structural Position	Classification[a]	Common Mutations[b]
10	10's loop	Minor	L10F/I/R/V/Y
16	10's loop	Other	?
20	Core domain	Minor	K20I/M/R/T/V
24	Active site region	Minor	L24F/I
33	Core domain	Minor	L33F/I/V
46	Flap region	Major	M46I/L/V
54	Flap region	Major	I54A/L/M/S/T/V
58	Core domain	Minor	Q58E
63	60's loop	Other	L63P
71	Core domain	Other	A71I/L/T/V
77	Core domain	Other	V77I
82	Active site region	Major	V82A/C/F/I/L/M/S/T
90	Core domain	Major	L90M

a Mutations were classified accordingly: ***major*** or primary resistance mutations are known to directly cause drug resistance, ***minor*** or accessory resistance mutations arise at a later stage to compensate for the changes associated with major mutations, and ***other*** refers to mutations or polymorphisms classified as neither a major or minor resistance mutation.

b Mutations obtained from the Stanford HIV drug resistance list.

Several codons in Gag found associated with PI resistance and/or located at the Gag CSs including E365, S373, A374, A431, R452 and Q474 were identified. Of interest was codon 431 which is commonly associated with Gag CS resistance mutation A431V (see chapter two, Figure 2.3). Although A431 only correlated with polymorphic codon 77 in PR, it had a significant role in a pathway leading to LPV/r failure which is discussed in the Gag-PR network below.

3.3.3 Most probable network in PR

Using BNL we determined the most probable network for PR under drug selection pressure (Figure 3.3A). Twenty variables including the LPV/r treatment node were present in the final dataset. Of these, six were LPV-associated resistance mutations and 14 were WT/polymorphic AA residues. The BNL revealed two probable pathways to LPV/r resistance. The first pathway involved a direct association between treatment and 54V as parent. Through 54V, the pathway revealed direct associations between 10F and 82A+46I, with the combination 46I+54V+82A commonly seen in LPV failure. A direct inverse connection between resistance mutation 76V and

WT L24 as well as an indirect inverse connection between 82A and WT Q58 connects 76V to triple combination 46I+54V+82A. The second pathway connected 82A with LPV/r via the subtype C signature mutation L93 as child.

3.3.4 Most probable network for Gag-PR

The most probable Gag-PR network comprised 35 variables including the LPV/r treatment node (Figure 3.3B). Together with the six PRMs described in Figure 1A, 10 variables from Gag was included in the final dataset. Interestingly, instead of altering the interactions observed in PR (Figure 3.3A), the Gag mutations added to the existing network and mostly correlated with other Gag mutations. In terms of resistance pathways, a direct association was observed between LPV and 54V via P63 (through an inverse association with 63L). While 54V maintained its association with 46I+82A, Gag mutation 431V correlated with this combination. Moreover, a strong direct association linked 10F with 431V. Unlike its dependency on Q58 and L24 in the PR network, L76V was also found associated with 46I, occurring after the 46I+54V+82A combination. Noteworthy was Gag polymorphic variable 69K which formed a parent arc with the LPV treatment node.

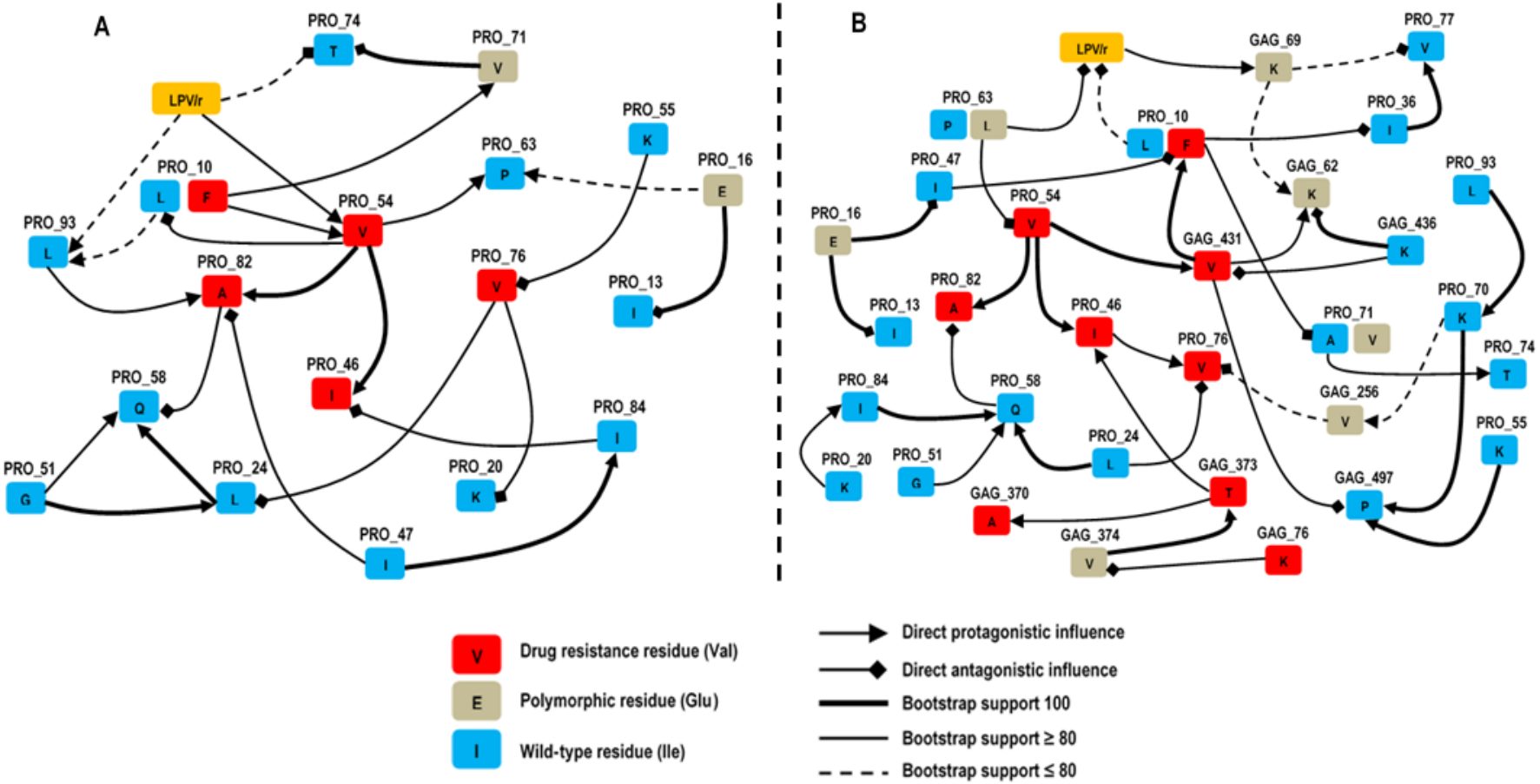


Figure 3.3 Most probable network showing associations between drug resistance mutations, polymorphisms, wild-type residues and Lopinavir (LPV/r) treatment in **(A)** protease and **(B)** Gag-protease. Arc thickness represents bootstrap support. **Note:** Arc direction does not represent the order of accumulation of mutations or causality but may indicate a multivariable effect in the network.

3.4 Discussion

Using a combination of phylogenetic and Bayesian statistical inferences, we identified several positively selected and coevolving Gag-PR sites under drug selection pressure and used this information to construct Bayesian networks to identify pathways to LPV/r resistance.

In the context of evolution, reconstructing ancestral phylogeny is crucial since phylogenetic trees can provide information on the evolutionary history of genes (Soltis and Soltis, 2003). Maximum-likelihood and Bayesian methods use probabilistic approaches to estimate tree topologies. Usually, statistical nodal support is calculated by random alignment resampling with replacement in a technique known as bootstrapping. Although acceptable, interpreting bootstrap values is challenging (Hassan et al., 2017). Instead of inferring the single most probable tree observed in ML, Bayesian methods produce a set of credible trees (0.95 CI[17]) that reflects the posterior probability topology given the data (Paraskevis et al., 2004; Hassan et al., 2017). In comparison to bootstrapping, Bayesian posterior probability has been suggested to be less biased in phylogenetic accuracy (Hassan et al., 2017). Therefore, in this study, both ML and BI methods were used to construct the phylogenetic trees. To obtain an effective sample size (ESS) >200 in BI, the Markov chain Monte Carlo (MCMC) length was set to 10,000,000. With only a 6.33% difference between the ML and BI topologies, the trees were considered acceptable for the positive selection and CAPS analyses as discussed below.

Positive selection is the acquisition of advantageous mutations. These signature mutational patterns contribute towards functional protein shifts (Qian et al., 2017) and has been widely studied using codon substitution models (Romero et al., 2016). In this study, a comparison of different site models revealed that models allowing for positive selection (M2a and M8) fit the data better than their counterparts (M1a and M7). Although most sites were under a strong purifying selection, several of the positively selected sites were mostly located in Gag. Similarly, CAPS revealed that most coevolving sites were found in Gag than PR. Taken together, this suggests that Gag can tolerate mutational changes better than PR. Interestingly, this theory was corroborated in a study conducted by Darapaneni et al. (2015) where 61% of the viral PR was conserved, particularly within the dimer interface, active site and flap regions. This suggests that AA changes are restricted to specified regions in PR to maintain the structural integrity of the enzyme during cleavage. Contrastingly, Gag displays high natural variability (Li et al., 2013), particularly at the CSs (Torrecilla et al., 2014). Prabu-Jeyabalan et al. (2002) demonstrated that

[17] Confidence Interval

substrate recognition by PR is determined by the asymmetric structure of the CSs and not necessarily the AA sequence. This theory, coined the "substrate envelope" hypothesis was based on studies involving inactive PR variants complexed with peptide substrates (Prabu-Jeyabalan et al., 2002; Tie et al., 2005). Although few, codons located at the CSs including E365 (CA|p2), S373 (p2|NC) A431 (NC|p1) and R452 (p1|p6) were identified in this study. Additionally, Özen et al. (2014) showed that the "conserved dynamic behaviour" of PR's active site, together with complimentary enzyme-substrate mutations increases substrate recognition. Considering the location of the PRMs identified in these sequences (Table 3.1), this suggests that PR and CS mutations work together to alter the substrate envelope of PR, increasing Gag proteolysis and providing a competitive advantage over the PIs.

To understand the resistance dynamics of specific mutations at various coevolving sites, we explored pathways to resistance in PR and observed the multivariable effect of what happens when Gag mutations are then introduced. In this study, the combination of M46I+I54V+V82A selected by PR is unsurprising since this is the three most common substitutions seen in LPV/r failure (Barber et al., 2012; Van Zyl et al., 2013; Grossman et al., 2014). As previously mentioned, in tandem with this combination, the introduction of L76V can also lead to DRV/r cross-resistance (Tang and Shafer, 2012). Although several studies have demonstrated the impact of genetic diversity on pathways to drug resistance (Sylla et al., 2008; Martinez-Cajas et al., 2009; Barber et al., 2012), the stepwise accumulation and multivariable interaction of mutations in pathways to resistance is largely unknown. In this study, two probable pathways in PR leading to LPV/r failure were described, one beginning with I54V and the other with V82A. Interestingly, Zhang et al. (2010) described a conditional structural independence between 46+54+82, where 46 and 54 are mutually independent given 82. Therefore, the authors posited that the sequential order of mutations would be either 46-82-54, 82-46-54 or 82-54-46. However, a separate study conducted by Zhang et al. (1997) described an order of 54-46-82 and 46-54-82. Although both studies reiterate these dependencies, our study suggests that the deciding factor as to whether I54V or V82A is selected first is dependent on L10F, where if I54V is selected first then L10F is also selected. However, if V82A is selected first followed by I54V then either the L10F mutation or the WT L10 could be favoured. Although more work would be required to confirm this theory, L10F could be functionally important. Previously, Martinez-Picado et al. (1999) showed that viruses harbouring the L10R mutation together with M48I+L63P+V82T+I84V replicated at the same rate as the WT, thereby suggesting its role in restoring viral fitness. However, the role of L10F when the M46I+I54V+V82A combination is selected together with L76V as seen in this study, is still unclear. Interestingly, Wong-Sam et al. (2018) showed that L76V reduces

hydrophobic contacts with I45/47 in PR's flaps and D30/T74 in PR's core, corresponding with protein instability. However, the authors postulated that reduced contacts caused by Val76 can allow for greater flap mobility. Furthermore, Louis et al. (2011) showed that the co-selection of M46I counteracts this instability by enhancing proteolytic cleavage in subtype B. This suggests that L76V occurs to modify flap dynamics to maintain cleavage with the altered Gag CSs whilst maintaining resistance to LPV/r.

Interestingly, the pathways to resistance changes significantly when other PIs are investigated. For example, Deforche et al. (2006) showed that though L10F, M46I and I54V were present in the network, only major PRMs D30N, N88S and L90M connected with the NFV node as parent. However, using a temporal nodes Bayesian network, Hernandez-Leal et al. (2000) showed that when IDV was present, a direct pathway existed between M46I, I54V and V82A through L90M as parent. Thus, the selection of specific mutations and intricate changes in the pathways to resistance are attributed to the drug itself. However, these networks may still change if Gag mutations are also introduced.

In the Gag-PR network a combination of L10F+M46I+I54V+V82A and A431V was observed. The A431V Gag mutation can cause resistance to all PIs except DRV (Fun et al., 2012) even in the absence of PRMs (Nijhuis et al., 2007). Moreover, its level of resistance is comparable to that of a single major PRM such as V82A (Nijhuis et al., 2009) and it can improve Gag cleavage regardless of the protease variant (Kolli et al., 2009). Although A431V has been associated with V82A (Prabu-Jeyabalan et al., 2004), a direct relationship was observed with I54V in this study. This could suggest that A431V and its link to V82A is dependent on the presence of I54V and M46I. This is consistent with the findings observed by Singh (2015) wherein A431V was mostly found in viruses with PRMs (M46I+I54V+V82A) than those without. This suggests that the if PRMs are selected, A431V serves to act in a structural compensatory role at a later stage.

Of note was non-CS mutation Q69K (chapter two; Figure 2.5). In this study, Q69K indirectly associated with A431V via G62K and V77 in PR. Singh (2015) reported that Q69K increased viral replication by 13% in the presence of A431V and M46I+I54V+V82A suggesting that it occurs to further compensate for changes associated with these mutations. Additionally, although CAPS identified the coevolving pair of 77 and 431, an indirect inverse association between these two codons was noted in the network. Interestingly, Gupta et al. (2015) demonstrated that a PR double mutant (V77I+L33F) resulted in a more stable structure with altered flap dynamics in the presence of Nelfinavir (NFV) in comparison to the WT. This could suggest that while A431V

works to restore cleavage, V77I works to stabilize the mutant PR by changing the open/closed conformation of the flaps to compensate for the modified CS. However, more work is required to elucidate the relationship between A431V+V77I in the presence of LPV/r.

3.5 Conclusions

Since mutations share unequal importance in drug resistance outcomes, the complexity of mutational patterns is challenging to interpret for treatment purposes. This study showed that strong dependencies exist between the evolutionary selection of PRMs and those identified in Gag. In particular, the combination of L10F+M46I+I54V+L76V+V82A together with A431V in Gag was shown to follow a specific pathway in patients receiving a LPV/r-inclusive treatment regimen. However, to fully understand the resistance dynamics between Gag and PR, a true understanding regarding the effects of PR's mutational patterns in the presence of PI drugs without Gag is first required. Therefore, the next chapter investigated the structural implications of two multi-drug resistant protease sequences in the presence of LPV as well as DRV (since DRV is administered in 3rd-line ARV therapy).

Chapter four

Examining DRV and LPV binding in the presence of multi-drug resistance mutational patterns in HIV-1

4.1 Introduction

As described in chapter one of the literature review (section 1.9.4), the HIV-1 PIs were developed to mimic the substrate transition state to inhibit Gag cleavage during maturation (Lv et al., 2015). However, PR's ability to acquire multiple resistance mutations at various positions can cause PI cross-resistance (Rhee et al., 2010). This phenomenon has led to suboptimal binding in several on-market drugs. Amino acid variation at positions 10, 46, 54, 82, 84 and 90 comprise some of the most commonly associated drug resistance mutations in the viral PR (Wensing et al., 2017).

In our study, BNL (section 3.3.3 of chapter three) found associations between mutations L10F, M46I, I54V, L76V and V82A. These mutations (L10F, M46I, I54V, V82A) have serious implications on LPV (Boender et al., 2016; Paredes et al., 2017) and DRV drug binding, particularly if L76V is simultaneously co-selected, as it can enhance the level of drug resistance (Young et al., 2010; Louis et al., 2011; Boender et al., 2016). Although phenotypic and genotypic tests can be employed to evaluate resistance mutations, these assays only provide information on the experimental or functional aspects of the virus (Hu et al., 2011). Consequently, information evaluating the structural impact of such a complex combination of mutations are not clearly understood.

Furthermore, since it has been shown that various mutations have distinct effects in subtype C versus subtype B (Kiguoya et al., 2017), and that certain AAs are selected more frequently than others, as shown in Chapter two of this study, the question arises as to why and how are mutations selected?

Therefore, the purpose of this chapter was to investigate the structural implications of a complex combination of PRMs on LPV and DRV binding in HIV-1 subtype C using MD simulations and binding-free energy scores.

4.2 Methods

4.2.1 Multi-drug resistant PR sequences

Two PR sequences comprising multi-drug resistant (MDR) phenotypes were retrieved from the PCS cohort. These sequences have specific resistance profiles and was selected based on the information obtained from the Bayesian networks in Chapter three as described in Table 4.1 below.

Table 4.1 Sequence identities with their specific resistance profiles.

Sequence Identity	Comment	Protease resistance profile
Wild-type	AY772699	None
PCS124	MDR; four PI mutations	L10F, M46I, I54V, V82A
PCS069	MDR; five PI mutations	L10F, M46I, I54V, L76V, V82A

Note: PCS124 (accession number: MK446206) and PCS069 (accession number: MK446189).

Additionally, both PCS069 and PCS124 was also selected because they possess the A431V Gag CS mutation which was shown to be important in the LPV resistance pathways in chapter three. However, for the purpose of this chapter, only the PR sequences together with the drugs (i.e. LPV and DRV) will be structurally explored. A graphical representation of the methods used to analyze the structural implications of the PI resistance mutations on drug binding is indicated in Figure 4.1 below.

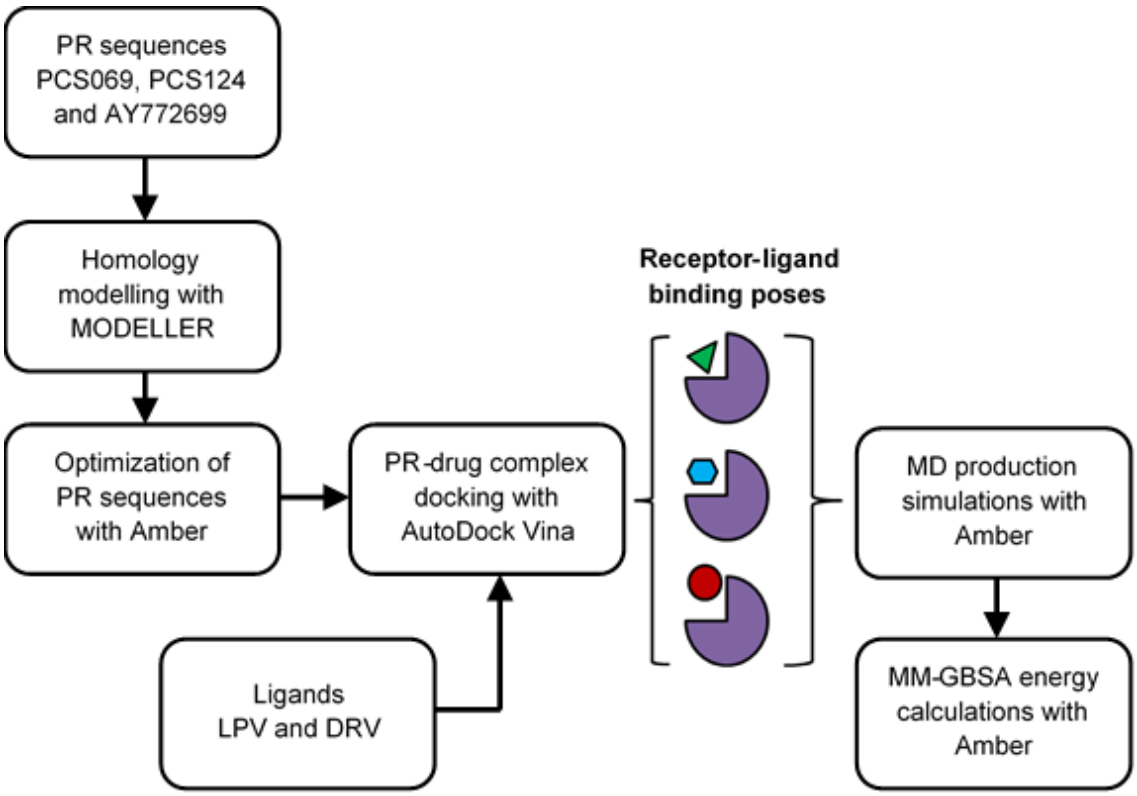


Figure 4.1 Illustration showing the sequential steps used to investigate the structural implications of resistance mutations on drug binding.

4.2.2 Homology modelling

Homology modelling of the viral PR was conducted in MODELLER v.9.16 (Webb and Sali, 2014). To construct the tertiary structures a suitable template was first selected. Accordingly, the atomic coordinates for an HIV-1 subtype C PR bound to NFV (PDB ID: 2R5Q) was retrieved from the PDB (Coman et al., 2008). The 2R5Q PR was resolved by X-ray diffraction to 2.3 Å and did not contain any PI drug resistance mutations. Prior to modelling, the ligand (NFV) was removed and the template and target sequences (viral PR) were aligned in MAFFT (Katoh and Stadley, 2013).

4.2.3 MD optimization of the PR models

The PR sequences were optimized using the MD package Amber 14 (Case, 2014). Topology and coordinate files were generated using the LeaP program. The ff03.r1 Amber force field for proteins was applied to the system. The system was explicitly solvated by the TIP3P water molecules with a margin of 12.0 Å. The SHAKE algorithm was used to restrain the covalent hydrogen bonds (Ryckaert et al. 1977) and the particle mesh ewald (PME) method with a 10.0 Å cutoff was used to calculate the Coulomb (electrostatic) interactions in the system (Darden et al. 1993). Prior to equilibration, a two-step minimization was performed as follows: (i) minimization of the waters with 3,000 steps of steepest-descent and 2,000 steps of conjugated gradient and (ii) minimization of the entire system with 7,000 steps of steepest-descent and 3,000 steps of conjugated gradient minimizations. Following minimization, the entire system was gradually heated to from 0 to 300 K over 50 picoseconds (ps) using the constant volume and normal temperature (NVT) mechanics. Thereafter, the system switched to the isothermal isobaric (constant pressure and normal temperature; NPT) mechanics for a further 50 ps. Finally, five nanoseconds (ns) of MD simulations were carried out. Additionally, the temperature of the system was monitored using the Langevin dynamics thermostat. Additionally, when force constraints were applied, the atomic residues were restrained at 2.0 kcal/mol.

4.2.4 Molecular docking

The optimized PR structures and ligands (DRV and LPV) were docked to generate the receptor-ligand complex and obtain the various binding poses. Molecular docking was carried out using AutoDock Vina (Trott and Olson, 2010). The binding poses with the lowest binding scores were selected for the final MD production simulations.

4.2.5 MD production simulations

Final MD production simulations were performed on the PR-drug complexes as described in section 4.2.3 with some modifications. Briefly, the ligands were parametrized using the AM1-BCC charges in Amber's Antechamber program. The General Amber force field (GAFF) was applied to the receptor-ligand system. Lastly, final simulations were carried out over 100 ns (10,000,000 cycles). Snapshots were taken every 5,000th generation to ensure that an ensemble of uncorrelated frames were obtained.

4.2.6 MM-GBSA calculations

The binding-free energies of the receptor-ligand complexes were calculated using the MM-GBSA method (Kollman et al., 2000) in Amber. Free-energies were calculated over 2,000 snapshots from the last 80–100 ns production trajectories. The energy calculations were used to compare the level of drug binding between the MDR and WT PRs. Therefore, entropy calculations were not included in our analyses as comparisons were made between the same proteins (i.e. the viral PR).

4.3 Results

4.3.1 Prediction of theoretical tertiary structures

The 2R5Q subtype C PR was shown to have a high sequence similarity to the consensus C (calculated using all the subtype C treatment exposed PR sequences with the Los Alamos Consensus Maker Tool) and the AY772699 reference sequence used in this study as shown in Figure 4.2 below. With a 92% sequence similarity between the AY772699 reference sequence and 2R5Q, the template was considered acceptable for molecular modelling of the MDR PR sequences. Additionally, the template had no gaps and therefore did not require any gap penalties or manual adjustments to be made.

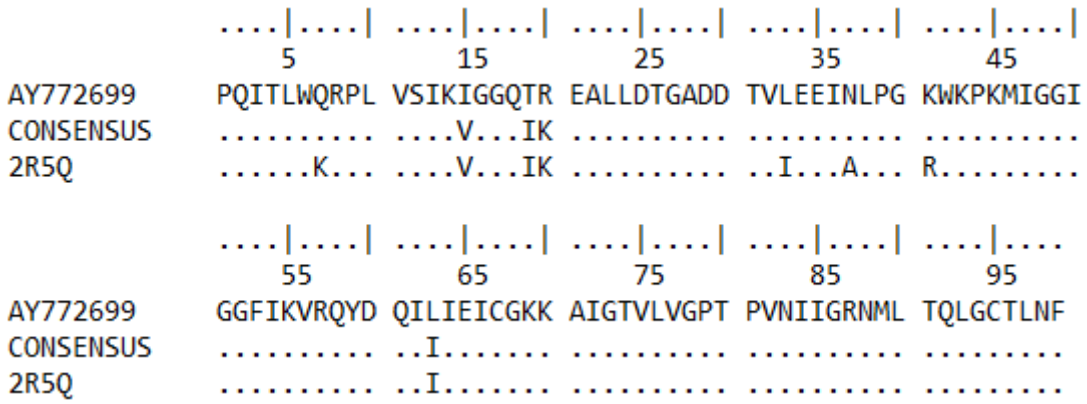


Figure 4.2 Codon alignment showing sequence similarity between the AY772699 reference sequence, consensus C and the 2R5Q template used for the prediction of the HIV-1 viral PR.

4.3.2 Assessing model refinement and the optimal binding poses

To assess the MD properties and level of equilibration, system energies and structural stability of the PR models were evaluated, as depicted in Figure 4.3 below. Accordingly, Figure 4.3a–c demonstrated that the kinetic, potential and total energies gradually increased, corresponding to the 50 ps heating phase from 0 to 300 K. Of note, the kinetic energy remained constant thus indicating that the thermostat, which monitors kinetic energy to assess temperature through the system was functional. Additionally, the potential and total energy initially increased during constant volume mechanics (NVT) and then decreased when the system switched to constant pressure (NPT) before stabilizing indicating equilibrium had been reached.

To evaluate the stability of the structures, a root mean square deviation (RMSD) calculation was performed for each PR model as shown in Figure 4.3d–f. Of the three models, the WT (Figure 4.3d) appeared the most stable with a relatively lower RMSD than the PCS124 (Figure 4.3e) and PCS069 (Figure 4.3f) models. Interestingly, the PCS124 PR model depicted greater stability over the five ns simulation in comparison to PCS069.

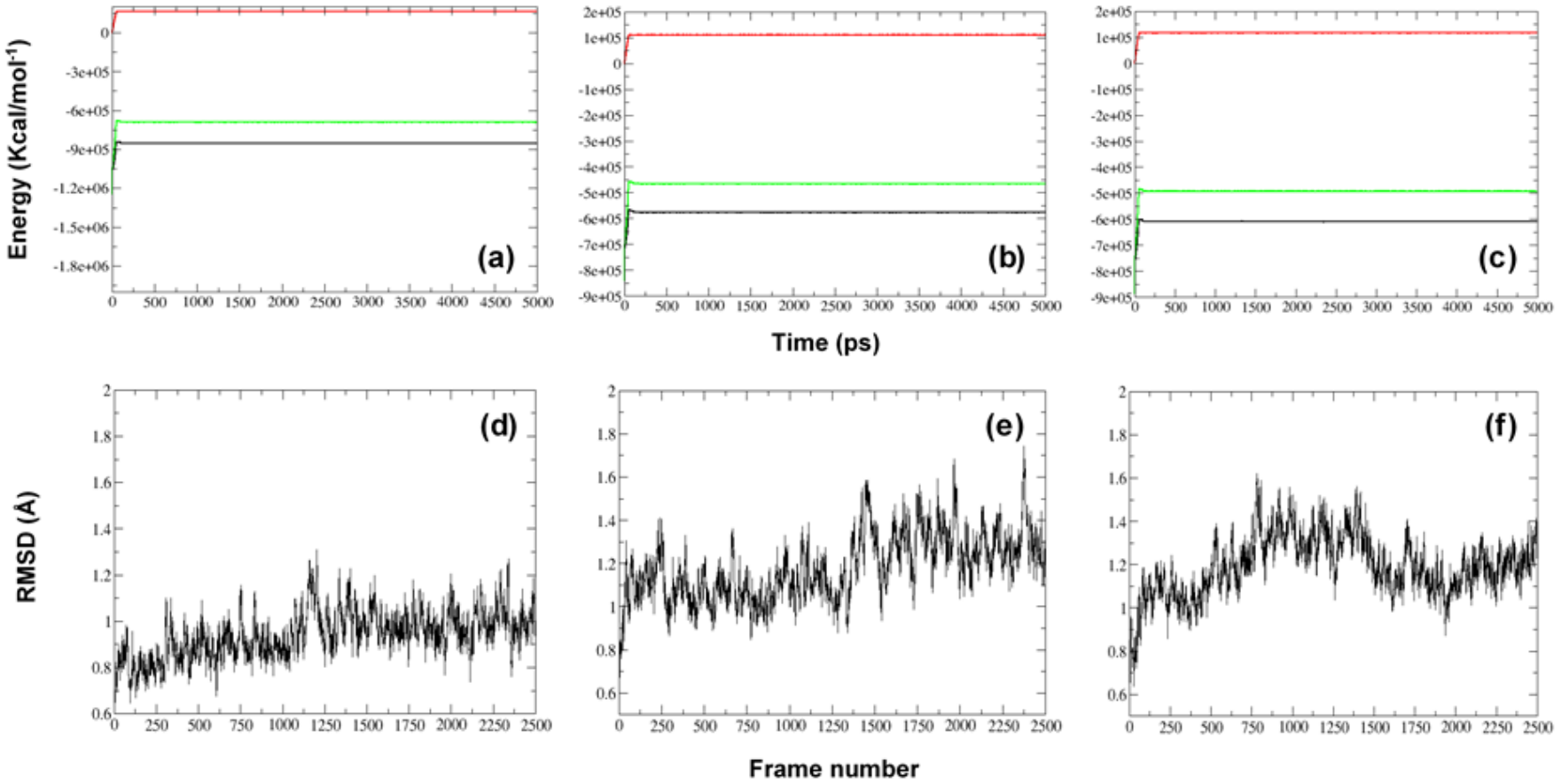


Figure 4.3 System and structural assessment of the HIV-1 WT and mutant PR models. Figure 4.3 (a)–(c) and (d)–(f) represents the energy terms and the RMSD for the wild-type (AY772699) and mutant protease (PCS124 and PCS069), respectively.

To generate the PR-drug complexes for the MD production simulations, we performed molecular docking analyses with AutoDock Vina. The search space evaluated eight different binding poses for each PR-drug complex with varying binding scores. Therefore, the models with the lowest binding scores were selected for downstream analyses. The binding scores for all models ranged from -7.4 to -9.8 kcal/mol. Furthermore, the ligands were flexibly docked and allowed for torsions and/or rotatable bonds to be considered. Consequently, the drugs were "neatly" docked into PR's substrate cavity and varied through the WT and mutants as seen in Figure 4.4 below.

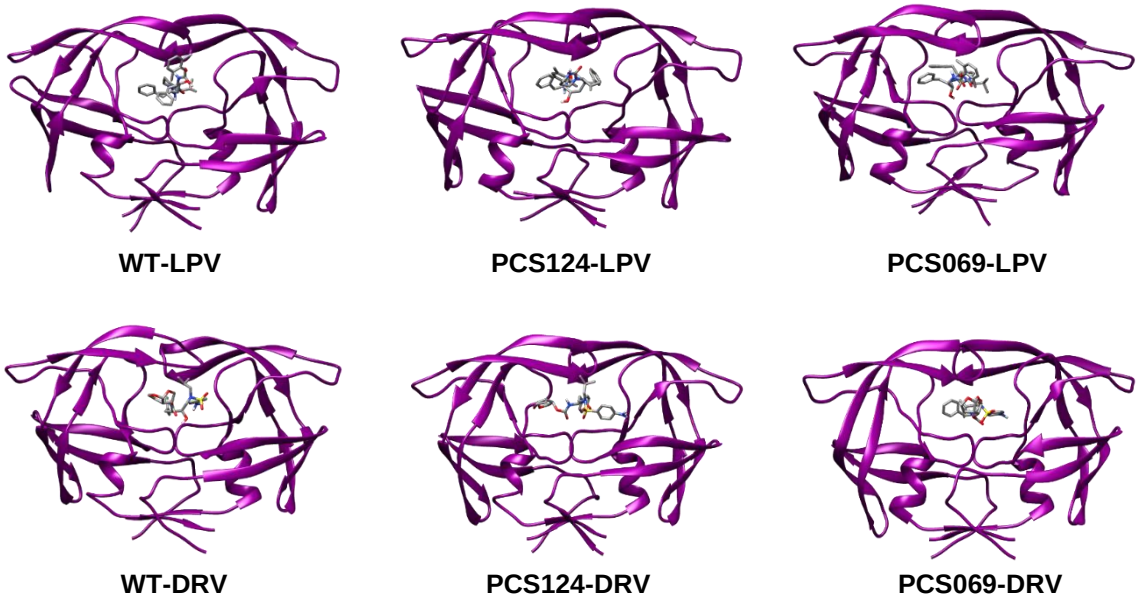


Figure 4.4 Binding poses generated for the HIV-1 LPV- and DRV-PR bound complexes over the 100 ns MD simulation.

4.3.3 Evaluating PR's stability

To assess the structural stability of the PR-drug systems, RMSD plots were generated over the 100 ns simulation as depicted in Figure 4.5. Accordingly, the LPV complexes (Figure 4.5a) were most stable in comparison to the DRV-bound PR mutants (Figure 4.5b). However, when comparing the WT and mutants in both the LPV and DRV models, the WT complexes (black) displayed greater structural stability over the 100 ns (Figure 4.5a–b). Interestingly, between the LPV-bound mutants, the PCS124-LPV complex (red) displayed the closet structural stability to the WT whilst the PCS069-LPV mutant (green) only stabilized around 65 ns. Contrastingly, the DRV-bound PCS124 mutant only reached a relatively stable equilibrium around 80–100 ns. In the DRV-bound PCS069 mutant a plateau was observed between 35 and 55 ns as well as in the last 90–100 ns of the simulation.

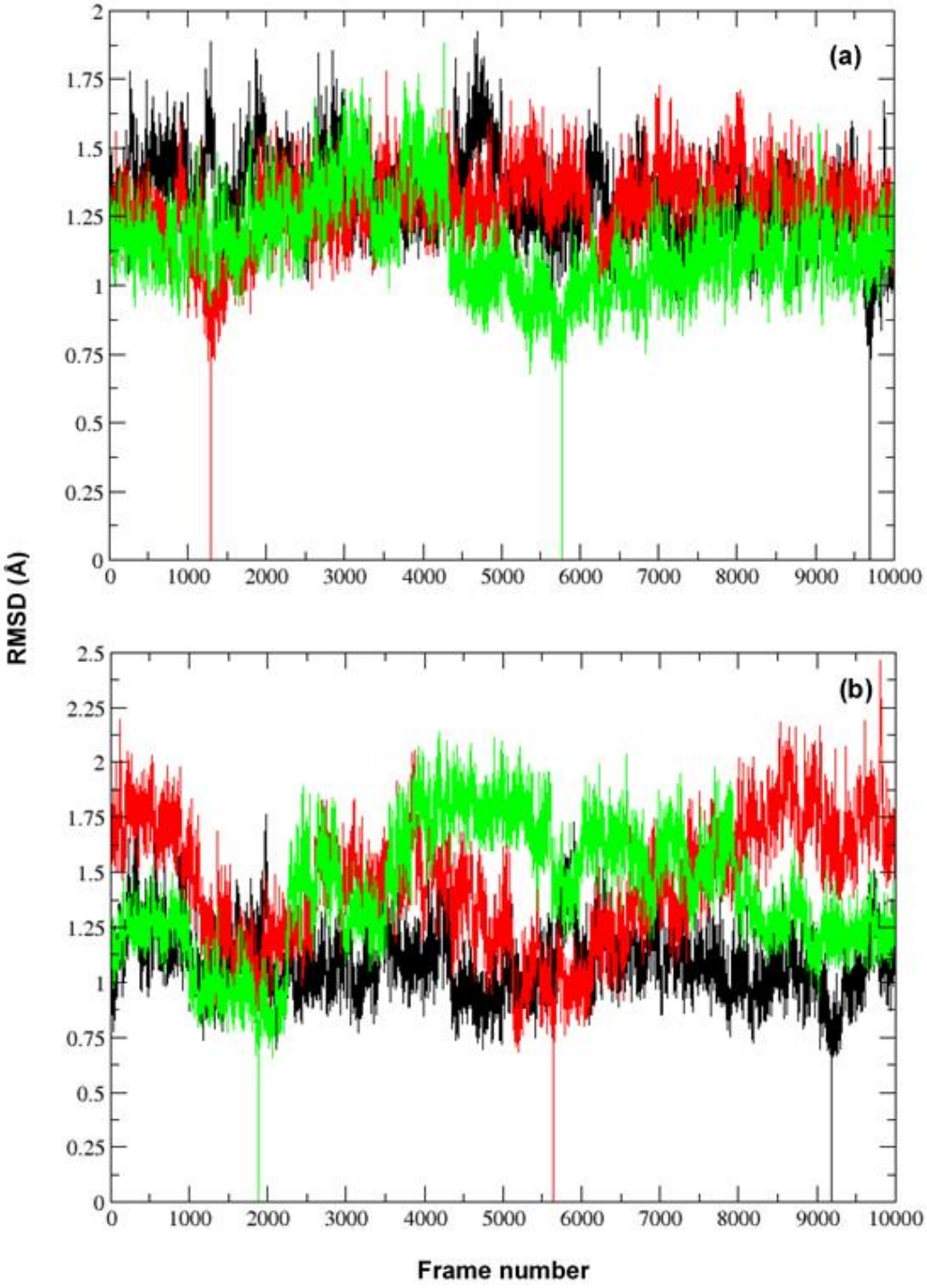


Figure 4.5 Time series RMSD plots of the (a) LPV- and (b) DRV-bound complexes. **Note:** the HIV-1 WT, PCS124 and PCS069 mutants are represented by the black, red and green lines, respectively.

Interestingly, the DRV- to LPV-bound fluctuations in the PR mutants was further reiterated by the flexibility of PR's AA residues, as observed by the root mean square fluctuation (RMSF) plot

in Figure 4.6 below. In particular, the mutant-DRV complexes displayed greater particle flexibility in comparison to the WT. Contrastingly, in the LPV-bound WT model, greater fluctuations were observed around the flap (Figure 4.6a; dotted circles) and elbow regions. Additionally, particle variation between the PR's subunits were also seen (Figure 4.6a and b; purple circles).

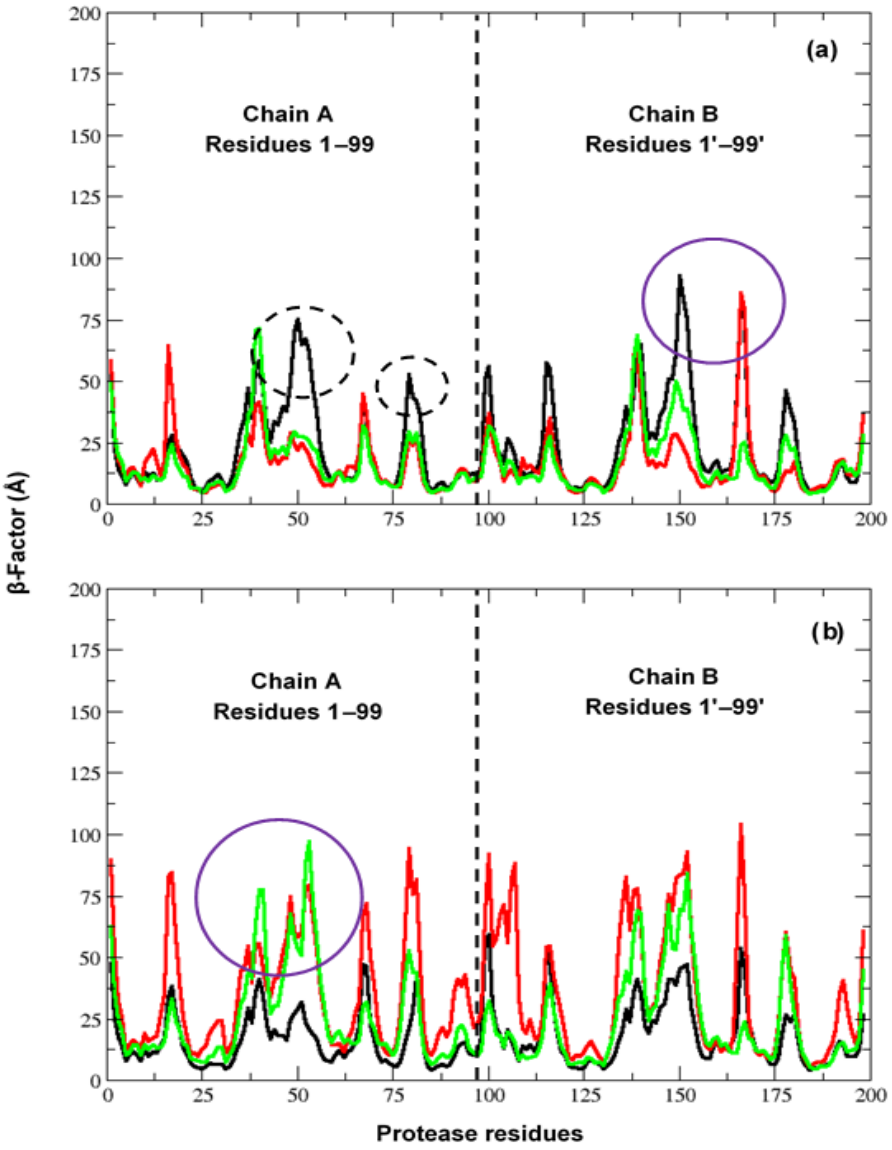


Figure 4.6 RMSF analysis of the (a) LPV- and (b) DRV-bound complexes. **Note:** the HIV-1 WT, PCS124 and PCS069 mutants are represented by the black, red and green lines, respectively.

In addition to residue fluctuations, the distance between the PRMs and active site residue D25 was altered as shown in Table 4.2. The LPV-bound PCS124 model displayed a greater distance with AA residues 46 (21.159 Å) and 82 (10.084 Å). Contrastingly, in the DRV-bound PCS124

model, residues 10 (8.745 Å), 46 (19.929 Å) and 82 (7.955 Å) were found at smaller distance when compared to the PCS069-DRV model (9.714 Å; 19.804 Å and 8.886 Å, respectively). Interestingly, when compared with the WT, the distance between the AAs in the WT and D25 was greater for most residues in both DRV-bound PR models. Contrastingly, only residues 54 and 82 in the LPV-bound PCS124 and PCS069 models, respectively displayed a smaller distance to the WT. Finally, the angles of the PRMs and thereby the rotation of the AA side chains between the models were also altered.

