Investigating the role of Gag in protease inhibitor susceptibility amongst West African HIV-1 subtypes

Rawlings Walshak Philemon Datir

Division of Infection and Immunity

University College London

A thesis submitted for the degree of Doctor of Philosophy

August 2020

Declaration

I, Rawlings W.P Datir confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

Abstract

HIV-1 Gag contributes to susceptibility of protease inhibitors (PIs) in the absence of known resistance mutations in the protease gene. For the majority of HIV-infected patients worldwide, PIs are the second, and last-line of therapy. Clinically, only around 20% of individuals who fail PI regimen develop major resistance mutations in protease. We previously showed that full length Gag-protease-derived phenotypic susceptibility to PIs differed between HIV-1 CRF02_AG and subtype G-infected patients who went on to successfully suppress viral replication versus those who experienced virological failure of boosted lopinavir monotherapy as first-line treatment in a clinical trial. We hypothesised therefore that baseline PI susceptibility by Gag-protease phenotyping could be used to predict treatment outcomes for patients on second line, boosted-PI treatment in the real-world clinical setting in Nigeria, where subtypes CRF02_AG/G dominate the epidemic. We used clinical and demographic data; HIV-1subtype, sex, age, viral load, duration of treatment and baseline CD4 count to match individuals who experienced second-line failure with ritonavir-boosted PI-based ART ('baseline failures') to those who achieved virological response ('baseline successes') with virological failure defined by viral load <400 copies of HIV-1 RNA/mL by week 48. Using a single replication-cycle assay, we carried out in vitro phenotypic susceptibility testing of patient-derived viruses from these two groups. We found no impact of baseline HIV-1 Gag-protease-derived phenotypic susceptibility on outcomes of PI-based second-line ART, treatment outcome could not be predicted using baseline susceptibility alone.

Secondly, we sought to explore the role of mutation in Gag-protease genotypic and phenotypic changes within patients who failed PI-based regimens without known drug resistance associated protease mutations in order to identify novel determinants of PI resistance. We used longitudinal samples collected at baseline, and at virological failure to explore the role of Gag mutations. Using target enrichment and next generation sequencing (NGS), followed by haplotype reconstruction and phenotypic drug assays and phylogenetic analysis, we reported for the first time a four-amino acid mutation signature in HIV-1,

CRF02_AG matrix (S126del, H127del, T122A and G123E) which confer reduced susceptibility to the PI, lopinavir and atazanavir.

Our multi-pronged genotypic and phenotypic approach to document emergence and temporal dynamics of a novel protease inhibitor resistance signature in HIV-1 matrix domain reveals the interplay between Gag associated resistance and fitness.

Impact Statement

In 2019, an estimated 690,000 people died of HIV/AIDS-related illnesses worldwide. About 25.4 million people were on antiretroviral therapy in 2019, compared with just 6.4 million in 2009. Sub-Saharan Africa bears the highest burden of HIV/AIDS as 67.4% of the estimated 38 million people living with HIV are from that region, with 54.5% in East & Southern Africa and 12.9% in West & Central Africa.

As antiretroviral therapy (ART) scale up progresses globally in the absence of universal viral load monitoring, significant numbers of persons living with HIV (PLWH) are experiencing virological failure (VF) with emergent drug resistance. Transmitted/pre-treatment resistance is on the increase. Protease inhibitors (PIs) are used exclusively as second-line treatment for patients who have developed resistance to and have failed first line treatment. Studies demonstrate that the detection of major canonical protease mutations is around 20% in PLWH treated with PI-containing combination ART raising the question of how virologic failure occurs in the remaining cases. Inadequate adherence to medication has been implicated but determinants of susceptibility outside the protease gene have also been considered. Mutations in HIV-1 envelope and gag genes have been suggested to impact susceptibility to PIs although the determinants of failure still remain elusive. Determinants of PI failure has been further complicated by subtype-specific differences in HIV-1 response to antiretroviral drugs. For example, it was previously shown that CRF02_AG and G subtypes appear inherently more resistant to some PIs.

In this present study, we provided further evidence on the role of Gag in protease inhibitor susceptibility. Our findings also contributed to the knowledge that naturally occurring Gag polymorphisms in CRF02_AG and G subtypes may be responsible for the inherently reduced susceptibility to some protease inhibitors. More importantly, we identified for the first time, previously undescribed amino acid mutation signatures in HIV-1 matrix, which reduces susceptibility to the protease inhibitors, lopinavir and atazanavir in the absence of major drug

resistance-associated mutations in the protease gene. Our findings from this study have served as strong support by providing preliminary data for a 5-year US-National Institute of Allergy and Infectious Diseases (NIAID)-NIH grant (R01AI147331) to our collaborators at the Institute of Human Virology Nigeria. The grant titled: Impact of non-B HIV-1 Subtypes on second-line Protease Inhibitor Regimens in Africa (INSPIRE) would expand on the findings in this project on a larger scale, in a real-world clinical setting.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A	Alanine (amino acid)
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Treatment
APV	Amprenavir
ATV	Atazanavir
AZT	Zidovudine
bp	Base pairs
CA	Capsid subunit
c ART	Combination Antiretroviral Treatment
cm	centimetre
CSM	Cleavage Site Mutations
D	Aspartic acid (amino acid)
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic Acid
DRV	Darunavir
d4T	Stavudine
E	Glutamic acid (amino acid)
EM	Electron microscopy
Env	Envelope
G	Glycine (amino acid)
GFP	Green fluorescent protein
gp	Glycoprotein
HIV	Human immunodeficiency virus
I	Isoleucine (amino acid)
IDV	Indinavir
IF	Immunofluorescence
INSTI	Integrase Strand Transfer Inhibitor
K	Lysine (amino acid)
KB	Kilobases
L	Leucine (amino acid)
LPV	Lopinavir
MA	Matrix subunit
ml	millilitre
mm	millimetre
N	Asparagine (amino acid)
NC	Nucleocapsid subunit
NCSM	Non-Cleavage Site Mutations
ng	Nanogram
nM	Nanomolar
nm	Nanometer
nt	Nucleotide
OD	Optical density
P	Proline (amino acid)
PBS	Phosphate-buffered saline
pbs	Primer binding site
PI	Protease inhibitor

Pol	Polymerase
Pro	Protease
Q	Glutamine (amino acid)
R	Arginine (amino acid)
RC	Replicative capacity
RT	Reverse transcriptase
RTV	Retroviral Vector
RNA	Ribonucleic Acid
S	Serine (amino acid)
SIV	Simian immunodeficiency virus
SQV	saquinavir
T	Threonine (amino acid)
TPV	Tipranavir
UV	Ultraviolet
V	Valine (amino acid) OR vaults
WHO	World Health Organization
X	Times
µg	microgram
µl	microlitre
3TC	Lamivudine

CHAPTER 1: Introduction and Literature Review

1.1   The History of the Discovery of HIV

In 1981, a rare form of *Pneumocystis carinii* Pneumonia (PCP) was identified among a group of young gay men in Los Angeles, USA[1]. This form of pneumonia was relatively rare and ordinarily asymptomatic. Similar clusters of PCP was soon identified among other groups of gay men in New York and San Francisco[2-4], this time around, Kaposi Sarcoma (a rare infection-related cancer) was also found among the group. Kaposi Sarcoma and *Pneumocystis carinii* Pneumonia (PCP) are usually asymptomatic in immunocompetent individuals, but among this group of young men, the mortality rate was astoundingly high. Medical examination of these patients revealed a common denominator – all patients had depleted CD4+ T-cells, leading to markedly high reductions in the cellular immune response. These rare conditions were later found to arise due to severe immunodeficiency, but the cause remained unknown [5]. Later on, these diseases were reported amongst heterosexual individuals and patients who have received multiple blood transfusions as well as intravenous drug users. Additionally, the prevalence of the disease was particularly high among USA immigrants from Haiti[6-9].

The term acquired immune deficiency syndrome (AIDS) was introduced by the Centres for Disease Control (CDC) in 1982 to describe this new disease. The CDC defined AIDS as: "the occurrence of a disease likely to be caused by diminished cell-mediated immunity, such as PCP or KS, where no other reason for immunodeficiency could be identified"[10]. The epidemiology of AIDS suggested that it was caused by an infection; however, the causative agent was still not known.

At the Institut Pasteur in Paris, France, Luc Montaignier's team isolated a novel retrovirus from a patient suffering from lymphadenopathy in 1983. This novel virus was named Lymphadenopathy Associated Virus (LAV)[11]. Meanwhile, at the National Institutes of Health (NIH) in the USA, while working on samples from AIDS patients, Robert Gallo's team isolated a novel virus from the T-cells of 48 patients with AIDS or at risk of developing AIDS. This novel virus was called human T-lymphotropic virus type III (HTLV-III) [12]. The suggestion then was that

these two novel viruses (LAV and HTLV-III) were the possible causal agents of AIDS. It did not take long to realize that the viruses described by Montaignier and Gallo were indeed the same virus[13]. In 1986, the causal agent of AIDS was renamed as Human Immunodeficiency Virus (HIV) by the International Committee on the Taxonomy of Viruses[14].

1.2 The Burden of HIV/AIDS Globally, and in Nigeria

An estimated 75.7 million people have been infected with HIV, and 32.7 million have died from AIDS-related illnesses since the beginning of the epidemic. At the end of 2019, an estimated 38.0 million people were living with HIV worldwide, with 1.7 million new infections[15] – as summarized in Table 1.1. In 2019, estimated deaths from AIDS-related illnesses worldwide was 690,000. Although this is a huge number, it was low when compared to the estimated 1.1 million AIDS-related deaths in 2010 and 1.7 million deaths in 2004. This decline could be attributed to improved access to antiretroviral therapy which has led to an increase in the life expectancy of infected individuals. At the end of 2019, 25.4 million people were estimated to be on antiretroviral therapy compared with just 6.4 million in 2009. AIDS-related deaths have been reduced by 60% since the peak in 2004[15].

Sub-Saharan Africa bears the highest burden of HIV/AIDS as 67.4% of the estimated 38 million people living with HIV are from that region, with 54.5% in East & Southern Africa and 12.9% in West & Central Africa[15].

In Nigeria, a country of about 200 million people, the 2012 national HIV prevalence rate in the general population, according to the National HIV&AIDS and Reproductive Health Survey (NARHS) was 3.4%[16]. In 2018, the Nigeria National HIV/AIDS Indicator and Impact Survey (NAIIS) was carried out based on a revised and enhanced methodology. About 1.9 million people were living with HIV (national prevalence of 1.5%), 130,000 new infections and 53,000 deaths from AIDS-related illness[17,18]. From the same survey, 67% of people living with HIV knew their status, 53% of people living with HIV were on treatment with 42% of them virologically suppressed (viral load <1000 copies/ml).

Table 1.1:   2019 Regional Data of Global HIV burden

Region	People living with HIV 2019	New HIV infections 2019			AIDS-related deaths 2019	People accessing treatment 2019
		Total	Aged 15+	Aged 0–14		
Eastern and southern Africa	20.7 million [18.4 million– 23.0 million]	730 000 [580 000– 940 000]	660 000 [520 000– 850 000]	74 000 [50 000– 120 000]	300 000 [230 000– 390 000]	15.0 million [14.4 million– 15.1 million]
Asia and the Pacific	5.8 million [4.3 million– 7.2 million]	300 000 [210 000– 390 000]	280 000 [200 000– 370 000]	15 000 [8900– 22 000]	160 000 [94 000– 240 000]	3.5 million [3.4 million– 3.5 million]
Western and central Africa	4.9 million [3.9 million– 6.2 million]	240 000 [150 000– 390 000]	190 000 [120 000– 310 000]	52 000 [30 000– 83 000]	140 000 [100 000– 210 000]	2.9 million [2.8 million– 2.9 million]
Latin America	2.1 million [1.4 million– 2.8 million]	120 000 [73 000– 180 000]	120 000 [71 000– 170 000]	3400 [2100– 5500]	37 000 [23 000– 56 000]	1.3 million [1.2 million– 1.3 million]
The Caribbean	330 000 [270 000– 400 000]	13 000 [8700– 19 000]	12 000 [8 000– 17 000]	960 [610– 1500]	6900 [4900– 10 000]	210 000 [200 000– 210 000]
Middle East and North Africa	240 000 [170 000– 400 000]	20 000 [11 000– 38 000]	18 000 [9500– 36 000]	1400 [920– 2100]	8000 [4900– 14 000]	92 000 [88 000– 93 000]
Eastern Europe and central Asia	1.7 million [1.4 million– 1.9 million]	170 000 [140 000– 190 000]	160 000 [140 000– 190 000]	—*	35 000 [26 000– 45 000]	730 000 [710 000– 740 000]
Western and central Europe and N America	2.2 million [1.7 million– 2.6 million]	65 000 [49 000– 87 000]	65 000 [48 000– 87 000]	—*	12 000 [8700– 19 000]	1.8 million [1.7 million– 1.8 million]
Global totals	38.0 million [31.6 million– 44.5 million]	1.7 million [1.2 million- 2.2 million]	1.5 million [1.1 million– 2.0 million]	150 000 [94 000– 240 000]	690 000 [500 000– 970 000]	25.4 million [24.5 million– 25.6 million]

* Estimates for children are not published because of the small numbers.

Source: UNAIDS, 2020 available at

https://www.unaids.org/en/resources/fact-sheet

1.3   Genetic diversity and geographical distribution of HIV-1

The International Committee on Taxonomy of Viruses classifies HIV as belonging to the family Retroviridae[19]. Retroviridae family can be sub-divided into; Spumavirinae and Orthoretrovirinae (where the genus Lentivirus belongs) based upon morphology and genome sequence. Retroviruses are in group six of the Baltimore classification system, with all retroviruses having a positive-sense single-stranded RNA (ssRNA) genome[20]. A prolonged incubation time between infection and disease characterises lentiviruses [21]. The genome of lentiviruses is carried as two non-identical, single strands of positive-sense RNA. For replication to occur, the genomic RNA must be reverse transcribed into a linear double-stranded DNA copy. The DNA is then integrated into the host cell genome [22]. The "central dogma" of molecular biology is that genetic information is transferred in a hierarchy: DNA$\rightarrow$ RNA$\rightarrow$protein[23]. With lentiviruses, this central dogma is however challenged. Lentiviruses cause chronic persistent infection in equines, primates, bovines, ovines and felines[21,24-26].

HIV is broadly divided into HIV-1 (responsible for the majority of global infections) and HIV-2 (which is almost limited to a few countries in West Africa). HIV-1 viruses are further divided into four groups: Group Major (M), group Outlier (O), group N (non-M, non-O) and group P (pending identification of further human cases). HIV-1 group M viruses are responsible for the majority of infections worldwide. HIV-1 groups M and N originated from SIV infecting chimpanzees (*Pan troglodytes troglodytes*)[27] and O and P from SIV infecting wild gorillas[28,29]. Viruses from groups N and P are the most recently described groups. They were reported to be infecting a small number of individuals in the West-African country of Cameroon[30,31]. Similarly, viruses from group O were also first identified in Cameroon, where they cause up to 2% of the total number of infections in that country [32].

HIV-1 group M has evolved to be the pandemic strain and exhibits a substantial degree of diversity. Nine (9) different subtypes have been classified within this group. These subtypes are: A, B, C, D, F, G, H, J and K. The variation in amino acid level between these subtypes ranges between 17 -35% [33-35]. Four (4) sub-

sub types were previously classified; A1, A2, F1 and F2 [36]. Recently, in 2018, Désiré and colleagues proposed the addition of seven (7) more sub-sub types [37].

When a patient is dually infected with two different subtypes, recombination can occur between these viruses *in vivo*. Such recombinations give rise to the production of unique recombinant forms (URFs)[38]. Whenever URFs are reported in at least three individuals who are considered epidemiologically distinct, the URFs are then called Circulating Recombinant Forms (CRFs)[39]. CRFs are numbered according to the order that they were discovered or reported of their discovery.

Figure 1.1. shows the global distribution of HIV-1 subtypes. Interestingly, global subtype distribution is such that specific subtypes predominate particular geographical regions, with West and Central Africa showing the highest diversity. Subtype C predominated in southern Africa and accounted for about half of all HIV-1 infections worldwide in 2010–15. In the western world (North America, the Caribbean, Latin America, Western and Central Europe and Australia), B viruses are the dominant subtypes whereas subtype A and various CRFs dominate in Asia. In the 2010 -2015 period, subtypes B and A were responsible for 12·1% and 10·3% of infections, respectively, followed by circulating recombinant forms (CRFs), CRF02_AG and CRF01_AE, subtype G, and subtype D respectively. Subtypes F, H, J, and K combined accounted for 0·9% of infections globally. Other CRFs accounted for 3·7%, bringing the proportion of all CRFs to 16·7%. Unique recombinant forms (URFs) constituted 6·1%, resulting in recombinants accounting for 22·8% of all global HIV-1 infections[35]. The HIV-1 circulating recombinant, CRF02_AG accounts for about 50% of all HIV-1 infections in West Africa[32,38,40-51], with subtype G as the second most prevalent, accounting for nearly 30% of infections in the region[52].

