Coevolution of mutations in HIV-1 Env and Gag-PR genes: Implications for the development of Protease Inhibitors resistance

By

Ntombikhona Fortunate Maphumulo

Submitted in fulfillment 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

2023

Supervisor: Prof. Michelle L. Gordon

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 2017 to July 2023 under the supervision of Prof 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:  Date :_19 July 2023_____

Ntombikhona Maphumulo (Student)

Signed: _____

Prof Michelle Gordon (Supervisor)

Declaration

I, Miss Ntombikhona Fortunate Maphumulo 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

Ntombikhona Maphumulo

Publications

Ntombikhona F. Maphumulo and Michelle L. Gordon (2021). Potential Associations of Mutations within the HIV-1 Env and Gag Genes Conferring Protease Inhibitor (PI) Drug Resistance. *Microbiol. Res. 2021, 12, 967–977. https://doi.org/10.3390/ microbiolres12040071*

Author Contributions: Prof. Michelle L. Gordon and I conceptualized the study. I did the investigation,

Dedication

I dedicate this thesis to myself owoMashimane, oZukuzela, oSondamase, oDlakela, owaseNgoleleni. A difficult journey it was, but you made it through when all odds were against you. The fact that you made it thus far is a testimony that "*no weapon that is formed against you shall prosper*".

"He lifted me up from the pit of despair, out of the miry clay; He set my feet upon a rock, and made my footsteps firm" Psalms 40:2

Acknowledgments

Far and foremost, to God almighty for his grace that has carried me thus far.

I want to express my gratitude to my supervisor Prof. Michelle Gordon for her guidance, encouragement, and teachings, throughout my degree.

I want to extend my appreciation to Dr. S Eche, for patiently teaching me molecular modelling techniques, and to Dr. E San James for his assistance when I needed it.

To my family, bo-Dlakela, wouldn't ask for a better family than you guys, the love, support, and prayers were all well received, and will always appreciate and love you all.

To My mom V.N. Maphumulo, your actions screamed love and unity, and you taught us so much with so few words. On behalf of Abazukulu bakamaMchunu sithi Siyabonga Fuze, Mashiyamahle, may your teachings be instilled in our hearts forever.

To a friend, my soulmate Dr. Sinaye Ngcapu thank you so much for all that you have contributed. Stay blessed and I love you.

To my hearts, Sinesihle and Lusizo, wishing you God's guidance and protection. I Love you my kiddos and will always hold you in my heart.

Babili_oNkiyankiya_aMandlaphi_Ngcosana

Table of Contents

List of figures

Figure 1.1 : HIV-1 full genome shows three major genes that code for different proteins and enzymes, a figure taken directly from (Shukla and Chauhan, 2019). 4
Figure 1.2: Representation of the HIV-1 life cycle steps starting from binding and fusion into the host cell, reverse transcription, integration, replication, assembly, and budding. Then a new immature HIV-1 virus is released and infects more cells (Shukla and Chauhan, 2019). 6
Figure 1.3:Structure of HIV-1 Env gp120, **A**) signaling peptide (yellow), Constant region C1-C5 (peach), Variable region V1-V5 (coral), disulfide bridge and CD4 binding loop (blue). **B**) showing the variable loops (V1 to V5) and bridging sheet, α-helix are shown in purple, β-sheet in cyan and loop in gray. 8
Figure 1.4: Schematic representation and structure of HIV-1 gp41 adopted from (A Pantophlet, 2010). (**A**) AA sequence and their residue numbers in the gp41 sequences. (**B**)3 major domains: an ectodomain (from FP to C-HR), a transmembrane region, and a cytoplasmic tail. (**C**) Ribbon representation of the gp41 structure. 9
Figure 1. 5 : Evolution of coreceptor use, showing early stage to late stage of infection. Taken directly from (Flynn et al., 2019)..... 10
Figure 1.6: Models for Env incorporation into HIV virions. The passive incorporation model shows no interaction between Gag and Env. The Gag–Env co-targeting model shows that Env is enriched in virus assembly domains through lipid raft (brown) association. The direct and indirect interaction of Gag-Env is also shown in this figure, taken directly from (Murphy and Saad, 2020)..... 13

List of Tables

No table of figures entries found.

Abbreviations and Symbols

α	Alpha
Å	Angstrom
β	Beta
µL	Microlitres
ml	Millilitres
3D	Three-dimensional
AIDS	Acquired Immunodeficiency Syndrome
AA	Amino acid
AMBER	Assisted Model Building with Energy Refinement
APV	Amprenavir
ATV	Atazanavir
BN	Bayesian Network
BLOSUM	Blocks substitution matrix.

BNAbs	Broadly neutralizing anti-HIV antibodies.
C1-C5	Constant domains
CA	Capsid
CAPS	Coevolution Analysis for Protein Sequences
CHARMM	Chemistry of Harvard Macromolecular Mechanics
CT	Cytoplasmic tail
CHR	C-terminal heptad repeat
DsDNA	Double-stranded Deoxyribonucleic acid
DNA	Deoxyribonucleic acid
DRV	Darunavir
Env	Envelope
ER	Endoplasmic reticulum.
FDA	Food and Drug Administration
FP	Fusion peptide
FPPR	Fusion peptide proximal region
Gag-PR	Gag-protease
GP	Glycoproteins
GROMACS	Groningen Machine for Chemical Simulations
HAART	Highly active antiretroviral therapy
HR1 and HR2	Heptad Repeat 1 and 2
HIV-1	Human Immunodeficiency virus type1
IDV	Indinavir
LLP	Lentiviral lytic peptides
LPV	Lopinavir
LPV/r	Lopinavir/Ritonavir
MA	Matrix
MD	Molecular dynamics
MVC	Maraviroc
NABs	Neutralizing antibodies.
NC	Nucleocapsid
NFV	Nelfinavir

NHR	N-terminal heptad repeat
Ns	Nanosecond
Pr55Gag	Gag precursor protein
Pr160Gag-Pol	Gag-Pol precursor protein
PLWH	patients living with HIV
PCS	Protease Cleavage Site
PCR	Polymerase chain reaction
PDB	Protein databank
PR	Protease
PIs	Protease inhibitors
Ps	Pica seconds
RMSF	Root mean square fluctuation.
RMSD	Root mean square deviation.
RTV	Ritonavir
SsRNA	Single-stranded Ribonucleic Acid
SU	Surface
T20	Enfuvirtide
TM	Transmembrane
TPV	Tipranavir
V1-V5	Variable domains
VDW	Van-der-Waals
WT	Wild type

Abstract

Cross-resistance in PIs-exposure has been reported to be driven by Gag, however recent studies suggest Env also contributes to PIs resistance. Although studies have reported gp41 mutations in PI failures, the impact of the full-length Env on PI resistance remains unclear. We investigate the prevalence of subtype C Env mutations in patients failing PIs, the coevolution of Env mutations with Gag-PR mutations and determine whether mutations in gp120 as a result of PI resistance, affect co-receptor usage. Lastly, determine the structural changes in the Env during PI failure.

The study used generated sequences from subtype C infected patients failing LPV/r inclusive treatment and HIV-1 subtype C drug-naïve sequences downloaded from the Los Alamos HIV database to compare the frequency of Env mutations in patients failing LPV/r. Bayesian network probability was applied to determine the relationship between mutations occurring within the Env and Gag-PR regions and LPV/r treatment. Furthermore, Los Alamos sequence database tools, Geno2Pheno[coreceptor], and Molecular dynamics simulations were used to demonstrate structural changes and to understand how gp120 mutations affect co-receptor usage. Lastly, Molecular dynamics simulation followed by Ring server was used to determine the structural changes caused by mutations in gp41, and gag mutations, and to look for interaction (hydrogen bond contact and VDW) between mutations and the nearby residues.

Thirty-five mutations in the Env region had significantly higher frequencies in LPV/r treated patients. Env mutations were shown to coevolve with Gag-PR and they form a potential pathway to LPV/r resistance. Gp120 sequences from the PIs treated patients showed to modulate viral entry by protecting the virus from antibody recognition through the increased length in V1/V2 and V5 variable loops and the number of N-glycosylation sites observed in VI/V2. Results further showed that gp120 mutations could modulate viral entry through coreceptor switching induced by a higher charge in the V3 region, mutations in coreceptor-specific sites, and those that interact with the coreceptor binding site. Three mutations in gp41 (D632E-HR2, I688-TM, and P724Q/S-CT) were shown to influence folding by stabilizing the β-turns in their respective regions.

Env coevolution with Gag-PR (mainly MA and CA) was shown through the pathway to LPV/r resistance. The gp120 mutations were also shown to contribute to viral entry through coreceptor usage and immune escape, while gp41 and Gag mutations played a role in stabilizing the α-helices which might influence the fusion of the virus. Further investigations using site-directed mutagenesis are needed to determine the effect of mutations on replication capacity.

Chapter One

Introduction and literature review

1. Introduction

South Africa continues to be the world's largest AIDS epidemic's epicenter for HIV-1 pandemic. Even though HIV-1 is controlled in South Africa, the number of cases among young people who are still affected by the pandemic is still high. In 2022, an estimated 13.9% population is living with HIV-1 infection in South Africa. As a result, it's critical that South Africa continue to treat HIV-1 as a public health emergency (Allinder and Fleischman, 2019, Zuma et al., 2022).

The use of combination Antiretroviral therapy (cART) has successfully reduced human immunodeficiency virus type 1 (HIV-1) viremia and improved immune function in many treated patients living with HIV (PLWH), resulting in great improvements in medical morbidity and life expectancy (UNAIDS, 2020). However, the emergence of HIV-1 drug resistance in patients with poor ART adherence or using a suboptimal drug regimen, and/or lack of viral load monitoring is common in low and middle-income countries. Acquired and transmitted drug resistance continues to be a major public health problem in many parts of the world (particularly in resource-limited settings) (UNAIDs and Organization, 2011, Hedt et al., 2011). Generally, virological failure as the consequence of HIV-1 drug resistance is usually caused by mutations that arise in the viral regions targeted by the antiretroviral drugs. However, recent data shows that virological failure can also occur in the absence of drug-target gene mutations, particularly in PLWH actively taking protease inhibitors (PI) (Gallego et al., 2001, Pulido et al., 2012).

Gag, a substrate of protease (PR), was the first gene identified, outside of PR, to influence the effectiveness of PI-containing drug regimens (Parry et al., 2011, Giandhari et al., 2016). The PIs are a class of anti-HIV-1 drugs that inhibit the activity of the HIV-1 PR enzyme, which is responsible for the production of infectious viral particles. The viral PR enzyme does not cleave the envelope; its role is to cleave the group-specific antigen (Gag) precursor protein (Pr55Gag) and the Gag-Pol precursor protein (Pr160Gag-Pol), and it's also responsible for the production of reverse transcriptase (RT) and integrase enzymes by cleaving the Pr160Gag-Pol (LORI et al., 1988, Rabi et al., 2013). Mutations at the cleavage sites of Gag have been shown to confer PI resistance in the absence of PR mutations (Robinson et al., 2000, Pettit et al., 2002). In addition to compensating for viral fitness, mutations at the Gag nucleocapsid and p1 cleavage site have also been associated with reduced susceptibility of PI in treatment-experienced patients. Non-cleavage mutations, including amino acid insertions close to the Gag cleave site of PI-resistant HIV-1 variants, have been reported to restore the otherwise compromised activity of the mutant PR (van Maarseveen et al., 2012, Fun et al., 2012). Interestingly,

Rabi et al (2013) have implicated envelope (Env) glycoprotein gene as likely contributor of PI resistance.

The HIV-1 Env is an important structural component of the virus comprised of glycoproteins (surface (SU)-gp120 and transmembrane (TM)-gp41) that facilitate virus entry into target cells by binding to host cell surface receptors (CD4) and the coreceptors (CCR5 or CXCR4), and driving membrane fusion (Benureau et al., 2016, Rawat et al., 2005). Gp120 regulates viral tropism by binding to the target-cell receptors, while gp41 promotes fusion between viral and cellular membranes, releasing the viral nucleocapsid into the cytoplasm (Tedbury and Freed, 2014). According to Rabi et al.,2013, the *env* gene is likely to be involved in the development of PI resistance, since *env* genes isolated from PI treatment failures were able to significantly reduce the activity of PIs (up to 10-fold) in the presence of wild-type PR and Gag (Rabi et al., 2013). Furthermore, This was later supported by Coetzer et al (2017), who identified mutations in gp41 associated with PI failure in the absence of known resistance mutations in PR (Coetzer et al., 2017). Despite the growing amount of data available demonstrate that mutations in *gag* and *env* genes can confer resistance to PIs, less is known about the Env-mediated drug resistance in PI-exposure isolates and whether these co-evolve with Gag and PR mutations.

1.1 HIV-1 virus

Acquired Immunodeficiency Syndrome (AIDS) was first identified in 1981 and was first associated with men who have sex with men, blood transfusion recipients, and drug users (Barré-Sinoussi et al., 2013, Gottlieb et al., 1981). In 1983, the Human Immunodeficiency virus type 1 (HIV-1) was shown to be the causative agent of AIDS (Gallo and Montagnier, 2003, Barré-Sinoussi et al., 2013). HIV-1 is a genus member of lentiviruses that belong to a group of retroviruses (Engelman and Cherepanov, 2012). Currently there are 4 groups of HIV-1: the “main or major” group M which is the leading cause of HIV-1 infections globally, while “outlier” group O and recently discovered groups N and P remain in West Africa (Sharp and Hahn, 2011). There is enormous genetic diversity within group M itself, with at least nine distinct subtypes (A, B, C, D,F, G, H, J, and K) and circulating recombinant forms (Hemelaar, 2012). Subtypes B and C are the most predominant and responsible for most infections globally (Giovanetti et al., 2020). According to Giovanetti et al. (2020), subtype C predominates in South Africa, the Horn of Africa, and India (Lole et al., 1999, Giovanetti et al., 2020).

The structure of HIV-1 is distinct from that of the other family members; HIV-1 is composed of two single-stranded RNA molecules encased in a conical capsid that houses the p24 viral protein and has

a diameter of about 120 nm and a roughly spherical shape (Blood, 2016, McGovern et al., 2002). The full HIV-1 genome is encoded on one long single-stranded Ribonucleic Acid (ssRNA) that comprises of nine genes (*gag, pol, env, vif, vpr, rev, vpu, tat*, and *nef*) (Fig 1.1) (Shukla and Chauhan, 2019). Of these nine genes within the retroviral genome, there are three major genes namely the *gag*, followed by the *pol* and *env* genes (Freed, 2001). The *gag* gene encodes four structural proteins of the matrix (p17), capsid (p24), nucleocapsid (p7), and p6 (Freed, 1998, Waheed and Freed, 2012). The Pol region encodes all viral enzymes, presented by PR (p11), reverse transcriptase (p66/p51), and Integrase (p32). The Env glycoprotein encodes the gp160 glycoprotein, which is subsequently cleaved into surface gp120 and transmembrane gp41 (Frankel and Young, 1998).

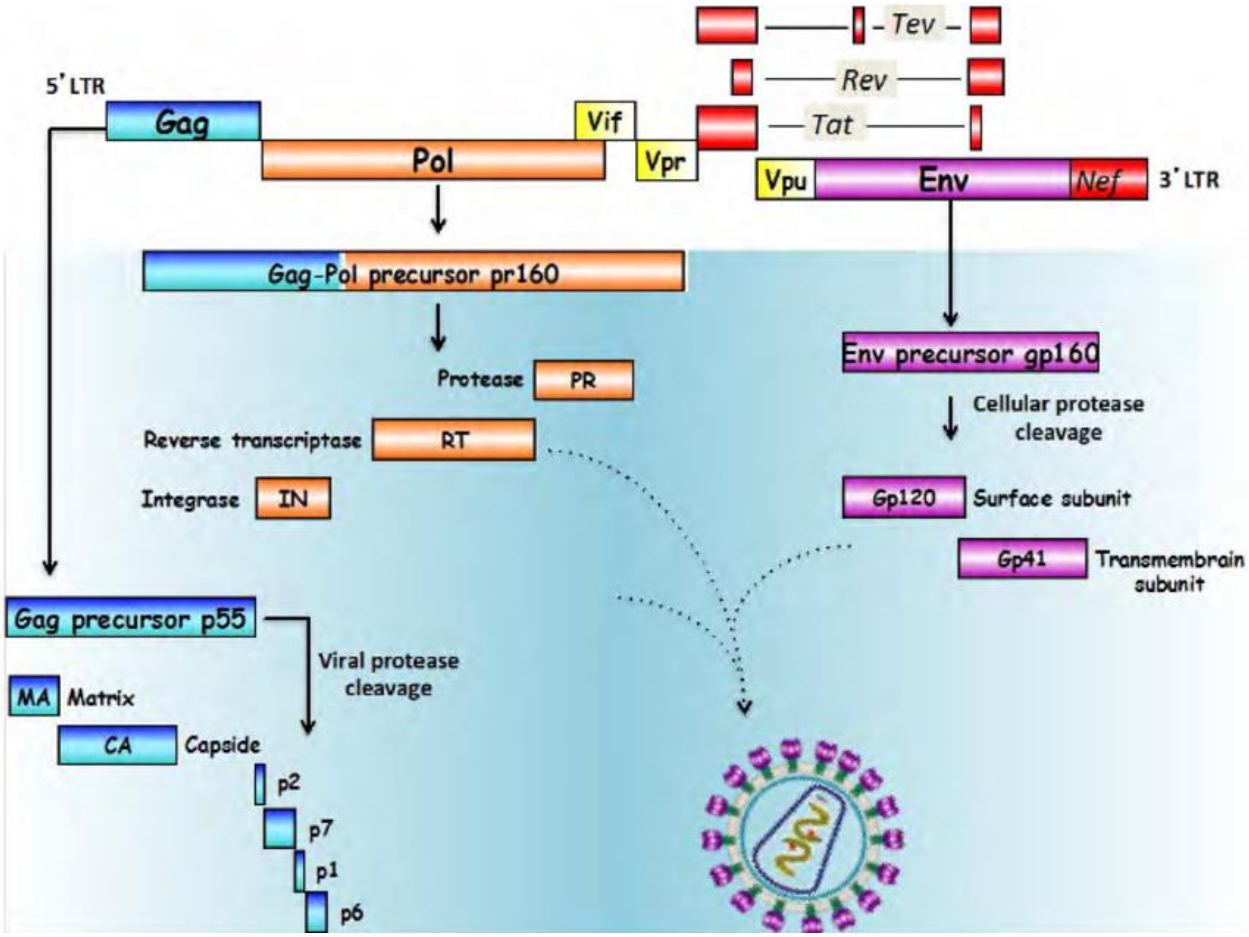


Figure 1.1 : HIV-1 full genome shows three major genes that code for different proteins and enzymes, a figure taken directly from (Shukla and Chauhan, 2019).

1.2 HIV-1 entry replication cycle

Entry to the cell is through the interaction of gp160 spikes on the HIV-1 viral Env with CD4 and chemokine co-receptors CCR5 or CXCR4 on the target cell surface (Eckert and Kim, 2001). During entry, the viral gp120 binds to the cellular receptor CD4 protein of the target cell. The gp120 undergoes a structural change, exposing the chemokine coreceptor binding domain of the gp120, allowing it to bind with the target chemokine coreceptors (Shrivastava and LaLonde, 2010, Korkut and Hendrickson, 2012).

This binding induces an interaction with the gp41 transmembrane proteins, Heptad Repeat 1 and 2 (HR1 and HR2) with each forming a central triple-stranded α-helical coiled-coil core, resulting in the formation of a stable six-helix bundle structure, and the fusion of the viral and host cell membranes with the release of the nucleocapsid containing the viral RNA into the cytoplasm of the host cell. Reverse transcriptase (RT) which act as RNA-dependent DNA polymerase, DNA-dependent DNA polymerase is then synthesizes viral RNA into double-stranded Deoxyribonucleic Acid (dsDNA) or pro-viral Deoxyribonucleic Acid (DNA). The RT also cleaves the RNA template strand of RNA/DNA hybrids using its RNase H activity (Hu and Hughes, 2012).

Integrase then facilitates the integration of the newly created viral DNA into the host chromosome, where it remains for the duration of the cell's existence (fig 1.2) (Freed, 2001, Saxena and Chitti, 2016, Melikyan et al., 2000, Chan et al., 1998). Integrase catalyses activity of the 3’-processing and a strand transfer reaction leading to the insertion of the processed viral DNA into the target DNA. The integrated viral DNA is then transcribed into mRNA by RNA polymerase. The transcribed mRNA is then spliced and transported from the nucleus into the cytoplasm where it is translated into viral proteins and enzymes. The virus particles are assembled from the newly synthesized proteins and become immature virions (Engelman and Cherepanov, 2012). The polyproteins within the virion are then cleaved by the viral PR to produce mature infectious virus particles (Marsden and Zack, 2013).

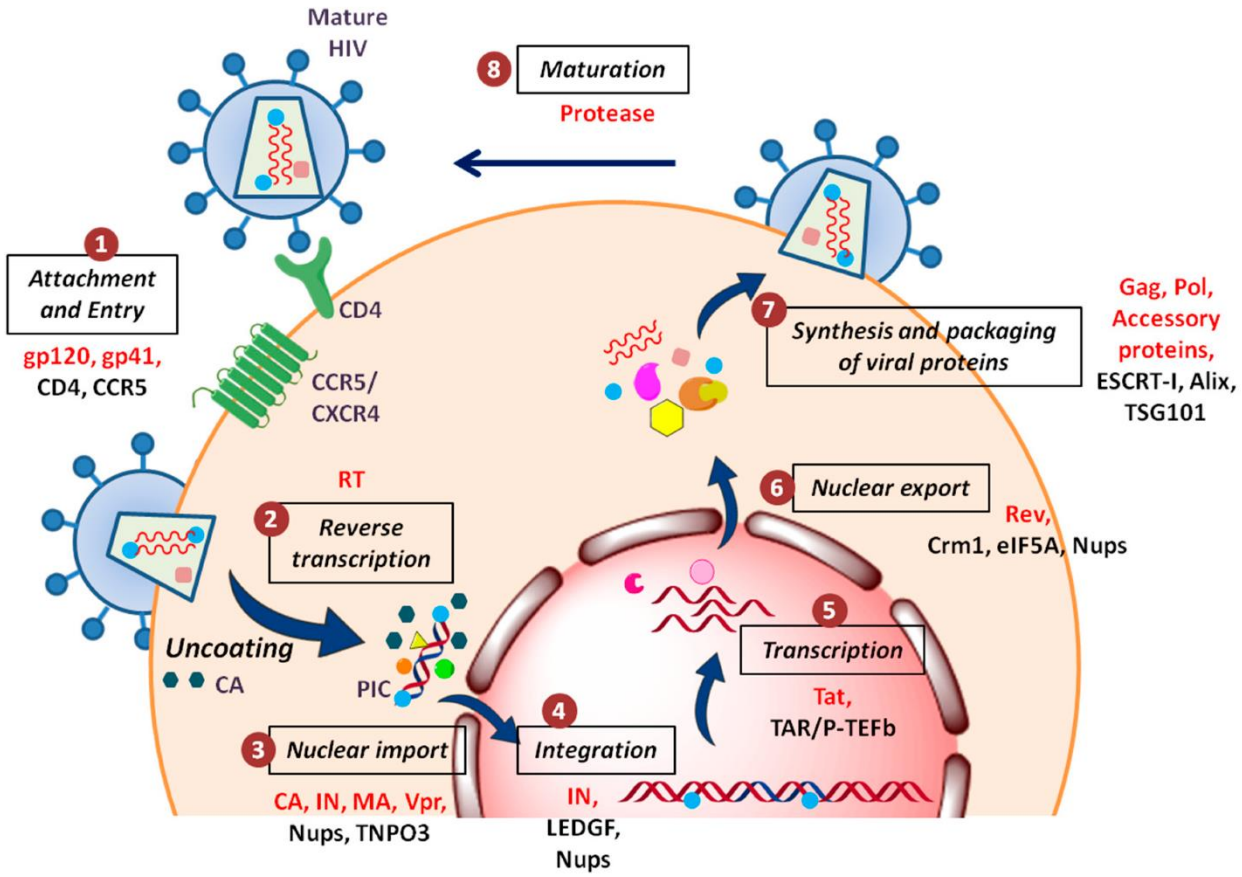


Figure 1.2: Representation of the HIV-1 life cycle steps starting from binding and fusion into the host cell, reverse transcription, integration, replication, assembly, and budding. Then a new immature HIV-1 virus is released and infects more cells (Shukla and Chauhan, 2019).

1.3 HIV-1 full-length envelope

HIV-1 Env gene encodes a single protein, gp160, that is synthesized into glycoproteins and forms a significant part of the outer layer of the virus. Gp160 helps the virus adhere and interact with the host cell's surface membrane by presenting itself as viral membrane spikes consisting of 3 linked molecules of gp120 and attached to the membrane by gp41 (Gabuzda et al., 1991). Studies have also reported that the Env transcript enters the endoplasmic reticulum (ER), where it rests for an indeterminate period and then is translated into the polyprotein precursor gp160 that is active through the Golgi complex (Stein and Engleman, 1990, Abbadessa et al., 2015). Proteolytic processing of HIV-1 gp160 to gp120 and gp41 occurs in the cis and medial cisternae of the RER-Golgi complex during intracellular transport and is essential for the attainment of membrane-fusion and viral infectivity (Bosch and Pawlita, 1990, Stein and Engleman, 1990, Willey et al., 1988, McCune et al., 1988).

1.3.1 Structure and function of gp120

The surface Env gp120 is heavily glycosylated at approximately 25 glycosylation sites with carbohydrates accounting for approximately 50% of the molecular weight. Based on the CD4-bound structure of Env, gp120 is organized into three parts: inner domain, outer domain, and bridging sheet (Pantophlet and Burton, 2006). The bridging sheet is composed of a four-stranded antiparallel β-sheet (β2, β3, β20, β21) (fig 1.3b). The bridging sheet linked the inner and outer domains, with β2 and β3 forming part of the inner domain, while β20 and β21 form part of the outer domain (Shrivastava et al., 2012).

The mature gp120 consists of five variable domains (V1-V5) and five relatively constant domains (C1-C5) (fig 1.3a) (Willey et al., 1986). The constant domains of gp120 form an integral portion of the core and principal determinants (particularly C1, C3, and C4) that bind the CD4 receptor. Variable regions, except for V5, are held together by nine disulfide bridges that are covalently bound by 18 Cys residues. The V1- V2 loop is the most variable, ranging from 50-90 amino acid (AA) long, whereas the V4 and V5 loop ranges from 19-44 AA and 14-36 AA, respectively. Additionally, a shorter V1/V2 domain of Env increases HIV-1 fusion and Env incorporation, whereas a longer V1/V2 reduces the fusion and incorporation of Env (Cavrois et al., 2014). The length of V1-V2 is associated with coreceptor switching or CXCR4 use and is also known to impair CD4 binding, which leads to an increase in the poor entry of HIV-1 Env via CCR5 and CXCR4. Delobel et al, (2005), reported that the CCR5 to CXCR4 switch could occur despite effective HAART, and the patient with the predominant CXCR4 variants during HAART, either from baseline or after the switch, tend to have lower CD4+ T-cell counts on HAART compared to the patients with predominant CCR5 variants (Delobel et al., 2005).

Furthermore, the V3 loop is short (34-35 AA) and less variable but is a major determinant of HIV-1 tropism (Curlin et al., 2010, Zolla-Pazner and Cardozo, 2010, Thielen et al., 2010). The V3 loop has previously been reported to play an important role in membrane fusion, coreceptor specificity, and dominant epitopes recognized by neutralizing antibodies (Freed et al., 1991). Sequence changes in the V3 loop play a crucial role in coreceptor switching from CCR5 to CXCR4 (Pollakis et al., 2004, Shioda et al., 1994). Coreceptor switching is driven by the loss of entry fitness via CCR5 and increased CD4 binding (Coetzer et al., 2008). Although it has been postulated that deletion of the V4 loop may impair proper gp160 protein folding, the role of V4 and V5 remains unclear (Moore et al., 2008, Rong et al., 2007).

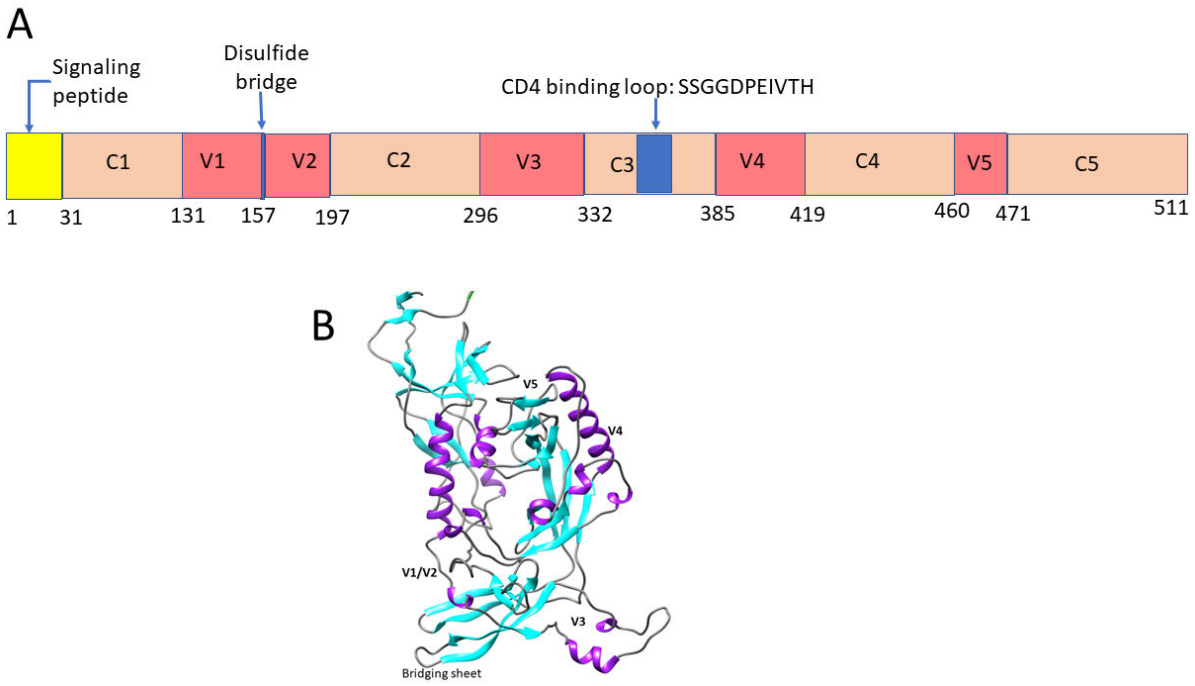


Figure 1.3:Structure of HIV-1 Env gp120, **A**) signaling peptide (yellow), Constant region C1-C5 (peach), Variable region V1-V5 (coral), disulfide bridge and CD4 binding loop (blue). **B**) showing the variable loops (V1 to V5) and bridging sheet, α-helix are shown in purple, β-sheet in cyan and loop in gray.

1.3.2 Structure and function of gp41

The gp41 transmembrane mediates the viral fusion of Env with the host cell membrane. The gp41 consists of three major domains: extracellular (ectodomain), transmembrane, and C-terminal cytoplasmic tail (CT) (Checkley et al., 2011). The ectodomain comprises three functional regions: fusion peptide (FP), fusion peptide proximal region (FPPR), N-terminal heptad repeat (NHR), and C-terminal heptad repeat (CHR) (fig 1.4). In addition to these regions, is the membrane-proximal external region (MPER), a conformationally-flexible region that connects the CHR to the TM region. The MPER is a highly conserved motif in the HIV-1 gp41 that contains an epitope that binds to broadly neutralizing antibodies and therefore become a target region for drugs and vaccine design (A Yi et al., 2016, Liu et al., 2018). The NHR and CHR regions are located in an antiparallel orientation and are linked by a loop region that plays a critical role in fusion (Bär and Alizon, 2004). The cytoplasmic tail of gp41 regulates both Env incorporation and fusion, and it contains three lentiviral lytic peptides (LLP) regions that interact with the viral membrane and regulate fusion (Viard et al., 2008, Wyss et al., 2005). The highly conserved gp41 TM domain is anchored in the Env lipid bilayer

and induces the fusion of the viral and cellular membrane (Wilen et al., 2012, Shang et al., 2008, Hollingsworth IV et al., 2018).

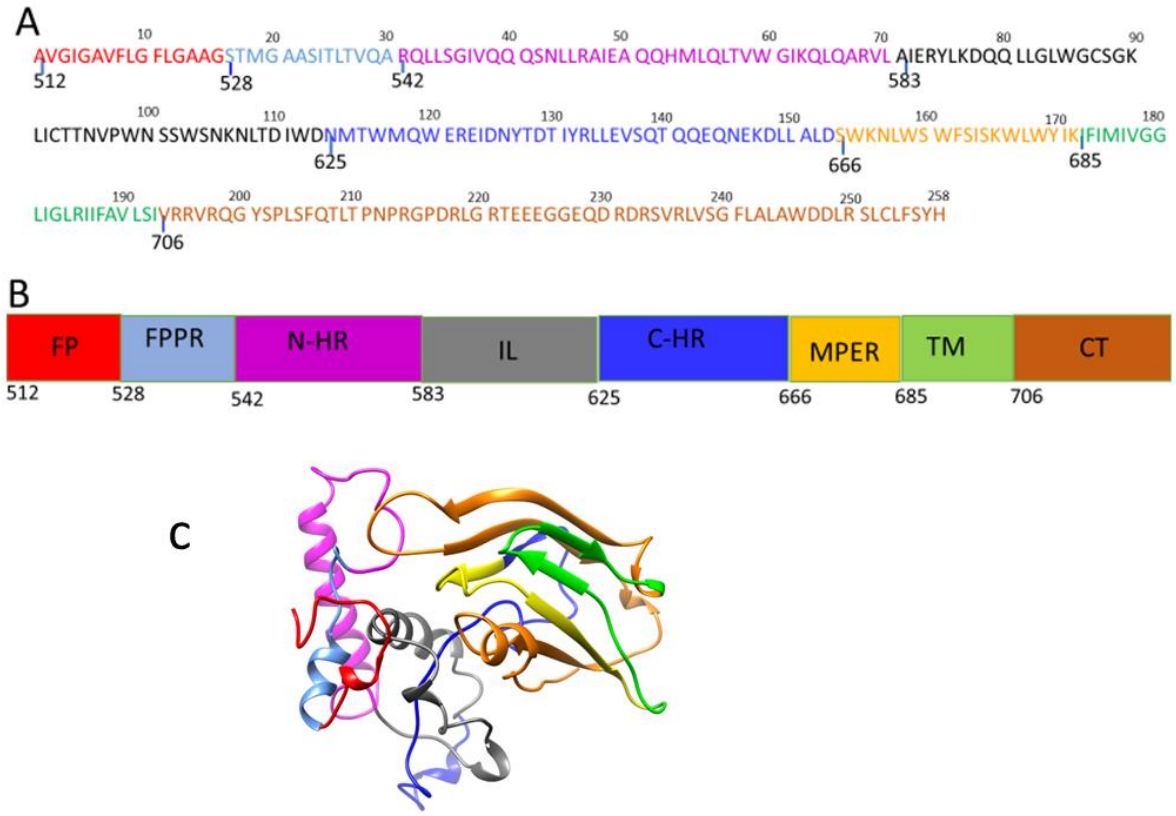


Figure 1.4: Schematic representation and structure of HIV-1 gp41 adopted from (A Pantophlet, 2010). (**A**) AA sequence and their residue numbers in the gp41 sequences. (**B**)3 major domains: an ectodomain (from FP to C-HR), a transmembrane region, and a cytoplasmic tail. (**C**) Ribbon representation of the gp41 structure.

1.3.3 Co-receptor switching

In the early stages of subtype C infection, the CCR5 virus isolates dominate HIV-1, whereas CXCR4 is selected later in chronic infection (fig 1.5) and is associated with a faster decline in CD4 cell count (Moore et al., 2004, Mugwagwa and Witten, 2007, Zhang et al., 2010). Coreceptor switching in subtype C is limited and is driven by the loss of entry fitness via CCR5 and increased CD4 binding (Coetzer et al., 2008, Kotokwe et al., 2023). Furthermore, the length of V1-V2 is associated with coreceptor switching or CXCR4 use, and is also known to impair CD4 binding, which leads to the poor entry of HIV-1 Env via CCR5 or CXCR4 (Delobel et al., 2005). HIV-1 subtype C is known to have reduced R5X4 compared to other subtypes, and this is because coreceptor switching in subtype C requires more mutations than other subtypes (Coetzer et al., 2011). In-line with other subtype C,

Matume et al (2020) reported the 3-5-fold increase of X4 virus in late chronic patients (Matume et al., 2020).

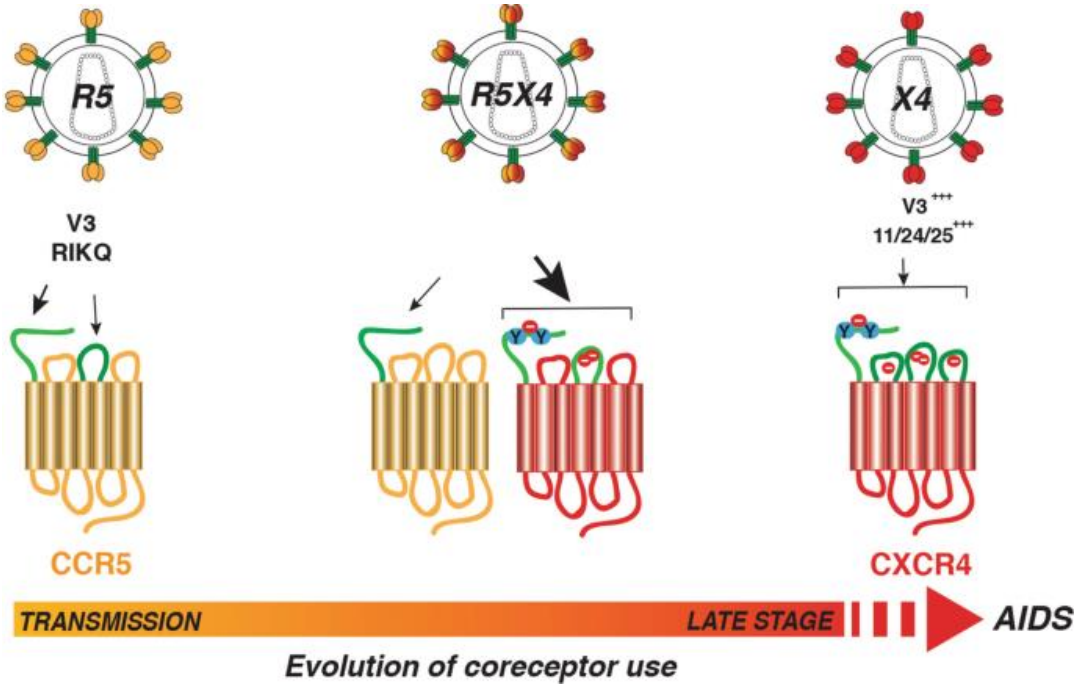


Figure 1. 5 : Evolution of coreceptor use, showing early stage to late stage of infection. Taken directly from (Flynn et al., 2019).