Table 4.2 Distance between active site residue D25 and the PRMs in the HIV-1 WT and mutant drug models (all measurements in Å).

Codons	Wild-type*		PCS124#		PCS069#, ǂ	
	LPV	DRV	LPV	DRV	LPV	DRV
10	8.047	11.154	9.164	8.745[a]	10.801	9.714[a]
46	18.222	18.633	21.159	19.929	19.228	19.804
54	17.789	15.477	16.930[b]	15.322[a]	18.578	16.217
76	11.903	12.996	-	-	11.925	11.759[a]
82	9.361	9.676	10.084	7.955[a]	8.884[b]	8.886[a]

***** Wild-type amino acids = L10, M46, I54, L76 and V82.
# Drug resistance amino acids = L10F, M46I, I54V, L76V and V82A.
ǂ L76V in PCS069 only.
[a] Wild-type distance is greater than the DRV mutants.
[b] Wild-type distance is greater than the LPV mutants.

4.3.4 Predicted binding-free energies for the PR-drug models

To evaluate the level of drug binding when bound to the mutant PR, the overall binding-free energies for each complex was calculated as depicted in Table 4.3 below. In general, the LPV-bound PR complexes displayed lower binding energies and thus a stronger structural stability in comparison to the DRV-bound models. Interestingly, the greatest binding stability was observed for the PCS069-LPV model (-23.4128 kcal/mol) where a -0.9826 kcal/mol difference was seen between the mutant and WT. However, when compared with the PCS124 model, a -3.4417 kcal/mol energy difference was observed.

Contrastingly, the WT-DRV model (-19.3166 kcal/mol) depicted the strongest binding mode in comparison to the DRV-bound PR mutants. Similarly, to the LPV-PR models, the PCS069-DRV

model (-17.9286 kcal/mol) generated a further energetically-favourable binding mode than the PCS124-DRV bound model (-15.8152 kcal/mol).

Table 4.3 Binding-free energies of the HIV-1 PR-drug complexes calculated over 2,000 snapshots in kcal/mol.

Complex	ΔE_{VDW}	ΔE_{EEL}	ΔG_{GB}	ΔG_{SURF}	ΔG_{gas}	ΔG_{solv}	ΔG_{bind}
Lopinavir							
WT-LPV	-29.5581	-10.4458	21.4441	-3.8704	-40.0039	17.5737	-22.4302
PCS124-LPV	-29.1449	-2.6835	15.6074	-3.7501	-31.8283	11.8573	-19.9711
PCS069-LPV	-28.7231	-18.2620	26.7036	-3.1313	-46.9851	23.5723	-23.4128
Darunavir							
WT-DRV	-26.3212	-12.6814	23.2752	-3.5891	-39.0026	19.6861	-19.3166
PCS124-DRV	-24.9921	-3.1009	15.2767	-2.9989	-28.0930	12.2778	-15.8152
PCS069-DRV	-27.9972	-9.4031	22.6892	-3.2175	-37.4003	19.4717	-17.9286

Note: ΔE_{VDW} = van der Waals contributions; ΔE_{EEL} = electrostatic contributions; ΔG_{GB} = polar contributions; ΔG_{SURF} = non-polar contributions; ΔG_{gas} = free energy in gas phase; ΔG_{solv} = solvation free energy; ΔG_{bind} = binding free energy.

4.3.5 PR-drug interactions

To explore the interactions between the inhibitor and PR, we constructed two-dimensional (2D) interaction maps for each drug complex as shown in Figures 4.7–4.12 below. Expectedly, the WT-LPV model revealed a hydrogen bond between catalytic residue D25 and LPV (Figure 4.7). Moreover, three additional hydrogen bonds were seen with residues A28, G48 and G49. Of note, van der Waals (vdW) interactions with V82/V82' (V181)[18] was observed. Additionally, three alkyl (V32, I47 and I84) and one π-alkyl (I50) interaction involving sites at which LPV-associated PRMs were noted.

[18] In nature, one chain in PR contains 1–99 AAs. Therefore, chain two starts from 100–198. In practice, we can also refer to the second chain AA numbers as prime, for example V82A in chain A and V82A' in chain B.

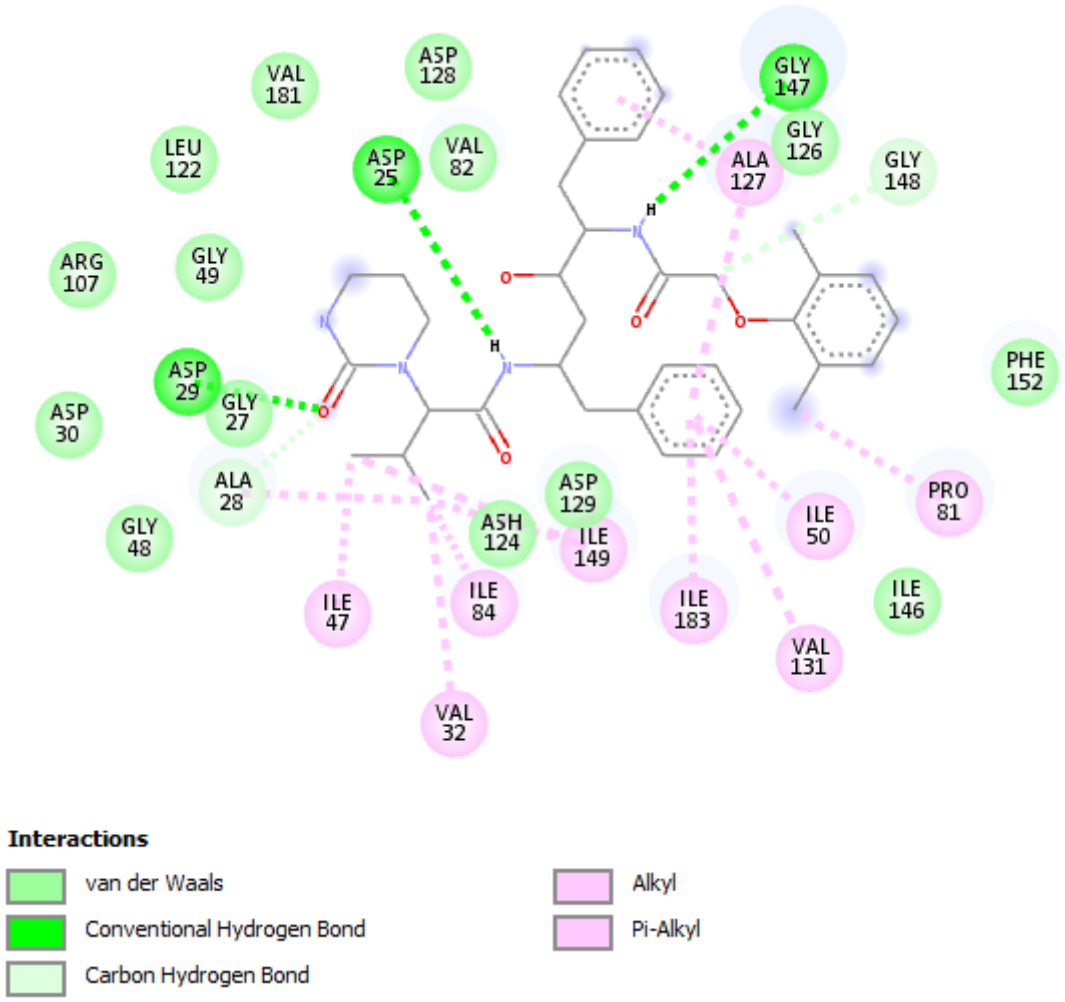


Figure 4.7 Amino acid interaction map of the HIV-1 WT complexed with LPV.

Figures 4.8 to 4.9 represent the PCS124-LPV and PCS069-LPV models, respectively. Interestingly, in the PCS124-LPV model (Figure 4.8), residues I50, G49 and catalytic triad AA G27 hydrogen bonded with LPV's oxygen atoms. Moreover, while V82A formed an alkyl bond with LPV, V82A' formed a π-alkyl bond with LPV's methylbenzene ring. In the PCS069-LPV model (Figure 4.9), conventional hydrogen bonds were seen with D25 and G48 whilst codon R8' (R107) formed a double hydrogen bond with two oxygen atoms in LPV. Notably, in addition to V82A/V82A', minor PRM L10F' (L109F) formed a π-alkyl interaction with the drug's methylbenzene. Interestingly, L10F in chain A was not involved in ligand binding.

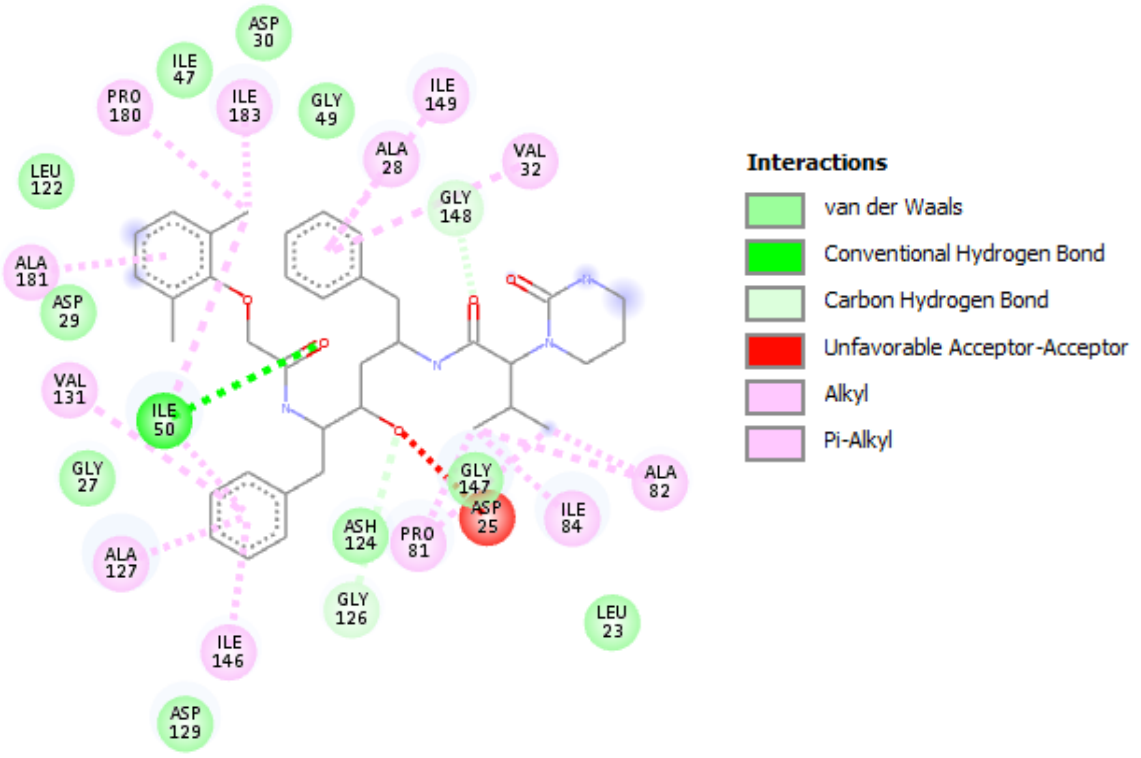


Figure 4.8 Amino acid interaction map of the HIV-1 isolate PCS124 complexed with LPV.

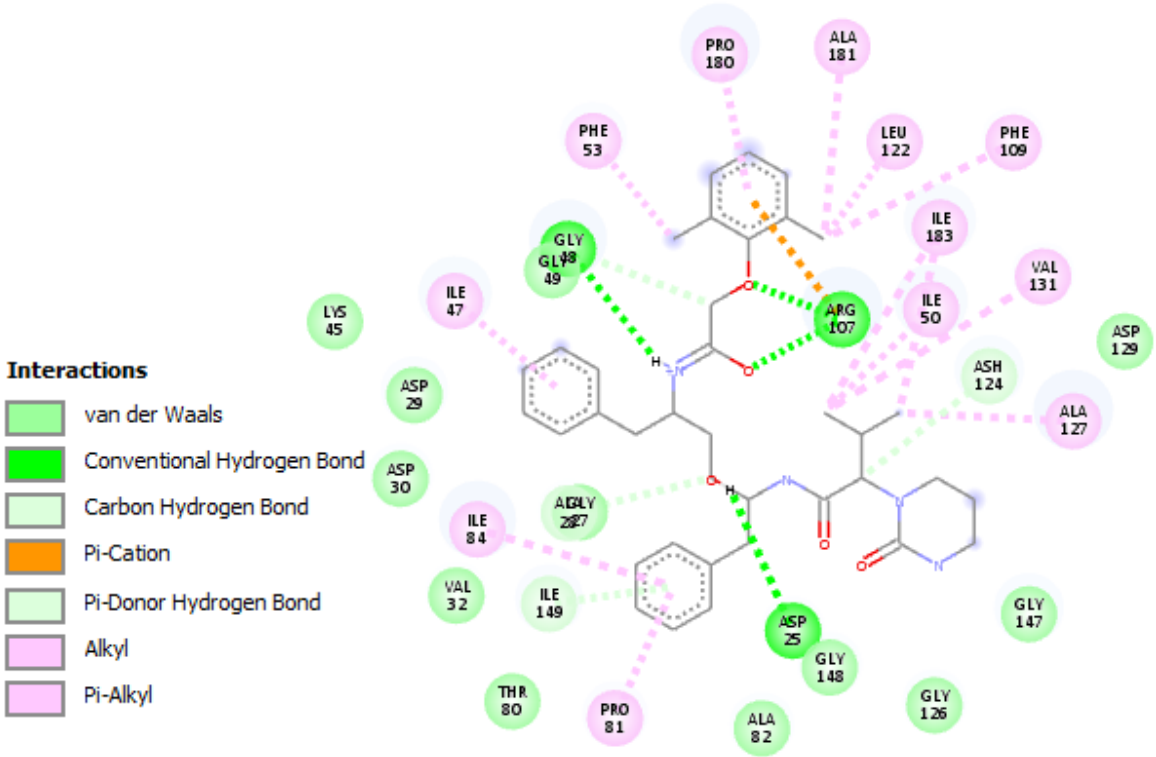


Figure 4.9 Amino acid interaction map of the HIV-1 isolate PCS069 complexed with LPV.

Unlike in the WT-LPV model, the D25 in the WT-DRV model (Figure 4.10) was attractively charged to a nitrogen atom in DRV. In addition, A28 and D29/30 formed hydrogen bonds with the 4-aminobenzenesulfonamide of DRV whilst G27' (G126) hydrogen bonded with the oxygen atom in the acetate group of the (3R,3aS,6aR)-Hexahydrofuro[2,3-*b*]furan-3-yl (bis-furyl) molecule. Similarly, to the WT-LPV model, resistance associated sites V32, I47, I50 V82 and I84 also interacted with DRV.

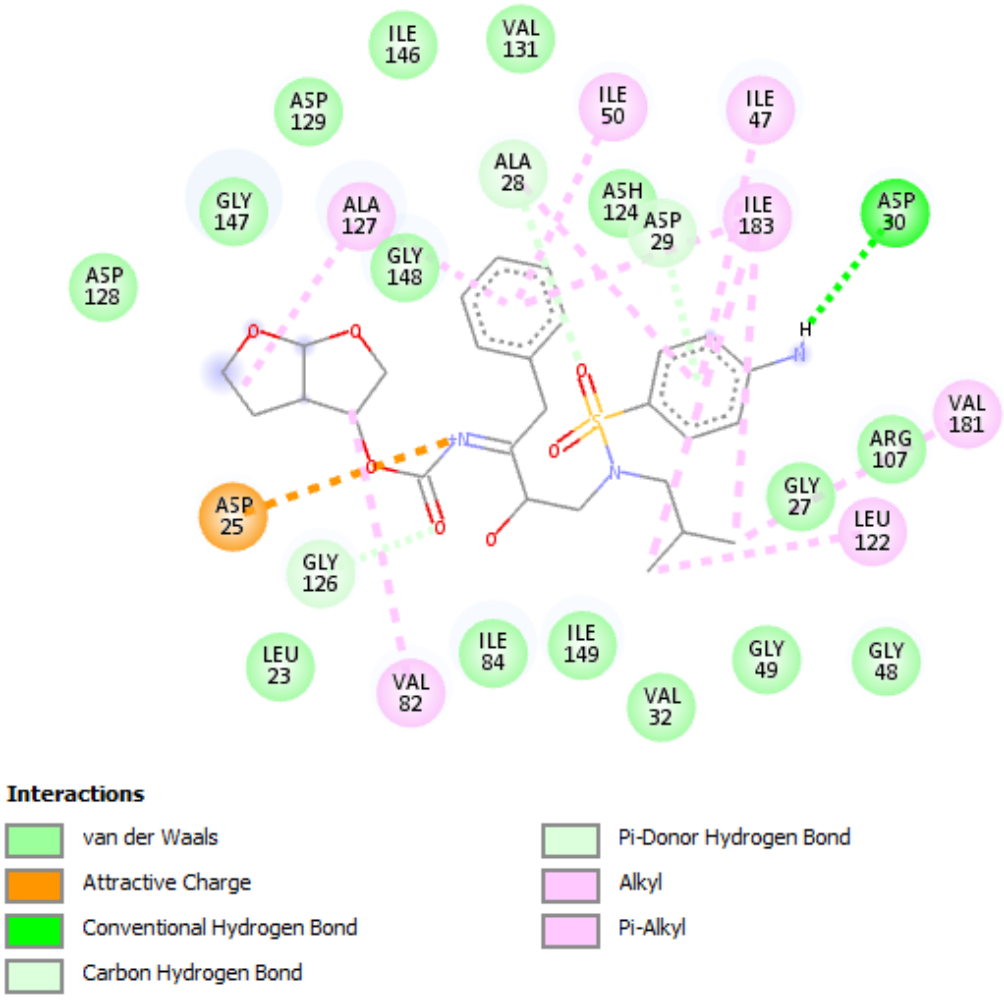


Figure 4.10 Amino acid interaction map of the HIV-1 WT complexed with DRV.

Interactions between DRV and the mutant PR models are depicted in Figures 4.11–4.12. Accordingly, in the PCS124-DRV model (Figure 4.11), R8 and G48' (G147) formed conventional hydrogen bonds with the 4-aminobenzenesulfonamide molecule and the nitrogen atom of the bis-furyl acetate. Furthermore, R8 also formed a π-cation interaction with the benzene ring of 4-aminobenzenesulfonamide. Residue I50 formed a π-donor hydrogen bond with DRV's aromatic

benzene. In comparison to the WT-DRV model, catalytic residue D25 lost its hydrogen bond and formed a vdW interaction with DRV instead. Moreover, only L10F and V82A formed vdW and π-alkyl interactions with DRV, respectively.

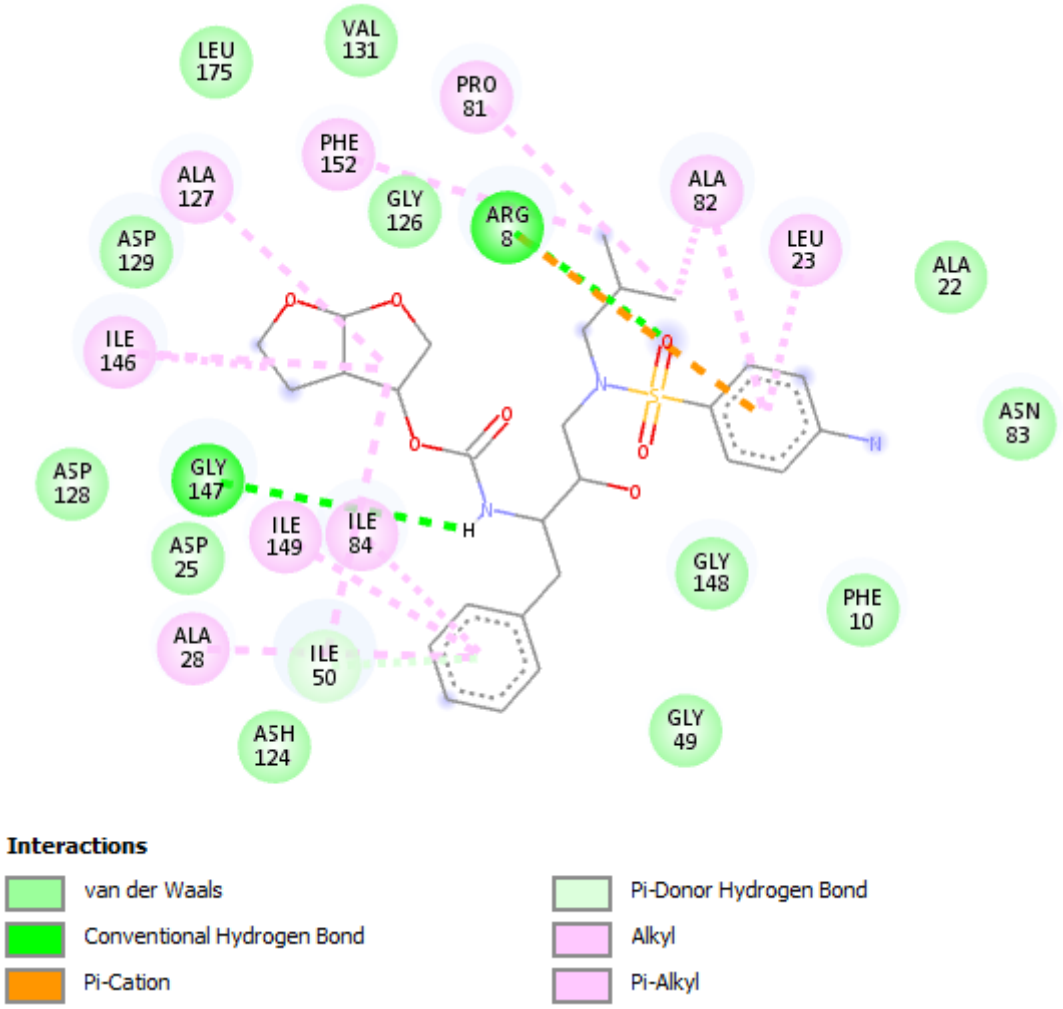


Figure 4.11 Amino acid interaction map of the HIV-1 isolate PCS124 complexed with DRV.

The PCS069-DRV model (Figure 4.12) revealed carbon hydrogen bonds with residues D30' (D129) and G48' (G147) as well as the heterocyclic furan of DRV's bis-furyl acetate. In a similar manner to the PCS124-DRV model (Figure 4.11), D25 only retained vdW interactions with DRV. Interestingly, the PCS069-DRV model was the only complex to have lost its interactions with codon 82 in PR. Contrastingly, it was also the only model to gain a vdW interaction with I54V.

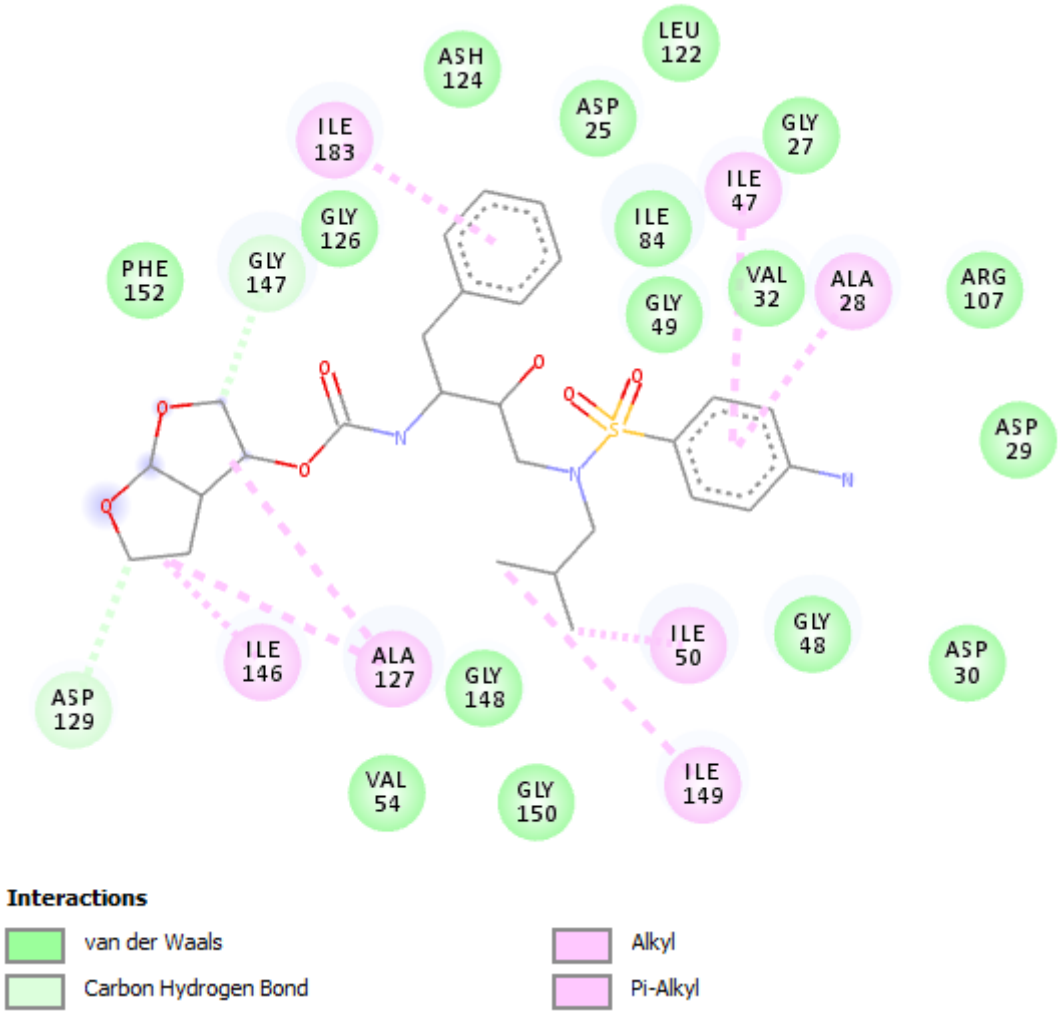


Figure 4.12 Amino acid interaction map of the HIV-1 isolate PCS069 complexed with DRV.

4.4 Discussion

Biological function is largely dependent on molecular dynamics of protein systems. Therefore, drug resistance mutations can alter the mechanism by which the HIV-1 PR cleaves in nature (Cai et al., 2014). Here, we predicted three molecular models corresponding to two PR mutants (PCS124 and PCS069) comprising major PRMs and a structure resembling the subtype C WT PR (AY772699). Homology modelling is an important tool that has been used to investigate molecular structural dynamics when experimental data is not available (Yuan and Xu, 2018) and several studies have used this approach to explore the dynamical behaviour of the HIV-1 PR (Lockhat et al., 2016; Amamuddy et al., 2018; Delino et al., 2018; Nayak et al., 2019).

Prior to docking experiments, the PR structures were refined to produce a favourable energy state. Therefore, we used explicit MD simulations to mimic the physiological state observed in nature. In this study the level of protein stability between the models appeared to be WT > PCS124 > PCS069. This is unsurprising since in the absence of drug therapy, the WT is functionally fit (Quiñones-Mateus and Arts, 2003; Hu and Kuritzkes, 2014). This suggests that the WT PR represents a "simpler", thermodynamically stable form of PR in comparison the mutants when PIs are not present. Thereafter, AutoDock Vina generated favourable docking poses that were within the confines of PR's substrate envelope. Additionally, subtle differences in the orientation of the ligands may be due to the flexible docking carried out by AutoDock. Flexible docking algorithms are more accurate in predicting binding modes since it allows movement of the rotatable bonds in the ligand (Pagadala et al., 2017). Therefore, as multiple AA side chains are affected by the movement of the backbone atoms of the protein (Pagadala et al., 2017), it is conceivable that the ligand should also be flexible enough to accommodate this movement allowing for greater binding affinity and more precise binding.

To evaluate the structural changes induced by the PRMs on drug binding, we performed 100 ns MD production simulations on the LPV- and DRV-bound PR models. Accordingly, the RMSD plots revealed that the WT models were significantly more stable than either of the drug-bound mutants. Proteins are dynamically flexible macromolecules that can exhibit large structural changes over of time (Haspel et al., 2010). Consequently, the PCS124/069 mutants may have folded into multiple alternative structures to continually find the optimal energy minima to evade drug binding while remaining active for Gag cleavage. In contrast, this suggests that the WT easily folded and maintained a stable native state since it did not comprise any mutational changes in the active site or flap regions and was thereby susceptible to drug binding in comparison to the mutants. Of interest was the discrepancy between the stability of the LPV- versus DRV-bound mutants. Here, the RMSD plots revealed greater stability with LPV rather than DRV. A study conducted by Dierynck et al. (2007) showed that MDR PRs comprising 10–14 PI resistance mutations resulted in reduced binding affinity coupled with limited antiviral activity for LPV. Contrastingly, for DRV, the study showed that weaker antiviral activity was not observed even with a 1,000-fold decrease in binding affinity. Consequently, this suggests that the LPV-associated PI mutations highlighted in this study displays greater structural evasiveness to LPV than DRV. Furthermore, the Dierynck study observed that the number of DRV-associated mutations also correlated with the level of drug binding. Therefore, this theory may also translate to the number of mutations selected by the mutants as observed with RMSD fluctuations between the quatro-PCS124 and penta-PCS069 PR models.

Furthermore, a link between the RMSD and RMSF plots regarding the DRV-bound mutants were also observed. In nature, active proteins can rapidly manipulate their local fluctuations, adopt a specific conformation and transition between conformations (Kmiecik et al., 2018). In this instance, the DRV-bound mutants showed larger fluctuations in PR's AA residues as compared to LPV. In the WT models, fluctuations at the flaps and elbow suggests that particle motion at these regions occurred to allow for the open/closed conformation of PR and to accommodate movement of the ligand and thereby binding through the 100 ns simulation. Generally, flap flexibility is restricted when PR is bound to an inhibitor (Trylska et al., 2007). However, Piana et al. (2002) showed that residues 24–30 and 45–55 in the viral PR correlated with substrate motion. Moreover, it has become acceptable that interaction of PR's flaps with a substrate influences the interval and frequency of flap opening (Trylska et al., 2007; Goldfarb et al., 2015; Yu et al., 2017). Interestingly, the RMSF plots also revealed a level of asymmetry between the two PR chains. This phenomenon has been previously observed in several studies evaluating the structural conformation of the HIV-1 PR (Lexa and Carlson, 2011; Huang et al., 2017; Paulsen et al., 2017). Cai et al. (2014) postulated that the asymmetrical ligand prompts asymmetry in the otherwise symmetrical PR. Thus, the two chains could behave differently in the mutant PRs to the evade PI drugs. Moreover, differences in the interatomic distances between catalytic residue D25 and the mutant/WT models reiterate a mechanism for drug resistance. Specifically, reduced interatomic distance between D25 and the PCS069 model was consistent where the inclusion of L76V promoted reduced hydrophobic contacts and greater flexibility as an alternative mechanism for drug resistance (Wong-Sam et al., 2018). Contrastingly, the WT AA residues were distally located to D25. This discrepancy could be linked to the sequential accumulation of mutations in the PCS124/069 mutants and the positioning of AA side chains. For instance, as described in chapter three, if L76V is selected then there is a high probability of the co-selection of M46I to counteract the changes induced by the substitution of leucine to valine (Louis et al., 2011). Additionally, since torsion angles phi and psi are important factors in controlling protein folding (Saravanan and Selvaraj, 2017), suggesting that differences in the interatomic distance with the altered torsion angles occurred to maintain conformational flexibility of PR when the PRMs are present.

Thereafter, we performed binding-free energy calculations to assess the level of drug binding. Considering the above observations, it was expected that DRV would bind to PR at a lesser extent that LPV. More specifically, with only a -2.4447 kcal/mol average energy difference between the two mutants relative to the WT, the PRMs did not have a significant effect on drug binding. Previous studies have shown that viral PRs with either 19 (Agniswamy et al., 2012) and 21 (Louis et al., 2011; Louis et al., 2013) mutations were required for a major reduction in the loss of

hydrophobic contacts to occur. However, this study showed that the PCS069 model portrayed lower binding energies in comparison to the PCS124 models. This infers that resistance comprising the L76V pattern could be due to specific AA interaction with the drug or it is substrate related.

Finally, to explore the detailed PR-drug interactions, several interaction maps were constructed to observe specific residues in PR involved in ligand binding. These analyses revealed that residues R8, D25, A28, G27, G48, G49 and I50 in the WT and mutants formed important hydrogen bonds with LPV. In the HIV-1 PR aspartic residue 25 deprotonates a water molecule to produce a tetrahedral intermediate while the second aspartic (D25') facilities Gag cleavage (Chaudhury and Gray, 2009). Consequently, due to its catalytic activity mutational changes from D25 to A, Y, H or N eliminates PR's enzymatic activity (Huang and Chen, 2013). Thus, a strong hydrogen bond linking D25 to a ligand in the WT PR is expected. Outwardly, the unfavourable interaction observed between D25 and LPV in the PCS124 mutant (Figure 4.8) suggests that PR has "recognized its Trojan horse" and structurally conformed to directly inhibit enzymatic catalysis and drug binding. Apart from D25, it has been shown that residues R8 and G27 at the base of the active site and G49 at the flap tips stabilize the PR dimer (Weber, 1990; Wlodawer and Gustchina, 2000) while A28 and I50 are important AA residues for substrate recognition (Scott et al., 2000). Of note, Ceccherini-Silberstein et al. (2004) showed that apart from G16, G48 and G73, the remaining PR glycine residues are highly conserved. This suggests that binding to the glycine residues causes subtle rearrangements in the PR structure that are also complimentary to the changes induced by the PRMs.

Interestingly, codons V82/V82' were found to have vdW interactions in the WT-LPV complex, whereas alkyl/π-alkyl interactions with LPV were observed in the PCS124 and PCS069 mutants. Other bond types including alkyl and π-alkyl bonds can improve the hydrophobic bonding between a receptor and ligand (Arthur and Uzairu, 2019). Therefore, the switch to alkyl/π-alkyl bonds from vdW interactions in the mutants suggests a stronger association with the mutant alanine at position 82 than valine. Weber and Agniswamy (2009) demonstrated that in the presence of SQV, IDV and DRV, the alanine substitution shifted residues 81 and 82 to compensate for the loss of interactions induced by the smaller side chain. In addition, a study conducted by Liu et al. (2008) revealed that in a PR comprising a single I54V mutation, significant conformational changes observed in the flap region were coupled with structural changes in the 80's loop. In a similar manner, Nakashima et al. (2016) revealed that double PR mutant I47V/I50V induced structural flexibility to the enzyme. Since M46I is found in the PR's flaps, it

is sufficient to assume that this mutation can also influence PR flexibility. Therefore, in this study, the stronger bond coupled with the shift in main chain residues suggests that partial enzymatic rigidity was induced to inhibit the drug whilst retaining its ability to cleave Gag. Furthermore, this bonding effect could also be correlated with L10F which formed a π-alkyl bond with LPV in the PCS069 mutant. Agniswamy et al. (2013) showed that loss of contact between PR and novel PIs were due the displacement of a water molecule by L10F. Furthermore, since only one L10F residue was involved in ligand binding, this asymmetry could aid in maintaining PR's flexibility by not restricting the enzyme with phenylalanine's large bulky aromatic rings in both chains.

Like the LPV models, specific AA residues including R8, G27, A28, D29/30, G48 and I50 were all involved in DRV-PR interactions. Interestingly, in a study evaluating the binding differences between DRV and its structural analog APV, Hou et al. (2009) found that stable interactions were observed between specific AAs (A29, D30, G48) and the bis-furyl and 4-aminobenzenesulfonamide moieties of DRV. This agreed with the findings observed in the current study (Figures 10–12). Furthermore, the bis-furyl (or bis-THF) moiety was designed to produce a network of hydrogen bonds with the main chain atoms of the viral PR (Agniswamy et al., 2015). Our analyses revealed more vdW interaction instead, particularly in the L76V-PCS069 mutant corresponding to the binding-free energies. Contrastingly, in addition to several hydrogen (A28, D29/D30, and G27) and alkyl/π-alkyl bonds (Figure 4.10), the WT-DRV model also revealed a strong attractive charge between D25 and DRV. This indicates that the level of drug binding in the WT is much stronger in comparison to the mutants which is also supported by our GBSA calculations. Further reiterating this point was the loss of strong bonds between DRV and D25 in both the PCS124 and PCS069 mutants as only vdW interactions were noted. Therefore, a loss of contact between PR and these chemical moieties (bis-furyl and 4-aminobenzenesulfonamide) is the main reason for DRV resistance (Raugi et al., 2016).

Finally, our study revealed a vdW interaction between I54V and the PCS069-DRV mutant (Figure 4.12). As vdW interactions are considered for short-ranges this suggests that the flaps were pulled down closer to the drug. This agreed with a study conducted by Agniswamy et al. (2019) where the flap tips curled downwards in a DRV-bound I54V PR mutant. Flap "curling" has been linked to the open/closed states of PR when DRV is present. Yu et al. (2015) demonstrated that flexible curling of the flap tips into PR induces an open state. This observation suggests that PR utilizes mutations in its flaps to induce a semi-open conformation coupled with a loss of contact to minimize atomic bonds and inhibit drug binding.

4.5 Conclusions

The native structure of proteins is complex where it can transition between several states to adapt to changes within the genome, or its environment (Gidalevitz et al., 2011). In this study we evaluated the structural implications caused by LPV and DRV resistance associated mutations L10F, M46I, I54V, L76V and V82A when complexed with the drugs. Our analyses indicate that these mutations cause significant structural flexibility to the viral PR particularly within the flap regions. Moreover, it was also seen that the viral PR was relatively more susceptible to LPV in comparison to DRV. Furthermore, since L76V contradicted itself by showing stronger drug binding when it was included, the impact of this mutation in this context was still unclear. Therefore, chapter five investigated the structural mechanism by which the mutant PR's cleave the NC|p1 Gag CS that comprised the A431V resistance mutation.

Chapter five

Structural impact of the HIV-1 NC|p1 A431V Gag cleavage site mutation on the PCS124 and PCS069 protease mutants

5.1 Introduction

As previously discussed in chapter two, under drug selection pressure pathogens can effectively and rapidly accumulate mutations in various drug targets (Hooper and Jacoby, 2015; Kennedy and Read, 2017; Nguyen et al., 2018). Whilst most drug resistance mutations are clustered close to the active site (King et al., 2004), the HIV-1 PR has been shown to acquire mutations in distal locations, such as the flexible flaps and substrate cleft to evade drug binding (Ragland et al., 2014; Flynn et al., 2017). Though these resistant PR's are functionally weaker, they are still able to cleave the substrate and thereby outcompete the inhibitor for hydrolysis (Kantor et al., 2002; Rhee et al., 2003).

Incidentally, mutations arising in other proteins, such as Gag are being frequently reported (Fehér et al., 2002; Dam et al., 2009; Pillay et al., 2014; Teto et al., 2017; Su et al., 2018; Williams et al., 2019). The NC|p1 A431V Gag mutation is of interest, since it is associated with patients failing PI therapy (Lambert-Niclot et al., 2008; Giandhari et al., 2015). Additionally, in agreement with other studies (Verheyen et al., 2006; Malet et al., 2007), our analyses revealed direct associations between major LPV PRMs and A431V (chapter three).

While Gag-PR coevolution has been reported (Doyon et al., 1996; Zhang et al., 1997; Kolli et al., 2006; Dam et al., 2009), the mechanism by which hydrolysis is restored and the molecular basis in which these Gag CS mutations are selected is unclear.

Therefore, continuing from chapter four, this chapter investigated the mechanism by which the A431V Gag CS mutation structurally interacts with the mutant PR to properly maintain enzymatic cleavage under drug selection pressure.

5.2 Methods

5.2.1 Modelling of the NC|p1 Gag CS viral sequences

To maintain consistency and gain insightful knowledge on the co-selection of mutations in Gag-PR, the PCS069 and PCS124 viral sequences utilized in Chapter four was carried forward. As previously discussed, these viral sequences co-selected the NC|p1 CS mutation, A431V. The template selected for homology modelling (PDB ID: 2FNS) was an HIV-1 WT NC|p1 substrate complexed with an inactive HIV-1 subtype B PR (Prabu-Jeyabalan et al., 2006). The sequence region spanned from residues R429 to G435 as shown in Figure 5.1 below.

While templates spanning the entire target region is advantageous, there are currently only two entries for the NC|p1 CS in the PDB[19]. Additionally, as this region is relatively smaller (10 AAs) and does not comprise any tricky loops in comparison to other proteins, the 2FNS template is considered acceptable for use.

	428	429	430	431	432	433	434	435	436	437
Subtype B	E	R	Q	A	N	F	L	G	K	I
AY772699	.	.	.	.	.	.	.	.	.	.
2FNS	-	.	.	.	.	.	.	.	-	-
PCS124	.	.	.	V	.	.	.	.	.	.
PCS069	.	.	.	V	.	.	.	.	.	.

Figure 5.1 Sequence alignment of the HIV-1 subtype B, AY772699 reference, the 2FNS template and the target sequences. **Note:** the PCS124 and PCS069 NC|p1 Gag cleavage sites only have the A431V resistance mutation.

Thereafter, sequence alignment and homology modelling were carried out as described in section 4.2.2 of chapter four.

5.2.2 Geometric optimization of the ligand

In comparison to the drugs (LPV and DRV), the Gag CSs are molecularly larger and somewhat bulkier due to the AA side chains. Furthermore, because these ligands were theoretically modelled, optimizing their tertiary structure is imperative. Therefore, to ensure that the Gag CSs are geometrically and stereo-chemically accurate, a geometric optimization was carried out. Briefly, the Gag CSs were optimized using a semi-empirical quantum mechanics (qm) simulation with the sqm program in AmberTools v.14 (Case, 2014). Simulations were optimized at the Hamiltonian level using the AM1 qm-theory. The number of self consistent field (SCF) iteration cycles required for the geometry to converge and stereo-chemical equilibration to be reached was 100. Thereafter, the Antechamber program was used to prepare the files for Leap.

5.2.3 Docking, MD simulations and binding-free energies

Once ligand optimization had been achieved, the docking, PR-Gag production simulations and calculation of binding-free energies were performed according to sections 4.2.4 to 4.2.6 of chapter four, respectively.

[19] http://www.rcsb.org/ : search criterion NC-p1

5.3 Results

5.3.1 Gag-PR conformational stability

The time series RMSD for the PCS124-A431V and PCS069-A431V models are depicted in Figure 5.2. As observed, the PCS124-A431V model maintained a relatively stable complex from 50–100 ns whilst the PCS069-A431V model showed greatest structural equilibrium at approximately 48–77 ns and then again from 88–100 ns. Although definitive equilibrium was reached at specific intervals, the overall A431V-bound PCS124 and PCS069 models depicted very similar RMSD values over the entirety of the production simulations.