Although the global and regional distribution of HIV-1 depends on many factors—such as HIV-1 prevalence, global mobility, socioeconomic factors, differences in the characteristics of diverse strains, and differences in access to antiretroviral treatment and clinical care—in several cases, the proportion of recombinants

increases over different subtypes. This increase could be due to differences in transmissibility or replicative capacity of recombinant viruses versus the so-called pure subtypes[53].
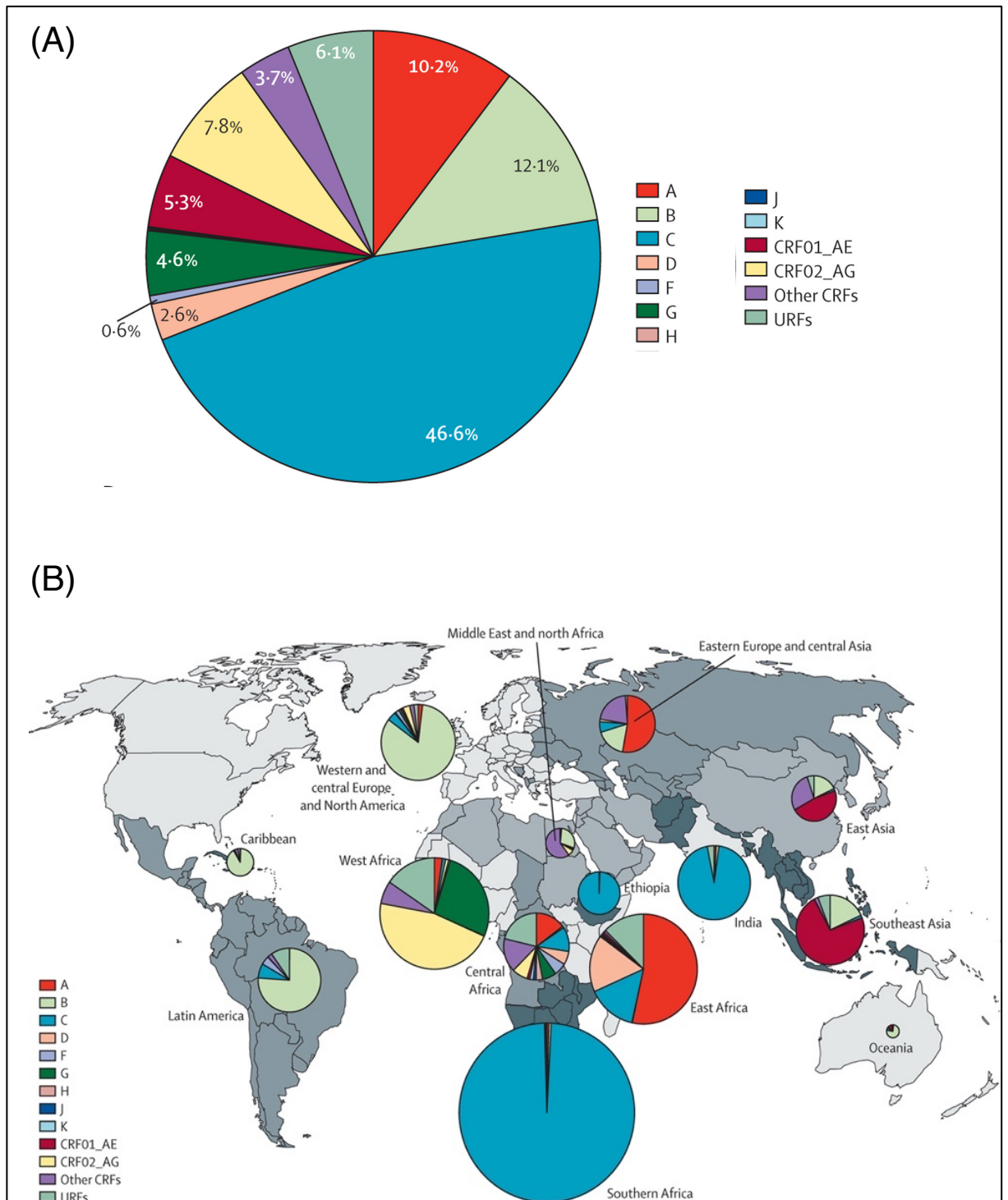
Figure 1.1: Distribution of HIV-1 Subtypes, CRFs and URFs, 2010 - 2015
(A) The pie chart represents the global frequency of each subtype (B) Global distribution by regions. The distribution of subtypes shows a remarkable difference between geographical regions across the world. At 47%, Southern Africa (predominated by subtype C) had the majority of infections worldwide. While CRF02_AG and G subtypes (occurring in West and Central Africa) accounted for 13% infections worldwide. (adapted from Hemelaar, 2019)[35].

First described in 1986, the origin of HIV-2 is thought to be from sooty mangabey monkeys rather than chimpanzees. Infection with HIV-2 have primarily been reported in the West African countries of Senegal, Guinea Bissau, Mali, Mauritania, Nigeria and Sierra Leone [54], with a few cases reported in western Europe and America[54,55]. HIV-2 is divided into groups A–H, with groups A and B being the most prevalent. Recombinations between HIV-2 groups are rare, although they have been described[56].

HIV-2 is less pathogenic than HIV-1, which results in a prolonged period until the signs of immunodeficiency and AIDS develop. Mother-to-child transmission rate of around 2 - 7%, compared to 10–40% in those infected with HIV-1. It is less infectious and most individuals have a lower viral load in comparison with those infected with HIV-1[54].

1.4   Clinical HIV infection

HIV is present in the blood and bodily fluids of an infected individual, so can be transmitted through sexual contact, parenterally, perinatally or through breastfeeding. Shortly after infection, individuals become viraemic, and at this stage, HIV is detectable in the plasma by nucleic acid amplification of viral RNA or detection of the viral core protein p24. About 4–6 weeks after infection, antibodies to HIV become detectable. Most infected individuals seroconvert by three months and become HIV antibody-positive; rarely, this takes up to 6 months[57]. Cells which express CD4 and chemokine receptors (CCR5 and CXCR4) such as dendritic cells, macrophages, T-cells and monocytes are targeted by HIV. Few reports have also been published showing that HIV may also infect cells which lack these receptors (such as astrocytes and renal epithelial cells) [58,59]. Other researchers have, however, not been able to replicate these studies. When infected, the natural course of disease progression proceeds differently in different individuals. In the absence of any treatment, some infected persons maintain high CD4 cell counts and suppressed viral loads. Although similar, the underlying mechanistic processes leading to long-term non-progression and viral control also differs between individuals[60].

HIV infection occurs when a small number of viruses (founder viruses) are transmitted from an infected person to an uninfected person. Initially, there are high levels of circulating HIV, the result of rapid replication in infected cells. During this period, patients can be symptomatic with features of primary HIV infection. Initial transmission is typically followed by a period of rapid viral replication and massive inflammatory response by the host[61,62]. After specific antibodies have developed, viral levels decline to reach a steady state known as "virologic set point" or setpoint viral load (SPVL). Patients may generally remain asymptomatic for several years. Over time, CD4 lymphocyte numbers gradually decline because of viral killing, apoptosis and activation of CD8 lymphocytes. CD4 cell levels eventually decline to a point where cell-mediated immunity is affected, and the individual becomes susceptible to opportunistic infections, HIV-associated nephropathy, dementia and cancers[63].

Adult HIV patients are generally classified into four clinical stages by the world health organization (WHO) with stage one being the early asymptomatic stage and stage four being the AIDS stage. The peak of HIV RNA in the blood is reached six weeks post-infection after which RNA level begins to decline until the SPVL is established 9-12 weeks post-infection. Typically, a prolonged period of clinical latency follows. This could range from a couple of months, to a few years. There is a progressive loss of CD4 T cells in the blood, which eventually leads to immunodeficiency. This progressive depletion of CD4 T cells leads to catastrophic immune dysfunction, which is the hallmark of HIV-infection. The most significant observed depletion of CD4 T cells has been shown to happen early in the gastrointestinal tract. Even with treatment, gastrointestinal tract T-cells have been reported only to recover minimally[64]. The latent (sometimes asymptomatic phase) may last up to ten years, and when not treated, progresses to stage two.

Clinical stage two of HIV infection is usually characterised by unexplained weight loss (<10% of presumed or measured body weight), persistent (and generalized) lymphadenopathy, oral lesions such as thrush and oral hairy leukoplakia, hypo proliferative anaemia and thrombocytopenia and reactivation of varicella-zoster

(shingles). Other symptoms may include recurrent respiratory tract infections (sinusitis, tonsillitis, otitis media, pharyngitis).

The third stage is typically characterised by additional manifestations such as unexplained severe weight loss (>10% of presumed or measured body weight), unexplained chronic diarrhoea for longer than one month, unexplained persistent fever (intermittent or constant for longer than one month). Also, bacterial infections, including pulmonary tuberculosis, pyelonephritis, pneumonia, empyema meningitis and osteomyelitis, may manifest. Others may include acute necrotizing ulcerative stomatitis, gingivitis or periodontitis, unexplained anaemia neutropaenia and sometimes chronic thrombocytopaenia.

During the fourth stage (severely symptomatic stage), clinical AIDS-defining symptoms manifest. These include (but not limited to): HIV wasting syndrome, Pneumocystis (jirovecii) pneumonia, recurrent severe bacterial pneumonia, chronic herpes simplex infection (orolabial, genital or anorectal of more than one month's duration or visceral at any site). Others include Oesophageal candidiasis, extrapulmonary tuberculosis, Kaposi sarcoma, cytomegalovirus infection, central nervous system toxoplasmosis, HIV encephalopathy *et cetera*[65].
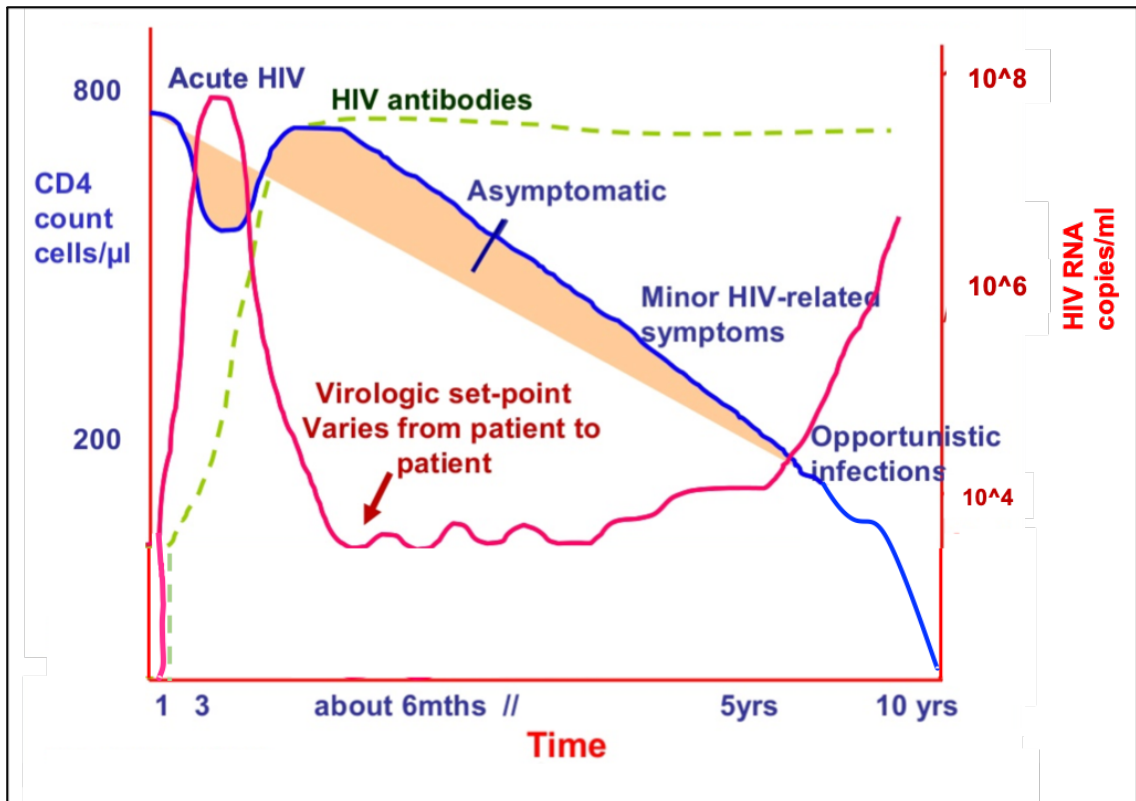
Figure 1.2 The natural course of untreated HIV infection

At the onset of infection, there is a burst of viremia (pink line), which is then inhibited by the onset of host immune responses. HIV antibodies (green dashed lines) begin to appear between 2-6 weeks. Equilibrium is reached between the virus and the host at virologic set point. There is a steady decline in CD4 T-cell (blue lines) count. At low CD4 T-cell counts, opportunistic infections begin to appear, viral load rises and continues to rise until death if there is no treatment.

1.5   The Biology of HIV-1

1.5.1   Genome organization of Retroviruses

Three major structural genes are common to all lentiviruses. These are: "group-specific antigen" (gag), DNA polymerase (pol) and envelope (env). Lentiviruses may also possess regulatory genes; transactivator of transcription (tat) and regulator of expression of virion proteins (rev) as well as virus-specific accessory genes. The pol encodes the reverse transcriptase (RT), integrase (IN) and protease (PR) enzymes, the env encodes the surface and transmembrane proteins of the viral envelope. All retroviral Gag proteins consist of three major domains: Matrix (MA), which mediates binding to membranes and targets Gag to appropriate assembly sites in the cell; Capsid (CA), which mediates lattice-forming protein-protein interactions in both the immature and mature capsids; and Nucleocapsid (NC), which contains one or two zinc knuckles that bind and package the viral RNA genome. These three domains are connected by spacer peptides or additional domains, which vary across different species[66].

HIV-1 Virion structure

Like other retroviruses, letiviruses are enveloped viruses. Mature HIV virions have a spherical morphology of ~110nm in diameter[67]. The outer envelope is composed of a lipid bilayer with viral envelope glycoproteins. Some cellular components are also incorporated since the envelope is mainly derived from the infected cell. The matrix protein (MA), which is anchored to the internal surface of the lipid bilayer by N-terminal myristoyl associated groups is surrounded by the envelope [21,68]. A dense cone-shaped nucleocapsid core is contained in the matrix. Viral genomic RNA molecules, viral proteins, as well as some cellular factors, are packaged in this matrix shell. Within the virion, the HIV genome is present as two identical single-stranded RNA molecules. In persistently infected cells, it exists as proviral double-stranded DNA [69]. It has an approximate genome size of 9.2kb.

The HIV genome is shown in a linear schematic in Figure 1.3. The viral protease or host cell proteases act to process the polyprotein products of the three primary genes into mature particle-associated proteins. The viral protease processes the

55-kDa Gag precursor (Pr55Gag) and the 160-kDa Gag-Pol precursor (Pr66Gag-Pol). The HIV protease has also been shown to cleave a single site in Nef [21,70]. Host cell proteases process the160-kDa Env precursor. The additional proteins: Tat, Rev, Vpr, Vif, Vpu and Nef are primary translation products of messenger RNA (mRNA) splicing[71].
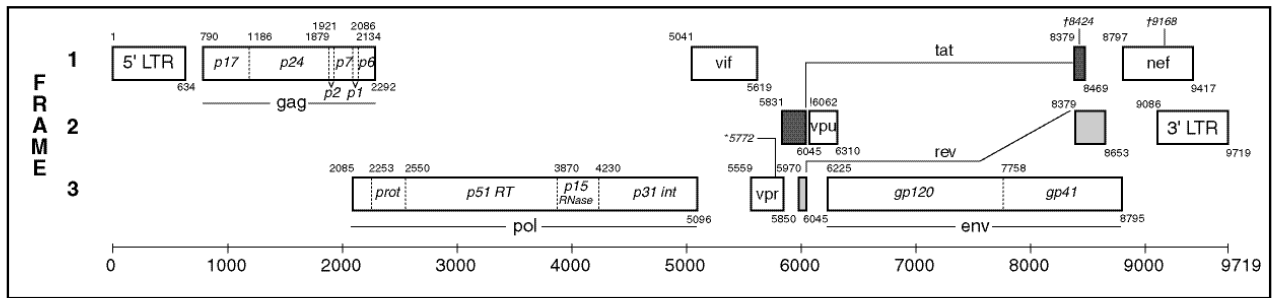
Figure 1.3: Linear schematic of the HIV-1 genome
A linear representation of the HIV-1 genome is shown. Open reading frames (ORFs) and the non-coding, regulatory 5' and 3' LTRs are shown as boxes. Reading frames are labelled 1, 2, 3 on the left and box position indicates the reading frame of each gene. Maturation cleavage sites for Gag, Pol and Env polyproteins are shown as dashed lines. The spliced exons (tat and rev) are shown as shaded rectangles. (Figure is taken from Los Alamos National Laboratory HIV-1 Gene Map 2017)[72].

As stated above, all retroviruses consist of three genes; gag, pol and env. These genes encode the structural proteins. Additional proteins such as Tat, Rev, Negative factor (Nef), Viral infectivity factor (Vif), Viral Protein U (Vpu) and Viral protein r (Vpr) are encoded by the more complex retroviruses such as SIV and HIV. These additional proteins perform regulatory or accessory roles.