1.3.4 HIV-1 Env glycans

HIV-1 gp120 is a heavily glycosylated protein with nearly 50% of its molecular weight comprising of glycan moieties (Poignard et al., 2001). Gp41 is also glycosylated, but to a much lesser extent than gp120. N-linked glycosylation sites (PNGs), which are added to asparagine residues, target an amino acid sequence pattern that is defined by an Asn-X-Ser/Thr motif, where X can represent any amino acid other than proline (Dutta et al., 2017). The N-linked glycan structure is altered in the endoplasmic reticulum and then the Golgi apparatus by first deleting some sugar residues (trimming), then re-glycosylation by adding sugar residues (processing), such as galactose, fucose, or sialic acid, to yield its mature form. There are average of 93 PNGs on Env, which vary significantly between gp120 HIV-1 strains (ranging between 18-33) (Wang et al., 2013). The PNGs perform a variety of functions, such as assisting in the folding, oligomerization and processing of gp120 that take place during CD4 and co-receptor binding (Wang et al., 2013). Additionally, by preventing neutralizing antibodies from accessing the antigenic polypeptide surface of the HIV-1 Env, PNGs can sterically protect the virus against antibody neutralization (Doores, 2015). Furthermore, viruses that

have gp120 glycans may aid in viral dispersion when they are taken in by DC-SIGN lectins, which are specific to dendritic cells and intercellular adhesion molecule-3-grabbing non-integrins (Alen et al., 2012).

1.4 Env-Gag Interaction

1.4.1 Gag gene

HIV-1 *env* and *gag* genes are two major structural proteins of the virus that are important for the assembly, budding, and infectivity of the virus (Murakami, 2008). The mature *gag* gene plays an essential role in several steps of the HIV-1 replication cycle, notably during the assembly, budding, and release of infectious viral particles (Klingler et al., 2020). It also plays a role in coordinating the interplay between viral components to generate particles that contain all elements required for viral maturity and infectivity (Murphy and Saad, 2020). The *gag* gene consists of a polyprotein precursor $Pr55^{Gag}$ that is cleaved by the viral PR to the mature gag proteins: matrix (p17 MA), capsid (p24 CA), nucleocapsid (p7 NC), p6, and two spacer peptides, SP1 (originally called p2) and SP2 (originally called p1) (Freed, 2001). Gag proteins are synthesized in the cytoplasm and subsequently bind to the membrane via the MA domain. In the MA domain, Gag proteins interact with PIP_2 (plasma-membrane-specific acidic phospholipid, phosphatidylinositol-(4,5)-bisphosphate) and Env proteins, which are crucial for the efficient membrane binding, plasma membrane targeting, Gag-Gag multimerization, incorporation of the viral genome and Env proteins into virions, and budding (Murakami, 2008).

1.4.2 Role of gag in HIV-1 envelope formation

The Env glycoproteins are synthesized in ribosomes attached to the endoplasmic reticulum membrane, while Gag and Gag-Pol precursors are synthesized in free ribosomes in the cytoplasm (Muriaux and Darlix, 2010). The Env glycoprotein travels through the secretory pathway to the plasma membrane. During trafficking, Env translates into the precursor gp160 protein, which is then cleaved into gp120 and gp41 subunits by a cellular PR called furin (Muriaux and Darlix, 2010, Pasquato et al., 2007). The gp120 and gp41 complexes are incorporated into lipid rafts where virion assembly occurs (Muriaux and Darlix, 2010).

The Env glycoprotein is incorporated into the virions during budding from the plasma membrane (Freed and Martin, 1995). This incorporation is mediated by the matrix (MA) domain of Gag by interacting with the gp41 cytoplasmic tail (CT) during HIV-1 particle assembly (Yu et al., 1992). The Env trimer requires MA trimers for incorporation since disruption of the MA trimer causes the loss of Env incorporation and viral infectivity (Tedbury et al., 2016, Tedbury and Freed, 2014).

Studies have proposed several mechanisms by which *env* genes are incorporated into virions. These include passive incorporation, direct Gag-Env interaction, Gag-Env co-targeting, and indirect Gag-Env interaction. The passive incorporation model allows for the incorporation of host cellular and membrane proteins at the cell surface into virus particles without significant sorting and this model does not include Gag (Murphy and Saad, 2020). The Gag–Env co-targeting model involves the recruitment of both genes to assembly sites on the plasma membrane (PM) based on their independent association with lipid raft-like microdomains fig.1.6 (Checkley et al., 2011, Murphy and Saad, 2020). The direct Gag–Env interaction model involves the interaction between the HIV-1 Env gp41 cytoplasmic tail (CT) and the MA domain of Gag during HIV-1 particle assembly (Yu et al., 1992), while the indirect Gag–Env interaction model involves Env incorporation that is mediated by a cellular co-factor that binds both Env and Gag (Checkley et al., 2011).

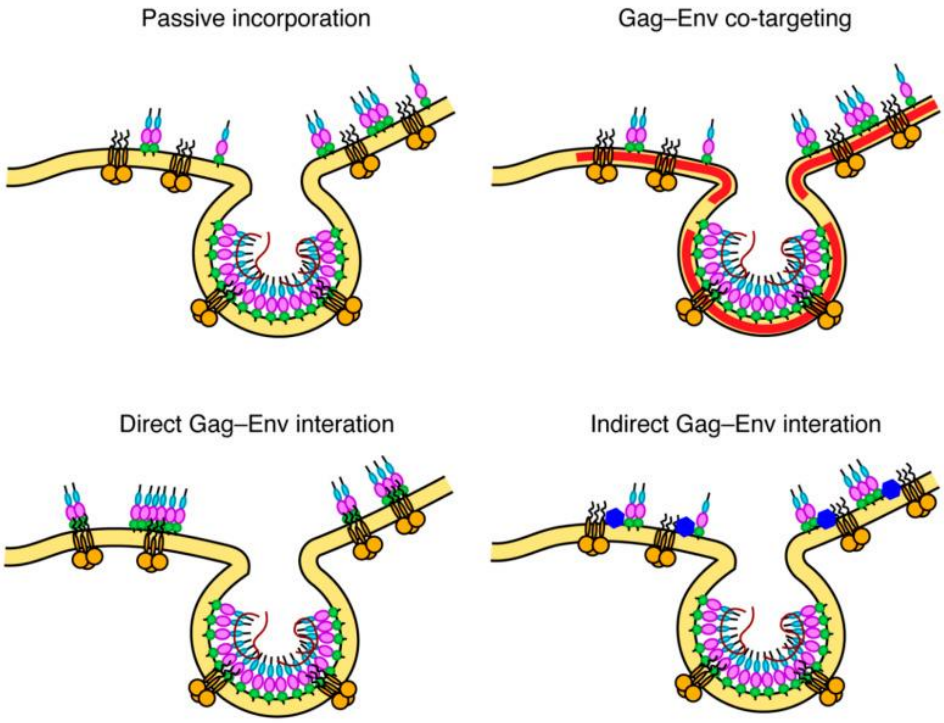


Figure 1.6: Models for Env incorporation into HIV virions. The passive incorporation model shows no interaction between Gag and Env. The Gag–Env co-targeting model shows that Env is enriched in virus assembly domains through lipid raft (brown) association. The direct and indirect interaction of Gag-Env is also shown in this figure, taken directly from (Murphy and Saad, 2020).

1.5 Role of Env glycoprotein mutations on viral fitness and efficacy of anti-viral drugs.

1.5.1 Envelope glycoprotein mutations affecting viral infectivity.

Studies have suggested that HIV-1 exhibits a high level of genetic diversity that could cause the *env* gene to rapidly evolve during infection (Santoro and Perno, 2013, Andrews and Rowland-Jones, 2017, Dang et al., 2020). The rapid evolution of the *env* gene could enhance the capacity of HIV-1 to escape immune surveillance under selection pressure caused by anti-viral drugs (Wang et al., 2022). Mutations in most gp120 regions have been associated with replication capacity. Mutations in hydrophobic regions found to be important for receptor binding have been associated with the reduced binding affinity of the Env gp120 protein on CD4 receptors (Prabakaran et al., 2007, Olshevsky et al., 1990).

Studies found that mutations in the conserved C2 domain were associated with increased viral replication capacity while mutations in the conserved C3 domain were associated with low viral load

(Dang et al., 2020). Similarly, mutations in the V1/V2 loop were shown to lead to an evasion of the immune response which helped boost viral replication (Saunders et al., 2005), while mutations in the variable V3 loop were homogeneous and deletions in the variable V4 loop was associated with increased viral replication (Dang et al., 2020). These findings suggest that genetic variations in different regions of the HIV-1 Env might be involved in increased viral infectivity and replication capacity.

1.5.2 Impact of Env glycoprotein mutations on HIV-1 entry inhibitors

Over the years, researchers have developed entry inhibitors that target and inhibit different stages of the entry process. These include blocking the (i) binding of the viral glycoprotein gp120 to the CD4 cell receptor (PRO-542 [CD4-immunoglobulin G2 (IgG2)], TNX-355, BMS-806, and CADA); (ii) attachment of the gp120 to CCR5 or CXCR4 [SCH-C, vicriviroc (SHC-D), aplaviroc (GW873140), maraviroc], and (iii) the gp41-mediated fusion of the viral and cellular membranes [Enfuvirtide (T-20)] (Lobritz et al., 2010, Westby and van der Ryst, 2005). Many entry inhibitors are currently in clinical development or clinical use, with Enfuvirtide being the only fusion inhibitor approved by the US Food and Drug Administration for clinical use since 2003 and is currently not available in many limited-resourced countries.

Although clinical data on resistance to entry inhibitors is scarce, *in vitro* studies suggest changes in amino acids (Trp 112, Thr 257, Ser 375, Phe 382, and Met 426) of gp120 are implicated in resistance to BMS-806 and BMS-155. In addition, natural genetic variations in HIV-1 Env gp120 of subtype C and recombinant subtype AE (CRF01_AE) viruses may account for resistance to these drugs (Briz et al., 2006). Furthermore, the shift in coreceptor usage and rapid evolution of the *env* gene that promotes interaction between gp120 and the coreceptor to occur despite the presence of the inhibitor is the main resistance pathway associated with CCR5 or CXCR4 inhibitor failure (Coetzer et al., 2008). Specific mutations (V38A/E) in HIV-1 Env gp41 have been shown to confer resistance to Enfuvirtide (Ray et al., 2007). While there is evidence of resistance in vitro, there is a need for clinical studies to assess the susceptibility of different HIV-1 variants to entry inhibitors, including their role in cross-resistance to other non-fusion inhibitors.

1.6 The contribution of Env glycoprotein mutations to HIV-1 Protease inhibitor resistance

1.6.1 HIV-1 Protease

HIV-1 PR is a homodimeric enzyme consisting of two polypeptide chains that are 99 AA long. The enzyme is responsible for the cleavage of gag and gag-pol polyproteins for mature virions to be formed (Fun et al., 2012). The active site (binding pockets) is located between the two identical subunits and is covered by two flexible β-hairpin flaps which need to open to allow the substrates to access the active site. The HIV-1 protease enzyme activity can be inhibited by blocking the active site of the protease (Miller et al., 1989, Wensing et al., 2010).

1.6.1.2 Protease Inhibitors and drug resistance

HIV-1 PR is a homodimeric enzyme consisting of two polypeptide chains that are necessary for cleavages of amino acids in the gag and gag-pol polyproteins for mature virions to be formed (Fun et al., 2012). Cleavage of the gag polyproteins is necessary and sufficient for the assembly of virus-like particles including the release of essential enzymes such as PR, reverse transcriptase, RT-RNase H, and integrase. HIV-1 PIs inhibit cleavage of the gag and gag-pol polyproteins thereby preventing viral maturation and blocking the infectivity of budding virions (Flexner, 1998). There are nine FDA-approved PIs, namely: Saquinavir (SQV), Indinavir (IDV), Ritonavir (RTV), Nelfinavir (NFV), Amprenavir (APV), Lopinavir (LPV), Atazanavir (ATV), Tipranavir (TPV) and Darunavir (DRV). With the exception of NFV, all PI drugs are manufactured in combination with ritonavir (RTV) as a booster.

Resistance to PIs occurs in a sequential process, resulting in major and minor mutations (Nijhuis et al., 1999). Primary resistance mutations to the PIs arise in and around the substrate cleft of PR which reduces the binding affinity of the inhibitor to the mutant PR (Markowitz et al., 1998, Baxter et al., 2016). In addition, primary mutations can slightly reduce the binding affinity of the binding site for the Gag and Gag-pol substrates, resulting in a deficit in the replicative ability of the virus (Gulnik et al., 1995, Croteau et al., 1997). Secondary resistance mutations tend to occur in combination with other primary mutatios (Baxter et al., 2016, Berkhout, 1999). These secondary resistance mutations can compensate for the decreased activity of the mutant PR by increasing replication capacity and viral fitness by restoring viral PR activity (Baxter et al., 2016, Berkhout, 1999).

The most common PI drug resistance mutations reported in South African include M46I, I54V, L76V, and V82A (Feucht et al., 2014, Makatini et al., 2019, Obasa et al., 2021, Chimukangara et al., 2022), although many patients did not harbour any PI resistance mutations (Wallis et al., 2011, Obasa et al., 2020), suggesting that non-adherence could be contributing to PI failure. However, reduced susceptibility to PIs in the absence of any major PI resistance-associated mutations in the PR region, suggesting a role for mutations outside the PR gene (Gallego et al., 2001, Pulido et al., 2012, Sutherland et al., 2015).

1.6.2 The role of Gag and Env mutations to Protease Inhibitor resistance

1.6.2.1 Gag mutations to Protease Inhibitor resistance.

Studies have shown that PIs are often hampered by drug resistance mutations in the viral PR as well as in the viral substrate, Gag (Su et al., 2019). Alone and together with mutations in the PR gene, gag mutations at both cleavage sites and non-cleavage sites were found to partially restore the replication capacity and confer PI resistance (Gupta et al., 2010, Nijhuis et al., 2007). Gag cleavage site mutations at p1/p6 (L449F) and NC/p1 (Q430, A431V and L449F) have been shown to interact with the PR mutation I84V (Dam et al., 2009, Prado et al., 2002). In addition, gag mutations A431V and I437V were associated with PR mutation V82A (Clavel and Mammano, 2010). Furthermore, gag mutations (A431V,I437V and P453L) together with PR (I50V and V82A) have been found to compensate for the loss of viral fitness and contributed to the development of resistance to PIs (Su et al., 2019, Malet et al., 2007). Gag mutations at non-cleavage sites (H219Q and R409K), in combination with PR mutations (L75R, I84V) were also associated with PI resistance (Gatanaga et al., 2002).

1.6.2.2 Env mutations to Protease Inhibitor resistance

Studies have demonstrated that the mechanism for PI resistance may be mediated by mutations in the HIV-1 *env* gene (Rabi et al., 2013). Amino acid changes in the HIV-1 Env gp41 Heptad repeat (HR) (607T and 641L) and cytoplasmic tail (CT) (721I) have been associated with PI failure (Coetzer et al., 2017). In addition, it has been reported that gp41 mutations may also be associated with viral failure in patients taking the PI-based regimen (Perrier et al., 2019, Castain et al., 2019). *In vitro* studies through mutagenesis have demonstrated that Env mutations can increase the cell-cell spread and alter the stability of the HIIV Env gp120-gp41 interaction, resulting in resistance to multiple drug resistance (Hikichi et al., 2021). However, this remains unclear whether similar findings can be observed in clinical HIV-1 isolates. Recently, using a Bayesian Network (BN), we demonstrated an

interaction between known Gag-PR mutations with *env* gene mutations in sequences of patients virologically failing a PI-based regimen (Maphumulo and Gordon, 2021). However further investigation using phenotypic and site-directed mutagenesis (SDM) assays are needed to understand the effects of Env mutations alone and in combination with gag mutations on viral fitness and PI efficacy. It also remains unclear whether Env mutations directly interact with PR mutations.

1.7 Computational methods

Computational methods such as Coevolution methods, Bayesian network analysis, molecular docking, molecular dynamics, and 3D homology modeling are being used more frequently as tools for drug resistance surveillance (Kirchmair et al., 2011, Das et al., 2009, Huang et al., 1998).

1.7.1 Coevolution methods

A statistical inference technique called Direct Coupling Analysis (DCA) is used to infer direct co-evolutionary couplings between pairs of residues in multiple sequence alignments. Compared to DCA, GREMLIN is more accurate (Kamisetty et al., 2013). However, it is difficult to carry out large-scale analyses using GREMLIN the online server-based tools. In CAPS, the co-evolutionary networks connecting amino acid sites within a protein are computationally compared, and the correlated evolutionary variation of the amino acid site is measured. The CAPS technique analyses the transition probability scores between pairs of sequences at 2 sites using the blocks substitution matrix (BLOSUM)(Fares and Travers, 2006). On the other side, by undertaking fully automated blind prediction evaluations, the Automated Model Evaluation (CAMEO) platform improves the CASP experiment. CAMEO offers a variety of scores for the purpose of comparing different aspects of structure prediction algorithms (Haas et al., 2018). As a result, CAPS and CAMEO are the most effective coevolutionary tools.

1.7.2 Bayesian Network

The statistical correlation between multiple variables can be described by a probabilistic model known as a Bayesian Network (BN), which is represented by a direct acyclic graph (Pearl, 1998). BNs are learned from the observation of the data, and produce a network that has the most correlations and the fewest arcs (Heckerman et al., 1995, Deforche et al., 2006). The direct conditional or unconditional dependencies between the variables are reflected in the network (figure 1.7)

(Myllymäki et al., 2002). The interdependence represented by arcs in-between the variables indicates unconditional dependency, whereas the absence of an arc represents conditional dependency (Myllymäki et al., 2002).

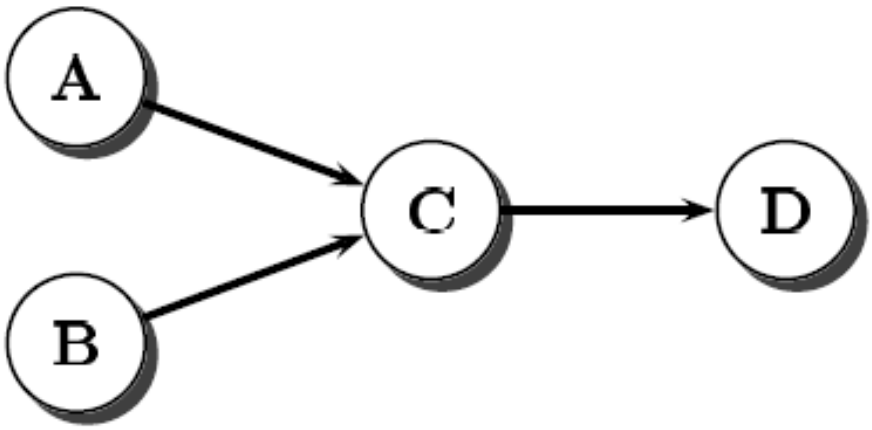


Figure 1.7: A representation of conditional and unconditional dependencies in the Bayesian Network, taken directly from (Myllymäki et al., 2002).

1.7.3 Homology modelling

Homology modelling is a quick, and inexpensive way to determine the three-dimensional (3D) structure of a protein from its AA sequences, using a comparable high-resolution protein structure available from publicly accessible databases (Bishop et al., 2008; (Muhammed and Aki-Yalcin, 2019). Homology modelling has be used to infer an undefined protein sequence's biological activity and biochemical function based on the similarity between the sequence and well-known protein structures (Shi et al., 2001). The basic steps to homology modeling is as follows: (1) Choose the best PDB templates that work with the target sequence to be modeled. (2) Align the template structures with the desired sequence. (3) Construct the 3D structure using MODELLER or Swiss Model. (4) Evaluate the model.

1.7.4 Molecular dynamics (MD) Simulations

Molecular dynamics simulations have made it possible to understand the interaction of biological molecules and probe the structure and dynamics of individual molecules, and also provide atomistic details of biomolecular motions (Weng et al., 2019, Sakuraba and Kono, 2016). The MD simulations can capture a variety of biomolecular processes, such as ligand binding, protein folding, and show

the positions of all atoms (Hollingsworth and Dror, 2018). MD simulations can also be used to test the accuracy of the modelled structure or even refine the structure (Hollingsworth and Dror, 2018). Simulations are sometimes useful in refining protein homology models (Mirjalili and Feig, 2013).

For MD simulations, a variety of online and offline tools are readily available: Chemistry of Harvard Macromolecular Mechanics (CHARMM), Groningen Machine for Chemical Simulations (GROMACS), and Assisted Model Building with Energy Refinement (AMBER) (Gajula et al., 2016, Hollingsworth and Dror, 2018).

1.8. Study significance, research aims, and objectives.

Failure of second line PI inclusive regimens has implications for future treatment options, especially in resource limited countries. While the prevalence of second line failure has reached 18.0 per 100 person years in southern Sub-Saharan Africa (Edessa et al., 2019), more than 30% of patients who failed a PI-based treatment regimen in South Africa did not harbor any resistance associated mutations in PR (Wallis et al., 2011, Meintjes and Maartens, 2012, Obasa et al., 2020) . Fun et al (2012) described mutations in Gag at both cleavage and non-cleavage sites that are associated with PI resistance. Gp41 in Env has also been suggested to contribute to PI failure, with reported mutations mainly in the HR and CT regions (Rabi et al., 2013, Perrier et al., 2019, Manasa et al., 2017, Coetzer et al., 2017, Castain et al., 2019). However, no study has described the interaction of both gag (MA and CA) and full-length Env in the development of PI resistance. Thus, the focus of this study was to determine the frequency of gp120 and gp41 mutations in patients failing PIs in subtype C South Africa and understand if these mutations coevolve with Gag-PR mutations and their structural implications.

1.8.1 Aims:

1. To determine the effect of protease inhibitor therapy on envelope variation.
2. To understand the effect of protease inhibitor resistance mutations on viral entry.
3. To determine the structural changes that occur in the envelope during PI failure.

1.8.2 Study Objectives:

1. To identify mutations in HIV-1 subtype C Env associated with PI failure.

2. To identify HIV-1 subtype C Env, Gag, and protease co-evolution in PI treatment failure.

3. To determine whether mutations in gp120 associated with PI resistance, affect co-receptor usage.

3. To use homology modelling to construct models of Env during PI failure.

1.9 Thesis outline

Five chapters make up the thesis as follows:

Chapter one comprises the introduction, literature review, Aims, and objectives of this study.

Chapter two published research paper titled "Potential Associations of Mutations within the HIV-1 Env and Gag Genes Conferring Protease Inhibitor (PI) Drug Resistance".

Chapter three manuscript titled "Gp120 mutations facilitate viral entry through immune escape and coreceptor switching during PI failure".

Chapter four manuscript titled "Structural analysis of gp41 mutations that co-evolve with Gag".

Chapter five General discussion, conclusion and study limitations, and future recommendations.

Chapter Two

Potential Associations of Mutations within the HIV-1 Env and Gag Genes Conferring Protease Inhibitor (PI) Drug Resistance

2.1. Introduction

While the dispensing of HIV-1 protease inhibitors (PIs) as part of a routine second-line regimen has been a significant turning point in the management of HIV-1 (Ghosh et al., 2008), the development of resistance to PIs is also increasing (Kim and Baxter, 2008). PIs are known to inhibit the activity of the HIV-1 protease (PR) enzyme responsible for the proteolytic processing of HIV structural Gag and enzymatic Pol polyprotein components (Pettit et al., 1994). The viral protease (PR) enzyme cleaves the Gag precursor protein (Pr55Gag) and the Gag-Pol precursor protein (Pr160Gag-Pol), resulting in virion maturation (Pettit et al., 1994, Rabi et al., 2013, LORI et al., 1988). This proteolytic process prevents the formation and maturation of infectious HIV particles. The efficacy of the PI is limited by the emergence of resistance mutations that are potentially caused by poor compliance, subtherapeutic systemic levels of the drug, or prolonged treatment with one PI-based regimen during virologic failure (Zdanowicz, 2006).

Interestingly, >20% of patients failing a PI-based regimen do not harbor any resistance associated mutations in the protease (PR) domain (Gallego et al., 2001, Obasa et al., 2020, Wallis et al., 2011). While non-adherence has been identified as a key player, some studies have suggested that Gag mutations can indirectly affect the Gag cleavage site and drive PI resistance without any mutations in the protease region (Perry, 2014, Alfadhli et al., 2016). These mutations were commonly found close to the cleavage sites, although non-cleavage site mutations have also been linked with PI resistance (Codoñer et al., 2017, Fun et al., 2012). Verheyen et al. (2006) have linked Gag precursor p7–p1 (A431V, K436R, and I437V) and p1–p6 (L449P, P452S, and P453L) cleavage site mutations with resistance to PIs (Malet et al., 2007, Nijhuis et al., 2007, Verheyen et al., 2006). These mutations are thought to compensate for any enzymatic impairment of protease caused by the loss of van der Waals interactions between the inhibitor and binding sites (Fun et al., 2012, Malet et al., 2007, Nijhuis et al., 2007, Kletenkov et al., 2017, Larrouy et al., 2010, Gatanaga et al., 2002, Kožíšek et al., 2012).

Others have suggested that the mechanism for PI failure may involve mutations in the *env* gene (Coetzer et al., 2017, Rabi et al., 2013). Coetzer et al. (2017) reported gp41 mutations from both the Heptad repeat (HR) (607T and 641L) and cytoplasmic tail (CT) (721I) that potentially contribute to PI failure (Coetzer et al., 2017). Two studies that looked at virological failure during PI exposure both reported that mutations in the CT impacted PI susceptibility (Castain et al., 2019, Perrier et al., 2019). Furthermore, Env mutations have been suggested to promote cell to cell transmission,

leading to high level drug resistance mutations in ARV target genes; by doing so, they increase the level of resistance to a broad panel of ARVs in vitro (Van Duyne et al., 2019, Hikichi et al., 2021). The HIV-1 Env is said to concurrently evolve to escape from both neutralizing antibodies (NABs) and ARVs (Hikichi et al., 2021).

Although studies have reported the emergence of Gag and Env mutations in patients failing PIs (Castain et al., 2019, Coetzer et al., 2017, Perrier et al., 2019), to date, no one has reported on the coevolution of Gag and Env in these patients. Here, we investigated the prevalence of Env mutations associated with PI treatment failure and the possibility of coevolution with mutations within Gag, and protease, in full-length Env and Gag-PR sequences from HIV-1 subtype C infected individuals failing a PI-inclusive regimen.

2.2. Materials and Methods

2.2.1. Study Cohort

This retrospective study used 24 stored plasma samples obtained from virologically failing PI treated patients enrolled in the Protease Cleavage Site (PCS) study (2009–2013) at Mc Cords and King Edward VIII hospitals Durban, South Africa (Marie and Gordon, 2019). All enrolled patients received Lopinavir/Ritonavir therapy for at least six months and had plasma HIV-1 RNA levels >1000 copies/mL. In addition, sequences from 344 subtype C drug-naïve isolates were downloaded from the Los Alamos HIV-1 Database (http://hiv-web.lanl.gov) accessed on 10 September 2020. Written informed consent was obtained from all study participants. The Biomedical Research Ethics Committee approved the study of the University of KwaZulu-Natal (BREC NO: 678/17).

2.2.2. Amplification and sequencing analyses of the Env domain

Viral RNAs were extracted from 140 µL of plasma using the QIAamp RNA kit (QIAGEN Services, Inc., Germantown, MD, USA) and reverse-transcribed using a Thermoscript RT-PCR kit (Invitrogen, Carlsbad, CA, USA). A 1.7 kb Gag-Protease product was amplified by nested PCR using the Takara Ex Taq HS enzyme kit. Specific primers, Gag +1 5 0 -GAGATCTCTCGACGCAGGAC-30 (HXB2 nucleotide: 675 to 697, forward primer) and 3 00rvp 50 -GGAGTGTTATat GGATTTTCAGGCCCAATT-30 (HXB2 positions: 2696 to 2725, reverse primer), were used for the first-round PCR, at 55 ◦C for 30 min (cDNA synthesis) and 94 ◦C for 2 min (initial denaturation), followed by 35 cycles of 94 ◦C for 15 s (denaturation), 55 ◦C for 30 s (annealing) and 68 ◦C for 2 min (extension), and ended with a 5 min

incubation at 68 ◦C (final extension). Long fwd 50 -GAC TCG GCT TGC TGA AGC GCG CAC GGC AAG AGG CGA GGG GCG ACT GGT GAGTAC GCC AAA AAT TTT GAC TAG CGG AGG CTA GAA GGA GAGAGA TGG G-30 (HXB2 nucleotide: 695 to 794, forward primer) and Long rev 50 -GGC CCA ATT TTT GAA ATT TTT CCT TCC TTT TCC ATT TCT GTA CAA ATT TCT ACT AAT GCT TTT ATT TTT TCT GTC AAT GGC CAT TGT TTA ACT TTT G-30 (HXB2 nucleotide: 2706 to 2805, reverse primer) were used for nested PCR at the following conditions: 94 ◦C for 2 min (initial denaturation), 40 cycles of 94 ◦C for 30 s (denaturation), 60 ◦C for 30 s (annealing) and 72 ◦C for 2 min (extension) followed by a 7 min hold at 72 ◦C (final extension). Primers were complementary to subtype B, gag-protease-deleted NL4-3 plasmid (pNL4-3Δgag-protease) on either side. Amplicons were electrophoresed in a 1% agarose gel to confirm the presence of the 1.7 kb amplicon corresponding to the gag-PR gene. The size of the product was determined using the GeneRuler™ 1 kb DNA Ladder (Fermentas International Inc., Waltham, MA, USA), as shown in Supplementary Figure S1.

Next, we amplified an Env fragment spanning HXB2 positions 5580 to 8586 by nested PCR using first-round primers Rec2F 50 -GATAAAGCCACCTTTGCCTAGT-30 (HXB2 positions: 5514) and Env2 50 -TTCTAGGTCTCGAGATACTGCT-30 (HXB2 positions: 8889), second-round primers Env1 50 -AAGGGCCACAGAGGGAGCCATA-30 (HXB2 positions: 5580) and E270R, 50 -GCGTCCCAGAAGTTCCACAA-30 (HXB2 positions: 8566) and PCR conditions as previously described (Rangel et al., 2003). Sequences were generated using the Gag-Protease sequencing primers 50 -CTT GTC TAG GGC TTC CTT GGT-30 (HXB2 position: 1078–1098), 5 0 -CTT CAG ACA GGA ACA GAGGA-30 (991–1010), 50 -GGT TCT CTC ATC TGG CCT GG3 0 (1462–1481), 50 -CCT TGC CAC AGT TGA AAC ATT T-30 (1960–1981), 50 -TAG AAG AAA TGA TGA CAG-30 (1817–1834), 50 -CTA ATA CTGTATCAT CTG CTC CTG T-30 (2328–2353), 5 0 -CCT GGC TTT AAT TTT TAC TGG-30 (2196–2268). Cycling conditions were as follows: 94 ◦C for 2 min (initial denaturation), 35 cycles of 94 ◦C for 15 s (denaturation), 55 ◦C for 30 s (annealing) and 68 ◦C for 4 min (extension) followed by a 10 min hold at 68 ◦C (final extension). Amplification of the *env* region was confirmed by gel electrophoresis (shown in Supplementary Figure S2).

The gp160 Env was sequenced using population-based forward and reverse primers (shown in Supplementary Table S1) and the BigDye v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA, USA), and run on an Applied Biosystems (ABI) 3130xl automated sequencer. Sequences were assembled and edited using the Applied Biosystems SeqScape® Software and aligned against subtype C reference.

2.2.3. Statistical analysis

Multiple sequence alignments were generated using the ClustalX program (http: //www.clustal.org/) accessed on 16 September 2020, and manually edited using BIOEDIT (Ibis Biosciences, An Abbott Company, Sylmar, CA, USA). The Virulign tool (https:// github.com/rega-cev/virulign) accessed on 11 December 2020, was used to determine the amino acid substitution at each position in Gag-PR and Env on PI exposure and naïve isolate sequences. Fisher's exact test was performed for all positions to identify mutations significantly more prevalent (p-value < 0.05) in PI-experienced patients.

2.2.4. Bayesian network

Bayesian Network learning was performed using B-Course (http:/www.b-course) accessed on May 2021, which modelled direct and indirect associations between Gag, PR, and Env mutations and exposure to PI-treatment. Only significant mutations identified using the Fisher's exact test were included in the model together with known PI-resistance associated mutations. As previously described, associations between mutations learned and graphical representations were done (Deforche et al., 2007, Theys et al., 2010). Non-parametric bootstrap analysis was calculated by running 5000 replicates to derive network robustness. Only interactions with bootstrap support over 65% were included in the network.

2.2.5. Coevolution using CAPS.

Gag-PR-Env coevolution analysis was performed using the Coevolution Analysis for Protein Sequences (CAPS, http://bioinf.gen.tcd.ie/caps/) accessed on 26 October 2020 program. CAPS identifies groups of coevolving pairs with a correlation coefficient > 0.5 by using amino acid (AA) sites to compare the correlated variance of their evolutionary rate using the probability scores between two pairs of aligned sequences (Fares and Travers, 2006). The evolutionary rate was estimated using the blocks substitution matrix (BLOSUM). The distribution of 5000 randomly sampled values was used to identify coevolving codons.

2.6 Positive selection

Positive selection pressure in Gag -PR-Env sites was assessed using the rate ratio of non-synonymous to synonymous (ω) substitutions in the CODEML program of the PAML software package (version 4.9) (Yang et al., 2005). Analysis was performed using three classes ω=<1, ω=1, ω=>1 assuming

purifying, neutral and positive selection, respectively. The ω value and likelihood estimates were calculated for three different codon-based ML pairs of site models: M0 (one ω) vs. M3 (discrete), M1a (nearly neutral) vs. M2a (positive selection) and M7 (beta) vs. M8 (beta and ω=>1). Comparison of M0 vs. M3 is a test of site rate variation, while M1 vs. M2 is for positive selection. The likelihood ratio test was used to evaluate the best-fitting model for the data (Anisimova et al., 2001). The Bayes Empirical Bayes method was used to identify specific sites under positive selection.

2.3. Results

2.3.1. Participant Characteristics

Twenty-four Gag-PR-Env sequences from 24 patients failing a PI-inclusive treatment regimen were available for analysis, of whom 14 were females and 10 were males, with the median age of 35 years (interquartile range (IQR) 17–38 years). The median viral load was 4.84 log10 copies/mL (IQR 4.12–5.51). All patients were infected with HIV subtype C as established by the Virulign tool.

2.3.2 Prevalence of Envelope mutations in LPV/r-experienced versus ART-naïve subtype C sequences

Twenty-four Env sequences were available for analysis. The prevalence of amino acids at each codon position was calculated by comparing sequences from LPV/r-treated and ART-naïve isolates. All sequences were confirmed as subtype C using the Rega subtyping tool. Figure 2.1 shows a total of thirty-five mutations identified in the *env* gene, with significantly higher frequencies in LPV/r-experienced patients compared to those that were ART-naïve. These were: I19LSV (32% vs. 13%), S110EN (16% vs. 4%), D167AGKNR (16% vs. 2%), P183AEKQRS (48% vs. 20%), Y191CFHNS (12% vs. 1%), N195HIKRST (28% vs. 8%), N230DT (20% vs. 16%), Q258EP (8% vs. 1%), N262LI (8% vs. 0%), T283IKNQSV (28% vs. 9%), V286AIM (16% vs. 3%), I294LMN (8% vs. 0%), S306DGNR (24% vs. 5%), P313DI (8% vs. 0%), Q315AKR (32% vs. 3%), I323KNRTV (12% vs. 1%), A329DIPT (16% vs. 0%), C331RS (8% vs. 0%), W338EG (8% vs. 0%), H374KLPR (8% vs. 1%), F382V (8% vs. 1%), F383LV (8% vs. 0%), L390QV (8% vs. 0%), N425DHR (20% vs. 4%), L453EIM (20% vs. 4%), R469KTW (8% vs. 1%), V489IL (16% vs. 1%), V505AEMT (8% vs. 1%), E507GQRW (28% vs. 10%), S534AL (12% vs. 2%), T536AMSV (32% vs. 10%), Q567KR (8% vs. 1%), D632EG (24% vs. 8%), I688TV (12% vs. 1%), P724LQRSX (32% vs. 12%). In contrast, the K63RT (16% vs. 50%), T132NQ (24% vs. 97%), T138NQ (36% vs. 80%), and V182EIKLMNT (12% vs. 56%) mutations were found at a lower frequency in the PI-experienced versus ART-naïve sequences.

2.3.3 Positive selected sites in Envelope and coevolution with Gag -PR residues

Here, we determined which amino acid changes were under positive selection pressure following treatment failure using PAML (Fares and Travers, 2006). Figure 2.2 shows PAML analysis of amino acid changes from participants failing an LPV/r-inclusive treatment regimen and ART-naïve participants that were positively selected. A total of twelve sites were positively selected: V85, T138, Y146, M147, E150, K151, G152, D277, G321, K322, N463, and T465. None of these sites were previously associated with entry and fusion inhibitor or LPV/r failure.