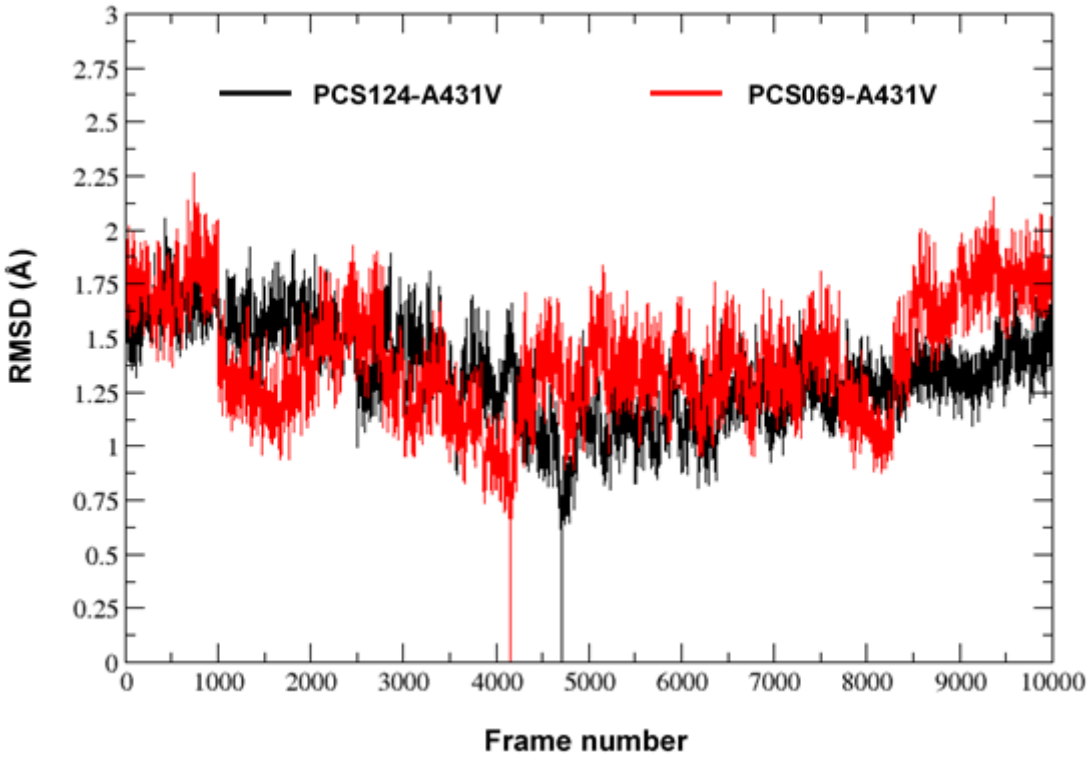


Figure 5.2 Time series RMSD plots of the HIV-1 isolate A431V-PR bound models.

To evaluate specific residue fluctuations in PR, we constructed RMSF plots as seen in Figure 5.3 below. Interestingly, similar trends in particle flexibility of the A431V-PR mutants was observed with the DRV-bound PR models (chapter four; Figure 4.6b) in comparison to the LPV-PR complexes (chapter four; Figure 4.6a). Moreover, the PCS069-A431V model displayed increased particle flexibility over more AA residues in chain B than A thus indicating a greater level of asymmetry in the L76V-inclusive mutant (Figure 5.3; dotted circle).

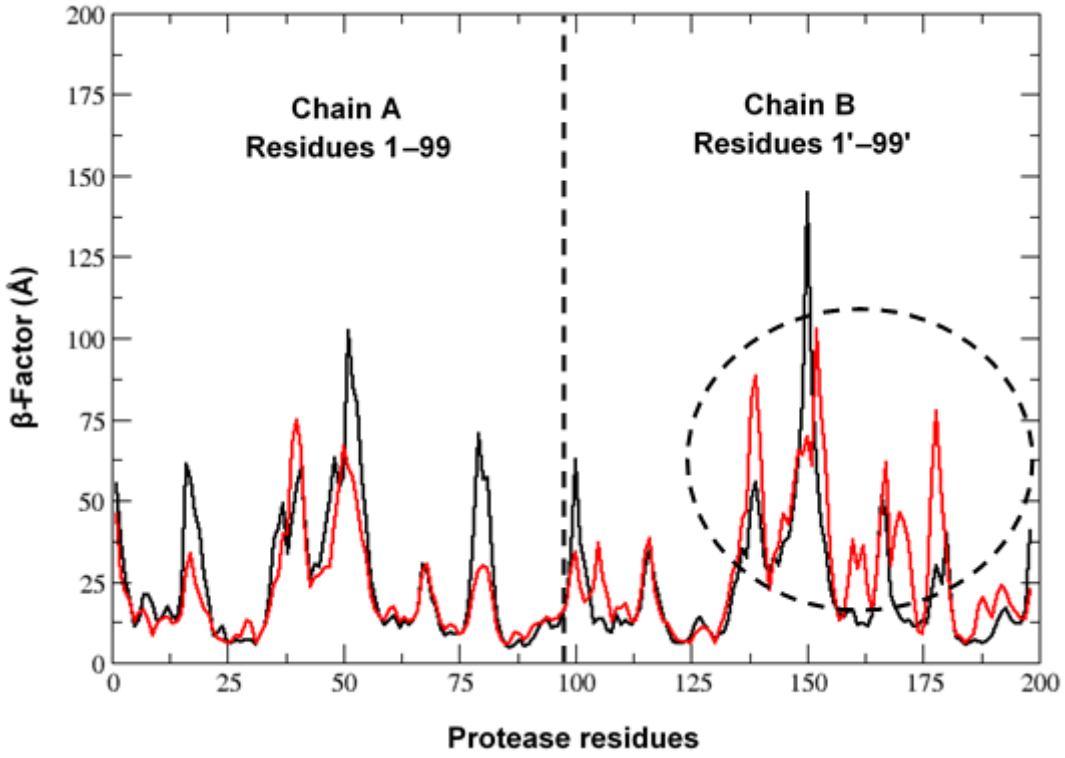
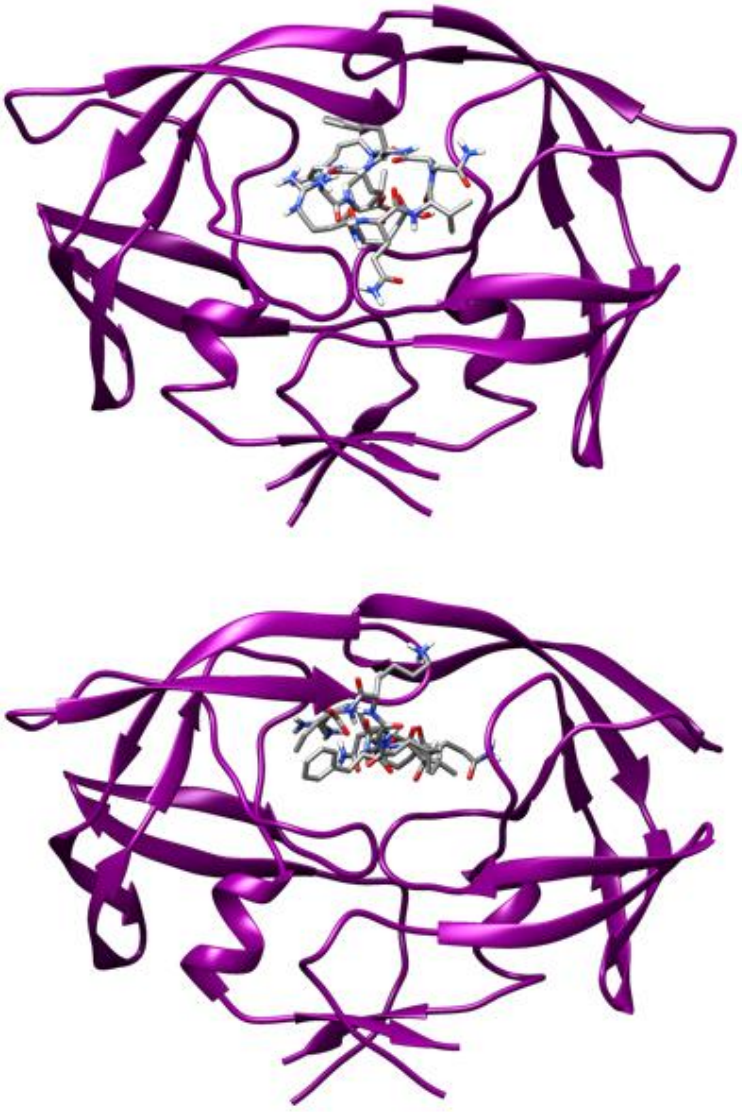


Figure 5.3 RMSF analysis of the A431V-PR bound mutants. **Note:** the HIV-1 PCS124-A431V and PCS069-A431V models are represented by the solid black and red lines, respectively.

5.3.2 NC|p1 binding to the PR mutants

Expectedly, the NC|p1 Gag CSs are stereo-chemically larger than their PI antagonists. As such, the CSs encompass a greater surface area within the PR's substrate cavity, as seen in Figure 5.4 below. Although, the NC|p1 CSs comprise several bulky side chains, such as phenylalanine's aromatic benzene ring at position 433, the AA side chains are rotatable as seen by the alternative conformations in the PCS124 and PCS069 mutant models (Figure 5.4).

The A431V NC|p1 mutants exhibited significantly stronger binding (Figure 5.4) as compared to the LPV- and DRV-bound PR mutants (chapter four; Table 4.3). Specifically, when comparing the most energetically-favourable binding mode, there was a 3-fold energy difference between the A431V mutant (-62.4605 kcal/mol) and the drug-bound mutant (PCS069-LPV: -23.4128 kcal/mol). Interestingly, the L76V-inclusive PCS069 model exhibited a greater binding affinity (-62.4605 kcal/mol) for the NC|p1 mutant in which a -12 kcal/mol energy difference was observed in comparison to the PCS124 model (-50.3386 kcal/mol).



PCS124-A431V	
ΔE_{VDW}	-58.4105
ΔE_{EEL}	-5.8792
ΔG_{GB}	21.7349
ΔG_{SURF}	-7.7838
ΔG_{gas}	-64.2897
ΔG_{solv}	13.9511
ΔG_{bind}	-50.3386

PCS069-A431V	
ΔE_{VDW}	-71.4542
ΔE_{EEL}	-7.1198
ΔG_{GB}	25.4574
ΔG_{SURF}	-9.3439
ΔG_{gas}	78.5740
ΔG_{solv}	16.1135
ΔG_{bind}	-62.4605

Figure 5.4 MM-GBSA binding-free energies of the A431V NC|p1 Gag CSs bound to the PCS124 and PCS069 HIV-1 PR mutants. **Note 1:** all predicted energies were calculated over 2,000 snapshots and are represented in kcal/mol. **Note 2:** ΔE_{VDW} = van der Waals contributions; ΔE_{EEL} = electrostatic contributions; ΔG_{GB} = polar contributions; ΔG_{SURF} = non-polar contributions; ΔG_{gas} = free energy in gas phase; ΔG_{solv} = solvation free energy; ΔG_{bind} = binding free energy.

5.3.3 Gag-PR interaction maps

To evaluate the specific interactions between the NC|p1 Gag CS and the PR mutants, we constructed Gag-PR interaction maps as depicted in Figures 5.5–5.6 below. Accordingly, the PCS124-A431V model revealed hydrogen bonds between residues A28_PR and K436_Gag, G48_PR and the nitrogen atom connecting residues L434_Gag and G435_Gag as well as a double hydrogen bond between PR's R8' (R107) and the oxygen atoms of Gag residues E428 and Q430.

Additionally, a hydrogen bond between D25_PR and the carbon atom of I437 in Gag was also observed. Moreover, residue D29 in PR was attractively charged to R429 in Gag. Of note, whilst L10F_PR formed an alkyl bond with A431V in Gag, G27 and I47 in PR also formed close vdW interactions with V431. In addition, major PRMs V82A' (V181A) and M46I formed vdW interactions with residues R429 and F433 in Gag, respectively. Furthermore, a π-π T-shaped stacking[20] was formed with F53 in PR and F433 in Gag. Interestingly, in addition to substrate cleft residue V32 in PR, L76_PR formed vdW interactions with K436_Gag whilst both I50/I50' (I149) formed close vdW interactions with each other, G48_PR, L76_PR and the NC|p1 substrate.

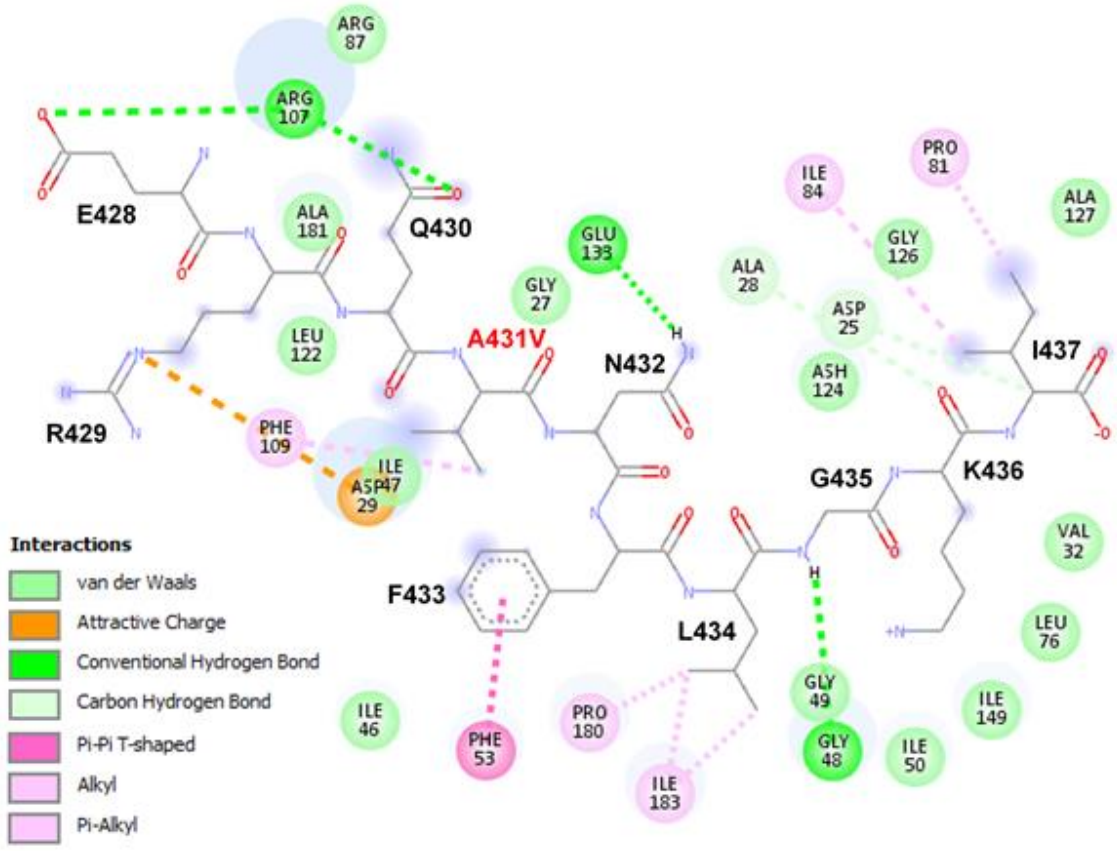


Figure 5.5 Amino acid interaction map showing the association between the HIV-1 isolate PCS124 protease and the mutated NC|p1 Gag cleavage site. **Note:** the A431V Gag mutation is indicated in red.

Figure 5.6 represents the interaction map for the PCS069-A431V model. When compared to the PCS124 substrate-bound model, it was evident that the PCS069 PR mutant formed significantly

[20] π-stacking refers to non-covalent, attractive interactions between two aromatic rings.

more interactions with the substrate. In particular, 12 hydrogen bonds were formed between the PCS069 residues and the NC|p1 CS. Of these, PR residue R8' formed three hydrogen bonds with the oxygen atoms in residues E428, R429 and Q430 in Gag. Interestingly, while D29_PR formed two hydrogen bonds with both oxygen atoms in Q430_Gag, D30_PR also formed an additional hydrogen bond with the second oxygen atom in Q430_Gag. Moreover, WT AA G48 in PR hydrogen bonded with the oxygen atom of the A431V mutation. Finally, residues G49' (G148), I50' (I149) and P79' (P180) also hydrogen bonded with the mutated NC|p1 substrate.

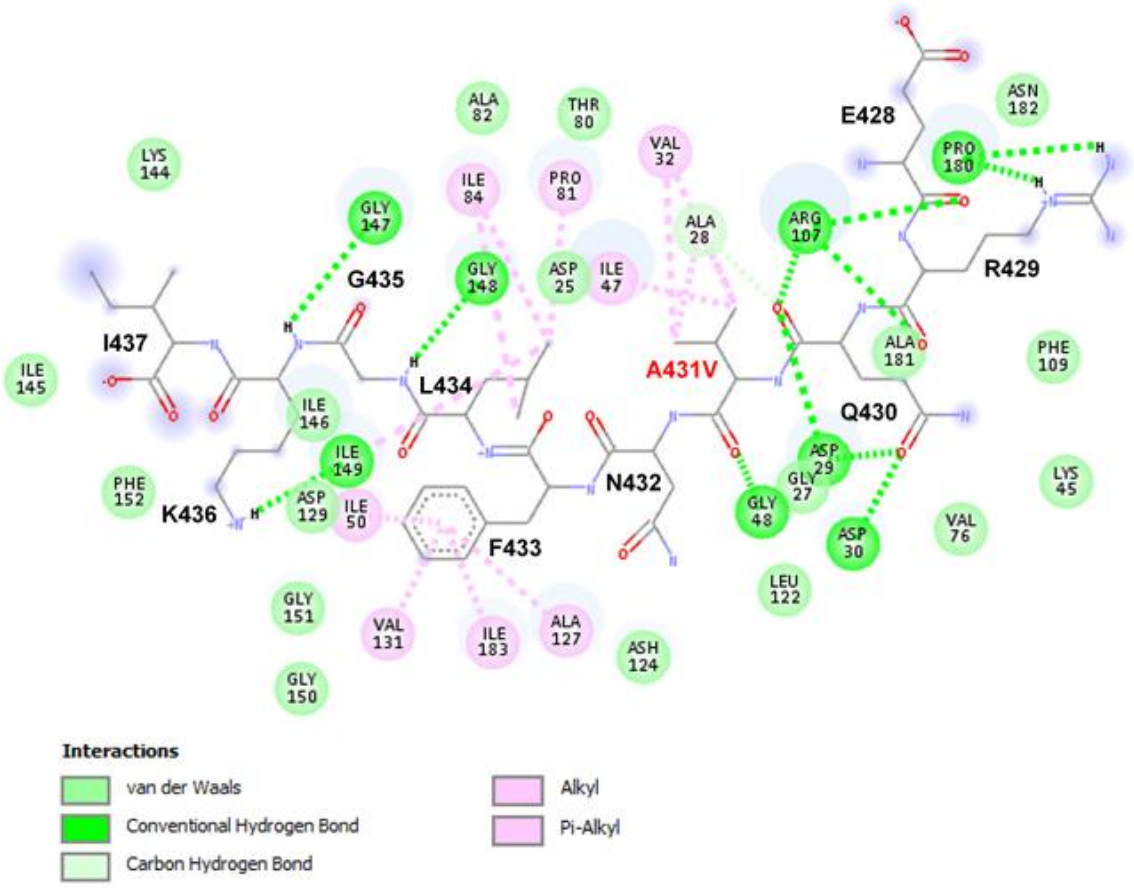


Figure 5.6 Amino acid interaction map showing the association between the HIV-1 isolate PCS069 protease and the mutated NC|p1 Gag CS. **Note:** the A431V Gag mutation is indicated in red.

Furthermore, whilst V82A also formed a vdW interaction with the substrate, V82A' (V181A) was proximal to residues R429 and Q430 in Gag (Figure 5.5). Moreover, minor PRM L10F' (L109F) was closely located to V82A'. Interestingly, the switch from leucine to valine of residue 76 in PR positioned the mutant valine closer to Q430_Gag instead of K436_Gag as seen in the PCS124-

A431V model. Additionally, several interactions were observed with A431V, namely, multiple alkyl bonds with WT residues V32 and I47 in PR. Finally, whilst D25 did not hydrogen bond with the mutated NC|p1 substrate, it formed close vdW interactions to I47 in PR which in turn was closely bonded to A431V in Gag.

5.4 Discussion

Generally, protein folding depends on a multiple factor process and elucidating the mechanism by which they fold provides a useful knowledge on their physiological role in nature (Wu et al., 2015). While our understanding on protein folding has vastly improved over the years, the exact mechanism by which folding occurs *in vivo* is not well understood (Clark and Ugrinov, 2009). Though "the protein folding problem" has proved to be a challenge for half a century[21] (Li et al., 2017), the molecular structures of small proteins are well predicted by computational means (Dill et al., 2008). Here, we evaluated the RMSD and RMSF of the mutated Gag-PR complex over a 100 ns MD simulation. Our analyses revealed that although the PCS124/069 models did fold and re-fold to alternate structures, greater conformational stability was observed over the 100 ns, particularly for the PCS124-A431V model which maintained relative stability from 50 ns onwards. However, the PCS069-A431V model stabilized for 28 ns prior to re-folding and restoring stability in the last 12 ns of the simulation. Goodchild et al. (2011) theorized that proteins can alter their folding in response to the environment whilst regulating protein functionality. Furthermore, a study conducted by Porter and Looger (2018) identified several proteins from the PDB that switch folds, thus having more than one conformation. In this context, this suggests that apart from drastic changes that would render the enzyme inactive, PR can alter its conformation to adapt to its mutated substrate as well as the drug-pressured environment. Additionally, Alexander et al. (2009) observed that a single AA mutation can equate to two folded states of a protein. Therefore, it is highly plausible that the PCS mutants, having four to five PI mutations can exist in several states to acclimate to the drug-pressured environment. Moreover, Agozzino and Dill (2018) found that least stable proteins can rapidly adapt to changing conditions. This suggests that the changing stability of the PCS069 mutant, having five PI mutations at functionally important sites maintains a less favourable structural conformation to continually adapt in favour of the NC|p1 CS binding as opposed to the drugs.

Interestingly, the RMSF data corroborates the RMSD findings and the aforementioned protein folding hypotheses. As greater particle flexibility was comparable to the DRV-bound PR models,

[21] The protein folding problem questions how an AA sequence dictates the tertiary structure of protein.

this suggests that the same mechanism employed to evade drug binding is used to maintain substrate cleavage. Specifically, as previously discussed, changing the AA torsion angles can alter protein conformity (Thukral et al., 2007; Ponnuraj and Saravanan, 2017; Saravanan and Selvaraj, 2017). Additionally, the level of asymmetry in the PCS069 model indicates that each chain is rapidly regulated either to constrict or flex to allow movement for the highly rotatable, bulky Gag CS. Therefore, it is conceivable that the asymmetrical AA fluctuations in a homodimeric protein may evade drug binding through the rapid regulation of its subunits.

While it is known that Gag cleavage occurs over several sequential steps as discussed in chapter one, the exact structural mechanism for this process is debatable (Mattei et al., 2018). Even though the PIs were designed to out-compete the natural substrate (Spearman, 2016), the HIV-1 Gag, consisting of approximately 500 AAs (Fun et al., 2012), is a large protein in nature. Therefore, it is also likely that the 10 bulky AAs recognized by PR for cleavage is also molecularly large. Separate studies conducted by Prabu-Jeyabalan et al. (2002) and King et al. (2004) showed that the PR's substrates occupies a consensus volume within the binding cavity. Özen et al. (2011) hypothesized that the ability of the CSs to fit within the substrate envelope is also determined by the substrate size and dynamics. This suggests that PR's capability to incorporate Gag essentially depends on the peptide's rotamers and ultimately its flexibility. Furthermore, the G435 and K436 residues contributes to the fundamentally flexible nature of the NC|p1 substrate (Özen et al., 2012).

Additionally, this flexibility may also contribute to the stronger binding observed in the PCS-A431V complexes as compared to the drug-bound PR models. The increased binding affinity for the NC|p1 CS also reiterates that the PRMs were compatibly selected to accommodate Gag. Moreover, the 12 kcal/mol difference between the PCS124 and PCS069 models indicates a selective advantage to acquiring a 4^{th} major PRM in favour of three (M46I+I54V+V82A). Therefore, it is accepted that via alternate mechanisms, patterns of mutations act synergistically to evade drugs (Agniswamy et al., 2016).

Lastly, to explore the intricate interactions between the mutated Gag and PR, 2D interaction maps were constructed. Our analyses revealed that key hydrogen bonds involving several Gag residues (E428, Q430, N432, L434 and G435) across the substrate was observed in the PCS124-A431V model. While substrate dynamical flexibility allows for its incorporation (Özen et al., 2011), this suggests that the formation of hydrogen bonds arose to stabilize the CS, keeping it relatively static for cleavage to occur. Furthermore, our analyses also revealed an important hydrogen bond with

PR's catalytic D25 and the carbon atom of isoleucine at position 437 in the NC|p1 Gag CS. This bond suggests that the CS rotated into and downwards of PR's substrate cavity to catalyze the reaction. Incidentally, this could be seen Figure 5.3 where the isoleucine extended down PR. Of note, Gag Q430, G435, K436 and I437 in Gag are also sites at which drug resistance/exposure associated mutations occur (Doyon et al., 1996; Gonzalez et al., 2004; Ghosn et al., 2011; Knops et al., 2011). As such, these regions in the substrate are variable and can therefore allow for alternate substrate poses to develop when PR is unable to actively accommodate the CS. Interestingly, the close interaction of the A431V Gag mutation with G27_PR and I47_PR indicate that the slightly larger valine substitution in place of alanine brings the active site and flaps closer together thus constricting the available space and allowing for greater binding affinity. Furthermore, the proximity of M46I in PR and the phenylalanine of F433 in Gag suggests that coordination between these two residues can regulate movement within the flaps. Of note was the interaction of the I50's to each other and the substrate. Tóth and Borics (2006) showed that the enzyme can exist in open, semi-open and curled conformations during a single MD simulation and that the distance between I50's (flap tips) indicate which conformation was favoured at the time. Finally, the bond between L10F_PR and A431V_Gag which was also observed in our Gag-PR BN (chapter three), suggests that coordination between these two residues are important for substrate recognition and linkage.

The increase of hydrogen bonds between the PCS069 PR and the NC|p1 substrate suggests that the L76V mutation in PR significantly contributes toward substrate recognition and favourbale cleavage rather than evading drug binding directly. Generally, hydrogen bonds facilitate protein-ligand binding through the displacement of receptor water molecules (Chen et al., 2016). In essence, either the water molecules are displaced by the ligand or are subtly shifted (Huggins and Tidor, 2011). While water molecules are important mediators in protein-ligand binding (Brenk et al., 2006), Chen et al. (1998) showed that in an enzyme-inhibitor complex, the ligand displaced an active site water molecule which created favourable inhibitor orientation. Thus, in our study, the extensive formation of hydrogen bonds suggests that protein water displacement may have occurred to properly orient the ligand within the substrate cavity. Additionally, these data also revealed that the L76V PR mutation also closely interacted with Q430 in Gag. Incidentally, Q430_Gag was also involved in four out of the 12 hydrogen bonds. A study conducted by van Maarseveen et al. (2012) on NC|p1 CS efficiency on resistance revealed that AA position of the substrate significant correlated with the difference between the NC|p1 430–435 residues outside PR's cavity and within the active site. Therefore, Q430 as well as A431V in Gag can alter PR's active site dynamics for efficient substrate cleavage.

While the hydrogen bonds between the PCS069 PR and NC|p1 CS hold the substrate in place, the CS seemingly "sits" at the top, closer to PR's flaps (Figure 5.3). This observation was pronounced by the hydrogen bond linking A431V in Gag to G48 in PR as well as the alkyl bond between I47_PR and A431V_Gag. Studies have shown that the positioning of PR's flaps correlated with its sensitivity to the PIs (Wlodawer et al., 1989; Yedidi et al., 2014) suggesting that the same mechanism can be employed when bound to the mutated NC|p1 substrate. Additionally, a study conducted by Khan et al. (2018) suggested that switching between flap positioning during inhibitor binding requires additional changes in PR's conformation which may result in significant energy fluctuations. In this study, conformational changes in other parts of PR were observed by interactions between PRMs L10F, V82A and L76V (Figure 5.5; dotted black lines). This suggests that when L76V is present together with A431V_Gag, the mutational dynamics between this PR resistance combination (L10F+M46I+I54V+V82A) changes to constrict these regions in PR and allow for efficient substrate processing in favour of drug binding. Contrastingly, the switch from hydrogen bonding to vdW interactions between D25_PR and the substrate indicates an indirect mechanism of cleavage which is consistent with studies evaluating L76V, as previously discussed (chapter four). Noteworthy, while vdW interactions are considered weak, these forces are often important in the interaction and shape of molecules (Atkins and de Paula, 2006). Therefore, in this instance, cleavage is coordinated by the flaps and strong vdW interactions rather than direct active site dynamics.

5.5 Conclusions

In summary, these data indicate that the mutated PR depicted great affinity for the NC|p1 Gag CS. Specifically, the A431V Gag mutation coordinated several PR residues to aid in substrate recognition and efficient binding. In addition, the PRMs actively work together to provide a compatible conformation that can accommodate Gag. Particularly, PRM L76V plays an important role in coordinating PR resistance dynamics, suggesting that its role is closely related to CS recognition rather than association with the drugs. Finally, these data revealed that constricting and flexing specific regions in PR while allowing flexible movement of the substrate can allow for multiple, complex mechanisms of resistance to occur.

Chapter six

General discussion, conclusions, future recommendations and study limitations

6.1 General discussion

With an estimated 1.8 million new HIV infections reported in 2017 equating to approximately 5,000 infections per day, HIV remains a serious public health risk (UNAIDS, 2018). Although more than a 55% reduction of AIDS-related deaths[22] have been reported since 2004 due to ARV therapy (HIV.gov, 2019), drug resistance and thereby the virus's ability to actively replicate despite the drugs remains a serious challenge (Ngo-Giang-Huong and Aghokeng, 2019). While the PIs show effective potency against the virus (Titanji et al., 2013), PR is able to accumulate resistance mutations in various regions, thus thwarting drug binding (Henderson et al., 2011). Furthermore, several resistance associated mutations reported at cleavage and non-CS positions in PR's natural substrate Gag, have also been reported in several studies (Maguire et al., 2002; Verheyen et al., 2006; Kolli et al., 2009; Parry et al., 2011; Shibata et al., 2011). While these mutations can improve viral fitness in the presence of PRMs (Fun et al., 2012), they have also been shown to directly infer resistance in the absence of PRMs (Parry et al., 2011). Therefore, it is becoming increasingly apparent that Gag and PR share evolutionary mechanisms to synergistically act against PI binding.

Most studies have investigated subtype B viruses even though subtype C remains the most prevalent subtype globally (Alcalde et al., 2012; Costa et al., 2014; Castley et al., 2017). Though several studies based on functional assays have provided useful information on the effect of resistance mutations found in Gag and PR (Brann et al., 2006; Matsunaga et al., 2015; Zhu et al., 2015; Giandhari et al., 2016), the selection of combinations of resistance mutations and its impact on the structure is lacking (Zhang et al., 2010; Costa et al., 2014). It is therefore important to improve our understanding on the pathways to resistance and the complexities associated with treatment failure.

Therefore, the current study investigated Gag-PR coevolution under drug selection pressure and the structural implications of patterns of resistance mutations in HIV-1 subtype C. To do this, we first identified the selection and frequency of mutations in Gag and PR using 1,972 and 130 PR and Gag sequences, respectively. Thereafter, we performed positive selection and coevolution analyses to construct Bayesian networks for elucidating probable pathways to resistance. Lastly, using the knowledge gained from BNL, we investigated the structural changes induced by the combination of resistance mutations on PI and Gag CS binding using homology modelling, docking and MD simulations.

[22] Globally, in 2018, approximately 770,000 people died as a result of AIDS-related deaths, compared to the 1.2 and 1.7 million deaths observed in 2010 and 2004, respectively.

Of the 1,972 PR sequences, only 964 had major PRMs. Site-wise, the highest percentage of mutations occurred at positions 46 (19%), 54 (23%) and 82 (25%). Additionally, certain AAs were selected more frequently over others, for example M46I occurred in 16% of the dataset as opposed to M46L (3%), thus suggesting the role of AAs in protein functionality. The same trend could be seen for I54V (22%) and V82A (23%). Incidentally, the M46I+I54V+V82A combination, which was also the most frequent pattern observed in this study, is commonly associated with failure to several PIs including the 2nd-line drug, LPV (Barber et al., 2012). Although, L76V only occurred in 7% of the dataset, it was part of the second most recurring combination (M46I+I54V+L76V+V82A) in this study. Amongst the minor PRMs, L10F/I, A71V and T74S were highly prevalent in the dataset. While the exact impact of minor resistance mutations on PR is arguably unclear (Nijhuis et al., 1999; Henderson et al., 2011; Wensing et al., 2017), it is accepted that their presence can either improve viral fitness or increase the level of drug resistance in the presence of other major PRMs (Scherrer et al., 2012).

Of the 30 Gag resistance/exposure associated mutations (Fun et al., 2012), only A431V, L449F, R452K, K62R, N373P, I401V, R409K and S451T occurred at significantly ($p<0.01$) higher frequencies in the treatment associated dataset. Of these, A431V (NC|p1), L449F (p1|p6), R452K (p1|p6), N373P (p2|NC) and S451T (p1|p6) are Gag CS mutations. Furthermore, this study also identified several presumptive Gag resistance associated mutations across multiple Gag domains. These included T53N, S54A, Q69K, D122T, S125D and N126E in matrix, D260E, I256V, D319E and R335K in capsid as well as P485A and L498S in p6. While the selection of Gag CS mutations directly alters the CS thereby allowing for unhindered substrate processing (Fun et al., 2012), non-CS mutations allow for distal changes in the structure to occur (Özen et al., 2014). Since Gag has numerous functional roles (chapter one; section 1.10), it is conceivable that these mutations are selected to preserve Gag functionality rather than resistance development itself. These data provided useful information on the prevalence and patterns of resistance in PR and Gag. Bearing this in mind, we investigated the evolutionary pathways of resistance in chapter three.

To evaluate the pathways leading resistance, we firstly identified the positively selected and coevolving sites in Gag-PR. To do this, we compiled 58 complete Gag-PR sequences from the PCS cohort. Accordingly, an overlapping coevolving group of 10, 46, 54, and 82 was observed suggesting that these positions either directly or indirectly interact with each other. Moreover, of these, codons 10 and 82 were also positively selected indicating that mutations at these positions provide a selective advantage to the enzyme. While several Gag mutations (E365, S373, A374, A431, R452 and Q474) were identified in the coevolution analysis. Of note was the NC|p1 Gag

CS mutation A431V, since it also occurred in 26% of the Gag dataset. This mutation has been associated with V82A, M46I and L76V in previous studies (Bally et al., 2000; Myint et al., 2004; Nijhuis et al., 2009). In addition to conferring resistance alone (Dam et al., 2009), it has also been shown to improve viral fitness in the presence of PRMs in subtype B (Mammano et al., 2000; Cote et al., 2001).

To the best of our knowledge, the evaluation of Gag-PR resistance mutations in pathways to LPV resistance using BNL has not been shown before. This study proposed two probable pathways to LPV resistance. The first pathway connected I54V to L10F, M46I and V82A while inverse associations with L24 and Q58 connected L76V to M46I+I54V+V82A. The second pathway connected V82A with LPV via L93 in PR. We postulated that the deciding factor as to whether I54V or V82A is selected first in the pathway also depends on selection of L10F. Therefore, if L10F is selected then I54V is also selected. While conditional dependencies between 46, 54 and 82 have been observed (Zhang et al., 2010), a recent study conducted by Vasavi et al. (2017) revealed that in a double PR mutant comprising L10F/N88S, the binding affinity of PR to NFV was lowered in the presence of these mutations. This suggest that L10F may play an important role in regulating resistance in the presence of major PRMs. Interestingly, in the Gag-PR network, the interactions between the PRMs remained unchanged whilst the Gag mutations essentially added to the existing network. Of note, A431V correlated with M46I+V82A via I54V. It also interacted with L10F. This suggests that associations between A431V and V82A depends on the selection of M46I and I54V which in turn depends on L10F. With numerous studies highlighting Gag-PR coevolution (Shibata et al., 2011; Kolli et al., 2014; Deshmukh et al., 2016; Codoñer et al., 2017), it is evident that the basis on which resistance occurs is dynamic and complex. Consequently, the structural mechanism by which these resistance mutations act on resistance is certainly unclear (Lin, 2016). Particularly, in the context of this study, the structural changes of L10F+M46I+I54V+L76V+V82A in subtype C is elusive. Therefore, chapter four evaluated the structural implications of this combination on LPV and DRV binding.

Based on the data from chapter two and the BNL which especially highlights L76V as a 4th mutation, we selected two PR sequences, one with L76V (PCS069) and one without (PCS124) in addition to the L10F+M46I+I54V+V82A from the PCS cohort to study further. The MD simulations revealed that the LPV-bound PR mutants were relatively stable over the 100 ns simulation in comparison to the DRV-bound models. Expectedly, LPV revealed stronger binding to the PR mutants. Although DRV is considered extremely potent as it comfortably fits into PR's substrate envelope, resistance can occur through the accumulation of several PRMs (Lockbaum

et al., 2018). Our data revealed that a loss of contact between DRV's essential chemical moieties, bis-furyl and 4-aminobenzenesulfonamide contributes largely to DRV inefficacy which is consistent with findings observed elsewhere (Raugi et al., 2016). LPV resistance on the other hand seems to be readily affected by the direct structural changes induced by the drug resistance mutations as seen in this study. While it is evident that these dynamical shifts in PR can allow for a reduction in binding affinity of the drugs, it is unclear how these mutations would be advantageous during cleavage. Furthermore, the presence and link of the NC|p1 A431V Gag mutation to these PRMs in terms of structure is unclear. Therefore, chapter five investigated interactions of the mutant NC|p1 CS on the MDR PR models.

The structural interactions of these specific PRMs coupled with the mutated NC|p1 Gag CS (A431V) has not been shown prior to the current study. These data revealed that MDR PR models had a stronger affinity for the CS in comparison to the PI drugs. This was particularly evident from the binding scores where the PCS124 and PCS069 models had a binding energy of -50.3386 kcal/mol and -62.4605 kcal/mol, respectively. This was a 3-fold energy difference between the lowest binding score of drug-bound model. Furthermore, it was seen that A431V coordinated several residues in PR to improve substrate recognition and binding whilst the PRMs work to alter PR's conformation to accommodate Gag. Additionally, it was suggested that L76V in PR provides a mechanistic advantage for substrate recognition and binding as opposed to directly inhibiting the drug as shown in other studies (Louis et al., 2011; Wong-Sam et al., 2018). Therefore, this study highlights that L76V utilizes an additional novel approach in PI drug resistance. Lastly, to our best knowledge, docking of the theoretical PR and Gag CSs to evaluate drug resistance and substrate binding has not been previously shown.

6.2 Conclusions

The current study investigated the intimate interactions between complex resistance patterns in Gag and PR under drug selection pressure. As observed, Gag and PR readily accumulate mutations in various regions to outcompete the inhibitors in favour of substrate binding. Importantly, the selection of mutations is not random but rather advantageous to the structural complexity of the enzyme-substrate complex. In particular, the mutant PR overcomes drug binding by switching conformations through mutations in or near the active and flap regions. Consequently, a balance between enzyme rigidity and flexibility through mutations and chain asymmetry aids the virus in restoring cleavage. Further complicating the process is the role of the substrate. Additionally, proper incorporation of the A431V Gag CS within PR's cavity depended

on movement of the substrate. Consequently, the AA side chains can rotate to aid in substrate recognition. These data highlight that the functional aspects of resistance are highly correlated with the structural dynamics of the enzyme and substrate.

6.3 Future recommendations and limitations

Some aspects of this study were limited and are hereby recommended for future investigation:

1. The Gag dataset utilized in this study was small. However, it is accepted there are limited viral sequences available that were genotyped from patients receiving ARV therapy. Increasing the sample size to evaluate a larger picture of Gag resistance dynamics is needed.
2. Functional assays investigating the step-wise accumulation of the mutations reported in this study should be performed to present a complete picture on pathways to resistance.
3. As some subtype C signature mutations were present in the BNL, a study evaluating these signature mutations in subtype C vs subtype B should be performed to identify possible subtype differences in pathways to resistance.
4. This study only evaluated the A431V NC|p1 Gag CS mutation because of its importance in the Bayesian networks, however, mutations at the remaining Gag CSs should be further studied to elucidate their individual roles in resistance in the presence of the PRMs highlighted in this study.
5. As this study did not look at the WT, the NC|p1 CS should also be evaluated without the A431V mutation in the presence of the PRMs to elucidate how cleavage might occur without this mutation.
6. Since several non-CS mutations were identified in the Gag-PR BNL, such as Q69K and I256V, the structural impacts of these mutations should be studied further to identify its role in pathways to resistance.
7. Accelerated molecular dynamics should be performed on a millisecond scale to evaluate the folding of the mutant Gag-PR proteins.
8. Using the information obtained in this study as a foundation, a computer-based programme can be developed to predict viral replication capacity on a structural level.

References

Adachi M, Ohhara T, Kurihara K, Tamada T, Honjo E, Okazaki N, Arai S, Shoyama Y, Kimura K, Matsumura H, Sugiyama S, Adachi H, Takano K, Mori Y, Hidaka K, Kimura T, Hayashi Y, Kiso Y and Kuroki R. 2009. Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proc. Natl. Acad. Sci*., 106, 4641–4646.

Agniswamy J, Kneller DW, Brothers R, Wang Y-F, Harrison RW and Weber IT. 2019. Highly drug-resistant HIV-1 protease mutant PRS17 shows enhanced binding to substrate analogues. *ACS Omega*, 4, 8707−8719.

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

Agniswamy J, Louis JM, Shen C-H, Yashchuk S, Ghosh AK and Weber IT. 2015. Substituted Bis-THF protease inhibitors with improved potency against highly resistant mature HIV-1 protease PR20. *J. Med. Chem*., 58, 5088–5095.

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

Agniswamy J, Shen C-H, Wang Y-F, Ghosh AK, Rao KV, Xu C-H, Sayer JM, Louis JM and Weber IT. 2013. Extreme multidrug resistant HIV-1 protease with 20 mutations is resistant to novel protease inhibitors with P1′-pyrrolidinone or P2-tris-tetrahydrofuran. *J. Med. Chem*., 56, 4017–4027.

Agozzino L and Dill KA. 2018. Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proc. Natl. Acad. Sci. USA*, 115, 9092–9097.

AIDSMAP. 2014. What is drug resistance? http://www.aidsmap.com/What-is-drug-resistance/page/1327026/. [Accessed 20th June].

Alcalde R, Guimaraes ML, Duarte AJ and Casseb J. 2012. Clinical, epidemiological and molecular features of the HIV-1 subtype C and recombinant forms that are circulating in the city of Sao Paulo, Brazil. Virol. J., 9, 156.

Alexander PA, He Y, Chen Y, Orban J and Bryan PN. 2009. A minimal sequence code for switching protein structure and function. *Proc. Natl. Acad. Sci. USA*, 106, 21149–21154.

Alexandre KB, Gray ES, Pantophlet R, Moore PL, Mcmahon JB, Chakauya E, O'Keefe BR, Chikwamba R and Morris L. 2011. Binding of the mannose-specific lectin, griffithsin, to HIV-1 gp120 exposes the CD4-binding site. *J. Virol*., 85, 9039–9050.

Alizon S and Magnus C. 2012. Modelling the course of an HIV infection: insights from ecology and evolution. *Viruses*, 4, 1984–2013.

Amamuddy OS, Bishop NT and Bishop ÖT. 2018. Characterizing early drug resistance-related events using geometric ensembles from HIV protease dynamics. *Sci. Rep*., 8, 17938.

Anfinsen CB, Haber E, Sela M amd Jr WF. 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA*, 47, 1309–1314.

Aoki M, Das D, Hayashi H, Aoki-Ogata H, Takamatsu Y, Ghosh AK and Mitsuya H. 2018. Mechanism of Darunavir (DRV)'s high genetic barrier to HIV-1 resistance: a key V32I substitution in protease rarely occurs, but once it occurs, it predisposes HIV-1 to develop DRV resistance. *mBio*, 9, e02425–17.

Aoki M, Venzon DJ, Koh Y, Aoki-Ogata H, Miyakawa T, Yoshimura K, Maeda K and Mitsuya H. Non-cleavage site Gag mutations in Amprenavir-resistant human immunodeficiency virus

type 1 (HIV-1) predispose HIV-1 to rapid acquisition of Amprenavir resistance but delay development of resistance to other protease inhibitors. *J. Virol*., 83, 3059–3068.

Argudo D, Bethel NP, Marcoline FV, Wolgemuth CW and Grabe M. 2017. New continuum approaches for determining protein-induced membrane deformations. *Biophys. J*., 112, 2159-2172.