Structural proteins which are required for the formation of viral particles is encoded by the gag gene. The env gene encodes envelope glycoproteins which are required by the virus to bind to host cells while pol encodes the enzymes required for the HIV-1 life cycle. The regulatory and accessory genes perform different functions to increase viral transcription and fitness. Vpr has several roles which include the induction of cell cycle arrest[73], nuclear transport of the pre-integration complex (PIC)[74] and the coactivation of HIV LTR[75]. Rev functions to regulate the export of transcribed genomic RNA from the nucleus while Tat activates the process of transcription[76].

HIV-1 relies on Vif to overcome the potent antiviral function of APOBEC3G (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G). By counteracting the function of APOBEC, the infectivity of viral particles is enhanced [77]. Nef is involved in infectivity and pathogenicity and has been shown to downregulate the expression of cell surface receptors, including CD4. It is necessary for the maintenance of high viral loads[78]. Vpu is an integral membrane protein that counteracts innate restriction by host tetherin[79,80], and it is involved in the degradation of CD4 in the endoplasmic reticulum. It also plays a role by enhancing the release of virions from the plasma membrane of HIV-infected cells[81].

1.5.3 The HIV-1 replication cycle

Conventionally, the replication cycle of HIV-1 is broadly divided into two phases - the first or early phase and the late phase. In its entirety, the replication cycle is completed in around 24 hours[82].

The early phase involves binding and fusion, entry into the host cell. Cell entry is followed by reverse transcription (the process by which an RNA template is synthesized into DNA), and integration of the viral DNA into the host genome.

The late phase comprises intracellular trafficking, nuclear export, assembly, budding and maturation. After maturation, the virus can infect new cells and progress to the early phase of the next cycle.

1.5.4   Early Phases

1.5.4.1   Viral Entry

Entry begins with the adhesion of the virus to the host cell by attaching to and binding to host cells. For this fusion to occur, the surface subunit envelope glycoprotein of the virus (called gp120) interacts with CD4 on the cellular plasma membrane. By this interaction, a conformational change is triggered in the gp120. This change in conformation then exposes the coreceptor binding site, which was previously occluded. When the cellular co-receptor binds, the transmembrane region of viral glycoprotein undergoes a change in its conformation thereby revealing the gp41 fusion peptide. The fusion peptide is then inserted into the plasma membrane of the target cell. Subsequently, two heptad-repeat regions (HR1 and HR2) within gp41 folds into a thermostable six-helix bundle (6HB) structure. Formation of the 6HB drives viral and cellular membrane fusion, which is needed for pore formation and virus entry[83-86].

1.5.4.1.1   Target Cell Receptors

For HIV-1 to enter into target cells, the viral membrane is required to fuse with the host cell's plasma membrane. The gp160 viral envelope protein interacts with the cellular receptors of the host. The virus binds to the CD4 (primary receptor) and either one of two co-receptors - CXCR4 or CCR5.

CD4 is a member of the immunoglobulin superfamily and is a type-1 integral membrane glycoprotein[87]. It serves as a ligand for Major Histocompatibility Complex (MHC) class II molecules, inducing T cell activation via a signalling cascade. It is predominantly expressed on the surface of T-helper cells. It is also an IL-16 receptor, a cytokine that has chemoattractant activity for CD4 + lymphocytes, monocytes and eosinophils[83].

CXCR4 and CCR5 belong to a family of chemokine receptors with a seven-transmembrane domain linked to a G-protein. They play a role in the trafficking of immune cells by facilitating entry into cells. CCR5 is expressed on either resting or activated T cells, while CXCR4 is expressed on a wide range of cells (both within and outside the immune system). They play a significant role in determining viral tropism.

The ability of HIV to replicate either in macrophages or T-cell lines is the hallmark of cellular tropism. HIV-1 viruses are described as "M-tropic" (M for macrophage) if they infect and establish infection in macrophages and peripheral blood mononuclear cells (PBMCs). A typical characteristic of these viruses is that they do not induce the formation of giant multinucleated cells (syncytia) hence are called Non-syncytium inducing (NSI) strains. Viruses that are able to infect T-cells and PBMCs are known as T-tropic. They induce syncytia and are sometimes referred to as syncytium inducing (SI) strains. Viruses either utilize CCR5 (R5 tropism) or CXCR4 (R4 tropism) to gain entry into cells. Some viruses (known as dual tropic) can enter cells using either CCR5 or CXCR4 or can enter by utilizing both coreceptors (R5X4 tropism)[88].

At the early stages of infection, a homogenous population of R5 viruses dominate, as the disease progresses, R5 to X4 switch may occur. Studies have established that the transition from CCR5 utilizing (R5) phenotype to a CXCR4 (X4) phenotype is associated with increased viral replication kinetics and progression to AIDS[83,89-91].
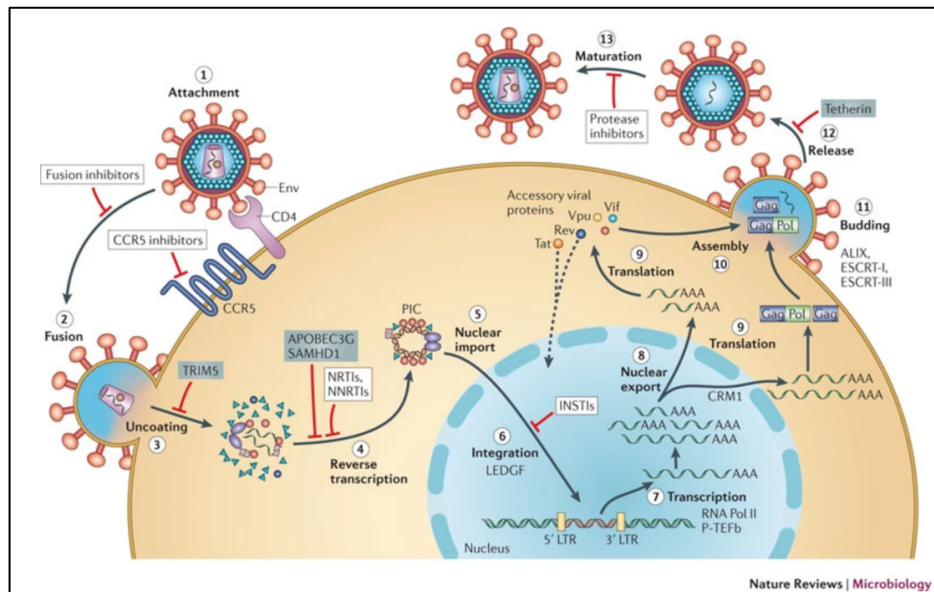
Figure 1.4: The HIV-1 life cycle and involvement of host proteins
The viral Envelope engages and binds to the surface of the CD4 and the CCR5 co-receptor
(1), then membrane fusion (2) and entry ensues. Reverse transcription and uncoating of the
capsid at occur at the nuclear pore, prior to nuclear import (5). (There is still some controversy
over the timing of reverse transcription, and uncoating). Once inside the nucleus, cDNA is
integrated (6) into the host genome. Proviral transcription then occurs (7), followed by viral
mRNAs being exported from the nucleus (8) and translation of viral proteins (9). Viral
assembly occurs at the plasma membrane (10). Finally, the virus buds from the cell (11) and
is released. This is followed by maturation. The matured virion is now capable of beginning
another round of infection.(Adapted from Engelman and Cherepanov, 2012)[92]

1.5.4.1.2   Viral envelope proteins

The HIV-1 env gene encodes the Env glycoproteins, which play an essential role in the virus replication cycle by mediating the attachment and fusion between viral and cellular membranes during the entry process cells[93]. The Env glycoproteins are synthesised as a single polypeptide precursor (gp160) which then undergoes post-translational modifications including N-linked glycosylation[94,95]. It is subsequently cleaved by furin cellular proteases in the Golgi apparatus[94,96]. This cleavage results in in the production of two subunits - gp120 and gp41[97]. Within the viral envelope, Env is trimeric with three gp120s each noncovalently bound to a gp41 subunit. The Env protein is anchored to the virus membrane by gp41 which is a transmembrane glycoprotein, whilst gp120 is a surface glycoprotein responsible for receptor binding[97]. The sequence of gp120 can be divided into five conserved regions (C1-C5), and five variable regions (V1-V5), The interactions of gp120 and coreceptors that are involved in tropism are located in V1/V2 and V3 loops in an area known as the bridging sheet[83,98]. During virus assembly, the gp120/gp41 complex is incorporated as heterotrimeric spikes into the lipid bilayer of nascent virions. These gp120/ gp41 complexes then initiate the infection process by binding receptor and coreceptor on the surface of target cells[93],[83,98].

1.5.4.2   Virus uncoating

Upon entry, the viral capsid utilizes microtubules and the actin cytoskeleton to transport itself within the cell[99]. A unique characteristic of HIV is its ability to enter the nucleus of a non-proliferating cell which enables it to infect terminally differentiated macrophages. Other retroviruses do not possess this distinct characteristic. For example, spleen necrosis virus (SNV) and murine leukaemia virus (MLV) require mitosis for productive infection and cannot integrate into non-diving cells[100].

The process which occurs in the cytoplasm by which the viral capsid disassembles is known as uncoating. The exact mechanism by which the capsid core uncoats and the transport of the viral nucleic acid into the nucleus are still controversial. Despite   numerous   studies   aimed   at   understanding   this

mechanism, the sequence of events between uncoating and reverse transcription still remains unclear. Also controversial is the exact location within the cell where uncoating takes place[101]. Opinion remains divided over the process. For example, it was suggested by Hilditch and Towers that by virtue of the size of HIV capsid in relation to the size of the nuclear pore, intact capsid cannot enter the nucleus, therefore, uncoating of the viral cDNA must therefore occur before nuclear entry[102]. A recent study by Burdick and colleagues, however, disagreed with this hypothesis. Using direct labelling and quantification of the viral capsid protein associated with infectious viral cores that produced transcriptionally active proviruses, they showed that intact (or nearly intact) viral cores enter the nucleus. This entry occurred through a mechanism involving interactions with host protein cleavage and polyadenylation specificity factor 6 (CPSF6)[103].

It has been shown that host factors are essential in directing the HIV uncoating and nuclear entry processes. It was suggested that the use of host co-factors enables evasion of the innate immune response. In the presence of CA mutations which prevent interactions with the host cofactors used for uncoating and nuclear entry, innate sensors are triggered, and an antiviral state is induced[104].

1.5.4.3   Reverse Transcription

A unique characteristic possessed by all retroviridae is the ability to copy single-stranded viral RNA into double-stranded DNA before integration. Reverse transcription is the process through which DNA is synthesised from a single-stranded RNA template. The reverse transcription process is made possible by the HIV-1 reverse transcriptase (RT) enzyme. The newly synthesized DNA then serves as the substrate for integration into the host genome.

1.5.4.3.1   The Reverse Transcriptase Enzyme

The reverse transcriptase has three functions; (i) an RNA dependent DNA polymerase activity to synthesize a single-stranded DNA copy from the RNA template; (ii) a DNA-dependent DNA polymerase activity to complete the synthesis of the newly transcribed cDNA and (iii) an RNase H activity which is required to degrade the RNA template from the RNA-DNA hybrid[105].

It comprises of two subunits (p66 and p51). These subunits share a common N-terminus. The p66 subunit contains two spatially distinct active sites for the separate activities of RT, whilst the p51 subunit has a structural role[106]. The p51 subunit lacks the RNase H domain and does not contribute to either of the catalytic activities of RT. Although isolated RNase H is not active and requires the presence of p51 to restore activity, it is thought that residues located in the C-terminal region of p51 and p66 are essential for RNase H cleavages[107]. Crystal structures of p66 domain of RT initially bound to inhibitors and also unbound have revealed a structure that resembles a right hand – as can be seen in Figure 1.5 This has led to the use of nomenclature based on that similarity; the subdomains have been named fingers, palm, thumb and the connection subdomain, which links the polymerase domain to the RNase H domain.
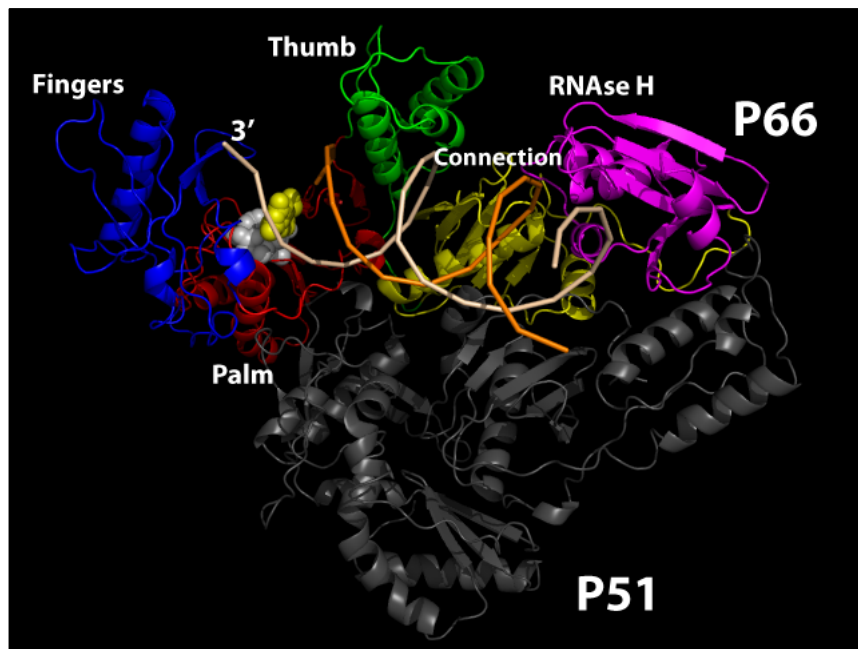
Figure 1.6 The structure of HIV-1 Reverse Transcriptase as revealed by X-ray crystallography

The reverse transcriptase is a heterodimeric protein consisting of p51 and p66 subunits. The p51 is shown in grey while the p66 monomer sub-domains are also shown in different colours: fingers (blue), palm (red), thumb (green), connection (yellow), RNAseH (magenta). Also shown are the nucleic acid template (beige) and primer (orange) and incoming dNTP shown in yellow spacefill mode. The enzyme active site which consists of three catalytic aspartates, D110, D185, and D186 is shown in white spacefill mode. (Adapted from Huang et al., 1998)[108].