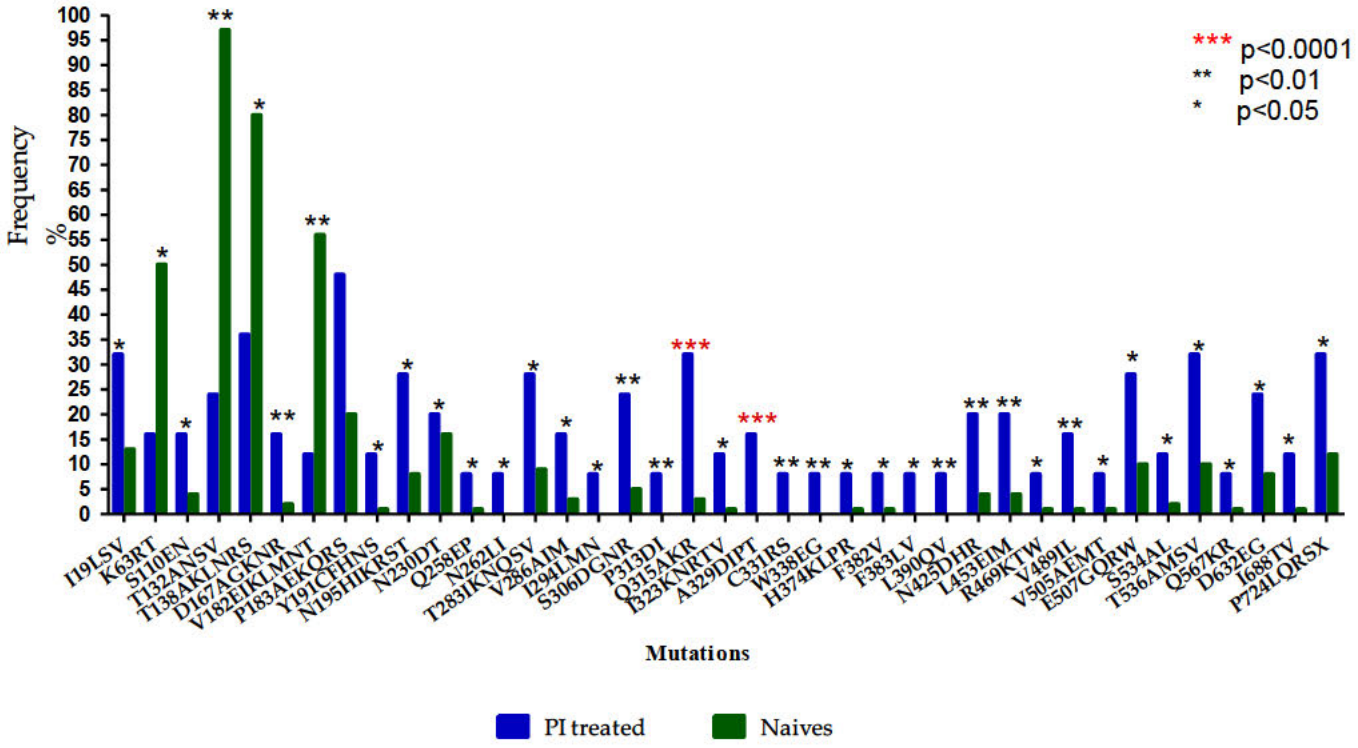


Figure 2.1: Frequency (%) of env mutations in ART-naïve versus LPV/r-treated subtype C isolates. Asterisks (*): *** represents p < 0.0001, ** represents p < 0.01, and * represents p < 0.05.

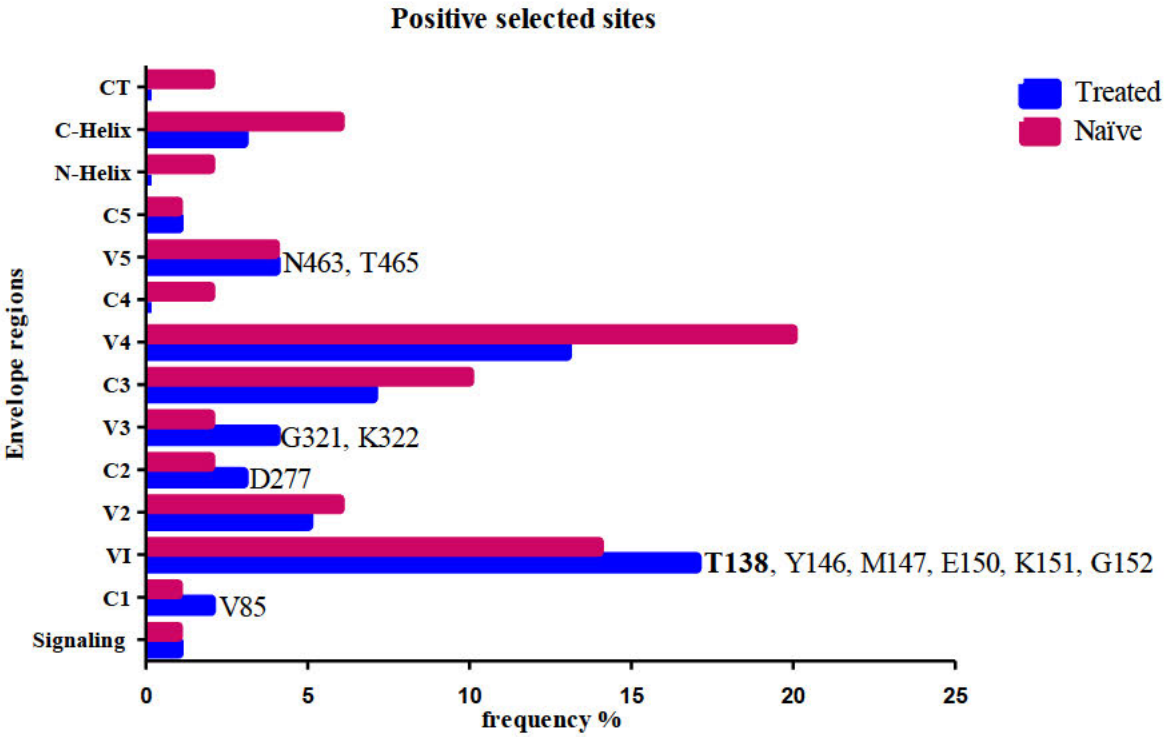


Figure 2.2: Positively selected amino acid codons in Env in LPV/r-treated and naïve isolates with the Bayesian probability of p > 99%. Blue shows the positively selected codons in the treated isolates, while maroon indicates the positively selected codons in the naïve isolates. Codon positions identified only in treated and not in naïve isolates are shown at the relevant env region. T138 is indicated in bold font and has a higher frequency in the treated isolates.

Coevolution Analyses for Protein Sequences (CAPS) was used to determine whether amino acids in Gag, PR, and Env sites coevolved in patients failing an LPV/r-inclusive treatment regimen. All positions with amino acid variation greater than 1% were included in the analysis. The coevolution analysis identified seven amino acid coevolving pairs, but these did not correlate significantly. The coevolving pairs were mostly seen between Env and Gag sites (Env: 11, 23, 29,142, 147, 336, 389, 400, 496, 534, 617, and 668; Gag: 108, 225, 256, 335, 372, 388, 478, and 480), while only one was observed between Env and PR (400-Env and 76-PR) sites.

2.3.4 Interactions between Envelope and Gag -PR mutations with LPV/r-treatment

Using Bayesian network learning, we explored the interactions between Gag, PR, and Env mutations in isolates exposed to LPV/r treatment. Only statistically significant positions according to Fisher's exact test were included in the analysis. Figure 2.3A shows mutations in the *env* gp120 region (S110N-Env, T132S-Env, T138S-Env, P183Q-Env, and Q315R-Env) that were indirectly associated with exposure to LPV/r treatment via known Gag-PR mutations (Q69K-Gag, S111I-Gag, I256V-Gag, and

V77I-PR). More specifically, P183Q-Env was indirectly associated with LPV/r treatment via interaction with V77I-PR. T132S-Env also showed a robust connection with Q69K-Gag mutation, which is further related to K20R-PR mutation. Both T138S-Env and S110N-Env showed a strong association with I256V-Gag. Interestingly, a robust direct connection with LPV/r exposure was observed between a known Gag-PR mutation (R76K-Gag) and wildtype variants at codon positions in Gag (Q182-Gag-PR-WT) and PR that have been linked to drug resistance (I50-PR-WT, V77-PR-WT, and L90-PR-WT).

Figure 2.3B shows the association between Gag, PR, Env gp41 mutations, and LPV/r treatment experience. There was a strong interaction between T536M-Env, S543A-Env, I688V-Env, and P724S-Env and Gag-PR mutations (S12T-PR, P63L-PR, Q182S-Gag, and I256V-Gag). Specifically, P724S-Env was directly associated with P63L-PR and Q182SGag and indirectly associated with A431V found at a Gag cleavage site. Another direct association was seen between S534A-Env and I256V-Gag. Both T536M-Env and I688V-Env were directly associated with S12T-PR. There was no direct interaction between treatment experience and Env gp41 mutations.

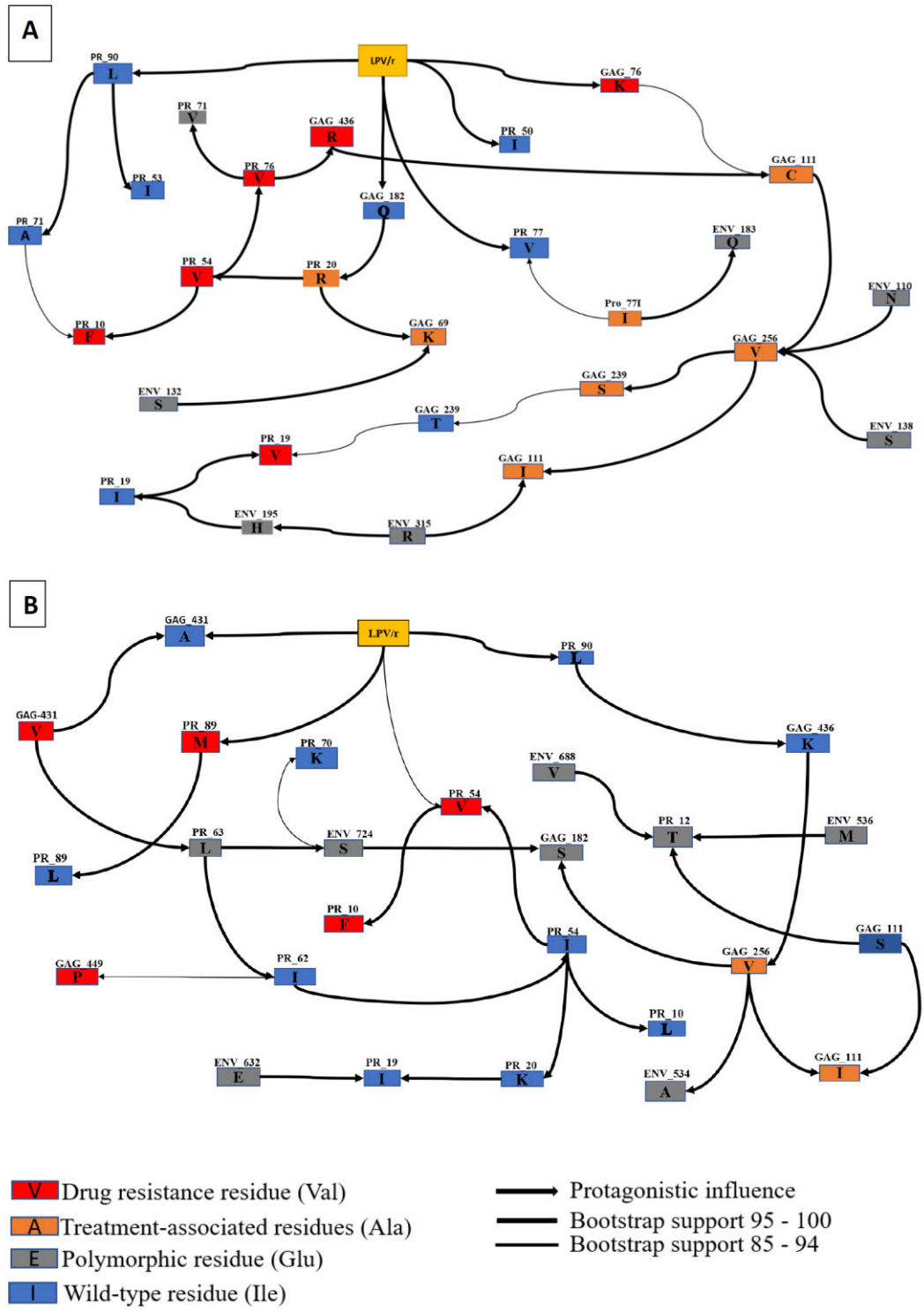


Figure 2.3: Annotated Bayesian network based on 5000 bootstraps, showing the association between nodes indicating Env and Gag mutations in subtype C LPV/r-treated sequences. Here, we determined the association between (**A**) Env gp120, Gag, PR mutations, and treatment experience and (**B**) gp41, Gag, PR, and treatment experience. The color code of nodes is defined based on their link with LPV/r: drug resistance variants (red), treatment associated variants (orange) and natural occurring: Polymorphic variants (gray), wild-type variants (blue). The arcs represent a direct dependency between corresponding nodes, and the thickness of an arc is in proportion to bootstrap support. The arc direction does not represent the accumulation of mutations or causal meaning but may indicate a multivariable effect in the network.

2.4. Discussion

Mutations in the *env* gene have previously been associated with resistance to PIs (Rabi et al., 2013); however, no studies have been conducted using subtype C PI failures, nor have they investigated their role in combination with the Gag. In this study, LPV/r-resistance associated mutations in Gag, PR, and Env were characterized in HIV-1 subtype C patients from KZN, South Africa, who were failing an LPV/r-inclusive treatment regimen. We identified several amino acids in the *env* gene that were associated with LPV/r failure. We also found potential pathways leading to LPV/r resistance that involved a combination of Gag, PR, and Env mutations.

Significantly higher frequencies of AA changes in the gp120 region of the LPV/r experienced group were commonly identified. These were mostly in the variable loops (V2: D167AR, V182I, P183QS, Y191F and N195H, and V3: S306R, P313DI, Q315R, I323KNR, A329DT, and C331RS); however, increased frequencies were also observed in the constant regions (C2: N230DT, Q258EP, N262LI, T283I, V286I and I294N, and C3: W338EG, H374KR, F382V, and F383LV). Amino acid changes in these regions can, therefore, potentially increase the infectivity and replication capacity of the virus (Dang et al., 2020).

In gp41, sequences from LPV/r failures harbored significantly higher frequencies of Heptad repeat (HR) mutations at codons 534, 536, 567, and 632. Consistent with our findings, an HIV-1 subtype A study also observed higher levels of mutations in HR in LPV/r failures (Coetzer et al., 2017). Furthermore, higher mutation frequencies were also seen in the cytoplasmic tail (CT); this is in line with other studies that reported CT mutations associated with virological failure in PI-experienced participants (Perrier et al., 2019, Coetzer et al., 2017, Castain et al., 2019). As the CT plays a functional role in Env incorporation during virion assembly, this suggests that mutations in this region may influence the efficiency of Env incorporation (Murphy et al., 2017).

Using Bayesian network learning, we found potential pathways leading to LPV/r resistance that involved a combination of Env and Gag mutations. I256V appeared to be an essential mutation as it was shown to be directly associated with three Env mutations (S110N-Env, T138S-Env, and S543A-Env) and indirectly associated with Q315R-Env via S111I-Gag and P724S-Env via Q182S-Gag. Although less is known about the I256V-Gag mutation, it has been previously associated with LPV/r failure in subtype C studies (Ntale, 2012, Singh, 2015). It has also been associated with reducing drug susceptibility to benzodiazepine and benzimidazole in Gag subtype B studies (Li et al., 2013).

Mutations T138S-Env and S110N-Env showed a strong interaction with I256V-Gag, forming a pathway (T138S-Env + S110N-Env + I256V + S111C + R76K) to LPV/r exposure. A study by Singh et al. (2015) identified I256V, S111C, and R76K in Gag following LPV/r treatment in subtype C patients in South Africa (Singh, 2015). Interestingly, T138 was also positively selected in our PAML analysis. Although a wildtype amino acid, T138, is known to play an essential role in antibody binding, and although the change from Threonine (T) to Serine (S) weakened the binding, T138S did not inhibit antibody binding (Hyser et al., 2008).

Another potential pathway for LPV/r resistance included Env mutations Q315R-Env and the same Gag mutations I256V, S111C, and R76K, with the addition of S111I. While Q315R has been identified as a naturally occurring polymorphism in subtype A (Ratcliff et al., 2013), its role as a drug-resistance-associated mutation in subtype C requires more mechanistic studies. Mutation T132S-Env showed a robust connection with Q69K-Gag, which has previously been associated with reduced susceptibility to PIs (Giandhari et al., 2016).

Interestingly, known PI-resistance-associated mutations K20R-PR, I54V-PR, L10F-PR, and L76V-PR were not directly connected to LPV/r exposure in the network, suggesting that immune selection pressure may also be in play. While not directly associated with LPV/r resistance, Env gp41 (S543A-Env and P724SEnv) mutations were associated with Gag (Q182S-Gag and I256V-Gag) mutations. Env mutation P724S was strongly associated with Q182S-Gag, which was further associated with I256V. S534A-Env was also directly associated with I256V-Gag, further supporting the role of I256V-Gag in connecting Env in the pathway to drug resistance. Interestingly, most of the mutations in Gag connected to Env mutations were located in the capsid domain (such as I256V), which is similar to other studies (Coetzer et al., 2017).

Another potential network to LPV/r resistance involved mutations found in the *env* in combination with the *gag* and *pr* region. Env mutation P183Q-Env was directly associated with V77I-PR, previously associated with PI drug resistance(Doualla-Bell et al., 2006, Mata-Munguía et al., 2014, Santoro and Perno, 2013). In addition, Env mutations T536M and I688V showed a strong direct association with PR mutation S12TPR, which is known to be selected in LPV/r- and RTV-treated patients in subtype C (Giandhari et al., 2016). Interestingly, the P724S Env mutation was linked to LPV/r failure via the P63L-PR and A431V-Gag pathways. The A431V Gag mutation located in the CS of Gag is known to confer resistance to all PIs except Darunavir (DRV) (Nijhuis et al., 2007), while P63L-

PR was observed at baseline in patients initiating antiretroviral therapy (Lopinavir/r, Lamivudine, and Zidovudine) (Nijhuis et al., 2007)[. These findings suggest that there might be other pathways to LPV/r treatment failure that involve Gag (Matrix and Capsid), minor PR, and Env mutations.

2.5. Conclusions

We found a high prevalence of Env mutations in HIV-1 subtype C associated with LPV/r treatment failure. The majority of these associations were in combination with a Gag (Matrix and Capsid) and/or PR mutation. Further investigations using site-directed mutagenesis need to be conducted to determine whether Env mutations alone can affect LPV/r efficacy.

Chapter Three

Gp120 mutations facilitate viral entry through immune escape and coreceptor switching during PI failure.

3.1. Introduction

Despite treatment adherence and maintenance of the HIV-1 virus, virological failure has been observed in patients taking PI in the absence of drug target gene mutations (Gallego et al., 2001, Pulido et al., 2012), indicating that mutations outside the protease gene can contribute to PI resistance (Rabi et al., 2013). Studies have demonstrated that the *env* gene can promote PI resistance (Rabi et al., 2013, Perrier et al., 2019). Furthermore, mutations in the HIV-1 gp120 *env* gene (C1, V1, V2, and V3) have been reported to be associated with PI failure (Maphumulo and Gordon, 2021). Apart from C1, these (variable) regions are critical for viral entry, and the coreceptor binding site is concealed by these regions (Briz et al., 2006). The V1/V2 and V3 regions are located on the outer surface of gp120 and can participate in phenotypic changes of HIV-1 (Ogert et al., 2001).

V1V2 and V3 are also recognized as targets for neutralizing antibodies and can impact viral coreceptor usage (Pollakis et al., 2001). Structures have shown that V1V2 and V3 regions function as key nanodevices to create the Env structure for immune escape compatible with the maintenance of viral infectivity (Yokoyama et al., 2016). Depending on the conformation of these loops, gp120 binds to CCR5 or CXCR4 coreceptor. Subtype B uses both CCR5 and CXCR4, and the switch in tropism occurs in 50% of patients (Schuitemaker et al., 2011). In contrast, subtype C uses CCR5 primarily in the early and late stages of infection, and the proportion of chronic infection uses CXCR4 (SMEATON et al., 2011, Jakobsen et al., 2013) (Lynch et al., 2009). This is because the coreceptor switching in subtype C requires more mutations than other subtypes (Coetzer et al., 2011).

HIV-1 gp120 is one of the most extensively glycosylated proteins with 23 or 24 N-linked glycosylation sites (MYERS and LENROOT, 1992, Pollakis et al., 2001). It's comprised of 50% glycan moiety, which influences Env conformation and affects viral entry, infectivity, and antibody recognition during CD4 and co-receptor binding that mediate membrane fusion and cell entry of HIV-1 (Wang et al., 2013, Leonard et al., 1990), (Pollakis et al., 2001). The VIV2 domain has the most variable loop length with a high number of glycosylated sites, whereas V3 is less variable in relation to loop length (Cavrois et al., 2014). Furthermore, the virus's transition from the R5 to X4 phenotype was significantly influenced by the loss of an N-linked glycosylation site in the V3 region (Pollakis et al., 2001).

Previous studies have reported the gp120 structure and changes in the structure caused by Env inhibitor failures (Prabakaran et al., 2007, Kwon et al., 2012). However, the impact of structural changes in gp120 caused by Env mutations associated with PI remains unknown. Here we

demonstrate how these changes affect co-receptor usage and will help us understand the effect of protease inhibitor resistance mutations on viral entry.

3.2. Methodology

3.2.1. Sequence analysis

The impact of PI treatment on gp120 was investigated by comparing envelope sequences obtained from virologically failing PI-treated patients enrolled in the Protease Cleavage Site (PCS) study (2009–2013) at McCord and King Edward VIII hospitals Durban, South Africa (Marie and Gordon, 2019). All enrolled patients received Lopinavir/Ritonavir therapy for at least six months and had plasma HIV-1 RNA levels > 1000 copies/mL. In addition, sequences from 344 subtype C drug-naïve isolates were downloaded from the Los Alamos HIV-1 Database (http://hiv-web.lanl.gov).

3.2.2. Homology modelling

Homology models of six mutant HIV-1 envelope proteins from PI-treated patients and one wild-type (WT) sample were obtained using the SWISS-MODEL web server (https://swissmodel.expasy.org. Template selection criteria were based on the best resolution and highest sequence identity. The crystal structure of HIV-1 concC_Base0 prefusion env trimer in complex with a human antibody fragment 3H109L and 35O25 variants at 3.5 angstroms (6CK9) was used as a template.

3.2.3. Molecular Dynamic (MD) Simulations

Molecular dynamics simulations were performed using AMBER 18, which was accessed at the Centre for High-Performance Computing (http://www.chpc.ac.za). The GPU version of the Pmemd engine provided with the AMBER 18 package was used for MD simulation. The FF14SB Amber force field was used to describe the amino acid residues of the protein. Amino acid residues of the proteins were renumbered based on the dimeric form of the enzyme from 1 to 511. The system was explicitly solvated by the TIP3P water molecules with a margin of 12.0 Å. Prior to equilibration, a two-step minimization was performed. Partial minimization of 2000 steps with an applied restraint potential of 500 kcal/mol for both solutes were carried out, this was performed for 1000 steps using the steepest descent method and then followed by 1000 steps of conjugate gradients. Furthermore, full minimization of 1000 steps were further performed by conjugate gradient algorithm without restraint. The whole system was then gradually heated from 0K to 300K, executed for 50ps, in a manner that the systems maintained a fixed number of atoms and fixed volume. Also, the

temperature was monitored using the Langevin thermostat with a collision frequency of 1.0ps, with a potential harmonic restraint of 10kcal/mol. After heating, the entire system was equilibrated at a constant temperature of 300K, with the additional features such as several atoms and pressure also kept constant mimicking an isobaric-isothermal ensemble (NTP) as mentioned in (Kehinde et al., 2019). followed by the equilibration estimating 500 ps of each system was conducted. Finally, the Molecular dynamic simulations were performed for 100ns, and for each simulation, the SHAKE algorithms were employed to constrict the bonds of hydrogen atoms. As mentioned by Kehinde et al (2019), the simulations coincide with the NTP, with randomized seeding, the constant pressure-coupling of 1 bar maintained by the Berendsen barostat, a pressure-coupling constant of 2ps, a temperature of 300K, and Langevin thermostat with a collision frequency of 1.0ps (Kehinde et al., 2019).

3.2.4. Post-Dynamic Analysis

The root mean square fluctuation (RMSF), root mean square deviation (RMSD), and radius of gyration (ROG) were performed using the CPPTRAJ modules implemented in Amber18. Origin data analysis software (Origin Lab, Northampton, MA) was used to generate all the graphs. Structures were viewed and analyzed in the UCSF Chimera software package https://www.rbvi.ucsf.edu/chimera (Pettersen et al., 2004), and Discovery studio Visualizer Software, 2021 https://3ds.com/products-services/biovia/products.

3.2.5 GP120 Characteristics

We used the online Los Alamos sequence database tools to determine the characteristics of the HIV-1 variable regions (gp120- V1 to V5 loops). This online tool provides results of the HIV N-linked glycosylation site, the loop length, and the V3 loop net charge (NC) where we used default settings that have computed with KRH = (+) and DE = (-). The following link was used: https://www.hiv.lanl.gov/content/sequence/VAR_REG_CHAR/index.html. The Kruskal Wallis tests were used to compare the deferences between treated and naïve.

3.2.6 Coreceptor switch

HIV-1 V3 sequences were submitted to Geno2Pheno[coreceptor] (https://coreceptor.geno2pheno.org/), for classification of coreceptor usage (CCR5-using or CXCR4-using. A false-positive rate (FPR) cut-off of 2.5% and 20% was used (Poveda et al., 2012). In addition, the 11–25 rule was applied to define coreceptor tropism (Arif et al., 2017).

3.3. Results

3.3.1 The Length of the variable loops in the PI-treated vs naïve groups

Differences in loop lengths were compared between the PI-treated and naïve groups. Using the Kruskal-Wallis test, the V1/V2 and V5 lengths were significantly higher in the treated group (P-values of 5.611e-05 and 2.2e-15 respectively), as shown in Fig 3.1a. As expected, V1V2 and V5 hypervariable regions were also significantly higher. Interestingly, the V4 hypervariable region was the only one that was significantly lower in the treated group (P-value = 0.002242).

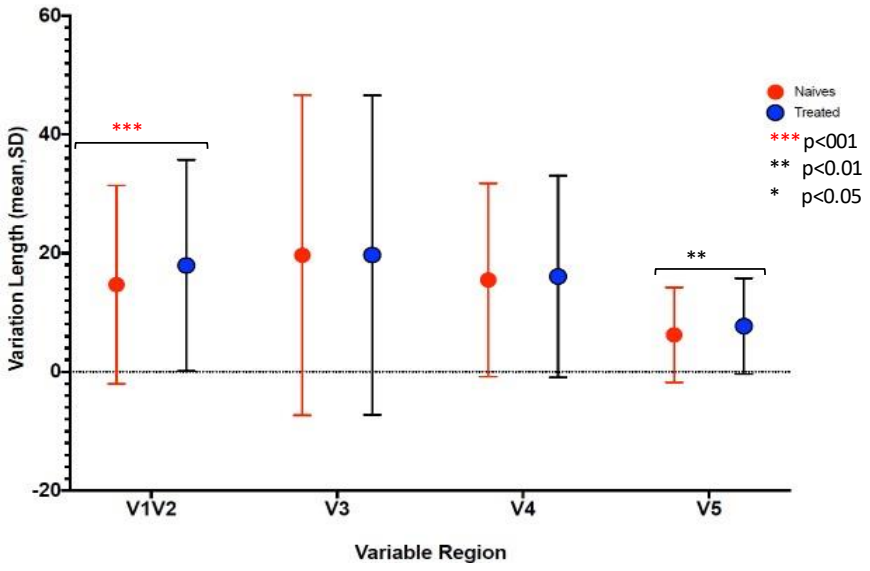
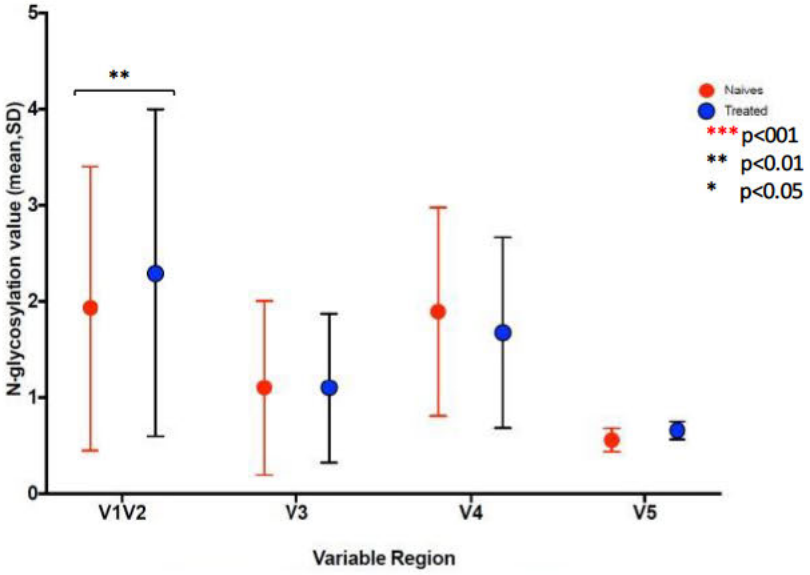


Figure 3.1: Frequency distribution of variable region in treated vs naïve. GP120 characteristics, were obtained in the Los Alamos and applied in the R-programme to calculate the P-value and showed the variable length frequency distribution between treated and naïve using mean and standard deviation. V1/V2 and V5 lengths were significantly higher in the treated group with the P-values of 5.611e-05 and 2.2e-15 respectively.

3.3.2 Functional sites in gp120

Interestingly, putative glycosylation sites were found in most regions of the envelope, except C5 and V5. The number of N- glycosylation sites for V1/V2 were significantly higher (P-value = 0.008727) in the treated group as shown in fig 3.2a. Putative phosphorylation sites that were seen in the constant regions (C1 and C3) usually used the SXXD/E motive while the variable regions used the TXXR/E motive. Putative phosphorylation sites were seen at constant regions AA 110, 115, 264, and 365 and variable regions 161 and 303. Of note, the S110N mutation resulted in the loss of this phosphorylation site which in turn resulted in the acquisition of an N-link glycan. Conserved N-myristoylation sites were mainly seen in C1 and C2.

A



B

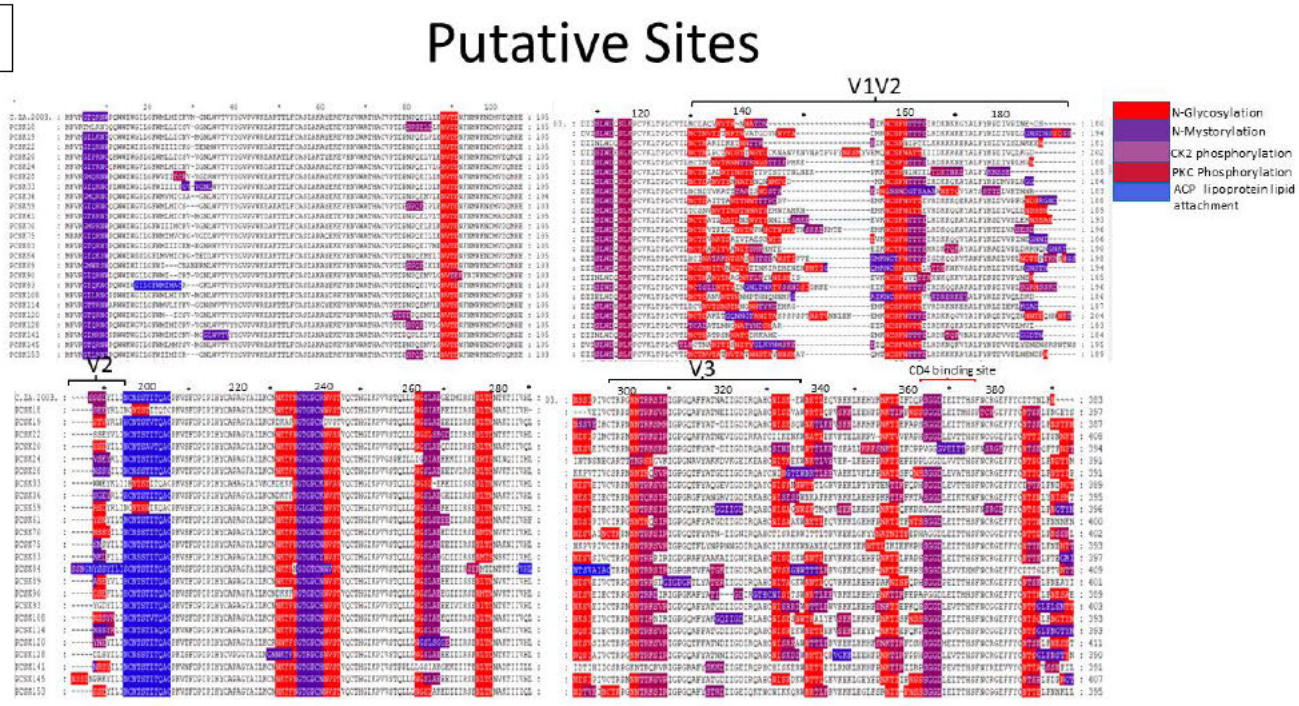


Figure 3.2: A) Frequency distribution of glycosylation sites in treated versus naïve, calculated in R-programme showing mean and standard deviation. The number of N- glycosylation sites for V1/V2 were significantly higher with the P-value = 0.008727. **B)** Alignment of the subtype C PI-treated sequences and the naïve sequence. Putative sites were determined using PROSITE in GENEDOC. Functional sites are shown by the colored boxes: red - N-glycosylation site; purple - myristoylation site; plum – CK2 phosphorylation magenta - PKC phosphorylation; blue - ACP-lipoprotein lipid attached site.

3.3.3. Coreceptor prediction

Using a 20% FPR, 37.5% of the treated group were predicted to be CXCR4 viruses and this decreased to 20.8% when using an FPR of 2.5%; this was still significantly higher than the naïve group 1.2% (p=0.0002). A similar result was obtained when using the 11/25 rule. The V3 charge was significantly higher in the treated versus naïve group (p= 0.015).

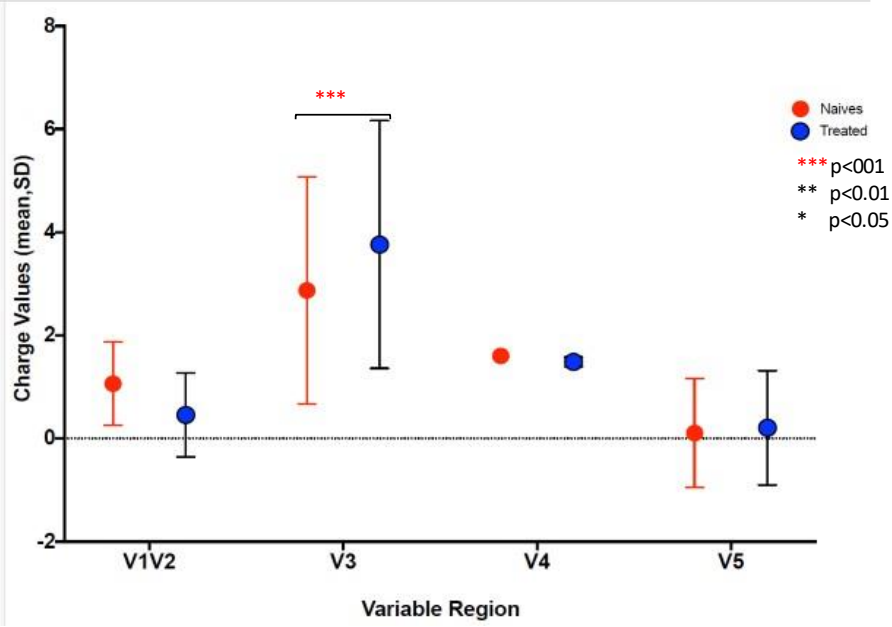


Figure 3.3: Frequency distribution of GP120 charge in treated versus naïve, showing mean and standard deviation. Treated is shown in blue and naïve is shown in red and the p-value is in black.

Table 3.1*:* Coreceptor prediction in subtype C sequences, using a FPR of 20% and 2.5%.