Arthur DE and Uzairu A. 2019. Molecular docking studies on the interaction of NCI anticancer analogues with human Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit. *J. King Saud. Univ. Sci*., doi.org/10.1016/j.jksus.2019.01.011.

Arts EJ and Hazuda DJ. 2012. HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med*., 2, a007161.

Atkins P and de Paula J. 2006. Physical chemistry for the life sciences, 2nd edition. W.H. Freeman and Company, New York, pp. 1–618.

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

Barber TJ, Harrison L, Asboe D, Williams I, Kirk S, Gilson R, Bansi L, Pillay D and Dunn D. 2012. Frequency and patterns of protease gene resistance mutations in HIV-infected patients treated with Lopinavir/Ritonavir as their first protease inhibitor. *J. Antimicrob. Chemother*., 67, 995–1000.

Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W and Montagnier L. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*., 220, 868-871.

Barré-Sinoussi F, Ross AL and Delfraissy J-F. 2013. Past, present and future: 30 years of HIV research. *Nat. Rev. Microbiol*., 11, 877–883.

Baxter JD, Chasanov WM and Adams JL. 2016. An update on HIV-1 protease inhibitor resistance. *J. AIDS Clin. Res*., 7, 6.

Bell NM and Lever AML. 2013. HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol*., 21, 136–144.

Berkowitz R, Fisher J and Goff SP. 1996. RNA packaging. In Krausslich, HG (eds.), *Current Topics in Microbiology and Immunology: RNA Packaging*. Springer Verlag, Berlin, Germany, pp. 177–218.

Berthet-Colominas C, Monaco S, Novelli A, Sibï G, Mallet F and Cusack S. 1999. Head-to-tail dimers and inter-domain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. *EMBO J*., 18, 1124–1136.

Bessong PO. 2008. Polymorphisms in HIV-1 subtype C proteases and the potential impact on protease inhibitors. *Trop. Med. Int. Health*, 13, 144–151.

Bharat TAM, Davey NE, Ulbrich P, Riches JD, de Marco A, Rumlova M, Sachse C, Ruml T and Briggs JAG. 2012. Structure of the immature retroviral capsid at 8Å resolution by cryo-electron microscopy. *Nature*., 487, 385–389.

Bharat TAM, Menendez LRC, Hagen WJH, Lux V, Igonet S, Schorb M, Schur FKM, Kräusslich H-G and Briggs JAG. 2014. Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly. *Proc. Natl. Acad. Sci*., 22, 8233–8238.

Boender TS, Hamers RL, Ondoa P, Wellington M, Chimbetete C, Siwale M, Maksimos EEFL, Balinda SN, Kityo CM, Adeyemo TA, Akanmu AS, Mandaliya K, Botes ME, Stevens W,

Rinke de Wit TF and Sigaloff KCE. 2016. Protease inhibitor resistance in the first 3 years of second-line antiretroviral therapy for HIV-1 in Sub-Saharan Africa. *J. Infect. Dis*., 214, 873–883.

Borman AM, Paulous S and Clavel F. 1996. Resistance of human immunodeficiency virus type 1 to protease inhibitors: selection of resistance mutations in the presence and absence of the drug. *J. Gen. Virol*., 77, 419–426.

Borsetti A, Ohagen A and Gottlinger HG. 1998. The C-terminal half of the human immunodeficiency virus type 1 Gag precursor is sufficient for efficient particle assembly. *J. Virol*., 72, 9313–9317.

Brann TW, Dewar RL, Jiang MK, Shah A, Nagashima K, Metcalf JA, Falloon J, Lane HC and Imamichi T. 2006. Functional correlation between a novel amino acid insertion at codon 19 in the Protease of human immunodeficiency virus type 1 and polymorphism in the p1/p6 Gag cleavage site in drug resistance and replication fitness. *J. Virol*., 80, 6136-6145.

Brenk R, Vetter SW, Boyce SE, Goodin DB and Shoichet BK. 2006. Probing molecular docking in a charged model binding site. *J. Mol. Biol*., 357, 1449–14470.

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

Budambula V, Musumba FO, Webale MK, Kahiga TM, Ongecha-Owuor F, Kiarie JN, Sowayi GA, Ahmed AA, Ouma C and Were T. 2015. HIV-1 protease inhibitor drug resistance in Kenyan antiretroviral treatment-naïve and -experienced injection drug users and non-drug users. *AIDS Res. Ther*., 12, 27.

Buonaguro L, Tornesello ML and Buonaguro FM. 2007. HIV-1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J. Virol*., 81, 10209–10219.

Burnett JC, Miller-Jensen K, Shah PS, Arkin AP and Schaffer DV. 2009. Control of stochastic gene expression by host factors at the HIV promoter. *PLoS Pathog*., 5, e1000260.

Butler IF, Pandrea I, Marx PA and Apetrei C. 2007. HIV genetic diversity: biological and public health consequences. *Curr. HIV. Res*., 5, 23-45.

Cai Y, Myint W, Paulsen JL, Schiffer CA, Ishima R and Yilmaz NK. 2014. Drug resistance mutations alter dynamics of inhibitor-bound HIV-1 protease. *J. Chem. Theory Comput*., 10, 3438–3448.

Cairns TM and Craven RC. 2001. Viral DNA synthesis defects in assembly-competent Rous sarcoma virus CA mutants. *J. Virol*., 75, 242–250.

Callebaut C, Stray K, Tsai L, Williams M, Yang Z-Y, Cannizzaro C, Leavitt SA, Liu X, Wang K, Murray BP, Mulato A, Hatada M, Priskich T, Parkin N, Swaminathan S, Lee W, He G-X, Xu L and Cihlar T. 2011. In vitro characterization of GS-8374, a novel phosphonate-containing inhibitor of HIV-1 protease with a favorable resistance profile. *Antimicrob. Agents Chemother*., 55, 1366–1376.

Campbell-Yesufu OT and Gandhi RT. 2011. Update on human immunodeficiency virus (HIV)-2 infection. *Clin. Infect. Dis*., 52, 780–787.

Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham TE III, Darden TA, Duke RE, Gohlke H, Goetz AW, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossváry I, Kovalenko A, Lee TS, LeGrand S, Luchko T, Luo R, Madej B, Merz KM, Paesani F, Roe DR, Roitberg A, Sagui C, Salomon-Ferrer R, Seabra G, Simmerling CL, Smith W, Swails J, Walker RC, Wang J, Wolf RM, Wu X and Kollman PA. 2014. AMBER 14. University of California, San Francisco.

Castley A, Sawleshwarkar S, Varma R, Herring B, Thapa K, Dwyer D, Chibo D, Nguyen N, Hawke K, Ratcliff R, Garsia R, Kelleher A, Nolan D, The Australian Molecular Epidemiology Network-HIV (AMEN-HIV). 2017. A national study of the molecular epidemiology of HIV-1 in Australia 2005–2012. *PLoS ONE*, 12, e0170601.

Ceccherini-Silberstein F, Erba F, Gago F, Bertoli A, Forbici F, Bellocchi MC, Gori C, d'Arrigo R, Marcon L, Balotta C, Antinori A, d'Arminio Monforte A and Perno C-F. 2004. Identification of the minimal conserved structure of HIV-1 protease in the presence and absence of drug pressure. *AIDS*, 18, 11–19.

CDC (Centre for Disease Control and Prevention). 1981. Pneumocystis pneumonia --- Los Angeles. *MMWR Morb. Mortal Wkly. Rep*., 30, 250–252.

Chakrabarti S and Panchenko AR. 2010. Structural and functional roles of coevolved sites in proteins. *PLoS ONE*, 5, e8591.

Chang MW and Torbett BE. 2011. Accessory mutations maintain stability in drug resistant HIV-1 protease. *J. Mol. Biol*., 410, 756–760.

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

Chen CR and Makhatadze GI. 2017. Molecular determinant of the effects of hydrostatic pressure on protein folding stability. *Nature Commun*., 8, 14561.

Chen D, Oezguen N, Urvil P, Ferguson C, Dann SM and Savidge TC. 2016. Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Sci. Adv*., 2, e1501240.

Chen JM, Xu SL, Wawrzak Z, Basarab GS and Jordan DB. 1998. Structure-based design of potent inhibitors of scytalone dehydratase: displacement of a water molecule from the active site. *Biochemistry*, 51, 17735–17744.

Cheung NJ and Yu W. 2018. *De novo* protein structure prediction using ultra-fast molecular dynamics simulation. *PLoS ONE*, 13, e0205819.

Chica RA. 2018. Designer proteins activate fluorescence. *Nature*, 561, 471–472.

Chimukangara B, Lessells RJ, Rheed S-Y, Giandhari J, Kharsany ABM, Naidoo K, Lewis L, Cawood C, Khanyile D, Ayalew KA, Diallo K, Samuel R, Hunth G, Vandormaela A, Stray-Pedersenk B, Gordon M, Makadzange T, Kiepiela P, Ramjee G, Ledwaba J, Kalimashe M, Morris L, Parikh U, Mellors JW, Shafer RW , Katzenstein D, Moodley P, Gupta RK, Pillay D, Karim SSA and de Oliveira T. 2019. Trends in pretreatment HIV-1 drug resistance in antiretroviral therapy-naive adults in South Africa, 2000–2016: a pooled sequence analysis. *EClinicalMedicine*, 9, 26–34.

Chin BS. 2017. Molecular epidemiology of human immunodeficiency virus. *Infect. Chemother*., 49, 1–9.

Choudhury S, Everitt L, Pettit SC, Kaplan AH. 2003. Mutagenesis of the dimer interface residues of tethered and untethered HIV-1 protease result in differential activity and suggest multiple mechanisms of compensation. *Virol*., 307, 204–212.

Ciemny M, Kurcinski M, Kamel K, Kolinski A, Alam N, Schueler-Furman O and Kmiecik S. 2018. Protein–peptide docking: opportunities and challenges. *Drug Discov. Today*, 23, 1530–1537.

Clark PL and Ugrinov KG. 2009. Chapter 24 measuring cotranslational folding of nascent polypeptide chains on ribosomes. In *Methods in Enzymology*. Academic Press, Indiana, USA, pp. 567–590.

Clavel F and Mammano F. 2010. Role of Gag in HIV resistance to protease inhibitors. *Viruses*, 2, 1411–1426.

Codoñer FM, Peña R, Blanch-Lombarte O, Jimenez-Moyano E, Pino M, Vollbrecht T, Clotet B, Martinez-Picado J, Draenert J and Prado JG. 2017. Gag-protease coevolution analyses define novel structural surfaces in the HIV-1 matrix and capsid involved in resistance to protease inhibitors. *Sci. Rep*., 7, 3717.

Coffin J, Haase A, Levy JA, Montagnier L, Oroszlan S, Teich N, Temin H, Toyoshima K, Varmus H, Vogt P and Weiss R. 1986. What to call the AIDS virus? *Nature*, 321, 10.

Cohen CJ. 2006. Successful HIV treatment: lessons learned. *J. Manag. Care Pharm*., 12, 6–11.

Collins JR, Burt SK and Erickson JW. 1995. Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics. *Nat. Struct. Biol*., 2, 334–338.

Coman RM, Robbins AH, Fernandez MA, Gilliland CT, Sochet AA, Goodenow MM, McKenna R and Dunn BM. 2008. The contribution of naturally occurring polymorphisms in altering the biochemical and structural characteristics of HIV-1 subtype C protease. *Biochemistry*, 47, 731–743.

Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero J, Rhodes A, Robbins HL, Roth E, Shivaprakash M, Titus D, Yang T, Tepplert H, Squires KE, Deutsch PJ and Emini EA. 1995. *In vivo* emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature*., 374, 569–571.

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

Costa MGS, Benetti-Barbosa TG, Desdouits N, Blondel A, Bisch PM, Pascutti PG and Batista PR. 2014. Impact of M36I polymorphism on the interaction of HIV-1 protease with its substrates: insights from molecular dynamics. *BMC Genomics*, 15, S5.

Cote HC, Brumme ZL and Harrigan PR. 2001. Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by Indinavir, Ritonavir, and/or Saquinavir. *J. Virol*., 75, 589–594.

Court R, Gordon M, Cohen K, Stewart A, Gosnell B, Wiesner L and Maartens G. 2016. Random Lopinavir concentrations predict resistance on Lopinavir-based antiretroviral therapy. *Int. J. Antimicrob. Agents*, 48, 158–162.

Craveur P, Joseph AP, Rebehmed J and de Brevern AG. 2013. β-bulges: extensive structural analyses of β-sheets irregularities. *Protein Sci*., 22, 1366–1378.

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

Darapaneni VC, Sakhamuri J and Darapaneni V. 2015. Conservation analysis of HIV-1 protein sequences reveal potential drug binding sites: a case of viral protein U and protease. *Am. J. Current Virol*., 1, 201500607.

Darden T, York D and Pedersen L. 1993. Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys*., 98, 10089.

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

Daw MA, El-Bouzedi A, Ahmed MO, Dau AA and in association with the Libyan Study Group of Hepatitis & HIV. 2017. Molecular and epidemiological characterization of HIV-1 subtypes among Libyan patients. *BMC Res. Notes*, 10, 170.

DeBenedictis EP and Keten S. 2019. Mechanical unfolding of alpha- and beta-helical protein motifs. *Soft Matter*, 15, 1243–1252.

Deforche K, Silander T, Camacho R, Grossman Z, Soares MA, Van Laethem K, Kantor R, Moreau Y and Vandamme A-M. 2006. Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. *Bioinformatics*, 22, 2975–2979.

Delino NS, Aoki M, Hayashi H, Hattori S, Chang SB, Takamatsu Y, Martyr CD, Das D, Ghosh AK and Mitsuya H. 2018. GRL-079, a novel HIV-1 protease inhibitor, is extremely potent against multidrug-resistant HIV-1 variants and has a high genetic barrier against the emergence of resistant variants. *Antimicrob. Agents Chemother*., 62, e02060-17.

De Luca A, Flandre P, Dunn D, Zazzi M, Wensing A, Santoro MM, Günthard HF, Wittkop L, Kordossis T, Garcia F, Castagna A, Cozzi-Lepri A, Churchill D, De Wit S, Fuchs W, Imaz A, Mussini C, Obel N, Perno CF, Roca B, Reiss P, Schülter E, Torti C, van Sighem A, Zangerle R and Descamps D. 2016. Improved Darunavir genotypic mutation score predicting treatment response for HIV-1 subtype B and non-B infected patients receiving a salvage regimen. *J. Antimicrob. Chemother*., 71, 1352–1360.

De Oliveira, Engelbrecht TSS, van Rensburg JE, Gordon M, Bishop K, Megede RJ, Barnett SW and 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.

Deshmukh L, Louisa JM, Ghirlando R and Clore GM. 2016. Transient HIV-1 Gag–protease interactions revealed by paramagnetic NMR suggest origins of compensatory drug resistance mutations. *Proc. Natl. Acad. Sci. USA*, 113, 12456–12461.

Dierynck I, De WM, Gustin E, Keuleers I, Vandersmissen J, Hallenberger S and Hertogs K. 2007. Binding kinetics of darunavir to human immunodeficiency virus type 1 protease explain the potent antiviral activity and high genetic barrier. *J. Virol*., 81, 13845–13851.

Dill KA, Ozkan SB, Shell MC and Weikl TR. 2008. The protein folding problem. *Annu. Rev. Biophys*., 37, 289–316.

Doherty RS, De Oliveira T, Seebregts C, Danaviah S, Gordon M and Cassol S. 2005. BioAfrica's HIV-1 proteomics resource: combining protein data with bioinformatics tools. *Retrovirol*., 2, 18.

Doherty KM, Nakka P, King BM, Rhee S-Y, Holmes SP, Shafer RW and Radhakrishnan ML. 2011. A multifaceted analysis of HIV-1 protease multidrug resistance phenotypes. *BMC Bioinformatics*, 12, 477.

Douek DC, Brenchley JM, Betts MR, Ambrozak DR, Hill BJ, Okamoto Y, Casazza JP, Kuruppu J, Kunstman K, Wolinsky S, et al. 2002. HIV preferentially infects HIV-specific CD4+ T cells. Nature, 417, 95–98.

Doyon L, Croteau G, Thibeault D, Poulin F, Pilote L and Lamarre D. 1996. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J. Virol*., 70, 3763–3769.

Dudley DM, Bailey AL, Mehta SH, Hughes AL, Kirk GD, Westergaard RP and O'Connor DH. 2014. Cross-clade simultaneous HIV drug resistance genotyping for reverse transcriptase, protease, and integrase inhibitor mutations by Illumina MiSeq. *Retrovirol*., 11, 122.

Dwyer DS. 2001. Electronic properties of the amino acid side chains contribute to the structural preferences in protein folding. *J. Biomol. Struct. Dyn*., 18, 881–892.

Dykes C and Demeter LM. 2007. Clinical significance of human immunodeficiency virus type 1 replication fitness. *Clin. Microbiol. Rev*., 20, 550–578.

Eberle J and Gürtler L. 2012. HIV types, groups, subtypes and recombinant forms: errors in replication, selection pressure and quasispecies. *Intervirol*., 55, 79–83.

Eisenberg D. 2003. The discovery of the α-helix and β-sheet, the principal structural features of proteins. *Proc. Natl. Acad. Sci. USA*, 100, 11207–11210.

Etta EM, Mavhandu L, Manhaeve C, McGonigle K, Jackson P, Rekosh D, Hammarskjold M-L, Bessong PO and Tebit DM. 2017. High level of HIV-1 drug resistance mutations in patients with unsuppressed viral loads in rural northern South Africa. *AIDS Res. Ther*., 14, 36.

Falkenhagen A and Joshi S. 2018. HIV entry and its inhibition by bifunctional antiviral proteins. *Mol. Ther. Nuc. Acids*, 13, 347–364.

Fang CT, Chang YY, Hsu HM, Twu SJ, Chen KT, Lin CC, Huang LYL, Chen MY, Hwang JS, Wang JD and Chuang CY. 2007. Life expectancy of patients with newly-diagnosed hiv infection in the era of highly active antiretroviral therapy. *QJM*, 100, 97–105.

Fares MA and Travers SAA. 2006. A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics*, 173, 9–23.

Farhadi T. 2018. Advances in protein tertiary structure prediction. *BBRJ*, 2, 20–25.

Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pepin J, Posada D, Peeters M, Pybus OG and Lemey P. 2014. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science*., 346, 56–61.

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

Fehér A, Weber IT, Bagossi P, Boross P, Mahalingam B, Louis JM, Copeland TD, Torshin IY, Harrison RW and Tözsér J. 2002. Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. *Eur. J. Biochem*., 269, 4114–4120.

Fiorentini S, Marini E, Caracciolo S and Caruso A. 2006. Functions of the HIV-1 matrix protein p17. *New Microbiol*., 29, 1–10.

Flynn WF, Haldane A, Torbett BE and Levy RM. 2017. Inference of epistatic effects leading to entrenchment and drug resistance in HIV-1 protease. *Mol. Biol. Evol*., 34, 1291–1306.

Foley B, Leitner T, Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, Wolinsky S and Korber B. 2017. HIV sequence compendium 2017. Publisher: Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico. LA-UR-17-25240.

Forster MJ, Mulloy B and Nermut MV. 2000. Molecular modelling study of HIV p17gag (MA) protein shell utilising data from electron microscopy and x-ray crystallography. *J. Mol. Biol*., 298, 841–857.

Freed EO. 1998. HIV-1 Gag proteins: diverse functions in the virus life cycle. *Virol*., 251, 1–15.

Freed EO. 2015. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol*., doi:10.1038/nrmicro3490.

Frenkel D and Smit B. 2002. Understanding Molecular Simulation. From Algorithms to Applications, 2nd Ed. Academic Press, New York.

Friedman-Kien A, Laubenstein L, Marmor M, Hymes K, Green J, Ragaz A, Gottleib J, Muggia F, Demopoulos R, Weintraub M. 1981. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. MMWR Morb. Mortal Wkly. Rep., 30, 305–308.

Fu Y, Zhao J and Chen Z. Insights into the molecular mechanisms of protein-ligand interactions by molecular docking and molecular dynamics simulation: a case of oligopeptide binding protein. *Comput. and Math. Meth. Med*., 3502514.

Fun A, Wensing AMJ, Verheyen J and Nijhuis M. 2012. Human immunodeficiency virus gag and protease: partners in resistance. *Retrovirol*., 9, 63.

Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M. Haynes BF, Palker TJ. Redfield R, Oleske J, Safai B. 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*, 224, 500-513.

Gamble TR, Yoo S, Vajdos FF, von Schwedler UK, Worthylake DK, Wang H, McCutcheon JP, Sundquist WI and Hill CP. 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science*., 278, 849–853.

Ganser-Pornillos BK, Cheng A and Yeager M. 2007. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell*, 131, 70–79.

Gatanaga H, Suzuki Y, Tsang H, Yoshimura K, Kavlick MF, Nagashima K, Gorelick RJ, Mardy S, Tang C, Summers MF and 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.

Gelderblom HR, Hausmann EH, Ozel M, Pauli G and Koch MA. 1987. Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. Virol., 156, 171–176.

Genheden S and Ryde U. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov*., 10, 449–461.

Ghosn J, Delaugerre C, Flandre P, Galimand J, Cohen-Codar I, Raffi F, Delfraissy J-F, Rouzioux C and Chaix M-L. 2011. Polymorphism in Gag gene cleavage sites of HIV-1 non-B subtype and virological outcome of a first-line lopinavir/ritonavir single drug regimen. *PLoS ONE*, 6, e24798.

Giandhari J, Basson AE, Coovadia A, Kuhn L, Abrams EJ, Strehlau R, Morris L and Hunt GM. 2015. Genetic changes in HIV-1 Gag-protease associated with protease inhibitor-based therapy failure in pediatric patients. *AIDS Res. Hum. Retroviruses*, 31, 776–782.

Giandhari J, Basson AE, Sutherland K, Parry CM, Cane PA, Coovadia A, Kuhn L, Hunt G and Morris L. 2016. Contribution of Gag and protease to HIV-1 phenotypic drug resistance in pediatric patients failing protease inhibitor-based therapy. *Antimicrob. Agents Chemother*., 6, 2248–2256.

Gidalevitz T, Prahlad V and Morimoto RI. 2011. The stress of protein misfolding: from single cells to multicellular organisms. *Cold Spring Harb. Perspect. Biol*., 3, a009704.

Girnary R, King L, Robinson L, Elston R and Brierley I. 2007. Structure-function analysis of the ribosomal frameshifting signal of two human immunodeficiency virus type 1 isolates with increased resistance to viral protease inhibitors. *J. Gen. Virol*., 88, 226–235.

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

Godbey WT. 2014. Chapter 2 – Proteins. In Godbey WT (eds.), *An Introduction to Biotechnology*. Woodhead Publishing, New Orleans, Louisiana, pp.9–33.

Godet J, Boudier C, Humbert N, Ivanyi-Nagy R, Darlix J-L and Mély Y. 2012. Comparative nucleic acid chaperone properties of the nucleocapsid protein NCp7 and Tat protein of HIV-1. *Virus Res*., 169, 349–360.

Goedert JJ and Gallo RC. 1985. Epidemiological evidence that HTLV-III is the AIDS agent. *Eur. J. Epidemiol*., 1, 155–159.

Goldfarb NE, Ohanessian M, Biswas S, McGee Jr TD, Mahon BP, Ostrov DA, Garcia J, Tang Y, McKenna R, Roitberg A and Dunn BM. 2015. Defective hydrophobic sliding mechanism and active site expansion in HIV-1 protease drug resistant variant Gly48Thr/Leu89Met: mechanisms for the loss of saquinavir binding potency. *Biochemistry*, 54, 422−433.

Gonzalez R, Masquelier B, Fleury H, Lacroix B, Troesch A, Vernet G and Telles JN. 2004. Detection of human immunodeficiency virus type 1 antiretroviral resistance mutations by high-density DNA probe arrays. *J. Clin. Microbiol*., 42, 2907–2912.

Gonzalez LMF, Santos AF, Abecasis AB, Van Laethem K, Soares EA, Deforche K, Tanuri A, Camacho R, Vandamme A-M and Soares MA. 2008. Impact of HIV-1 protease mutations A71V/T and T74S on M89I/V-mediated protease inhibitor resistance in subtype G isolates. *J. Antimicrob. Chemother*., 61, 1201–1204.

Goodchild SC, Curmi PMG and Brown LJ. 2011. Structural gymnastics of multifunctional metamorphic proteins. *Biophys. Rev*., 3, 143.

Gordon M, De Oliveira T, Bishop K, Coovadia HM, Madurai L, Engelbrecht S, van Rensburg JE, Mosam A, Smith A and Cassol S. 2003. Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies. *J. Virol*., 77, 2587–2599.

Gottlinger HG, Dorfman T, Sodroski JG and Haseltine WA. 1991. Effect of mutations affecting the p6 *gag* protein on human immunodeficiency virus particle release. *Proc. Natl. Acad. Sci*., 88, 3195–3199.

Graham MJ, Drake AJ, Djorgovski SG, Mahabal AA and Donalek C. Using conditional entropy to identify periodicity. *MNRAS*, 434,2629–2635.

Gres AT, Kirby KA, KewalRamani VN, Tanner JJ, Pornillos O and Sarafianos SG. 2015. X-ray structures of native HIV-1 capsid protein reveal conformational variability. *Science*, 349, 99–103.

Gross I, Hohenberg H, Wilk T, Wiegers K, Grättinger M, Müller B, Fuller S and Kräusslich H-G. 2000. A conformational switch controlling HIV-1 morphogenesis. *EMBO J*., 19, 103–113.

Grossman Z, Paxinos EE, Averbuch D, Maayan S, Parkin NT, Engelhard D, Lorber M, Istomin V, Shaked Y, Mendelson E, Ram D, Petropoulos CJ and Schapiro JM. 2004. Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to Nelfinavir. *Antimicrob. Agents Chemother*., 48, 2159–2165.

Grossman Z, Schapiro JM, Levy I, Elbirt D, Chowers M, Riesenberg K, Olstein-Pops K, Shahar E, Istomin V, Asher I, Gottessman B-S, Shemer Y, Elinav H, Hassoun G, Rosenberg S, Averbuch D, Machleb-Guri K, Kra-Oz Z, Radian-Sade S, Rudich H, Ram D, Maayan S, Agmon-Levin N and Sthoeger Z. 2014. Comparable long-term efficacy of Lopinavir/Ritonavir and similar drug-resistance profiles in different HIV-1 subtypes. *PLoS ONE*, 9, e86239.

Gu W-G, Zhang X and Yuan J-F. 2014. Anti-HIV drug development through computational methods. *AAPS J*., 16, 674–680.

Günthard HF and Scherrer AU. 2016. HIV-1 subtype C, Tenofovir, and the relationship with treatment failure and drug resistance. *J. Infect. Dis*., 214, 1289–1291.

Guo J, Henderson LE, Bess J, Kane B and Levin JG. 1997. Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J. Virol*., 71, 5178–5188.

Gupta A, Jamal S, Goyal S, Jain R, Wahi D and Grover A. 2015. Structural studies on molecular mechanisms of Nelfinavir resistance caused by non-active site mutation V77I in HIV-1 protease. *BMC Bioinformatics*, 16, S10.

Gupta RK, Kohli A, McCormick AL, Towers GJ, Pillay D and Parry CM. 2010. Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays. *AIDS*, 24, 1651–1655.

Haimov B and Srebnik S. 2016. A closer look into the α-helix basin. *Sci. Rep*., 6, 38341.

Hall TA. 1999. BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser*., 41, 95–98.

Haspel N, Moll M, Baker ML, Chiu W and Kavraki LE. 2010. Tracing conformational changes in proteins. *BMC Struct. Biol*., 10, S1.

Hassana AS, Pybusc OG, Sandersa EJ, Alberte J and Esbjörnsson J. 2017. Defining HIV-1 transmission clusters based on sequence data. *AIDS*, 31, 1211–1222.

Henderson GJ, Lee S-K, Irlbeck DM, Harris J, Kline M, Pollom E, Parkin N and Swanstrom R. 2011. Interplay between single resistance-associated mutations in the HIV-1 protease and viral infectivity, protease activity, and inhibitor sensitivity. *Antimicrob. Agents Chemother*., 623–633.

Henn A, Flateau C and Gallien S. 2017. Primary HIV infection: clinical presentation, testing, and treatment. *Curr. Infect. Dis. Rep*., 19, 37.

Henrich TJ and Kuritzkes DR. 2013. HIV-1 entry inhibitors: recent development and clinical use. *Curr. Opin. Virol*., 3, 51–57.

Hertogs K, Bloor S, Kemp SD, Van den Eynde C, Alcorn TM, Pauwels R, Van Houtte M, Staszewski S, Miller V and Larder BA. 2000. Phenotypic and genotypic analysis of clinical HIV-1 isolates reveals extensive protease inhibitor cross resistance: a survey of over 6000 samples. *AIDS*, 14, 1203– 1210.

Hill MK, Bellamy-McIntyre A, Vella LJ, Campbell SM, Marshall JA, Tachedjian G and Mak J. 2007. Alteration of the proline at position 7 of the HIV-1 spacer peptide p1 suppresses viral infectivity in a strain dependent manner. *Curr. HIV Res*., 5, 69–78.

Hill MK, Shehu-Xhilaga M, Crowe SM and Mak J. 2002. Proline residues within spacer peptide p1 are important for human immunodeficiency virus type 1 infectivity, protein processing, and genomic RNA dimer stability. *J. Virol*., 76, 11245–11253.

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

Ho SK, Coman RM, Bunger JC, Rose SL, O'Brien P, Munoz I, Dunn BM, Sleasman JW and Goodenow MM. 2008. Drug-associated changes in amino acid residues in Gag p2, $p7^{NC}$, and $p6^{Gag}/p6^{Pol}$ in human immunodeficiency virus type 1 (HIV-1) display a dominant effect on replicative fitness and drug response. *Virol*., 378, 272–281.

Hooper DC and Jacoby GA. 2015. Mechanisms of drug resistance: quinolone resistance. *Ann. N. Y. Acad. Sci*., 1354, 12–31.

Hou T, Zhang W, Wang J and Wang W. 2009. Predicting drug resistance of the HIV-1 protease using molecular interaction energy components. *Proteins*, 74, 837–846.

Hu L, Huang T, Liu X-J and Cai Y-D. 2011. Predicting protein phenotypes based on protein-protein interaction network. *PLoS ONE*, 6, e17668.

Hu W-S and Hughes SH. 2012. HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med*., 2, a006882.

Hu Z and Kuritzkes DR. 2014. Altered viral fitness and drug susceptibility in HIV-1 carrying mutations that confer resistance to nonnucleoside reverse transcriptase and integrase strand transfer inhibitors. *J. Virol*., 88, 9268–9276.

Huang L and Chen C. 2013. Understanding HIV-1 protease autoprocessing for novel therapeutic development. *Future Med. Chem*., 5, 11.

Huang S-W, Li W-Y, Wang W-H, Lin Y-T, Chou C-H, Chen M, Huang HD, Chen Y-H, Lu P-L, Wang S-F, Oka S, Chen YMA. 2017. Characterization of the drug resistance profiles of patients infected with CRF07_BC using phenotypic assay and ultra-deep pyrosequencing. *PLoS ONE*, 12, e0170420.

Huggins DJ and Tidor B. 2011. Systematic placement of structural water molecules for improved scoring of protein–ligand interactions. *Protein Eng. Des. Sel*., 24, 777–789.

Iacob SA, Iacob DG and Jugulete G. 2017. Improving the adherence to antiretroviral therapy, a difficult but essential task for a successful HIV treatment—clinical points of view and practical considerations. *Front. Pharmacol*., 8, 831.

Iordanskiy S, Waltke M, Feng Y and Wood C. 2010. Subtype-associated differences in HIV-1 reverse transcription affect the viral replication. *Retrovirol*., 7, 85.

Ivanov D, Stone JR, Maki JL, Collins T and Wagner G. 2005. Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. *Mol. Cell*, 17, 137–143.

Janke U, Kulke M, Buchholz I, Geist N, Langel W and Delcea M. Drug-induced activation of integrin alpha IIb beta 3 leads to minor localized structural changes. *PLoS ONE*, 14, e0214969.

Jiao D, Golubkov PA, Darden TA and Ren P. 2008. Calculation of protein–ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. USA*., 105, 6290–6295.

Kalish ML, Robbins KE, Pieniazek D, Schaefer A, Nzilambi N, Quinn TC, St. Louis ME, Youngpairoj AS, Phillips J, Jaffe HW and Folks TM. 2004. Recombinant viruses and early global HIV-1 epidemic. *Emerg. Infect. Dis*., 10, 1227–1234.

Kantor R, Zijenah LS, Shafer RW, Mutetwa S, Johnston E, Lloyd R, von Lieven A, Israelski D and Katzenstein DA. 2002. HIV-1 subtype C reverse transcriptase and protease genotypes in Zimbabwean patients failing antiretroviral therapy. *AIDS Res. Hum. Retroviruses*, 18, 1407–1413.

Kasamba I, Baisley K, Mayanja BN, Maher D and Grosskurth H. 2012. The impact of antiretroviral treatment on mortality trends of HIV-positive adults in rural Uganda: a longitudinal population-based study, 1999–2009. *Trop. Med. Int. Health*, 17, 66–73.

Katoh K and Standley DM. 2013. MAFFT multiple alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol*., 30, 772–780.

Kennedy DA and Read AF. 2017. Why does drug resistance readily evolve but vaccine resistance does not? *Proc. R. Soc. B*., 284, 20162562.

Khan SN, Persons JD, Guerrero M, Ilina TV, Oda M and Ishima R. 2018. An inhibitor-interaction intermediate of HIV-1 protease, revealed by isothermal titration calorimetry and NMR spectroscopy. *bioRxiv*. doi:http://dx.doi.org/10.1101/404996.

Kieken F, Arnoult E, Barbault F, Paquet F, Huynh-Dinh T, Paoletti J, Genest D and Lancelot G. 2002. HIV-1(Lai) genomic RNA: combined used of NMR and molecular dynamics simulation for studying the structure and internal dynamics of a mutated SL1 hairpin. *Eur. Biophys. J. Biophy*., 31, 521–531.

Kiguoya MW, Mann JK, Chopera D, Gounder K, Lee GQ, Hunt PW, Martin JN, Ball TB, Kimani J, Brumme ZL, Brockman MA and Ndung'u T. 2017. Subtype-specific differences in Gag-

protease-driven replication capacity are consistent with intersubtype differences in HIV-1 disease progression. *J. Virol*., 91, e00253.

King NM, Prabu-Jeyabalan M, Nalivaika EA and Schiffer CA. 2004. Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chem. Biol*., 11, 1333–1338.

Kitchen DB, Decornez H, Furr JR and Bajorath J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov*., 3, 935–949.

Kmiecik S, Kouza M, Badaczewska-Dawid AE, Kloczkowski A and Kolinski. 2018. Modeling of protein structural flexibility and large-scale dynamics: coarse-grained simulations and elastic network models. *Int. J. Mol. Sci*.,19, 3496.

Knops E, Brakier-Gingras L, Schülter E, Pfister H, Kaiser R and Verheyen J. 2011. Mutational patterns in the frameshift-regulating site of HIV-1 selected by protease inhibitors. *Med. Microbiol. Immunol*., 201, 213–218.

Koh Y, Das D, Leschenko S, Nakata H, Ogata-Aoki H, Amano M, Nakayama M, Ghosh AK and Mitsuya H. 2008. GRL-02031, a novel nonpeptidic protease inhibitor (PI) containing a stereochemically defined fused cyclopentanyltetrahydrofuran potent against multi-PI-resistant human immunodeficiency virus type 1 in vitro. *Antimicrob. Agents Chemother*., 53, 997–1006.

Kolli M, Lastere S and Schiffer CA. 2006. Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virol*., 347, 405-409.

Kolli M, Özen A, Kurt-Yilmaz N and Schiffer CA. 2014. HIV-1 Protease-substrate coevolution in nelfinavir resistance. *J. Virol*., 88, 7145–7154.

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

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA and Cheatham, III TE. 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res*., 33, 889–897.

Kožíšek M, Henke S, Šašková KG, Jacobs GB, Schuch A, Buchholz B, Müller V, Kräusslich H-G, Řezáčová P, Konvalinka J and Bodem J. 2012. Mutations in HIV-1 gag and pol compensate for the loss of viral fitness caused by a highly mutated protease. *Antimicrob. Agents Chemother*., 56, 4320–4330.

Kurth AE, Celum C, Baetem JE, Vermund SH and Wasserheit JN. 2011. Combination HIV prevention: significance, challenges, and opportunities. *Curr. HIV/AIDS Rep*., 8, 62–72.

Kyaw NTT, Kumar AMV, Oo MM, Oo HN, Kyaw KWY, Thiha S, Aung TK, Win T, Mon YY and Harries AD. 2017. Long-term outcomes of second-line antiretroviral treatment in an adult and adolescent cohort in Myanmar. *Global Health Action*, 10, 1290916.

Lambert-Niclot S, Flandre P, Malet I, Canestri A, Soulié C, Tubiana R, Brunet C, Wirden M, Katlama C, Calvez V and Marcelin A-G. 2008. Impact of Gag mutations on selection of darunavir resistance mutations in HIV-1 protease. *J. Antimicrob. Chemother*., 62, 905–908.

Landia A, Mazzoldi A, Andreoni C, Bianchi M, Cavallini A, Laurino M, Ricotti L, Iuliano R, Matteoli B and Ceccherini-Nelli L. 2008. Modelling and control of HIV dynamics. *Comput. Methods Programs Biomed.*, 89, 162–168.

Larrouy L, Chazallon C, Landman R, Capitant C, Peytavin G, Collin G, Charpentier C, Storto A, Pialoux G, Katlama C, Girard PM, Yeni P, Aboulker JP,Brun-Vezinet F and Descamps D. 2010. Gag mutations can impact virological response to dual-boosted protease inhibitor

combinations in antiretroviral-naïve HIV-infected patients. *Antimicrob. Agents Chemother*., 54, 2910–2919.

Lau KA and Wong JJL. 2013. Current trends of HIV recombination worldwide. *Infect. Dis. Rep*., 5, s1e4

Lessells RJ, Katzenstein DK and de Oliveira T. 2012. Are subtype differences important in HIV drug resistance? *Curr. Opin. Virol*., 2, 636–643.

Levison JH, Orrell C, Losina E, Lu Z, Freedberg KA and Wood R. 2011. Early outcomes and the virologic impact of delayed treatment switching on second-line therapy in an antiretroviral roll-out program in South Africa. *Antivir. Ther*., 16, 853–861.

Lexa KW and Carlson HA. 2011. Binding to the open conformation of HIV-1 protease: binding the open HIVp flaps. *Proteins*, 79, 2282–2290.

Li B, Fooksa M, Heinze S and Meiler J. 2017. Finding the needle in the haystack: towards solving the protein-folding problem computationally. *Crit. Rev. Biochem. Mol. Biol*., 53, 1–28.

Li G, Verheyen J, Rhee S-Y, Voet A, Vandamme A-M and Theys K. 2013. Functional conservation of HIV-1 Gag: implications for rational drug design. *Retrovirol*., 10, 126.

Li G, Verheyen J, Theys K, Piampongsant S, Van Laethem K and Vandamme A-M. 2014. HIV-1 Gag C-terminal amino acid substitutions emerging under selective pressure of protease inhibitors in patient populations infected with different HIV-1 subtypes. *Retrovirol*., 11, 79.

Lin K-H. 2016. Viral proteases as drug targets and the mechanisms of drug resistance: a dissertation. University of Massachusetts Medical School, Boston, United States.

Lingappa JR, Reed JC, Tanaka M, Chutiraka K and Robinson BA. 2014. How HIV-1 Gag assembles in cells: putting together pieces of the puzzle. *Virus Res*., 139, 89–107.

Lipkowitz KB. 1990. Reviews m computational chemistry. CH Publishers, New York, United States.

Liu F, Kovalevsky AY, Tie Y, Ghosh AK, Harrison RW and Weber IT. 2008. Effect of flap mutations on structure of HIV-1 protease and inhibition by Saquinavir and Darunavir. *J. Mol. Biol*., 381, 102–115.

Lockbaum GJ, Leidner F, Rusere LN, Henes M, Kosovrasti K, Nalivaika EA, Ali A, Yilmaz NK and Schiffer CA. 2018. Structural adaptation of Darunavir analogs against primary resistance mutations in HIV-1 protease. *BioRxiv*. doi:http://dx.doi.org/10.1101/406637.

Lockhat HA, Silva J, Alves CN, Govender T, Lameira J, Maguire GEM, Sayed Y and Kruger HG. 2016. Binding free energy calculations of nine FDA approved protease inhibitors against HIV-1 subtype C I36T↑T containing 100 amino acids per monomer. *Chem. Biol. Drug Des*., 87, 487–498.

Lopalco L. 2010. CCR5: from natural resistance to a new anti-HIV strategy. *Viruses*, 2, 574-600.

Lopes CAF, Soares MA, Falci DR and Sprinz E. 2015. The evolving genotypic profile of HIV-1 mutations related to antiretroviral treatment in the north region of Brazil. *BioMed Res. Int*., 738528.

Louis JM, Tözsér J, Roche J, Matúz K, Aniana A and Sayer JM. 2013. Enhanced stability of monomer fold correlates with extreme drug resistance of HIV-1 protease. *Biochemistry*, 52, 7678–7688.

Louis JM, Zhang Y, Sayer JM, Wang Y-F, Harrison RW and Weber IT. 2011. Drug resistance mutation L76V decreases the dimer stability and rate of autoprocessing of HIV-1 protease by reducing internal hydrophobic contacts. *Biochemistry*, 50, 4786–4795.

Lu K, Heng X and Summers F. 2011. Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol*., 410, 609–633.

Lu X, Zhao H, Zhang Y, Wang W, Zhao C, Li Y, Ma L, Cui Z and Chen S. 2017. HIV-1 drug-resistant mutations and related risk factors among HIV-1-positive individuals experiencing treatment failure in Hebei Province, China. *AIDS Res. Ther*., 14, 4.

Lv Z, Cju Y and Wang Y. 2015. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV/AIDS Res. Palliative Care*, 7, 95–104.

Lynch RM, Shen T, Gnanakaran S and Derdeyn CA. 2009. Appreciating HIV type 1 diversity: subtype differences in Env. *AIDS Res. Hum. Retroviruses*, 25, 237–248.

Maguire MF, Guinea R, Griffin P, Macmanus S, Elston RC, Wolfram J, Richards N, Hanlon MH, Porter DJ, Wrin T, Parkin N, Tisdale M, Furfine E, Petropoulos C, Snowden BW and Kleim JP. 2002. Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J. Virol*., 76, 7398–7406.

Maillard PV, Zoete V, Michielin O and Trono D. 2011. Homology-based identification of capsid determinants that protect HIV1 from human TRIM5α restriction. *J. Biol. Chem*., 286, 8128–8140.

Mailler E, Bernacchi S, Marquet R, Paillart J-C, Vivet-Boudou V and Smyth RP. 2016. The life-cycle of the HIV-1 Gag–RNA complex. *Viruses*, 8, 248.

Malet I, Roquebert B, Dalban C, Wirden M, Amellal B, Agher R, Simon A, Katlama C, Costagliola D, Calvez V, Marcelin AG. 2007. Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. *J. Infect*., 54, 367–374.

Mammano F, Trouplin V, Zennou V and Clavel F. 2000. Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. *J. Virol*., 74, 8524–8531.

Marie V and Gordon M. 2019. Gag-protease coevolution shapes the outcome of lopinavir-inclusive treatment regimens in chronically infected HIV-1 subtype C patients. *Bioinformatics*. doi:10.1093/bioinformatics/btz076.

Martinez-Cajas JL, Pai NP, Klein MB and Wainberg MA. 2009. Differences in resistance mutations among HIV-1 non-subtype B infections: a systematic review of evidence (1996–2008). *J. Int. AIDS Society*, 12, 11.

Martinez-Picado J, Savara AV, Sutton L and D'aquila. 1999. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J. Virol*., 73, 3744–3752.