1.5.4.3.2   The process of Reverse Transcription

The process of reverse transcription is schematically represented in Figure 1.6

At the 5' end of the RNA, there is a primer binding site (PBS) sequence, where for HIV-1 a lysine 3 (Lys3) transfer RNA (tRNA) is bound. This tRNA acts as the primer to DNA synthesis (Fig 1.6A). The negative strand of DNA is synthesised by the RT enzyme from the PBS towards the 5' end, creating the minus-strand strong stop DNA[109]. (Fig 1.6B). The production of strong stop DNA occurs in the viral particle before entry into a cell[110]. The RNase H function of RT degrades RNA in an RNA: DNA hybrid (Fig 1.6C). First-strand transfer then occurs, resulting in the minus-strand strong-stop DNA being translocated to the 3' end of the RNA where the complimentary repeat sequence (R) can anneal (Fig 1.6D). In a viral particle, two copies of the HIV RNA genome are present. At this point, template switching can occur with the rest of the negative strand of DNA being synthesised from either of the two genomes[111]. This results in genetic recombination, creating HIV-1 diversity and aiding in immune escape[112]. Again, synthesis is followed by the RNase H function removing all RNA in a hybrid except for the purine-rich polypurine tract (PPT), which is resistant to degradation (Fig 1.6D). The PPT now acts as the primer for the synthesis of the positive strand of DNA [113]. The positive strand DNA gets synthesised in a 5' to 3' direction through the U3, R, U5 regions and the first 18 nucleotides of the tRNA, creating the plus-strand strong stop DNA (Fig 1.6E). The RNase H activity degrades all RNA in a hybrid now including the PPT and the tRNA, bar for a single A ribonucleotide[114]. This exposes the PBS sequence on the plus sense strong stop DNA, allowing second strand transfer to occur with the PBS annealing to its complimentary sequence at the other end of the positive strand of DNA (Fig 1.6F). The synthesis of both strands is completed, forming a linear dsDNA flanked by LTR's (Fig 1.6G). This RT process is highly complicated, with first and second strand transfers suggesting the requirement for a certain level of genome organisation for this to efficiently occur.
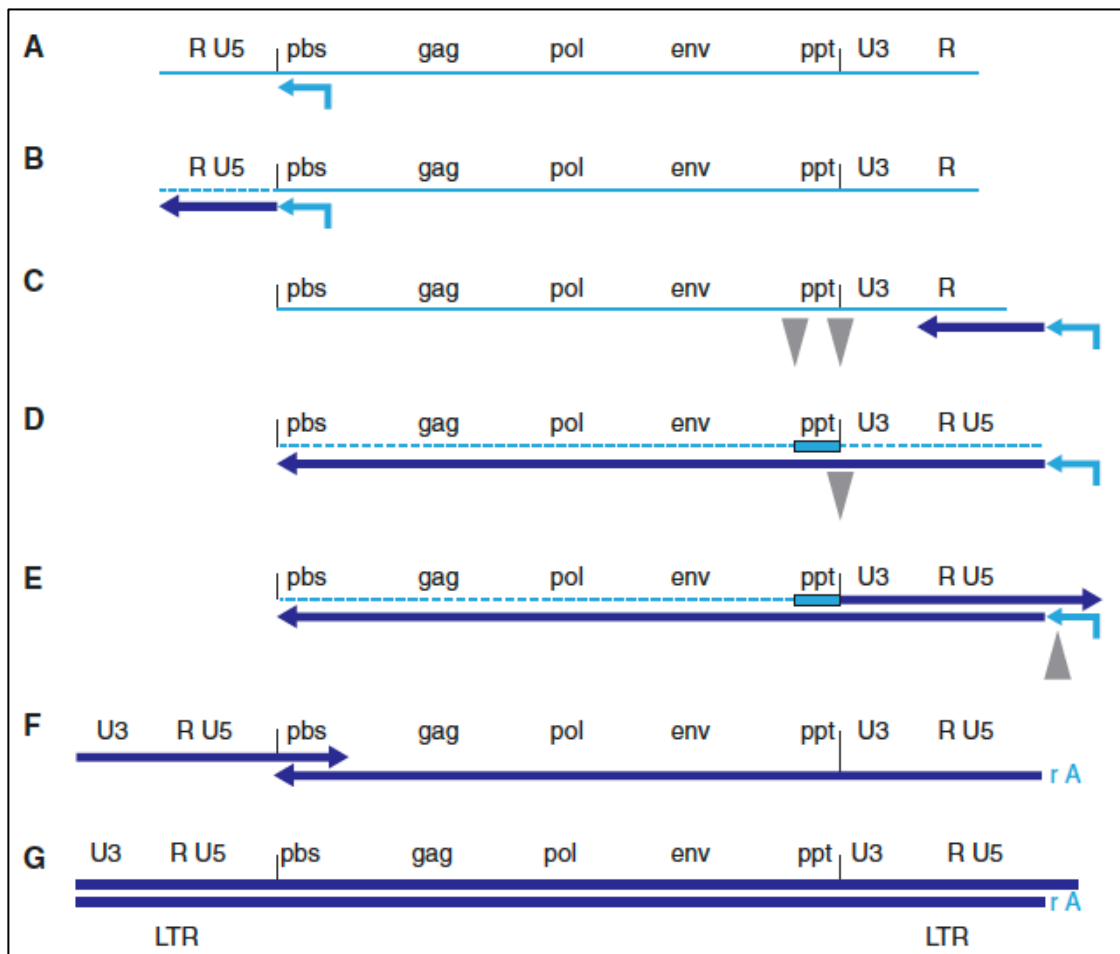
Figure 1.6: The stages of HIV reverse transcription
Conversion of the single-stranded RNA genome of a retrovirus into double-stranded DNA occurs stepwise as follows: (A). tRNA (light blue arrow) binds to primer binding site (pbs) near the 5' end. (B). RT then initiates reverse transcription, generating minus-strand DNA (dark blue arrow) and the RNase-H activity of RT degrades the RNA template (light blue, dashed line). (C). Minus strand transfer then occurs between the R sequences at both ends of the genome (first strand transfer) (D). After this transfer, minus-strand synthesis continues along the length of the genome. As DNA synthesis proceeds, so does RNase H degradation of RNA genome template. PPT (light blue box) is resistant to RNase H degradation and acts as a primer for positive-strand DNA synthesis (E). Plus-strand synthesis continues until the first 18 nucleotides of the tRNA are copied, allowing RNase H cleavage to remove the tRNA primer. (F) the cleavage of tRNA allows second strand transfer (G) extension of plus and minus strands leads to the synthesis of complete double stranded linear viral DNA). (Figure adapted from Hu and Hughes, 2012)[112].

1.5.4.3 Nuclear entry

HIV-1, as well as many other viruses that depend on entry into the nucleus for replication, have developed an evolutionary strategy to dock and translocate through the nuclear pore complex (NPC). In particular, the nuclear pore is not a static window, but it is a dynamic structure involved in many vital cellular functions, such as nuclear import/export, gene regulation, chromatin organization and genome stability[115].

Prior to integration, the viral cDNA is translocated into the nucleus as part of a large nucleoprotein complex, the pre-integration complex (PIC), which contains both viral (i.e., P17, P24, RT, integrase and Vpr)[116] and cellular proteins. HIV-1 and other lentiviruses have the unusual ability to infect non-dividing cells, which implies that the PIC must enter the nucleus through an intact nuclear membrane. In contrast, other retroviruses require the disintegration of the nuclear membrane during mitosis to gain access to the nuclear components.

1.5.4.4 Integration

Following entry of HIV into a host cell, the virus synthesizes a double-stranded (ds) DNA copy of its RNA genome[117]. The viral DNA is then irreversibly inserted into the host genome. This process is called integration.

1.5.4.4.1 HIV integrase enzyme

Integrase is encoded by the C-terminal region of the HIV-1 *pol* gene. It is a 32 kDa, 288 amino acid-long protein that is conventionally divided into three structural and functional domains, namely N-terminal domain (NTD), catalytic core domain (CCD), and C-terminal domain (CTD).

As reviewed by Mbisa et al[118] the NTD encompasses IN residues 1–49 and is made up of a triplet of α-helices, containing a double histidine/cysteine (H12-H16-C40-C43) zinc-binding motif. The motif plays a role in the dimerization of IN monomers and the binding of cellular factors[119]. Integrase residues 50–212 make up the CCD which contains the active site composed of a triad of acidic residues D64, D116, and E152, also called the DDE motif. The motif is essential for the

coordination of divalent metal ions (Mg2+ or Mn2+) that are essential for IN enzymatic functions.[119]. The CTD, residues 213–288, contains SH3 domains that non-specifically bind to DNA [120]. The C-terminal domain (amino acids 213-288) has a very high affinity for DNA and thus plays a role in binding to host DNA [121] so has been implicated in the stability of the integrase/DNA complex[122].

It is believed that the functional entity of HIV-1 integrase is a tetramer assembled from two symmetrical dimers each bound to one of the viral DNA ends[123]. Elucidation of the structure of full-length HIV-1 IN and its mode of action has recently benefited from the determination of the crystal structure of prototype foamy virus (PFV) IN tetramer in complex with 3′ processed viral DNA ends[123,124]. The IN–DNA complex is called an intasome and is the minimal structure required for integration into target DNA.

1.5.4.4.2   The Mechanism of Integration

Integrase catalyses 3' end processing and strand transfer of the viral DNA. Upon viral DNA synthesis, IN multimerises on the nascent DNA resulting in the formation of a DNA:IN complex known as intasome[124]. A functional intasome contains four IN proteins in which only two IN molecules contact the DNA[124]. 3'-end processing of the DNA by IN results in removal of two nucleotides from each 3'end of the blunt-ended viral DNA, revealing 3'hydroxyl groups that can be joined to the target DNA.

The second step of integration involves DNA strand transfer, whereby the viral DNA ends are inserted into the target DNA. The 3' hydroxyl groups carry out a nucleophilic attack on the phosphodiester bond in the opposite strands of the target DNA, followed by the covalent joining of 3' ends to the 5' ends of the target DNA. Five (5) nucleotides separate the sites of attack on the target DNA, resulting in five nucleotides long single strand gaps on the target DNA and two nucleotides long overhangs at the 5' ends of the viral DNA which are filled by the cellular DNA repair machinery.

After entry into the nucleus the viral DNA can also undergo circularization into 2-LTR and 1-LTR circles. 2-LTR circle formation is dependent on the host cell non-homologous DNA end-joining (NHEJ) components Ku70/80, ligase IV and

XRCC4 [125]. 1-LTR circles are formed by a recombination event between the LTRs and depend on the host MRN complex (Mre11, Rad50, and NBS1)[126].

1.5.5 Late Phases

The late phase of the HIV-1 replication cycle comprises viral gene expression, intracellular trafficking, transcription, nuclear export, assembly, budding and maturation.

1.5.5.1 Transcription and translation of viral genes

The late phase of the HIV-1 replication cycle begins with transcription from the previously integrated provirus. The transcription process is regulated by the viral transactivator protein (Tat). About 40 differentially spliced species of messenger RNA (mRNA) transcripts are produced from the HIV-1 transcription process. These mRNA transcripts are of three types: spliced, partially spliced or unspliced. In order to be exported from the nucleus, unspliced mRNAs require the viral protein Rev whereas spliced mRNAs are readily exported from the nucleus. Located at both the 5' and 3' end of the HIV-1 genome are long terminal repeats (LTRs). The 5' LTR is dominant and acts as the promoter for transcription. It is only when the 5' LTR is defective that the 3' LTR functions as the promoter[127,128]. The HIV-1 LTR is divided into three regions, U3, R and U5. The process of transcription is initiated from the R region in the 5'LTR and terminates in the R region of the 3'LTR. Four elements that regulate viral transcription are present in the U3 region. The cellular transcription factors (such as NF-κB, Sp1 and NFAT) bind to the modulatory and enhancer elements at specific binding sites[129]. The core promoter element contains the TATA box. The TAR element binds the HIV-1 protein Tat, which enhances proviral transcription.

HIV-1 transcription is initiated when cellular transcription factors bind the 5'LTR and recruit histone modifying enzymes, allowing recruitment of the RNA polymerase II (RNAPII)[130]. Repressive proteins such as negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF) halts the transcription process shortly after starting[131,132]. The short mRNA which was transcribed before the halt

adopts a stem-loop structure known as the transactivation region (TAR). The viral Tat protein binds TAR and stimulates transcription elongation by recruiting the cellular positive transcription elongation factor b (P-TEFb)[133]. P-TEFb contains a catalytic subunit, cyclin-dependent kinase 9 (CDK9), and a regulatory subunit, cyclin T1. P-TEFb recruitment results in phosphorylation of NELF and DSIF[134]. Phosphorylated NELF dissociates from the complex and DSIF phosphorylation blocks its inhibitory activity on transcription elongation. Furthermore, P-TEFb phosphorylates serine residues at position 2 and 5 in the C-terminus of the RNAPII, which enhances its processivity. These phosphorylation events result in a Tat dependent positive feedback loop resulting in accumulation of Tat and sustained HIV-1 transcription (as reviewed by Khan, 2019)[135].

Like cellular mRNA, viral mRNA is capped at the 5' end, introns are removed by splicing, and a poly-A tail is added at the 3' end before it is exported from the nucleus. HIV-1 mRNA processing is coupled with its transcription. Tat promotes capping of the viral mRNA due to RNAPII phosphorylation. RNA guanylyltransferase (RNA GT) is recruited to the viral transcript by the phosphorylated Ser2 in RNAPII. Phosphorylation of the Ser5 is required for activation of the RNA GT[136]. RNA triphosphatase cleaves the 5' triphosphate into a diphosphate which is then capped with GMP by RNA GT and methylated by guanine N7methyltransferase. HIV-1 contains two polyadenylation signals. One at the 5' end and one at the 3' end of the genome. The 5' polyadenylation signal is suppressed due to its proximity to the transcription initiation site and by binding of U1 snRNP[137,138]. The usage of the 3' signal is enhanced by the presence of an upstream enhancer motif. Cleavage and polyadenylation specificity factor (CPSF) binds to this enhancer motif and cleaves mRNA between the AAUAAA motif and the downstream U or GU rich region. A Poly A tail is then added by Poly(A) polymerase using ATP as a substrate.

Splicing is a process in which introns are removed from mRNA transcripts by a large multicomponent ribonucleoprotein complex known as the splicosome. Spliceosome contains five small nuclear ribonucleoprotein complexes (snRNPs),

U1, U2, U4, U5, and U6, which assemble onto the transcript[135]. U1 snRNP recognises the 5' splice site and U2 snRNP recognises the 3' splice site. Splicing is regulated by sequences in the exons and introns known as exonic and intronic splicing enhancers or silencers (ESE and ISE respectively). Serine-arginine (SR) rich family of splicing activators are recruited by ESEs rich in purine. SR proteins bind ESEs and stabilize the binding of the core splicing factors at the splice sites. HIV-1 transcript splicing results in about 40 differentially spliced mRNA transcripts due to alternative usage of the multiple 5' and 3' splice sites[139].

Partially spliced transcripts are 4kb long and encode Env, Vpu, Vpr and Vif. 2kb transcripts are fully spliced and encode Tat, Rev and Nef. 50% of the viral mRNAs are unspliced of 9kb length. Unspliced mRNA codes for the Gag and GagPol precursor polyproteins that are packaged into budding virions alongside the viral genome[135]. HIV-1 employs various strategies to exploit the cellular translation machinery for efficient synthesis of its proteins[140]. Translation occurs mainly by the Cap-dependent scanning method in which the 5' end of the capped mRNA recruits the 40S ribosomal subunit[141]. The 40S ribosomal subunit scans towards the 3' end and recruits the 60S ribosomal subunit when a start codon (AUG) is encountered. For the translation of the Pol gene, a programmed frameshift occurs at the Gag stop codon resulting in the production of GagPol precursor protein. This event is crucial for the production of viral enzymes. Env and Nef genes are translated due to a leaky scanning mechanism of Vpu and rev start codons by the ribosomes. Furthermore, HIV-1 uses structural RNA elements such as internal ribosome entry (IRES) to initiate translation. The IRES drives translation by directly recruiting the 40S ribosomal subunit in a cap-independent manner. Two IRES elements have been found in the HIV-1 genome. The IRES present in the 5'UTR seems to be active under oxidative stress and when the cap-dependent translation is blocked. The IRES present in the Gag coding mRNA drives expression of a 40kDa Gag isoform in addition to the Pr55[Gag] (as reviewed by Khan, 2019)[135].

1.5.5.2.1   Virion assembly

The HIV-1 assembly process is a multi-complex mechanism that takes place at the host cell plasma membrane[142]. In T-cells, viral assembly (and budding), takes place on plasma membranes while in macrophages, it takes place on endosomal membranes [143,144]. HIV-1 assembly is driven by oligomerization of the Gag polyprotein at the plasma membrane of an infected cell, leading to membrane envelopment and budding of an immature virus particle[145]. The process of particle assembly is driven by Gag proteins and is achieved by directing virion budding, controlling virion size as well as ensuring the inclusion of other vital components.

The matrix (MA) domain of Gag contains highly basic sequence residues located between positions 16 to 31. This region has been recognized as being responsible for targeting Gag to the plasma membrane once it is synthesized in the cytosol. These residues form a cluster on the surface of the MA globular domain that is conserved among retroviruses.

At the amino terminus of MA is a myristic acid which is attached via covalent bonds to the basic residues. It is essential for viral budding and anchors Gag to the plasma membrane lipid bilayer[146]. Amino acids 29 and 31 of MA are particularly necessary for membrane targeting as substitution results in localisation of Gag to intracellular compartments[147]. Gag targets PI(4,5)P2 [phosphatidylinositol-(4,5)-bisphosphate] within the plasma membrane and in its absence, Gag localises to internal cellular membranes[147,148].

Studies using fluorescence imaging techniques (FRET) have shown that multimerisation of full-length Gag occurs at the plasma membrane[149]. Viral RNA and Gag subunits CA and NC have been implicated in Gag multimerisation. Viral RNA has been shown to promote Gag multimerisation through its interactions with NC, so it is thought that NC does not mediate multimerisation directly[150]. Mutations in the C terminal domain of CA have been shown to affect assembly[150,151].

Two copies of the viral RNA genome are incorporated into each virion from the full-length, unspliced viral RNA present in the cytoplasm. NC is involved in genomic RNA recruitment and RNA dimerisation is required for its incorporation into virions (as reviewed by Moore and Hu, 2009)[152]. Genomic RNA encodes a packaging signal, ψ, in the 5' region of Gag which labels it for packaging into the virion. Ψ encodes 4 stem loop structures (SL1-4) which interact with two zinc finger motifs of Gag NC to bring about RNA encapsidation. However, most of the 5' UTR is involved in RNA genome encapsidation (as reviewed by Sundquist and Krausslich, 2012)[153].

The viral envelope proteins, gp120 and gp41, are also sequestered to the host cell membrane for incorporation in the viral membrane. The Env gp160 precursor is synthesised in the rough endoplasmic reticulum and cleaved into two subunits by cellular furin proteases during its transport to the cell surface via the golgi apparatus. The process of Env recruitment is not yet fully understood, but it has been shown to require the MA subunit of Gag (as reviewed by Sundquist and Krausslich, 2012)[153]. A mutation at position 49 of MA decreases the incorporation of Env into the virion, but deletion of the gp41 C terminal domain reverses this effect, implying that the interaction occurs between MA and the C terminus of Env[154]. Studies have shown that cleavage of Gag, in particular the separation of MA, during maturation is essential for the formation of a single focus of Env, required for viral infectivity[155].

A number of viral accessory proteins are also packaged into the assembling virion as they are required in the early stages of the life cycle in the next infected cell. Vpr is incorporated via direct interaction with the Gag p6 subunit, which is sufficient for its inclusion (as reviewed by Sundquist and Krausslich, 2012)[153].