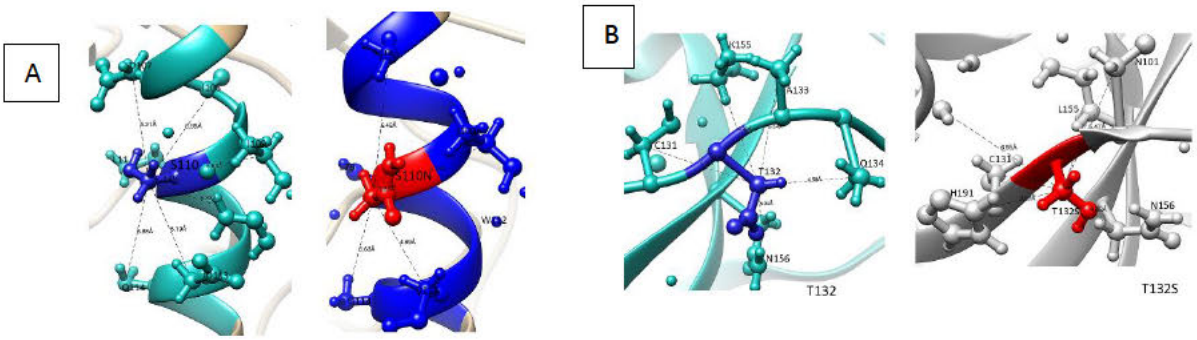
	Coreceptor Prediction		
Sample ID	Predicted Coreceptor (20% FPR)	Predicted Coreceptor (2.5% FPR)	11/25 Rule tropism
PCSK18	R5	R5	R5
PCSK19	R5	R5	R5
PCSK20	R5	R5	R5
PCSK22	R5	R5	R5
PCSK24	X4	R5	R5
PCSK28	R5	R5	R5
PCSK33	R5	R5	R5
PCSK36	X4	X4	X4
PCSK59	R5	R5	R5
PCSK61	R5	R5	R5
PCSK70	R5	R5	R5
PCSK75	X4	R5	R5
PCSK83	X4	X4	X4
PCSK84	X4	R5	X4
PCSK89	X4	X4	X4
PCSK90	X4	X4	R5
PCSK93	R5	R5	R5
PCSK108	R5	R5	R5
PCSK114	R5	R5	R5
PCSK120	R5	R5	R5
PCSK128	R5	R5	R5
PCSK141	X4	X4	X4
PCSK145	R5	R5	R5
PCSK153	X4	R5	R5

3.3.4. Effect of potential PI-associated resistance mutations on the structure of gp120

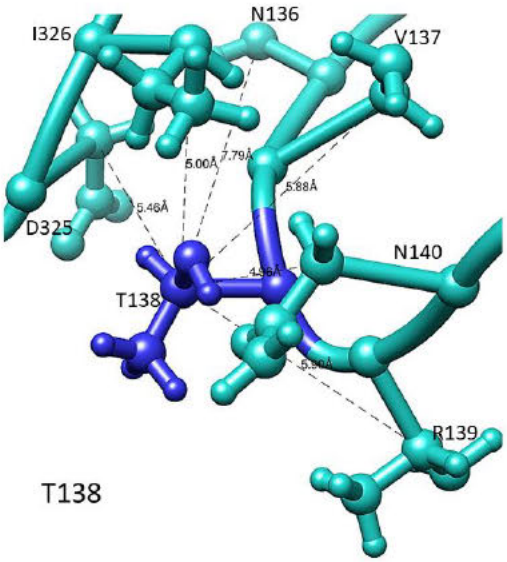
A previous study from our laboratory using a Bayesian Network reported mutations in an envelope that potentially contribute to PI failure (C1-S110N; V1-T132S, T138S; V2-P183S; V2- N195Hand V3-Q315R). Fig3.4a-d illustrates how these mutations affect envelope structure. We also provide information on their interactions with other residues within a 5-radius range (Appendix 3.1), as well as the number and distance of hydrogen bonds. The S110 interacts with positions E106 – S115 and R429, while the mutant S110N interacts with the same positions and loses contact with R429 which is the contact residue in the C4 region. T132 and T132S interact with similar AAs, WT T132 showed additional interaction with S158, N186, and D187 while T132S interacts with additional AAs V134 and Y191 as shown in figure 3.4b. Although mutation T138S and WT T138 interact with the same AAs, T138S interacted with additional AAs (N141, G142, N143, N152, E153, R327 & Q328). P183,

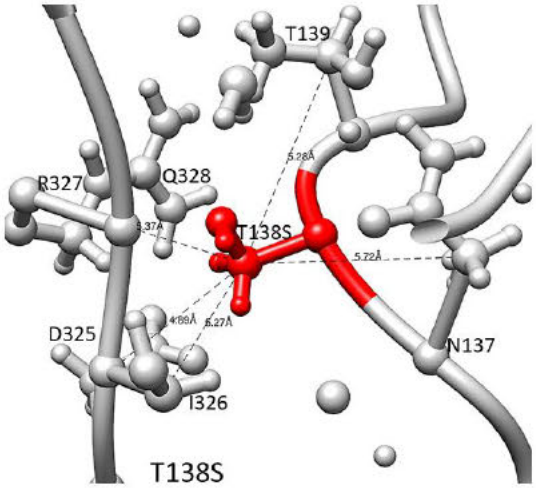
P183Q, and P183S also interact with similar AAs, P183Q showed additional interaction with N186 and S189, whereas P183S showed additional interaction with G186, Y193, and I194. Furthermore, only P183S contains the hydrogen bond interaction with G185. Both N195 and mutation N195H interact with AAs at the coreceptor-specific site and CD4 contact residues in V2 and C4 respectively and show the hydrogen bond interaction with S198 and S199. Q315 and Q315R are shown to interact with the same position. In addition, Q315R interacted with (T123 and P124). They both form two hydrogen bonds with I311, and Q315R forms an additional hydrogen bond with K121 (2.92 Å) found at the coreceptor binding site. From the superimposed V3 and C3 structures shown in fig3.4g, the R5 viruses showed conserved V3 and C3 structures while X4 viruses showed more variability between their respective structures which could impact the efficacy of the coreceptor and CD4 binding state.

Lastly, we looked at the molecular surfaces on the discovery studio visualizer (solvent-solid-parent color) of the modelled proteins fig3.4h. Structures from two of the isolates (K19 and K36) showed an open pocket around the coreceptor binding site outside of V3 (S375, F376 & N377); this was not visible in the other structures. This opening exposes the coreceptor CCR5/CXCR4 binding sites and usually occurs after gp120 has bound to the CD4 receptor.

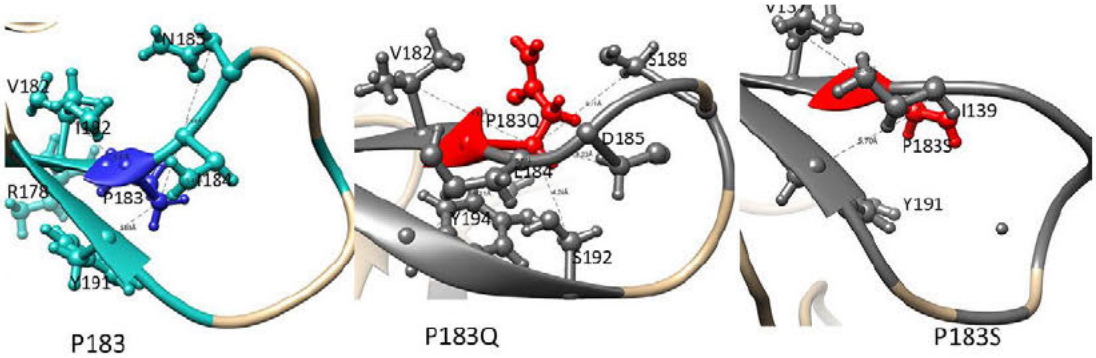


C

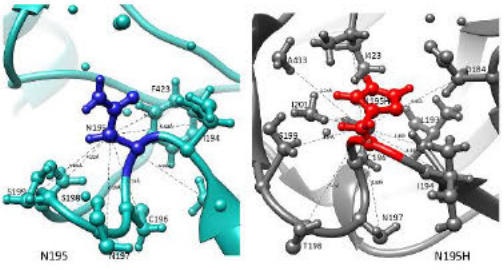

N136
I326
V137
D325
N140
T138
R139
T138


T139
R327
Q328
T138S
D325
N137
I326
T138S

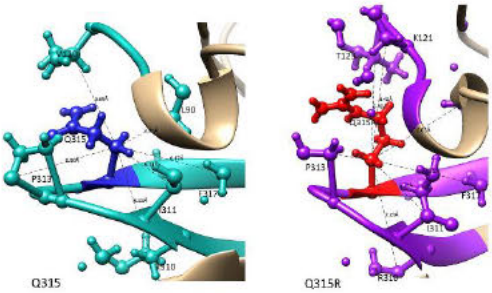
D


V182
P183
R178
V191
P183
V182
S188
P183Q
D185
L184
Y194
S192
P183Q
I139
P183S
Y191
P183S

E


N195
N195H

F


Q315
Q315R

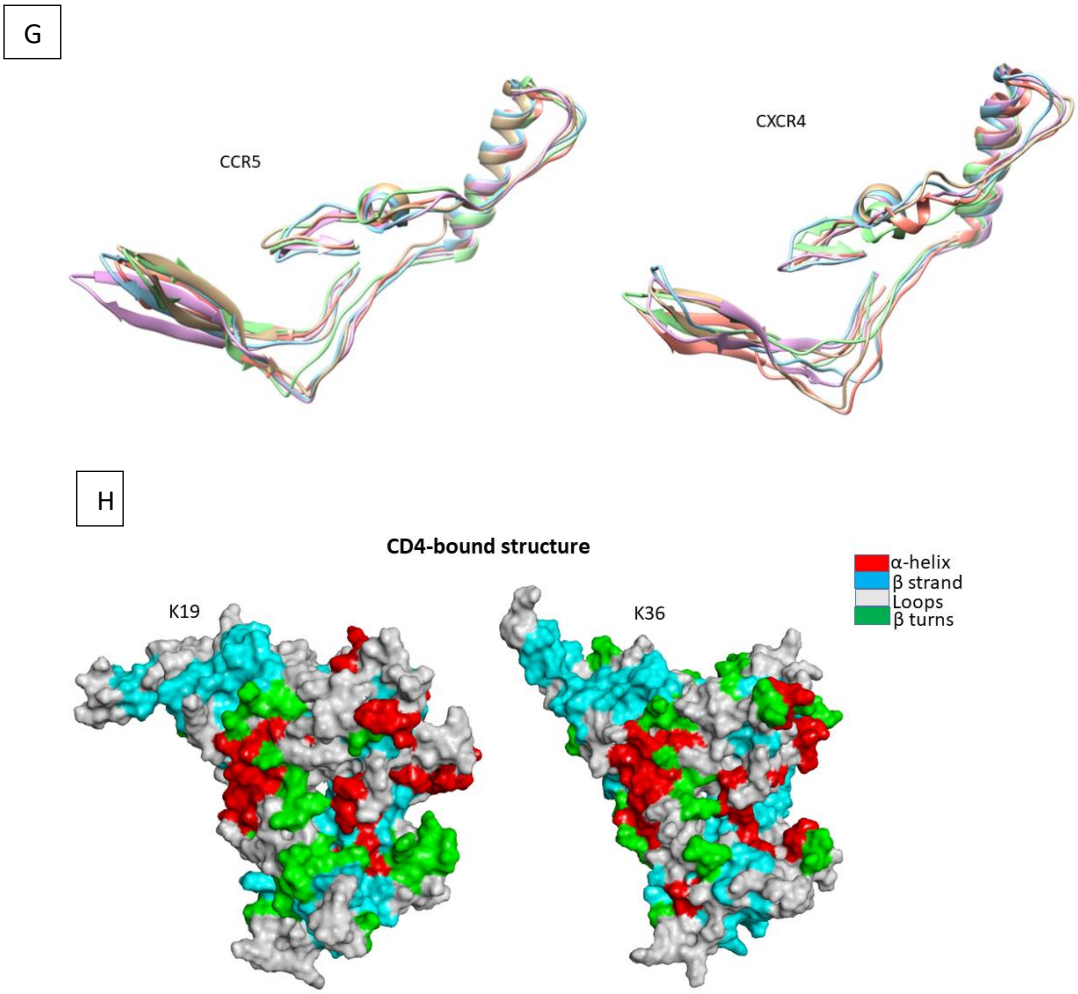


Figure 3.4: The interaction of mutations that potentially contribute to PI failure with other AA residues within a 5-angstrom radius. WT residues are highlighted in blue, and mutations are presented in red ball and stick for (**A)** S110N in the C1 region, (**B)** T132S and (**C)** T138S in the V1 region, (**D)** P183S and (**E)** N195H in the V2 region, and (**F)** Q315R in the V3 region. **G)** The structure of V3 C3 showing the GPGR Tip of the V3 loop and CD4 binding sites in the superimpose protein for CCR5 and CXCR4 usage. **H)** The structural surface shows the opening on the CD4-bound structure.

3.4. Discussion

Using various online tools, we've studied the gp120 characteristics, coreceptor switching, and structural model. According to the characteristics of gp120, the lengthened V1V2 loop seen in the treated group and the corresponding increase in N-glycosylation sites may have protected these viruses against CD4-binding-site-directed neutralizing antibodies, possibly by shielding the underlying region of envelope glycoprotein from antibody recognition (van Gils et al., 2011). N-linked glycosylation is recognized as a major mechanism for minimizing virus-neutralizing antibody response and is important for the induction of virus entry and fusion and has been shown to

promote a switch to CXCR4 usage in disease progression (Pollakis et al., 2001). Furthermore, the increased length of V5 may increase resistance to neutralization by reducing antibody binding affinity through steric hindrance (Guo et al., 2012). The V4 hypervariable region was observed to affect the coreceptor usage, the process of independent exchange of hypervariable regions, and the screening of chimeric envelope for the ability to infect both lymphocytes and macrophages, showed V4 hypervariable region to suppress CXCR4 mediated entry(Ghaffari et al., 2005, Smyth et al., 1998). In this study, we showed the shorter length of the V4 hypervariable region, and these were found in 67% of CXCR4 using sequences however, its role remains unclear.

N-myristoylation and phosphorylation sites were found in the very conserved area in the constant region. However, those with mutations that caused a loss of phosphorylation did not show any change in the structure, while phosphorylation has been associated with protein regulation and stability,(Garcia-Garcia et al., 2016, Giroud et al., 2011) in this instance, this does not appear to be the case. On the other hand, conserved N-myristoylation could play a role in signal transduction, intracellular host, and budding (Wang et al., 2021, Maurer-Stroh and Eisenhaber, 2004, Udenwobele et al., 2017).

Pastor et al (2006) stated that mutations in the V1/V2 region directly or indirectly improved CCR5-mediated entry (Pastore et al., 2006). In this study, the V1 mutation T138S was observed to interact with the receptor binding site and may enhance the interaction of gp120 with receptors to facilitate entry (Chaillon et al., 2011). Furthermore, V2 positions P183 and N195 are found within the co-receptor-specific sites, and the changes in the V2 regions P183S/Q and N195H are expected to lead to the development of CD4-independent entry and binding of gp120 to the CCR5, However, variants with an alteration in asparagine reduces CCR5 binding (Kolchinsky et al., 2001). This was observed in this study where the of N195H (4/6=67%) mutation was frequently detected with CXCR4. In addition, the sequences consisting of both V1 T138S and V2 N195H mutations showed the opening in the coreceptor binding site on the env surface structure. This could be a result of CD4-bound, where V1V2 rotates outward in the opening state (Wang et al., 2013). One of the mutations was in the conserved region of V3-tip GPGQ. The Q315R mutation alone is sufficient to alter the co-receptor usage, and minimal changes in V3 sequences are known to result in co-receptor switching from CCR5 to CXCR4 (Pastore et al., 2006).

The CCR5-tropism viruses are typically less positively charged than those of CXCR4 viruses. An increase in V3 net charge can convert CCR5 tropic viruses into CXCR4 tropic viruses (Yokoyama et al., 2012, Shioda et al., 1994). Pollakis et al (2001) reported that the high charge in V3 and the loss of the N-linked glycosylation site within the V3 region can lead to a complete switch from CCR5 to CXCR4 virus (Pollakis et al., 2001). Although, our results showed an increase in the V3 net charge of the treated versus naïve group which will influence to the co-receptor switch, no changes in N-glycosylation were observed in both groups within the V3 region. Consistent with the charge results, the coreceptor tropism change from CCR5 to CXCR4 were also observed using Geno2pheno analysis. This was consistent with the studies that reported that the mutations in the coreceptor binding site reduced the binding efficiency of gp120 to CCR5 (Rizzuto and Sodroski, 2000, Reeves et al., 2004).

Our results showed that 38% (9/24) of the sequences were CXCR4 viruses, and their structure was shown to be altered compared to the conserved structure showing CCR5 viruses. The non-conserved structure of the CXCR4 virus structure was due to the substitutions of residues in the tip and after the β-hairpin. The GPGR V3 tip may impact predicted β-turn and co-receptor engagement at this site (Wang et al., 2021). The frequency of the coreceptor switch for naïve patients was much lower than that of the treated due to the findings that the coreceptor switch between diagnosis and starting antiretroviral therapy is rare (Mortier et al., 2013). Although this study has shown interesting results in the occurrence of gp120 mutations in the PI-treated group, the effects of coreceptor switch in the presence of these mutations, and their effect on coreceptor usage is still unknown. Furthermore, the mechanism of entry will help us understand if the virus containing these mutations fuses at the plasma membrane or spreads within the host via cell-to-cell fusion.

3.5. Conclusion

The study suggests that gp120 mutations could facilitate PI resistance by increasing the length of V1V2 and increasing the overall number of N-glycosylation sites. Env mutations in PI-treated isolates further facilitate viral entry through immune escape and coreceptor switching induced by higher charge in the V3 region, and mutations in coreceptor-specific sites. Facilitating the viral entry could possibly allow the virus to have an effect in replication capacity. Therefore, further studies need to be done to investigate if env mutations directly modulate PI efficacy.

Chapter Four

Structural analysis of gp41 mutations that co-evolve with Gag.

4.1 Introduction

Previous investigations of PI failure in the absence of mutations in the PR region suggested that failure could be driven by mutations in Gag (Fun et al., 2012, Parry et al., 2011). Studies on Subtype C reported Gag mutations (Q69K, K95R, R286K, N374S, E428D, A431, K436R, I437V, L449P, R452K, P453L, S465F, D480E, and Y484P) that were associated with PI-failure and were located in the cleavage as well as non-cleavage sites (Kelly Pillay et al., 2014, Giandhari et al., 2016, Marie and Gordon, 2019, Obasa et al., 2020). These Gag mutations were thought to compensate for an impaired protease, indirectly decreasing the susceptibility of viruses to the PIs (Gatanaga et al., 2002, Malet et al., 2007, Nijhuis et al., 2007, Larrouy et al., 2010, Kletenkov et al., 2017).

Rabi et al (2019) was the first to suggest that mutations in the HIV-1 envelope could also contribute to PI failure, and this has since been supported by other researchers who have reported PI-associated mutations in gp41 of Env (Manasa et al., 2017, Coetzer et al., 2017, Castain et al., 2019). From these reports, most mutations were found in the CT (L721I, I781T, T818, V832I, and I/L836F) (Coetzer et al., 2017, Manasa et al., 2017, Perrier et al., 2019, Castain et al., 2019). This is not surprising, as it is well known that the MA domain of Gag and the CT of gp41 interact during HIV-1 particle assembly (Freed and Martin, 1995, Dorfman et al., 1994, Postler and Desrosiers, 2013) and that this interaction also regulates viral infectivity, since the CT is only able to fuse with target cells during cell entry once the MA and NC domains in Gag are cleaved (Wyma et al., 2004).

In chapter 2 we reported novel gp41 mutations that could potentially contribute to PI failure. Further to that, our group reported Gag mutations Q69K, R76K, Y79F, S111C/I, T239A/S, I256V, A431V, K436R, and P453L associated with LPV/r treatment failure (Singh, 2015, Marie and Gordon, 2019), and most of these mutations (Q69K, R76K, S111C/I, T239S, I256V, and A431V) were also shown to co-evolve with mutations in gp41 in the pathway to LPV/r resistance (Maphumulo and Gordon et al (2021). Here, we constructed homology models of gp41 and Gag to investigate how these co-evolving mutations affect their protein structures during PI failure.

4.2 Methods

4.2.1 Sample selection

Twenty-four samples obtained from virologically failing PI-treated patients and sequenced as described in Chapter 2. Fifteen samples were selected for structural analysis. Of these 15 isolates, 11 harboured the mutations of interest in the gp41 region and did not harbour any PI mutations in the PR region. A WT isolate whose gp41 sequence was most similar to the consensus subtype C was selected for comparison. Samples selected for the Gag structures had at least six out of the seven mutations that were reported in chapter 2 and included a mixture of those with and without PI mutations.

4.2.2 Homology modelling

Homology modelling was conducted using SWISS-MODEL web server (https://swissmodel.expasy.org. The crystal structure of the HIV-1 envelope trimer 16055 NFL TD CC (T569G) in complex with Fabs 35022 and PGT124 (PDBID: 5UM8) was used to model the N-terminal regions of gp41. The membrane-proximal external region, transmembrane domain, and cytoplasmic tail was modelled using the PDBID: 7LOI as a template. Template selection criteria for Gag was based on the best resolution and highest sequence identity with the query sequence. The structure of the N-terminal 283-residue fragment of the HIV-1 Gag polyprotein (1L6N) was used as a template.

4.2.3 Molecular Dynamic (MD) Simulations

Molecular dynamics simulations were performed as previously described (Eche et al., 2021). Briefly, MD simulations were performed using the FF14SB Amber force field in Amber 18 and was run using the GPU version of the Pmemd engine accessed at the Centre for High-Performance Computing (CHPC) (http://www.chpc.ac.za). Amino acid residues of the gp41 and Gag proteins were renumbered based on the protein length from 1 to 341 and 1 to 283 respectively. The system was explicitly solvated by the TIP3P water molecules with a margin of 12.0 Å. Prior to equilibration, a two-step minimization was performed. Partial minimization of 2000 steps with an applied restraint potential of 500 kcal/mol for both solutes were carried out, followed by full minimization of 1000 steps. The whole system was then gradually heated from 0K to 300K for 50ps, followed by equilibration of the entire system at a constant temperature of 300K. The MD simulations were performed for a total of 100ns.

4.2.4 Post-Simulation Analysis

The root mean square fluctuation (RMSF), root mean square deviation (RMSD), and radius of gyration (ROG) were performed using the CPPTRAJ modules implemented in Amber18. Snapshots were analysed every 10000 frames using CPPTRAJ Ring server (https://ring.biocomputingup.it/) was used to identify salt bridges, hydrogen bonds, disulphide bridges. The structures were viewed and analyzed in UCSF Chimera (Petterson et al., 2004).

4.3 Results

4.3.1 Gp41 mutation pattern

Five gp41 mutations that were found at significantly higher frequencies in PI failures (as described in Chapter 2) were further investigated: T536A/M (T536A-22% and T536M-4%), S543A (13%), D632E (22%), I688V (9%), and P724Q/S (P724Q-22% and P724S-9%) (Fig 4.1). P724Q (13%), D632E (9%) and S534A (9%) occurred without any other mutations in gp41. Interestingly, T536A always occurred in combination with at least one other gp41 mutation: P724S (9%), D632E (4%) and I688V (9%). Of note, P724S only occurred together with T536A.

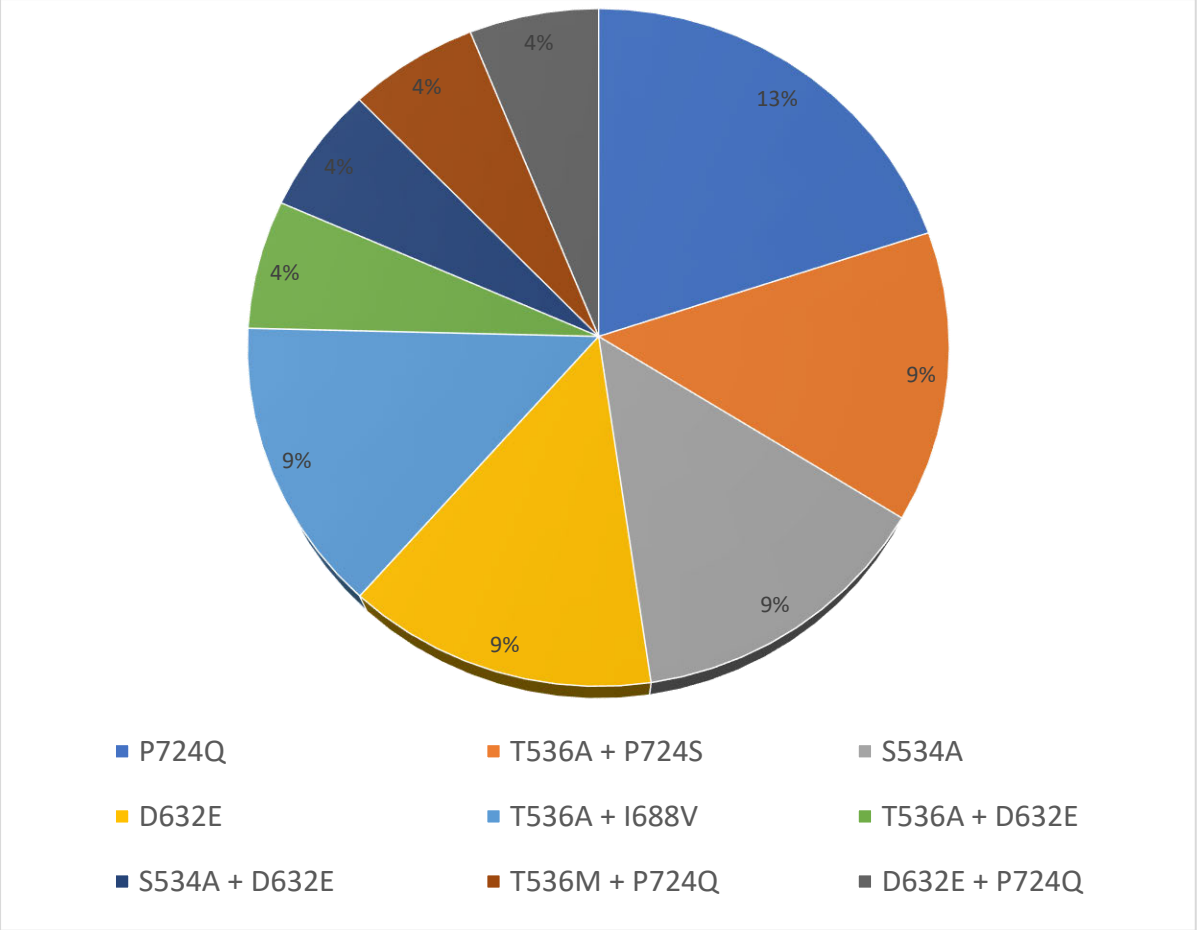


Figure 4.1: Pie chart showing the gp41 mutational pattern.

4.3.2 GP41 interactions

Representative structures containing the mutations identified in the BN are shown in this chapter. Due to the availability of template structures for the C-terminal region of gp41, the modelled structures were divided into the ectodomain (fig 4.2 & 4.3) and the TM/CT (fig 4.4). The lowest energy structure for each sample was chosen as it represents the more stable protein complex. The dynamic stability of MD simulations was evaluated using RMSD (appendix 4.1). Furthermore, all snapshot structures were superimposed and are also shown in the appendix 4.2. Hydrogen bond interactions from both the Ring server and Chimera are shown in Table 4.1.

Table 4.2: The gp41 mutations and the residues they interact with through hydrogen bond. The hydrogen bonds distance taken at the threshold of less than, or equal to 3.5Å. Mutations that have PR mutations are highlighted bold.

PID	Mutations	H-bond (3.5Å)	H-bond in Chimera
WT	S534	L537 & T538	T538
	T536	A533	A532 & Q540
	D632	W628 & N636R	M629 & N636R
	I688	I684, F685, G691 & L692	I684, G691 & L692
	P724	NONE	NONE
PCSK75	S534A	L602	G531 & L602
PCSK93	D632E	NONE	W628 & T639
PCSK36	P724Q	L721T, R744 & R761	L721T, R729, D743, R744 & R761
PCSK70	P724Q	P727L & D728	P727L & D728
PCSK145	P724Q	E734, E735 & R747	L721I, R725, E734, E735, D743 & R747
PCSK24	S534A	M530	M530 & G600
	D632E	W628 & N636D	W628
PCSK28	T536A	A532, A533 & Q540	A532 & Q540
	I688V	I684 & F685	I684, F685 & L692
PCSK83	T536A	Q540	A533 & V539A
	P724S	L721T	L721T
PCSK84	T536A	A532, A533, V539 & Q540	A532 & Q540
	D632E	N636E & Y643	W628, T639 & Y643
PCSK114	T536A	NONE	A532, V539A & Q540
	P724S	E735 & E736	P722 & E736
PCSK120	D632E	W628	W628 & R633
	P724Q	L727, D728, L730 & E731	G726, L727, D728 & L730
PCSK33	**D632E**	**R585, W628 & M629**	**R585 & W628**
PCSK61	**S534A**	**M530 & L602**	**L602**
PCSK89	**T536M**	**V539 & Q540**	**Q540**
	P724Q	**S716 & L721A**	**L721A**
PCSK108	**T536A**	**A525, A532, V539 & Q540**	**A532 & Q540**
	I688V	**I684, F685, L692, G693 & L694**	**I684 & L692**

From the Ring Server and Chimera, S534A formed a hydrogen bond with M530 (also within the FPPR) in one sample, with L602 in the loop region in a second sample and with both M530 and L602 in a third sample Fig 4.2 shows how the S534A mutation caused the hydrogen bond formed between the WT S534 and L537 and T538 in the FPPR (shown in the pink α-helix, WT fig 4.2a) to shift to L602 at the beginning of the loop region (shown in grey). Furthermore, S534A disrupted the interaction of T536 and A532 that were seen in the WT. The hydrogen bonds seen between the WT T536 and A532 and Q540 in 4/5 of the samples remained the same in the T536A mutant.

All 4 samples with D632E formed hydrogen bonds with W628 (in combination with other residues), which was also seen with D632, which formed hydrogen bonds with W628, R629 and N639R. Of

note, D632 in HR2 formed an ionic interaction with R585 in the loop region in 5/15 samples, table shown in appendix 4.3. Interestingly, those with D632E and that did not have any PI resistance mutations in PR also did not show this ionic interaction. In the WT sample, an ionic interaction was seen between K601 in the loop region and E654 in HR2 that was not seen in any of the mutant samples (fig 4.3). The I688V mutation occurred in the α-helix of the TM region. In both samples with I688V, hydrogen bonds were formed with I684, F685 and L692; these hydrogen bonds were also seen in the WT structure (fig 4.4b).

Lastly, P724Q/S occurred in the loop in the CT (fig 4.4c). P724Q formed hydrogen bonds with L721I, L727 and D728, and intermolecular interaction was also seen between P724Q of chain 1 and E735, D743 and R747 of chain 3.P724Q/S interacted with L721A/I/T in 4/7 samples, of these four samples two further interact with D743, two of seven samples interact with L727 and D728 and the last interaction was with E735 in different chains (2 sample); WT P724 did not form any hydrogen bonds with other residues.

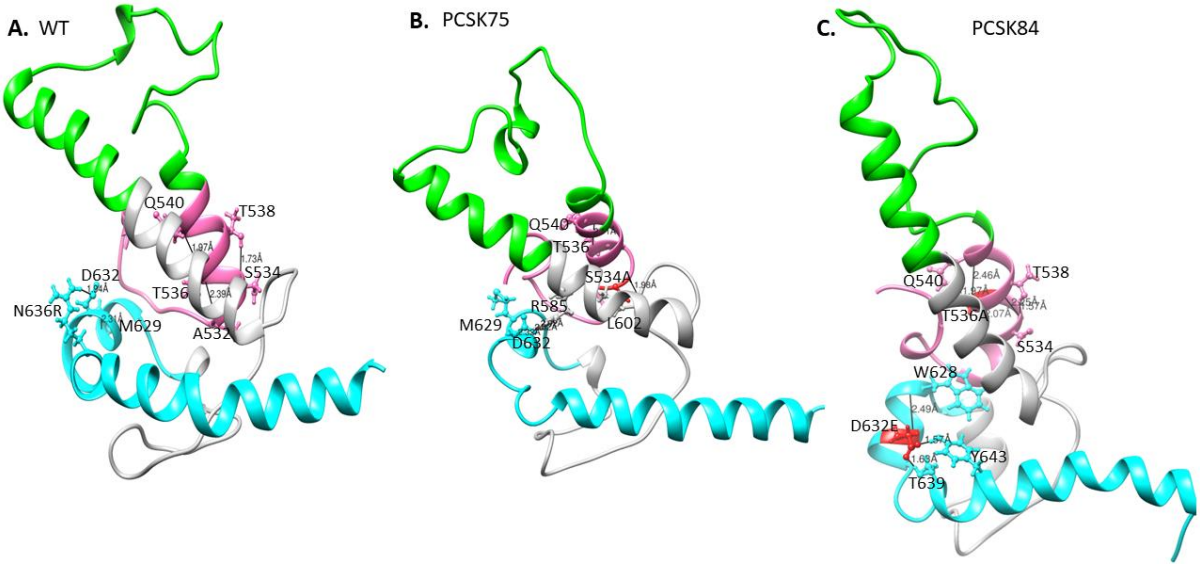


Figure 4.2: Part of the ectodomain region showing the **A**) WT structure residues S534 and T536 interacting with residues T538, A532 and Q540, and D632 interacting with M629 and N636R respectively, **B**) PCSK75 showing S534A interacting with residues within the FPPR and in the loop region, and **C**) PCSK84 showing T536A interacting with residues within the FPPR and D632E located in HR2 interacting with residues within the HR2. The colour presentation is: FP (pink), HR1 (green), and HR2 (cyan). Mutations are shown in red ball and stick.

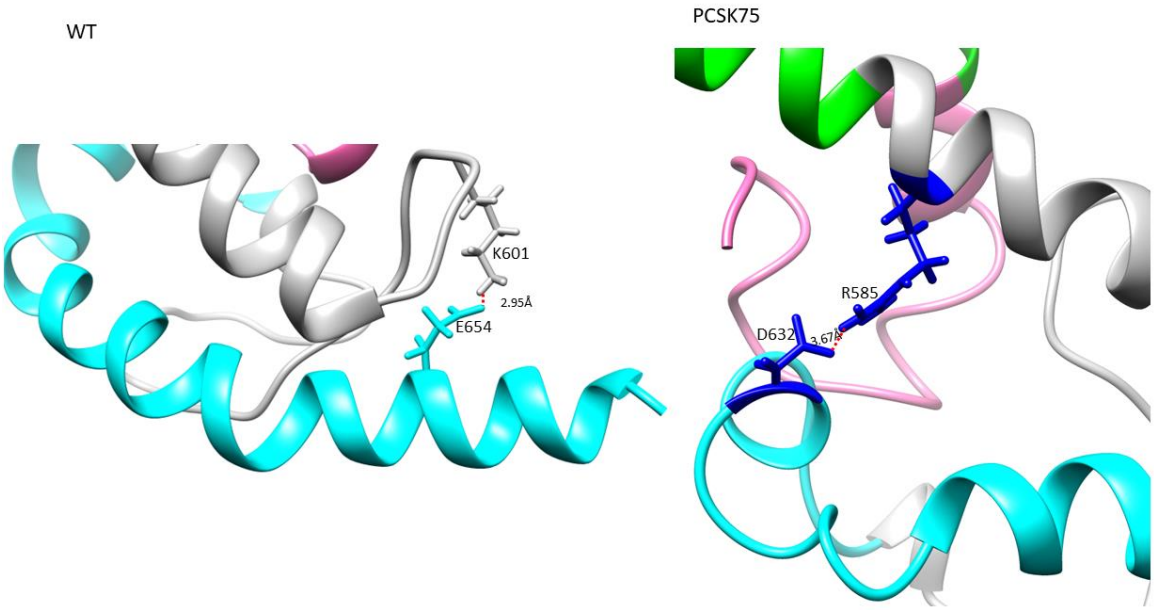
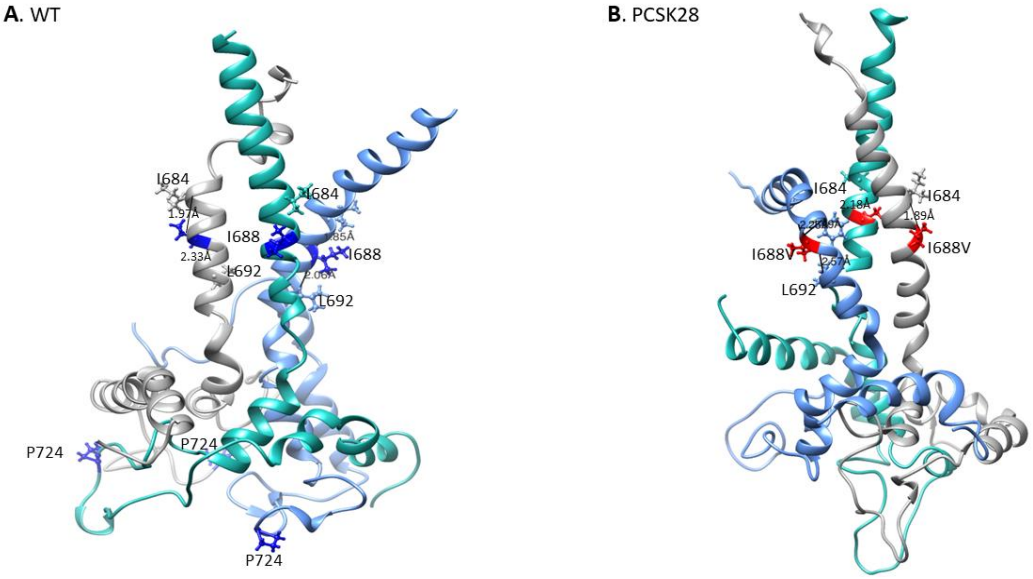


Figure 4.3: The Ectodomain structure showing a salt bridge between residues K601 and E654 in the WT, and PCSK75 showing a salt bridge between D632 and R585. Grey indicates the loop region, cyan is HR2, and pink is the FP region.



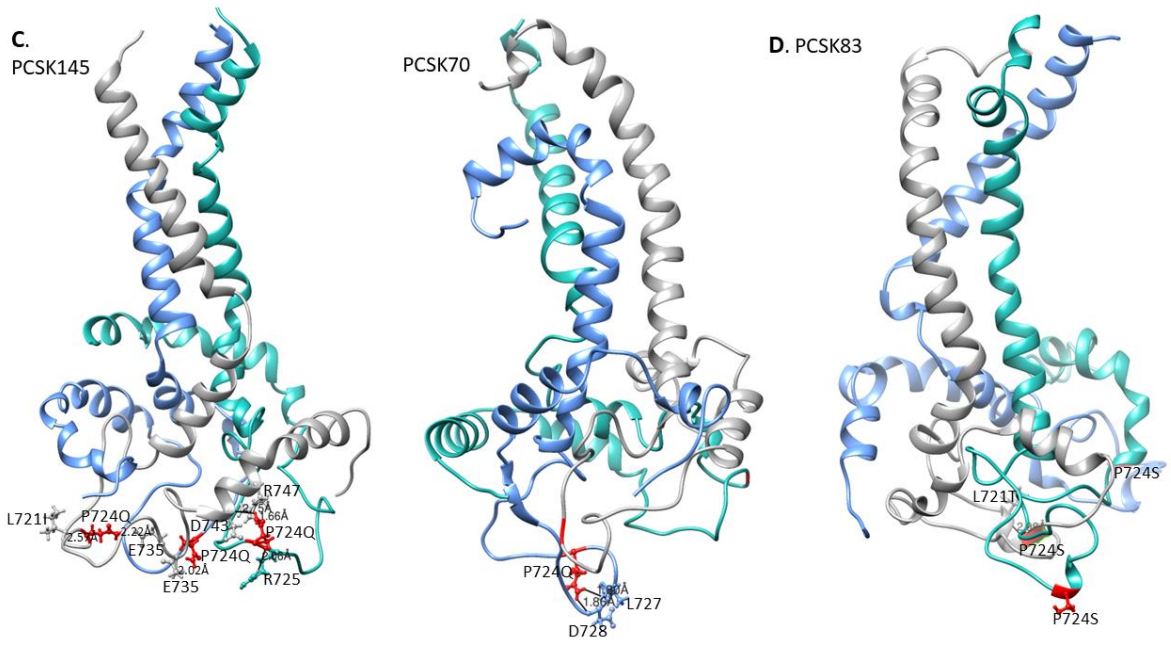


Figure 4.4: The structure of gp41 presenting MPER in purple, TM in blue, CT in green. Mutations are shown in red ball and stick. The structure of **A)** the WT, **B)** PCSK28 showing the TM mutation I688V interacting with MPER residues, **C)** PCSK70 and PCSK145 structures showing the P724Q mutation with no interaction and **D)** PCSK83 structure showing the P724S mutation.