Masse S, Lu X, Dekhtyar T, Lu L, Koev G, Gao F, Mo H, Kempf D, Bernstein B, Hanna GJ and Molla A. 2007. In vitro selection and characterization of human immunodeficiency virus type 2 with decreased susceptibility to Lopinavir. *Antimicrob. Agents Chemother*., 51, 3075–3080.

Matsunaga S, Masaoka T, Sawasaki T, Morishita R, Iwatani Y, Tatsumi M, Endo Y, Yamamoto N, Sugiura W and Ryo A. 2015. A cell-free enzymatic activity assay for the evaluation of HIV-1 drug resistance to protease inhibitors. *Front. Microbiol*., 6, 1220.

Mattei S, Tana A, Glass B, Müller B, Kräusslich H-G and Briggs JAG. 2018. High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation. *Proc. Natl. Acad. Sci. USA*, 115, 9401–9410.

Meintjes G, Moorhouse MA, Carmona S, Davies N, Dlamini S, van Vuuren C, Manzini T, Mathe M, Moosa Y, Nash J, Nel J, Pakade Y, Woods J, Van Zyl G, Conradie F and Venter F. 2017. Adult antiretroviral therapy guidelines 2017. *South Afr. J. HIV. Med*., 18, a776.

McCormick-Davis C, Dalton SB, Singh DK and Stephens EB. 2000. Comparison of Vpu sequences from diverse geographical isolates of HIV type 1 identifies the presence of highly

variable domains, additional invariant amino acids, and a signature sequence motif common to subtype C isolates. *AIDS Res. Hum. Retroviruses*, 16, 1089–1095.

McDermott J, Farrell L, Ross R and Barklis E. 1996. Structural analysis of human immunodeficiency virus type 1 Gag protein interactions, using cysteine-specific reagents. *J. Virol*., 70, 5106–5114.

McKinnon JE, Delgado R, Pulido F, Shao W, Arribas JR and Mellors JW. 2011. Single genome sequencing of HIV-1 Gag and protease to assess protease inhibitor resistance at virologic failure of simplified vs. standard maintenance therapy with Lopinavir/Ritonavir in the OK04 trial. *Antivir. Ther*., 16, 725–732.

Nakashima M, Ode H, Suzuki K, Fujino M, Maejima M, Kimura Y, Masaoka T, Hattori J, Matsuda M, Hachiya A, Yokomaku Y, Suzuki A, Watanabe N, Sugiura W and Iwatani Y. 2016. Unique flap conformation in an HIV-1 protease with high-level darunavir resistance. *Front. Microbiol*., 7, 61.

Ndung'u T, Renjifo B and Essex M. 2001. Construction and analysis of an infectious human immunodeficiency virus type 1 subtype C molecular clone. *J. Virol*., 75, 4964–4972.

Midde NM, Patters BJ, Rao PSS, Cory TJ and Kumar S. 2016. Investigational protease inhibitors as antiretroviral therapies. *Expert Opin. Investig. Drugs*, 25, 1189–1200.

Mo H, Parkin N, Stewart KD, Lu L, Dekhtyar T, Kempf DJ and Molla A. 2007. Identification and structural characterization of I84C and I84A mutations that are associated with high-level resistance to human immunodeficiency virus protease inhibitors and impair viral replication. *Antimicrob. Agents Chemother*., 51, 732–735.

Mudgal MM, Birudukota N and Doke MA. 2018. Applications of click chemistry in the development of HIV protease inhibitors. *Int. J. Med. Chem*. doi:https://doi.org/10.1155/2018/2946730.

Muhammed MT and Aki-Yalcin E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem. Biol. Drug Des*., 93, 12–20.

Myint L, Matsuda M, Matsuda Z, Yokomaku Y, Chiba T, Okano A, Yamada K and Sugiura W. 2004. Gag non-cleavage site mutations contribute to full recovery of viral fitness in protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob. Agents Chemother*., 48, 444–452.

Myllymäki P, Silander T, Tirri H and Uronen P. 2002. B-Course: a web-based tutorial for Bayesian and causal data analysis. *Int. J. Art. Intell. Tools*, 11, 396–387.

NDOH. 2019. 2019 ART clinical guidelines for the management of HIV in adults, pregnancy, adolescents, children, infants and neonates. National Department of Health, South Africa, pp. 1–24.

Nayak C, Chandra I, Singh SK. 2019. An *in silico* pharmacological approach toward the discovery of potent inhibitors to combat drug resistance HIV-1 protease variants. *J. Cell. Biochem.*, 120, 9063–9081.

Ngo-Giang-Huonga N and Aghokeng AF. 2019. HIV drug resistance in resource-limited countries: threat for HIV elimination. *EClinicalMedicine*, 9, 3–4.

Nguyen QH, Contamin L, Nguyen TVA, A-L Bañuls. 2018. Insights into the processes that drive the evolution of drug resistance in Mycobacterium tuberculosis. *Evol. Appl*., 11, 1498–1511.

NIH. 2018. Drugs that fight HIV: a reference guide for prescription HIV-1 medications. United States Department of Health and Human Services, 18, 7628.

Nijhuis M, van Maarseveen NM, Lastere S, Schipper P, Coakley E, Glass B, Rovenska M, de Jong D, Chappey C, Goedegebuure IW, Heilek-Snyder G, Dulude D, Cammack N, Brakier-

Gingras L, Konvalinka J, Parkin N, Kräusslich H-G, Brun-Vezinet F, Boucher CAB. 2007. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med*., 4, e36.

Nijhuis M, Schuurman R, de Jong D, Erickson J, Gustchina E, Albert J, Schipper P, Gulnik S and Boucher CA. 1999. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS*, 1999, 13, 2349–2359.

Nijhuis M, Wensing AMJ, Bierman WFW, de Jong D, Kagan R, Fun A, Jaspers CAJJ, Schurink KAM, van Agtmael MA and Boucher CAB. 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.

Ning J, Erdemci-Tandogan G, Yufenyuy EL, Wagner J, Himes BA, Zhao G, Aiken C, Zandi R and Zhang P. 2016. *In vitro* protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway. *Nat. Commun*., 7, 13689.

Nisole S and Saïb A. 2004. Early steps of retrovirus replicative cycle. *Retrovirol*., 1, 9.

Novikova M, Adams LJ, Fontana J, Gres AT, Balasubramaniam M, Winkler DC, Kudchodkar SB, Soheilian F, Sarafianos SG, Steven AC and Freed EO. 2018. Identification of a structural element in HIV-1 Gag required for virus particle assembly and maturation. *mBio*, 9, e01567-18.

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

Özen A, Haliloğlu T and Schiffer CA. 2012. HIV-1 protease and substrate coevolution validates the substrate envelope as the substrate recognition pattern. *J. Chem. Theory Comput*., 8, 2.

Özen A, Lin K-H, Yilmaz NK and Schiffer CA. 2014. Structural basis and distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease. *Proc. Natl. Acad. Sci. USA*, 111, 15993–15998.

Pagadala NS, Syed K and Tuszynski J. 2017. Software for molecular docking: a review. *Biophys. Rev*., 9, 91–102.

Pandrea I, Sodora DL, Silvestri G and Apetrei C. 2008. Into the wild: simian immunodeficiency virus (SIV) infection in natural hosts. *Trends Immunol*., 29, 419–428.

Paraskevis D, Magiorkinis E, Magiorkinis G, Kiosses VG, Lemey P, Vandamme A-M, Rambaut A and Hatzakis A. 2004. Phylogenetic reconstruction of a known HIV-1 CRF04_cpx transmission network using maximum likelihood and Bayesian methods. *J. Mol. Evol*., 59, 709–717.

Paredes R, Tzou PL, van Zyl G, Barrow G, Camacho R, Carmona S, Grant PM, Gupta RK, Hamers RL, Harrigan PR, Jordan MR, Kantor R, Katzenstein DA, Kuritzkes DR, Maldarelli F, Otelea D, Wallis CL, Schapiro JM and Shafer RW. 2017. Collaborative update of a rule-based expert system for HIV-1 genotypic resistance test interpretation. *PLoS ONE*, 12, e0181357.

Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, Laughlin CA, Saif LJ and Daszak P. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev*., 72, 457–470.

Parry CM, Kohli A, Boinett CJ, Towers GJ, McCormick AL and Pillay D. 2009. Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1. *J. Virol*., 83, 9094–9101.

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

Paulsen JL, Leidner F, Ragland DA, Yilmaz NK and Schiffer CA. 2017. Interdependence of inhibitor recognition in HIV-1 protease. *J. Chem. Theory Comput*., 13, 2300−2309.

Pearl J. 1998. Graphical models for probabilistic and causal reasoning. In Gabbay, D.M. and Smets, P. (eds), *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 1: Quantified Representation of Uncertainty and Imprecision*. Kluwer Academic Publishers, Dordrecht, pp. 367–389.

Peeters M, Mundeke SA, Ngole EM and Delaporte E. 2010. Origin of HIV/AIDS and risk for ongoing zoonotic transmissions from nonhuman primates to humans. HIV Ther., 4, 387–390.

Pelley JW. 2007. Protein structure and function. In Pelley JW (eds.), *Elsevier's Integrated Biochemistry*. Mosby, Lubbock, Texas, pp.19–28.

Perricone U, Gulotta MR, Lombino J, Parrino B, Cascioferro S, Diana P, Cirrincione G and Padova A. An overview of recent molecular dynamics applications as medicinal chemistry tools for the undruggable site challenge. Med. Chem. Commun., 9, 920–936.

Pessôa R and Sanabani SS. 2017. High prevalence of HIV-1 transmitted drug resistance mutations from proviral DNA massively parallel sequencing data of therapy naïve chronically infected Brazilian blood donors. *PLoS ONE,* 12, e0185559.

Piana S, Carloni P and Rothlisberger U. 2002. Drug resistance in HIV-1 protease: flexibility-assisted mechanism of compensatory mutations. *Protein Sci*., 11, 2393–2402.

Pillay K. 2015. The impact of the p7/p1 cleavage site mutations on replication capacity and drug resistance in HIV-1 subtype C. Master of Science, University of KwaZulu-Natal, South Africa.

Pillay SK, Singh U, Singh A, Gordon M and Ndungu T. 2014. Gag drug resistance mutations in HIV-1 subtype C patients, failing a protease inhibitor inclusive treatment regimen, with detectable lopinavir levels. *J. Int. AIDS Soc*., 17, 19784.

Ponnuraj K and Saravanan KM. 2017. Dihedral angle preferences of DNA and RNA binding amino acid residues in proteins. *Int. J. Biol. Macromol*., 97, 434-439.

Porter LL and Looger LL. 2018. Extant fold-switching proteins are widespread. *Proc. Natl. Acad. Sci. USA*, 115, 5968–5973.

Prabu-Jeyabalan M, Nalivaika EA, King NM and Schiffer CA. 2004. Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J. Virol*., 78, 12446–12454.

Prabu-Jeyabalan M, Nalivaika EA, Romano K and Schiffer CA. 2006. Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *J. Virol*., 80, 3607–3616.

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

Prado JG, Wrin T, Beauchaine J, Ruiz L, Petropoulos CJ, Frost SD, Clotet B, D'Aquila RT and Martinez-Picado J. 2002. Amprenavir-resistant HIV-1 exhibits Lopinavir cross-resistance and reduced replication capacity. *AIDS*, 16, 1009–1017.

Prieto-Martínez FD, Arciniega M and Medina-Franco JL. 2018. Molecular docking: current advances and challenges. *TIP Rev. Esp. Cienc. Quím. Biol*., 21, 1–23.

Qian J, Liu Y, Chao N, Ma C, Chen Q, Sun J and Wu Y. 2017. Positive selection and functional divergence of farnesyl pyrophosphate synthase genes in plants. *BMC Mol*. Biol., 18, 3.

Quiñones-Mateus ME and Arts E. 2003. Fitness of drug resistant HIV-1: Methodology and clinical implications. *Drug Resist. Updat*., 5, 224–233.

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

Race E. 2001. Cross-resistance within the protease inhibitor class. *Antivir. Ther*., 6, 29–36.

Ragland DA, Nalivaika EA, Nalam MNL, Prachanronarong KL, Cao H, Bandaranayake RM, Cai Y, Kurt-Yilmaz N and Schiffer CA. 2014. Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease. *J. Am. Chem. Soc*., 136, 11956−11963.

Rao VS, Srinivas K, Sujini GN and Kumar GN. 2014. Protein-protein interaction detection: methods and analysis. *Int. J. Proteomics*, 2014, 147648.

Raugi DN, Smith RA and Gottlieb GS. 2016. Four amino acid changes in HIV-2 protease confer class-wide sensitivity to protease inhibitors. *J. Virol*., 90, 1062–1069.

Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J and Shafer RW. 2003. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res*., 31, 298–303.

Rhee S-Y, Taylor J, Fessel WJ, Kaufman D, Towner W, Troia P, Ruane P, Hellinger J, Shirvani V, Zolopa A and Shafer RW. 2010. HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrob. Agents Chemother*., 54, 4253–4261.

Richman DD, Little SJ, Smith DM, Wrin T, Petropoulos C and Wong JK. 2004. HIV evolution and escape. Trans. *Am. Clin. Climatol. Assoc*., 115, 289–303.

Robb ML, Eller LA, Kibuuka H, Rono K, Maganga L, Nitayaphan S, Kroon E, Sawe FK, Sinei S, Sriplienchan S, Jagodzinski LL, Malia J, Manak M, de Souza MS, Tovanabutra S, Sanders-Buell E, Rolland M, Dorsey-Spitz J, Eller MA, Milazzo M, Li Q, Lewandowski A, Wu H, Swann E, O'Connell RJ, Peel S, Dawson P, Kim JH and Michael NL. 2016. Prospective study of acute HIV-1 infection in adults in East Africa and Thailand. *N. Engl. J. Med*., 374, 2120–2130.

Rodenburg CM, Li Y, Trask SA, Chen Y, Decker J, Robertson DL, Kalish ML, Shaw GM, Allen S, Hahn BH and Gao F. Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. AIDS Res. Hum. Retroviruses, 17, 161-168.

Rodriguez-Rodriguez L, Tsuchihashi Z, Fuentes GM, Bambara RA and Fay PJ. 1995. Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro. *J. Biol. Chem*., 270, 15005–15011.

Romero PE, Weigand AM and Pfenninger M. 2016. Positive selection on panpulmonate mitogenomes provide new clues on adaptations to terrestrial life. *BMC Evol. Biol*., 16, 164.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hönha S, Larget B, Liu L, Suchard MA and Hulesenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol*., 61, 539–542.

Ryckaert JP, Ciccoti C and Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-Alkanes. *J. Comput. Phys*., 23, 327–341.

Sael L, Chitale M and Kihara D. 2012. Structure- and sequence-based function prediction for non-homologous proteins. *J. Struct. Funct. Genomics*, 13, 111–123.

Saleh NA, Elhaes H and Ibrahim M. 2017. Chapter 2 – Design and development of some viral protease inhibitors QSAR and molecular modeling studies. In Gupta SP (eds) *Viral Proteases and Their Inhibitors*. Elsevier B.V. Publishers, Amsterdam, pp. 25–58.

Salmaso V. 2018. Exploring protein flexibility during docking to investigate ligand-target recognition. Doctor of Philosophy, University of Padova, Padova.

Salmaso V and Moro S. 2018. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. *Front. Pharmacol*., 9, 923.

Santoro MM and Perno CF. 2013. HIV-1 genetic variability and clinical implications. *ISRN Microbiol*., 481314.

Santos AF and Soares MA. 2010. HIV genetic diversity and drug resistance. *Viruses*, 2, 503–531.

Saravanan KM and Selvaraj S. 2017. Dihedral angle preferences of amino acid residues forming various non-local interactions in proteins. *J. Biol. Phys*., 43, 265–278.

Scherrer AU, Ledergerber B, von Wyl V, Boni J, Yerly S, Klimkait T, Cellerai C, Furrer H, Calmy A, Cavassini M, Elzi L, Vernazza PL, Bernasconi E and Günthard HF. 2012. Minor protease inhibitor mutations at baseline do not increase the risk for a virological failure in HIV-1 subtype B infected patients. *PLoS ONE*, 7, e37983.

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

Shafer RW. 2017. Human immunodeficiency virus type 1 drug resistance mutations update. *J. Infect. Dis*., 216, 843–846.

Shafer RW, Rhee S-Y, Pillay D, Miller V, Sandstrom P, Schapiro JM, Kuritzkes DR and Bennett D. 2007. HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS*, 21, 215–223.

Shah S, Alexaki A, Pirrone V, Dahiya S, Nonnemacher MR and Wigdahl B. 2014. Functional properties of the HIV-1 long terminal repeat containing single-nucleotide polymorphisms in Sp site III and CCAAT/enhancer binding protein site I. *Virol. J*., 11, 92.

Sham HL, Kempf DJ, Molla A, Marsh KC, Kumar GN, Chen C-M, Kati W, Stewart K, Lal R, Hsu A, Betebenner D, Korneyeva M, Vasavanonda S, McDonald E, Saldivar A, Wideburg N, Chen X, Niu P, Park C, Jayanti V, Grabowski B, Granneman GR, Sun E, Japour AJ, Leonard JM, Plattner JJ and Norbeck DW. 1998. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrob. Agents Chemother*., 42, 3218–3224.

Sharp PM and Hahn BH. 2011. Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* 1, a006841.

Shen C, Yu X, Harrison RW and Weber IT. 2016. Automated prediction of HIV drug resistance from genotype data. *BMC Bioinformatics*, 17, 278.

Shibata J, Sugiura W, Ode H, Iwatani Y, Sato H, Tsang H, Matsuda M, Hasegawa N, Ren F and Tanaka H. 2011. Within-host co-evolution of Gag P453L and protease D30N/N88D demonstrates virological advantage in a highly Protease inhibitor-exposed HIV-1 case. *Antiviral Res*., 90, 33– 41.

Silver N, Paynter M, McAllister G, Atchley M, Sayir C, Short J, Winner D, Alouani DJ, Sharkey FH, Bergefall K, Templeton K, Carrington D and Quiñones-Mateu ME. 2018. Characterization of minority HIV-1 drug resistant variants in the United Kingdom following the verification of a deep sequencing-based HIV-1 genotyping and tropism assay. *AIDS Res. Ther*., 15, 18.

Sing T and Däumer M. 2006. Chapter 6 Interpretation algorithms. In Geretti AM (eds.), *Antiretroviral Resistance in Clinical Practice*. Mediscript, London, United Kingdom.

Singh S. 2015. Acquired and transmitted drug resistance in HIV-1 subtype C: implications of novel mutations on replication capacity, cleavage and drug susceptibility. Doctor of Philosophy, University of KwaZulu-Natal, South Africa.

Smith RL, de Boer R, Brul S, Budovskaya Y and van der Spek H. 2013. Premature and accelerated aging: HIV or HAART? *Front. Gen.*, 3, 328.

Soares RO, Torres PHM, da Silva ML and Pascutti PG. 2016. Unraveling HIV protease flaps dynamics by constant pH molecular dynamics simulations. *J. Struc. Biol*., 195, 216–226.

Solbak SMO, Reksten TR, Hahn F, Wray V, Henklein P, Henklein P, Halskau O, Schubert U and Fossen T. 2013. HIV-1 p6 – a structured to flexible multifunctional membrane-interacting protein. *Biochimica et Biophysica Acta*, 1828, 816–823.

Soltis DE and Soltis PS. 2003. The role of phylogenetics in comparative genetics. *Plant Physiol*., 132, 1790–1800.

Spearman P. 2016. HIV-1 Gag as an antiviral target: development of assembly and maturation inhibitors. *Curr. Top. Med. Chem*., 16, 1154–1166.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.

STATS SA. 2018. Mid-year population estimates. Statistics South Africa, Pretoria, South Africa.

Stray KM, Callebaut C, Glass B, Tsai L, Xu L, Müller B, Kräusslich H-G and Cihlar T. 2013. Mutations in multiple domains of Gag drive the emergence of in vitro resistance to the phosphonate-containing HIV-1 protease inhibitor GS-8374. *J. Virol*., 87, 454–463.

Stys D, Blaha I and Strop P. 1993. Structural and functional studies in vitro on the p6 protein from the HIV-1 *gag* open reading frame. *Biochim. Biophys. Acta*, 1182, 157–161.

Su CT-T, Koh DW-S and Gan SK-E. 2019. Reviewing HIV-1 Gag mutations in protease inhibitors resistance: insights for possible novel Gag inhibitor designs. *Molecules*, 24, 3243.

Su CT-T, Kwoh C-K, Verma CS and Gan SK-E. 2018. Modeling the full-length HIV-1 Gag polyprotein reveals the role of its p6 subunit in viral maturation and the effect of non-cleavage site mutations in protease drug resistance. *J. Biomol. Struct. Dyn*., 16, 4366–4377.

Subramani A and Floudas CA. Structure prediction of loops with fixed and flexible stems. *J. Phys. Chem. B*., 116, 6670–6682.

Suguna K, Padlan EA, Smith CW, Carlson WD and Davies DR. 1987. Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action. *Proc. Natl. Acad. Sci. USA*, 84, 7009–7013.

Sui H, Gui T, Jia L, Guo W, Han J, Liu Y, Bao Z, Li H, Li J and Li L. 2014. Different frequencies of drug resistance mutations among HIV-1 subtypes circulating in China: a comprehensive study. *PLoS ONE*, 9, e91803.

Summers MF, Henderson LE, Chance MR, Bess JW Jr, South TL, Blake PR, Sagi I, Perez-Alvarado G, Sowder RC 3rd, Hare DR and Arthur LO. 1992. Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci*., 1, 563–574.

Sundquist WI and Kräusslich H-G. 2012. HIV-1 assembly, budding and maturation. *Cold Spring Harb. Perspect. Med*., 2, a006924.

Sutherland KA, Mbisa JL, Ghosn J, Chaix M-L, Cohen-Codar I, Hue S, Delfraissy J-F, Delaugerre C and Gupta RK. 2014. Phenotypic characterization of virological failure following Lopinavir/Ritonavir monotherapy using full-length gag–protease genes. *J. Antimicrob. Chemother*., 69, 3340–3348.

Sylla M, Chamberland A, Boileau C, Traoré HA, Ag-Aboubacrine S, Cissé M, Koala S, Drabo J, Diallo I, Niamba P, Tremblay-Sher D, Machouf N, Rashed S, Nickle DC, Nguyen VK and Tremblay CL. 2008. Characterization of drug resistance in antiretroviral-treated patients

infected with HIV-1 CRF02_AG and AGK subtypes in Mali and Burkina Faso. *Antivir. Ther*., 13, 141–148.

Tamiya S, Mardy S, Kavlick MF, Yoshimura K, and 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.

Tang MW and Shafer RW. 2012. HIV-1 antiretroviral resistance scientific principles and clinical applications. *Drugs*, 72, e1-e25.

Tarasova O, Poroikov V and Veselovsky A. 2018. Molecular docking studies of HIV-1 resistance to reverse transcriptase inhibitors: mini-review. *Molecules*, 23, 1233–1249.

Taylor BS, Hunt G, Abrams EJ, Coovadia A, Meyers T, Sherman G, Strehlau R, Morris L and Kuhn L. 2011. Rapid development of antiretroviral drug resistance mutations in HIV-infected children less than two years of age initiating protease inhibitor-based therapy in South Africa. *AIDS Res. Human Retroviruses*, 27, 945–956.

Teilum K, Olsen JG and Kragelund BB. 2009. Functional aspects of protein flexibility. *Cell. Mol. Life Sci*., 66, 2231–2247.

Teto G, Tagny CT, Mbanya D, Fonsah JY, Fokam J, Nchindap E, Kenmogne L, Njamnshi AK and Kanmogne GD. 2017. Gag P2/NC and pol genetic diversity, polymorphism, and drug resistance mutations in HIV-1 CRF02_AG- and non-CRF02_AG infected patients in Yaoundé, Cameroon. *Sci. Reports*, 7, 14136.

Thierry S, Munir S, Thierry E, Subra F, Leh H, Zamborlini A, Saenz D, Levy DN, Lesbats P, Saïb A, Parissi V, Poeschla E, Deprez E and Delelis O. 2015. Integrase inhibitor reversal dynamics indicate unintegrated HIV-1 DNA initiate *de novo* integration. *Retrovirol*., 12, 24.

Thukral L, Shenoy SR, Bhushan K and Jayaram B. 2007. *ProRegIn*: a regularity index for the selection of native-like tertiary structures of proteins. *J. Biosci*., 32, 71.

Tie Y, Boross PI, Wang Y-F, Gaddis L, Liu F, Chen X, Tozser J, Harrison RW and Weber IT. 2005. Molecular basis for substrate recognition and drug resistance from 1.1 to 1.6 Å resolution crystal structures of HIV-1 protease mutants with substrate analogs. *FEBS J*., 272, 5265–5277.

Titanji BH, Aasa-Chapman M, Pillay D and Jolly C. 2013. Protease inhibitors effectively block cell-to-cell spread of HIV-1 between T cells. *Retrovirol*., 10, 161.

Todd MJ, Semo N and Freire E. 1998. The structural stability of the HIV-1 protease. *J. Mol. Biol*., 283, 475–488.

Tomasini MD, Johnson DS, Mincer JS and Simon SM. 2018. Modeling the dynamics and kinetics of HIV-1 Gag during viral assembly. *PLoS ONE*, 13, e0196133.

Tong J, Wu Y and Bai M and Zhan P. 2017. 3D-QSAR and molecular docking studies on HIV protease inhibitors. *J. Mol. Struct*., 1129, 17–22.

Torrecilla E, Delicado TL and Holguín A. 2014. New findings in cleavage sites variability across groups, subtypes and recombinants of human immunodeficiency virus type 1. *PLoS ONE*, 9, e88099.

Toth G and Borics A.2006. Flap opening mechanism of HIV-1 protease. *J. Mol. Graph. Model*., 24, 465–474.

Travers SAA, Tully DC, McCormack GP and Fares MA. 2007. A study of the coevolutionary patterns operating within the env gene of the HIV-1 group M subtypes. *Mol. Biol. Evol*., 24, 2787–2801.

Tremblay CL. 2008. Combating HIV resistance – focus on darunavir. *Ther. Clin. Risk Manag.*, 4, 759–765

Trott O and Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.*, 31, 455–461.

Trylska J, Tozzini V, Chang CA and McCammon JA. 2007. HIV-1 Protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophys. J.*, 92, 4179–4187.

Tsai H-C, Chen I-T, Wu K-S, Tseng Y-T, Sy C-L, Chen J-K, Lee SS-J and Chen Y-S. 2017. High rate of HIV-1 drug resistance in treatment failure patients in Taiwan, 2009–2014. *Infect. Drug Resist.*, 10, 343–352.

UNAIDS (United Nations AIDS). 2018. Global HIV and AIDS statistics – 2018 fact sheet. https://www.unaids.org/en/resources/fact-sheet. [Accessed 20th June].

van Maarseveen NM, Andersson D, Lepšík, Fun A, Schipper PJ, de Jong D, Boucher CAB and Nijhuis M. 2012. Modulation of HIV-1 Gag NC/p1 cleavage efficiency affects protease inhibitor resistance and viral replicative capacity. *Retrovirol*., 9, 29.

Van Zyl GU, Liu TF, Claassen M, Engelbrecht S, de Oliveira T, Preiser W, Wood NT, Travers S, Shafer RW. 2013. Trends in genotypic HIV-1 antiretroviral resistance between 2006 and 2012 in South African patients receiving first- and second-line antiretroviral treatment regimens. *PLoS ONE*, 8, e67188.

Vasavi CS, Tamizhselvi R and Munusami P. 2017. Drug resistance mechanism of L10F, L10F/N88S and L90M mutations in CRF01_AE HIV-1 protease: molecular dynamics simulations and binding free energy calculations. *J. Mol. Graph. Model*., 75, 390–402.

Veenhoff LM, Heuberger EHML and Poolman B. 2002. Quaternary structure and function of transport proteins. *Trends Biochem. Sci*., 27, 242–249.

Velazquez-Campoy A, Muzammil S, Ohtaka H, Schon A, Vega S and Freire E. 2003. Structural and thermodynamic basis of resistance to HIV-1 protease inhibition: implications for inhibitor design. *Curr. Drug Targets Infect. Disord*., 3, 311–328.

Venkatakrishnan B, Palii M-L, Agbandje-McKenna M and McKenna R. 2012. Mining the protein data bank to differentiate error from structural variation in clustered static structures: an examination of HIV protease. *Viruses*, 4, 348–362.

Verheyen J, Knops E, Kupfer B, Hamouda O, Somogyi S, Schuldenzucker U, Hoffmann D, Kaiser R, Pfister H and Kucherer C. 2009. Prevalence of C-terminal *gag* cleavage site mutations in HIV from therapy-naıve patients. *J. Infect.*, 58, 61-67.

Verheyen J, Litau E, Sing T, Däumer M, Balduin M, Oette M, Fätkenheuer G, Rockstroh JK, Schuldenzucker U, Hoffmann D, Pfister H and 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.

Vora J, Patel S, Sinha S, Sharma S, Srivastava A, Chhabria M and Shrivastava N. Molecular docking, QSAR and ADMET based mining of natural compounds against prime targets of HIV. *J. Biomol. Struct. Dyn.*, 37, 131–146.

Votteler J, Neumann L, Hahn S, Hahn F, Rauch P, Schmidt K, Studtrucker N, Solbak SMO, Fossen T, Henklein P, Ott DE, Holland G, Bannert N and Schubert U. 2011. Highly conserved serine residue 40 in HIV-1 p6 regulates capsid processing and virus core assembly. *Retrovirol*., 8, 11.

Wainberg MA and Brenner BG. 2012. The impact of HIV genetic polymorphisms and subtype differences on the occurrence of resistance to antiretroviral drugs. *Mol. Biol. Int*., 256982.

Wallis CL, Mellors JW, Venter WDF, Sanne I and Stevens W. 2011. Protease inhibitor resistance is uncommon in HIV-1 subtype C infected patients on failing second-line Lopinavir/r-containing antiretroviral therapy in South Africa. *AIDS Res. Treatment*, 769627.

Warnke D, Barreto J and Temesgen Z. 2007. Antiretroviral drugs. *J. Clin. Pharmacol*., 47, 1570–1579.

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL and Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nat*., 460, 711–716.

Webb B and Sali A. 2014. Protein structure modeling with MODELLER. *Methods Mol. Biol*., 1137, 1–15.

Weber IT. 1990. Evaluation of homology modeling of HIV protease. *Proteins*, 7, 172–184.

Weber IT and Agniswamy J. 2009. HIV-1 protease: structural perspectives on drug resistance. *Viruses*, 1, 1110-1136.

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

Wensing AM, Calvez V, Günthard HF, Johnson VA, Paredes R, Pillay D, Shafer RW and Richman DD. 2017. 2017 Update of the Drug Resistance Mutations in HIV-1. *Top. Antivir. Med*., 24, 132–133.

Wensing AMJ, van Maarseveen NM and Nijhuis M. 2010. Fifteen years of HIV Protease inhibitors: raising the barrier to resistance. *Antiviral. Res*., 85, 59–74.

Williams A, Basson A, Achilonu I, Dirr HW, Morris L and Sayed Y. 2019. Double trouble? Gag in conjunction with double insert in HIV protease contributes to reduced DRV susceptibility. *Biochem. J*., 476, 375–384.

Williams KC and Burdo TH. 2009. HIV and SIV infection – the role of cellular restriction and immune responses in viral replication and pathogenesis. *APMIS*, 117, 400–412.

Wills JW and Craven RC. 1991. Form, function, and use of retroviral *gag* proteins. *AIDS*, 5, 639–654.

Wlodawer A and Gustchina A. 2000. Structural and biochemical studies of retroviral proteases. *Biochim. Biophys. Acta*, 1477, 16–34.

Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J and Kent SB. 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Sci*., 245, 616–621.

Wong-Sam A, Wang Y-F, Zhang Y, Ghosh AK, Harrison RW and Weber IT. 2018. Drug resistance mutation L76V alters nonpolar interactions at the flap-core interface of HIV-1 protease. *ACS Omega*, 3, 12132–12140.

Wright DW. 2011. Molecular dynamics simulation of drug resistance in HIV-1 protease and reverse transcriptase. Doctor of Philosophy, University College London, United Kingdom.

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

Wu X, Yang G and Zhou L. 2015. The multi-factor coupled protein folding: insights from molecular dynamics simulations. In Basak SC, Restrepo G and Villaveces JL. (eds.), *Advances in Mathematical Chemistry and Applications*. Bentham Science Publishers, United Arab Emirates, pp. 265–299.

Xiao P, Li J, Fu G, Zhou Y, Huan X and Yang H. 2017. Geographic distribution and temporal trends of HIV-1 subtypes through heterosexual transmission in China: a systematic review and meta-analysis. *Int. J. Environ. Res. Public Health*, 14, 830.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol*., 24, 1586–1591.

Yang H, Nkeze J and Zhao RY. 2012. Effects of HIV-1 protease on cellular functions and their potential applications in antiretroviral therapy. *Cell Biosci.*, 2, 32.

Yates PJ, Hazen R, St. Clair M, Boone L, Tisdale M and Elston RC. 2006. In vitro development of resistance to human immunodeficiency virus protease inhibitor GW640385. *Antimicrob. Agents Chemother*., 50, 1092–1095.

Yedidi RS, Proteasa G, Martin PD, Liu Z, Vickrey JF, Kovari IA and Kovari LC. 2014. A multi-drug resistant HIV-1 protease is resistant to the dimerization inhibitory activity of TLF-PafF. *J. Mol. Graph Model*., 53, 105–111.

Young TP, Parkin NT, Stawiski E, Pilot-Matias T, Trinh R, Kempf DJ and Norton M. 2010. Prevalence, mutation patterns, and effects on protease inhibitor susceptibility of the L76V mutation in HIV-1 protease. *Antimicrob. Agents Chemother*., 54, 4903–4906.

Yu Y, Wang J, Chen Z, Wang G, Shao Q, Shi J and Zhu W. 2017. Structural insights into HIV-1 protease flap opening processes and key intermediates. *RSC Adv*., 7, 45121.

Yu Y, Wang J, Shao Q, Shi J and Zhu W. 2015. Effects of drug-resistant mutations on the dynamic properties of HIV-1 protease and inhibition by Amprenavir and Darunavir. *Sci. Reports*, 5, 10517.

Yuan X and Xu Y. 2018. Recent trends and applications of molecular modeling in GPCR–ligand recognition and structure-based drug design. *Int. J. Mol. Sci*., 19, 2105.

Zhang J, Houb T, Wangc W and Liua JS. 2010. Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance. *Proc. Natl. Acad. Sci. USA*, 107, 1321–1326.

Zhang Y-M, Imamichi H, Imamichi T, Lane HC, Falloon J, Vasudevachari MB and Salzman NP. 1997. Drug resistance during indinavir therapy is caused by mutations in the protease gene and its Gag substrate cleavage sites. *J. Virol*., 71, 6662–6670.

Zhang S, Kaplan AH and Tropsha A. 2008. HIV-1 Protease function and structure studies with the simplicial neighborhood analysis of protein packing (SNAPP) method. *Proteins*, 73, 742–753.

Zhang B, Li J and Lü Q. 2018. Prediction of 8-state protein secondary structures by a novel deep learning architecture. *BMC Bioinformatics*, 19, 293.

Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C and Zhang P. 2013. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*., 497, 643–646.

Zhu Q, Yu Z, Kabashima T, Yin S, Dragusha S, El-Mahdy AFM, Ejupi V, Shibata T and Kai M. 2015. Fluorometric assay for phenotypic differentiation of drug-resistant HIV mutants. *Sci, Rep*., 5, 10323

Appendix A: Publication

This paper was published as a Discovery Note in the journal *Bioinformatics*, btz079. It explored Gag-PR coevolution in pathways to resistance.

Bioinformatics, 2019, 1–5
doi: 10.1093/bioinformatics/btz076
Advance Access Publication Date: 12 February 2019
Discovery Note

OXFORD

Sequence analysis

Gag-protease coevolution shapes the outcome of lopinavir-inclusive treatment regimens in chronically infected HIV-1 subtype C patients

V. Marie and M. Gordon*

KwaZulu-Natal Research Innovation and Sequencing Platform, University of KwaZulu-Natal, Durban, South Africa

*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on November 17, 2018; revised on January 3, 2019; editorial decision on January 25, 2019; accepted on February 11, 2019

Abstract

Motivation: Commonly, protease inhibitor failure is characterized by the development of multiple protease resistance mutations (PRMs). While the impact of PRMs on therapy failure are understood, the introduction of Gag mutations with protease remains largely unclear.

Results: Here, we utilized phylogenetic analyses and Bayesian network learning as tools to understand Gag-protease coevolution and elucidate the pathways leading to Lopinavir failure in HIV-1 subtype C infected patients. Our analyses indicate that while PRMs coevolve in response to drug selection pressure within protease, the Gag mutations added to the existing network while specifically interacting with known Lopinavir failure PRMs. Additionally, the selection of mutations at specific positions in Gag-protease suggests that these coevolving mutational changes occurs to maintain structural integrity during Gag cleavage.

Contact: tarinm@ukzn.ac.za

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Appendix B: Protease and Gag sequence accession numbers

Accession numbers for treatment exposed protease sequences

KF526221	KF135163	KF135023	JN132268	KC423757	JN638200	AY529558	KF135091
KF526227	AY529608	KF135139	JQ480293	KC423705	AY677427	KF134957	KF134930
KF526212	KF135033	KF135029	JN638132	KC423284	FJ445708	KF134958	AY900811
KF526218	AY677473	KF134948	JN638225	KC423763	KC423104	KF135149	KF135083
KF526207	AY529534	KC423753	KC423404	KC423440	KC423562	KC424259	KF135084
KF526220	KF135158	KF135015	KC951632	JN132390	JN132221	AY529578	KF135083
KF526226	KF135159	KF134985	KC423124	JN381594	JQ480163	AY529598	KF134938
AY677485	KF135092	KC423906	KC423703	KF135071	KC424241	AY529599	KF135165
AY529577	KF135150	AY529533	EU308099	KF135072	JN638092	JN700932	KF135162
JF960549	KC423127	KF135028	JQ480160	KF135031	KF134959	JN638218	AY900818
AY529541	KF135044	KF134951	JQ480224	KF135032	KF793146	AY529592	JN638106
AY529542	KF135045	KF134936	KC424212	AY529613	KC423799	KF793134	JN638215
AY529543	KF135046	AY900815	JF960548	KC422880	JN638162	KC423788	JN638070
AY529544	AY529582	AF358747	JF960554	KC423650	AY529554	KF135146	JQ480220
AY529545	KF135030	JN638156	KC423550	AY900795	AY529555	KF135106	KC423714
AY529573	KC423878	KC423388	KC423158	HM593308	AY901184	KF134949	JQ480222
KF134975	KF135066	KC423813	JQ480273	AY529581	AY529561	KF134935	AY529590
KF134994	KF134950	JN638193	KC424332	GU324863	AY529550	KF135120	KF134960
KF134963	KC424133	KC424143	KC423006	JN638208	AY529562	KF135121	JQ480274
MH568933	KC424348	KC422859	KC423841	KC424268	AY529563	KF135076	KF134988
KF135003	KF134953	KC423761	AY900794	JN381618	AY529549	KF135156	JQ480297
GU324870	KF135049	KC424273	AY589912	KC422999	AY529548	KF135157	AY677471
KF134941	KF135151	JQ480225	KC423893	AY090850	AY529567	AY529593	KC424322
KF134941	KF135152	AY090845	HM623531	FJ445712	AY529568	JN638207	KC423458
KF135040	KF135078	FJ445719	JF960579	KC423358	AY529569	KF134997	KC423099
KF135147	KF135001	AY589876	KC423437	AY900770	AY529610	AY529547	JN381572
KF135004	KF135055	HQ843636	KC423449	KC424150	AY529587	KC423644	KC423119
AF358760	KF134939	JQ480283	JN132295	KC424037	JN132219	JN087532	KC423957
AF358750	EF116342	JQ480287	JF960550	JQ480157	KF134969	EU854548	AY461494
KF241499	KF135009	KT032035	KC951639	JN638187	KC422967	EU854547	AY461495
AY677402	EF548178	EU854490	KC423376	AY461496	AY529529	EU854562	KC423384
KF134937	KF135068	EU854489	KT032022	FJ445727	KF135086	EU854561	KF135052
AY900745	KF135010	EU854488	JN638145	AY900805	AY529556	EU854546	KF135127
KF135138	AY529606	EU854487	KC951637	JN638077	AY529557	KF134929	KF135008
KC422954	JN393305	JQ480279	JN381576	EF186986	KF766541	JN132231	KC423688
KF135167	KC423340	KC423211	AJ577964	JN638167	DQ826648	KC422988	KC423140
KF135128	KC424059	KC422931	AJ577965	KT351807	EF186971	HM623520	KF793177
KF135129	KC423540	KC423359	KT032031	JN638096	JQ480275	AY090849	EU308093
KF135131	KC424306	KC424330	AY275752	KC424044	EF186966	AY677450	KC423243
KF135133	KC423491	KC423042	GQ401298	JN381596	JQ480210	JQ480162	JQ480227
JN638206	KC852928	KC423070	GQ401324	KC423565	KC423652	KC423883	KC424349
JN393296	AY677460	KC423483	JQ430823	KC424373	KC423064	KC423856	KC423153

JN132267	KC423610	AY901099	AY677386	JN638204	JQ480289	KC423829	JQ480226
JN132340	KC422992	KC424067	GQ401321	KC423232	KC423228	KC423895	KC424042
AY461501	JN132381	KF135067	GQ401303	KC422911	KF135163	AY090844	KC424319
KC423529	JN132393	KC424028	AY313368	JN638151	KC424186	FJ445702	KC423748
JN638083	KC423303	KC422920	AY900895	KC423380	KC422797	JN381584	KC424328
JQ480282	KC852931	KC423193	GQ401308	KC424381	JN132364	AY090843	KC423160
KC423693	AY589917	KF135105	GQ401299	DQ826651	EU854551	EU854536	KC424340
KC424271	KF135016	JN381616	AY275739	KC204774	EU854550	EU854535	KC423600
KC423150	JN638088	KT032025	AY275720	KC951636	KC422895	EU854534	JN132236
KC423334	KC423685	KC423586	AY390191	KC424224	KC423041	JF960535	KC423853
KJ081337	AY677436	JF960569	AY275748	KC423329	KC423180	AY900772	KC424191
KC423915	KF134982	KC424064	GQ401307	KC423956	KC423614	KC423605	KC424167
KC423885	KC423417	JF960555	GQ401304	JQ480214	KC016209	KC423386	KC424305
EU854504	KC424209	KC423155	GQ401320	JN638074	KC423660	FJ445728	KC423360
EU854502	KC852933	KC424064	GQ401316	JN132229	KC423744	JN381607	KC424354
EU854503	KC423046	KC423467	GQ401322	JN393294	JQ361673	JN381608	HM120057
KC423069	KC424327	KC424098	GQ401319	KF134976	KC424096	KC423653	JF960544
JQ480280	KC423830	JQ480278	GQ401302	AF331699	KC424182	KT032051	KC424147
JN638114	KC424285	KC422873	GQ401305	JF960574	JN638174	JN638169	KC424147
KC424116	KC423146	KT032030	GQ401333	KC423297	DQ826627	KC423185	JN381614
AF338992	JN638179	KC422861	KT351848	KC423828	DQ826638	JF960567	KC422852
AY900778	KF134979	KC423210	AY275735	EF186961	DQ826653	KC951635	JQ361665
KC422864	HM623525	JN132224	AY275730	DQ826628	HQ843518	KC423768	AY090840
JN638203	JN132261	KC424184	AY900730	EF186975	JN638069	JN638113	KC423182
JN132310	KT032027	AY900766	KC424377	FJ983395	KC424258	JN381587	KC422839
JN132396	JN638117	KC423217	GQ401301	KF766542	KC423350	JN381588	KC423355
JN132344	DQ826657	EF186968	KC016282	EF186981	FJ445721	KC424197	KC423363
KC422906	KC423495	KF134961	DQ826629	DQ826646	KC423289	KC423545	KF134971
KC423576	JN381567	KF134962	KJ544553	AJ577947	KC423995	KT032019	KC424356
AY900844	JN132387	EF186965	EF186976	JQ480167	JF960545	KC423071	KC424054
JN381615	KC423020	EF186960	DQ826665	EF186982	KC423677	JN132309	KC424130
AY714411	JN381609	DQ826641	EF186973	KF135012	KF134995	KC423547	AY461491
AY090856	JN381610	EF186956	DQ826654	KC424407	JX982979	KC424334	AY589932
JF960528	KC422984	DQ826649	KC016236	KC423188	JF960561	JN638199	KC423456
KT032024	JQ361668	DQ826662	EF186970	JN132380	KC423270	AY589898	JQ480276
KF134993	JQ480166	EF186962	EF186958	JN132394	AY900767	JN381574	KC423815
JN638130	JN638223	DQ826626	DQ826664	JQ480158	KT351808	EF449991	AY714457
KC424420	KC423455	DQ826639	EF186969	JN638194	KT351809	EU854510	JN638101
AY589925	KF134929	EF186988	DQ826645	DQ826663	JN638141	EU854509	JN638087
KC422857	JN638082	DQ826666	JN185242	EF186957	JN132220	KC423326	KC423382
KC423511	KC424259	DQ826630	KF134943	JN381560	KC423671	JN132299	JN638110
KC424329	KC422869	DQ826632	DQ826660	JN381559	KC423636	JN381591	HM601455
KC424297	KC422944	EF186984	KC423811	HM623523	KC423942	KC422940	JN638095