The general notion is that viral RNA only triggers Gag assembly and is dispensable for subsequent assembly. However, in recent studies, Yang and colleagues utilized Single-molecule localization microscopy (SMLM) to investigate the mechanism by which the HIV-1 viral RNA (vRNA) mediates the

assembly of thousands of Gag proteins into a virus particle at the plasma membrane. They established that vRNA is indispensable throughout the assembly, scaffolding the formation of assembly intermediates and maintaining their architectures via balancing of external forces acting on the assembly environment. Their results suggest that beyond inducing Gag to form low-ordered multimer basic assembly units, vRNA is essential in scaffolding and maintaining the stability of the subsequent assembly process[156].

More recently, studies showed that HIV-1 exploits the Gag-Pol ribosomal frameshift mechanism in support of Pol expression as a Gag-Pol fusion protein at relatively low levels, resulting in the promotion of Gag assembly and Pol incorporation into virus particles. HIV-1 virus assembly and replication benefit from this Gag-Pol ribosomal frameshift mechanism[157].

1.5.5.2.2 Virion budding and release

The pathway to HIV-1 budding process is represented in Figure 1.8

The Endosomal sorting complexes required for transport (ESCRTs) are a family of cellular proteins comprising ESCRT-0, -I, -II, and -III complexes. Recruitment of the ESCRT system by HIV is one of the best-documented examples of the comprehensive way in which a virus hijacks a normal cellular process[158]. HIV-1 uses the ESCRT protein pathway to bud from infected cells. The ESCRT complex is involved in multivesicular body biogenesis, cytokinesis, and macroautophagy[159],[160]. ESCRT-I and ESCRT-II are essential for the formation of the bud whereas ESCRT-III is required for scission of the bud[161]. More recent studies also clarify the role of ESCRT-II in the late stages of HIV replication and reinforce the notion that ESCRT-II plays an integral part during this process as it does in sorting ubiquitinated cargos and in cytokinesis[162]. Vps4 is involved in recycling of the ESCRT proteins after budding[163].

The carboxy-terminus (p6 domain) in Gag contains two late domains: The Pro-Thr/Ser-Ala-Pro [P(T/S)AP] and the Tyr-Pro-Xn-Leu (YPXnL) motif (where X is any amino acid and n=1–3 residues).   The P(T/S)AP domain recruits tumour

susceptibility gene 101 (TSG101) while the YPXnL domain recruits ALG2-interacting protein X (ALIX)[164,165]. ESCRT complexes and Vps4 are subsequently recruited. The ESCRT-III complex proteins form concentric rings (circular arrays or spirals) at the base of the bud that is then constricted, resulting in the scission of the bud[166]. ESCRT complex-mediated budding and scission is triggered by ubiquitination of the cargoes[167].
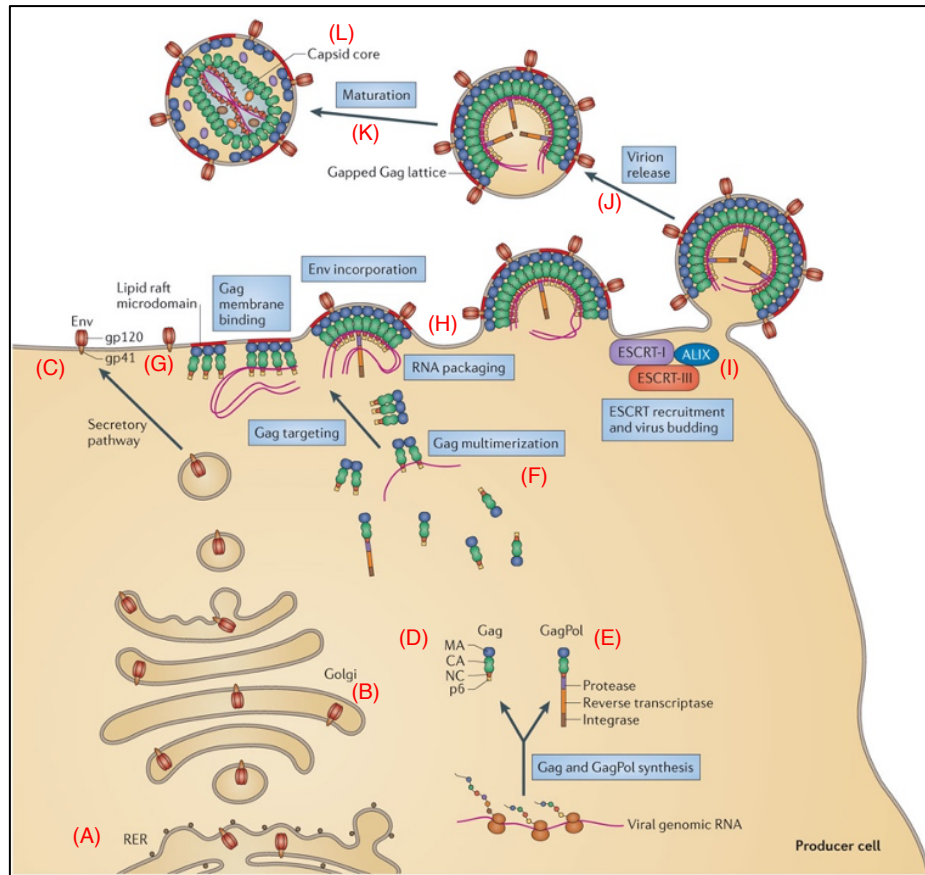
Figure 1.7: The late stages of the HIV-1 replication cycle

The viral envelope (Env) glycoproteins are trafficked from the rough endoplasmic reticulum (RER) to the Golgi via the secretory pathway (A - B). They are then moved (in vesicles) to the plasma membrane (C). From full-length viral RNA, Gag precursor polyprotein is synthesized in the cytosol (D). A programmed frameshifting event during the translation of Gag-encoding viral RNA results in the synthesis of GagPol precursor polyprotein (E). Viral genomic RNA is then recruited by Gag, multimerizes (F) and is trafficked to the plasma membrane (G). Gag then inserts its amino-terminal myristate into the lipid bilayer and anchors to the plasma membrane (H). The assembling particle incorporates Env and then recruits endosomal sorting complex required for transport I (ESCRT-I) and the ESCRT-associated factor ALIX (I). As the budding process proceeds, the ESCRT-III and vacuolar protein sorting 4 (VPS4) complexes are recruited and drive the membrane scission reaction that leads to particle release (J). The proteolytic cleavage of the Gag and GagPol polyprotein precursors by the viral protease is then triggered, in a process known as maturation (K), leading to the formation of the conical capsid core (L) Adapted from Freed, E.O., 2015 [168]

1.5.5.2.2   Virion maturation

The final stage of the HIV-1 viral lifecycle is the maturation of immature virions. The maturation step is essential for the newly assembled virion to be infectious. During maturation, the immature virion experiences dramatic morphological changes and becomes infectious. Maturation begins concomitant with or immediately after budding and is driven by the HIV-1 protease cleavage of Gag and Gag-Pol polyproteins[169]. Maturation is the process by which the virion develops its final infectious structure. The Gag polyprotein coordinates these stages in a process that requires the coordinated synthesis of all viral proteins as domains of the Gag and Gag-Pol polyproteins. This ensures that components are made in correct proportions[170].

During retrovirus maturation, cleavage of the precursor structural Gag polyprotein by the viral protease induces architectural rearrangement of the virus particle from an immature into a mature, infectious form. The structural rearrangement encapsidates the viral RNA genome in a fullerene capsid, producing a diffusible viral core that can initiate infection upon entry into the cytoplasm of a host cell[66].

1.6   The HIV-1 Protease Enzyme

Protease is a dimeric aspartyl protease composed of two identical polypeptides of 99 amino acids each (as shown in Figure 1.9). The mature HIV-1 protease (PR) is synthesized as part of a large Gag-Pol polyprotein precursor It is responsible for its release from the precursor (termed autoprocessing), thus promoting controlled proteolysis of the viral Gag and Gag-Pol polyproteins into the mature structural and functional proteins required for virus assembly, maturation and propagation[171,172]. The triplet Asp-Thr-Gly (from positions 25-27) form the active site of each protease monomer. There are two flexible flaps in the HIV-1 protease. These flaps are dynamic in solution hence enable the entry of inhibitors and substrate into the binding cleft situated above the active site[173].

It is still not fully understood how protease auto-cleaves from the Gag-Pol polyprotein; however, it is known that the dimerisation of Gag-Pol is required for

the activation of protease. Protease autocleavage can be controlled *in vitro* by the oxidation of cysteine residues 67 and 95[174]. Once dimerization is completed, *cis* cleavage of the protease from within the Gag-Pol polyprotein occurs, before protease cleaves the remainder of the Gag-Pol polyprotein in *trans* at specific cleavage sites. Studies show that specific cleavage sites are recognised by their secondary structure and not their amino acid sequence[175,176].

flap domain (44-57)

tip of the flap (49-52)

active site

elbow (37-43)

catalytic aspartates (25)

10's loop (15-18)

60's loop (66-69)

core domain

dimerization region (N 1-4, C 96-99)
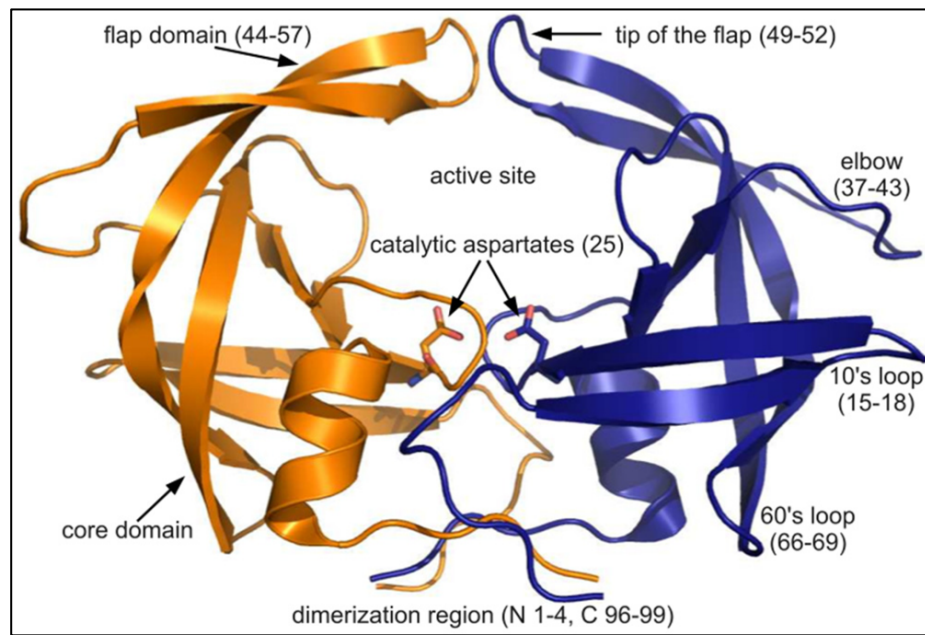
Figure 1.9 The HIV-1 Protease structure

Ribbon diagram of the HIV-1 protease structure showing one monomer in orange and the second monomer in blue. Regions of protease structure are labelled, and relevant residue numbers are given in parenthesis. The dynamic flaps that enable the entry of substrate and inhibitors into the substrate-binding cleft are labelled. The active site of the enzyme is also labelled, showing the catalytic aspartates on each monomer. Adapted from Ventakakrishnan et al., 2012. [177].

1.7   The HIV-1 Gag Polyprotein and its Role in the HIV-1 Replication Cycle

HIV-1 Gag is a structural gene whose precursor (designated as Gag Pr55Gag) has a molecular weight of 55kDa and is cleaved into mature proteins – Matrix (MA), Capsid (CA), nucleocapsid proteins (NC), p6 and spacer proteins p1 and p2 by the viral protease encoded by HIV-1 pol gene[178]. The Gag precursor is thought to have assembly and membrane targeting functions, whereas the mature proteins are involved in uncoating and disassembly[178]. There are approximately 2000 copies of Gag in a single HIV-1 virus particle[179] that comprises about 50% of the viral mass[180].

The Gag proteins of HIV-1 are central players in virus particle assembly, release, and maturation, and also function in the establishment of a productive infection[181,182]. The linear structure and function of each of the individual Gag domains are described briefly and summarized in Figure 1.10.

1.7.1   Gag Domains

1.7.1.1   Matrix (MA, p17)

MA is mainly responsible for membrane targeting and binding. The N-terminus of MA is co-translationally modified by N-terminal myristoylation; this modification is essential for the membrane transport and binding function of MA[146,183].

The MA domain of HIV-1 Gag plays critical roles in virus assembly by targeting the Gag precursor to the plasma membrane and directing the incorporation of the viral envelope (Env) glycoprotein into virions[184]. The positively charged basic residues on MA are thought to interact with the negatively charged residues on the inner surface of the plasma membrane, thus stabilising membrane association[146,183]. The overall structure of MA is thought to be trimeric, and this may be functionally relevant for assembly. It is also involved in the incorporation of the full-length envelope into virions; a direct interaction takes place between the cytoplasmic tail of gp41and matrix[178]. Recently, Kroupa and colleagues observed that Gag dimerization and the highly basic region in the matrix domain contribute significantly to the specificity of viral RNA binding[185].

1.7.1.2 Capsid (CA, p24)

The CA proteins form a cone-shaped structure located centrally in the mature virus, surrounding the complex required for proviral DNA synthesis and integration. A region located in CA forms an exposed loop structure which is responsible for the binding of cyclophilin A. The C-terminal domain contains a region which displays significant homology amongst retroviruses, known as the major homology region (MHR), mutations in this area produce assembly, maturation and infectivity defects in progeny viruses[178]. The dimerization domain of CA also interacts with the MHR region, forming hydrogen bonds which contribute by stabilising the structure of CA[21].

More recent studies have shown that CA protects HIV-1 complexes from degradation, mediates docking at the nuclear pore before uncoating, and determines the depth of nuclear penetration *en route* to integration[186].

1.7.1.3 Nucleocapsid (NC, p7)

The NC protein, like other Gag proteins, serves a multifunctional role in the HIV life cycle. Its primary role is in the binding of RNA, particularly to an area in the 5'-UTR, which contains a specific stretch of RNA that forms several stem-loop structures thought to be essential for the packaging of viral RNA into the virion. The packaging signal or Ψ, is 120 nucleotides long and is located between the 5'-LTR and the Gag start codon. The analogous region to the packaging signal in the 3' repeat region contains the polyadenylation signal[187]. The RNA/NC interaction occurs primarily through secondary structure interactions, as there is little sequence conservation of this region[188]. A flexible basic linker region joins the two zinc finger motifs of NC. Mutations that abolish zinc binding also result in loss of infectivity and diminished genome encapsidation[146]. The NC/RNA interactions also include the coating of viral RNA in the mature virion and binding of the tRNA^{lys} primer. All these interactions can be attributed to the 'nucleic acid chaperone' function of NC, which may also carry out functional roles in reverse transcription and integration. The N-terminal domain of mature NC is thought to play a role in Gag multimerisation by binding to Pr55^{Gag}, and through interactions with host proteins, is also thought to be involved in assembly and transport[21].

1.7.1.4   p6 Protein

The C-terminal protein to be released from the Pr55^{Gag} precursor is the proline rich p6, which is only found in primate lentiviruses[21]. It plays an essential role in the final budding step and also by incorporating Vpr. A motif within Gag known as the late domain, is thought to interact with host endosomal sorting machinery. The amino acid sequence P(T/S)APP near the N-terminus of the protein has been shown to promote budding of the virion from the membrane by binding to components of the ESCRT (endosomal sorting complex required for transport) I, II and III pathway[189]. The trans frame protein p6^{pol} has been implicated in the regulation of protease activity truncation of the C-terminal portion of p6 are thought to increase Gag processing [187,190].
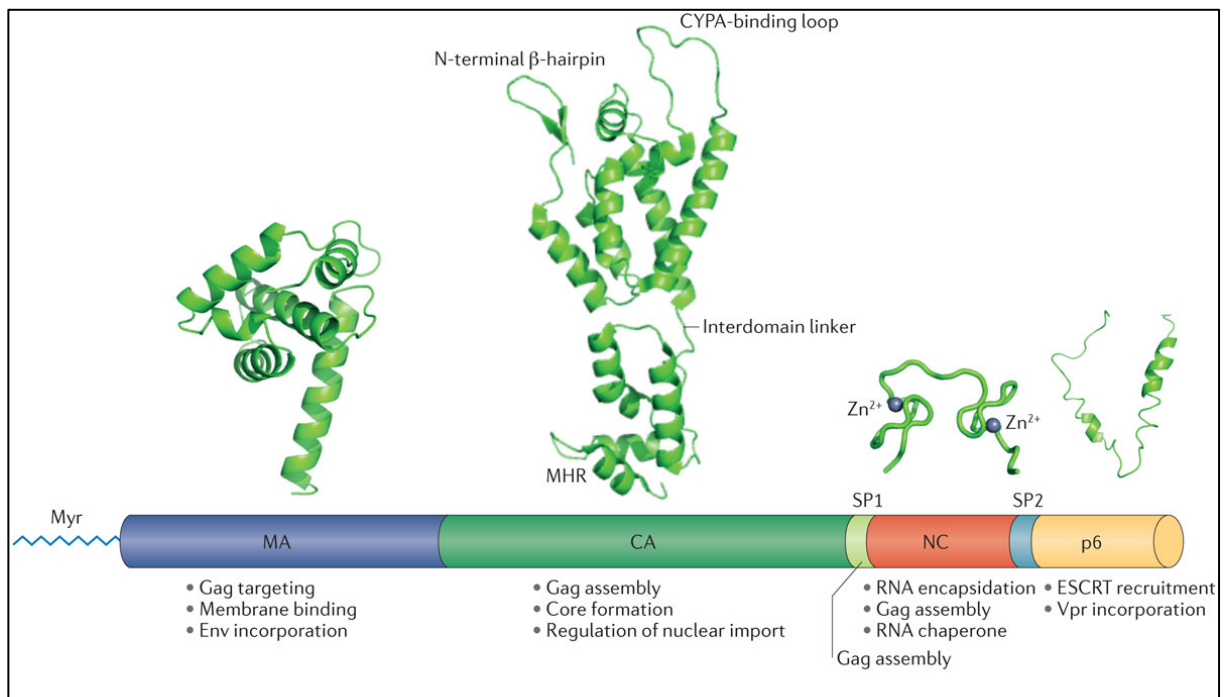
Figure 1.10: Gag structure and functions

The 4 domains (MA, CA, NC and p6) and the spacer peptides (SP1 and SP2) are shown in different colours. The ribbon diagram of each domain is shown above the linear sketch. Also shown are summaries of the major function(s) of each of domain. (Adapted from Freed at al.,2015 [168]).