4.3.3 Gag mutations associated with PI-failure.

We previously reported Gag mutations in the MA and CA domain that are potentially associated with PI failure and interact with gp41 (Chapter 2). Here we show the impact of these Gag mutations on the protein structure. Interestingly, most MA and CA mutations occurred in the alpha-helices (fig 4.5). Of note, both the wild type (Q69) and mutant (Q69K) formed a hydrogen bond with residue Q65 and residue Q65 & P66 respectively. Mutants with either R76K and S111C lost a hydrogen bond, while Y79F and I256V mutants did not have any change in hydrogen bonding compared to the WT. However, the length of the bond for the WT-I256 and mutant-I256V were 3.086Å and 2.062Å respectively. Mutants with T239S in the CA domain formed a hydrogen bond with Q117 in the MA domain as shown in Table 4.2.

Table 4.8: The Gag mutations and the residues they interact with through hydrogen bonds and VDW bonds. Taken at the distance threshold of less or equal to 3.5Å for hydrogen bond and the distance threshold for VDW is 0.5Å. Mutations with PR mutations are shown in bold.

PID	Mutations	H-bond (3.5Å)	VDW (0.5Å)
WT	Q69	Q65	
	R76	G71, T72, E73, F79, N80	
	Y79	V82, A83	
	S111	E107, Q108, K114, T115	
	T239	S234	
	I256	W249	
PCSK18	Q69K	Q65, S66, Q196, D200	Q65
	R76K	E71, T72, E73, Y79F, N80	E73, N80, E207
	Y79F	L75, R76K, V82, A83	
	S111I	G107, Q108, A115	E107
	I256V	W249	W249
PCSK24	Y79F	A83	Q65, L68, L75
PCSK75	Q69K	P66	
	I256V	W249	
PCSK84	Q69K		
	Y79F	V82, L75, A83	Q65, L68, L101
PCSK93	R76K	T72, E73, N80	E73
	S111C	E107, Q108, L114, A115T	
PCSK70	R76K	T72, E73, N80	G71
	I256V	W249	P231, W249
PCSK128	Q69K	P66S	
	I256V	W249	
PCSK59	Q69K		T240
	R76K	T72, E73, N80	
	S111C	T115	
PCSK83	R76K	G71,Q69, T72,F79, N80	N80
	Y79F	L75, R76K, V82, A83	L101, L68
	S111C	E107	W249
PCSK114	I256V	W249	
PCSK19	**I256V**	**P252, W260, R259**	**V251, W260**
PCSK33	**Q69K**		
	R76K	**T72, E73, Y79, N80**	**G71, N80**
	I256V	**W249**	
PCSK61	**Q69K**		
	R76K	**E72, N80**	
	I256V	**W249**	**I261**
PCSK89	**R76K**	**S72, Y79, N80**	
	S111C	**E107**	
PCSK90	**Q69K**	**E65, P66**	
PCSK108	**Q69K**		
	R76K	**G71, T72, E73, N80**	
	S111C	**E107, A115M**	
	T239S	**Q117, S234**	
	I256V	**W249, I261**	
PCSK153	**Q69K**	**L283**	**I282**
	R76K	**T72, E73, Y79F, N80**	**E73**
	Y79F	**L75, R76K, V82, A83**	**L101**
	I256V	**W249**	

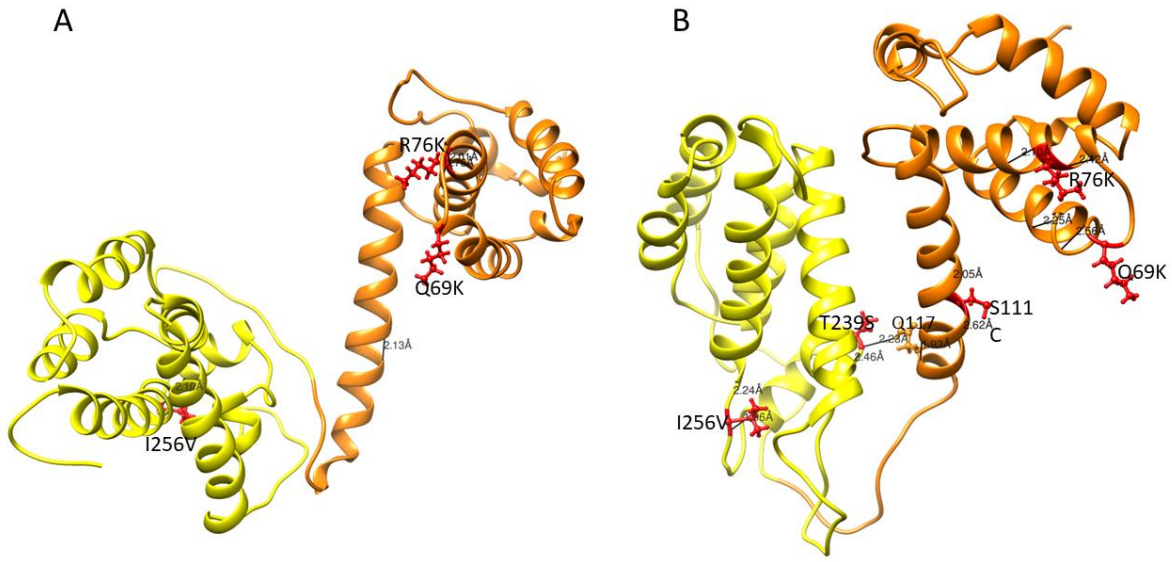


Figure 4.5: Gag structure presenting the MA and CA in orange and yellow respectively. Mutations are showed in red ball and stick **A**) PCSK70 structure showing Q69K, R76K and I256V, this patient did not horbour mutations in the PR gene. **B**) PCSK108 showing Q69K, R76K and I256V, as well as S111C and T239S. This patient harboured both PR and mutations.

4.4 Discussion

In chapter 2, mutations in gp41 and Gag of HIV-1 subtype C were shown to coevolve in patients failing a PI-inclusive regimen (Coetzer et al., 2017, Maphumulo and Gordon, 2021). In this chapter, the effects of those mutations on their respective protein structures were investigated. Overall, most of the mutations associated with PI resistance were located in the α-helices, which could be related to constraints on folding during the fusion step of viral replication (Melikyan, 2008). In addition, key mutation S534A that have been previously shown to reduce fusion with target cells showed a change in hydrogen bonding pattern with residues within FPPR, and bind with residues in the loop region, suggesting that its affect the fusion (Lu et al., 2019). Furthermore, these changes suggests that this mutant would favour the cell-to-cell route of infection rather than via membrane fusion. The T536A mutation gained the hydrogen bond within the FPPR, suggesting that this mutation stabilizes the FPPR α-helix. Interestingly, T536A did not occur in combination with S534A, suggesting an antagonistic relationship between these two mutations. In the CT, P724Q/S was found in the highly immunogenic loop known as the Kennedy epitope which is expressed on the exterior of free infectious viral particles (Cleveland et al., 2000, Cleveland et al., 2003, Santos da Silva et al., 2013), while P724S was also associated with PI failure in the BN in Chapter 2, and suggests a link between the development of ARV resistance and immune escape.

S534A and T536A are both found in the FPPR, which is an important region involved in the association of gp120 with gp41 (Freed et al., 1990, Cao et al., 1993). Lu et al. (2021) showed that S534 in the FPPR formed hydrogen bonds with N656 in HR2. They suggested that this interaction stabilized the formation of the Env trimer, and that this interaction was crucial for viral fusion (Markosyan et al., 2003, Lu et al., 2019, Lu et al., 2021). In this study, S534 did not show any interaction with residues in HR2. Instead, S534A formed a hydrogen bond with L602, found in the disulphide loop of the loop region that connects HR1 and HR2. This loop is the site of non-covalent contacts with gp120 C5 that are essential for entry and are identified as the furin recognition sites (Cao et al., 1993, Caffrey, 2001, Sen et al., 2007, A Yi et al., 2016). This interaction suggests that S534A might affect the gp120-gp41 interaction. Interestingly, neither S534A nor T536A formed any interactions with HR, suggesting that these mutations could negatively affect the association of gp120-gp41. Hikichi et al (2021) has shown that where viral entry is inhibited, the virus reverts to a predominant cell-to-cell mode of infection (Hikichi et al., 2021). This suggests that in viruses under drug selection pressure, they will preferentially use the cell-to-cell route of infection in the presence of these mutations (Freed et al., 1990, Lu et al., 2019, Lu et al., 2021, Hikichi et al., 2021).

Lan et al (2022), reported HIV-1 cell-to-cell transfer to be significantly more effective compared to normal replication (Lan et al., 2022). Cell-to-cell improved viral spread efficiency by preventing neutralizing antibodies from interfering with the entry and overcoming the restrictive effects of several anti-viral restriction factors. The main cell-to-cell mechanism is the structure called Viral Synapse (VS), which is formed between infected and uninfected cells and enhanced the HIV-1 infection (Wang et al., 2017). This method of viral transmission also has a negative impact on the HIV-1 inhibitors (Abela et al., 2012, Malbec et al., 2014, Dufloo et al., 2018, Pedro et al., 2019, Lan et al., 2022). Although our sequence data is showing that these mutations are not found in the same isolate, single or combination of S534A & T536A, can significantly reduce HIV-1 infectivity but not affecting viral production (Lu et al., 2019).

A salt bridge between D632 and K574 was previously reported and was associated with viral entry, inhibition and the stabilisation of the interaction between the α-helixes of HR1 and HR2 (Jiang and Debnath, 2000, He et al., 2008). In this study, this intramolecular salt bridge was seen between D632 and R585 (5/15 sequences) in the loop region and is most likely responsible for the same function. The movement of the salt bridge appears to be subtype specific, although there are no signature

mutations at these positions. This suggests that other mutations are involved in this structural difference between the subtypes. Interestingly, sequences with D632E but did not harbour any PI associated resistance mutations did not form this salt bridge; however they did maintain the same hydrogen bonds (D632E with W628 and N636) as the WT, which supports Jiang and Debnath et al (2000) findings that mutants with D632E retain the same activity as the D632 WT and does not affect viral infectivity (Jiang and Debnath, 2000, He et al., 2008).

The I688V interaction with same residues as WT. Although the distance between I688 and I584 in MPER region showed the longer distance from that of the WT, whereas its interaction with L692 within the TM formed the shorter distance compared to the WT, suggesting a stronger folding within the TM then with MPER. The trimer presented the WT P724 without hydrogen bond formation and mutations P724Q/S appeared to form both intramolecular and intermolecular hydrogen bond with L721A/T, L727, D728, and E735 and D743 respectively. The intramolecular interaction presented stability in retaining the hydrophilic loop to help maintain the β-turn, whereas the intermolecular interaction suggests the strengthening of the trimer interface.

Structural changes were also observed in Gag mutants. The Q69 residue was previously reported to be in the α3/4 loop and contributed to structural flexibility in Gag (Codoñer et al., 2017). In this study, mutants with Q69K shown in the α4/5 loop appeared to be longer suggesting that it occupied more volume than the WT. Our α-helix numbering were double checked and they matched that of the template (1L6N). Matrix mutations R76K and S111C lost a hydrogen bond, suggesting that there was increased flexible around the α-helix. This is in line with a study that reported a loss of hydrogen bonds to be associated with flexibility around the helix which could increase the affinity or accessibility of MA-CA cleavage sites for PR (Parry et al., 2011, Fun et al., 2012). Moreover, the loss of hydrogen bonds in the α-helix can allow better access of solvent to the carbonyl group which is the functional group where carbon double-bonded to an oxygen (Woolfson and Williams, 1990). Interestingly, Y79F mutants did not gain or lose any hydrogen bonds, suggesting that the presence of this mutation did not contribute to any changes in the α-helix. The T239S (Capsid) mutation has been previously reported to stabilize the MA-CA structure and maintain protein folding by gaining a hydrogen bond between the two domains (Hubbard and Haider, 2010). The I256V mutant showed a small difference in the hydrogen bond distance in CA-CA interaction compared to the WT. The shorter distance suggested a stronger interaction in α-helix structure, which could influence viral assembly and maturation (Fun et al., 2012, Saito and Yamashita, 2021). These results showed minor

changes in the structure which led to structural flexibility within the α-helices in MA and retain of structural folding in CA.

4.5. Conclusion

Most gp41 mutations identified in this study affect the flexibility of the protein. The S534A mutation might affect the fusion process, favouring a cell-cell route of transmission. Interestingly, all Gag mutations also contributed to the flexibility of the protein which could in turn affect MA-CA cleavage. While the study showed the effect of mutations in Gag and gp41 separately, it was not able to show if these occurred simultaneously, which is a limitation of the study. Furthermore, this study did not conduct functional assays such as replication capacity to confirm the effect of Env mutations in combination with Gag mutations on viral fitness and PI efficacy. Lastly, the sequences generated for this study only display the N-terminal portion of the CT in gp41, and the effects of mutations in the C-terminal portion are not taken in-to account.

Chapter Five

General Discussion, Conclusion, Limitations and Future work

General discussion

Protease inhibitors are a class of anti-HIV drugs that inhibit the activity of HIV-1 PR, an enzyme that cleaves the Gag and Gag-Pol precursor proteins during viral maturation (Yang et al., 2012, Arts and Hazuda, 2012). Protease inhibitor resistance develops in a sequential process, with most major (can cause resistance on their own) resistance mutations found in the active site of the PR domain, leading to reduced PI binding affinity. This also causes decreased binding affinity of the natural substrate, Gag, which in turn can affect viral replication capacity (Nijhuis et al., 2007). To compensate for this, secondary mutations (which can only cause resistance in the presence of major mutations) arise and improve the processing effectiveness of the resistant enzyme (Borman et al., 1996, Nijhuis et al., 1999). However, several studies have reported patients experiencing virologic failure (VL >1000cpm) without resistance-associated mutations in the PR region, that were unrelated to non-adherence issues. This suggested a role for other HIV-1 genes in the development of PI resistance (Pulido et al., 2012, Gallego et al., 2001). Mutations in the *gag* (Alfadhli et al., 2016, Parry et al., 2011) as well as *env* genes have since been shown to contribute to the development of PI resistance (Rabi et al., 2013). However, most of these investigations involving *env* have focused on gp41 (Perrier et al., 2019, Coetzer et al., 2017, Castain et al., 2019). None of these studies have reported on gp120 mutations in the context of PI resistance in HIV-1 subtype C, or the coevolution of all three genes (*env, gag* and *PR*) during the development of PI resistance. Therefore, in this study we identified mutations in HIV-1 subtype C Env that were related to PI therapy failure and suggested pathways to PI resistance that involved Env and Gag mutations, including HIV-1 subtype C signature polymorphisms in PR, and highlighted the effects of these mutations on the structure of Env.

To identify mutations associated with PI treatment failure, Env sequences generated from subtype C infected patients failing an LPV/r inclusive treatment were compared to downloaded HIV-1 subtype C drug-naïve sequences from the Los Alamos HIV database. Mutations in gp120 were seen in the variable loops V2 (D167AR, V182I, P183QS, Y191F and N195H) and V3 (S306R, P313DI, Q315R, I323KNR, A329DT, and C331RS) and constant regions C2 (N230DT, Q258EP, N262LI, T283I, V286I and I294N) and C3 (W338EG, H374KR, F382V, and F383LV). Mutations in V1/V2 as well as C2 and C3 have previously been associated with increased viral infectivity and replication capacity (Yokoyama et al., 2016, Dang et al., 2020). It is therefore possible that the Env mutations identified in the context of PI resistance, may serve a similar function. gp41 resistance-associated mutations were found throughout the region, including FPPR (S534A and T536A/M), HR (Q567K/R and D632E), TM (I688V), and CT (P724Q/S). Other studies have reported mutations in the CT, although not at the same amino acid position (Coetzer et al., 2017, Castain et al., 2019, Perrier et al., 2019). These differences could

be related to differences in subtype, since Coetzer et al (2017) reported the 721I N-terminal CT mutation in subtype A, while other studies from subtype B and CRF02_AG reported I781T (I270T) and V832 (V321I) found in the C-terminal of the CT (Castain et al., 2019, Perrier et al., 2019). This is the first report of mutations in the FPPR and TM regions associated with PI failure.

Using Bayesian network learning, our study also assessed the possible coevolution of the Gag-PR and Env regions during the development of PI resistance. Two potential pathways to LPV/r resistance were found. The first one involved a combination of Env and Gag mutations (I256V, S111C and R76K), where I256V in Gag was shown to be directly associated with three Env mutations (S110N-Env, T138S-Env, and S534A-Env), resulting in a pathway (Env mutation + I256V + S111C + R76K) that lead to LPV/r resistance. In addition, Q315R-Env at the V3 loop tip, formed the same pathway with the Gag mutations I256V, S111C, and R76K, with the addition of S111I which formed an indirect association between I256V and Q315R. Gp41 mutations (S534A and P724S) were also associated with the I256V Gag mutation. These data suggest that Env mutations are not directly associated with LPV/r failure; instead, they are indirectly associated through Gag (MA & CA) mutations. These Gag mutations have been previously associated with PI resistance in subtype C patients, although not specifically in patients without mutations in the PR region (Ntale, 2012, Singh, 2015). The second potential pathway to LPV/r resistance involved mutations in Env and Gag including the V77I minor mutation in the PR, which has been previously associated with PI drug resistance in both subtype B and C (Mata-Munguía et al., 2014, Doualla-Bell et al., 2006), and two signature subtype C mutations, S12T and P63L. S12T has been shown to be selected in LPV/r- and RTV-treated patients in subtype C (Giandhari et al., 2016). Again, although Env mutations were not directly associated with LPV/r resistance in these pathways, it is possible that they may exert pressure through the Gag and minor PR mutations.

To get a clear understanding of the implications of the Env mutations on the structure and function of Env, gp120 and gp41 were investigated separately as each glycoprotein has its own unique function. In chapter 3, the effects of mutations in gp120 were described with an emphasis on the effects on viral entry and evasion of the immune system. Taken together, the increased number of N-glycosylation sites in the V1/V2 region, increased sequence variation in V1/V2 and V5, as well as the increased V1/V2 loop length in the treated group vs naïve group indicate that these mutations could potentially protect the virus against neutralization by CD4-binding antibodies (Pollakis et al., 2001, Wang et al., 2013, van Gils et al., 2011). In addition P183S/Q and N195H in V2 both occurred at

coreceptor-specific sites, and could play a crucial role in coreceptor usage (Thielen et al., 2010). In fact, 38% (9/24) of the treated sequences were predicted to be CXCR4 viruses, and of those with N195H (6/9) four were predicted to be CXCR4 viruses. Similarly, those with the V3-Q315R mutation found at the tip of the V3 loop were also predicted to be CXCR4 coreceptor using viruses. The evidence for a coreceptor switch in the treated group was strengthened by the significantly higher charge in V3 (p-value=0.015) in the treated group vs naïve group. CXCR4 usage mediated by a higher V3 charge was previously reported in clade C viruses (Pollakis et al., 2001).

A closer look at the effects of these mutations on the Env structure, particularly in the V1/V2 regions (V1: T138S and V2: N195H) showed a change to the opening of the coreceptor binding site. Interestingly, T138S was positively selected in sequences that do not have PI mutations in the PR region. Of note, the CXCR4 predicted viruses (that also harboured the N195H mutation) showed structural alterations in the GPGR tip, which could impact the β-turn and co-receptor engagement at that site (Pastore et al., 2006). When taken together, these results suggest that gp120 mutations modulate viral entry activity by (i) Protecting the HIV-1 from antibody recognition and leading to immune escape, (ii) coreceptor switching, induced by mutations at coreceptor-specific sites and mutations interacting with the coreceptor binding sites.

Mutations in gp41 have previously been associated with PI resistance (Coetzer et al., 2017, Castain et al., 2019, Perrier et al., 2019), and have also been linked to mutations in the MA of Gag. This is not surprising, as the CT of gp41 interacts with the MA in the formation of the Env during budding (Freed and Martin, 1995). Therefore, the impact of Gag and gp41 co-evolving mutations on their respective structures was further explored. Of note, mutations T536A in the FPPR and P724Q/S in the CT were found most frequently in our sequences (22% and 31% respectively), with 13% of sequences harbouring the P724Q mutation alone; and T536A and P724S (9%) often occurred together. Furthermore, P724S was shown in the BN to be associated with treatment via Gag specific mutations and were also positively selected in sequences without PI mutations in the PR region. Interestingly, P724Q/S occurs within the KS and the hydrophobic loop of the KS has been associated with antibody binding (Freed and Martin, 1995, Dorfman et al., 1994, Cosson, 1996, Santos da Silva et al., 2013, Cleveland et al., 2003, Alfadhli et al., 2019, Murphy et al., 2017). Therefore, the P724Q/S mutations and their interactions with other KS residues (L721A/T, PS727 and D728) may impair antibody binding and recognition and could lead to increased viral replication.

Of note, S534A mutants formed hydrogen bonds with residue M530 which is the key residue that forms part of the gp120-gp41 interface, and prefusion state (Kumar et al., 2019), and L602, which is in the fusion domain, and disulphide loop site of non-covalent interactions with gp120 (Jacobs et al., 2005). This suggests that this mutation might prevent the cleavage of gp160, hence preventing the fusion process from occurring. It is possible that, in order to make up for the loss of viral fitness caused by this mutation, these viruses may choose cell-to-cell contact as their preferred means of transmission. According to previous reports, when alterations in the gp120-gp41 interface occurs, the virus may alternatively spread via the cell-cell route of transmission (Zhong et al., 2013, Pedro et al., 2019, Hikichi et al., 2021).

Codoner et al (2017), reported the coevolution of MA/CA and PR residues in the development of PI resistance. These MA/CA residues were demonstrated in the helices to either stabilizes or contribute to flexibility of the structure (Codoñer et al., 2017). This study looked at the structural changes that occurred in the MA and CA mutations that were shown to coevolve with Env mutations following PI failure without PR mutations. Interestingly, the two Gag MA mutations (R76K and S111C) that co-evolved with the CA mutation I256V and Env mutations, both caused the loss of a hydrogen bond, potentially increasing flexibility within the α-helix 4 and 5 respectively. Parry et al (2011) hypothesized that the loss of hydrogen bonds and increased flexibility, increased the affinity or availability of the MA-CA cleavage site with respect to the protease (Parry et al., 2011).

MA mutation Q69K found in the bend (residues: R L68, T70, G71 and T72) displayed a longer side chain compared to the Q69 wildtype, suggesting that it occupies a larger volume (Khan and Vihinen, 2007). It is unclear, though, how this mutation inhabiting a larger volume would affect the processing. The CA mutation T239S could contribute to the stability of the MA-CA structure by forming a hydrogen bond with Q117 in MA. Another CA mutation, I256V, retained the hydrogen bond present in the WT; however, the distance between the WT residues (I256 and W249) was shorter. This shorter distance could contribute to compact α-helix within the CA.

Conclusion

This is the first study that reports and proposes that gp120 indirectly facilitates PI resistance through acquiring mutations that increase the length of V1/V2, consequently increasing the overall number of N-glycosylation sites, as well as increasing the overall charge in V3. In this way, these mutations facilitate viral entry through immune escape, and drive the switch to CXCR4 viruses which could

possibly enhance viral replication. Moreover, it has been demonstrated that the gp41 mutations reported here affect the flexibility of the protein. The S534A mutation have an effect on fusion by interacting with the gp120-gp41 interface, while mutations in the Kennedy epitope are more related to immune escape. Overall, Env mutations contribute to PI drug resistance through coevolution with mutations in Gag which are known to contribute to PI failure and compensate for the loss of viral fitness, suggesting an interplay between immune escape and the development of drug resistance. This study will give a theoretical understanding regarding the role played by regions outside of protease (gag and env) in the development of protease inhibitor resistance especially in subtype C. Therefore, suggests that region outside protease should be included in drug resistance testing, certainly in the interpretation of drug resistance profiles.

Limitations and future work

Limitations

There are several limitations in this study, including the small sample size. In addition, this study did not conduct functional assays such as replication capacity and drug susceptibility assays that confirm the effect of Env mutations alone and in combination with Gag mutations on viral fitness and PI efficacy. Furthermore, the sequences generated for this study only display the N-terminal portion of the CT in gp41, which could mean that the C-terminal domain of CT mutations that could contribute to PI failure are not included in this study. Lastly, this study reports the effect of mutations that were reported to co-evolve with gag and PR. The mutations that were significantly high in treated and did not show any co-evolution were not further investigated, and therefore not represented. Although these mutations did not coevolve with Gag-PR, their significant increase indicates the possibility to contribute to the effects of replication capacity in their respective genes.

Future work:

1. To perform replication capacity and drug susceptibility assays on these isolates. This will be achieved by cloning amplified env and Gag genes isolate from PI-treated viruses into a plasmid backbone and investigating the effects of these on replication capacity and susceptibility to PI (LPV).
2. Investigations using site-directed mutagenesis to determine whether Env mutations alone can affect LPV/r efficacy. The 5 mutations identified will be inserted into a pnl4.3 backbone and replication capacity and drug susceptibility assays will be performed. Combinations of these mutations with mutations in PR and Gag will also be investigated in PR and Gag will also be investigated to determine the relationship between mutations in env, Gag and PR.

3. Invitro study to gain insight into how mutations reported in chapter 3 facilitate viral entry through immune escape and coreceptor switching. Studying the mechanism of entry in the presence of these mutations will help us understand if the virus fuses at the plasma membrane or spreads within the host via cell-to-cell contact.
4. Understanding the impact of FPPR mutations in gp120-gp41 association during PI resistance. The FPPR mutations will be modelled in the same structure to clearly understand their impact on the association of gp120 with gp41.
5. Investigate the effect of Gag and Env mutations on HIV-1 assembly and budding in PI failures, and this will aid in understanding the mechanism of coevolution in Env and Gag through PI resistance.

References

COETZER, M., LEDINGHAM, L., DIERO, L., KEMBOI, E., ORIDO, M. & KANTOR, R. 2017. Gp41 and Gag amino acids linked to HIV-1 protease inhibitor-based second-line failure in HIV-1 subtype A from Western Kenya. *Journal of the International AIDS Society,* 20**,** e25024.

Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, et al. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. Engl J Med. 2012;366:1275–86.

A PANTOPHLET, R. 2010. Antibody epitope exposure and neutralization of HIV-1. *Current pharmaceutical design,* 16**,** 3729-3743.

A YI, H., C FOCHTMAN, B., C RIZZO, R. & JACOBS, A. 2016. Inhibition of HIV entry by targeting the envelope transmembrane subunit gp41. *Current HIV research,* 14**,** 283-294.

ABBADESSA, D., SMURTHWAITE, C. A., REED, C. W. & WOLKOWICZ, R. 2015. A Single-Cell Platform for Monitoring Viral Proteolytic Cleavage in Different Cellular Compartments. *Biochemistry Insights,* 8**,** BCI. S30379.

ABELA, I. A., BERLINGER, L., SCHANZ, M., REYNELL, L., GÜNTHARD, H. F., RUSERT, P. & TRKOLA, A. 2012. Cell-cell transmission enables HIV-1 to evade inhibition by potent CD4bs directed antibodies. *PLoS pathogens,* 8**,** e1002634.

ALEN, M. M., DALLMEIER, K., BALZARINI, J., NEYTS, J. & SCHOLS, D. 2012. Crucial role of the N-glycans on the viral E-envelope glycoprotein in DC-SIGN-mediated dengue virus infection. *Antiviral research,* 96**,** 280-287.

ALFADHLI, A., MACK, A., RITCHIE, C., CYLINDER, I., HARPER, L., TEDBURY, P. R., FREED, E. O. & BARKLIS, E. 2016. Trimer Enhancement Mutation Effects on HIV-1 Matrix Protein Binding Activities. *Journal of virology,* 90**,** 5657-5664.

ALFADHLI, A., STAUBUS, A. O., TEDBURY, P. R., NOVIKOVA, M., FREED, E. O. & BARKLIS, E. 2019. Analysis of HIV-1 matrix-envelope cytoplasmic tail interactions. *Journal of virology,* 93**,** e01079-19.

ALLINDER, S. M. & FLEISCHMAN, J. 2019. The world's largest HIV epidemic in crisis: HIV in South Africa. *Center for Strategic and International Studies, April,* 2**,** 2019.

ANDREWS, S. M. & ROWLAND-JONES, S. 2017. Recent advances in understanding HIV evolution. *F1000Research,* 6.

ANISIMOVA, M., BIELAWSKI, J. P. & YANG, Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular biology and evolution,* 18**,** 1585-1592.

ARIF, M. S., HUNTER, J., LEDA, A. R., ZUKUROV, J. P. L., SAMER, S., CAMARGO, M., GALINSKAS, J., KALLAS, E. G., KOMNINAKIS, S. V. & JANINI, L. M. 2017. Pace of coreceptor tropism switch in HIV-1-infected individuals after recent infection. *Journal of virology,* 91**,** e00793-17.

ARTS, E. J. & HAZUDA, D. J. 2012. HIV-1 antiretroviral drug therapy. *Cold Spring Harbor perspectives in medicine,* 2**,** a007161.

BÄR, S. & ALIZON, M. 2004. Role of the ectodomain of the gp41 transmembrane envelope protein of human immunodeficiency virus type 1 in late steps of the membrane fusion process. *Journal of virology,* 78**,** 811-820.

BARRÉ-SINOUSSI, F., ROSS, A. L. & DELFRAISSY, J.-F. 2013. Past, present and future: 30 years of HIV research. *Nature Reviews Microbiology,* 11**,** 877-883.

BAXTER, J., CHASANOV, W. & ADAMS, J. 2016. An update on HIV-1 protease inhibitor resistance. *Journal of AIDS and Clinical Research,* 7.

BENUREAU, Y., COLIN, P., STAROPOLI, I., GONZALEZ, N., GARCIA-PEREZ, J., ALCAMI, J., ARENZANA-SEISDEDOS, F. & LAGANE, B. 2016. Guidelines for cloning, expression, purification and functional characterization of primary HIV-1 envelope glycoproteins. *Journal of Virological Methods,* 236**,** 184-195.

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

BLOOD, G. A. C. 2016. Human immunodeficiency virus (HIV). *Transfusion Medicine and Hemotherapy,* 43**,** 203.

BORMAN, A. M., PAULOUS, S. & 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. *Journal of General Virology,* 77**,** 419-426.

BOSCH, V. & PAWLITA, M. 1990. Mutational analysis of the human immunodeficiency virus type 1 env gene product proteolytic cleavage site. *Journal of virology,* 64**,** 2337-2344.

BRIZ, V., POVEDA, E. & SORIANO, V. 2006. HIV entry inhibitors: mechanisms of action and resistance pathways. *Journal of Antimicrobial Chemotherapy,* 57**,** 619-627.

CAFFREY, M. 2001. Model for the structure of the HIV gp41 ectodomain: insight into the intermolecular interactions of the gp41 loop. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease,* 1536**,** 116-122.

CAO, J., BERGERON, L., HELSETH, E., THALI, M., REPKE, H. & SODROSKI, J. 1993. Effects of amino acid changes in the extracellular domain of the human immunodeficiency virus type 1 gp41 envelope glycoprotein. *Journal of virology,* 67**,** 2747-2755.

CASTAIN, L., PERRIER, M., CHARPENTIER, C., PALICH, R., DESIRE, N., WIRDEN, M., DESCAMPS, D., SAYON, S., LANDMAN, R. & VALANTIN, M.-A. 2019. New mechanisms of resistance in virological failure to protease inhibitors: selection of non-described protease, Gag and Gp41 mutations. *Journal of Antimicrobial Chemotherapy,* 74**,** 2019-2023.

CAVROIS, M., NEIDLEMAN, J., SANTIAGO, M. L., DERDEYN, C. A., HUNTER, E. & GREENE, W. C. 2014. Enhanced fusion and virion incorporation for HIV-1 subtype C envelope glycoproteins with compact V1/V2 domains. *Journal of virology,* 88**,** 2083-2094.

CHAILLON, A., BRAIBANT, M., MOREAU, T., THENIN, S., MOREAU, A., AUTRAN, B. & BARIN, F. 2011. The V1V2 domain and an N-linked glycosylation site in the V3 loop of the HIV-1 envelope glycoprotein modulate neutralization sensitivity to the human broadly neutralizing antibody 2G12. *Journal of virology,* 85**,** 3642-3648.

CHAN, D. C., CHUTKOWSKI, C. T. & KIM, P. S. 1998. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proceedings of the National Academy of Sciences,* 95**,** 15613-15617.

CHECKLEY, M. A., LUTTGE, B. G. & FREED, E. O. 2011. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *Journal of molecular biology,* 410**,** 582-608.

CHIMUKANGARA, B., LESSELLS, R. J., SARTORIUS, B., GOUNDER, L., MANYANA, S., PILLAY, M., SINGH, L., GIANDHARI, J., GOVENDER, K. & SAMUEL, R. 2022. HIV-1 drug resistance in adults and adolescents on protease inhibitor-based antiretroviral therapy in KwaZulu-Natal Province, South Africa. *Journal of Global Antimicrobial Resistance,* 29**,** 468-475.

CLAVEL, F. & MAMMANO, F. 2010. Role of Gag in HIV resistance to protease inhibitors. *Viruses,* 2**,** 1411-1426.

CLEVELAND, S. M., BURATTI, E., JONES, T. D., NORTH, P., BARALLE, F., MCLAIN, L., MCINERNEY, T., DURRANI, Z. & DIMMOCK, N. J. 2000. Immunogenic and antigenic dominance of a nonneutralizing epitope over a highly conserved neutralizing epitope in the gp41 envelope glycoprotein of human immunodeficiency virus type 1: its deletion leads to a strong neutralizing response. *Virology,* 266**,** 66-78.

CLEVELAND, S. M., MCLAIN, L., CHEUNG, L., JONES, T. D., HOLLIER, M. & DIMMOCK, N. J. 2003. A region of the C-terminal tail of the gp41 envelope glycoprotein of human immunodeficiency virus type 1 contains a neutralizing epitope: evidence for its exposure on the surface of the virion. *Journal of general virology,* 84**,** 591-602.

CODOÑER, F. M., PEÑA, R., BLANCH-LOMBARTE, O., JIMENEZ-MOYANO, E., PINO, M., VOLLBRECHT, T., CLOTET, B., MARTINEZ-PICADO, J., DRAENERT, R. & PRADO, J. G. 2017. Gag-protease coevolution analyses define novel structural surfaces in the HIV-1 matrix and capsid involved in resistance to Protease Inhibitors. *Scientific reports,* 7**,** 1-10.

COETZER, M., LEDINGHAM, L., DIERO, L., KEMBOI, E., ORIDO, M. & KANTOR, R. 2017. Gp41 and Gag amino acids linked to HIV-1 protease inhibitor-based second-line failure in HIV-1 subtype A from Western Kenya. *Journal of the international AIDS Society,* 20**,** e25024.

COETZER, M., NEDELLEC, R., CILLIERS, T., MEYERS, T., MORRIS, L. & MOSIER, D. E. 2011. Extreme genetic divergence is required for coreceptor switching in HIV-1 subtype C. *Journal of acquired immune deficiency syndromes (1999),* 56**,** 9.

COETZER, M., NEDELLEC, R., SALKOWITZ, J., MCLAUGHLIN, S., LIU, Y., HEATH, L., MULLINS, J. I. & MOSIER, D. E. 2008. Evolution of CCR5 use before and during coreceptor switching. *Journal of virology,* 82**,** 11758-11766.

COSSON, P. 1996. Direct interaction between the envelope and matrix proteins of HIV-1. *The EMBO journal,* 15**,** 5783.

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

CURLIN, M. E., ZIONI, R., HAWES, S. E., LIU, Y., DENG, W., GOTTLIEB, G. S., ZHU, T. & MULLINS, J. I. 2010. HIV-1 envelope subregion length variation during disease progression. *PLoS Pathog,* 6**,** e1001228.

DAM, E., QUERCIA, R., GLASS, B., DESCAMPS, D., LAUNAY, O., DUVAL, X., KRÄUSSLICH, H.-G., HANCE, A. J., CLAVEL, F. & GROUP, A. S. 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.

DANG, L. V. P., PHAM, H. V., DINH, T. T., NGUYEN, T. H., VU, Q. T. H., VU, N. T. P., LE, P. T. B., NGUYEN, L. V., LE, H. T. & VU, P. T. 2020. Characterization of envelope sequence of HIV virus in children infected with HIV in Vietnam. *SAGE Open Medicine,* 8**,** 2050312120937198.

DAS, K., BANDWAR, R. P., WHITE, K. L., FENG, J. Y., SARAFIANOS, S. G., TUSKE, S., TU, X., CLARK, A. D., BOYER, P. L. & HOU, X. 2009. Structural basis for the role of the K65R mutation in HIV-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance. *Journal of Biological Chemistry,* 284**,** 35092-35100.

DEFORCHE, K., CAMACHO, R., GROSSMAN, Z., SILANDER, T., SOARES, M., MOREAU, Y., SHAFER, R. W., VAN LAETHEM, K., CARVALHO, A. & WYNHOVEN, B. 2007. Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. *Infection, genetics and evolution,* 7**,** 382-390.