KC422991	JN132271	DQ826621	KT351811	KF135168	EU308077	JN132379	JN638115
KC424149	JQ480294	EF186974	KT351810	DQ826667	HM623533	KC423168	JN132238
JN638231	JN638161	EF186987	KC423525	EF186985	KC423512	JN381623	KC423707
KC423663	AY090841	DQ826668	KJ544534	DQ826625	KC423289	JN381624	JN132341
KC422862	KC424333	DQ826624	KJ544535	DQ826622	KC424263	JN381568	JN132227
KC423212	KC423420	EF186955	KC016234	DQ826635	KC424132	JF960529	KC424311
JQ480291	KC423851	DQ826633	JN381583	KC423695	KC951633	KC423962	KC423308
KC424200	FJ445709	DQ826652	KC423875	KC424276	KC423043	KC423840	KC423946
KF793181	AY090847	DQ826647	JN132264	KC423247	KC423402	KC422961	JF960557
KF793157	AY090852	DQ826644	DQ826643	KC423324	AY461502	KC423379	JN638178
KC423432	GU324882	DQ826656	DQ826659	KC423817	KC424119	KC423396	KC424344
KC423964	EF186979	EF186926	EF186963	HM623527	KC423780	KF135054	KC423898
AY589882	DQ826669	EF186983	DQ826637	KC423745	JN638072	JN638220	KC424257
KC423594	EF186967	KC016240	DQ826636	KC422803	KC423687	KC422941	KC423037
KC422943	EF186927	KC016219	DQ826658	JN638152	KC423690	KC424251	KF135006
JF960552	EF186959	DQ826620	DQ826642	AY090860	JN638166	KC423970	JQ480223
JN638227	AY797580	DQ826623	AY900793	FJ445701	JN638104	KC422953	KC423094
JN132216	AY529532	JN132384	JQ480211	AY529591	KC423030	JN087540	KC423764
JN638158	KC423936	AY677411	JN132258	AY900773	KC423987	JN087539	JN638103
JN700929	KF135019	JN381611	AY900789	AY677447	KC424133	JN381564	KC423127
KC423939	JN638160	JN132240	KC423286	AY900776	KC423234	KC423937	KC423913
KC424019	KC423634	KC423892	JN638116	KC016226	KC423414	JN087541	KC424294
JF960575	JN132342	JQ480219	EF549036	KC016211	KC423531	EU854508	KC424190
KC423305	KC423721	KC423271	AY090854	KC424279	KT032036	EU854506	KC423012
JN393306	GU324879	JQ480271	FJ445707	AY589931	KC423723	EU854507	JF960563
KC423227	JN132218	JN381586	KC016237	JN087542	KC424321	KC423148	KT032038
KC423273	KC423436	JN381585	EF186928	KC423222	KC423704	GU271172	KF134981
JN132266	JQ480286	EU854527	EF186980	KC423114	JN132239	GU271173	KC423176
KF793185	KC422976	EU854528	KF135060	KT032058	KC422869	KC016227	KC424080
JN132260	KC423648	KC422853	KF135061	KC423322	JN393295	KC016197	JN638066
JN638108	KC424033	AY677448	KF135067	KC423889	KC016213	KC016246	AY529605
KC423380	JN132230	KC423474	AY900833	KT032033	KT032057	KJ176652	AY090853
JN638159	JN638226	KC423767	KF135112	KC423106	KC423888	JQ361677	AY589910
JN132228	KC422982	KC423134	KF135005	JN087536	KC423667	KC422863	AY677453
KC423737	KC423835	KT032037	JN087535	JN087547	KT032060	KC424396	JN132385
KC423603	AY589913	AY589936	KC422997	EU854553	AY589899	KC423137	JN132305
AY589907	JQ480228	JN638071	KC423161	KC423577	KT032055	KC423008	JN132235
KT351819	FJ445709	KC951661	AY677479	KC016243	KC423765	KT032056	EU854500
KT351818	KC424094	JQ480292	AY900786	JN087537	KC423603	AY589908	EU854499
AY275733	KC424156	JQ480215	AJ577870	KT032053	EU854498	KF134952	KF793156
KC423454	KC424376	KC422877	JF960542	JN087531	JN087545	AY900762	JQ480288
JN132255	KC423054	JN638097	AY677426	KT032050	EU854496	AY714443	JQ480296
KC422971	KC424087	KC423943	AY900769	KC423718	EU854497	EF525661	KC423933

JN132363	KC423963	DQ826631	JN638163	KT032034	KC424348	KC423100	JQ480216
JN638175	JN638142	JN638143	AY589937	JN087533	AY589904	KC423192	JN638191
JN132237	JN638073	KC423063	AY677417	KT032012	EU854483	KC423977	KC423983
JN638094	KC424181	KC422916	FJ445724	KC423147	EU854482	KC424335	GQ401316
KC423126	JN638222	KC422855	AY589894	KC016212	JN087546	KC423684	KF135160
EU308079	AY677400	JN638186	AY900768	KT032061	EU854481	KC423858	EU308081
KC951650	JF960526	JN638098	EF186929	KT032043	HM623509	KC423735	KT032062
AY529531	KC423588	JN638105	EF186978	KC424424	KT032040	JN638229	KC424352
FM251950	KC423852	KC423191	KC423062	JN638224	JN638230	KC423276	EU854559
AY677381	KC422850	GU324885	KC423802	EU854525	JN638181	KC423774	JN638109
JN638177	KC423055	KC423878	KC423948	EU854524	KC424345	JN381590	KC423839
AY589920	KC423429	GU324884	KC423503	KC423669	KC423559	JN638136	KC422948
KC422908	EU854480	KC424092	JN638079	KC422867	JN393304	JN393303	JN638148
KF527077	KC424418	JN638067	JQ480229	KC424144	JN132256	JN638091	JN381606
KF527173	JQ480218	KC424152	JN638164	KC423649	KC422806	KF134934	JN638144
KF527176	KC424275	KC422807	JN700927	KC424196	KC424177	JN132294	JN638090
KF527067	JN638168	JN638075	AY589870	JN638147	KC424244	KT032014	JN638137
KF527180	JN381619	KC423257	JN132226	KC424303	JN638198	JN638099	JN638131
KF527192	JN132270	JN638085	JN381579	KC423945	KC423639	JN638213	KC423075
KF527185	JN132343	AY589889	JN132392	KC423416	JN381605	KC423424	EU854541
KF527189	KC423891	JN638081	JN638146	JN638112	JQ480232	JN638202	EU854542
KF527181	KC424246	KC423755	JN638212	KC423854	KC423451	KT032029	KC423315
KF527061	KC423139	KC424136	KC423787	JN638124	JN638197	KC951811	KT032046
KF527187	KC424035	KC422935	KC423034	JF960537	JN638171	JN638221	KC423490
KF527191	KC423221	KC423971	KC422983	JN638084	KC423149	AY901102	KC422979
JN638150	KC424078	KC423101	AY901107	JN381571	JN638189	KC423828	KC423855
KF134990	JN087538	JN393301	JN638139	KF793182	KC424062	JF960558	JN638182
KC424004	JN087534	KC422898	KC423860	JN638219	KC422965	JF960566	JN638176
KC423378	KC423722	JQ480165	KC423563	KC423818	KC423694	KC424221	KC423725
KT032054	KC423225	KC423795	KF793138	KC424406	KC423504	JN132269	KC424108
KC424114	KC423291	KC424323	JN638126	KC423758	KC852930	JN638102	KC423089
KC423004	KC422825	KC423373	JN700931	KC423762	HM623503	KT032002	KC423216
KC423568	KC423385	EU854532	KC423750	JF960572	KC423061	KC422836	KC423198
EU854523	KC422843	EU854530	JN638209	JN381557	EU854519	KC423253	KC423927
EU854522	KC423876	EU854533	JN638185	KC423341	EU854520	KJ081322	JQ480212
KC424079	JQ480209	KC423864	EU308066	EU854494	JN638118	JN381612	JN087543
JF960547	KC422917	KC424397	KT032032	EU854495	KC423258	JN132292	JN087544
KC423387	KC423868	KC423256	KC423201	JN381603	JQ480208	KC423832	JN638120
KC423444	JN381578	KC423174	KC423398	KC423904	KC422831	JN638180	KC424312
JF960549	JN381577	KC424101	KC424291	JN638210	JN132389	KF793139	JN638119
KC424138	KT032052	KC424166	KT032042	JN638216	AY900781	EU854557	KC423053
KC424172	KC423651	JF960546	EU854526	KC423683	JF960530	EU854558	KC424179
JF960577	KC423558	JN700933	AY529559	JF960536	KF793171	EU854539	KC951646

KC423190	JN638125	JN638183	KC423553	KF793121	JN381581	EU854538	KC423300
JN638192	JN638076	KC852932	KC424362	KF793136	KC424009	EU854543	KC422984
JN638190	KC424400	JN393292	KC423471	JN132257	KC424126	EU854545	JQ480290
KC424280	KC423260	JN132225	KC424178	EF549368	AY090858	JQ361682	JN638214
KC423949	KC423036	KF793174	KC422977	EU854515	JN132388	KC423896	JN638086
JN638165	KC423708	AY677456	KF135107	EU854514	JN638133	KC423173	KF135070
KC424359	KC423082	KT032041	JF960534	EU854513	KC423018	JQ480168	AY529595
KC423279	KC423920	AF425391	KC423357	EU854512	AY589897	KC424111	KF135140
KC424416	JN132293	KC423117	KC423804	KC424255	JQ480284	JQ480272	KF135119
KF135110	HM119721	JQ480185	KC951640	KC423292	KC423831	KC423235	KF135153
KC423554	KT032047	JN132215	AY677441	JN638170	KC951634	KC422921	KF135154
KC424238	JN638172	KC423142	JQ480217	KC423797	HM119983	KC423175	KF135075
KC424109	KC423361	KC424338	AY900765	KC424093	JN381562	KC423661	KF134956
KC423112	JQ480164	HM623529	DQ826640	KC423686	JN638127	KC424000	KF134999
JN132391	JN132262	JN638129	JN381566	JQ480298	KT351817	KT351815	AY529594
KC423167	KC423497	KC424384	JN381565	KC423116	KF135053	JQ480159	AY529619
JN381604	KC423781	KC423290	JN638111	KC424090	JN132382	KT032028	KF135114
JN638153	KC423098	JQ480295	KC423323	KT032048	KC423812	KC423816	KF135144
AY900763	JN132263	JN381598	JN381600	EF548284	JN132217	KF135133	AY529602
JN381561	KC423519	JN381599	JN381601	JQ480231	KC424112	AF339009	AY529570
JF960565	JN132259	AY589906	KC423502	KC423249	KC424135	KC424081	AY529571
KC424117	KC423026	KC423903	JN381570	AY140821	AY900807	AY714447	JN132333
KC424405	KC213589	JN132265	KC423729	AY140823	AY900761	KC951668	KF134970
KC422875	KC423742	JN393297	AY900849	AY900833	AY900800	JN638128	AY529583
KC424401	EU854484	KC423181	JQ480281	AY900803	FJ445710	KC423215	KF135118
JF960576	EU854485	KC424236	KC423372	AY900757	JF960559	JQ480207	AY529609
KC424176	JN132298	KC424107	KC424093	JN638196	JQ480285	JN638068	KP411835
KC424317	FJ445729	KC423239	KC423686	KT032039	KC422800	KC424222	AY900816
EU854518	KC423973	JQ361684	AY090851	JQ480230	AY677425	KF134990	AY529588
EU854516	KF135123	KC424387	AY900779	AY900845	KF135022	KF135014	AY529572
EU854517	KC422796	KC423873	KC423906	JQ480221	KC424422	JN638080	AY529596
JQ480213	DQ826650	JN132307	KC423044	KC423655	KC423633	JN638107	KF135064
JN638121	EF186964	KF135013	KC424371	KC852929	EU854537	AY714404	KF135065
KF135002	KF134973	KF135097	KF135085	AY529566	AY529617	KC423832	KF135089
AY529537	GQ401326	KF135098	KC423340	AY900813	AY529616	KF135018	AY529530
AY529538	GQ401327	KF135099	KC424373	KC424217	AY529621	KF135103	KF135033
AY529607	AY677445	KF135117	KF135165	KF135047	KF135173	KF135104	AY900885
KF134934	KF135141	AY529579	KF135082	KF135048	KF135174	KF134989	KF135094
KF135055	KF135142	AY529564	KF135155	KF135024	KF135175	AY529612	KF135095
KF135077	AY529600	AY529575	KF135110	KF135025	KF135080	AY529620	KF135096
KF135172	AY529601	AY529574	KF135111	AY529576	AY529618	KF135164	AY010422
KC424004	KF134954	AY529565	KF135116	KF134987	AY529622	KF135143	AY010425

Accession numbers for naïve protease sequences

AB081151	AB285773	AB285817	AB640496	AM071421	AY007739	AY136977	AY165216
AB081152	AB285774	AB285818	AF527205	AM071422	AY007740	AY136978	AY165219
AB081154	AB285775	AB285819	AF527207	AM071423	AY007741	AY136979	AY165220
AB081155	AB285776	AB285820	AF527211	AM071424	AY007742	AY136980	AY165223
AB081156	AB285777	AB285821	AF527213	AM071425	AY007743	AY136981	AY165224
AB081157	AB285778	AB285822	AF527217	AM071426	AY007744	AY136982	AY165225
AB081158	AB285779	AB285823	AF527222	AM071427	AY007746	AY136983	AY165226
AB081159	AB285780	AB285824	AF527230	AM071428	AY036312	AY136984	AY165230
AB081161	AB285781	AB285825	AF527240	AM071429	AY036325	AY136985	AY165235
AB081162	AB285782	AB285826	AF527242	AM071430	AY036329	AY136986	AY165252
AB081163	AB285783	AB285827	AF527246	AM071431	AY036377	AY136987	AY165253
AB081164	AB285784	AB285828	AF527248	AM071432	AY036391	AY136988	AY165255
AB081165	AB285786	AB285829	AF527250	AM071433	AY036394	AY136989	AY165257
AB081166	AB285787	AB285831	AF527256	AM071434	AY036395	AY136990	AY165260
AB081167	AB285788	AB285832	AF527264	AM071435	AY036397	AY136991	AY165263
AB081168	AB285789	AB285833	AF527266	AM071436	AY036399	AY136992	AY165266
AB081169	AB285790	AB285834	AF527272	AM071437	AY036401	AY136993	AY165272
AB081170	AB285791	AB285835	AF527277	AM071438	AY036404	AY136994	AY165281
AB081171	AB285792	AB285836	AF527284	AM071439	AY036405	AY136995	AY196498
AB081172	AB285793	AB285837	AF527286	AM071440	AY036406	AY136996	AY196499
AB081173	AB285794	AB285838	AF527292	AM071441	AY036408	AY136997	AY196500
AB081175	AB285795	AB285839	AF527294	AM071442	AY136954	AY136998	AY196501
AB285752	AB285796	AB285840	AF527305	AM071443	AY136955	AY136999	AY196502
AB285753	AB285797	AB285841	AF527340	AM071444	AY136956	AY137000	AY196503
AB285754	AB285798	AB285842	AJ287005	AM071445	AY136957	AY137001	AY196504
AB285755	AB285799	AB285843	AM040985	AM071446	AY136958	AY137002	AY196505
AB285756	AB285800	AB356162	AM040991	AM071447	AY136959	AY137003	AY196506
AB285757	AB285801	AB356179	AM041002	AM071448	AY136960	AY137004	AY196507
AB285758	AB285802	AB356180	AM041004	AM071449	AY136961	AY137005	AY196508
AB285759	AB285803	AB356214	AM041011	AM071450	AY136962	AY137006	AY196509
AB285760	AB285804	AB356260	AM041013	AM071451	AY136963	AY137007	AY196510
AB285761	AB285805	AB356269	AM041017	AM071452	AY136964	AY137008	AY196511
AB285762	AB285806	AB356345	AM041034	AM071453	AY136965	AY165186	AY196512
AB285763	AB285807	AB356375	AM041050	AM071454	AY136966	AY165187	AY196513
AB285764	AB285808	AB442231	AM071412	AY007730	AY136967	AY165196	AY196514
AB285765	AB285809	AB442233	AM071413	AY007731	AY136968	AY165205	AY196515
AB285766	AB285810	AB442259	AM071414	AY007732	AY136969	AY165207	AY196516
AB285767	AB285811	AB640061	AM071415	AY007733	AY136970	AY165208	AY196517
AB285768	AB285812	AB640066	AM071416	AY007734	AY136971	AY165209	AY213437
AB285769	AB285813	AB640073	AM071417	AY007735	AY136972	AY165211	AY213521
AB285770	AB285814	AB640195	AM071418	AY007736	AY136973	AY165213	AY213522
AB285771	AB285815	AB640318	AM071419	AY007737	AY136975	AY165214	AY213525

AB285772	AB285816	AB640380	AM071420	AY007738	AY136976	AY165215	AY213542
AY213550	AY463228	AY515763	AY515809	AY585266	AY746362	AY829282	AY829326
AY228556	AY463229	AY515764	AY515810	AY585267	AY746365	AY829283	AY829327
AY228557	AY463230	AY515765	AY515811	AY585268	AY746368	AY829284	AY829328
AY267321	AY463231	AY515766	AY515812	AY589839	AY746371	AY829285	AY829329
AY331723	AY463232	AY515767	AY515813	AY589840	AY746374	AY829286	AY829330
AY359695	AY463233	AY515768	AY515814	AY589841	AY746376	AY829287	AY829331
AY359696	AY463234	AY515769	AY515815	AY589842	AY746379	AY829288	AY829332
AY359697	AY463236	AY515771	AY515816	AY589843	AY746382	AY829290	AY829333
AY390076	AY463237	AY515772	AY515817	AY589844	AY746385	AY829291	AY829334
AY390079	AY515728	AY515773	AY515818	AY589845	AY746386	AY829292	AY829335
AY390080	AY515729	AY515774	AY515819	AY589846	AY746389	AY829293	AY829336
AY390081	AY515730	AY515775	AY515820	AY589847	AY749171	AY829294	AY829337
AY390188	AY515731	AY515776	AY515821	AY589848	AY749174	AY829295	AY829338
AY390190	AY515732	AY515777	AY515822	AY589849	AY749177	AY829296	AY878054
AY390193	AY515733	AY515778	AY515823	AY589850	AY772690	AY829297	AY878055
AY390194	AY515734	AY515779	AY515824	AY589851	AY772691	AY829298	AY878056
AY390196	AY515735	AY515780	AY515825	AY589852	AY772692	AY829299	AY878057
AY390198	AY515736	AY515781	AY515826	AY589853	AY772693	AY829300	AY878058
AY435389	AY515737	AY515782	AY515827	AY589854	AY772694	AY829301	AY878059
AY444208	AY515738	AY515783	AY515828	AY589855	AY772695	AY829302	AY878060
AY444209	AY515739	AY515784	AY515829	AY589856	AY772696	AY829303	AY878061
AY444210	AY515740	AY515785	AY515830	AY589857	AY772698	AY829304	AY878062
AY444211	AY515741	AY515786	AY515831	AY589858	AY772699	AY829305	AY878063
AY444212	AY515742	AY515787	AY515832	AY589859	AY772700	AY829306	AY878064
AY444213	AY515744	AY515789	AY515833	AY589860	AY787492	AY829307	AY878065
AY444214	AY515745	AY515791	AY515834	AY589861	AY787522	AY829308	AY878068
AY444215	AY515746	AY515792	AY515835	AY589862	AY787532	AY829309	AY878070
AY444216	AY515747	AY515793	AY515836	AY589863	AY787535	AY829310	AY878071
AY444217	AY515748	AY515794	AY515837	AY589864	AY787536	AY829311	AY878072
AY444218	AY515749	AY515795	AY515838	AY589865	AY829268	AY829312	AY900740
AY444219	AY515750	AY515796	AY515839	AY589866	AY829269	AY829313	AY900752
AY444220	AY515751	AY515797	AY515840	AY589867	AY829270	AY829314	AY900759
AY463217	AY515752	AY515798	AY515841	AY589868	AY829271	AY829315	AY900760
AY463218	AY515753	AY515799	AY515842	AY589869	AY829272	AY829316	AY900774
AY463219	AY515754	AY515800	AY515843	AY611649	AY829273	AY829317	AY900775
AY463220	AY515755	AY515801	AY515844	AY611650	AY829274	AY829318	AY900777
AY463221	AY515756	AY515802	AY515845	AY611651	AY829275	AY829319	AY900780
AY463222	AY515757	AY515803	AY515846	AY703908	AY829276	AY829320	AY900782
AY463223	AY515758	AY515804	AY515847	AY703909	AY829277	AY829321	AY900785
AY463224	AY515759	AY515805	AY515848	AY703910	AY829278	AY829322	AY900787
AY463225	AY515760	AY515806	AY515849	AY703911	AY829279	AY829323	AY900797
AY463226	AY515761	AY515807	AY585264	AY739911	AY829280	AY829324	AY900806

AY463227	AY515762	AY515808	AY585265	AY739913	AY829281	AY829325	AY900832
AY900834	DQ011172	DQ093589	DQ222250	DQ275652	DQ369988	DQ445633	DQ660009
AY900835	DQ011173	DQ093590	DQ222251	DQ275653	DQ369989	DQ445634	DQ660013
AY900836	DQ011174	DQ093591	DQ222252	DQ275654	DQ369990	DQ445635	DQ660014
AY900837	DQ011175	DQ093592	DQ222253	DQ275655	DQ369991	DQ445637	DQ660015
AY900839	DQ011176	DQ093593	DQ222254	DQ275656	DQ369992	DQ659965	DQ660016
AY900843	DQ011177	DQ093594	DQ222255	DQ275657	DQ369993	DQ659966	DQ660017
AY900850	DQ011178	DQ093595	DQ222256	DQ275658	DQ369994	DQ659967	DQ660019
AY900853	DQ011179	DQ093596	DQ222257	DQ275659	DQ369995	DQ659968	DQ660020
AY900855	DQ011180	DQ093597	DQ222258	DQ275660	DQ369996	DQ659969	DQ660021
AY900858	DQ013271	DQ093598	DQ222259	DQ275661	DQ369997	DQ659970	DQ660022
AY900914	DQ013277	DQ093599	DQ222260	DQ275664	DQ396364	DQ659972	DQ660023
AY900917	DQ056302	DQ093600	DQ222261	DQ351216	DQ396365	DQ659973	DQ660024
AY900920	DQ056404	DQ093601	DQ222262	DQ351217	DQ396367	DQ659974	DQ660025
AY900936	DQ056405	DQ093602	DQ222263	DQ351218	DQ396368	DQ659975	DQ660026
AY900953	DQ056406	DQ093604	DQ222264	DQ351219	DQ396369	DQ659976	DQ660027
AY900954	DQ056408	DQ093605	DQ222265	DQ351220	DQ396370	DQ659977	DQ660028
AY900955	DQ056409	DQ093607	DQ222266	DQ351221	DQ396371	DQ659978	DQ660029
AY900956	DQ056410	DQ164104	DQ222267	DQ351222	DQ396372	DQ659979	DQ660031
AY900957	DQ056411	DQ164106	DQ222268	DQ351223	DQ396373	DQ659980	DQ660032
AY901965	DQ056412	DQ164107	DQ222269	DQ351224	DQ396374	DQ659982	DQ660033
AY901966	DQ056413	DQ164108	DQ222270	DQ351225	DQ396375	DQ659983	DQ660035
AY901967	DQ056414	DQ164109	DQ222271	DQ351226	DQ396376	DQ659984	DQ660036
AY901968	DQ056415	DQ164110	DQ222272	DQ351227	DQ396377	DQ659985	DQ660037
AY901969	DQ056416	DQ164111	DQ222273	DQ351228	DQ396378	DQ659986	DQ660038
AY901970	DQ056417	DQ164113	DQ222274	DQ351229	DQ396379	DQ659987	DQ660039
AY901971	DQ056418	DQ164114	DQ222275	DQ351230	DQ396380	DQ659988	DQ660041
AY901972	DQ060209	DQ164115	DQ222276	DQ351231	DQ396381	DQ659989	DQ660042
AY901973	DQ060212	DQ164117	DQ222277	DQ351232	DQ396382	DQ659990	DQ660043
AY901974	DQ060213	DQ164118	DQ222278	DQ351234	DQ396383	DQ659991	DQ660045
AY901975	DQ060215	DQ164119	DQ222279	DQ351235	DQ396384	DQ659992	DQ660046
AY901976	DQ060216	DQ164121	DQ222280	DQ351237	DQ396385	DQ659993	DQ660047
AY901977	DQ060218	DQ164122	DQ222281	DQ369976	DQ396386	DQ659994	DQ660048
AY901978	DQ060219	DQ164126	DQ222282	DQ369977	DQ396387	DQ659995	DQ660050
AY901979	DQ060220	DQ164127	DQ275642	DQ369978	DQ396388	DQ659996	DQ660051
AY901980	DQ060221	DQ164129	DQ275643	DQ369979	DQ396389	DQ659998	DQ660052
AY901981	DQ060226	DQ205288	DQ275644	DQ369980	DQ396390	DQ659999	DQ660053
AY945738	DQ060227	DQ222243	DQ275645	DQ369981	DQ396391	DQ660000	DQ660054
DQ011165	DQ060228	DQ222244	DQ275646	DQ369982	DQ396392	DQ660001	DQ660055
DQ011166	DQ070553	DQ222245	DQ275647	DQ369983	DQ396393	DQ660003	DQ660056
DQ011167	DQ093585	DQ222246	DQ275648	DQ369984	DQ396394	DQ660004	DQ660057
DQ011169	DQ093586	DQ222247	DQ275649	DQ369985	DQ396395	DQ660005	DQ660058
DQ011170	DQ093587	DQ222248	DQ275650	DQ369986	DQ396399	DQ660007	DQ660059

DQ011171	DQ093588	DQ222249	DQ275651	DQ369987	DQ445632	DQ660008	DQ660060
DQ660061	EF120336	EF379168	EF407723	EF491938	EF602206	EF602250	EF602297
DQ660062	EF120337	EF379171	EF407724	EF491939	EF602207	EF602251	EF602298
DQ660063	EF122539	EF379174	EF407725	EF491940	EF602208	EF602252	EF602299
DQ660064	EF143496	EF379176	EF407726	EF491941	EF602209	EF602253	EF602301
DQ660065	EF143499	EF379177	EF407727	EF491942	EF602210	EF602254	EF602582
DQ660066	EF143501	EF379178	EF407728	EF491943	EF602211	EF602255	EF602583
DQ660067	EF143502	EF379181	EF407729	EF491944	EF602212	EF602256	EF602584
DQ660068	EF143518	EF379191	EF407730	EF491945	EF602213	EF602257	EF602585
DQ660069	EF143521	EF379193	EF407731	EF491946	EF602214	EF602258	EF602586
DQ660070	EF143527	EF379195	EF407732	EF491947	EF602215	EF602259	EF602587
DQ660071	EF143529	EF379198	EF407733	EF491948	EF602216	EF602260	EF602588
DQ660073	EF143532	EF379199	EF407734	EF491950	EF602217	EF602261	EF602589
DQ660074	EF143533	EF379203	EF407735	EF602172	EF602218	EF602263	EF602590
DQ660075	EF157888	EF379204	EF407736	EF602173	EF602219	EF602264	EF602591
DQ666390	EF186933	EF379206	EF407737	EF602174	EF602220	EF602265	EF602592
DQ826595	EF186934	EF379207	EF407738	EF602175	EF602221	EF602266	EF602593
DQ826596	EF186935	EF379208	EF407739	EF602176	EF602222	EF602267	EF602594
DQ826597	EF186936	EF379209	EF407740	EF602177	EF602223	EF602268	EF602595
DQ826598	EF186938	EF379211	EF407741	EF602180	EF602224	EF602269	EF602596
DQ826599	EF186939	EF379212	EF407742	EF602181	EF602225	EF602270	EF602597
DQ826600	EF186940	EF379213	EF407743	EF602182	EF602226	EF602271	EF602598
DQ826601	EF186942	EF407701	EF407744	EF602183	EF602227	EF602272	EF602599
DQ826602	EF186952	EF407702	EF407745	EF602184	EF602228	EF602273	EF602600
DQ826603	EF186953	EF407703	EF407746	EF602185	EF602229	EF602274	EF602601
DQ826604	EF186954	EF407704	EF407747	EF602186	EF602230	EF602275	EF602602
DQ826605	EF195277	EF407705	EF407748	EF602187	EF602231	EF602277	EF602603
DQ826606	EF368549	EF407706	EF407749	EF602188	EF602232	EF602278	EF602604
DQ826607	EF368642	EF407707	EF407750	EF602189	EF602233	EF602279	EF602605
DQ826608	EF368697	EF407708	EF407751	EF602190	EF602235	EF602280	EF602606
DQ826609	EF368731	EF407709	EF469243	EF602191	EF602236	EF602282	EF602607
DQ826610	EF368743	EF407710	EF491925	EF602192	EF602237	EF602283	EF602608
DQ826611	EF368805	EF407711	EF491926	EF602193	EF602238	EF602284	EF602609
DQ826612	EF368807	EF407712	EF491927	EF602195	EF602239	EF602285	EF602610
DQ826613	EF368846	EF407713	EF491928	EF602196	EF602240	EF602286	EF602611
DQ826614	EF368925	EF407714	EF491929	EF602197	EF602241	EF602288	EF602612
DQ826615	EF369040	EF407715	EF491930	EF602198	EF602242	EF602289	EF602613
DQ826616	EF369043	EF407716	EF491931	EF602199	EF602243	EF602290	EF602614
DQ826617	EF369057	EF407717	EF491932	EF602200	EF602244	EF602291	EF602615
DQ826618	EF369085	EF407718	EF491933	EF602201	EF602245	EF602292	EF602616
DQ826619	EF369163	EF407719	EF491934	EF602202	EF602246	EF602293	EF602617
EF064480	EF369173	EF407720	EF491935	EF602203	EF602247	EF602294	EF602618
EF120244	EF379154	EF407721	EF491936	EF602204	EF602248	EF602295	EF602619

EF120254	EF379159	EF407722	EF491937	EF602205	EF602249	EF602296	EF602620
EF602621	EU110070	EU244654	EU248348	EU449520	EU683754	EU683800	FJ199618
EF602622	EU110071	EU244655	EU248356	EU449521	EU683755	EU683801	FJ199621
EF602623	EU110072	EU244656	EU248376	EU449522	EU683756	EU693681	FJ199628
EF602624	EU110073	EU244657	EU248393	EU449523	EU683757	EU693717	FJ199639
EF602625	EU110074	EU244658	EU248397	EU449524	EU683758	EU693892	FJ199641
EF602626	EU110075	EU244659	EU248441	EU449525	EU683759	EU744912	FJ199646
EF602627	EU110078	EU244660	EU248453	EU449526	EU683761	EU744913	FJ199662
EF602628	EU110079	EU244661	EU248485	EU449527	EU683762	EU744914	FJ199668
EU030410	EU110080	EU244662	EU248490	EU449528	EU683763	EU786673	FJ199671
EU030412	EU110083	EU244663	EU248493	EU449529	EU683764	EU786681	FJ199673
EU030413	EU158865	EU244664	EU248516	EU449530	EU683765	EU822552	FJ199676
EU030414	EU158866	EU244665	EU248523	EU449531	EU683766	EU822566	FJ199680
EU030415	EU158867	EU244666	EU248524	EU449532	EU683767	EU822638	FJ199683
EU030416	EU158868	EU244667	EU248543	EU449533	EU683768	EU822697	FJ199689
EU030417	EU158869	EU244668	EU248545	EU449534	EU683769	EU822704	FJ199691
EU030418	EU158870	EU244669	EU248558	EU449535	EU683770	EU822714	FJ199692
EU106117	EU158871	EU244670	EU248561	EU449536	EU683771	EU856363	FJ199700
EU106118	EU158872	EU244671	EU248569	EU449537	EU683772	EU884500	FJ199703
EU106119	EU158873	EU244672	EU248584	EU449538	EU683773	FJ030660	FJ199707
EU106120	EU158874	EU244673	EU248586	EU449539	EU683775	FJ030696	FJ199710
EU106121	EU158875	EU244674	EU255374	EU449540	EU683776	FJ199532	FJ199712
EU106122	EU158876	EU244675	EU255465	EU449542	EU683777	FJ199533	FJ199714
EU106123	EU158877	EU244676	EU255509	EU449543	EU683778	FJ199537	FJ199719
EU106124	EU158878	EU244677	EU255513	EU449544	EU683780	FJ199546	FJ199722
EU106125	EU158879	EU244678	EU293445	EU449545	EU683781	FJ199552	FJ199726
EU106126	EU158880	EU244679	EU293446	EU449546	EU683782	FJ199553	FJ199732
EU110053	EU158881	EU244680	EU293447	EU449547	EU683783	FJ199560	FJ199734
EU110054	EU158882	EU244681	EU293448	EU449548	EU683784	FJ199561	FJ199735
EU110055	EU158885	EU244682	EU293449	EU449550	EU683785	FJ199563	FJ199736
EU110056	EU158886	EU244683	EU293450	EU449551	EU683786	FJ199565	FJ199738
EU110057	EU158887	EU244684	EU306757	EU449552	EU683787	FJ199569	FJ199741
EU110058	EU158888	EU244685	EU312784	EU449553	EU683788	FJ199570	FJ199742
EU110059	EU158889	EU244686	EU345720	EU449554	EU683789	FJ199571	FJ199744
EU110060	EU158890	EU244687	EU345911	EU484322	EU683790	FJ199572	FJ199745
EU110061	EU158891	EU244688	EU345933	EU484323	EU683791	FJ199573	FJ199749
EU110062	EU158892	EU244689	EU345935	EU484324	EU683792	FJ199575	FJ199750
EU110063	EU244647	EU244690	EU345945	EU673374	EU683793	FJ199576	FJ199759
EU110064	EU244648	EU244691	EU429990	EU673391	EU683794	FJ199578	FJ199760
EU110065	EU244649	EU244692	EU447776	EU673400	EU683795	FJ199580	FJ199764
EU110066	EU244650	EU244693	EU447778	EU673406	EU683796	FJ199586	FJ199765
EU110067	EU244651	EU248293	EU447779	EU673407	EU683797	FJ199589	FJ199766
EU110068	EU244652	EU248298	EU447780	EU683752	EU683798	FJ199594	FJ199768

EU110069	EU244653	EU248333	EU447781	EU683753	EU683799	FJ199610	FJ199773
FJ199775	FJ199845	FJ498263	FJ498306	FJ498349	FJ498393	FJ498437	FJ498483
FJ199776	FJ199863	FJ498264	FJ498307	FJ498350	FJ498394	FJ498438	FJ498484
FJ199779	FJ199864	FJ498265	FJ498308	FJ498351	FJ498395	FJ498439	FJ498486
FJ199780	FJ199868	FJ498266	FJ498309	FJ498352	FJ498396	FJ498440	FJ498487
FJ199781	FJ199878	FJ498267	FJ498310	FJ498353	FJ498397	FJ498441	FJ498488
FJ199782	FJ199907	FJ498268	FJ498311	FJ498354	FJ498398	FJ498442	FJ498489
FJ199786	FJ199915	FJ498269	FJ498312	FJ498355	FJ498399	FJ498443	FJ498490
FJ199787	FJ199943	FJ498270	FJ498313	FJ498356	FJ498400	FJ498444	FJ498491
FJ199790	FJ199966	FJ498271	FJ498314	FJ498357	FJ498401	FJ498445	FJ498492
FJ199792	FJ199970	FJ498272	FJ498315	FJ498358	FJ498402	FJ498446	FJ498493
FJ199794	FJ199973	FJ498273	FJ498316	FJ498359	FJ498403	FJ498447	FJ498494
FJ199795	FJ199978	FJ498274	FJ498317	FJ498360	FJ498404	FJ498448	FJ498495
FJ199797	FJ199983	FJ498275	FJ498318	FJ498361	FJ498405	FJ498449	FJ498496
FJ199798	FJ199985	FJ498276	FJ498319	FJ498362	FJ498406	FJ498450	FJ498498
FJ199799	FJ199986	FJ498277	FJ498320	FJ498363	FJ498408	FJ498451	FJ498499
FJ199800	FJ360877	FJ498278	FJ498321	FJ498364	FJ498409	FJ498452	FJ498500
FJ199803	FJ360878	FJ498279	FJ498322	FJ498365	FJ498410	FJ498453	FJ498501
FJ199804	FJ360879	FJ498280	FJ498323	FJ498366	FJ498411	FJ498455	FJ498503
FJ199807	FJ360880	FJ498281	FJ498324	FJ498367	FJ498412	FJ498456	FJ498504
FJ199808	FJ360881	FJ498282	FJ498325	FJ498368	FJ498413	FJ498457	FJ498505
FJ199809	FJ360882	FJ498283	FJ498326	FJ498369	FJ498414	FJ498458	FJ498506
FJ199812	FJ360883	FJ498284	FJ498327	FJ498371	FJ498415	FJ498459	FJ498507
FJ199815	FJ481652	FJ498285	FJ498328	FJ498372	FJ498416	FJ498460	FJ498508
FJ199816	FJ481654	FJ498286	FJ498329	FJ498373	FJ498417	FJ498461	FJ498509
FJ199817	FJ481655	FJ498287	FJ498330	FJ498374	FJ498418	FJ498462	FJ498510
FJ199818	FJ498244	FJ498288	FJ498331	FJ498375	FJ498419	FJ498463	FJ498511
FJ199820	FJ498245	FJ498289	FJ498332	FJ498376	FJ498420	FJ498464	FJ498512
FJ199821	FJ498246	FJ498290	FJ498333	FJ498377	FJ498421	FJ498465	FJ498513
FJ199824	FJ498247	FJ498291	FJ498334	FJ498378	FJ498422	FJ498466	FJ498514
FJ199825	FJ498248	FJ498292	FJ498335	FJ498379	FJ498423	FJ498467	FJ498515
FJ199826	FJ498249	FJ498293	FJ498336	FJ498380	FJ498424	FJ498468	FJ498516
FJ199827	FJ498250	FJ498294	FJ498337	FJ498381	FJ498425	FJ498469	FJ498517
FJ199828	FJ498251	FJ498295	FJ498338	FJ498382	FJ498426	FJ498470	FJ498518
FJ199830	FJ498253	FJ498296	FJ498339	FJ498383	FJ498427	FJ498471	FJ498519
FJ199831	FJ498254	FJ498297	FJ498340	FJ498384	FJ498428	FJ498472	FJ498520
FJ199835	FJ498255	FJ498298	FJ498341	FJ498385	FJ498429	FJ498473	FJ498521
FJ199837	FJ498256	FJ498299	FJ498342	FJ498386	FJ498430	FJ498474	FJ498522
FJ199838	FJ498257	FJ498300	FJ498343	FJ498387	FJ498431	FJ498475	FJ498524
FJ199839	FJ498258	FJ498301	FJ498344	FJ498388	FJ498432	FJ498476	FJ498525
FJ199841	FJ498259	FJ498302	FJ498345	FJ498389	FJ498433	FJ498478	FJ498526
FJ199842	FJ498260	FJ498303	FJ498346	FJ498390	FJ498434	FJ498480	FJ498527
FJ199843	FJ498261	FJ498304	FJ498347	FJ498391	FJ498435	FJ498481	FJ498529

FJ199844	FJ498262	FJ498305	FJ498348	FJ498392	FJ498436	FJ498482	FJ498530
FJ498531	FJ595394	FJ595442	FJ878959	GQ294621	GQ399875	GQ400595	GQ427134
FJ498532	FJ595395	FJ595443	FJ878960	GQ372203	GQ399890	GQ400599	GQ427135
FJ498533	FJ595396	FJ595444	FJ878961	GQ398837	GQ399932	GQ400608	GQ427136
FJ498534	FJ595397	FJ595445	FJ878962	GQ398838	GQ399970	GQ400615	GQ427137
FJ498535	FJ595398	FJ595446	FJ878963	GQ398881	GQ399972	GQ400619	GQ427138
FJ498536	FJ595399	FJ595447	FJ878964	GQ398909	GQ399981	GQ400644	GQ427139
FJ498537	FJ595400	FJ595448	FJ878965	GQ398966	GQ399985	GQ400646	GQ427140
FJ498538	FJ595401	FJ595450	FJ878966	GQ399028	GQ400018	GQ400651	GQ433802
FJ498539	FJ595402	FJ595451	FJ878967	GQ399033	GQ400025	GQ400661	GQ433803
FJ498540	FJ595403	FJ595452	FJ878968	GQ399069	GQ400044	GQ400682	GQ433804
FJ498541	FJ595404	FJ595453	FJ878969	GQ399082	GQ400064	GQ400699	GQ433805
FJ498542	FJ595405	FJ595454	FJ878970	GQ399122	GQ400077	GQ400816	GQ433806
FJ498543	FJ595406	FJ595455	FM251949	GQ399127	GQ400095	GQ400909	GQ433807
FJ548791	FJ595407	FJ595456	FN557337	GQ399144	GQ400143	GQ400924	GQ433808
FJ594149	FJ595408	FJ595457	FN557338	GQ399200	GQ400145	GQ400942	GQ433809
FJ595365	FJ595409	FJ595458	FN599687	GQ399228	GQ400149	GQ400949	GQ433810
FJ595367	FJ595410	FJ595459	FN599688	GQ399252	GQ400151	GQ400950	GQ433811
FJ595368	FJ595411	FJ595460	FN599712	GQ399298	GQ400158	GQ400980	GQ433812
FJ595369	FJ595412	FJ595461	FN599718	GQ399345	GQ400161	GQ400993	GQ433813
FJ595370	FJ595414	FJ595462	FN599720	GQ399346	GQ400193	GQ401022	GQ433814
FJ595371	FJ595415	FJ670521	FN599737	GQ399364	GQ400211	GQ427111	GQ433815
FJ595372	FJ595416	FJ878937	FN599773	GQ399391	GQ400237	GQ427112	GQ433816
FJ595373	FJ595417	FJ878938	FN599776	GQ399394	GQ400273	GQ427113	GQ433817
FJ595374	FJ595418	FJ878939	FR666608	GQ399408	GQ400288	GQ427114	GQ433818
FJ595375	FJ595420	FJ878940	FR666613	GQ399414	GQ400298	GQ427115	GQ433819
FJ595376	FJ595421	FJ878941	FR666617	GQ399464	GQ400339	GQ427116	GQ433820
FJ595377	FJ595422	FJ878942	FR666620	GQ399501	GQ400369	GQ427117	GQ433821
FJ595378	FJ595423	FJ878943	FR666621	GQ399507	GQ400374	GQ427118	GQ433822
FJ595379	FJ595425	FJ878944	FR666622	GQ399517	GQ400394	GQ427119	GQ433823
FJ595380	FJ595426	FJ878945	FR666625	GQ399518	GQ400400	GQ427120	GQ433824
FJ595381	FJ595427	FJ878946	FR666631	GQ399556	GQ400428	GQ427121	GQ433825
FJ595382	FJ595428	FJ878947	FR666642	GQ399593	GQ400431	GQ427122	GQ433826
FJ595383	FJ595430	FJ878948	FR666649	GQ399616	GQ400463	GQ427123	GQ433827
FJ595384	FJ595431	FJ878949	FR666650	GQ399666	GQ400496	GQ427124	GQ433828
FJ595385	FJ595432	FJ878950	FR666658	GQ399679	GQ400501	GQ427125	GQ433829
FJ595386	FJ595434	FJ878951	FR666669	GQ399697	GQ400504	GQ427126	GQ433830
FJ595387	FJ595435	FJ878952	GQ240958	GQ399711	GQ400508	GQ427127	GQ433831
FJ595388	FJ595436	FJ878953	GQ241015	GQ399715	GQ400526	GQ427128	GQ433832
FJ595389	FJ595437	FJ878954	GQ241016	GQ399733	GQ400527	GQ427129	GQ433833
FJ595390	FJ595438	FJ878955	GQ241017	GQ399752	GQ400542	GQ427130	GQ433834
FJ595391	FJ595439	FJ878956	GQ241019	GQ399768	GQ400570	GQ427131	GQ433835
FJ595392	FJ595440	FJ878957	GQ290728	GQ399796	GQ400585	GQ427132	GQ433836