1.7.2 Protease cleavage of the Gag polyprotein

During or immediately after budding, the viral protease auto-activates and cleaves Gag at specific sites to initiate maturation. This results in disassembly of the Gag lattice and condensation of the released NC–RNA complex into a compact ribonucleoprotein particle, which presumably prepares the genome for reverse transcription and integration. Around 1,500 copies of the new CA proteins then assemble into the mature capsid that re-encapsulates the genomic complex and its associated replicative enzymes. This generates the retroviral core, which consists of the mature capsid and its contents[66].

The proteolytic cleavage of Pr55[Gag] precursor by the virally encoded protease gives rise to the release of the HIV-1 Gag proteins: matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p7), p6 and the spacers p1 (also known as SP2), p2 (also known as SP1) [191,192] as represented in the schematic in Figure 1.11. Also, when Gag-pol polyprotein is cleaved, integrase, reverse transcriptase, and protease enzymes are released. A site in Nef is also cleaved by protease although this cleavage of Nef does not appear to be functionally important [21].

It is not known precisely at what time cleavage occurs, what is known however is that cleavage occurs late enough into the budding process to ensure that all cleavage products remain in the budding virion.

The secondary structure, not the amino acid sequence of each site determines where cleavage will occur consequently, amino acid sequences between cleavage sites vary significantly [193]. Between the different cleavage site, the secondary structures are fairly well conserved. The different rates of cleavage at each site is attributed to the difference in the protruding side chains.

Studies have demonstrated that the amino acids at the cleavage site (four amino acids before the cleavage site and three after), which are in direct contact with protease regulate the rate of cleavage[194,195].

The rates of cleavage have also been shown to be affected by amino acid residues which are more distant from the cleavage site. For example, the fourth and fifth amino acids following the cleavage site p4' and p5', have been shown

to affect rates of cleavage[196,197]. These amino acid residues are mostly hydrophobic, and it has been shown that a hydrophobic residue at the amino acid position before the cleavage site (p1) is required for cleavage to occur[198].

Cleavage at sites within Gag proceeds in a precise, and tightly-controlled order. This order of cleavage is as a result of the variation in protruding side chains and amino acid sequences present. The order in which the Gag sites undergo cleavage is shown in Figure 1.11 along with the amino acid positions at each cleavage site.

The first cleavage event occurs at the C terminal of p2 (SP1), resulting in the release of two intermediates: MA-CA-p2 and NC-p1-p6[199]. After this first step, genomic RNA is condensed to the centre of the virion.

The second cleavage events occur between MA and CA (leading to the release of MA from CA-p2). Simultaneously, p1/p6 is cleaved, leading to the release of p6 from NC-p1. The cleavage steps at the second stage occur more slowly than the first cleavage step (occurrence is at a ten-fold lower rate than the first cleavage step)[200].

Finally, the p1 and p2 spacer peptides are separated from the NC and CA, respectively. This final, rate-limiting step proceeds at a much slower rate than the first two steps and occurs a rate several hundred-fold reduced in comparison to that of the first cleavage step[197]. It has been shown that for the mature core surrounding the viral RNA to be formed, p2 must separate from CA[145].

For infectious viruses to be produced, cleavage must be procced in the precise order. When Gag is incorrectly cleaved, immature virions are produced, with viral cores that appear by electron microscopy to be shaped like doughnuts. Virions with correctly cleaved Gag appear conical-shaped cores, characteristic of mature retroviruses[201]. The structural differences between the mature and immature virions are diagrammatically presented in the electron micrographs in Figure 1.12.
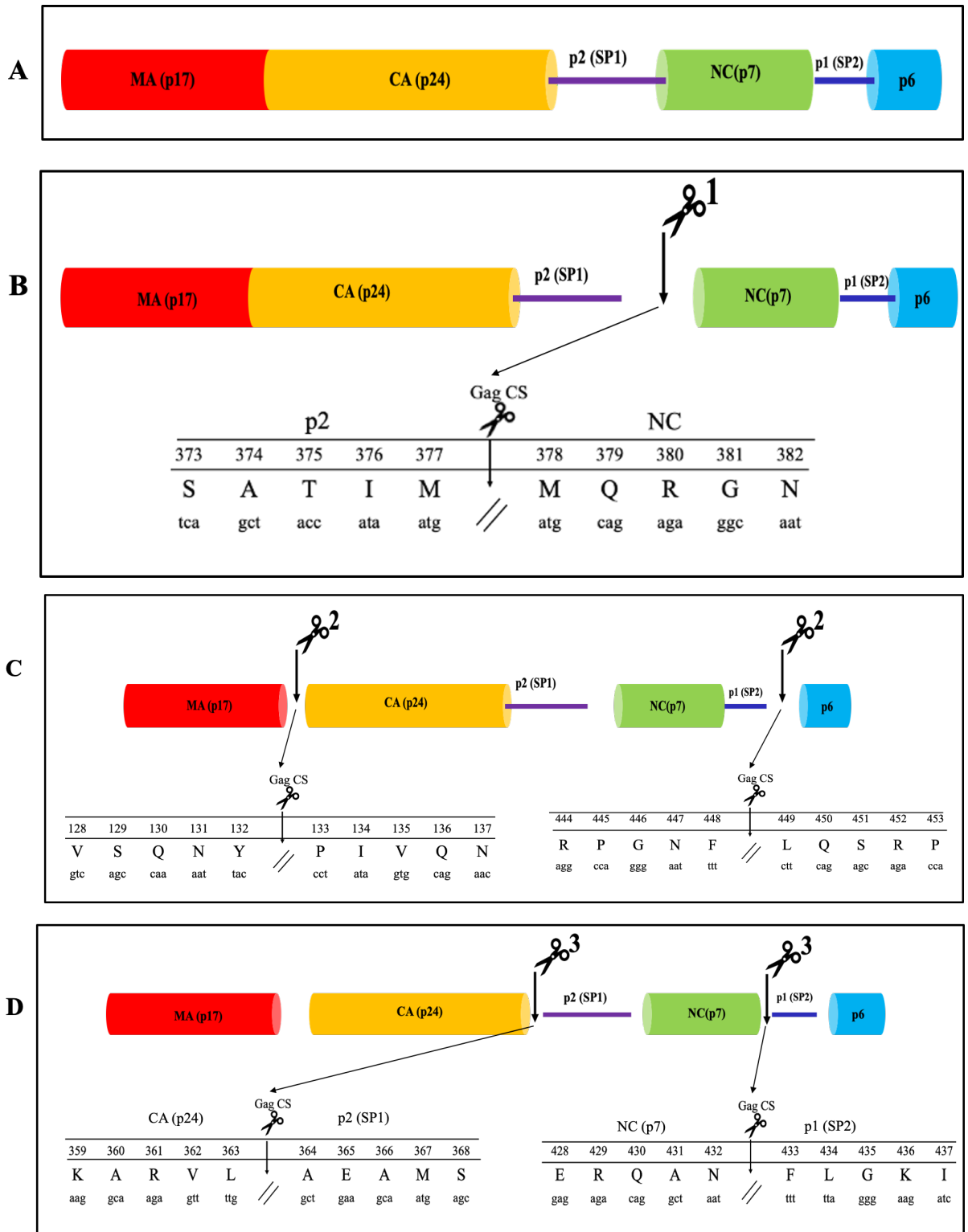
Figure 1.11 Ordered cleavage of the Gag polyprotein into its functional subunits by viral protease

Gag processing by protease cleavage is an essential part of the viral lifecycle. (A) Uncleaved Gag showing the four domains and the spacer peptides (B) The maturation cascade starts with the cleavage between spacer peptide 1 (SP1) and nucleocapsid (NC), (C) before cleavage between both matrix (MA) and capsid (CA) and SP2 and p6. (D) SP2 is removed from NC and the final stage of the maturation cleavage cascade is the removal of SP1 from CA.
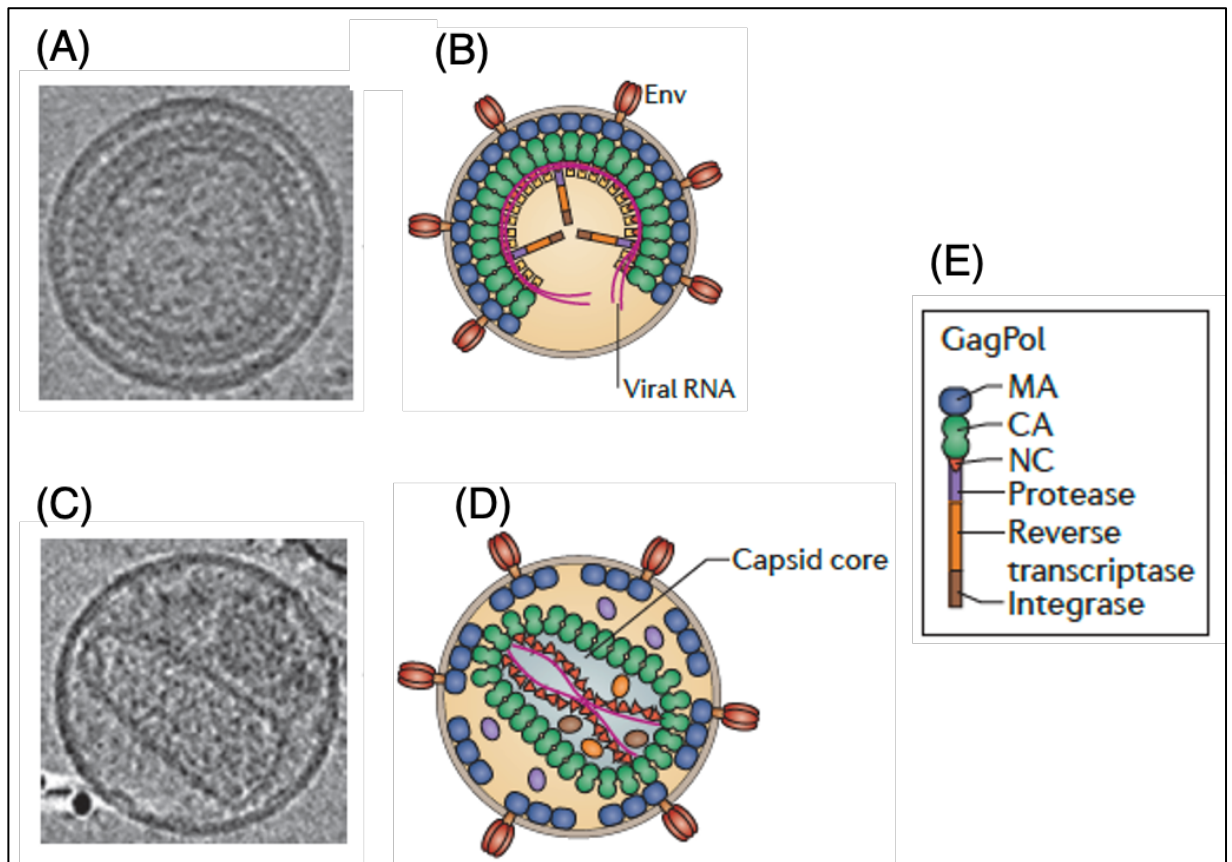
Figure 1.12: HIV-1 Maturation

Shown in this figure are (A). A cryo-electron tomograph and (B). a schematic representation of the immature virion. With (C) and (D) showing the mature HIV-1 virion. (E) shows a schematic of the Gag-pol polyprotein. Figures taken from: Eric Freed 2015[168]

1.9   Cell-to-cell spread

So far, the discussion of the HIV-1 replication cycle has been based on mature virion being released extracellularly and infecting a new target cell (cell-free infection). A second mechanism occurs in which the virus is not released into the extracellular environment but is transmitted through a direct cell-to-cell contact. In cell-to-cell spread, adhesive structures, called virological synapses (VSs) are formed between infected and uninfected T-cells. Cell-to-cell infection occurs through the formation of the virological synapse between infected (donor) and uninfected cells [202]. At the site of contact between cells, polarisation occurs between viral proteins and cellular organelles. Cell-to-cell spread is mediated by the viral Env protein in the infected cell and the CD4 receptor of the uninfected cell [203] and can reduce sensitivity to antiretroviral drugs.

Cell-to-cell infection is critical for efficient viral spread in vitro[204,205]; however, the extent of VS-mediated spread that occurs in vivo is unlear[206,207]. It may play a role in the spread of the virus in areas with dense populations of CD4+ cells such as lymph nodes.

By concentrating the release of viral particles at the site of cell-cell contact, HIV-1 cell-to-cell transmission increases the efficiency of viral spread several orders of magnitude compared to the dissemination of cell-free particles (a massive and very efficient infection that maybe 100–1,000-fold more efficient than infection carried out by cell-free viral particles[202,208-211]. It also provides protection from neutralizing antibodies; it overcomes the inhibitory effects of some anti-viral restriction factors such as tetherin, SAMHD1[212] and TRIM5α under certain conditions. Also, this mode of viral transmission influences the treatment and pathogenesis of the infection. Importantly, HIV-1 cell-to-cell transmission is a route that leads to the establishment of latent infection (as reviewed by Pedro et al., 2019)[213].

The persistence of low levels of viral replication in the presence of ART may be attributed to the cell-to-cell spread. This has raised questions as to how effective antiretroviral drugs are, in preventing this mode of transmission[214]. However, studies show that cell-to-cell spread can be effectively prevented by entry and

protease inhibitors [214],[215]. Another study showed that whilst protease inhibitors are equally effective at blocking cell-to-cell spread, certain reverse transcriptase inhibitors are between 4 and 20-fold less effective[216]. Cell-to-cell spread is also thought to enable immune evasion from the complement system and antibodies.

1.10   Host factors involved in the HIV-1 life cycle

Mammalian cells have evolved several mechanisms to prevent or block lentiviral infection and spread. This could be through the expression of several proteins whose function is to suppress viral replication. These proteins, called restriction factors, provide the first line of defence against infection. These proteins provide protection either as a component of or even preceding, innate antiviral responses. Figure 1.4. shows some of these restriction factors and where they act in the viral replication cycle.

Host restriction constitutes an intrinsic cellular defence. They impair HIV infection by blocking different steps of the HIV replication cycle. Viral restriction is so powerful that HIV needs to counteract these cellular factors for the establishment of a productive infection. For instance, viral accessory proteins Vif and Vpu inhibit APOBEC3G and tetherin antiviral functions, respectively [217]. HIV-1 can also avoid restriction factors through mutations as is the case for the HIV-1 capsid that evolved to escape from TRIM5α recognition.

Table 1 .2 summarises the functions of host restriction factors involved in the HIV life cycle.

Table 1.2: Summary of antiviral functions of host restriction factors involved in the HIV life cycle

Restriction Factor	Antiviral Function	HIV Counteraction	References
IFITM1/2/3	- Inhibit viral entry by reducing membrane fluidity - Negative imprinting of virions	Mutating Vpu and Env	[218, 219]
TRIM5α	- Premature uncoating - Targets viral capsid for proteasomal degradation	Viral capsids with reduced affinity for TRIM5α	[220]
APOBEC3G	- Deamination of cytidines to uracils during RT, creating hypermutated proviral DNA	HIV-1 Vif	[77]
SAMHD1	- Inhibits RT by decreasing the cellular pool of dNTPs - Degradation of HIV genomic RNA	HIV-2 Vpx	[221]
Mx2	- Inhibits HIV nuclear import - probably impairs uncoating		[222]
Tetherin (BST-2)	- HIV budding (viral entrapment) - Internalization and degradation of virions by the endosomal pathway	HIV-1 Vpu	[223,224]
CH25	- Inhibits viral entry by affecting membrane fluidity		[217]
ZAP	- Degradation of viral mRNA and inhibition of translation	HIV-Rev via Matrin3	[225,226]
SLFN11	- Inhibits HIV translation by preventing the change of tRNA pool composition		[227,228]
ISG15	- Inhibits HIV-1 release (ISGylation of Gag)		[229]
GBP5	- Decreases viral progeny infectivity by impairing incorporation of gp120 in budding viruses		[230]

1.11 Treatment of HIV-1 infection

There is currently no vaccine for HIV. Only two people have been cured of the disease; (i) the 'London patient' and (ii) the 'Berlin patient'.