DEFORCHE, K., SILANDER, T., CAMACHO, R., GROSSMAN, Z., SOARES, M., VAN LAETHEM, K., KANTOR, R., MOREAU, Y., VANDAMME, A.-M. & WORKGROUP, N.-B. 2006. Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. *Bioinformatics,* 22**,** 2975-2979.

DELOBEL, P., SANDRES-SAUNÉ, K., CAZABAT, M., PASQUIER, C., MARCHOU, B., MASSIP, P. & IZOPET, J. 2005. R5 to X4 switch of the predominant HIV-1 population in cellular reservoirs during effective highly active antiretroviral therapy. *JAIDS Journal of Acquired Immune Deficiency Syndromes,* 38**,** 382-392.

DOORES, K. J. 2015. The HIV glycan shield as a target for broadly neutralizing antibodies. *The FEBS journal,* 282**,** 4679-4691.

DORFMAN, T., MAMMANO, F., HASELTINE, W. A. & GÖTTLINGER, H. 1994. Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. *Journal of Virology,* 68**,** 1689-1696.

DOUALLA-BELL, F., AVALOS, A., GAOLATHE, T., MINE, M., GASEITSIWE, S., NDWAPI, N., NOVITSKY, V. A., BRENNER, B., OLIVEIRA, M. & MOISI, D. 2006. Impact of human immunodeficiency virus type 1 subtype C on drug resistance mutations in patients from Botswana failing a nelfinavir-containing regimen. *Antimicrobial Agents and Chemotherapy,* 50**,** 2210-2213.

DUFLOO, J., BRUEL, T. & SCHWARTZ, O. 2018. HIV-1 cell-to-cell transmission and broadly neutralizing antibodies. *Retrovirology,* 15**,** 1-14.

DUTTA, D., MANDAL, C. & MANDAL, C. 2017. Unusual glycosylation of proteins: Beyond the universal sequon and other amino acids. *Biochimica et Biophysica Acta (BBA)-General Subjects,* 1861**,** 3096-3108.

ECHE, S., KUMAR, A., SONELA, N. & GORDON, M. L. 2021. Acquired HIV-1 Protease Conformational Flexibility Associated with Lopinavir Failure May Shape the Outcome of Darunavir Therapy after Antiretroviral Therapy Switch. *Biomolecules,* 11**,** 489.

ECKERT, D. M. & KIM, P. S. 2001. Mechanisms of viral membrane fusion and its inhibition. *Annual review of biochemistry,* 70**,** 777-810.

EDESSA, D., SISAY, M. & ASEFA, F. 2019. Second-line HIV treatment failure in sub-Saharan Africa: A systematic review and meta-analysis. *PLoS One,* 14**,** e0220159.

ENGELMAN, A. & CHEREPANOV, P. 2012. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature Reviews Microbiology,* 10**,** 279-290.

FARES, M. A. & TRAVERS, S. A. 2006. A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics,* 173**,** 9-23.

FEUCHT, U. D., ROSSOUW, T., VAN DYK, G., FORSYTH, B. & KRUGER, M. 2014. Consequences of prior use of full-dose ritonavir as single protease inhibitor as part of combination antiretroviral regimens on the future therapy choices in HIV-1–infected children. *The Pediatric Infectious Disease Journal,* 33**,** e53-e59.

FLEXNER, C. 1998. HIV-protease inhibitors. *New England Journal of Medicine,* 338**,** 1281-1293.

FLYNN, J. K., GARTNER, M., LAUMAEA, A. & GORRY, P. R. 2019. Futuristic Methods for Determining HIV Co-receptor Use. *Global Virology III: Virology in the 21st Century***,** 625-663.

FRANKEL, A. D. & YOUNG, J. A. 1998. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem,* 67.

FREED, E., MYERS, D. & RISSER, R. 1991. Identification of the principal neutralizing determinant of human immunodeficiency virus type 1 as a fusion domain. *Journal of virology,* 65**,** 190-194.

FREED, E. O. 1998. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology,* 251**,** 1-15.

FREED, E. O. 2001. HIV-1 replication. *Somatic cell and molecular genetics,* 26**,** 13-33.

FREED, E. O. & MARTIN, M. A. 1995. Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. *Journal of virology,* 69**,** 1984-1989.

FREED, E. O., MYERS, D. J. & RISSER, R. 1990. Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. *Proceedings of the National Academy of Sciences,* 87**,** 4650-4654.

FUN, A., WENSING, A. M., VERHEYEN, J. & NIJHUIS, M. 2012. Human Immunodeficiency Virus Gag and protease: partners in resistance. *Retrovirology,* 9**,** 63.

GABUZDA, D., OLSHEVSKY, U., BERTANI, P., HASELTINE, W. A. & SODROSKI, J. 1991. Identification of membrane anchorage domains of the HIV-1 gp160 envelope glycoprotein precursor. *JAIDS Journal of Acquired Immune Deficiency Syndromes,* 4**,** 34-40.

GAJULA, M. P., KUMAR, A. & IJAQ, J. 2016. Protocol for molecular dynamics simulations of proteins. *Bio-protocol,* 6**,** e2051-e2051.

GALLEGO, O., DE MENDOZA, C., PÉREZ-ELÍAS, M. J., GUARDIOLA, J. M., PEDREIRA, J., DALMAU, D., GÓNZALEZ, J., MORENO, A., ARRIBAS, J. R. & RUBIO, A. 2001. Drug resistance in patients experiencing early virological failure under a triple combination including indinavir. *Aids,* 15**,** 1701-1706.

GALLO, R. C. & MONTAGNIER, L. 2003. The discovery of HIV as the cause of AIDS. *New England Journal of Medicine,* 349**,** 2283-2285.

GARCIA-GARCIA, T., PONCET, S., DEROUICHE, A., SHI, L., MIJAKOVIC, I. & NOIROT-GROS, M.-F. 2016. Role of protein phosphorylation in the regulation of cell cycle and DNA-related processes in bacteria. *Frontiers in Microbiology,* 7**,** 184.

GATANAGA, H., SUZUKI, Y., TSANG, H., YOSHIMURA, K., KAVLICK, M. F., NAGASHIMA, K., GORELICK, R. J., MARDY, S., TANG, C. & SUMMERS, M. F. 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. *Journal of Biological Chemistry,* 277**,** 5952-5961.
GHAFFARI, G., TUTTLE, D. L., BRIGGS, D., BURKHARDT, B. R., BHATT, D., ANDIMAN, W. A., SLEASMAN, J. W. & GOODENOW, M. M. 2005. Complex determinants in human immunodeficiency virus type 1 envelope gp120 mediate CXCR4-dependent infection of macrophages. *Journal of virology,* 79**,** 13250-13261.
GHOSH, A. K., CHAPSAL, B. D., WEBER, I. T. & MITSUYA, H. 2008. Design of HIV protease inhibitors targeting protein backbone: an effective strategy for combating drug resistance. *Accounts of chemical research,* 41**,** 78-86.
GIANDHARI, J., BASSON, A. E., SUTHERLAND, K., PARRY, C. M., CANE, P. A., COOVADIA, A., KUHN, L., HUNT, G. & MORRIS, L. 2016. Contribution of Gag and protease to HIV-1 phenotypic drug resistance in pediatric patients failing protease inhibitor-based therapy. *Antimicrobial agents and chemotherapy,* 60**,** 2248-2256.
GIOVANETTI, M., CICCOZZI, M., PAROLIN, C. & BORSETTI, A. 2020. Molecular epidemiology of HIV-1 in African countries: a comprehensive overview. *Pathogens,* 9**,** 1072.
GIROUD, C., CHAZAL, N. & BRIANT, L. 2011. Cellular kinases incorporated into HIV-1 particles: passive or active passengers? *Retrovirology,* 8**,** 1-14.
GOTTLIEB, M. S., SCHROFF, R., SCHANKER, H. M., WEISMAN, J. D., FAN, P. T., WOLF, R. A. & SAXON, A. 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *New England Journal of Medicine,* 305**,** 1425-1431.
GULNIK, S. V., SUVOROV, L. I., LIU, B., YU, B., ANDERSON, B., MITSUYA, H. & ERICKSON, J. W. 1995. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry,* 34**,** 9282-9287.
GUO, D., SHI, X., ARLEDGE, K., SONG, D., JIANG, L., FU, L., ZHANG, S., WANG, X. & ZHANG, L. 2012. V5 region in the HIV-1 envelope glycoprotein determines viral sensitivity to the broadly neutralizing monoclonal antibody VRC01. *Retrovirology,* 9**,** 1-1.
GUPTA, R. K., KOHLI, A., MCCORMICK, A. L., TOWERS, G. J., PILLAY, D. & PARRY, C. M. 2010. Full length HIV-1 gag determines protease inhibitor susceptibility within in vitro assays. *AIDS (London, England),* 24**,** 1651.
HAAS, J., BARBATO, A., BEHRINGER, D., STUDER, G., ROTH, S., BERTONI, M., MOSTAGUIR, K., GUMIENNY, R. & SCHWEDE, T. 2018. Continuous Automated Model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12. *Proteins: Structure, Function, and Bioinformatics,* 86**,** 387-398.
HE, Y., LIU, S., LI, J., LU, H., QI, Z., LIU, Z., DEBNATH, A. K. & JIANG, S. 2008. Conserved salt bridge between the N-and C-terminal heptad repeat regions of the human immunodeficiency virus type 1 gp41 core structure is critical for virus entry and inhibition. *Journal of virology,* 82**,** 11129-11139.
HECKERMAN, D., GEIGER, D. & CHICKERING, D. M. 1995. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning,* 20**,** 197-243.
HEDT, B. L., LAUFER, M. K. & COHEN, T. 2011. Drug resistance surveillance in resource-poor settings: current methods and considerations for TB, HIV, and malaria. *The American journal of tropical medicine and hygiene,* 84**,** 192.
HEMELAAR, J. 2012. The origin and diversity of the HIV-1 pandemic. *Trends Mol Med,* 18.
HIKICHI, Y., VAN DUYNE, R., PHAM, P., GROEBNER, J. L., WIEGAND, A., MELLORS, J. W., KEARNEY, M. F. & FREED, E. O. 2021. Mechanistic Analysis of the Broad Antiretroviral Resistance Conferred by HIV-1 Envelope Glycoprotein Mutations. *Mbio,* 12**,** e03134-20.
HOLLINGSWORTH IV, L. R., LEMKUL, J. A., BEVAN, D. R. & BROWN, A. M. 2018. HIV-1 env gp41 transmembrane domain dynamics are modulated by lipid, water, and ion interactions. *Biophysical journal,* 115**,** 84-94.

HOLLINGSWORTH, S. A. & DROR, R. O. 2018. Molecular dynamics simulation for all. *Neuron,* 99**,** 1129-1143.
HU, W.-S. & HUGHES, S. H. 2012. HIV-1 reverse transcription. *Cold Spring Harbor perspectives in medicine,* 2**,** a006882.
HUANG, H., CHOPRA, R., VERDINE, G. L. & HARRISON, S. C. 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science,* 282**,** 1669-1675.
HUBBARD, R. E. & HAIDER, M. K. 2010. Hydrogen bonds in proteins: role and strength. *eLS*.
HYSER, J. M., ZENG, C. Q.-Y., BEHARRY, Z., PALZKILL, T. & ESTES, M. K. 2008. Epitope mapping and use of epitope-specific antisera to characterize the VP5∗ binding site in rotavirus SA11 NSP4. *Virology,* 373**,** 211-228.
JACOBS, A., SEN, J., RONG, L. & CAFFREY, M. 2005. Alanine scanning mutants of the HIV gp41 loop. *Journal of Biological Chemistry,* 280**,** 27284-27288.
JAKOBSEN, M. R., CASHIN, K., ROCHE, M., STERJOVSKI, J., ELLETT, A., BORM, K., FLYNN, J., ERIKSTRUP, C., GOUILLOU, M. & GRAY, L. R. 2013. Longitudinal analysis of CCR5 and CXCR4 usage in a cohort of antiretroviral therapy-naive subjects with progressive HIV-1 subtype C infection. *PLoS One,* 8**,** e65950.
JIANG, S. & DEBNATH, A. K. 2000. A salt bridge between an N-terminal coiled coil of gp41 and an antiviral agent targeted to the gp41 core is important for anti-HIV-1 activity. *Biochemical and biophysical research communications,* 270**,** 153-157.
KAMISETTY, H., OVCHINNIKOV, S. & BAKER, D. 2013. Assessing the utility of coevolution-based residue–residue contact predictions in a sequence-and structure-rich era. *Proceedings of the National Academy of Sciences,* 110**,** 15674-15679.
KEHINDE, I., RAMHARACK, P., NLOOTO, M. & GORDON, M. 2019. The pharmacokinetic properties of HIV-1 protease inhibitors: A computational perspective on herbal phytochemicals. *Heliyon,* 5**,** e02565.
KELLY PILLAY, S., SINGH, U., SINGH, A., GORDON, M. & NDUNGU, T. 2014. Gag drug resistance mutations in HIV-1 subtype C patients, failing a protease inhibitor inclusive treatment regimen, with detectable lopinavir levels. *Journal of the International AIDS Society,* 17**,** 19784.
KHAN, S. & VIHINEN, M. 2007. Spectrum of disease-causing mutations in protein secondary structures. *BMC Structural Biology,* 7**,** 1-18.
KIM, R. & BAXTER, J. D. 2008. Protease inhibitor resistance update: where are we now? *AIDS patient care and STDs,* 22**,** 267-277.
KIRCHMAIR, J., DISTINTO, S., ROMAN LIEDL, K., MARKT, P., MARIA ROLLINGER, J., SCHUSTER, D., MARIA SPITZER, G. & WOLBER, G. 2011. Development of anti-viral agents using molecular modeling and virtual screening techniques. *Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders),* 11**,** 64-93.
KLETENKOV, K., HOFFMANN, D., BÖNI, J., YERLY, S., AUBERT, V., SCHÖNI-AFFOLTER, F., STRUCK, D., VERHEYEN, J., KLIMKAIT, T. & STUDY, S. H. C. 2017. Role of Gag mutations in PI resistance in the Swiss HIV cohort study: bystanders or contributors? *Journal of Antimicrobial Chemotherapy,* 72**,** 866-875.
KLINGLER, J., ANTON, H., RÉAL, E., ZEIGER, M., MOOG, C., MÉLY, Y. & BOUTANT, E. 2020. How HIV-1 gag manipulates its host cell proteins: A focus on interactors of the nucleocapsid domain. *Viruses,* 12**,** 888.
KOLCHINSKY, P., KIPRILOV, E., BARTLEY, P., RUBINSTEIN, R. & SODROSKI, J. 2001. Loss of a single N-linked glycan allows CD4-independent human immunodeficiency virus type 1 infection by altering the position of the gp120 V1/V2 variable loops. *Journal of virology,* 75**,** 3435-3443.
KORKUT, A. & HENDRICKSON, W. A. 2012. Structural plasticity and conformational transitions of HIV envelope glycoprotein gp120. *PloS one,* 7**,** e52170.

KOTOKWE, K., MOYO, S., ZAHRALBAN-STEELE, M., HOLME, M. P., MELAMU, P., KOOFHETHILE, C. K., CHOGA, W. T., MOHAMMED, T., NKHISANG, T. & MOKALENG, B. 2023. Prediction of Coreceptor Tropism in HIV-1 Subtype C in Botswana. *Viruses,* 15**,** 403.

KOŽÍŠEK, M., HENKE, S., ŠAŠKOVÁ, K. G., JACOBS, G. B., SCHUCH, A., BUCHHOLZ, B., MÜLLER, V., KRÄUSSLICH, H.-G., ŘEZÁČOVÁ, P. & KONVALINKA, J. 2012. Mutations in HIV-1 gag and pol compensate for the loss of viral fitness caused by a highly mutated protease. *Antimicrobial agents and chemotherapy,* 56**,** 4320-4330.

KUMAR, S., SARKAR, A., PUGACH, P., SANDERS, R. W., MOORE, J. P., WARD, A. B. & WILSON, I. A. 2019. Capturing the inherent structural dynamics of the HIV-1 envelope glycoprotein fusion peptide. *Nature communications,* 10**,** 763.

KWON, Y. D., FINZI, A., WU, X., DOGO-ISONAGIE, C., LEE, L. K., MOORE, L. R., SCHMIDT, S. D., STUCKEY, J., YANG, Y. & ZHOU, T. 2012. Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. *Proceedings of the National Academy of Sciences,* 109**,** 5663-5668.

LAN, J., LI, W., YU, R., SYED, F. & YU, Q. 2022. Cell-to-cell transmission of HIV-1 from provirus-activated cells to resting naïve and memory human primary CD4 T cells is highly efficient and requires CD4 and F-actin but not chemokine receptors. *Journal of Medical Virology,* 94**,** 5434-5450.

LARROUY, L., CHAZALLON, C., LANDMAN, R., CAPITANT, C., PEYTAVIN, G., COLLIN, G., CHARPENTIER, C., STORTO, A., PIALOUX, G. & KATLAMA, C. 2010. Gag mutations can impact virological response to dual-boosted protease inhibitor combinations in antiretroviral-naive HIV-infected patients. *Antimicrobial agents and chemotherapy,* 54**,** 2910-2919.

LEONARD, C. K., SPELLMAN, M. W., RIDDLE, L., HARRIS, R. J., THOMAS, J. N. & GREGORY, T. 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *Journal of Biological Chemistry,* 265**,** 10373-10382.

LI, G., VERHEYEN, J., RHEE, S. Y., VOET, A., VANDAMME, A. M. & THEYS, K. 2013. Functional conservation of HIV-1 gag: implications for rational drug design. *Retrovirology,* 10.

LIU, H., SU, X., SI, L., LU, L. & JIANG, S. 2018. The development of HIV vaccines targeting gp41 membrane-proximal external region (MPER): challenges and prospects. *Protein & cell,* 9**,** 596-615.

LOBRITZ, M. A., RATCLIFF, A. N. & ARTS, E. J. 2010. HIV-1 entry, inhibitors, and resistance. *Viruses,* 2**,** 1069-1105.

LOLE, K. S., BOLLINGER, R. C., PARANJAPE, R. S., GADKARI, D., KULKARNI, S. S., NOVAK, N. G., INGERSOLL, R., SHEPPARD, H. W. & RAY, S. C. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of virology,* 73**,** 152-160.

LORI, F., SCOVASSI, A. I., ZELLA, D., ACHILLI, G., CATTANEO, E., CASOLI, C. & BERTAZZONI, U. 1988. Enzymatically active forms of reverse transcriptase of the human immunodeficiency virus. *AIDS research and human retroviruses,* 4**,** 393-398.

LU, W., CHEN, S., YU, J., BEHRENS, R., WIGGINS, J., SHERER, N., LIU, S.-L., XIONG, Y., XIANG, S.-H. & WU, L. 2019. The polar region of the HIV-1 envelope protein determines viral fusion and infectivity by stabilizing the gp120-gp41 association. *Journal of virology,* 93**,** e02128-18.

LU, W., LI, T.-W., PHILLIPS, S. & WU, L. 2021. Reverted HIV-1 Mutants in CD4+ T-Cells Reveal Critical Residues in the Polar Region of Viral Envelope Glycoprotein. *Microbiology Spectrum,* 9**,** e01653-21.

LYNCH, R. M., SHEN, T., GNANAKARAN, S. & DERDEYN, C. A. 2009. Appreciating HIV type 1 diversity: subtype differences in Env. *AIDS research and human retroviruses,* 25**,** 237-248.

MAKATINI, Z., MDA, S., TOWOBOLA, O., MTHIYANE, H., MILES, P. & BLACKARD, J. 2019. Characterisation of protease resistance mutations in a South African paediatric cohort with virological failure, 2011-2017. *South African Medical Journal,* 109**,** 511-515.

MALBEC, M., MOUQUET, H. & SCHWARTZ, O. 2014. HIV-1 cell-to-cell transmission and broadly neutralizing antibodies. *Medecine sciences: M/S,* 30**,** 508-510.

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

MANASA, J., VARGHESE, V., POND, S. L. K., RHEE, S.-Y., TZOU, P. L., FESSEL, W. J., JANG, K. S., WHITE, E., RÖGNVALDSSON, T. & KATZENSTEIN, D. A. 2017. Evolution of gag and gp41 in patients receiving ritonavir-boosted protease inhibitors. *Scientific reports,* 7**,** 1-11.

MAPHUMULO, N. F. & GORDON, M. L. 2021. Potential Associations of Mutations within the HIV-1 Env and Gag Genes Conferring Protease Inhibitor (PI) Drug Resistance. *Microbiology Research,* 12**,** 967-977.

MARIE, V. & GORDON, M. 2019. Gag-protease coevolution shapes the outcome of lopinavir-inclusive treatment regimens in chronically infected HIV-1 subtype C patients. *Bioinformatics,* 35**,** 3219-3223.

MARKOSYAN, R. M., COHEN, F. S. & MELIKYAN, G. B. 2003. HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation. *Molecular biology of the cell,* 14**,** 926-938.

MARKOWITZ, M., CONANT, M., HURLEY, A., SCHLUGER, R., DURAN, M., PETERKIN, J., CHAPMAN, S., PATICK, A., HENDRICKS, A. & YUEN, G. J. 1998. A preliminary evaluation of nelfinavir mesylate, an inhibitor of human immunodeficiency virus (HIV)-1 protease, to treat HIV infection. *Journal of Infectious Diseases,* 177**,** 1533-1540.

MARSDEN, M. D. & ZACK, J. A. 2013. HIV/aids eradication. *Bioorganic & medicinal chemistry letters,* 23**,** 4003-4010.

MATA-MUNGUÍA, C., ESCOTO-DELGADILLO, M., TORRES-MENDOZA, B., FLORES-SOTO, M., VÁZQUEZ-TORRES, M., GÁLVEZ-GASTELUM, F., VINIEGRA-OSORIO, A., CASTILLERO-MANZANO, M. & VÁZQUEZ-VALLS, E. 2014. Natural polymorphisms and unusual mutations in HIV-1 protease with potential antiretroviral resistance: a bioinformatic analysis. *BMC bioinformatics,* 15**,** 1-17.

MATUME, N. D., TEBIT, D. M. & BESSONG, P. O. 2020. HIV-1 subtype C predicted co-receptor tropism in Africa: an individual sequence level meta-analysis. *AIDS research and therapy,* 17**,** 1-16.

MAURER-STROH, S. & EISENHABER, F. 2004. Myristoylation of viral and bacterial proteins. *Trends in microbiology,* 12**,** 178-185.

MCCUNE, J. M., RABIN, L. B., FEINBERG, M. B., LIEBERMAN, M., KOSEK, J. C., REYES, G. R. & WEISSMAN, I. L. 1988. Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell,* 53**,** 55-67.

MCGOVERN, S. L., CASELLI, E., GRIGORIEFF, N. & SHOICHET, B. K. 2002. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Journal of medicinal chemistry,* 45**,** 1712-1722.

MEINTJES, G. & MAARTENS, G. 2012. Guidelines for antiretroviral therapy in adults. *Southern African Journal of HIV Medicine,* 13**,** 114-133.

MELIKYAN, G. B. 2008. Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm. *Retrovirology,* 5**,** 1-13.

MELIKYAN, G. B., MARKOSYAN, R. M., BRENER, S. A., ROZENBERG, Y. & COHEN, F. S. 2000. Role of the cytoplasmic tail of ecotropic Moloney murine leukemia virus Env protein in fusion pore formation. *Journal of virology,* 74**,** 447-455.

MILLER, M., SCHNEIDER, J., SATHYANARAYANA, B. K., TOTH, M. V., MARSHALL, G. R., CLAWSON, L., SELK, L., KENT, S. B. & WLODAWER, A. 1989. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science,* 246**,** 1149-52.

MIRJALILI, V. & FEIG, M. 2013. Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. *Journal of chemical theory and computation,* 9**,** 1294-1303.
MOORE, J. P., KITCHEN, S. G., PUGACH, P. & ZACK, J. A. 2004. The CCR5 and CXCR4 coreceptors—central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. *AIDS research and human retroviruses,* 20**,** 111-126.
MOORE, P. L., GRAY, E. S., CHOGE, I. A., RANCHOBE, N., MLISANA, K., ABDOOL KARIM, S. S., WILLIAMSON, C. & MORRIS, L. 2008. The c3-v4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection. *Journal of virology,* 82**,** 1860-1869.
MORTIER, V., DAUWE, K., VANCOILLIE, L., STAELENS, D., VAN WANZEELE, F., VOGELAERS, D., VANDEKERCKHOVE, L., CHALMET, K. & VERHOFSTEDE, C. 2013. Frequency and predictors of HIV-1 co-receptor switch in treatment naive patients. *PLoS One,* 8**,** e80259.
MUGWAGWA, T. & WITTEN, G. 2007. Coreceptor switching in HIV-1 subtype B and subtype C. *Bulletin of mathematical biology,* 69**,** 55.
MUHAMMED, M. T. & AKI-YALCIN, E. 2019. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical biology & drug design,* 93**,** 12-20.
MURAKAMI, T. 2008. Roles of the interactions between Env and Gag proteins in the HIV-1 replication cycle. *Microbiology and immunology,* 52**,** 287-295.
MURIAUX, D. & DARLIX, J.-L. 2010. Properties and functions of the nucleocapsid protein in virus assembly. *RNA biology,* 7**,** 744-753.
MURPHY, R. E. & SAAD, J. S. 2020. The interplay between HIV-1 gag binding to the plasma membrane and env incorporation. *Viruses,* 12**,** 548.
MURPHY, R. E., SAMAL, A. B., VLACH, J. & SAAD, J. S. 2017. Solution structure and membrane interaction of the cytoplasmic tail of HIV-1 gp41 protein. *Structure,* 25**,** 1708-1718. e5.
MYERS, G. & LENROOT, R. 1992. HIV glycosylation: what does it portend? *AIDS research and human retroviruses,* 8**,** 1459-1460.
MYLLYMÄKI, P., SILANDER, T., TIRRI, H. & URONEN, P. 2002. B-course: A web-based tool for Bayesian and causal data analysis. *International Journal on Artificial Intelligence Tools,* 11**,** 369-387.
NIJHUIS, M., SCHUURMAN, R., DE JONG, D., ERICKSON, J., GUSTCHINA, E., ALBERT, J., SCHIPPER, P., GULNIK, S. & BOUCHER, C. A. 1999. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *Aids,* 13**,** 2349-2359.
NIJHUIS, M., VAN MAARSEVEEN, N. M., LASTERE, S., SCHIPPER, P., COAKLEY, E., GLASS, B., ROVENSKA, M., DE JONG, D., CHAPPEY, C. & GOEDEGEBUURE, I. W. 2007. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med,* 4**,** e36.
NTALE, R. S. 2012. The role of early cytotoxic lymphocyte (CTL) escape in the pathogenesis of HIV-1 subtype C infection.
OBASA, A. E., AMBIKAN, A. T., GUPTA, S., NEOGI, U. & JACOBS, G. B. 2021. Increased acquired protease inhibitor drug resistance mutations in minor HIV-1 quasispecies from infected patients suspected of failing on national second-line therapy in South Africa. *BMC Infectious Diseases,* 21**,** 1-8.
OBASA, A. E., MIKASI, S. G., BRADO, D., CLOETE, R., SINGH, K., NEOGI, U. & JACOBS, G. B. 2020. Drug resistance mutations against protease, reverse transcriptase and integrase inhibitors in people living with HIV-1 receiving boosted protease inhibitors in South Africa. *Frontiers in microbiology,* 11**,** 438.
OGERT, R. A., LEE, M. K., ROSS, W., BUCKLER-WHITE, A., MARTIN, M. A. & CHO, M. W. 2001. N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. *Journal of virology,* 75**,** 5998-6006.

OLSHEVSKY, U., HELSETH, E., FURMAN, C., LI, J., HASELTINE, W. & SODROSKI, J. 1990. Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. *Journal of virology,* 64**,** 5701-5707.

PANTOPHLET, R. & BURTON, D. R. 2006. GP120: target for neutralizing HIV-1 antibodies. *Annu. Rev. Immunol.,* 24**,** 739-769.

PARRY, C. M., KOLLI, M., MYERS, R. E., CANE, P. A., SCHIFFER, C. & PILLAY, D. 2011. Three residues in HIV-1 matrix contribute to protease inhibitor susceptibility and replication capacity. *Antimicrobial agents and chemotherapy,* 55**,** 1106-1113.

PASQUATO, A., DETTIN, M., BASAK, A., GAMBARETTO, R., TONIN, L., SEIDAH, N. & DI BELLO, C. 2007. Heparin enhances the furin cleavage of HIV-1 gp160 peptides. *FEBS letters,* 581**,** 5807-5813.

PASTORE, C., NEDELLEC, R., RAMOS, A., PONTOW, S., RATNER, L. & MOSIER, D. 2006. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. *Journal of virology,* 80**,** 750-758.

PEARL, J. 1998. Graphical models for probabilistic and causal reasoning. *Quantified Representation of Uncertainty and Imprecision.* Springer.

PEDRO, K. D., HENDERSON, A. J. & AGOSTO, L. M. 2019. Mechanisms of HIV-1 cell-to-cell transmission and the establishment of the latent reservoir. *Virus research,* 265**,** 115-121.

PERRIER, M., CASTAIN, L., REGAD, L., TODESCO, E., LANDMAN, R., VISSEAUX, B., YAZDANPANAH, Y., RODRIGUEZ, C., JOLY, V. & CALVEZ, V. 2019. HIV-1 protease, Gag and gp41 baseline substitutions associated with virological response to a PI-based regimen. *Journal of Antimicrobial Chemotherapy,* 74**,** 1679-1692.

PERRY, C. M. 2014. Elvitegravir/cobicistat/emtricitabine/tenofovir disoproxil fumarate single-tablet regimen (Stribild(R)): a review of its use in the management of HIV-1 infection in adults. *Drugs,* 74.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M. & MENG, E. C. 2004. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem,* 25.

PETTIT, S. C., HENDERSON, G. J., SCHIFFER, C. A. & SWANSTROM, R. 2002. Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. *Journal of virology,* 76**,** 10226-10233.

PETTIT, S. C., MOODY, M. D., WEHBIE, R. S., KAPLAN, A. H., NANTERMET, P. V., KLEIN, C. A. & SWANSTROM, R. 1994. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *Journal of virology,* 68**,** 8017-8027.

POIGNARD, P., SAPHIRE, E. O., PARREN, P. W. & BURTON, D. R. 2001. gp120: Biologic aspects of structural features. *Annual review of immunology,* 19**,** 253-274.

POLLAKIS, G., ABEBE, A., KLIPHUIS, A., CHALABY, M. I., BAKKER, M., MENGISTU, Y., BROUWER, M., GOUDSMIT, J., SCHUITEMAKER, H. & PAXTON, W. A. 2004. Phenotypic and genotypic comparisons of CCR5-and CXCR4-tropic human immunodeficiency virus type 1 biological clones isolated from subtype C-infected individuals. *Journal of virology,* 78**,** 2841-2852.

POLLAKIS, G., KANG, S., KLIPHUIS, A., CHALABY, M. I., GOUDSMIT, J. & PAXTON, W. A. 2001. N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *Journal of Biological Chemistry,* 276**,** 13433-13441.

POSTLER, T. S. & DESROSIERS, R. C. 2013. The tale of the long tail: the cytoplasmic domain of HIV-1 gp41. *Journal of virology,* 87**,** 2-15.

POVEDA, E., PAREDES, R., MORENO, S., ALCAMÍ, J., CÓRDOBA, J., DELGADO, R., GUTIÉRREZ, F., LLIBRE, J. M., GARCIA DELTORO, M. & HERNANDEZ-QUERO, J. 2012. Update on clinical and methodological recommendations for genotypic determination of HIV tropism to guide the usage of CCR5 antagonists. *AIDS Rev,* 14**,** 208-217.

PRABAKARAN, P., DIMITROV, A. S., FOUTS, T. R. & DIMITROV, D. S. 2007. Structure and function of the HIV envelope glycoprotein as entry mediator, vaccine immunogen, and target for inhibitors. *Advances in pharmacology,* 55**,** 33-97.

PRADO, J. G., WRIN, T., BEAUCHAINE, J., RUIZ, L., PETROPOULOS, C. J., FROST, S. D., CLOTET, B., RICHARD, T. & MARTINEZ-PICADO, J. 2002. Amprenavir-resistant HIV-1 exhibits lopinavir cross-resistance and reduced replication capacity. *Aids,* 16**,** 1009-1017.

PULIDO, F., ARRIBAS, J., HILL, A. & MOECKLINGHOFF, C. 2012. No evidence for evolution of protease inhibitor resistance from standard genotyping, after three years of treatment with darunavir/ritonavir, with or without nucleoside analogues. *AIDS research and human retroviruses,* 28**,** 1167-1169.

RABI, S. A., LAIRD, G. M., DURAND, C. M., LASKEY, S., SHAN, L., BAILEY, J. R., CHIOMA, S., MOORE, R. D. & SILICIANO, R. F. 2013. Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. *The Journal of clinical investigation,* 123**,** 3848-3860.

RANGEL, H. R., WEBER, J., CHAKRABORTY, B., GUTIERREZ, A., MAROTTA, M. L., MIRZA, M., KISER, P., MARTINEZ, M. A., ESTE, J. A. & QUINONES-MATEU, M. E. 2003. Role of the human immunodeficiency virus type 1 envelope gene in viral fitness. *Journal of virology,* 77**,** 9069-9073.

RATCLIFF, A. N., SHI, W. & ARTS, E. J. 2013. HIV-1 resistance to maraviroc conferred by a CD4 binding site mutation in the envelope glycoprotein gp120. *Journal of virology,* 87**,** 923-934.

RAWAT, S. S., JOHNSON, B. T. & PURI, A. 2005. Sphingolipids: modulators of HIV-1 infection and pathogenesis. *Bioscience reports,* 25**,** 329-343.

RAY, N., HARRISON, J. E., BLACKBURN, L. A., MARTIN, J. N., DEEKS, S. G. & DOMS, R. W. 2007. Clinical Resistance to Enfuvirtide Does Not Affect Susceptibility ofHuman Immunodeficiency Virus Type 1 to Other Classes of Entry Inhibitors. *Journal of virology,* 81**,** 3240-3250.

REEVES, J. D., MIAMIDIAN, J. L., BISCONE, M. J., LEE, F.-H., AHMAD, N., PIERSON, T. C. & DOMS, R. W. 2004. Impact of mutations in the coreceptor binding site on human immunodeficiency virus type 1 fusion, infection, and entry inhibitor sensitivity. *Journal of virology,* 78**,** 5476-5485.

RIZZUTO, C. & SODROSKI, J. 2000. Fine definition of a conserved CCR5-binding region on the human immunodeficiency virus type 1 glycoprotein 120. *AIDS research and human retroviruses,* 16**,** 741-749.

ROBINSON, L. H., MYERS, R. E., SNOWDEN, B. W., TISDALE, M. & BLAIR, E. D. 2000. HIV type 1 protease cleavage site mutations and viral fitness: implications for drug susceptibility phenotyping assays. *AIDS research and human retroviruses,* 16**,** 1149-1156.

RONG, R., GNANAKARAN, S., DECKER, J. M., BIBOLLET-RUCHE, F., TAYLOR, J., SFAKIANOS, J. N., MOKILI, J. L., MULDOON, M., MULENGA, J. & ALLEN, S. 2007. Unique mutational patterns in the envelope α2 amphipathic helix and acquisition of length in gp120 hypervariable domains are associated with resistance to autologous neutralization of subtype C human immunodeficiency virus type 1. *Journal of virology,* 81**,** 5658-5668.

SAITO, A. & YAMASHITA, M. 2021. HIV-1 capsid variability: viral exploitation and evasion of capsid-binding molecules. *Retrovirology,* 18**,** 32.

SAKURABA, S. & KONO, H. 2016. Spotting the difference in molecular dynamics simulations of biomolecules. *The Journal of chemical physics,* 145**,** 074116.

SANTORO, M. M. & PERNO, C. F. 2013. HIV-1 genetic variability and clinical implications. *International Scholarly Research Notices,* 2013.

SANTOS DA SILVA, E., MULINGE, M. & PEREZ BERCOFF, D. 2013. The frantic play of the concealed HIV envelope cytoplasmic tail. *Retrovirology,* 10**,** 1-24.

SAUNDERS, C. J., MCCAFFREY, R. A., ZHARKIKH, I., KRAFT, Z., MALENBAUM, S. E., BURKE, B., CHENG-MAYER, C. & STAMATATOS, L. 2005. The V1, V2, and V3 regions of the human immunodeficiency virus type 1 envelope differentially affect the viral phenotype in an isolate-dependent manner. *Journal of virology,* 79**,** 9069-9080.