FJ595393	FJ595441	FJ878958	GQ294616	GQ399801	GQ400592	GQ427133	GQ433837
GQ433838	GQ433882	GU188837	GU199556	GU201815	GU253426	HM191529	HM593130
GQ433839	GQ433883	GU188838	GU199557	GU201816	GU253428	HM191539	HM593131
GQ433840	GQ433884	GU188839	GU199558	GU201817	GU324862	HM191541	HM593132
GQ433841	GQ433885	GU188840	GU199559	GU201818	GU326106	HM191542	HM593133
GQ433842	GQ433886	GU188841	GU199560	GU201819	GU326122	HM191545	HM593136
GQ433843	GQ433887	GU188842	GU199561	GU201820	GU326127	HM191551	HM593149
GQ433844	GQ433888	GU188843	GU199562	GU201821	GU326128	HM191553	HM593157
GQ433845	GQ433889	GU188844	GU199563	GU201822	GU326172	HM191554	HM593158
GQ433846	GQ433890	GU188845	GU199564	GU201823	GU326187	HM191556	HM593162
GQ433847	GQ433891	GU188846	GU199565	GU201824	GU906868	HM468977	HM593166
GQ433848	GQ433892	GU188847	GU199566	GU201825	GU906870	HM572396	HM593173
GQ433849	GQ433893	GU188848	GU199567	GU201826	HG421606	HM572397	HM593177
GQ433850	GQ862311	GU188849	GU199568	GU214076	HG421607	HM572398	HM593186
GQ433851	GQ999989	GU199525	GU199569	GU214085	HG421655	HM572399	HM593187
GQ433852	GU188807	GU199526	GU199570	GU214086	HG421662	HM572400	HM593193
GQ433853	GU188808	GU199527	GU199571	GU253395	HG421665	HM572401	HM593194
GQ433854	GU188809	GU199528	GU199572	GU253396	HM037812	HM572402	HM593195
GQ433855	GU188810	GU199529	GU199573	GU253397	HM119611	HM572403	HM593201
GQ433856	GU188811	GU199530	GU199574	GU253399	HM119629	HM572404	HM593202
GQ433857	GU188812	GU199531	GU199575	GU253400	HM119650	HM572405	HM593204
GQ433858	GU188813	GU199532	GU199576	GU253401	HM119655	HM572406	HM593205
GQ433859	GU188814	GU199533	GU199577	GU253402	HM119656	HM572407	HM593206
GQ433860	GU188815	GU199534	GU199578	GU253403	HM119659	HM572408	HM593207
GQ433861	GU188816	GU199535	GU199579	GU253404	HM119667	HM572409	HM593211
GQ433862	GU188818	GU199536	GU199580	GU253405	HM119714	HM572410	HM593213
GQ433863	GU188819	GU199537	GU199582	GU253406	HM119754	HM572411	HM593215
GQ433864	GU188820	GU199538	GU201798	GU253407	HM119768	HM572413	HM593217
GQ433865	GU188821	GU199539	GU201799	GU253408	HM119815	HM572414	HM593218
GQ433866	GU188822	GU199540	GU201800	GU253410	HM119836	HM572415	HM593224
GQ433867	GU188823	GU199541	GU201801	GU253411	HM119841	HM572416	HM593226
GQ433868	GU188824	GU199542	GU201802	GU253412	HM119868	HM572417	HM593227
GQ433869	GU188825	GU199543	GU201803	GU253413	HM119876	HM593106	HM593228
GQ433870	GU188826	GU199544	GU201804	GU253414	HM119902	HM593107	HM593229
GQ433871	GU188827	GU199545	GU201805	GU253415	HM119932	HM593108	HM593230
GQ433872	GU188828	GU199546	GU201806	GU253416	HM119938	HM593109	HM593235
GQ433873	GU188829	GU199548	GU201807	GU253417	HM119953	HM593112	HM593237
GQ433874	GU188830	GU199549	GU201808	GU253418	HM119975	HM593113	HM593238
GQ433875	GU188831	GU199550	GU201809	GU253419	HM120023	HM593115	HM593243
GQ433876	GU188832	GU199551	GU201810	GU253420	HM120046	HM593121	HM593244
GQ433877	GU188833	GU199552	GU201811	GU253422	HM120061	HM593122	HM593246
GQ433878	GU188834	GU199553	GU201812	GU253423	HM120062	HM593126	HM593247
GQ433879	GU188835	GU199554	GU201813	GU253424	HM120091	HM593128	HM593253

GQ433880	GU188836	GU199555	GU201814	GU253425	HM120150	HM593129	HM593254
HM593255	HM593300	HM593346	HM593391	HM593438	HM593498	HQ166857	HQ453370
HM593256	HM593301	HM593347	HM593392	HM593440	HM593499	HQ166858	HQ453371
HM593257	HM593302	HM593348	HM593393	HM593441	HM593500	HQ166859	HQ453372
HM593258	HM593303	HM593349	HM593394	HM593442	HM593501	HQ166860	HQ453373
HM593259	HM593304	HM593350	HM593396	HM593443	HM593502	HQ166861	HQ453374
HM593260	HM593305	HM593351	HM593397	HM593444	HM593503	HQ166862	HQ453375
HM593261	HM593306	HM593352	HM593398	HM593445	HM593505	HQ166863	HQ453376
HM593262	HM593307	HM593353	HM593399	HM593446	HM593507	HQ166864	HQ453377
HM593263	HM593308	HM593354	HM593400	HM593447	HM593510	HQ166865	HQ453378
HM593265	HM593309	HM593355	HM593401	HM593448	HM627515	HQ166866	HQ453379
HM593266	HM593310	HM593356	HM593402	HM593449	HM752148	HQ166867	HQ453380
HM593267	HM593311	HM593357	HM593403	HM593450	HM752151	HQ166868	HQ453381
HM593268	HM593312	HM593358	HM593404	HM593451	HM752152	HQ166869	HQ453382
HM593269	HM593313	HM593359	HM593405	HM593452	HM752158	HQ166870	HQ453384
HM593270	HM593315	HM593360	HM593407	HM593453	HM752160	HQ166871	HQ453385
HM593271	HM593316	HM593361	HM593408	HM593454	HM752163	HQ166872	HQ453386
HM593272	HM593317	HM593362	HM593409	HM593455	HM752164	HQ166873	HQ453387
HM593273	HM593318	HM593363	HM593410	HM593456	HM752165	HQ166874	HQ453388
HM593274	HM593319	HM593364	HM593411	HM593457	HM752175	HQ166875	HQ453389
HM593275	HM593320	HM593366	HM593412	HM593458	HM752176	HQ166876	HQ453390
HM593276	HM593321	HM593367	HM593413	HM593459	HM752177	HQ166877	HQ453392
HM593277	HM593322	HM593368	HM593414	HM593460	HM752180	HQ166878	HQ453393
HM593278	HM593323	HM593369	HM593415	HM593461	HM752186	HQ166879	HQ453394
HM593279	HM593324	HM593370	HM593416	HM593462	HM752191	HQ166880	HQ539009
HM593280	HM593325	HM593371	HM593418	HM593463	HM854124	HQ166881	HQ539035
HM593281	HM593326	HM593372	HM593419	HM593464	HQ108352	HQ166882	HQ539197
HM593282	HM593327	HM593373	HM593420	HM593465	HQ108366	HQ166883	HQ588277
HM593283	HM593328	HM593374	HM593421	HM593466	HQ166841	HQ166884	HQ638705
HM593284	HM593329	HM593375	HM593422	HM593468	HQ166842	HQ166885	HQ638709
HM593285	HM593330	HM593377	HM593423	HM593470	HQ166843	HQ166886	HQ638713
HM593286	HM593331	HM593378	HM593424	HM593471	HQ166844	HQ166887	HQ638727
HM593287	HM593332	HM593379	HM593425	HM593473	HQ166845	HQ234749	HQ638739
HM593288	HM593333	HM593380	HM593426	HM593475	HQ166846	HQ234750	HQ696823
HM593289	HM593335	HM593381	HM593427	HM593478	HQ166847	HQ234751	HQ696824
HM593290	HM593336	HM593382	HM593429	HM593479	HQ166848	HQ234752	HQ696825
HM593291	HM593337	HM593383	HM593430	HM593483	HQ166849	HQ234753	HQ696826
HM593293	HM593338	HM593384	HM593431	HM593484	HQ166850	HQ234754	HQ696827
HM593294	HM593339	HM593385	HM593432	HM593485	HQ166851	HQ453364	HQ696828
HM593295	HM593340	HM593386	HM593433	HM593486	HQ166852	HQ453365	HQ702662
HM593296	HM593341	HM593387	HM593434	HM593489	HQ166853	HQ453366	HQ844570
HM593297	HM593342	HM593388	HM593435	HM593491	HQ166854	HQ453367	HQ844571
HM593298	HM593343	HM593389	HM593436	HM593493	HQ166855	HQ453368	HQ844572

HM593299	HM593344	HM593390	HM593437	HM593494	HQ166856	HQ453369	HQ844573
HQ844574	HQ916283	HQ993962	HQ994204	HQ994249	HQ994294	HQ994338	HQ994383
HQ844575	HQ916284	HQ993963	HQ994205	HQ994250	HQ994295	HQ994339	HQ994384
HQ844576	HQ916286	HQ993969	HQ994206	HQ994251	HQ994296	HQ994340	HQ994385
HQ844577	HQ916287	HQ994129	HQ994207	HQ994252	HQ994297	HQ994341	HQ994386
HQ844578	HQ916288	HQ994148	HQ994208	HQ994253	HQ994298	HQ994342	HQ994387
HQ844579	HQ916289	HQ994162	HQ994209	HQ994254	HQ994299	HQ994343	HQ994388
HQ844580	HQ916290	HQ994163	HQ994210	HQ994255	HQ994300	HQ994344	HQ994389
HQ844581	HQ916291	HQ994164	HQ994211	HQ994256	HQ994301	HQ994345	HQ994390
HQ844582	HQ993580	HQ994165	HQ994212	HQ994257	HQ994302	HQ994346	HQ994391
HQ844583	HQ993597	HQ994166	HQ994213	HQ994258	HQ994303	HQ994347	HQ994392
HQ844584	HQ993611	HQ994167	HQ994214	HQ994259	HQ994304	HQ994348	HQ994393
HQ844585	HQ993615	HQ994168	HQ994215	HQ994260	HQ994305	HQ994349	HQ994394
HQ844586	HQ993623	HQ994169	HQ994216	HQ994261	HQ994306	HQ994350	HQ994395
HQ844587	HQ993632	HQ994170	HQ994217	HQ994262	HQ994307	HQ994351	HQ994396
HQ844588	HQ993653	HQ994172	HQ994218	HQ994263	HQ994308	HQ994352	HQ994397
HQ844589	HQ993678	HQ994173	HQ994219	HQ994264	HQ994309	HQ994354	HQ994398
HQ844590	HQ993684	HQ994174	HQ994220	HQ994265	HQ994310	HQ994355	HQ994399
HQ844591	HQ993690	HQ994175	HQ994221	HQ994266	HQ994311	HQ994356	HQ994400
HQ844592	HQ993692	HQ994176	HQ994222	HQ994267	HQ994312	HQ994357	HQ994401
HQ844593	HQ993699	HQ994177	HQ994223	HQ994269	HQ994313	HQ994358	HQ994402
HQ844594	HQ993702	HQ994178	HQ994224	HQ994270	HQ994314	HQ994359	HQ994403
HQ844595	HQ993729	HQ994179	HQ994225	HQ994271	HQ994315	HQ994360	HQ994405
HQ844596	HQ993734	HQ994180	HQ994228	HQ994272	HQ994316	HQ994361	HQ994406
HQ844597	HQ993750	HQ994181	HQ994229	HQ994273	HQ994317	HQ994362	HQ994407
HQ844598	HQ993789	HQ994182	HQ994230	HQ994274	HQ994318	HQ994363	HQ994408
HQ844599	HQ993806	HQ994183	HQ994231	HQ994275	HQ994319	HQ994364	HQ994409
HQ844600	HQ993813	HQ994184	HQ994232	HQ994276	HQ994320	HQ994366	HQ994410
HQ844601	HQ993819	HQ994185	HQ994233	HQ994277	HQ994321	HQ994367	HQ994411
HQ844602	HQ993831	HQ994186	HQ994234	HQ994278	HQ994322	HQ994368	HQ994412
HQ844603	HQ993833	HQ994187	HQ994235	HQ994279	HQ994323	HQ994369	HQ994413
HQ844604	HQ993837	HQ994188	HQ994236	HQ994280	HQ994324	HQ994370	HQ994414
HQ844605	HQ993849	HQ994191	HQ994237	HQ994281	HQ994325	HQ994371	HQ994415
HQ844606	HQ993858	HQ994192	HQ994238	HQ994282	HQ994326	HQ994372	HQ994416
HQ844607	HQ993886	HQ994193	HQ994239	HQ994283	HQ994327	HQ994373	HQ994417
HQ844608	HQ993891	HQ994194	HQ994240	HQ994284	HQ994329	HQ994374	HQ994420
HQ844609	HQ993909	HQ994195	HQ994241	HQ994285	HQ994330	HQ994375	HQ994421
HQ844610	HQ993926	HQ994197	HQ994242	HQ994286	HQ994331	HQ994376	HQ994422
HQ844611	HQ993927	HQ994198	HQ994243	HQ994287	HQ994332	HQ994377	HQ994423
HQ844612	HQ993937	HQ994199	HQ994244	HQ994289	HQ994333	HQ994378	HQ994424
HQ844613	HQ993938	HQ994200	HQ994245	HQ994290	HQ994334	HQ994379	HQ994425
HQ844614	HQ993943	HQ994201	HQ994246	HQ994291	HQ994335	HQ994380	HQ994426
HQ844615	HQ993949	HQ994202	HQ994247	HQ994292	HQ994336	HQ994381	HQ994427

HQ844616	HQ993952	HQ994203	HQ994248	HQ994293	HQ994337	HQ994382	HQ994428
HQ994429	HQ994476	HQ994522	HQ994572	HQ994621	HQ994671	HQ994721	HQ994767
HQ994430	HQ994477	HQ994523	HQ994573	HQ994622	HQ994672	HQ994722	HQ994768
HQ994431	HQ994478	HQ994524	HQ994574	HQ994623	HQ994673	HQ994723	HQ994769
HQ994432	HQ994479	HQ994525	HQ994575	HQ994624	HQ994674	HQ994724	HQ994770
HQ994433	HQ994480	HQ994526	HQ994576	HQ994625	HQ994675	HQ994725	HQ994771
HQ994434	HQ994481	HQ994527	HQ994577	HQ994626	HQ994676	HQ994726	HQ994772
HQ994435	HQ994482	HQ994528	HQ994578	HQ994627	HQ994677	HQ994727	HQ994773
HQ994436	HQ994483	HQ994529	HQ994579	HQ994628	HQ994678	HQ994728	HQ994774
HQ994437	HQ994484	HQ994530	HQ994580	HQ994629	HQ994679	HQ994729	HQ994775
HQ994438	HQ994485	HQ994532	HQ994581	HQ994630	HQ994681	HQ994730	HQ994776
HQ994439	HQ994486	HQ994534	HQ994582	HQ994631	HQ994682	HQ994731	HQ994777
HQ994441	HQ994487	HQ994535	HQ994583	HQ994632	HQ994683	HQ994732	HQ994778
HQ994442	HQ994488	HQ994536	HQ994584	HQ994633	HQ994684	HQ994733	HQ994779
HQ994444	HQ994489	HQ994537	HQ994585	HQ994634	HQ994685	HQ994734	HQ994780
HQ994445	HQ994490	HQ994538	HQ994586	HQ994635	HQ994686	HQ994735	HQ994781
HQ994446	HQ994491	HQ994539	HQ994587	HQ994636	HQ994689	HQ994736	HQ994782
HQ994447	HQ994492	HQ994540	HQ994588	HQ994637	HQ994690	HQ994737	HQ994783
HQ994448	HQ994493	HQ994542	HQ994590	HQ994638	HQ994691	HQ994738	HQ994784
HQ994449	HQ994494	HQ994543	HQ994591	HQ994639	HQ994692	HQ994739	HQ994785
HQ994450	HQ994495	HQ994545	HQ994592	HQ994641	HQ994693	HQ994740	HQ994786
HQ994451	HQ994496	HQ994546	HQ994593	HQ994643	HQ994695	HQ994741	HQ994787
HQ994452	HQ994497	HQ994547	HQ994594	HQ994644	HQ994697	HQ994742	HQ994788
HQ994453	HQ994498	HQ994548	HQ994596	HQ994645	HQ994698	HQ994743	HQ994789
HQ994454	HQ994499	HQ994549	HQ994597	HQ994646	HQ994699	HQ994744	HQ994790
HQ994455	HQ994500	HQ994550	HQ994598	HQ994647	HQ994700	HQ994746	HQ994791
HQ994456	HQ994501	HQ994551	HQ994599	HQ994648	HQ994701	HQ994747	HQ994792
HQ994457	HQ994502	HQ994552	HQ994600	HQ994649	HQ994702	HQ994748	HQ994794
HQ994458	HQ994503	HQ994553	HQ994601	HQ994651	HQ994703	HQ994749	HQ994795
HQ994459	HQ994504	HQ994554	HQ994602	HQ994653	HQ994704	HQ994750	HQ994796
HQ994460	HQ994505	HQ994555	HQ994603	HQ994654	HQ994705	HQ994751	HQ994797
HQ994461	HQ994506	HQ994556	HQ994604	HQ994655	HQ994706	HQ994752	HQ994798
HQ994462	HQ994507	HQ994557	HQ994605	HQ994656	HQ994707	HQ994753	HQ994799
HQ994463	HQ994508	HQ994560	HQ994606	HQ994657	HQ994709	HQ994754	HQ994800
HQ994465	HQ994510	HQ994561	HQ994607	HQ994658	HQ994710	HQ994755	HQ994801
HQ994466	HQ994511	HQ994562	HQ994608	HQ994659	HQ994711	HQ994757	HQ994802
HQ994467	HQ994512	HQ994563	HQ994609	HQ994660	HQ994712	HQ994758	HQ994803
HQ994469	HQ994513	HQ994564	HQ994612	HQ994662	HQ994713	HQ994759	HQ994804
HQ994470	HQ994514	HQ994565	HQ994614	HQ994663	HQ994714	HQ994760	HQ994805
HQ994471	HQ994515	HQ994566	HQ994615	HQ994664	HQ994715	HQ994761	HQ994806
HQ994472	HQ994516	HQ994567	HQ994617	HQ994665	HQ994716	HQ994762	HQ994807
HQ994473	HQ994517	HQ994568	HQ994618	HQ994666	HQ994718	HQ994764	HQ994808
HQ994474	HQ994518	HQ994569	HQ994619	HQ994669	HQ994719	HQ994765	HQ994809

HQ994475	HQ994519	HQ994570	HQ994620	HQ994670	HQ994720	HQ994766	HQ994811
HQ994812	HQ994856	HQ994899	JF414421	JF414465	JF505601	JN010590	JN222644
HQ994813	HQ994857	HQ994900	JF414422	JF414466	JF773983	JN010592	JN628493
HQ994814	HQ994858	HQ994901	JF414423	JF414467	JF774000	JN010599	JN628500
HQ994815	HQ994859	HQ994902	JF414424	JF414468	JF774001	JN010601	JN628507
HQ994816	HQ994860	HQ994903	JF414425	JF487831	JF774014	JN010603	JN628518
HQ994817	HQ994861	HQ994904	JF414426	JF487832	JF774028	JN010606	JN639188
HQ994818	HQ994862	HQ994905	JF414427	JF487833	JF774043	JN010618	JN639189
HQ994819	HQ994863	HQ994906	JF414428	JF487834	JF774050	JN010620	JN639190
HQ994820	HQ994864	HQ994907	JF414429	JF487835	JF774051	JN010622	JN639191
HQ994821	HQ994865	HQ994908	JF414430	JF487836	JF774053	JN010624	JN639192
HQ994822	HQ994866	HQ994909	JF414432	JF487837	JF774055	JN010625	JN639193
HQ994823	HQ994867	HQ994910	JF414433	JF487838	JF906616	JN010629	JN639194
HQ994824	HQ994868	HQ994911	JF414434	JF487841	JF906695	JN010637	JN639195
HQ994825	HQ994869	HQ994912	JF414435	JF487843	JF906698	JN010640	JN639196
HQ994826	HQ994870	HQ994913	JF414436	JF487844	JN000018	JN010642	JN639197
HQ994827	HQ994871	HQ994914	JF414437	JF487851	JN010442	JN010644	JN639199
HQ994828	HQ994872	HQ994915	JF414438	JF487852	JN010446	JN010673	JN639200
HQ994829	HQ994873	HQ994916	JF414439	JF487853	JN010448	JN010697	JN639201
HQ994830	HQ994874	HQ994917	JF414440	JF487855	JN010449	JN010702	JN639202
HQ994831	HQ994875	HQ994932	JF414441	JF487856	JN010462	JN010704	JN639203
HQ994832	HQ994876	HQ994940	JF414442	JF487864	JN010463	JN010707	JN639205
HQ994833	HQ994877	HQ994996	JF414443	JF487867	JN010467	JN010710	JN639206
HQ994834	HQ994878	HQ995027	JF414444	JF487868	JN010470	JN010713	JN639207
HQ994835	HQ994879	HQ995037	JF414445	JF487875	JN010474	JN010715	JN639208
HQ994836	HQ994880	HQ995038	JF414446	JF487878	JN010476	JN010718	JN639209
HQ994837	HQ994881	HQ995112	JF414447	JF487881	JN010492	JN010720	JN639210
HQ994838	HQ994882	HQ995246	JF414448	JF487883	JN010494	JN010727	JN639211
HQ994839	HQ994883	HQ995288	JF414449	JF487885	JN010496	JN010729	JN639212
HQ994840	HQ994884	HQ995401	JF414450	JF487886	JN010500	JN010734	JN639213
HQ994842	HQ994885	HQ995414	JF414451	JF487887	JN010502	JN010739	JN639214
HQ994843	HQ994886	HQ995477	JF414452	JF487889	JN010508	JN010741	JN664970
HQ994844	HQ994887	HQ995493	JF414453	JF487891	JN010512	JN010743	JN664971
HQ994845	HQ994888	HQ995494	JF414454	JF487894	JN010517	JN010753	JN664972
HQ994846	HQ994889	HQ995495	JF414455	JF487905	JN010519	JN010758	JN664973
HQ994847	HQ994890	HQ995496	JF414456	JF487906	JN010521	JN010761	JN664974
HQ994848	HQ994891	HQ995497	JF414457	JF487907	JN010524	JN010764	JN664975
HQ994849	HQ994892	JF342273	JF414458	JF487908	JN010540	JN114121	JN664976
HQ994850	HQ994893	JF342277	JF414459	JF487909	JN010550	JN114216	JN664977
HQ994851	HQ994894	JF342290	JF414460	JF487910	JN010553	JN214931	JN664978
HQ994852	HQ994895	JF342301	JF414461	JF487911	JN010555	JN214962	JN664979
HQ994853	HQ994896	JF342317	JF414462	JF487912	JN010557	JN215115	JN664980
HQ994854	HQ994897	JF414419	JF414463	JF487913	JN010568	JN215125	JN664981

HQ994855	HQ994898	JF414420	JF414464	JF487914	JN010584	JN222612	JN664982
JN664983	JN665027	JQ658521	JQ658627	JQ658772	JX256211	JX300330	JX507823
JN664984	JN665028	JQ658522	JQ658629	JQ698679	JX256212	JX300339	JX507824
JN664985	JN665029	JQ658527	JQ658630	JQ698681	JX256213	JX300343	JX507825
JN664986	JN665030	JQ658529	JQ658632	JQ698683	JX256214	JX300363	JX507826
JN664987	JN665031	JQ658530	JQ658639	JQ698684	JX256215	JX300376	JX507827
JN664988	JN665032	JQ658531	JQ658642	JQ698687	JX256216	JX300427	JX507828
JN664989	JN665033	JQ658536	JQ658644	JQ698692	JX256218	JX300428	JX507829
JN664990	JN665034	JQ658538	JQ658651	JQ698695	JX299585	JX300450	JX507830
JN664991	JN665035	JQ658543	JQ658652	JQ698715	JX299590	JX300458	JX507831
JN664992	JN665036	JQ658544	JQ658653	JQ698722	JX299607	JX300526	JX507832
JN664993	JN665037	JQ658548	JQ658654	JQ698723	JX299624	JX300602	JX507833
JN664994	JN665038	JQ658551	JQ658658	JQ698725	JX299674	JX300649	JX507834
JN664995	JN665039	JQ658554	JQ658663	JQ698726	JX299706	JX300661	JX507835
JN664996	JN665040	JQ658557	JQ658670	JQ698731	JX299739	JX300755	JX507836
JN664997	JN665041	JQ658558	JQ658674	JQ698747	JX299741	JX300776	JX507837
JN664998	JQ071460	JQ658559	JQ658675	JQ698760	JX299743	JX300789	JX507838
JN664999	JQ071464	JQ658560	JQ658699	JQ698778	JX299827	JX300813	JX507839
JN665000	JQ071465	JQ658566	JQ658705	JQ698783	JX299847	JX300863	JX507840
JN665001	JQ071468	JQ658570	JQ658710	JQ698790	JX299877	JX300865	JX507841
JN665002	JQ071471	JQ658571	JQ658712	JQ698791	JX299904	JX300897	JX507842
JN665003	JQ071474	JQ658574	JQ658714	JQ698796	JX299916	JX300969	JX507843
JN665004	JQ071475	JQ658576	JQ658715	JQ698802	JX299940	JX300989	JX507844
JN665005	JQ071478	JQ658577	JQ658716	JQ698822	JX299990	JX300993	JX507845
JN665006	JQ071483	JQ658578	JQ658718	JQ698827	JX300049	JX301014	JX507846
JN665007	JQ071486	JQ658580	JQ658724	JQ698828	JX300071	JX301068	JX507847
JN665008	JQ071488	JQ658581	JQ658726	JQ698843	JX300090	JX301078	JX507848
JN665009	JQ071491	JQ658585	JQ658731	JQ750698	JX300105	JX301105	JX507849
JN665010	JQ266060	JQ658586	JQ658732	JX256193	JX300115	JX301120	JX507850
JN665011	JQ658476	JQ658589	JQ658734	JX256194	JX300118	JX499012	JX507851
JN665012	JQ658480	JQ658591	JQ658735	JX256196	JX300121	JX507809	JX507852
JN665014	JQ658488	JQ658592	JQ658736	JX256197	JX300129	JX507810	JX507853
JN665015	JQ658497	JQ658594	JQ658740	JX256198	JX300164	JX507811	JX507854
JN665016	JQ658498	JQ658599	JQ658744	JX256199	JX300176	JX507812	JX507855
JN665017	JQ658503	JQ658600	JQ658745	JX256200	JX300198	JX507813	JX507856
JN665018	JQ658504	JQ658604	JQ658751	JX256201	JX300222	JX507814	JX507857
JN665019	JQ658505	JQ658608	JQ658753	JX256202	JX300223	JX507815	JX507858
JN665020	JQ658507	JQ658615	JQ658754	JX256204	JX300233	JX507816	JX507859
JN665021	JQ658511	JQ658616	JQ658755	JX256205	JX300254	JX507817	JX507860
JN665022	JQ658512	JQ658617	JQ658756	JX256206	JX300261	JX507818	JX507861
JN665023	JQ658513	JQ658620	JQ658757	JX256207	JX300271	JX507819	JX507862
JN665024	JQ658514	JQ658621	JQ658766	JX256208	JX300322	JX507820	JX507863
JN665025	JQ658519	JQ658624	JQ658767	JX256209	JX300325	JX507821	JX507864

JN665026	JQ658520	JQ658626	JQ658769	JX256210	JX300326	JX507822	JX507865
JX507866	JX507909	KC537232	KC764457	KF026065	KF026111	KF026157	KF026202
JX507867	JX507910	KC537233	KC764458	KF026067	KF026112	KF026158	KF026204
JX507868	JX507911	KC537234	KC810411	KF026068	KF026113	KF026159	KF026207
JX507869	JX507912	KC537235	KC810488	KF026069	KF026114	KF026160	KF026209
JX507870	JX507913	KC537236	KC810511	KF026070	KF026116	KF026161	KF059615
JX507871	JX507914	KC537242	KC841660	KF026071	KF026117	KF026162	KF059617
JX507872	JX507915	KC537243	KC841663	KF026072	KF026118	KF026163	KF059702
JX507873	JX507916	KC740523	KC841674	KF026073	KF026119	KF026164	KF530909
JX507874	JX507917	KC764421	KC841688	KF026074	KF026120	KF026165	KF530955
JX507875	JX507918	KC764422	KC841702	KF026075	KF026121	KF026166	KF530988
JX507876	JX507919	KC764423	KC841719	KF026076	KF026122	KF026167	KF531001
JX507877	JX507920	KC764424	KC841733	KF026077	KF026123	KF026168	KF531053
JX507878	JX847172	KC764425	KC841738	KF026078	KF026126	KF026169	KF531072
JX507879	JX847173	KC764426	KC841744	KF026079	KF026127	KF026170	KF531144
JX507880	JX847174	KC764427	KC841756	KF026080	KF026128	KF026171	KF531156
JX507881	JX966498	KC764428	KC841757	KF026081	KF026129	KF026172	KF531186
JX507882	JX966501	KC764429	KC841762	KF026082	KF026130	KF026173	KF531197
JX507883	JX966511	KC764430	KC841768	KF026083	KF026131	KF026174	KF531283
JX507884	JX966535	KC764431	KC879721	KF026084	KF026132	KF026175	KF531287
JX507885	KC133219	KC764432	KC879722	KF026085	KF026133	KF026176	KF531293
JX507886	KC181291	KC764433	KC879723	KF026086	KF026134	KF026177	KF531337
JX507887	KC286155	KC764434	KC879725	KF026087	KF026135	KF026178	KF531507
JX507888	KC340200	KC764435	KC879726	KF026089	KF026136	KF026179	KF671784
JX507889	KC340392	KC764436	KC879727	KF026090	KF026137	KF026180	KF671785
JX507890	KC340393	KC764437	KC879728	KF026092	KF026138	KF026181	KF671790
JX507891	KC340422	KC764438	KC879729	KF026093	KF026139	KF026182	KF700961
JX507892	KC340491	KC764439	KC879730	KF026094	KF026140	KF026183	KF727691
JX507893	KC340492	KC764440	KC879731	KF026095	KF026141	KF026184	KF727692
JX507894	KC340804	KC764441	KC879732	KF026096	KF026142	KF026186	KF727693
JX507895	KC340887	KC764442	KC921843	KF026097	KF026143	KF026187	KF727694
JX507896	KC516875	KC764444	KC970885	KF026098	KF026144	KF026188	KF727695
JX507897	KC516889	KC764445	KC988060	KF026099	KF026145	KF026189	KF727696
JX507898	KC516939	KC764446	KC994221	KF026100	KF026146	KF026190	KF727697
JX507899	KC516954	KC764447	KC994246	KF026101	KF026147	KF026191	KF727698
JX507900	KC516970	KC764448	KC994280	KF026102	KF026148	KF026192	KF727699
JX507901	KC516987	KC764449	KC994361	KF026103	KF026149	KF026193	KF727700
JX507902	KC517017	KC764450	KC994370	KF026104	KF026150	KF026194	KF727701
JX507903	KC517029	KC764451	KC994382	KF026105	KF026151	KF026195	KF727702
JX507904	KC517037	KC764452	KF026059	KF026106	KF026152	KF026196	KF727703
JX507905	KC517038	KC764453	KF026060	KF026107	KF026153	KF026197	KF727704
JX507906	KC517042	KC764454	KF026061	KF026108	KF026154	KF026198	KF727705
JX507907	KC517043	KC764455	KF026063	KF026109	KF026155	KF026199	KF727706

JX507908	KC517051	KC764456	KF026064	KF026110	KF026156	KF026200	KF727707
KF727708	KF780982	KF781056	KF793166	KF806979	KF890232	KF927203	KF927364
KF727709	KF780984	KF781057	KF793168	KF806980	KF890233	KF927204	KF927365
KF727710	KF780985	KF781058	KF793170	KF806981	KF890234	KF927205	KF927366
KF727711	KF780987	KF781059	KF793173	KF806982	KF890235	KF927206	KF927367
KF727712	KF780988	KF781060	KF793178	KF806983	KF890236	KF927207	KF927368
KF727713	KF780989	KF781061	KF793179	KF853754	KF890237	KF927209	KF927370
KF727715	KF780992	KF781062	KF793180	KF853759	KF890238	KF927210	KF927371
KF727716	KF780995	KF781063	KF806938	KF853761	KF890239	KF927211	KF927372
KF727717	KF780996	KF781064	KF806939	KF853771	KF890240	KF927212	KF927373
KF727718	KF780998	KF781065	KF806940	KF853777	KF890241	KF927213	KF927374
KF727719	KF781001	KF781066	KF806941	KF853782	KF890243	KF927214	KF927376
KF727720	KF781003	KF781067	KF806943	KF853793	KF927170	KF927215	KF927379
KF727721	KF781006	KF781068	KF806944	KF853797	KF927171	KF927217	KF927380
KF727722	KF781007	KF781069	KF806945	KF853810	KF927172	KF927218	KF927381
KF727723	KF781009	KF781070	KF806946	KF853816	KF927173	KF927219	KF927383
KF727724	KF781010	KF782764	KF806947	KF853818	KF927174	KF927221	KF927384
KF727725	KF781011	KF782771	KF806948	KF853822	KF927175	KF927222	KF927385
KF727726	KF781013	KF782772	KF806949	KF853837	KF927176	KF927223	KF927387
KF727727	KF781018	KF793123	KF806950	KF853857	KF927177	KF927224	KF927388
KF727728	KF781020	KF793125	KF806951	KF853861	KF927178	KF927225	KF927390
KF727729	KF781021	KF793126	KF806952	KF853864	KF927179	KF927226	KF927393
KF727730	KF781023	KF793127	KF806954	KF853869	KF927180	KF927227	KF927394
KF727731	KF781024	KF793128	KF806955	KF853871	KF927181	KF927228	KF927396
KF727732	KF781025	KF793129	KF806956	KF853873	KF927182	KF927229	KF927397
KF766537	KF781026	KF793130	KF806958	KF853874	KF927183	KF927230	KF927398
KF766540	KF781027	KF793131	KF806959	KF853876	KF927184	KF927231	KF927399
KF768007	KF781028	KF793132	KF806960	KF853877	KF927185	KF927232	KF927402
KF780965	KF781031	KF793133	KF806961	KF853878	KF927186	KF927233	KF927403
KF780966	KF781032	KF793135	KF806962	KF853881	KF927188	KF927234	KF927404
KF780967	KF781033	KF793137	KF806963	KF853882	KF927189	KF927235	KF927405
KF780968	KF781034	KF793141	KF806965	KF890214	KF927190	KF927237	KF927406
KF780969	KF781035	KF793143	KF806966	KF890217	KF927191	KF927238	KF927407
KF780970	KF781036	KF793144	KF806968	KF890218	KF927192	KF927239	KF927408
KF780971	KF781037	KF793148	KF806969	KF890220	KF927193	KF927240	KF927409
KF780972	KF781039	KF793150	KF806970	KF890221	KF927194	KF927241	KF927410
KF780973	KF781040	KF793151	KF806971	KF890222	KF927195	KF927242	KF927411
KF780974	KF781041	KF793152	KF806972	KF890223	KF927196	KF927243	KF927412
KF780975	KF781043	KF793153	KF806973	KF890224	KF927197	KF927244	KF927413
KF780976	KF781046	KF793154	KF806974	KF890225	KF927198	KF927245	KF927414
KF780977	KF781049	KF793158	KF806975	KF890228	KF927199	KF927358	KF927415
KF780978	KF781050	KF793160	KF806976	KF890229	KF927200	KF927359	KF927416
KF780980	KF781052	KF793164	KF806977	KF890230	KF927201	KF927362	KF927417

KF780981	KF781055	KF793165	KF806978	KF890231	KF927202	KF927363	KF927418
KF927419	KJ770722	KJ807674	KJ807717	KM049920	KM050009	KM050084	KM050152
KF927420	KJ770805	KJ807675	KJ807718	KM049921	KM050010	KM050085	KM050154
KF927421	KJ770815	KJ807676	KJ807719	KM049922	KM050011	KM050086	KM050155
KF927422	KJ771012	KJ807677	KJ807720	KM049923	KM050012	KM050087	KM050160
KF927423	KJ771138	KJ807678	KJ807721	KM049924	KM050019	KM050088	KM050161
KF927424	KJ771184	KJ807679	KJ807722	KM049925	KM050020	KM050089	KM050176
KF927426	KJ771234	KJ807680	KJ807723	KM049926	KM050021	KM050090	KM050178
KF927427	KJ771303	KJ807681	KJ807724	KM049927	KM050022	KM050091	KM050179
KF927428	KJ771313	KJ807682	KJ807725	KM049928	KM050023	KM050092	KM050180
KF927429	KJ771420	KJ807683	KJ807726	KM049935	KM050024	KM050093	KM050181
KF927431	KJ771422	KJ807684	KJ807727	KM049936	KM050025	KM050094	KM050182
KF927432	KJ771486	KJ807685	KJ807728	KM049943	KM050026	KM050096	KM050183
KF927433	KJ771588	KJ807686	KJ807729	KM049944	KM050027	KM050097	KM050184
KJ194683	KJ771644	KJ807687	KJ807730	KM049946	KM050028	KM050105	KM050191
KJ194694	KJ771675	KJ807688	KJ807731	KM049947	KM050030	KM050106	KM050192
KJ364641	KJ771680	KJ807689	KJ906630	KM049948	KM050031	KM050107	KM050199
KJ467927	KJ771785	KJ807690	KJ906709	KM049949	KM050032	KM050112	KM050200
KJ467929	KJ771788	KJ807691	KJ906739	KM049950	KM050033	KM050113	KM050201
KJ467991	KJ807649	KJ807692	KJ906772	KM049951	KM050040	KM050114	KM050202
KJ467999	KJ807650	KJ807693	KJ906793	KM049957	KM050041	KM050115	KM050209
KJ561122	KJ807651	KJ807694	KJ906795	KM049958	KM050048	KM050116	KM050210
KJ561123	KJ807652	KJ807695	KJ906879	KM049963	KM050055	KM050117	KM050216
KJ561124	KJ807653	KJ807696	KJ906977	KM049964	KM050056	KM050118	KM050217
KJ561125	KJ807654	KJ807697	KJ906991	KM049965	KM050057	KM050119	KM050218
KJ561126	KJ807655	KJ807698	KJ907007	KM049966	KM050059	KM050120	KM050225
KJ561127	KJ807656	KJ807699	KJ907019	KM049973	KM050060	KM050121	KM050226
KJ561128	KJ807657	KJ807700	KJ907025	KM049980	KM050067	KM050122	KM050227
KJ561129	KJ807658	KJ807701	KJ907053	KM049981	KM050068	KM050123	KM050228
KJ561130	KJ807659	KJ807702	KJ907054	KM049982	KM050069	KM050124	KM050229
KJ561131	KJ807660	KJ807703	KJ907110	KM049983	KM050070	KM050125	KM050230
KJ561132	KJ807661	KJ807704	KJ907157	KM049990	KM050071	KM050126	KM050231
KJ561133	KJ807662	KJ807705	KJ907166	KM049991	KM050072	KM050127	KM050232
KJ561134	KJ807663	KJ807706	KJ907169	KM049992	KM050073	KM050130	KM050233
KJ561135	KJ807664	KJ807707	KJ907174	KM049999	KM050074	KM050131	KM050238
KJ561136	KJ807665	KJ807708	KM049900	KM050000	KM050075	KM050132	KM050240
KJ722081	KJ807666	KJ807709	KM049901	KM050001	KM050076	KM050133	KM050243
KJ723062	KJ807667	KJ807710	KM049902	KM050002	KM050077	KM050135	KM050248
KJ723387	KJ807668	KJ807711	KM049903	KM050003	KM050078	KM050136	KM050249
KJ724222	KJ807669	KJ807712	KM049904	KM050004	KM050079	KM050137	KM050254
KJ769722	KJ807670	KJ807713	KM049905	KM050005	KM050080	KM050144	KM050255
KJ770036	KJ807671	KJ807714	KM049911	KM050006	KM050081	KM050145	KM050256
KJ770134	KJ807672	KJ807715	KM049912	KM050007	KM050082	KM050150	KM050257

KJ770421	KJ807673	KJ807716	KM049919	KM050008	KM050083	KM050151	KM050264
KM050265	KM050371	KM050463	KM050570	KM050650	KM050709	KM050763	KP224495
KM050266	KM050372	KM050468	KM050571	KM050651	KM050710	KM050764	KP224496
KM050267	KM050373	KM050469	KM050572	KM050652	KM050711	KM050765	KP224500
KM050272	KM050374	KM050474	KM050573	KM050653	KM050712	KM050766	KP224501
KM050279	KM050375	KM050475	KM050575	KM050654	KM050717	KM050767	KP227524
KM050280	KM050380	KM050494	KM050576	KM050661	KM050718	KM057331	KP227525
KM050285	KM050381	KM050495	KM050577	KM050662	KM050719	KM057334	KP227526
KM050286	KM050382	KM050500	KM050578	KM050663	KM050720	KM057337	KP227527
KM050291	KM050383	KM050501	KM050579	KM050664	KM050721	KM057346	KP227528
KM050292	KM050388	KM050502	KM050580	KM050665	KM050722	KM057348	KP227529
KM050297	KM050389	KM050503	KM050581	KM050666	KM050723	KM057354	KP227530
KM050298	KM050394	KM050508	KM050588	KM050667	KM050724	KM057355	KP227531
KM050299	KM050395	KM050509	KM050589	KM050668	KM050725	KM057359	KP227532
KM050300	KM050400	KM050510	KM050596	KM050669	KM050726	KM218448	KP227533
KM050304	KM050406	KM050511	KM050597	KM050670	KM050727	KM218449	KP227534
KM050305	KM050407	KM050512	KM050601	KM050671	KM050728	KM218450	KP227535
KM050311	KM050408	KM050513	KM050602	KM050672	KM050729	KM218451	KP227536
KM050316	KM050409	KM050514	KM050609	KM050673	KM050736	KM218452	KP227537
KM050317	KM050410	KM050515	KM050610	KM050674	KM050737	KM218453	KP227538
KM050320	KM050411	KM050520	KM050617	KM050677	KM050739	KM218454	KP227539
KM050321	KM050416	KM050521	KM050618	KM050678	KM050740	KM218455	KP227540
KM050326	KM050417	KM050526	KM050623	KM050679	KM050741	KM218456	KP227541
KM050327	KM050422	KM050527	KM050624	KM050680	KM050742	KM218457	KP227542
KM050328	KM050423	KM050532	KM050625	KM050681	KM050743	KM218458	KP227543
KM050329	KM050428	KM050538	KM050631	KM050682	KM050744	KM218459	KP227544
KM050334	KM050429	KM050539	KM050632	KM050683	KM050745	KM218460	KP227545
KM050335	KM050430	KM050540	KM050633	KM050684	KM050746	KM283950	KP227546
KM050340	KM050431	KM050541	KM050634	KM050685	KM050747	KM284236	KP227547
KM050341	KM050437	KM050547	KM050635	KM050686	KM050748	KM284256	KP227548
KM050347	KM050438	KM050548	KM050636	KM050687	KM050749	KM284345	KP227549
KM050350	KM050439	KM050553	KM050637	KM050688	KM050750	KM284424	KP227550
KM050351	KM050440	KM050554	KM050638	KM050689	KM050751	KM284520	KP227552
KM050352	KM050441	KM050556	KM050639	KM050690	KM050752	KM820312	KP227553
KM050353	KM050442	KM050557	KM050640	KM050691	KM050753	KP224476	KP227554
KM050358	KM050447	KM050558	KM050641	KM050692	KM050754	KP224479	KP227555
KM050359	KM050448	KM050559	KM050642	KM050699	KM050755	KP224483	KP227556
KM050360	KM050449	KM050560	KM050643	KM050700	KM050756	KP224484	KP227557
KM050361	KM050450	KM050564	KM050644	KM050701	KM050757	KP224485	KP227558
KM050362	KM050455	KM050565	KM050645	KM050702	KM050758	KP224487	KP227559
KM050363	KM050456	KM050566	KM050646	KM050703	KM050759	KP224489	KP227560
KM050364	KM050460	KM050567	KM050647	KM050704	KM050760	KP224491	KP227561
KM050365	KM050461	KM050568	KM050648	KM050705	KM050761	KP224492	KP227563

KM050370	KM050462	KM050569	KM050649	KM050706	KM050762	KP224493	KP227564
KP227565	KP227609	KP411832	KT318906	KP227600	KP227644	KT318898	KT318928
KP227566	KP227610	KP411833	KT318907	KP227601	KP227646	KT318899	KT318929
KP227567	KP227611	KP411834	KT318908	KP227602	KP227647	KT318900	KT318930
KP227568	KP227612	KP411836	KT318909	KP227604	KP227648	KT318901	KT318931
KP227569	KP227613	KP411837	KT318910	KP227605	KP227649	KT318902	KT318932
KP227570	KP227614	KP411838	KT318911	KP227606	KP227650	KT318903	KT318933
KP227571	KP227615	KP411839	KT318912	KP227607	KP411830	KT318904	KT318934
KP227572	KP227616	KP688090	KT318913	KP227608	KP411831	KT318905	KT343876
KP227573	KP227617	KP688093	KT318914	KP227586	KP227630	KT318883	KT343882
KP227574	KP227618	KP688112	KT318915	KP227587	KP227631	KT318884	KT343890
KP227575	KP227619	KP688119	KT318916	KP227588	KP227632	KT318885	KT343896
KP227576	KP227620	KP688142	KT318917	KP227589	KP227633	KT318886	KT343898
KP227577	KP227621	KP688143	KT318918	KP227590	KP227634	KT318887	KT343901
KP227578	KP227622	KP968789	KT318919	KP227591	KP227635	KT318888	KP227599
KP227579	KP227623	KP968802	KT318920	KP227592	KP227636	KT318890	KP227643
KP227580	KP227624	KP968803	KT318922	KP227593	KP227637	KT318891	KT318897
KP227581	KP227625	KP968826	KT318923	KP227594	KP227638	KT318892	KT318896
KP227582	KP227626	KT022362	KT318924	KP227595	KP227639	KT318893	KP227642
KP227583	KP227627	KT022366	KT318925	KP227596	KP227640	KT318894	KP227598
KP227584	KP227628	KT022371	KT318926	KP227597	KP227641	KT318895	KT318927
KP227585	KP227629	KT191993	U52953				

Accession numbers for treated Gag sequences.