The 'London patient' underwent one haemopoietic stem-cell transplantation (HSCT) procedure with CCR5Δ32/Δ32 donor cells for the treatment of refractory Hodgkin lymphoma and achieved remission after 18-months[231], and at 30-months [232] off antiretroviral treatment. The 'Berlin patient' underwent two rounds of total body irradiation and allogeneic-HSCT with donor cells that did not express CCR5 (CCR5Δ32/Δ32) for treatment of acute myelogenous leukaemia [233-235].

Four years after HIV was discovered as the cause of AIDS, the first antiretroviral (ARV) was licensed for use. More antiretroviral agents were subsequently approved for use by the FDA. In the first decade of ARV drug therapy, these agents did not fundamentally change the destiny of those with HIV infection, although they could decrease virus load, increase CD4+ cell number and prolong survival over the short term. The major shortcomings were drug toxicity, drug resistance and high drug cost (as reviewed by Zhang, 2018) [236].

Combination-based ARV therapy (cART) was introduced in 1996. cART is a three-component treatment, composed of drugs with at least two independent mechanisms of action. This led to effectively sustained HIV suppression, significantly recovered immune function, markedly improved clinical symptoms and notably extended lifespan. In the third decade, with further development of ARV drugs and the availability of multiple ART regimens and fixed dose combinations (FDCs), acquired immune deficiency syndrome (AIDS) has become a chronic, manageable and infectious disease [236,237].

When viral load is suppressed by treatment, it has been shown that patients' CD4+ T cell count rises again towards that of normal levels, resulting in improved immune function (as reviewed by Arts and Hazuda, 2012)[238]. Combination Antiretroviral treatment (cART) has been shown to reduce transmission rates

between serodiscordant couples and between infected mothers and their uninfected babies. It also improves the prognosis of individual patients.

Studies conducted in diverse settings have demonstrated that effective use of antiretroviral therapy (ART) results in viral suppression and eliminates the risk of HIV transmission[239-241], the so-called Undetectable (Viral Load) = Untransmissible (U=U)[242].

As presented in section 1.5.3, the main steps of HIV viral replication include binding and entry, reverse transcription, integration, viral assembly, and budding. These steps form the basis for the targets of the different ARV drug classes: (i) (Nucleos(t)ide reverse transcriptase inhibitors (NRTIs), (ii) Non- Nucleoside reverse transcriptase inhibitors (NNRTIs), (iii) Protease inhibitors (PIs), (iv) Integrase strand transfer inhibitors (INSTIs) and (v) entry inhibitors (sub-divided as fusion inhibitors (FI) and inhibitors of co-receptor usage.

Figure 1.14 shows a schematic representation of the various drug classes and where they act in the viral replication cycle.
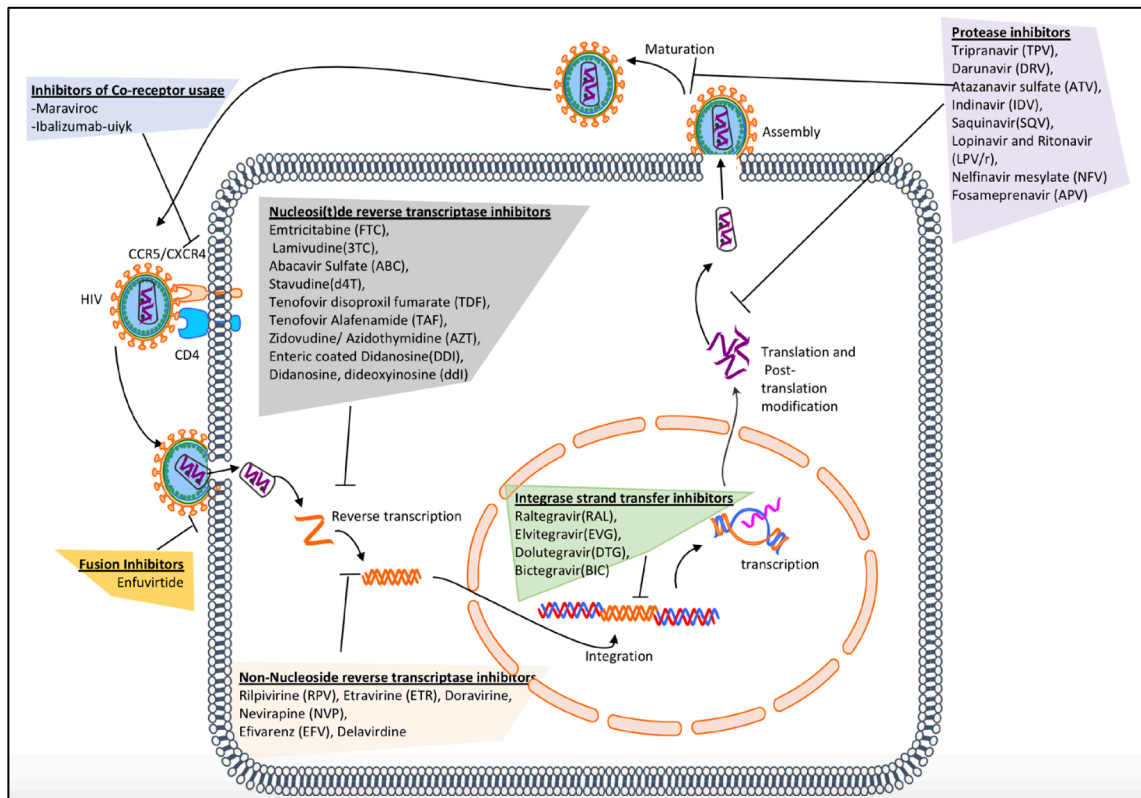
Figure 1.14: HIV replication cycle and antiretroviral drug target sites
FDA approved antiretroviral drugs for HIV treatment acts on different stages of the HIV-1 replication cycle. This figure shows the site of action of all approved HIV-1 drug classes. Adapted with permission from Njenda, DT, 2020 [243]

1.11.1 Entry and fusion inhibitors

Entry steps of the HIV-1 replication cycle are the targets for entry inhibitors. Two classes of entry inhibitors have received FDA approval. These are CCR5 co-receptor antagonists and fusion inhibitors

1.11.1.1 Fusion Inhibitors (FIs)

Currently, Enfuvirtide (also called T-20) is the only FI that is currently in use.

Fusion inhibitor (FI) drugs are delivered subcutaneously and have an extracellular mechanism of action upon entering the systemic system. FIs are in the form of short synthetic polypeptides (36 amino acids in length) and specifically, act as structural analogues that bind to the heptad region (HR-2) of HIV-1 gp41. HR-2 associates with HR-1 during hairpin formation to form the 6-helix bundle that facilitates attachment of the host cell and viral membranes [244,245,246,247].

Due to the need for a twice-daily administration at high doses, coupled with a very short half-life of less than four hours, Enfuvirtide has found limited use. Also, its subcutaneous mode of administration means it is expensive and can result in side effects at the injection site.

1.11.1.2 Chemokine receptor antagonists (CCR5 antagonists)

Maraviroc is the only drug in this class that has been approved by the FDA.

CCR5 antagonists are designed based on the knowledge that CCR5 is the co-receptor regularly involved in early infection. In contrast to many other drugs, maraviroc targets the host protein (CCR5), rather than the virus. It is a small molecule which selectively and reversibly binds the CCR5 coreceptor. This binding results in blocking the V3 loop interaction, thus inhibiting fusion of the cellular membranes. It is active against HIV-1 CCR5 tropic viruses and does not have any activity against CXCR4 tropic or dual/mixed tropic virus[248,249].

1.11.1.3   CD4-directed post-attachment inhibitors

The only FDA-approved drug in this class is Trogarzo (ibalizumab-uiyk) which was approved by the FDA in March, 2018 [250]. The drug is the first CD4-directed post-attachment HIV-1 inhibitor and the first humanised monoclonal antibody for the treatment of HIV/AIDS. It is approved in the USA for use as part of a combination antiretroviral regimen in heavily treatment-experienced patients with multidrug resistant (MDR) HIV-1 infection who have failed their regimens[251]. The recommended dose is a single intravenous 2000 mg loading dose followed by an intravenous maintenance dose of 800 mg once every 2 weeks[252].

The drug acts by binding to the extracellular domain 2 of the CD4 receptor. Its binding epitope is located at the interface between domains 1 and 2, opposite from the binding site for major histocompatibility complex class II molecules and gp120 attachment. Ibalizumab does not inhibit HIV gp120 attachment to CD4; however, its post-binding conformational effects block the gp120-CD4 complex from interacting with CCR5 or CXCR4 and thus prevents viral entry and fusion.

1.11.1.4   gp120 Attachment Inhibitors

The US FDA approved Rukobia (fostemsavir) in July 2020. It was approved as a new type of antiretroviral medication for adults living with HIV who have tried multiple HIV medications and whose HIV infection cannot be successfully treated with other therapies because of resistance, intolerance or safety considerations[253]. The recommended dosage of Rukobia is one 600-mg tablet taken orally twice daily with or without food[254].

Temsavir, the active metabolite of fostemsavir, binds directly to the viral envelope glycoprotein 120 (gp120), close to the CD4+ binding site. This novel mechanism of action locks gp120 into a closed state that prohibits the conformational change necessary for initial interaction between the virus and the surface receptors on CD4 cells, thereby preventing attachment and subsequent entry into host T cells and other immune cells[255-258].

1.11.2   Reverse Transcriptase Inhibitors

Reverse transcriptase inhibitors remain the central 'backbone' of antiretroviral therapy, since the first NRTI was introduced[259] (Zidovudine was the first RTI which got approval by the FDA in 1987). There are two distinct types of reverse transcriptase inhibitors, the nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs)

1.11.2.1   Nucleoside/Nucleotide reverse transcriptase inhibitors (NRTIs)

The nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) were the first class of antiretroviral drugs to be approved by the FDA. NRTIs are taken as prodrugs and must be taken into the host cell and phosphorylated before they become active. Once inside the host cell, cellular kinases will activate the drug [260]. The drug exerts its effect through its structure. NRTIs lack a 3'-hydroxyl group at the 2'-deoxyribosyl moiety and will have either a nucleoside or nucleotide as a base. Due to the missing 3'hydroxyl group, the NRTI prevents the formation of a 3'-5'-phosphodiester bond in growing DNA chains and can thus prevent replication of the virus. An interesting feature of these drugs is that their incorporation during RNA-dependent DNA or DNA-dependent DNA synthesis, inhibits the production of either positive or negative strands of the DNA [260,261]. Abacavir, Didanosine, Emtricitabine, Lamivudine, Stavudine, Tenofovir disoproxil fumarate, Tenofovir alafenamide AF, Zidovudine, Zalcitabine are some drugs in this class.

1.11.2.2   Non-nucleoside reverse transcriptase inhibitors

Drugs in this class include the following: Delavirdine, Efavirenz, Etravirine, Nevirapine, Rilpivirine and Doravirine. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are the second class of reverse transcriptase inhibitors. The primary mechanism of action is through the binding of the NNRTI to the reverse transcriptase in of a hydrophobic pocket proximal to the active site[262]. This binding creates a new spatial configuration of the substrate-binding site to reduce

the overall polymerase activity. By creating a different configuration, DNA synthesis becomes slowed overall. Because of the non-competitive inhibitor action of NNRTI, it is not effective against HIV-2 reverse transcriptase[261].

1.11.3 Integrase Inhibitors (INIs)

Raltegravir, Dolutegravir, Bictegravir, Elvitegravir are the drugs in this class, with Raltegravir being the first to be approved. These agents competitively inhibit the strand transfer reaction by binding metallic ions in the active site[263]. INIs belong to two major categories: (i) IN strand transfer inhibitors (INSTIs) that bind to the catalytic core domain of the enzyme to block the binding of the enzyme to dsDNA, and (ii) IN binding inhibitors (INBIs) that bind to the allosteric pocket of the IN and hamper the conformational changes required for strand transfer reaction. Currently, all USFDA-approved INIs belong to the group of INSTIs; whereas, therapeutic use of INBIs is still under the developmental phase[264]. INSTI enter the cell by passive diffusion or endocytic adsorption. Once in the cytoplasm, they can bind to the integrase enzyme, but only access the enzyme's active site once there has been a conformational change that is only induced after the enzyme catalyzes the 3'-end processing reaction.

Dolutegravir (DTG) belongs to the World Health Organization's (WHO) list of essential medicines and is currently thought to be the single most promising long-term anti-HIV medication available. In 2013, DTG received USFDA approval for third-line cART in patients who developed resistance to first- and second-line cART. DTG-containing regimens are now preferred first-line ART treatment in Low- and middle-income countries, as per WHO recommendations[265].

Mechanistically, DTG binds to the active site of HIV-1 IN which brings the β4-α2 loop of the enzyme into close contact with the extended linker region. This interaction leads to conformational changes that disengages IN from the deoxyadenosine present at the 3′ end of the virus. As a result, DTG chelates the divalent cations that are required for the enzyme function leading to inhibition of strand transfer reaction. The presence of a halobenzyl group in the structure

allows DTG to reach even farther in the active site, resulting in more stable interaction with the enzyme than RAL and EVG. This enables DTG to re-adjust to the conformational changes observed in RAL- and EVG- resistant mutants that make DTG capable of inhibiting IN mutant strains more effectively than other INIs[266] (as reviewed by Trivedi et al., 2020)[264].

1.11.4   Protease inhibitors

HIV-1 protease inhibitors are an essential component of current antiretroviral therapy (ART) which is responsible for dramatic improvement of life expectancy and mortality rates of HIV/AIDS patients in developing nations[267]. Nine protease inhibitors have so far been approved by the FDA for use for the treatment of HIV-1. These are generally classified as either first generation (Saquinavir, Ritonavir, Indinavir, Nelfinavir and, Fos(amprenavir), or second-generation PIs (Atazanavir, Darunavir, Lopinavir, Tipranavir).

As discussed previously, HIV protease catalyzes the hydrolysis of the Gag and Gag-Pol polyproteins at various cleavage sites which produce the structural proteins, What I was trying to get at here is that for you to include an explanation of mechanism of resistance of connection/RNaseH domain mutations i.e. they slow down RNaseH digestion of RNA genome template allowing time for inhibitor to be excised and polymerization to continue etc. Gag proteins, and the RT, IN, and PR enzymes for the new virion particles.

Protease Inhibitors are designed to mimic the natural substrates of the viral protease. They prevent the HIV-1 protease from cleaving the precursor proteins by precisely binding the active site of the virus protease. The hydroxyl group of the inhibitor interacts with the carboxyl group of the protease active site residues, Asp 25 and Asp 25′, by hydrogen bonds. The tight binding interactions of PI on the dimer surface of the homodimer HIV-1 protease active site inhibits the enzyme's functioning to execute aspartic acid-mediated cleavage of its

substrates[268]. The result of this inhibition affects the rate of viral assembly and leads to the production of defective and morphologically immature viral particles that are unable to establish another round of infection [269,270].