SAXENA, S. K. & CHITTI, S. V. 2016. Molecular Biology and Pathogenesis of Retroviruses. *Advances in Molecular Retrovirology.* IntechOpen.
SCHUITEMAKER, H., VAN'T WOUT, A. B. & LUSSO, P. 2011. Clinical significance of HIV-1 coreceptor usage. *Journal of translational medicine,* 9**,** 1-17.
SEN, J., JACOBS, A., JIANG, H., RONG, L. & CAFFREY, M. 2007. The disulfide loop of gp41 is critical to the furin recognition site of HIV gp160. *Protein science,* 16**,** 1236-1241.
SHANG, L., YUE, L. & HUNTER, E. 2008. Role of the membrane-spanning domain of human immunodeficiency virus type 1 envelope glycoprotein in cell-cell fusion and virus infection. *Journal of virology,* 82**,** 5417-5428.
SHARP, P. M. & HAHN, B. H. 2011. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine,* 1**,** a006841.
SHI, J., BLUNDELL, T. L. & MIZUGUCHI, K. 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of molecular biology,* 310**,** 243-257.
SHIODA, T., OKA, S., IDA, S., NOKIHARA, K., TORIYOSHI, H., MORI, S., TAKEBE, Y., KIMURA, S., SHIMADA, K. & NAGAI, Y. 1994. A naturally occurring single basic amino acid substitution in the V3 region of the human immunodeficiency virus type 1 env protein alters the cellular host range and antigenic structure of the virus. *Journal of virology,* 68**,** 7689-7696.
SHRIVASTAVA, I. & LALONDE, J. M. 2010. Fluctuation dynamics analysis of gp120 envelope protein reveals a topologically based communication network. *Proteins: Structure, Function, and Bioinformatics,* 78**,** 2935-2949.
SHRIVASTAVA, I. H., WENDEL, K. & LALONDE, J. M. 2012. Spontaneous rearrangement of the β20/β21 strands in simulations of unliganded HIV-1 glycoprotein, gp120. *Biochemistry,* 51**,** 7783-7793.
SHUKLA, E. & CHAUHAN, R. 2019. Host-HIV-1 interactome: a quest for novel therapeutic intervention. *Cells,* 8**,** 1155.
SINGH, U. 2015. *Acquired and transmitted drug resistance in HIV-1 subtype C: implications of novel mutations on replication capacity, cleavage and drug susceptibility.*
SMEATON, L. M., GIGUEL, F., NOVITSKY, V., SIKHULILE, M., MITCHELL, R. M., MAKHEMA, J., ESSEX, M., LOCKMAN, S. & KURITZKES, D. R. 2011. Prevalence and clinical associations of CXCR4-using HIV-1 among treatment-naive subtype C-infected women in Botswana. *Journal of acquired immune deficiency syndromes (1999),* 57**,** 46.
SMYTH, R. J., YI, Y., SINGH, A. & COLLMAN, R. G. 1998. Determinants of entry cofactor utilization and tropism in a dualtropic human immunodeficiency virus type 1 primary isolate. *Journal of virology,* 72**,** 4478-4484.
STEIN, B. & ENGLEMAN, E. 1990. Intracellular processing of the gp160 HIV-1 envelope precursor. Endoproteolytic cleavage occurs in a cis or medial compartment of the Golgi complex. *Journal of Biological Chemistry,* 265**,** 2640-2649.
SU, C. T.-T., KOH, D. W.-S. & GAN, S. K.-E. 2019. Reviewing HIV-1 Gag mutations in protease inhibitors resistance: insights for possible novel Gag inhibitor designs. *Molecules,* 24**,** 3243.
SUTHERLAND, K. A., PARRY, C. M., MCCORMICK, A., KAPAATA, A., LYAGOBA, F., KALEEBU, P., GILKS, C. F., GOODALL, R., SPYER, M. & KITYO, C. 2015. Evidence for reduced drug susceptibility without emergence of major protease mutations following protease inhibitor monotherapy failure in the SARA trial. *PloS one,* 10**,** e0137834.
TEDBURY, P. R. & FREED, E. O. 2014. The role of matrix in HIV-1 envelope glycoprotein incorporation. *Trends in microbiology,* 22**,** 372-378.
TEDBURY, P. R., NOVIKOVA, M., ABLAN, S. D. & FREED, E. O. 2016. Biochemical evidence of a role for matrix trimerization in HIV-1 envelope glycoprotein incorporation. *Proceedings of the National Academy of Sciences,* 113**,** E182-E190.

THEYS, K., DEFORCHE, K., LIBIN, P., CAMACHO, R. J., VAN LAETHEM, K. & VANDAMME, A.-M. 2010. Resistance pathways of human immunodeficiency virus type 1 against the combination of zidovudine and lamivudine. *Journal of General Virology,* 91**,** 1898-1908.

THIELEN, A., SICHTIG, N., KAISER, R., LAM, J., HARRIGAN, P. R. & LENGAUER, T. 2010. Improved prediction of HIV-1 coreceptor usage with sequence information from the second hypervariable loop of gp120. *The Journal of infectious diseases,* 202**,** 1435-1443.

UDENWOBELE, D. I., SU, R.-C., GOOD, S. V., BALL, T. B., VARMA SHRIVASTAV, S. & SHRIVASTAV, A. 2017. Myristoylation: An important protein modification in the immune response. *Frontiers in immunology,* 8**,** 751.

UNAIDS 2020. Seizing the moment: tackling entrenched inequalities to end epidemics. *Global AIDS Update*.

UNAIDS, U. & ORGANIZATION, W. H. 2011. Global HIV/AIDS response: epidemic update and health sector progress towards universal access: progress report 2011. *Global HIV/AIDS response: epidemic update and health sector progress towards universal access: progress report 2011.*

VAN DUYNE, R., KUO, L. S., PHAM, P., FUJII, K. & FREED, E. O. 2019. Mutations in the HIV-1 envelope glycoprotein can broadly rescue blocks at multiple steps in the virus replication cycle. *Proceedings of the National Academy of Sciences,* 116**,** 9040-9049.

VAN GILS, M. J., BUNNIK, E. M., BOESER-NUNNINK, B. D., BURGER, J. A., TERLOUW-KLEIN, M., VERWER, N. & SCHUITEMAKER, H. 2011. Longer V1V2 region with increased number of potential N-linked glycosylation sites in the HIV-1 envelope glycoprotein protects against HIV-specific neutralizing antibodies. *Journal of virology,* 85**,** 6986-6995.

VAN MAARSEVEEN, N. M., ANDERSSON, D., LEPŠÍK, M., FUN, A., SCHIPPER, P. J., DE JONG, D., BOUCHER, C. A. & NIJHUIS, M. 2012. Modulation of HIV-1 Gag NC/p1 cleavage efficiency affects protease inhibitor resistance and viral replicative capacity. *Retrovirology,* 9**,** 1-12.

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

VIARD, M., ABLAN, S. D., ZHOU, M., VEENSTRA, T. D., FREED, E. O., RAVIV, Y. & BLUMENTHAL, R. 2008. Photoinduced reactivity of the HIV-1 envelope glycoprotein with a membrane-embedded probe reveals insertion of portions of the HIV-1 Gp41 cytoplasmic tail into the viral membrane. *Biochemistry,* 47**,** 1977-1983.

WAHEED, A. A. & FREED, E. O. 2012. HIV type 1 Gag as a target for antiviral therapy. *AIDS research and human retroviruses,* 28**,** 54-75.

WALLIS, C. L., MELLORS, J. W., VENTER, W. D., SANNE, I. & 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 research and treatment,* 2011.

WANG, B., DAI, T., SUN, W., WEI, Y., REN, J., ZHANG, L., ZHANG, M. & ZHOU, F. 2021. Protein N-myristoylation: functions and mechanisms in control of innate immunity. *Cellular & Molecular Immunology,* 18**,** 878-888.

WANG, L., ENG, E. T., LAW, K., GORDON, R. E., RICE, W. J. & CHEN, B. K. 2017. Visualization of HIV T cell virological synapses and virus-containing compartments by three-dimensional correlative light and electron microscopy. *Journal of virology,* 91**,** e01605-16.

WANG, W., LI, Y., ZHANG, Z. & WEI, W. 2022. Human immunodeficiency virus-1 core: The Trojan horse in virus–host interaction. *Frontiers in Microbiology,* 13.

WANG, W., NIE, J., PROCHNOW, C., TRUONG, C., JIA, Z., WANG, S., CHEN, X. S. & WANG, Y. 2013. A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology,* 10**,** 1-14.

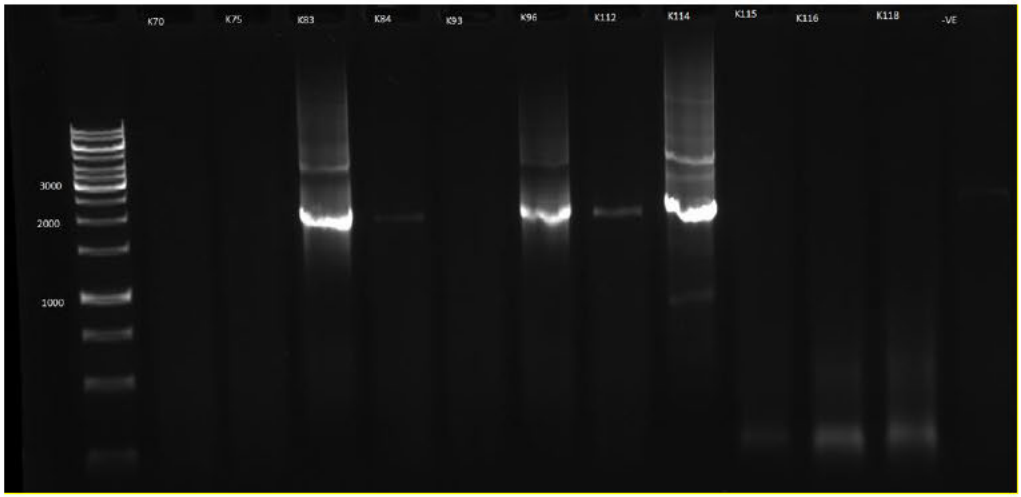
WENG, L., STOTT, S. L. & TONER, M. 2019. Exploring dynamics and structure of biomolecules, cryoprotectants, and water using molecular dynamics simulations: implications for biostabilization and biopreservation. *Annual review of biomedical engineering,* 21**,** 1-31.
WENSING, A. M., VAN MAARSEVEEN, N. M. & NIJHUIS, M. 2010. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res,* 85**,** 59-74.
WESTBY, M. & VAN DER RYST, E. 2005. CCR5 antagonists: host-targeted antivirals for the treatment of HIV infection. *Antiviral chemistry and chemotherapy,* 16**,** 339-354.
WILEN, C. B., TILTON, J. C. & DOMS, R. W. 2012. Molecular mechanisms of HIV entry. *Viral Molecular Machines.* Springer.
WILLEY, R. L., BONIFACINO, J. S., POTTS, B. J., MARTIN, M. A. & KLAUSNER, R. D. 1988. Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160. *Proceedings of the National Academy of Sciences,* 85**,** 9580-9584.
WILLEY, R. L., RUTLEDGE, R. A., DIAS, S., FOLKS, T., THEODORE, T., BUCKLER, C. E. & MARTIN, M. A. 1986. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. *Proceedings of the National Academy of Sciences,* 83**,** 5038-5042.
WOOLFSON, D. N. & WILLIAMS, D. H. 1990. The influence of proline residues on α-helical structure. *FEBS letters,* 277**,** 185-188.
WYMA, D. J., JIANG, J., SHI, J., ZHOU, J., LINEBERGER, J. E., MILLER, M. D. & AIKEN, C. 2004. Coupling of human immunodeficiency virus type 1 fusion to virion maturation: a novel role of the gp41 cytoplasmic tail. *Journal of virology,* 78**,** 3429-3435.
WYSS, S., DIMITROV, A. S., BARIBAUD, F., EDWARDS, T. G., BLUMENTHAL, R. & HOXIE, J. A. 2005. Regulation of human immunodeficiency virus type 1 envelope glycoprotein fusion by a membrane-interactive domain in the gp41 cytoplasmic tail. *Journal of virology,* 79**,** 12231-12241.
YANG, H., NKEZE, J. & ZHAO, R. Y. 2012. Effects of HIV-1 protease on cellular functions and their potential applications in antiretroviral therapy. *Cell & bioscience,* 2**,** 1-8.
YANG, Z., WONG, W. S. & NIELSEN, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular biology and evolution,* 22**,** 1107-1118.
YOKOYAMA, M., NAGANAWA, S., YOSHIMURA, K., MATSUSHITA, S. & SATO, H. 2012. Structural dynamics of HIV-1 envelope Gp120 outer domain with V3 loop. *PLoS One,* 7**,** e37530.
YOKOYAMA, M., NOMAGUCHI, M., DOI, N., KANDA, T., ADACHI, A. & SATO, H. 2016. In silico analysis of HIV-1 Env-gp120 reveals structural bases for viral adaptation in growth-restrictive cells. *Frontiers in microbiology,* 7**,** 110.
YU, X., YU, Q.-C., LEE, T. & ESSEX, M. 1992. The C terminus of human immunodeficiency virus type 1 matrix protein is involved in early steps of the virus life cycle. *Journal of virology,* 66**,** 5667-5670.
ZDANOWICZ, M. M. 2006. The pharmacology of HIV drug resistance. *American journal of pharmaceutical education,* 70.
ZHANG, H., TULLY, D. C., ZHANG, T., MORIYAMA, H., THOMPSON, J. & WOOD, C. 2010. Molecular determinants of HIV-1 subtype C coreceptor transition from R5 to R5X4. *Virology,* 407**,** 68-79.
ZHONG, P., AGOSTO, L. M., MUNRO, J. B. & MOTHES, W. 2013. Cell-to-cell transmission of viruses. *Current opinion in virology,* 3**,** 44-50.
ZOLLA-PAZNER, S. & CARDOZO, T. 2010. Structure-function relationships of HIV-1 envelope sequence-variable regions refocus vaccine design. *Nat Rev Immunol,* 10.
ZUMA, K., SIMBAYI, L., ZUNGU, N., MOYO, S., MARINDA, E., JOOSTE, S., NORTH, A., NADOL, P., AYNALEM, G. & IGUMBOR, E. 2022. The HIV epidemic in South Africa: Key findings from 2017 national population-based survey. *International Journal of Environmental Research and Public Health,* 19**,** 8125.

Appendix

Appendix 1: Mutational table showing PIDs and mutations in PR, Gag, and Env

PID	Protease	Gag	Envelope
PCSK18	NONE	Q69K, R76K, Y79F, S111I, Q182S, I256V	T138N
PCSK20	NONE	R76K, Y79F, I256VK, 436R, L449P	T138R
PCSK22	NONE	Q69K	T138S, P183Q
PCSK24	NONE	Y79F	Q315A, S534A, D632E
PCSK28	NONE	NONE	T536A, I688V
PCSK36	NONE	R76K, T239A	T138A, N195H, Q315R, P724Q
PCSK59	NONE	Q69K, R76K, S111C	T132S, P183Q
PCSK70	NONE	Q69K, R76K, I256V	P183A, P724Q
PCSK75	NONE	Q69K, V128I, I256V	T138S, S534A
PCSK83	NONE	R76K, Y79F, S111C, I256V	N195H, Q315R, T536A, P724S
PCSK84	NONE	Q69K, Y79F	T132A, T138S, N195S, Q315R, T536A, D632E
PCSK114	NONE	I256V	T138S, P183K, T536A, P724S
PCSK128	NONE	Q69K, I256V	T132A
PCSK120	NONE	Q69K, R76K, S111C, K436R	P183Q, D632E, P724Q
PCSK145	NONE	N/A	P724Q
PCSK93	NONE	R76K, S111C, P453L	T132A, D632E
PCSK19	V82A	I256V, L449F, R452K	S110N, T138K, P183S, N195H
PCSK61	M46I,I54V,L76V,V82A	Q69K, R76K, V128I, T239A, I256V, A431V, L449F	P183S, S534A
PCSK89	M46I,I54V,L76V,V82A	R76K, S111C, A431V	T132S, N195H, Q315R, T536M, P724Q
PCSK90	M46I,I50V,I54V,V82A	Q69K, A431V, P453L	S110N, N195H, Q315K
PCSK108	M46I,I54V,L76V,V82A	Q69K, R76K, S111C, I256V, T239S, A431V	P183Q, T536A, I688V
PCSK153	M46I,I54V, L76V,V82A, L24I	Q69K, R76K, Y79F, Q182S, I256V	NONE
PCSK33	M46I,F53L,I54V,L76V,V82A,L90M	Q69K, R76K, I256V, Q182S, K436R	P183E, N195I, D632E

Appendix 2: Supplementary material for chapter 2.



Supplementary figure 1: Gel picture, showing the amplification of Gag-PR. Gag-PR size 1.7 kb.

Supplementary figure 2.: Gel picture showing the amplification of envelope. Envelope size 3.5 kb.

S-table 1. Sequencing Primers.

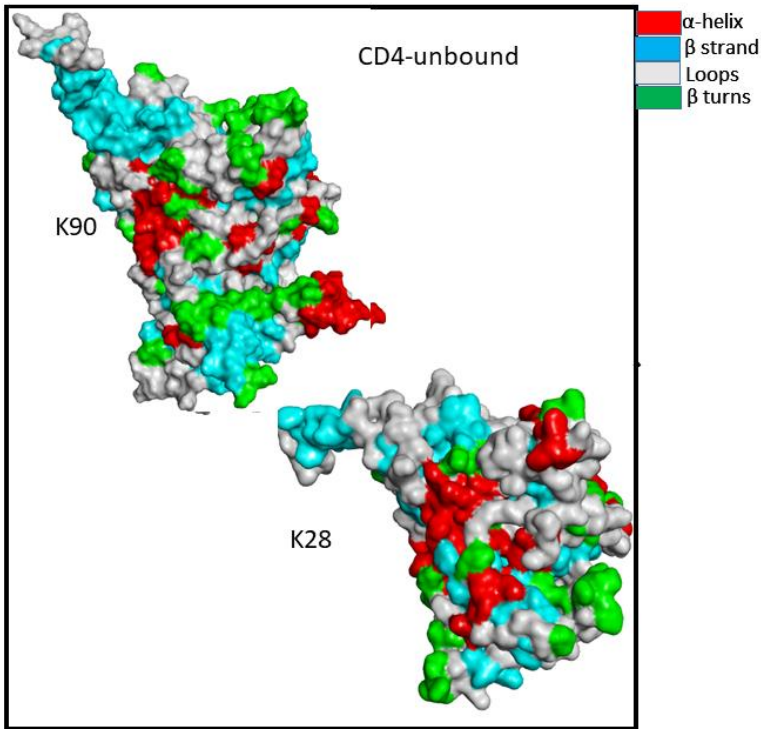
Forward Primers		
REC1F	5581 → 5602	5′- AGG GGC CGC AGA GGG AAC CAT A -3′
ENV-Z1F	6327 → 6353	5′- TGG GTC ACA GTC TAT TAT GGG GTA CCT- 3′
ENV-DF	7208 → 7226	5′- AGC ACA TTG TAA CAT TAG T- 3′
ENV-TUG	7859 → 7879	5′- GTC TGG TAT AGT GCA ACA GCA- 3′
Reverse Primers		
E270R	8566 ← 8585	5′- GCG TCC CAG AAG TTC CAC TG- 3′
ENV-ZAR	7629 ← 7652	5′- GTC CCT CAT ATC TCC TCC TCC TCT- 3′

Appendix 3: Supplementary material for chapter 3.

Appendix 3.1: Table S1: HIV-1 gp120 AA residues within a 5-radius range interacting with WT positions and mutations potentially contributing to PI failure.

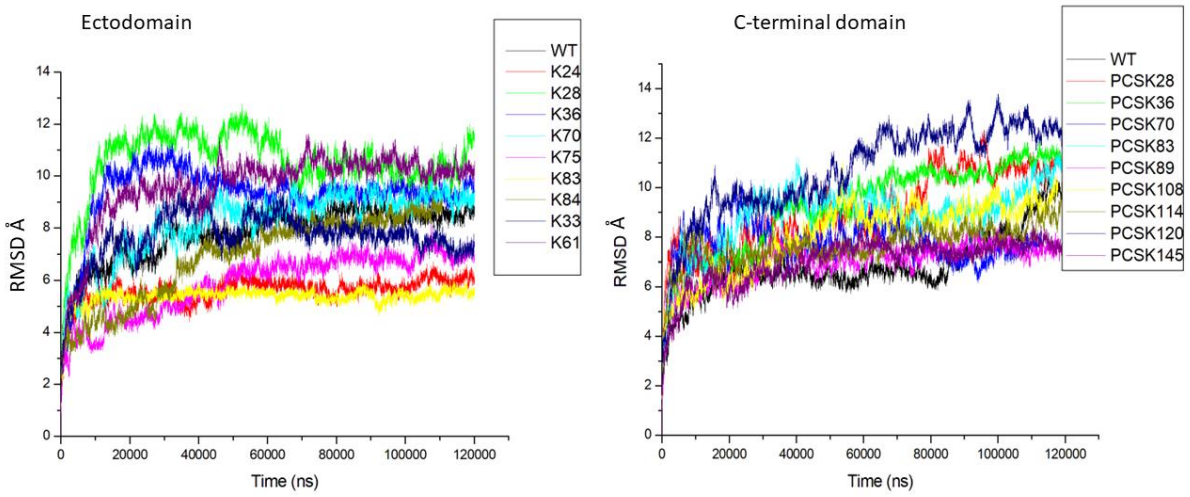
		AA interaction within 5Å										
S110	**S110N**	**T132**	**T132S**	**T138**	**T138S**	**P183**	**P183Q**	**P183S**	**N195**	**N195H**	**Q315**	**Q315R**
E106	E106	N130	T130	N136	S136	V181	I181	V181	D180	D180	V120	V120
D107	D107	C131	C131	V137	N137	V182	V182	V182	L193	L193	K121	K121
I108	I108	D133	N133	R139	T139	F184	L184	I184	I194	I194	L122	L122
I109	I109	M154	V134	N140	N140	G185	D185	G185	N195	C196	T163	T123
L111	L111	K155	M154	G324	N141	D187	N186	G186	C196	N197	I309	P124
W112	W112	N156	K155	D325	G142	S188	S187	D187	N197	T198	R310	T163
D113	D113	C157	N156	I326	N143	E190	S188	T188	T198	S199	I311	I309
Q114	Q114	S158	C157		N152	Y191	S189	N190	S199	T200	G312	R310
S115	S115	N186	H188		E153	R192	D190	S192	I201	I201	P313	I311
R429		D187	S189		D325		Y191	Y193	K421	I423	G314	G312
		S188	Y191		I326		R192	I194	I423	H425	A316	P313
		G189			R327				N425	Q428	F317	G314
					Q328				Q428	R429		G316
									G429	G431		F317
									G431	R432		
									A433	A433		

Appendix 3.2: The gp120 structural surface of the CD4-unbound, modelled using PDB:6CK9 as a template.

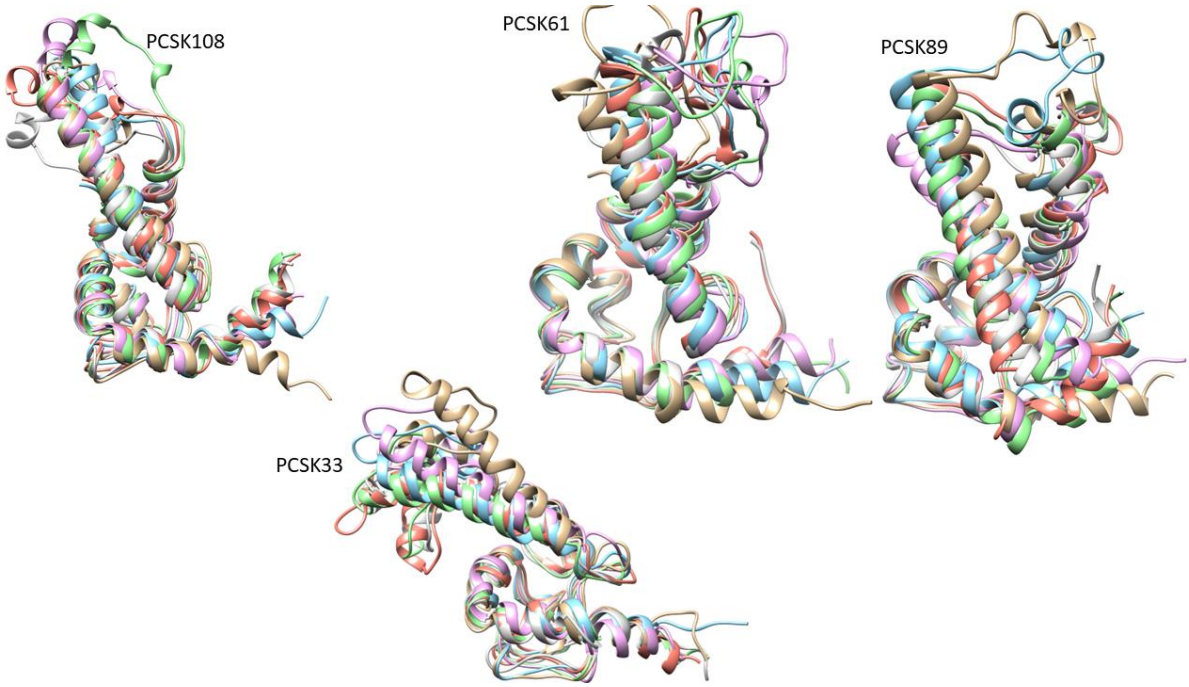


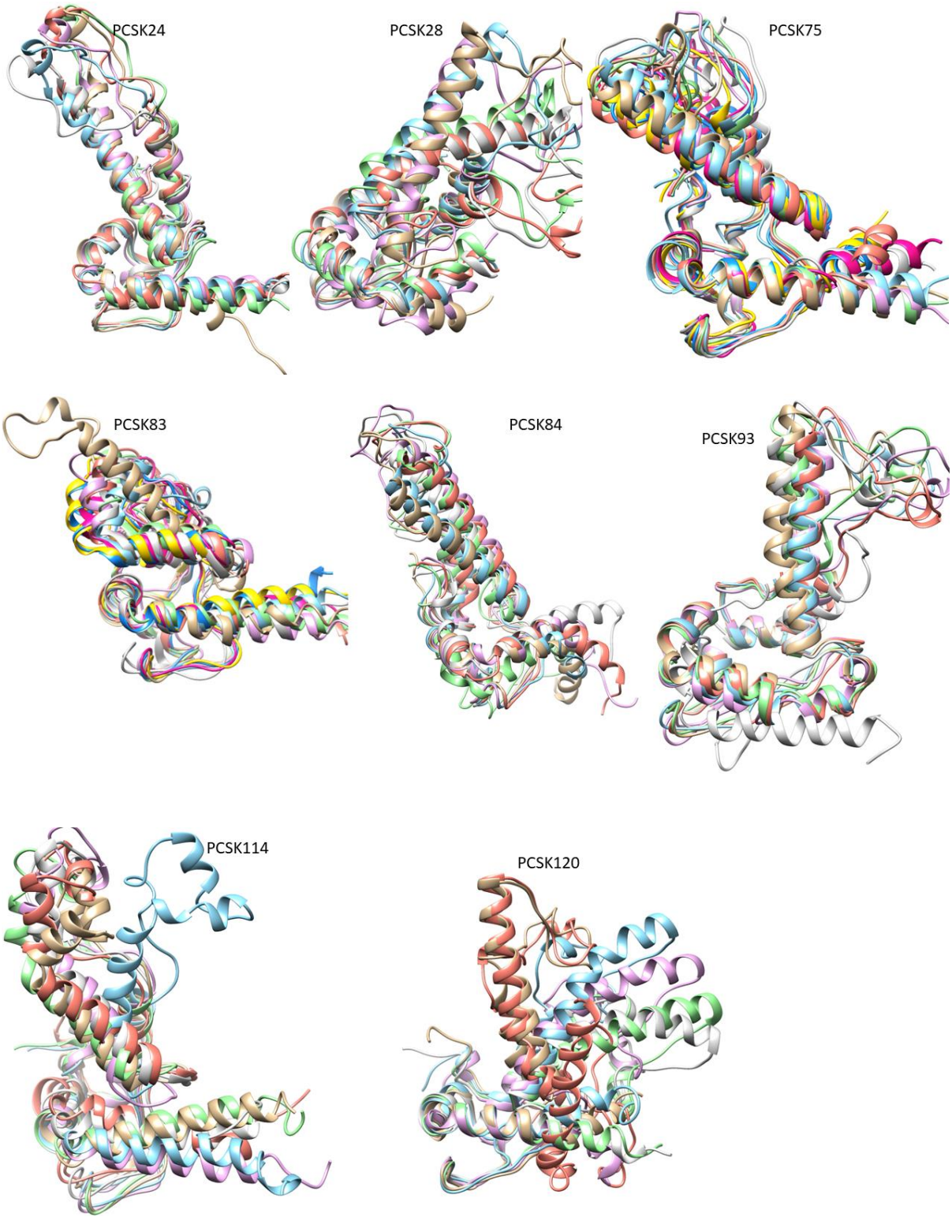
Appendix 4: Supplementary material for chapter 4

Appendix4.1: RMSD for gp41 ectodomain and C-terminal domains of gp41 (MPER and CT).



Appendix 4.2: Superimposed structures of ectodomain, modelled using PDB 5UM8 as a template.




PCSK24
PCSK28
PCSK75
PCSK83
PCSK84
PCSK93
PCSK114
PCSK120

Appendix 4.3: Salt bridges (ionic + hydrogen bonds) interaction table from ring server.

PID	HR1 & HR1	HR1 & Loop	Loop	Loop & HR2	HR2 & HR2
WT	R557-E560			K601-E654	R644-D648 & K655-E659
PCSK24					T626R-H643E
PCSK28					K655-Q658
PCSK33	R557-E560			R585-D632	
PCSK36					
PCSK61				R585-D589 & K601-E654	
PCSK70			E584-K588 & R585-D589	R585-D632	R644-E647
PCSK75				K585-D632	
PCSK83		R542-D589		R585-D640	
PCSK84	R557-E560		R585-D589		K655E-Q658K
PCSK89	R557-E560			R585-D632 & K601-E657	
PCSK93					K644-E647
PCSK108	R557-E560			R585-D632 & D589-R685	
PCSK114				K617-E634	
PCSK120			R585-D589	K601-E657	L641E-R644
PCSK145					R644-E647

Appendix 5: Biomedical Ethics recertification

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

09 January 2017

Ms MN Fortunate ([redacted])
School of Laboratory Medicine and Medical Sciences
College of Health Sciences
maphumulokhe@yahoo.com

Dear Ms Fortunate

Protocol: The role of HIV-1 envelope in the development of Protease Inhibitor drug resistant
Degree: PhD
BREC Ref No: BE678/17

EXPEDITED APPLICATION

A sub-committee of the Biomedical Research Ethics Committee has considered and noted your application received on 28 November 2017.

The study was provisionally approved pending appropriate responses to queries raised. Your response received on 21 December 2017 to BREC correspondence dated 20 December 2017 has been noted by a sub-committee of the Biomedical Research Ethics Committee. The conditions have now been met and the study is given **full ethics approval** and may begin as from 09 January 2018.

This approval is valid for one year from **09 January 2018.** To ensure uninterrupted approval of this study beyond the approval expiry date, an application for recertification must be submitted to BREC on the appropriate BREC form 2-3 months before the expiry date.

Any amendments to this study, unless urgently required to ensure safety of participants, must be approved by BREC prior to implementation.

Your acceptance of this approval denotes your compliance with South African National Research Ethics Guidelines (2015), South African National Good Clinical Practice Guidelines (2006) (if applicable) and with UKZN BREC ethics requirements as contained in the UKZN BREC Terms of Reference and Standard Operating Procedures, all available at http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx.

BREC is registered with the South African National Health Research Ethics Council (REC-290408-009). BREC has US Office for Human Research Protections (OHRP) Federal-wide Assurance (FWA 678).

The sub-committee's decision will be **RATIFIED** by a full Committee at its next meeting taking place on **13 February 2018.**

Appendix 6: Published article in the journal of microbiology research

microbiology research

MDPI

Article

Potential Associations of Mutations within the HIV-1 *Env* and *Gag* Genes Conferring Protease Inhibitor (PI) Drug Resistance

Ntombikhona F. Maphumulo and Michelle L. Gordon *

Department of Virology, Doris Duke Medical Research Institution, College of Health Sciences, University of KwaZulu-Natal, Durban 4013, South Africa; maphumulokhe@gmail.com
* Correspondence: tarinm@ukzn.ac.za; Tel.: +27-31-260-4498

Abstract: An increasing number of patients in Africa are experiencing virological failure on a second-line antiretroviral protease inhibitor (PI)-containing regimen, even without resistance-associated mutations in the protease region, suggesting a potential role of other genes in PI resistance. Here, we investigated the prevalence of mutations associated with Lopinavir/Ritonavir (LPV/r) failure in the Envelope gene and the possible coevolution with mutations within the Gag-protease (*gag-PR*) region. *Env* and *Gag-PR* sequences generated from 24 HIV-1 subtype C infected patients failing an LPV/r inclusive treatment regimen and 344 subtype C drug-naïve isolates downloaded from the Los Alamos Database were analyzed. Fisher's exact test was used to determine the differences in mutation frequency. Bayesian network probability was applied to determine the relationship between mutations occurring within the *env* and *gag-PR* regions and LPV/r treatment. Thirty-five mutations in the *env* region had significantly higher frequencies in LPV/r-treated patients. A combination of Env and Gag-PR mutations was associated with a potential pathway to LPV/r resistance. While Env mutations were not directly associated with LPV/r resistance, they may exert pressure through the Gag and minor PR mutation pathways. Further investigations using site-directed mutagenesis are needed to determine the impact of Env mutations alone and in combination with Gag-PR mutations on viral fitness and LPV/r efficacy.

Keywords: HIV-1; Envelope; Gag-protease; mutations; protease inhibitors

Citation: Maphumulo, N.F.; Gordon, M.L. Potential Associations of Mutations within the HIV-1 *Env* and *Gag* Genes Conferring Protease Inhibitor (PI) Drug Resistance. *Microbiol. Res.* **2021**, *12*, 967–977. https://doi.org/10.3390/microbiolres12040071

Academic Editor: Subhash Chand

Received: 9 October 2021
Accepted: 16 November 2021
Published: 15 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

While the dispensing of HIV-1 protease inhibitors (PIs) as part of a routine second-line regimen has been a significant turning point in the management of HIV-1 [1], the development of resistance to PIs is also increasing [2].

PIs are known to inhibit the activity of the HIV-1 protease (PR) enzyme responsible for the proteolytic processing of HIV structural Gag and enzymatic Pol polyprotein components [3]. The viral protease (PR) enzyme cleaves the Gag precursor protein ($Pr55^{Gag}$) and the Gag-Pol precursor protein ($Pr160^{Gag\text{-}Pol}$), resulting in virion maturation [3–5]. This proteolytic process prevents the formation and maturation of infectious HIV particles. The efficacy of the PI is limited by the emergence of resistance mutations that are potentially caused by poor compliance, subtherapeutic systemic levels of the drug, or prolonged treatment with one PI-based regimen during virologic failure [6].

Interestingly, >20% of patients failing a PI-based regimen do not harbor any resistance-associated mutations in the protease (PR) domain [7–9]. While non-adherence has been identified as a key player, some studies have suggested that Gag mutations can indirectly affect the Gag cleavage site and drive PI resistance without any mutations in the protease region [10,11]. These mutations were commonly found close to the cleavage sites, although non-cleavage site mutations have also been linked with PI resistance [12,13]. Verheyen et al. (2006) have linked Gag precursor p7–p1 (A431V, K436R, and I437V) and p1–p6 (L449P, P452S, and P453L) cleavage site mutations with resistance to PIs [14–16]. These mutations

are thought to compensate for any enzymatic impairment of protease caused by the loss of van der Waals interactions between the inhibitor and binding sites [12,14,15,17–20].

Others have suggested that the mechanism for PI failure may involve mutations in the Envelope (*env*) gene [5,21]. Coetzer et al. (2017) reported Gp41 mutations from both the Heptad repeat (HR) (607T and 641L) and cytoplasmic tail (CT) (721I) that potentially contribute to PI failure [21] Two studies that looked at virological failure during PI exposure both reported that mutations in the CT impacted PI susceptibility [22,23]. Furthermore, Env mutations have been suggested to promote cell to cell transmission, leading to high-level drug resistance mutations in ARV target genes; by doing so, they increase the level of resistance to a broad panel of ARVs in vitro [24,25]. The HIV-1 Envelope is said to concurrently evolve to escape from both neutralizing antibodies (NABs) and ARVs [25].

Although studies have reported the emergence of Gag and Env mutations in patients failing PIs [21–23], to date, no one has reported on the coevolution of Gag and Env in these patients. Here, we investigated the prevalence of Env mutations associated with PI treatment failure and the possibility of coevolution with mutations within Gag, and protease, in full-length Env and Gag-PR sequences from HIV-1 subtype C infected individuals failing a PI-inclusive regimen.

2. Materials and Methods

2.1. Study Cohort

This retrospective study used 24 stored plasma samples obtained from virologically failing PI-treated patients enrolled in the Protease Cleavage Site (PCS) study (2009–2013) at McCord and King Edward VIII hospitals Durban, South Africa [26]. All enrolled patients received Lopinavir/Ritonavir therapy for at least six months and had plasma HIV-1 RNA levels > 1000 copies/mL. In addition, sequences from 344 subtype C drug-naïve isolates were downloaded from the Los Alamos HIV-1 Database (http://hiv-web.lanl.gov) accessed on 10 September 2020. Written informed consent was obtained from all study participants. The Biomedical Research Ethics Committee of the University of KwaZulu-Natal (BREC NO: 678/17) approved the study.

2.2. Amplification and Sequencing Analyses of the Env Domain

Viral RNAs were extracted from 140 µL of plasma using the QIAamp RNA kit (QIAGEN Services, Inc., Germantown, MD, USA) and reverse-transcribed using a Thermoscript RT-PCR kit (Invitrogen, Carlsbad, CA, USA). A 1.7 kb Gag-Protease product was amplified by nested PCR using the Takara Ex Taq HS enzyme kit. Specific primers, Gag +1 5′-GAGATCTCTCGACGCAGGAC-3′ (HXB2 nucleotide: 675 to 697, forward primer) and 3″rvp 5′-GGAGTGTTATat GGATTTTCAGGCCCAATT-3′ (HXB2 positions: 2696 to 2725, reverse primer), were used for the first-round PCR, at 55 °C for 30 min (cDNA synthesis) and 94 °C for 2 min (initial denaturation), followed by 35 cycles of 94 °C for 15 s (denaturation), 55 °C for 30 s (annealing) and 68 °C for 2 min (extension), and ended with a 5 min incubation at 68 °C (final extension). Long fwd 5′-GAC TCG GCT TGC TGA AGC GCG CAC GGC AAG AGG CGA GGG GCG ACT GGT GAGTAC GCC AAA AAT TTT GAC TAG CGG AGG CTA GAA GGA GAGAGA TGG G-3′ (HXB2 nucleotide: 695 to 794, forward primer) and Long rev 5′-GGC CCA ATT TTT GAA ATT TTT CCT TCC TTT TCC ATT TCT GTA CAA ATT TCT ACT AAT GCT TTT ATT TTT TCT GTC AAT GGC CAT TGT TTA ACT TTT G-3′ (HXB2 nucleotide: 2706 to 2805, reverse primer) were used for nested PCR at the following conditions: 94 °C for 2 min (initial denaturation), 40 cycles of 94 °C for 30 s (denaturation), 60 °C for 30 s (annealing) and 72 °C for 2 min (extension) followed by a 7 min hold at 72 °C (final extension). Primers were complementary to subtype B, gag-protease-deleted NL4-3 plasmid (pNL4-3Δgag-protease) on either side. Amplicons were electrophoresed in a 1% agarose gel to confirm the presence of the 1.7 kb amplicon corresponding to the *gag-PR* gene. The size of the product was determined using the GeneRuler™ 1 kb DNA Ladder (Fermentas International Inc., Waltham, MA, USA), as shown in Supplementary Figure S1.