DQ351220	KF527065	KF527076	KF526225
KF766542	KF527071	KF527173	KF526209
KF766541	KF527072	KF526222	KF526214
KF527061	KF527074	KF526208	KF526217
KF527062	KF527075	KF526212	KF526218
KF526207	KF526220	HQ846903	

Accession numbers for the naïve Gag sequences.

AY228556	EU786673	FJ606409	HM593308	JF704309	JF704596	KM048445	KM048730
AY228557	EU786681	FJ606410	HM593309	JF704310	JF704597	KM048446	KM048731
AY463217	EU884500	FJ606411	HM593310	JF704311	JF704598	KM048447	KM048732
AY463218	FJ155203	FJ606412	HM593311	JF704312	JF704599	KM048448	KM048733
AY463219	FJ606114	FJ606413	HM593312	JF704313	JF704600	KM048449	KM048734
AY463220	FJ606116	FJ606414	HM593313	JF704314	JF704601	KM048450	KM048735
AY463221	FJ606117	FJ606415	HM593314	JF704315	JF704602	KM048451	KM048736

AY463222	FJ606118	FJ606416	HM593315	JF704316	JF704603	KM048452	KM048737
AY463223	FJ606119	FJ606417	HM593316	JF704317	JF704604	KM048453	KM048738
AY463224	FJ606120	FJ606418	HM593317	JF704318	JF704605	KM048454	KM048739
AY463225	FJ606121	FJ606419	HM593318	JF704319	JF704606	KM048455	KM048740
AY463226	FJ606122	FJ606420	HM593319	JF704320	JF704607	KM048456	KM048741
AY463227	FJ606123	FJ606421	HM593320	JF704321	JF704608	KM048457	KM048742
AY463228	FJ606124	FJ606422	HM593321	JF704322	JF704609	KM048458	KM048743
AY463229	FJ606125	FJ606423	HM593322	JF704323	JF704610	KM048459	KM048744
AY463230	FJ606126	FJ606424	HM593323	JF704324	JF704611	KM048460	KM048745
AY463231	FJ606127	FJ606425	HM593324	JF704325	JF704612	KM048461	KM048746
AY463232	FJ606128	FJ606426	HM593325	JF704326	JF704613	KM048462	KM048747
AY463233	FJ606129	FJ606427	HM593326	JF704327	JF704614	KM048463	KM048748
AY463234	FJ606130	FJ606428	HM593327	JF704328	JF704615	KM048464	KM048749
AY463236	FJ606131	FJ606429	HM593328	JF704329	JF704616	KM048465	KM048750
AY463237	FJ606132	FJ606430	HM593329	JF704330	JF704617	KM048466	KM048751
AY484419	FJ606133	FJ606431	HM593330	JF704331	JF704618	KM048467	KM048752
AY484420	FJ606134	FJ606432	HM593331	JF704332	JF704619	KM048468	KM048753
AY484421	FJ606135	FJ606433	HM593332	JF704333	JF704620	KM048469	KM048754
AY484422	FJ606136	FJ606434	HM593333	JF704334	JF704621	KM048470	KM048755
AY484423	FJ606137	FJ606435	HM593334	JF704335	JF704622	KM048472	KM048756
AY484424	FJ606138	FJ606436	HM593335	JF704336	JF704623	KM048473	KM048757
AY585264	FJ606141	FJ606437	HM593336	JF704337	JF704624	KM048474	KM048758
AY585265	FJ606142	FJ606438	HM593337	JF704338	JF704625	KM048475	KM048759
AY585266	FJ606143	FJ606439	HM593338	JF704339	JF704626	KM048476	KM048760
AY585267	FJ606144	FJ606440	HM593339	JF704340	JF704627	KM048477	KM048761
AY585268	FJ606145	FJ606441	HM593340	JF704341	JF704628	KM048478	KM048762
AY703908	FJ606146	FJ606442	HM593341	JF704342	JF704629	KM048479	KM048763
AY703909	FJ606147	FJ606443	HM593342	JF704343	JF704630	KM048480	KM048764
AY703910	FJ606148	FJ606444	HM593343	JF704344	JF704631	KM048481	KM048765
AY703911	FJ606149	FJ606445	HM593344	JF704345	JF704632	KM048482	KM048766
AY772690	FJ606150	FJ606446	HM593345	JF704346	JF704633	KM048483	KM048767
AY772691	FJ606151	FJ670521	HM593346	JF704347	JF704634	KM048484	KM048768
AY772692	FJ606152	FJ853501	HM593347	JF704348	JF704635	KM048485	KM048769
AY772693	FJ606153	FJ853502	HM593348	JF704349	JF704636	KM048486	KM048770
AY772694	FJ606154	FJ853503	HM593349	JF704350	JF704637	KM048487	KM048771
AY772695	FJ606155	FJ853504	HM593350	JF704351	JF704638	KM048488	KM048772
AY772696	FJ606156	FJ853507	HM593351	JF704352	JF704639	KM048489	KM048773
AY772698	FJ606157	FJ853512	HM593352	JF704353	JF704640	KM048490	KM048774
AY772699	FJ606158	FJ853514	HM593353	JF704354	JF704641	KM048491	KM048775
AY772700	FJ606160	FJ853515	HM593354	JF704355	JF704642	KM048492	KM048776
AY838565	FJ606162	FJ853519	HM593355	JF704356	JF704643	KM048493	KM048777
AY838566	FJ606163	FJ853520	HM593356	JF704357	JF704644	KM048494	KM048778
AY838567	FJ606164	FJ853522	HM593357	JF704358	JF704645	KM048495	KM048779

AY838568	FJ606165	FJ853523	HM593358	JF704359	JF704646	KM048496	KM048780
AY878054	FJ606166	FJ853524	HM593359	JF704360	JF704647	KM048497	KM048781
AY878055	FJ606167	FJ853525	HM593360	JF704361	JF704648	KM048498	KM048782
AY878056	FJ606168	FJ853528	HM593361	JF704362	JF704649	KM048499	KM048783
AY878057	FJ606169	FJ853529	HM593362	JF704363	JF704650	KM048500	KM048784
AY878058	FJ606170	FJ853532	HM593363	JF704364	JF704651	KM048501	KM048785
AY878059	FJ606171	FJ853539	HM593364	JF704365	JF704652	KM048502	KM048786
AY878060	FJ606172	FJ853542	HM593365	JF704366	JF704653	KM048503	KM048787
AY878061	FJ606173	FJ853549	HM593366	JF704367	JF704654	KM048504	KM048788
AY878062	FJ606174	FJ853551	HM593367	JF704368	JF704655	KM048505	KM048789
AY878063	FJ606175	FJ853553	HM593368	JF704369	JF704656	KM048506	KM048790
AY878064	FJ606176	FJ853554	HM593369	JF704370	JF704657	KM048507	KM048791
AY878065	FJ606177	FJ853555	HM593370	JF704371	JF704658	KM048508	KM048792
AY878068	FJ606178	FJ853557	HM593371	JF704372	JF704659	KM048509	KM048793
AY878070	FJ606179	FJ853559	HM593372	JF704373	JF704660	KM048510	KM048794
AY878071	FJ606180	FJ853560	HM593373	JF704374	JF704661	KM048511	KM048795
AY878072	FJ606181	FJ853561	HM593374	JF704375	JF704662	KM048512	KM048796
AY901965	FJ606182	FJ853562	HM593375	JF704376	JF704663	KM048513	KM048797
AY901966	FJ606183	FJ853563	HM593376	JF704377	JF704664	KM048514	KM048798
AY901967	FJ606184	FJ853564	HM593377	JF704378	JF704665	KM048515	KM048799
AY901968	FJ606185	FJ853566	HM593378	JF704379	JF704666	KM048516	KM048800
AY901969	FJ606186	FJ853567	HM593379	JF704380	JF704667	KM048517	KM048801
AY901970	FJ606187	FJ853569	HM593380	JF704381	JF704668	KM048518	KM048802
AY901971	FJ606189	FJ853573	HM593381	JF704382	JF704669	KM048519	KM048803
AY901972	FJ606190	FJ853574	HM593382	JF704383	JF704670	KM048520	KM048804
AY901973	FJ606191	FJ853575	HM593383	JF704384	JF704671	KM048521	KM048805
AY901974	FJ606192	FJ853578	HM593384	JF704385	JF704672	KM048522	KM048806
AY901975	FJ606193	FJ853579	HM593385	JF704386	JF704673	KM048523	KM048807
AY901976	FJ606194	FJ853580	HM593386	JF704387	JF704674	KM048524	KM048808
AY901977	FJ606195	FJ853581	HM593387	JF704388	JF704675	KM048525	KM048809
AY901978	FJ606196	FJ853586	HM593388	JF704389	JF704676	KM048526	KM048810
AY901979	FJ606197	FJ853590	HM593389	JF704390	JF704677	KM048527	KM048811
AY901980	FJ606198	GQ999989	HM593390	JF704391	JF704678	KM048528	KM048812
AY901981	FJ606199	HM593106	HM593391	JF704392	JF704679	KM048529	KM048813
DQ011165	FJ606200	HM593107	HM593392	JF704393	JF704680	KM048530	KM048814
DQ011166	FJ606201	HM593108	HM593393	JF704394	JF704681	KM048531	KM048815
DQ011167	FJ606202	HM593109	HM593394	JF704395	JF704682	KM048532	KM048816
DQ011169	FJ606203	HM593110	HM593395	JF704396	JF704683	KM048533	KM048821
DQ011170	FJ606204	HM593111	HM593396	JF704397	JF704684	KM048534	KM048822
DQ011171	FJ606205	HM593112	HM593397	JF704398	JF704685	KM048535	KM048823
DQ011172	FJ606206	HM593113	HM593398	JF704399	JF704686	KM048536	KM048824
DQ011173	FJ606207	HM593114	HM593399	JF704400	JF704687	KM048537	KM048825
DQ011174	FJ606208	HM593115	HM593400	JF704401	JF704688	KM048538	KM048826

DQ011175	FJ606209	HM593116	HM593401	JF704402	JF704689	KM048539	KM048827
DQ011176	FJ606210	HM593117	HM593402	JF704403	JF704690	KM048540	KM048828
DQ011177	FJ606211	HM593118	HM593403	JF704404	JF704691	KM048541	KM048829
DQ011178	FJ606212	HM593119	HM593404	JF704405	JF704692	KM048542	KM048830
DQ011179	FJ606213	HM593120	HM593405	JF704406	JF704693	KM048543	KM048831
DQ011180	FJ606214	HM593121	HM593406	JF704407	JF704694	KM048544	KM048832
DQ056404	FJ606215	HM593122	HM593407	JF704408	JF704695	KM048545	KM048833
DQ056405	FJ606216	HM593123	HM593408	JF704409	JF704696	KM048546	KM048834
DQ056406	FJ606217	HM593124	HM593409	JF704410	JF704697	KM048547	KM048835
DQ056408	FJ606218	HM593125	HM593410	JF704411	JF704698	KM048548	KM048836
DQ056409	FJ606219	HM593126	HM593411	JF704412	JF704699	KM048549	KM048837
DQ056410	FJ606220	HM593127	HM593412	JF704413	JF704700	KM048550	KM048838
DQ056411	FJ606221	HM593128	HM593413	JF704414	JF704701	KM048551	KM048839
DQ056412	FJ606222	HM593129	HM593414	JF704415	JF704702	KM048552	KM048840
DQ056413	FJ606223	HM593130	HM593415	JF704416	JF704703	KM048553	KM048841
DQ056414	FJ606224	HM593131	HM593416	JF704417	JF704704	KM048554	KM048842
DQ056415	FJ606225	HM593132	HM593417	JF704418	JF704705	KM048555	KM048843
DQ056416	FJ606226	HM593133	HM593418	JF704419	JF704706	KM048557	KM048844
DQ056417	FJ606227	HM593134	HM593419	JF704420	JF704707	KM048558	KM048845
DQ056418	FJ606228	HM593135	HM593420	JF704421	JF704708	KM048559	KM048846
DQ093585	FJ606229	HM593136	HM593421	JF704422	JF704709	KM048560	KM048847
DQ093586	FJ606230	HM593137	HM593422	JF704423	JF704710	KM048561	KM048848
DQ093587	FJ606231	HM593138	HM593423	JF704424	JF704711	KM048562	KM048849
DQ093588	FJ606232	HM593139	HM593424	JF704425	JQ219842	KM048563	KM048850
DQ093589	FJ606233	HM593140	HM593425	JF704426	JX140662	KM048564	KM048851
DQ093590	FJ606234	HM593141	HM593426	JF704427	KC149035	KM048565	KM048852
DQ093591	FJ606235	HM593142	HM593427	JF704428	KC149036	KM048566	KM048853
DQ093592	FJ606236	HM593143	HM593428	JF704429	KC149037	KM048567	KM048854
DQ093593	FJ606237	HM593144	HM593429	JF704430	KC149038	KM048568	KM048855
DQ093594	FJ606238	HM593145	HM593430	JF704431	KC149039	KM048569	KM048856
DQ093595	FJ606239	HM593146	HM593431	JF704432	KC149040	KM048570	KM048857
DQ093596	FJ606240	HM593147	HM593432	JF704433	KC149041	KM048571	KM048858
DQ093597	FJ606241	HM593148	HM593433	JF704434	KC149042	KM048572	KM048859
DQ093598	FJ606242	HM593149	HM593434	JF704435	KC149043	KM048573	KM048860
DQ093599	FJ606243	HM593150	HM593435	JF704436	KC149044	KM048574	KM048861
DQ093600	FJ606244	HM593152	HM593436	JF704437	KC149045	KM048575	KM048862
DQ093601	FJ606245	HM593153	HM593437	JF704438	KC149046	KM048576	KM048863
DQ093602	FJ606246	HM593154	HM593438	JF704439	KC149047	KM048577	KM048864
DQ093604	FJ606247	HM593155	HM593439	JF704440	KC149048	KM048578	KM048865
DQ093605	FJ606248	HM593156	HM593440	JF704441	KC149049	KM048579	KM048866
DQ093607	FJ606249	HM593157	HM593441	JF704442	KC149050	KM048580	KM048867
DQ164104	FJ606250	HM593158	HM593442	JF704443	KC149051	KM048581	KM048868
DQ164105	FJ606251	HM593159	HM593443	JF704444	KC149052	KM048582	KM048869

DQ164106	FJ606252	HM593160	HM593444	JF704445	KC149054	KM048583	KM048870
DQ164107	FJ606253	HM593161	HM593445	JF704446	KC149055	KM048584	KM048871
DQ164108	FJ606254	HM593162	HM593446	JF704447	KC149056	KM048585	KM048872
DQ164109	FJ606255	HM593163	HM593447	JF704448	KC149057	KM048586	KM048873
DQ164110	FJ606256	HM593164	HM593448	JF704449	KC149058	KM048587	KM048874
DQ164111	FJ606257	HM593165	HM593449	JF704450	KC149059	KM048588	KM048875
DQ164112	FJ606258	HM593166	HM593450	JF704451	KC149060	KM048589	KM048876
DQ164113	FJ606259	HM593167	HM593451	JF704452	KC149061	KM048590	KM048885
DQ164114	FJ606260	HM593168	HM593452	JF704453	KC149062	KM048591	KM048886
DQ164115	FJ606261	HM593169	HM593453	JF704454	KC149063	KM048592	KM048887
DQ164116	FJ606262	HM593170	HM593454	JF704455	KC149064	KM048593	KM048888
DQ164117	FJ606263	HM593171	HM593455	JF704456	KC149065	KM048594	KM048889
DQ164118	FJ606264	HM593172	HM593456	JF704457	KC149066	KM048595	KM048890
DQ164119	FJ606265	HM593173	HM593457	JF704458	KC149067	KM048596	KM048891
DQ164121	FJ606266	HM593174	HM593458	JF704459	KC149068	KM048597	KM048892
DQ164122	FJ606267	HM593175	HM593459	JF704460	KC149069	KM048598	KM048893
DQ164126	FJ606268	HM593176	HM593460	JF704461	KC149070	KM048599	KM048894
DQ164127	FJ606269	HM593177	HM593461	JF704462	KC149071	KM048600	KM048895
DQ164129	FJ606270	HM593178	HM593462	JF704464	KC149072	KM048601	KM048896
DQ275642	FJ606271	HM593179	HM593463	JF704465	KC149073	KM048602	KM048897
DQ275643	FJ606272	HM593180	HM593464	JF704466	KC149074	KM048603	KM048898
DQ275644	FJ606273	HM593181	HM593465	JF704467	KC149075	KM048604	KM048899
DQ275645	FJ606274	HM593182	HM593466	JF704468	KC149076	KM048605	KM048900
DQ275646	FJ606275	HM593183	HM593467	JF704469	KC149077	KM048606	KM048901
DQ275647	FJ606276	HM593184	HM593468	JF704470	KC149078	KM048607	KM048902
DQ275648	FJ606277	HM593185	HM593469	JF704471	KC149079	KM048608	KM048903
DQ275649	FJ606278	HM593186	HM593470	JF704472	KC149080	KM048609	KM048904
DQ275650	FJ606279	HM593187	HM593471	JF704473	KC149081	KM048610	KM048905
DQ275651	FJ606280	HM593188	HM593472	JF704474	KC149082	KM048611	KM048906
DQ275652	FJ606281	HM593189	HM593473	JF704475	KC149083	KM048612	KM048907
DQ275653	FJ606282	HM593190	HM593474	JF704476	KC149084	KM048613	KM048908
DQ275654	FJ606283	HM593191	HM593475	JF704477	KC149085	KM048614	KM048909
DQ275655	FJ606284	HM593192	HM593476	JF704478	KC149086	KM048615	KM048910
DQ275656	FJ606285	HM593193	HM593477	JF704479	KC149087	KM048616	KM048911
DQ275657	FJ606286	HM593194	HM593478	JF704480	KC149088	KM048617	KM048912
DQ275658	FJ606287	HM593195	HM593479	JF704481	KC149089	KM048618	KM048913
DQ275659	FJ606288	HM593196	HM593480	JF704482	KC149090	KM048619	KM048914
DQ275660	FJ606289	HM593197	HM593481	JF704483	KC149091	KM048620	KM048915
DQ275661	FJ606290	HM593199	HM593482	JF704484	KC149092	KM048621	KM048916
DQ275664	FJ606292	HM593200	HM593483	JF704485	KC149093	KM048622	KM048917
DQ351216	FJ606293	HM593201	HM593484	JF704486	KC149094	KM048623	KM048918
DQ351217	FJ606294	HM593202	HM593485	JF704487	KC149095	KM048624	KM048919
DQ351218	FJ606295	HM593203	HM593486	JF704488	KC149096	KM048625	KM048920

DQ351219	FJ606296	HM593204	HM593487	JF704489	KC149097	KM048626	KM048921
DQ351220	FJ606298	HM593205	HM593488	JF704490	KC149098	KM048627	KM048922
DQ351221	FJ606299	HM593206	HM593489	JF704492	KC149099	KM048628	KM048923
DQ351222	FJ606300	HM593207	HM593490	JF704494	KC149100	KM048629	KM048924
DQ351223	FJ606301	HM593208	HM593491	JF704496	KC149101	KM048630	KM048925
DQ351224	FJ606302	HM593209	HM593492	JF704497	KC149102	KM048631	KM048926
DQ351225	FJ606303	HM593210	HM593493	JF704498	KC149103	KM048632	KM048927
DQ351226	FJ606304	HM593211	HM593494	JF704499	KC149104	KM048633	KM048928
DQ351227	FJ606305	HM593212	HM593495	JF704500	KC149105	KM048634	KM048929
DQ351228	FJ606306	HM593213	HM593496	JF704501	KC149106	KM048635	KM048930
DQ351229	FJ606307	HM593214	HM593497	JF704502	KC149107	KM048636	KM048931
DQ351230	FJ606308	HM593215	HM593498	JF704503	KC149108	KM048637	KM048932
DQ351231	FJ606309	HM593216	HM593499	JF704504	KC149109	KM048638	KM048933
DQ351232	FJ606310	HM593217	HM593500	JF704505	KC149110	KM048639	KM048934
DQ351233	FJ606311	HM593218	HM593501	JF704506	KC149111	KM048640	KM048935
DQ351234	FJ606312	HM593219	HM593502	JF704507	KC149112	KM048641	KM048936
DQ351235	FJ606313	HM593220	HM593503	JF704508	KC149113	KM048642	KM048937
DQ351237	FJ606314	HM593221	HM593504	JF704509	KC149114	KM048643	KM048938
DQ369976	FJ606315	HM593222	HM593505	JF704510	KC149115	KM048644	KM048939
DQ369977	FJ606316	HM593223	HM593506	JF704511	KC149116	KM048645	KM048940
DQ369978	FJ606317	HM593224	HM593507	JF704512	KC149117	KM048646	KM048941
DQ369979	FJ606320	HM593225	HM593508	JF704513	KC149118	KM048647	KM048942
DQ369980	FJ606321	HM593226	HM593509	JF704514	KC149119	KM048648	KM048943
DQ369981	FJ606322	HM593227	HM593510	JF704515	KC149120	KM048649	KM048944
DQ369982	FJ606323	HM593228	HQ696823	JF704516	KC149121	KM048650	KM048945
DQ369983	FJ606324	HM593229	HQ696824	JF704517	KC149122	KM048651	KM048946
DQ369984	FJ606325	HM593230	HQ696825	JF704518	KC149123	KM048652	KM048947
DQ369985	FJ606326	HM593231	HQ696826	JF704519	KC149124	KM048653	KM048948
DQ369986	FJ606327	HM593232	HQ696827	JF704520	KC149125	KM048654	KM048949
DQ369987	FJ606328	HM593233	HQ696828	JF704521	KC149126	KM048655	KM048950
DQ369988	FJ606329	HM593234	JF704235	JF704522	KC149127	KM048656	KM048951
DQ369989	FJ606330	HM593235	JF704236	JF704523	KC149128	KM048657	KM048952
DQ369990	FJ606331	HM593236	JF704237	JF704524	KC149129	KM048658	KM048953
DQ369991	FJ606332	HM593237	JF704238	JF704525	KC149130	KM048659	KM048954
DQ369992	FJ606333	HM593238	JF704239	JF704526	KC149131	KM048660	KM048955
DQ369993	FJ606334	HM593239	JF704240	JF704527	KC149132	KM048661	KM048956
DQ369994	FJ606335	HM593240	JF704241	JF704528	KC149133	KM048662	KM048957
DQ369995	FJ606336	HM593241	JF704242	JF704529	KC149134	KM048663	KM048958
DQ369996	FJ606337	HM593242	JF704243	JF704530	KC149135	KM048664	KM048959
DQ369997	FJ606338	HM593243	JF704244	JF704531	KC149136	KM048665	KM048960
DQ396364	FJ606339	HM593244	JF704245	JF704532	KC149137	KM048666	KM048961
DQ396365	FJ606340	HM593245	JF704246	JF704533	KF578465	KM048667	KM048962
DQ396366	FJ606341	HM593246	JF704247	JF704534	KF578466	KM048668	KM048963

DQ396367	FJ606342	HM593247	JF704248	JF704535	KF578467	KM048669	KM048964
DQ396368	FJ606343	HM593248	JF704249	JF704536	KF766537	KM048670	KM048965
DQ396369	FJ606344	HM593249	JF704250	JF704537	KF766538	KM048671	KM048966
DQ396370	FJ606345	HM593250	JF704251	JF704538	KF766539	KM048672	KM048967
DQ396371	FJ606346	HM593251	JF704252	JF704539	KF766540	KM048673	KM048968
DQ396372	FJ606347	HM593252	JF704253	JF704540	KM048382	KM048674	KM048969
DQ396373	FJ606348	HM593253	JF704254	JF704541	KM048383	KM048675	KM048970
DQ396374	FJ606349	HM593254	JF704255	JF704542	KM048384	KM048676	KM048971
DQ396375	FJ606350	HM593255	JF704256	JF704543	KM048392	KM048677	KM048972
DQ396376	FJ606351	HM593256	JF704257	JF704544	KM048393	KM048678	KM048973
DQ396377	FJ606352	HM593257	JF704258	JF704545	KM048394	KM048679	KM048974
DQ396378	FJ606353	HM593258	JF704259	JF704546	KM048395	KM048680	KM048975
DQ396379	FJ606354	HM593259	JF704260	JF704547	KM048396	KM048681	KM048976
DQ396380	FJ606355	HM593260	JF704261	JF704548	KM048397	KM048682	KM048977
DQ396381	FJ606356	HM593261	JF704262	JF704549	KM048398	KM048683	KM048978
DQ396382	FJ606357	HM593262	JF704263	JF704550	KM048399	KM048684	KM048979
DQ396383	FJ606358	HM593263	JF704264	JF704551	KM048400	KM048685	KM048980
DQ396384	FJ606359	HM593264	JF704265	JF704552	KM048401	KM048686	KM048981
DQ396385	FJ606360	HM593265	JF704266	JF704553	KM048402	KM048687	KM048982
DQ396386	FJ606361	HM593266	JF704267	JF704554	KM048403	KM048688	KM048983
DQ396387	FJ606362	HM593267	JF704268	JF704555	KM048404	KM048689	KM048984
DQ396388	FJ606363	HM593268	JF704269	JF704556	KM048405	KM048690	KM048985
DQ396389	FJ606364	HM593269	JF704270	JF704557	KM048406	KM048691	KM048986
DQ396390	FJ606365	HM593270	JF704271	JF704558	KM048407	KM048692	KM048987
DQ396391	FJ606366	HM593271	JF704272	JF704559	KM048408	KM048693	KM048988
DQ396392	FJ606369	HM593272	JF704273	JF704560	KM048409	KM048694	KM048989
DQ396393	FJ606370	HM593273	JF704274	JF704561	KM048410	KM048695	KM048990
DQ396394	FJ606371	HM593274	JF704275	JF704562	KM048411	KM048696	KM048991
DQ396395	FJ606372	HM593275	JF704276	JF704563	KM048412	KM048697	KM048992
DQ396396	FJ606373	HM593276	JF704277	JF704564	KM048413	KM048698	KM048993
DQ396397	FJ606374	HM593277	JF704278	JF704565	KM048414	KM048699	KM048994
DQ396399	FJ606375	HM593278	JF704279	JF704566	KM048415	KM048700	KM048995
DQ445631	FJ606376	HM593279	JF704280	JF704567	KM048416	KM048701	KM048996
DQ445632	FJ606378	HM593280	JF704281	JF704568	KM048417	KM048702	KM048997
DQ445633	FJ606379	HM593281	JF704282	JF704569	KM048418	KM048703	KM048998
DQ445634	FJ606380	HM593282	JF704283	JF704570	KM048419	KM048704	KM048999
DQ445635	FJ606381	HM593283	JF704284	JF704571	KM048420	KM048705	KM049000
DQ445637	FJ606382	HM593284	JF704285	JF704572	KM048421	KM048706	KM049001
EF122539	FJ606383	HM593285	JF704286	JF704573	KM048422	KM048707	KM049002
EF469243	FJ606384	HM593286	JF704287	JF704574	KM048423	KM048708	KM049004
EU242016	FJ606385	HM593287	JF704288	JF704575	KM048424	KM048709	KM049005
EU242294	FJ606386	HM593288	JF704289	JF704576	KM048425	KM048710	KM049006
EU242369	FJ606387	HM593289	JF704290	JF704577	KM048426	KM048711	KM192379

EU242409	FJ606388	HM593290	JF704291	JF704578	KM048427	KM048712	KM192380
EU242415	FJ606389	HM593291	JF704292	JF704579	KM048428	KM048713	KM192381
EU242470	FJ606390	HM593292	JF704293	JF704580	KM048429	KM048714	KM192382
EU242490	FJ606391	HM593293	JF704294	JF704581	KM048430	KM048715	KM192383
EU293445	FJ606392	HM593294	JF704295	JF704582	KM048431	KM048716	KM192384
EU293446	FJ606393	HM593295	JF704296	JF704583	KM048432	KM048717	KM192385
EU293447	FJ606394	HM593296	JF704297	JF704584	KM048433	KM048718	KM192386
EU293448	FJ606395	HM593297	JF704298	JF704585	KM048434	KM048719	KM192387
EU293449	FJ606396	HM593298	JF704299	JF704586	KM048435	KM048720	KM192388
EU293450	FJ606397	HM593299	JF704300	JF704587	KM048436	KM048721	KM192389
EU347706	FJ606398	HM593300	JF704301	JF704588	KM048437	KM048722	KM192390
EU347707	FJ606399	HM593301	JF704302	JF704589	KM048438	KM048723	KM192391
EU347708	FJ606401	HM593302	JF704303	JF704590	KM048439	KM048724	KM192392
EU673411	FJ606402	HM593303	JF704304	JF704591	KM048440	KM048725	KM192393
EU673428	FJ606403	HM593304	JF704305	JF704592	KM048441	KM048726	KM192425
EU673437	FJ606404	HM593305	JF704306	JF704593	KM048442	KM048727	KM192426
EU673443	FJ606406	HM593306	JF704307	JF704594	KM048443	KM048728	KM192427
EU673444	FJ606407	HM593307	JF704308	JF704595	KM048444	KM048729	KM192428
KM192429							

Appendix C: Molecular Dynamic Scripts

Minimization of waters

```
 &cntrl
   imin=1,
   maxcyc=5000, ncyc=3000,
   ntb=1, cut=10.0, ntpr=5,
   ntr=1, restraintmask=':1-
198',
   restraint_wt=2.0
 &end
   /
```

Minimization of the whole system

```
 &cntrl
   imin=1,
   maxcyc=10000, ncyc=7000,
   ntb=1, ntr=0, cut=10.0,
ntpr=5
 &end
   /
```

Heating of proteins from 0K to 300K

```
 &cntrl
   imin=0, irest=0, ntx=1,
ig=-1,
   nstlim=25000, dt=0.002,
   ntc=2, ntf=2,
   cut=8.0, ntb=1,
   ntpr=500, ntwx=500,
   ntt=3, gamma_ln=2.0,
   tempi=0.0, temp0=300.0,
   ntr=1, restraintmask=':1-
198',
   restraint_wt=2.0,
   nmropt=1, ioutfm=1
 /
 &wt TYPE='TEMP0', istep1=0,
istep2=25000,
  value1=0.1, value2=300.0, /
 &wt TYPE='END' /
```

Equilibrate the density

```
 &cntrl
   imin=0,
   irest=1,
   ntx=5,
   nstlim=25000, dt=0.002,
   ntc=2, ntf=2,
   cut=8.0, ntb=2,
   ntp=1, taup=1.0,
   ntpr=500, ntwx=500,
   ntt=3, gamma_ln=2.0,
   temp0=300.0,
   ntr=1, restraintmask=':1-
198',
   restraint_wt=2.0,
   ig=-1,
   ioutfm=1
 &end
   /
```

Equilibrate

```
 &cntrl
   imin=0, irest=1, ntx=5,
   nstlim=1000000, dt=0.002,
   ntc=2, ntf=2,
   cut=10.0, ntb=2,
   ntp=1, taup=2.0,
   ntpr=1000, ntwx=1000,
   ntt=3, gamma_ln=2.0,
   temp0=300.0,
   ig=-1, ioutfm=1
 &end
   /
```

*****The same scripts were used to optimize the protein structure and for the final 100 ns enzyme-substrate MD simulations. Only the "**restraintmask**" flag (bold) was altered for the addition of the inhibitor.

Appendix D: Supplementary information

	30	32	33	46	47	48	50	54	76	82	84	88	90
Consensus	D	V	L	M	I	G	I	I	L	V	I	N	L
ATV/r		I	F	IL	V	VM	L	VTALM		ATFS	V	S	M
DRV/r		I	F		VA		V	LM	V	F	V		
FPV/r		I	F	IL	VA		V	VTALM	V	ATSF	V		M
IDV/r		I		IL	V			VTALM	V	AFTS	V	S	M
LPV/r		I	F	IL	VA	VM	V	VTALM	V	AFTS	V		M
NFV	N		F	IL	V	VM		VTALM		AFTS	V	DS	M
SQV/r						VM		VTALM		AT	V	S	M
TPV/r		I	F	IL	VA			VAM		TL	V		

Figure 1 Screenshot of the Stanford HIV Drug Resistance Database protease inhibitor (PI) resistance notes. The top row indicates the consensus amino acid and position. The following rows indicate each of the eight PIs and the major protease resistance mutations that occurs at a specific position.

Drug Resistance Interpretation: PR

PI Major Resistance Mutations:	**M46I, I54V, V82A**
PI Accessory Resistance Mutations:	None
Other Mutations:	None

Protease Inhibitors

atazanavir/r (ATV/r)	High-Level Resistance
darunavir/r (DRV/r)	Susceptible
lopinavir/r (LPV/r)	High-Level Resistance

Mutation Scoring: PR

PI	ATV/r	DRV/r	LPV/r
M46I	10	0	10
I54V	15	0	15
V82A	15	0	30
M46I + V82A	10	0	10
I54V + V82A	10	0	10
Total	60	0	75

Figure 2 Screenshot of output from the HIVdb algorithm showing the effects of M46I, I54V and V82A on Atazanavir/Ritonavir (ATV/r), Darunavir/Ritonavir (DRV/r) and Lopinavir/Ritonavir treatment (LPV/r).

Drug Resistance Interpretation: PR

PI Major Resistance Mutations:	**M46I, I54V, L76V, V82A**
PI Accessory Resistance Mutations:	None
Other Mutations:	None

Protease Inhibitors

atazanavir/r (ATV/r)	High-Level Resistance
darunavir/r (DRV/r)	Low-Level Resistance
lopinavir/r (LPV/r)	High-Level Resistance

Mutation Scoring: PR

PI	ATV/r	DRV/r	LPV/r
M46I	10	0	10
I54V	15	0	15
V82A	15	0	30
M46I + V82A	10	0	10
I54V + V82A	10	0	10
L76V	0	20	30
M46I + L76V	0	0	10
Total	60	20	115

Figure 3 Screenshot of the output from the HIVdb algorithm showing the effects of M46I, I54V, L76V and V82A on Atazanavir/Ritonavir (ATV/r), Darunavir/Ritonavir (DRV/r) and Lopinavir/Ritonavir (LPV/r) treatment.

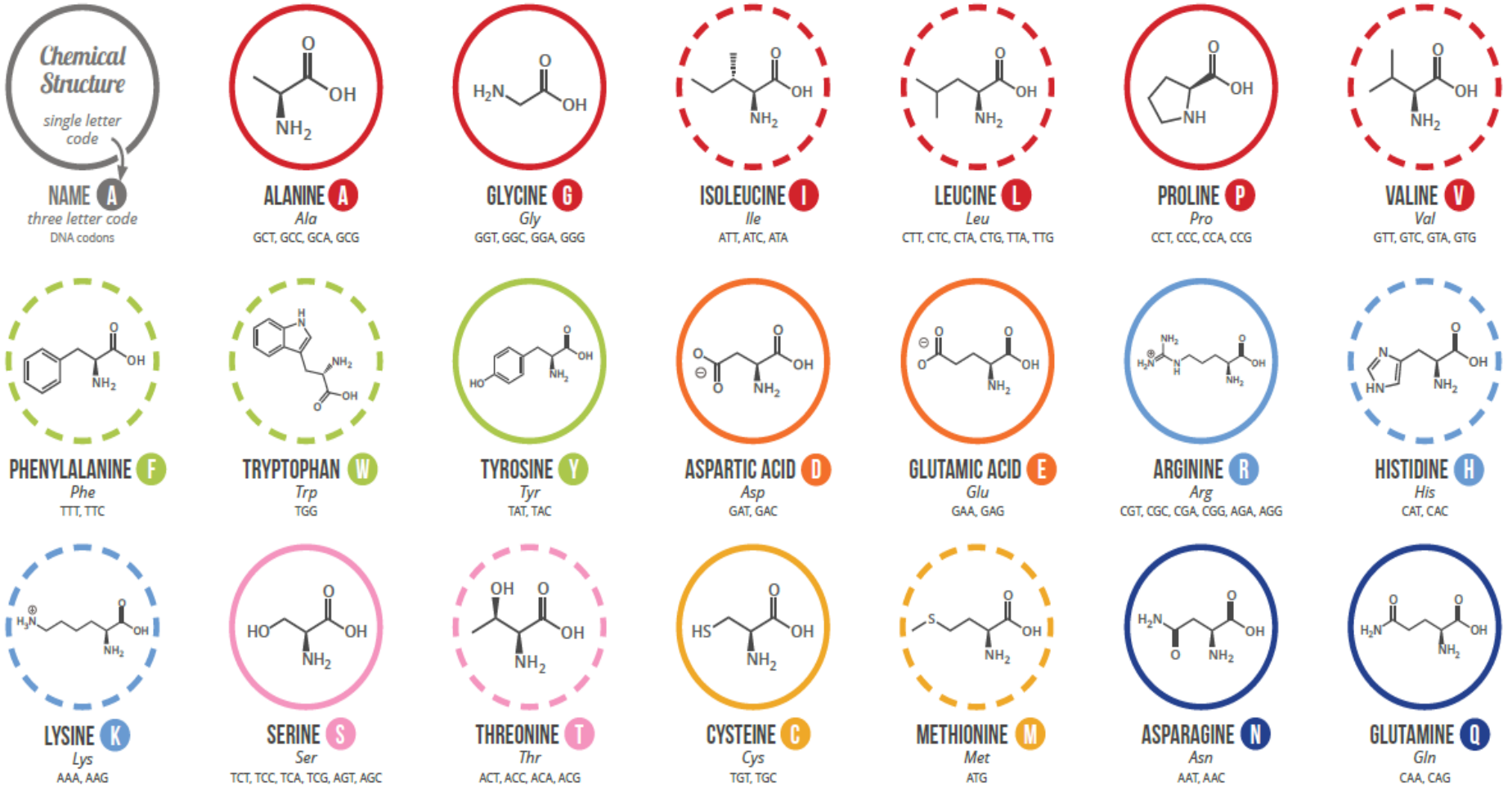


Figure 4 Illustration of the 20 common amino acids showing their structure, single and three letter code.

Appendix E: Ethics Certificate

UNIVERSITY OF
KWAZULU-NATAL
INYUVESI
YAKWAZULU-NATALI
RESEARCH OFFICE
Biomedical Research Ethics Administration
Westville Campus, Govan Mbeki Building
Private Bag X 54001
Durban
4000
KwaZulu-Natal, SOUTH AFRICA
Tel: 27 31 2604769 - Fax: 27 31 2604609
Email: BREC@ukzn.ac.za
Website http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx

20 November 2018

Ms V Marie (209501382)
HIV Pathogenesis Programme
School of Laboratory Medicine and Medical Sciences
veronna.marie@gmail.com

Dear Ms Marie

Protocol: Functional and structural analyses of gag and protease substitution in treated and naïve patients infected with HIV-1 subtype C. **Degree:** PhD
BREC reference number: BE446/15

RECERTIFICATION APPLICATION APPROVAL NOTICE

Approved:	08 January 2019
Expiration of Ethical Approval:	07 January 2020

I wish to advise you that your application for Recertification received 15 November 2018 for the above protocol has been **noted and approved** by a sub-committee of the Biomedical Research Ethics Committee (BREC) for another approval period. The start and end dates of this period are indicated above.

If any modifications or adverse events occur in the project before your next scheduled review, you must submit them to BREC for review. Except in emergency situations, no change to the protocol may be implemented until you have received written BREC approval for the change.

The committee will be notified of the above approval at its next meeting to be held on 11 December 2018.

Yours sincerely

Prof V Rambiritch
Chair: Biomedical Research Ethics Committee

cc: Leslie@ukzn.ac.za
cc: Pedzisai Gaza Gaza@ukzn.ac.za