Next, we amplified an Env fragment spanning HXB2 positions 5580 to 8586 by nested PCR using first-round primers Rec2F 5′-GATAAAGCCACCTTTGCCTAGT-3′ (HXB2 positions: 5514) and Env2 5′-TTCTAGGTCTCGAGATACTGCT-3′ (HXB2 positions: 8889), second-round primers Env1 5′-AAGGGCCACAGAGGGAGCCATA-3′ (HXB2 positions: 5580) and E270R, 5′-GCGTCCCAGAAGTTCCACAA-3′ (HXB2 positions: 8566) and PCR conditions as previously described [27]. Sequences were generated using the Gag-Protease sequencing primers 5′-CTT GTC TAG GGC TTC CTT GGT-3′ (HXB2 position: 1078–1098), 5′-CTT CAG ACA GGA ACA GAGGA-3′ (991–1010), 5′-GGT TCT CTC ATC TGG CCT GG-3′ (1462–1481), 5′-CCT TGC CAC AGT TGA AAC ATT T-3′ (1960–1981), 5′-TAG AAG AAA TGA TGA CAG-3′ (1817–1834), 5′-CTA ATA CTGTatCAT CTG CTC CTG T-3′ (2328–2353), 5′-CCT GGC TTT AAT TTT TAC TGG-3′ (2196–2268). Cycling conditions were as follows: 94 °C for 2 min (initial denaturation), 35 cycles of 94 °C for 15 s (denaturation), 55 °C for 30 s (annealing) and 68 °C for 4 min (extension) followed by a 10 min hold at 68 °C (final extension). Amplification of the *env* region was confirmed by gel electrophoresis (shown in Supplementary Figure S2)

The gp160 Env was sequenced using population-based forward and reverse primers (shown in Supplementary Table S1) and the BigDye v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA, USA), and run on an Applied Biosystems (ABI) 3130xl automated sequencer. Sequences were assembled and edited using the Applied Biosystems SeqScape® Software and aligned against subtype C reference.

2.3. Statistical Analysis

Multiple sequence alignments were generated using the ClustalX program (http://www.clustal.org/) accessed on 16 September 2020, and manually edited using BIOEDIT (Ibis Biosciences, An Abbott Company, Sylmar, CA, USA). The Virulign tool (https://github.com/rega-cev/virulign) accessed on 11 December 2020, was used to determine the amino acid substitution at each position in Gag-PR and Env on PI exposure and naïve isolate sequences. Fisher's exact test was performed for all positions to identify mutations significantly more prevalent (p-value < 0.05) in PI-experienced patients.

2.4. Bayesian Network

Bayesian network learning was performed using B-Course (http:/www.b-course) accessed on May 2021, which modeled direct and indirect associations between Gag, PR, and Env mutations and exposure to PI treatment. Only significant mutations identified using Fisher's exact test were included in the model together with known PI-resistance-associated mutations. As previously described, associations between mutations learned and graphical representations were evaluated [28,29]. Non-parametric bootstrap analysis was calculated by running 5000 replicates to derive network robustness. Only interactions with bootstrap support over 65% were included in the network.

2.5. Coevolution Using CAPS

Gag-PR-Env coevolution analysis was performed using the Coevolution Analysis for Protein Sequences (CAPS, http://bioinf.gen.tcd.ie/caps/) accessed on 26 October 2020 program. CAPS identifies groups of coevolving pairs with a correlation coefficient > 0.5 by using amino acid (AA) sites to compare the correlated variance of their evolutionary rate using the probability scores between two pairs of aligned sequences [30]. The evolutionary rate was estimated using the blocks substitution matrix (BLOSUM). The distribution of 5000 randomly sampled values was used to identify coevolving codons.

2.6. Positive Selection

Positive selection pressure in Gag-PR-Env sites was assessed using the rate ratio of non-synonymous to synonymous (ω) substitutions in the CODEML program of the PAML software package (version 4.9) [31]. Analysis was performed using three classes $\omega \leq 1$, $\omega = 1$, $\omega \geq 1$ assuming purifying, neutral, and positive selection, respectively. The ω value and likelihood estimates were calculated for three different codon-based ML pairs of site models: M0 (one ω) vs. M3 (discrete), M1a (nearly neutral) vs. M2a (positive selection), and M7 (beta) vs. M8 (beta and $\omega \geq 1$). Comparison of M0 vs. M3 is a test of site rate variation, while M1 vs. M2 is for positive selection. The likelihood ratio test was used to evaluate the best-fitting model for the data [32]. The Bayes Empirical Bayes method was used to identify specific sites under positive selection.

3. Results

3.1. Participant Characteristics

Twenty-four Gag-Prot-Env sequences from 24 patients failing a PI-inclusive treatment regimen were available for analysis, of whom 14 were females and 10 were males, with the median age of 35 years (interquartile range (IQR) 17–38 years). The median viral load was 4.84 $\log_{10}$ copies/mL (IQR 4.12–5.51). All patients were infected with HIV subtype C as established by the Virulign tool.

3.2. Prevalence of Envelope Mutations in LPV/r-Experienced versus ART-Naïve Subtype C Sequences

Twenty-four Env sequences were available for analysis. The prevalence of amino acids at each codon position was calculated by comparing sequences from LPV/r-treated and ART-naïve isolates. All sequences were confirmed as subtype C using the Rega subtyping tool. Figure 1 shows a total of thirty-five mutations identified in the *env* gene, with significantly higher frequencies in LPV/r-experienced patients compared to those that were ART-naïve. These were: I19LSV (32% vs. 13%), S110EN (16% vs. 4%), D167AGKNR (16% vs. 2%), P183AEKQRS (48% vs. 20%), Y191CFHNS (12% vs. 1%), N195HIKRST (28% vs. 8%), N230DT (20% vs. 16%), Q258EP (8% vs. 1%), N262LI (8% vs. 0%), T283IKNQSV (28% vs. 9%), V286AIM (16% vs. 3%), I294LMN (8% vs. 0%), S306DGNR (24% vs. 5%), P313DI (8% vs. 0%), Q315AKR (32% vs. 3%), I323KNRTV (12% vs. 1%), A329DIPT (16% vs. 0%), C331RS (8% vs. 0%), W338EG (8% vs. 0%), H374KLPR (8% vs. 1%), F382V (8% vs. 1%), F383LV (8% vs. 0%), L390QV (8% vs. 0%), N425DHR (20% vs. 4%), L453EIM (20% vs. 4%), R469KTW (8% vs. 1%), V489IL (16% vs. 1%), V505AEMT (8% vs. 1%), E507GQRW (28% vs. 10%), S534AL (12% vs. 2%), T536AMSV (32% vs. 10%), Q567KR (8% vs. 1%), D632EG (24% vs. 8%), I688TV (12% vs. 1%), P724LQRSX (32% vs. 12%). In contrast, the K63RT (16% vs. 50%), T132NQ (24% vs. 97%), T138NQ (36% vs. 80%), and V182EIKLMNT (12% vs. 56%) mutations were found at a lower frequency in the PI-experienced versus ART-naïve sequences.

3.3. Positive Selected Sites in Envelope and Coevolution with Gag-PR Residues

Here, we determined which amino acid changes were under positive selection pressure following treatment failure using PAML [30]. Figure 2 shows PAML analysis of amino acid changes from participants failing an LPV/r-inclusive treatment regimen and ART-naïve participants that were positively selected. A total of twelve sites were positively selected: V85, T138, Y146, M147, E150, K151, G152, D277, G321, K322, N463, and T465. None of these sites were previously associated with entry and fusion inhibitor or LPV/r failure.

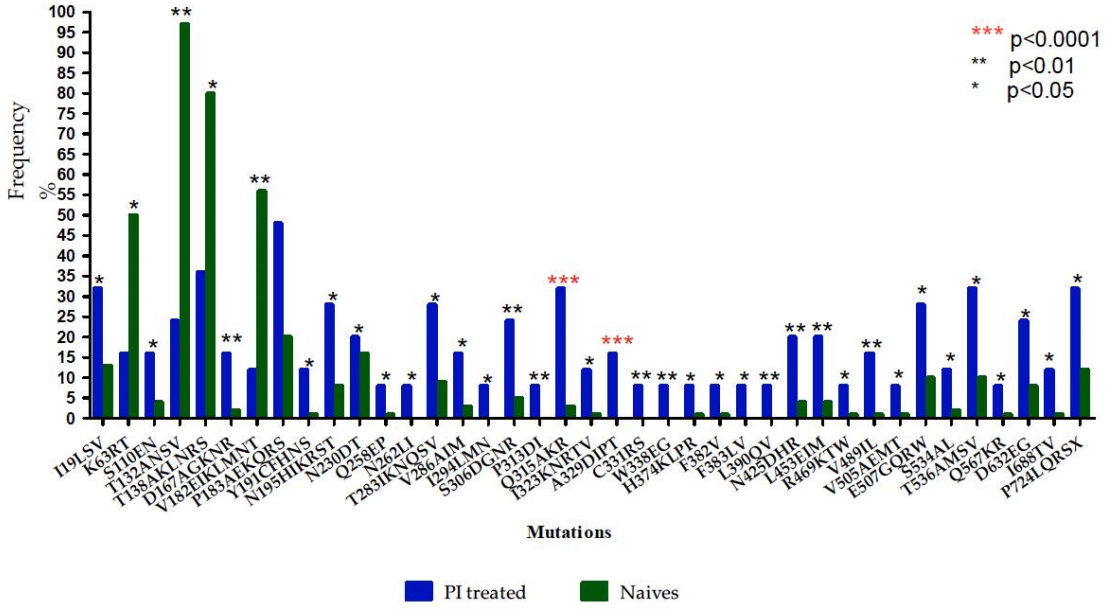


Figure 1. Frequency (%) of Env mutations in ART-naïve versus LPV/r-treated subtype C isolates. Asterisks (*): *** represents $p < 0.0001$, ** represents $p < 0.01$, and * represents $p < 0.05$.

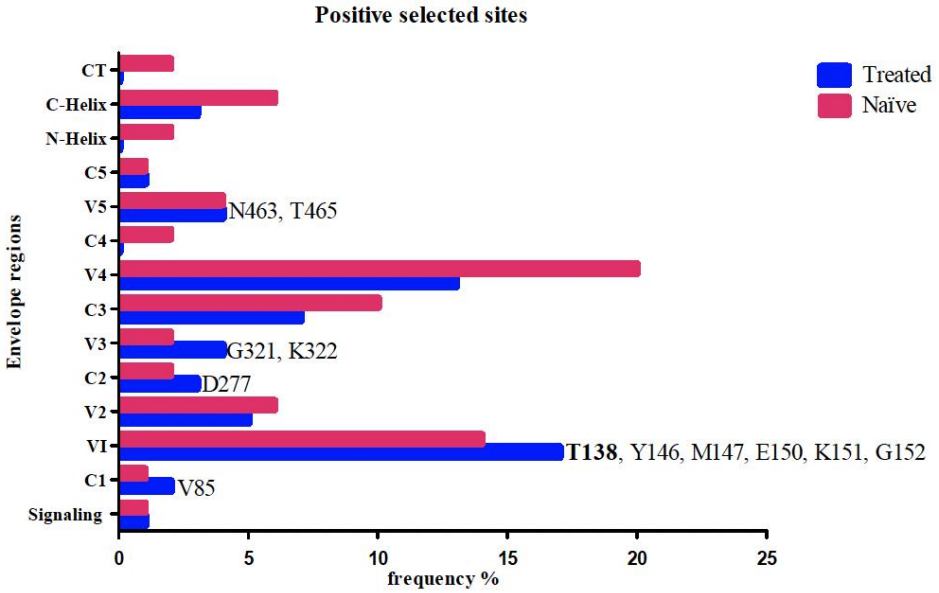


Figure 2. Positively selected amino acid codons in Env in LPV/r-treated and naïve isolates with the Bayesian probability of $p > 99\%$. Blue shows the positively selected codons in the treated isolates, while maroon indicates the positively selected codons in the naïve isolates. Codon positions identified only in treated and not in naïve isolates are shown at the relevant *env* region. T138 is indicated in bold font and has a higher frequency in the treated isolates.

Coevolution Analyses for Protein Sequences (CAPS) was used to determine whether amino acids in Gag, PR, and Env sites coevolved in patients failing an LPV/r-inclusive treatment regimen. All positions with amino acid variation greater than 1% were included in the analysis. The coevolution analysis identified seven amino acid coevolving pairs, but these did not correlate significantly. The coevolving pairs were mostly seen between Env and Gag sites (Env: 11, 23, 29,142, 147, 336, 389, 400, 496, 534, 617, and 668; Gag: 108, 225, 256, 335, 372, 388, 478, and 480), while only one was observed between Env and PR (400-Env and 76-PR) sites.

3.4. Interactions between Envelope and Gag-PR Mutations with LPV/r Treatment

Using Bayesian network learning, we explored the interactions between Gag, PR, and Env mutations in isolates exposed to LPV/r treatment. Only statistically significant positions according to Fisher's exact test were included in the analysis. Figure 3A shows mutations in the *env* Gp120 region (S110N-Env, T132S-Env, T138S-Env, P183Q-Env, and Q315R-Env) that were indirectly associated with exposure to LPV/r treatment via known Gag-PR mutations (Q69K-Gag, S111I-Gag, I256V-Gag, and V77I-PR). More specifically, P183Q-Env was indirectly associated with LPV/r treatment via interaction with V77I-PR. T132S-Env also showed a robust connection with Q69K-Gag mutation, which is further related to K20R-PR mutation. Both T138S-Env and S110N-Env showed a strong association with I256V-Gag. Interestingly, a robust direct connection with LPV/r exposure was observed between a known Gag-PR mutation (R76K-Gag) and wildtype variants at codon positions in Gag (Q182-Gag-PR-WT) and PR that have been linked to drug resistance (I50-PR-WT, V77-PR-WT, and L90-PR-WT).

Figure 3B shows the association between Gag, PR, Env Gp41 mutations, and LPV/r treatment experience. There was a strong interaction between T536M-Env, S543A-Env, I688V-Env, and P724S-Env and Gag-PR mutations (S12T-PR, P63L-PR, Q182S-Gag, and I256V-Gag). Specifically, P724S-Env was directly associated with P63L-PR and Q182S-Gag and indirectly associated with A431V found at a Gag cleavage site. Another direct association was seen between S534A-Env and I256V-Gag. Both T536M-Env and I688V-Env were directly associated with S12T-PR. There was no direct interaction between treatment experience and Env Gp41 mutations.

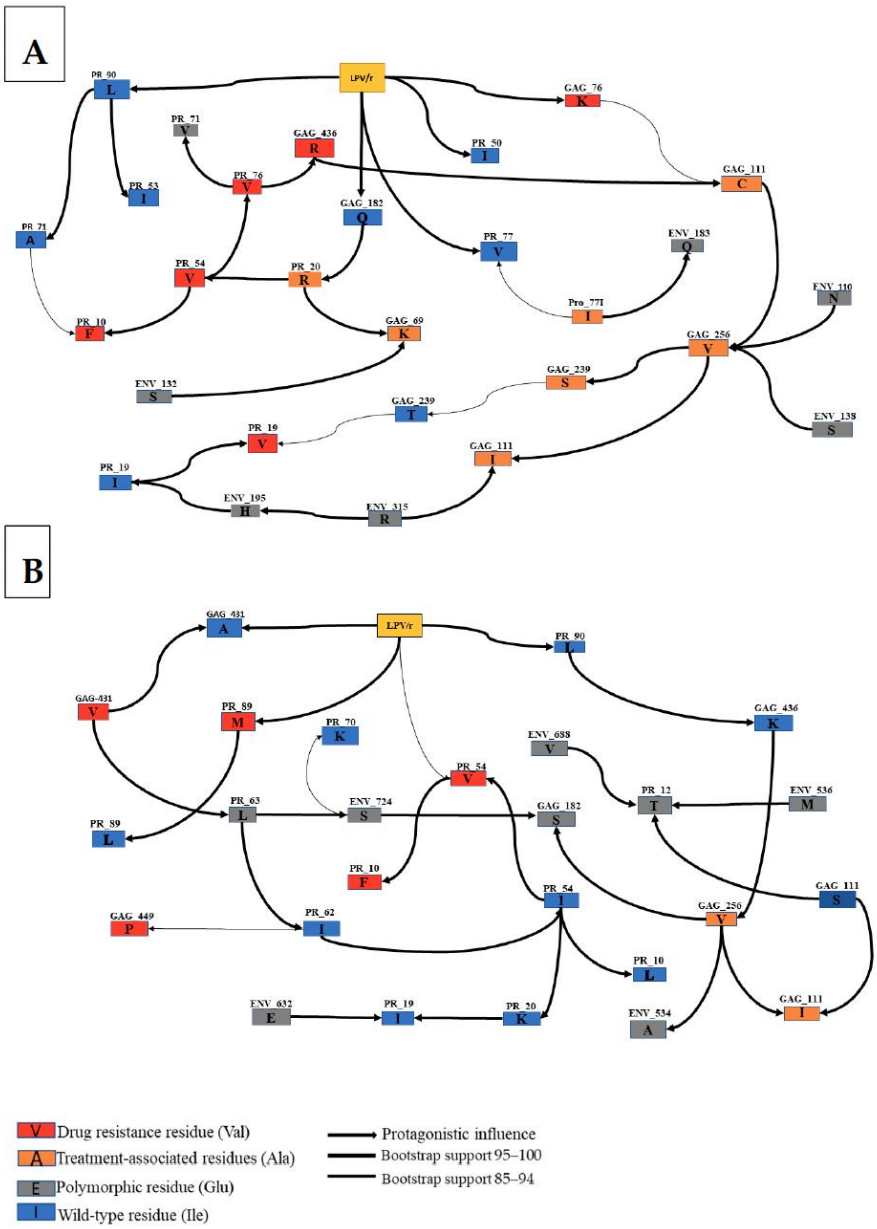


Figure 3. Annotated Bayesian network based on 5000 bootstraps, showing the association between nodes indicating Env and Gag mutations in subtype C LPV/r-treated sequences. Here, we determined the association between (**A**) Env Gp120, Gag, PR mutations, and treatment experience and (**B**) Gp41, Gag, PR, and treatment experience. The color code of nodes is defined based on their link with LPV/r: drug resistance variants (red), treatment associated variants (orange) and natural occurring: Polymorphic variants (gray), wild-type variants (blue). The arcs represent a direct dependency between corresponding nodes, and the thickness of an arc is in proportion to bootstrap support. The arc direction does not represent the accumulation of mutations or causal meaning but may indicate a multivariable effect in the network.

4. Discussion

Mutations in the *env* gene have previously been associated with resistance to PIs [5]; however, no studies have been conducted using subtype C PI failures, nor have they investigated their role in combination with the Gag. In this study, LPV/r-resistance-associated mutations in Gag, PR, and Env were characterized in HIV-1 subtype C patients from KZN, South Africa, who were failing an LPV/r-inclusive treatment regimen. We identified several amino acids in the *env* gene that were associated with LPV/r failure. We also found potential pathways leading to LPV/r resistance that involved a combination of Gag, PR, and Env mutations.

Significantly higher frequencies of AA changes in the GP120 region of the LPV/r-experienced group were commonly identified. These were mostly in the variable loops (V2: D167AR, V182I, P183QS, Y191F and N195H, and V3: S306R, P313DI, Q315R, I323KNR, A329DT, and C331RS); however, increased frequencies were also observed in the constant regions (C2: N230DT, Q258EP, N262LI, T283I, V286I and I294N, and C3: W338EG, H374KR, F382V, and F383LV). Amino acid changes in these regions can, therefore, potentially increase the infectivity and replication capacity of the virus [33].

In GP41, sequences from LPV/r failures harbored significantly higher frequencies of Heptad repeat (HR) mutations at codons 534, 536, 567, and 632. Consistent with our findings, an HIV-1 subtype A study also observed higher levels of mutations in HR in LPV/r failures [21]. Furthermore, higher mutation frequencies were also seen in the cytoplasmic tail (CT); this is in line with other studies that reported CT mutations associated with virological failure in PI-experienced participants [21–23]. As the CT plays a functional role in Env incorporation during virion assembly, this suggests that mutations in this region may influence the efficiency of Env incorporation [34].

Using Bayesian network learning, we found potential pathways leading to LPV/r resistance that involved a combination of Env and Gag mutations. I256V appeared to be an essential mutation as it was shown to be directly associated with three Env mutations (S110N-Env, T138S-Env, and S543A-Env) and indirectly associated with Q315R-Env via S111I-Gag and P724S-Env via Q182S-Gag. Although less is known about the I256V-Gag mutation, it has been previously associated with LPV/r failure in subtype C studies [35,36]. It has also been associated with reducing drug susceptibility to benzodiazepine and benzimidazole in Gag subtype B studies [37]. Mutations T138S-Env and S110N-Env showed a strong interaction with I256V-Gag, forming a pathway (T138S-Env + S110N-Env + I256V + S111C + R76K) to LPV/r exposure. A study by Singh et al. (2015) identified I256V, S111C, and R76K in Gag following LPV/r treatment in subtype C patients in South Africa [35]. Interestingly, T138 was also positively selected in our PAML analysis. Although a wildtype amino acid, T138, is known to play an essential role in antibody binding, and although the change from Threonine (T) to Serine (S) weakened the binding, T138S did not inhibit antibody binding [38].

Another potential pathway for LPV/r resistance included Env mutations Q315R-Env and the same Gag mutations I256V, S111C, and R76K, with the addition of S111I. While Q315R has been identified as a naturally occurring polymorphism in subtype A [39], its role as a drug-resistance-associated mutation in subtype C requires more mechanistic studies. Mutation T132S-Env showed a robust connection with Q69K-Gag, which has previously been associated with reduced susceptibility to PIs [40].

Interestingly, known PI-resistance-associated mutations K20R-PR, I54V-PR, L10F-PR, and L76V-PR were not directly connected to LPV/r exposure in the network, suggesting that immune selection pressure may also be in play.

While not directly associated with LPV/r resistance, Env Gp41 (S543A-Env and P724S-Env) mutations were associated with Gag (Q182S-Gag and I256V-Gag) mutations. Env mutation P724S was strongly associated with Q182S-Gag, which was further associated with I256V. S534A-Env was also directly associated with I256V-Gag, further supporting the role of I256V-Gag in connecting Env in the pathway to drug resistance. Interestingly, most

of the mutations in Gag connected to Env mutations were located in the capsid domain (such as I256V), which is similar to other studies [21].

Another potential network to LPV/r resistance involved mutations found in the *env* in combination with the *gag* and *PR* region. Env mutation P183Q-Env was directly associated with V77I-PR, previously associated with PI drug resistance [41–43]. In addition, Env mutations T536M and I688V showed a strong direct association with PR mutation S12T-PR, which is known to be selected in LPV/r- and RTV-treated patients in subtype C [40]. Interestingly, the P724S Env mutation was linked to LPV/r failure via the P63L-PR and A431V-Gag pathways. The A431V Gag mutation located in the CS of Gag is known to confer resistance to all PIs except Darunavir (DRV) [15], while P63L-PR was observed at baseline in patients initiating antiretroviral therapy (Lopinavir/r, Lamivudine, and Zidovudine) [15]. These findings suggest that there might be other pathways to LPV/r treatment failure that involve Gag (Matrix and Capsid), minor PR, and Env mutations.

5. Conclusions

We found a high prevalence of Env mutations in HIV-1 subtype C associated with LPV/r treatment failure. The majority of these associations were in combination with a Gag (Matrix and Capsid) and/or PR mutation. Further investigations using site-directed mutagenesis need to be conducted to determine whether Env mutations alone can affect LPV/r efficacy.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/microbiolres12040071/s1, Figure S1: Gel picture, showing the amplification of Gag-PR. Gag-PR size 1.7 kb. Figure S2: Gel picture showing the amplification of envelope. Envelope size 3.5 kb. Table S1: Sequencing Primers.

Author Contributions: Conceptualization, N.F.M., M.L.G.; data curation, N.F.M., M.L.G.; formal analysis, N.F.M.; funding acquisition, M.L.G.; investigation, N.F.M.; methodology, N.F.M.; resources, M.L.G.; writing of the manuscript, N.F.M., M.L.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Funding, Poliomyelitis Research Foundation, and the UKZN College of Health Sciences.

Institutional Review Board Statement: This study was approved by the University of KwaZulu-Natal Biomedical Research Ethics Committee (BREC NO. 678/17).

Informed Consent Statement: Consent was not applicable since the study was conducted using the stored plasma samples.

Data Availability Statement: Data are contained within the article and the Supplementary Material; further information can be obtained from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ghosh, A.K.; Chapsal, B.D.; Weber, I.T.; Mitsuya, H. Design of HIV protease inhibitors targeting protein backbone: An effective strategy for combating drug resistance. *Acc. Chem. Res.* **2008**, *41*, 78–86. [CrossRef]
2. Kim, R.; Baxter, J.D. Protease inhibitor resistance update: Where are we now? *AIDS Patient Care STDs* **2008**, *22*, 267–277. [CrossRef]
3. Pettit, S.C.; Moody, M.D.; Wehbie, R.S.; Kaplan, A.H.; Nantermet, P.V.; Klein, C.A.; Swanstrom, R. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.* **1994**, *68*, 8017–8027. [CrossRef] [PubMed]
4. Lori, F.; Scovassi, A.I.; Zella, D.; Achilli, G.; Cattaneo, E.; Casoli, C.; Bertazzoni, U. Enzymatically active forms of reverse transcriptase of the human immunodeficiency virus. *AIDS Res. Hum. Retrovir.* **1988**, *4*, 393–398. [CrossRef]
5. Rabi, S.A.; Laird, G.M.; Durand, C.M.; Laskey, S.; Shan, L.; Bailey, J.R.; Chioma, S.; Moore, R.D.; Siliciano, R.F. Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. *J. Clin. Investig.* **2013**, *123*, 3848–3860. [CrossRef] [PubMed]
6. Zdanowicz, M.M. The pharmacology of HIV drug resistance. *Am. J. Pharm. Educ.* **2006**, *70*, 100. [CrossRef] [PubMed]
7. Gallego, O.; de Mendoza, C.; Pérez-Elías, M.J.; Guardiola, J.M.; Pedreira, J.; Dalmau, D.; Gónzalez, J.; Moreno, A.; Arribas, J.R.; Rubio, A. Drug resistance in patients experiencing early virological failure under a triple combination including indinavir. *Aids* **2001**, *15*, 1701–1706. [CrossRef]

8. Wallis, C.L.; Mellors, J.W.; Venter, W.D.; Sanne, I.; Stevens, W. 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. Treat.* **2011**, *2011*, 769627. [CrossRef]
9. Obasa, A.E.; Mikasi, S.G.; Brado, D.; Cloete, R.; Singh, K.; Neogi, U.; Jacobs, G.B. Drug resistance mutations against protease, reverse transcriptase and integrase inhibitors in people living with HIV-1 receiving boosted protease inhibitors in South Africa. *Front. Microbiol.* **2020**, *11*, 438. [CrossRef]
10. Parry, C.M.; Kolli, M.; Myers, R.E.; Cane, P.A.; Schiffer, C.; Pillay, D. Three residues in HIV-1 matrix contribute to protease inhibitor susceptibility and replication capacity. *Antimicrob. Agents Chemother.* **2011**, *55*, 1106–1113. [CrossRef]
11. Alfadhli, A.; Mack, A.; Ritchie, C.; Cylinder, I.; Harper, L.; Tedbury, P.R.; Freed, E.O.; Barklis, E. Trimer Enhancement Mutation Effects on HIV-1 Matrix Protein Binding Activities. *J. Virol.* **2016**, *90*, 5657–5664. [CrossRef] [PubMed]
12. Fun, A.; Wensing, A.M.; Verheyen, J.; Nijhuis, M. Human Immunodeficiency Virus Gag and protease: Partners in resistance. *Retrovirology* **2012**, *9*, 63. [CrossRef] [PubMed]
13. Codoñer, F.M.; Peña, R.; Blanch-Lombarte, O.; Jimenez-Moyano, E.; Pino, M.; Vollbrecht, T.; Clotet, B.; Martinez-Picado, J.; Draenert, R.; Prado, J.G. Gag-protease coevolution analyses define novel structural surfaces in the HIV-1 matrix and capsid involved in resistance to Protease Inhibitors. *Sci. Rep.* **2017**, *7*, 3717. [CrossRef] [PubMed]
14. Malet, I.; Roquebert, B.; Dalban, C.; Wirden, M.; Amellal, B.; Agher, R.; Simon, A.; Katlama, C.; Costagliola, D.; Calvez, V. Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. *J. Infect.* **2007**, *54*, 367–374. [CrossRef]
15. Nijhuis, M.; Van Maarseveen, N.M.; Lastere, S.; Schipper, P.; Coakley, E.; Glass, B.; Rovenska, M.; De Jong, D.; Chappey, C.; Goedegebuure, I.W. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med.* **2007**, *4*, e36. [CrossRef] [PubMed]
16. Verheyen, J.; Litau, E.; Sing, T.; Daumer, M.; Balduin, M.; Oette, M.; Fatkenheuer, G.; Rockstroh, J.K.; Schuldenzucker, U.; Hoffmann, D. Compensatory mutations at the HIV cleavage sites p7/p1 and p1/p6-gag in therapy-naive and therapy-experienced patients. *Antivir. Ther.* **2006**, *11*, 879.
17. Kletenkov, K.; Hoffmann, D.; Böni, J.; Yerly, S.; Aubert, V.; Schöni-Affolter, F.; Struck, D.; Verheyen, J.; Klimkait, T.; Study, S.H.C. Role of Gag mutations in PI resistance in the Swiss HIV cohort study: Bystanders or contributors? *J. Antimicrob. Chemother.* **2017**, *72*, 866–875. [CrossRef]
18. Larrouy, L.; Chazallon, C.; Landman, R.; Capitant, C.; Peytavin, G.; Collin, G.; Charpentier, C.; Storto, A.; Pialoux, G.; Katlama, C. Gag mutations can impact virological response to dual-boosted protease inhibitor combinations in antiretroviral-naive HIV-infected patients. *Antimicrob. Agents Chemother.* **2010**, *54*, 2910–2919. [CrossRef]
19. Gatanaga, H.; Suzuki, Y.; Tsang, H.; Yoshimura, K.; Kavlick, M.F.; Nagashima, K.; Gorelick, R.J.; Mardy, S.; Tang, C.; Summers, M.F. 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.* **2002**, *277*, 5952–5961. [CrossRef]
20. Kožíšek, M.; Henke, S.; Šašková, K.G.; Jacobs, G.B.; Schuch, A.; Buchholz, B.; Müller, V.; Kräusslich, H.-G.; Řezáčová, P.; Konvalinka, J. Mutations in HIV-1 gag and pol compensate for the loss of viral fitness caused by a highly mutated protease. *Antimicrob. Agents Chemother.* **2012**, *56*, 4320–4330. [CrossRef]
21. Coetzer, M.; Ledingham, L.; Diero, L.; Kemboi, E.; Orido, M.; Kantor, R. Gp41 and Gag amino acids linked to HIV-1 protease inhibitor-based second-line failure in HIV-1 subtype A from Western Kenya. *J. Int. AIDS Soc.* **2017**, *20*, e25024. [CrossRef] [PubMed]
22. Castain, L.; Perrier, M.; Charpentier, C.; Palich, R.; Desire, N.; Wirden, M.; Descamps, D.; Sayon, S.; Landman, R.; Valantin, M.-A. New mechanisms of resistance in virological failure to protease inhibitors: Selection of non-described protease, Gag and Gp41 mutations. *J. Antimicrob. Chemother.* **2019**, *74*, 2019–2023. [CrossRef] [PubMed]
23. Perrier, M.; Castain, L.; Regad, L.; Todesco, E.; Landman, R.; Visseaux, B.; Yazdanpanah, Y.; Rodriguez, C.; Joly, V.; Calvez, V. HIV-1 protease, Gag and gp41 baseline substitutions associated with virological response to a PI-based regimen. *J. Antimicrob. Chemother.* **2019**, *74*, 1679–1692. [CrossRef]
24. Van Duyne, R.; Kuo, L.S.; Pham, P.; Fujii, K.; Freed, E.O. Mutations in the HIV-1 envelope glycoprotein can broadly rescue blocks at multiple steps in the virus replication cycle. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 9040–9049. [CrossRef]
25. Hikichi, Y.; Van Duyne, R.; Pham, P.; Groebner, J.L.; Wiegand, A.; Mellors, J.W.; Kearney, M.F.; Freed, E.O. Mechanistic Analysis of the Broad Antiretroviral Resistance Conferred by HIV-1 Envelope Glycoprotein Mutations. *Mbio* **2021**, *12*, e03134-20. [CrossRef]
26. Marie, V.; Gordon, M. Gag-protease coevolution shapes the outcome of lopinavir-inclusive treatment regimens in chronically infected HIV-1 subtype C patients. *Bioinformatics* **2019**, *35*, 3219–3223. [CrossRef] [PubMed]
27. Rangel, H.R.; Weber, J.; Chakraborty, B.; Gutierrez, A.; Marotta, M.L.; Mirza, M.; Kiser, P.; Martinez, M.A.; Este, J.A.; Quinones-Mateu, M.E. Role of the human immunodeficiency virus type 1 envelope gene in viral fitness. *J. Virol.* **2003**, *77*, 9069–9073. [CrossRef]
28. Deforche, K.; Camacho, R.; Grossman, Z.; Silander, T.; Soares, M.; Moreau, Y.; Shafer, R.W.; Va http://doi.org/10.1073/pnas.1820 Wynhoven, B. Bayesian network analysis of resistance pathways against HIV-1 protease inhibit 382–390. [CrossRef]
29. Theys, K.; Deforche, K.; Libin, P.; Camacho, R.J.; Van Laethem, K.; Vandamme, A.-M. Resistance pathways of human immunodeficiency virus type 1 against the combination of zidovudine and lamivudine. *J. Gen. Virol.* **2010**, *91*, 1898–1908. [CrossRef]

30. Fares, M.A.; Travers, S.A. A novel method for detecting intramolecular coevolution: Adding a further dimension to selective constraints analyses. *Genetics* **2006**, *173*, 9–23. [CrossRef]
31. Yang, Z.; Wong, W.S.; Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118. [CrossRef]
32. Anisimova, M.; Bielawski, J.P.; Yang, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **2001**, *18*, 1585–1592. [CrossRef]
33. Dang, L.V.P.; Pham, H.V.; Dinh, T.T.; Nguyen, T.H.; Vu, Q.T.H.; Vu, N.T.P.; Le, P.T.B.; Nguyen, L.V.; Le, H.T.; Vu, P.T. Characterization of envelope sequence of HIV virus in children infected with HIV in Vietnam. *SAGE Open Med.* **2020**, *8*, 2050312120937198. [CrossRef] [PubMed]
34. Murphy, R.E.; Samal, A.B.; Vlach, J.; Saad, J.S. Solution structure and membrane interaction of the cytoplasmic tail of HIV-1 gp41 protein. *Structure* **2017**, *25*, 1708–1718.e1705. [CrossRef] [PubMed]
35. Singh, U. Acquired and Transmitted Drug Resistance in HIV-1 Subtype C: Implications of Novel Mutations on Replication Capacity, Cleavage and Drug Susceptibility. Ph.D. Thesis, University of Kwazulu-Natal, Durban, South Africa, 2015.
36. Ntale, R.S. The Role of Early Cytotoxic Lymphocyte (CTL) Escape in the Pathogenesis of HIV-1 Subtype C Infection. Ph.D. Thesis, University of Cape Town, Cape Town, South Africa, 2012.
37. Li, G.; Verheyen, J.; Rhee, S.-Y.; Voet, A.; Vandamme, A.-M.; Theys, K. Functional conservation of HIV-1 Gag: Implications for rational drug design. *Retrovirology* **2013**, *10*, 126. [CrossRef] [PubMed]
38. Hyser, J.M.; Zeng, C.Q.-Y.; Beharry, Z.; Palzkill, T.; Estes, M.K. Epitope mapping and use of epitope-specific antisera to characterize the VP5* binding site in rotavirus SA11 NSP4. *Virology* **2008**, *373*, 211–228. [CrossRef]
39. Ratcliff, A.N.; Shi, W.; Arts, E.J. HIV-1 resistance to maraviroc conferred by a CD4 binding site mutation in the envelope glycoprotein gp120. *J. Virol.* **2013**, *87*, 923–934. [CrossRef]
40. Giandhari, J.; Basson, A.E.; Sutherland, K.; Parry, C.M.; Cane, P.A.; Coovadia, A.; Kuhn, L.; Hunt, G.; Morris, L. Contribution of Gag and protease to HIV-1 phenotypic drug resistance in pediatric patients failing protease inhibitor-based therapy. *Antimicrob. Agents Chemother.* **2016**, *60*, 2248–2256. [CrossRef]
41. Santoro, M.M.; Perno, C.F. HIV-1 genetic variability and clinical implications. *Int. Sch. Res. Not.* **2013**, *2013*, 481314. [CrossRef]
42. Mata-Munguía, C.; Escoto-Delgadillo, M.; Torres-Mendoza, B.; Flores-Soto, M.; Vázquez-Torres, M.; Gálvez-Gastelum, F.; Viniegra-Osorio, A.; Castillero-Manzano, M.; Vázquez-Valls, E. Natural polymorphisms and unusual mutations in HIV-1 protease with potential antiretroviral resistance: A bioinformatic analysis. *BMC Bioinform.* **2014**, *15*, 1–17. [CrossRef]
43. Doualla-Bell, F.; Avalos, A.; Gaolathe, T.; Mine, M.; Gaseitsiwe, S.; Ndwapi, N.; Novitsky, V.A.; Brenner, B.; Oliveira, M.; Moisi, D. Impact of human immunodeficiency virus type 1 subtype C on drug resistance mutations in patients from Botswana failing a nelfinavir-containing regimen. *Antimicrob. Agents Chemother.* **2006**, *50*, 2210–2213. [CrossRef] [PubMed]