All PIs, except Tipranavir (TPV), contain a hydroxy ethylene core which cannot be cleaved by the viral protease[271,272]. In its place, TPV contains a dihydropyrone ring as a central scaffold[273] as shown in Figure 1.15

Saquinavir was the first FDA-approved HIV protease inhibitor used in the treatment of patients with AIDS (in 1995). The mean 50% effective concentration (EC50) of saquinavir against HIV-1 in MT4 cells is 37.7 nM[274]. The adult dose is twice daily, saquinavir 1,000 mg in combination with ritonavir 100 mg. Few side effects related to saquinavir have been reported[260]. Saquinavir is not a preferred protease inhibitor regimen due to its low bioavailability[275].
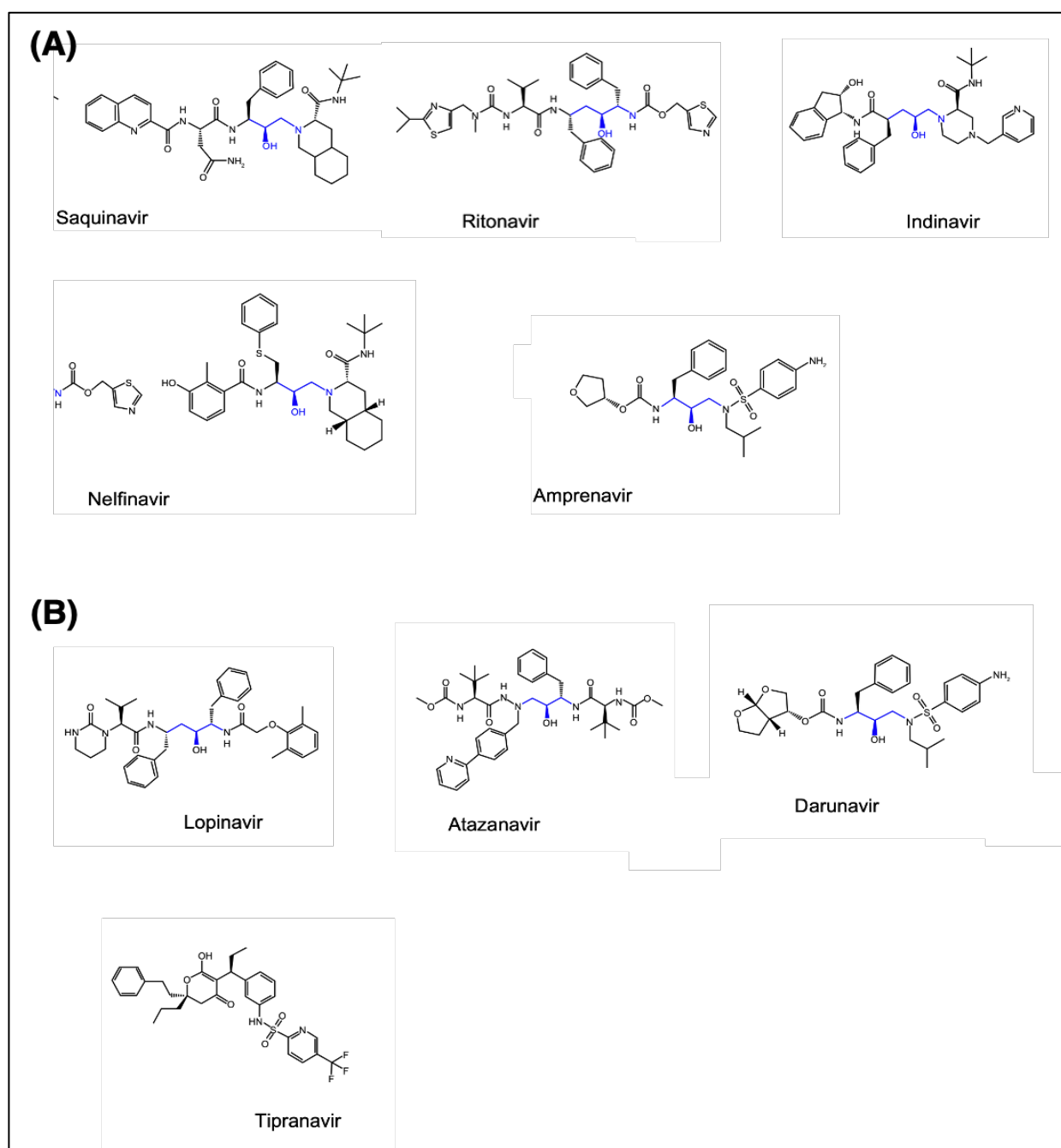
Figure 1.15: Chemical structures of each of the nine FDA-approved protease inhibitors

(A) first-generation and (B) second-generation HIV protease inhibitors. All FDA-Approved HIV-1 protease inhibitor shares the same structural similarities and a similar binding pattern with the presence of a hydroxy ethylene isostere core Tipranavir is the only member which has a dihydropyrone ring. The scaffolds are highlighted in blue. (Adapted from Lv et al., 2015)[276]

The next FDA-approved PI was Indinavir in 1996. An advantage of indinavir is its effective inhibition of both HIV-1 and HIV-2, while the disadvantage is the quick decrease in the concentration of circulating indinavir. The low plasma concentration of indinavir usually leads to treatment failures. The EC50 of indinavir is ~5.5 nM. Also, its low solubility may result in the development of kidney stones. Indinavir has a short-acting time and requires a dosage of 800 mg every 8 hours which has limited its use in clinical practice (as reviewed by Lv et al., 2015)[276]

Also approved for clinical use in 1996 was Ritonavir (RTV). It was initially designed as an HIV protease inhibitor (EC_{50} ~25 nM), but it was found later to boost the circulating concentration of other HIV protease inhibitors by inhibiting cytochrome P450 3A4 [277]. As a remarkably potent inhibitor of P450 3A4, a sub-therapeutic dose of ritonavir has been used to boost the plasma concentration of the second generation of HIV protease inhibitors, since HIV protease inhibitors are extensively metabolized by cytochrome P450 3A4[278]. Ritonavir-boosted protease inhibitor regimens require less frequent dosing, which benefits patients.

Nelfinavir (NFV) was approved in 1997. One terminus of the NFV molecule has the same DIQ group as SQV. The other terminus contains a 2-methyl-3-hydroxybenzamide group. The S-phenyl group at the P1 site was designed to magnify its potency. NFV EC_{50} is 30–60 nM, with preferred regimen of 1,250 mg orally, twice a day (as reviewed by Lv et al., 2015)[276].

Lopinavir (brand name: Kaletra), containing lopinavir and ritonavir, from Abbott Laboratories, was approved by the FDA in 2000 and was developed as a ritonavir-based agent. The core region of lopinavir, a hydroxyethylene dipeptide isostere, is the same as that of ritonavir. The P2 and P2′ group are altered in lopinavir relative to ritonavir. The 5-thiazolyl P2 group of ritonavir is replaced by a phenoxyacetyl group, and a six-member cyclic urea replaces the 2-isopropylthiazolyl P2′ group of ritonavir. In general, the new P2 and P2′ groups are smaller in order to decrease the contact with highly variable residues at the

82 site of HIV-1 protease[279]. The substitution of the P2 and P2′ groups improves the inhibitory potency of lopinavir against the drug-resistant variants of HIV-1 protease[279]. Lopinavir inhibits HIV protease activity, with the EC_{50} of ~17 nM. The dosage for adult patients is lopinavir 400 mg plus ritonavir 100 mg orally, twice a day.

Amprenavir was approved for clinical use in 1999. At the P1 site is a benzyl group with isobutyl group at the P1′ site. A sulfonamide connects the P1′ group and the phenyl amide P2′ group, with fewer chiral centres than do previous HIV PIs. This improvement simplifies chemical synthesis and increases oral availability[280]. The dosage of APV is 1,200 mg orally twice a day, with EC_{50} of 12–80 nM, and is less effective on HIV-2 protease than on HIV-1 protease[281].

The FDA approved fosamprenavir in 2003. It is the phosphate ester prodrug of amprenavir (APV) and is and is rapidly and extensively converted to amprenavir after oral administration. That metabolization increases the duration that APV is available, making it a 'slow-release version' of amprenavir and, thus, reducing the number of pills required versus standard amprenavir[282]. The recommended dose is fosamprenavir 1,400 mg in combination with ritonavir 100 mg orally, twice a day.

Atazanavir was also approved in 2003. It is an aza-dipeptide analogue, which exhibits potent anti-HIV activity. The EC_{50} of atazanavir in cell culture is 2.6–5.3 nM. A unique characteristic of atazanavir is the presence of a large phenylpyridyl P1 group that is asymmetric relative to its benzyl P1′ group. Atazanavir shows good oral bioavailability[283]. Thus, the benefit of this is once-a-day dosing with atazanavir 300 mg plus ritonavir 100 mg. Fewer side effects are associated with atazanavir than with other protease inhibitors[276,284,285].

Tipranavir is the only non-peptidomimetic HIV protease inhibitor and received approval from the FDA in 2005. Innovation in the tipranavir design is the functional substitution of the bridging water molecule connecting the inhibitor and protease flaps. The lactone oxygen atom of the dihydropyrone ring of tipranavir interacts directly with the Ile50 residues in the flap region of the HIV-1 protease.

The direct interaction stabilizes the protease-inhibitor complex. Tipranavir inhibits the HIV-1 protease that has developed resistance to other protease inhibitors (as reviewed by Lv et al., 2015)[276]. Tipranavir inhibits the replication of HIV-1 isolates, with EC_{50} ranging from 30–70 nM. The dosage is tipranavir 500 mg plus ritonavir 200 mg orally, twice a day[276].

Darunavir, approved in 2006, was the last HIV protease inhibitor approved by the FDA. The hydrogen bonds that DRV forms with the backbone of the HIV protease slows down the development of drug resistance[286]. The structure of darunavir is very similar to that of amprenavir, the only difference being that in darunavir, the P2 group bis-tetrahydrofuran replaces the tetrahydrofuran group of amprenavir. This change allows darunavir to have more hydrogen bonds with the Asp 29 residues of HIV protease[287].

Structure-based design strategies were used to develop darunavir with particular emphasis on maximizing active site interactions, especially promoting strong hydrogen-bonding interactions with the HIV-1 backbone atoms in the S2 and S2′ subsites. Its favourable properties include a picomolar affinity for protease, a high genetic barrier to resistance,[288,289] inhibition of precursor autoprocessing[171,290] and inhibition of protease dimerization[291]. Darunavir has an exceptional resistance profile due to a dual mechanism of action as it inhibits HIV-1 protease and inhibits dimerization of HIV-1 protease monomers[292].

The EC_{50} of DRV is as low as 1–2 nM; it can inhibit both HIV-1 and HIV-2 with high potency[281,293]. The dosage is darunavir 800 mg plus ritonavir 100 mg orally, once a day.

1.11.5 Maturation Inhibitors

HIV-1 maturation inhibitors are a class of small-molecule compounds that block a late step in the viral protease-mediated processing of the Gag polyprotein precursor, causing defective core condensation and the release of non-infectious virus particles from infected cells, thus blocking the spread of the infection to new cells [294].

The first maturation inhibitor to be investigated in detail was bevirimat (BVM), which was shown to prevent replication of infectious virions by interfering with CA–SP1 cleavage [295,296]. It acts by binding within six-helix bundles formed by a segment that spans the junction between the CA and spacer peptide 1 (SP1) subunits of Gag, and interferes with cleavage between CA and SP1 catalyzed by the HIV-1 protease [297].

Despite promising data in a phase IIa clinical trial, further development of bevirimat was suspended in 2010 due to bevirimat-resistance conferring Gag SP1 polymorphisms (at Gag amino acids 369, 370 and 371) present in approximately 50% of HIV-1-infected patients[298-301].

1.12 Resistance to Antiretroviral drugs

Resistance is the naturally occurring response of any microorganism facing the selective pressure of drugs, and HIV does not represent an exception to this rule. The high level of viral replication and turnover and the lack of a proofreading mechanism of HIV reverse transcriptase leads to the spontaneous generation of a large number of genetically distinct viral *quasispecies* coexisting in the same person.

The high error rate and low RNA affinity of HIV-1 reverse transcriptase and the high level of replication is a major contributor to the development of HIV drug resistance (HIVDR). With the HIV-1 reverse transcriptase (RT) enzyme lacking the 3' – 5' exonucleolytic proof-reading activity, the process of reverse transcription is highly error-prone. Several *in vitro,* cell culture studies have estimated the error rate to be as high as 3×10^{-5} per base per round [302-306]. In *in vivo* studies, Cuevas and colleagues used sequences derived from both intracellular viral DNA and plasma viral RNA and reported an estimated $(4.1 \pm 1.7) \times 10^{-3}$ mutations per base per cell the highest reported for any biological entity [307]. This translates to an arbitrary calculation of about 1-36 mutations occurring in each RNA genome target per round of replication [243].

Viral genes are usually the target for most antiretroviral drugs. When mutations occur in these genes, or at other areas of the viral genome, optimal drug function may be compromised. Hence drug resistance arises[308].

There are other factors which contribute to HIVDR. These include; drug-drug interactions, viral factors such as subtypes, sub-optimal drug levels in patients (, caused by non-compliance or by sub-optimal treatment regimens). These factors are summarized in Figure 1.15.

Drug resistance can either be acquired through drug selection pressure (acquired resistance) or transmitted from person to person (transmitted resistance). HIV-1 drug resistance mutations can be described as primary (major) or secondary, which encompasses additional descriptive terms, such as minor and accessory or compensatory mutations. Primary mutations are those that reduce drug susceptibility by themselves. Secondary (or accessory) mutations reduce drug susceptibility in combination with primary mutations or they improve the replicative fitness of virus isolates with a primary mutation. However, which mutations are considered primary and which are considered secondary are not strictly defined, and some mutations might be considered to be primary for one drug but secondary for another drug [309,310].

In addition to the risk posed by resistance to individual patients, the transmission of drug-resistant viruses leads to virus persistence in the absence of selective drug pressure. This complicates treatment of patients with newly acquired drug-resistant infections in the future[311,312].
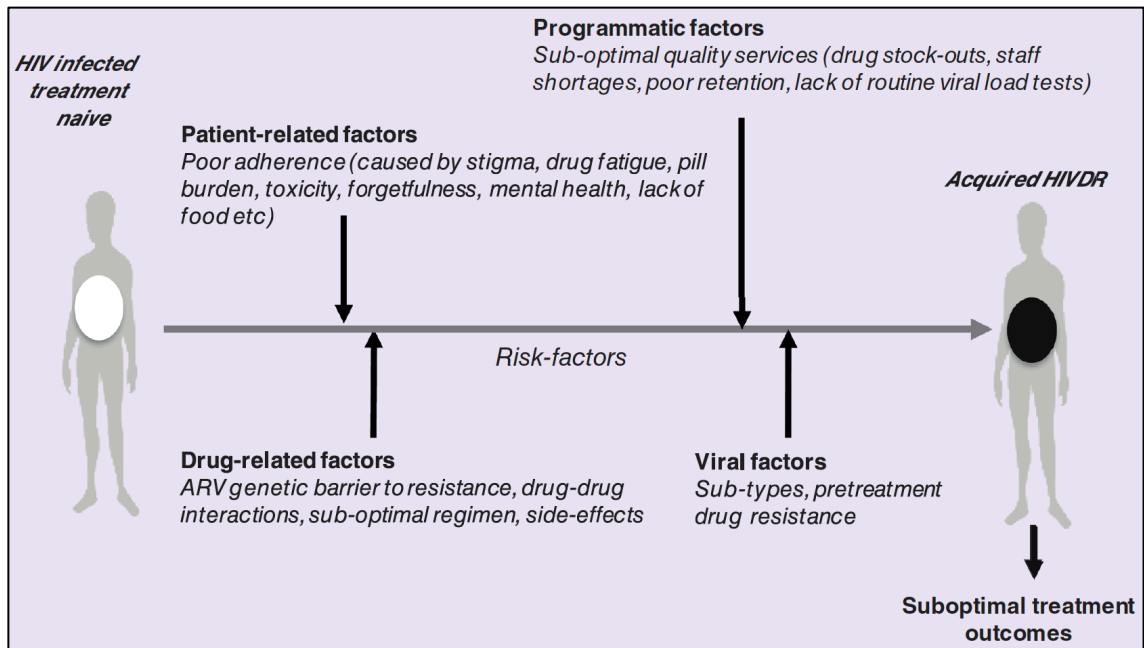
Figure 1.15: Factors associated with the emergence of acquired HIV drug resistance in antiretroviral therapy treated populations

Several factors could lead to the emergence of acquired drug resistance during cART. These include patient-related factors, drug-related factors, programmatic factors and viral factors. Image adapted from Inzaule et al., 2019)[313] – used with permission

1.12.1 Resistance to Entry Inhibitors

Reduced Ibalizumab susceptibility is associated with mutations that disrupt potential N-linked glycosylation sites (PNGS) in variable region 5 (V5) of HIV-1 envelope glycoproteins[314]. Loss of glycan on the V5 N-terminus of gp120 is considered a major determinant of ibalizumab resistance[315,316].

Amino acid substitutions occur in the 36-45 regions and result in significant loss of enfuvirtide activity[317]. Resistance arises as a result of mutations at the T20 binding site within gp41 including G36D, I37T, V38A/M, N42T/D and N43K[318]. These mutations reduce the affinity of Ibalizumab to gp41. However, they also confer reduced fitness as a result of decreased fusion efficiency. Compensatory mutations in the CHR domain of gp41 occur that can restore the viral fusion kinetics and thus viral fitness[319]. Susceptibility to enfuvirtide may be affected by mutations or polymorphisms in other regions of env (such as the HR2 region). Additionally, coreceptor usage and density may also affect susceptibility[320-322].

Resistance to Maraviroc occurs via two different mechanisms: The first mechanism is most likely through amino acid substitutions in the V3 loop of gp120. The second mechanism is not acquired resistance but rather the inability of phenotypic tropism assays to detect small quantities of CXCR4 virus that may be present, leading to overgrowth of CXCR4 virus in the presence of maraviroc and loss of viral control[248,323].

1.12.2 Resistance to reverse transcriptase inhibitors

1.12.2.1 Resistance to NRTIs

There are two mechanisms of NRTI resistance: (i) discriminatory mutations and (ii) primer unblocking mutations.

Discriminatory mutations enable the reverse transcriptase to discriminate between dideoxy-NRTI chain terminators and the cell's naturally produced dNTPs. This discrimination prevents NRTIs from being incorporated into a growing viral DNA[324]. Examples of these include; M184V, Q151M, and K65R mutations.

Primer unblocking mutations facilitates the phosphorolytic excision of an NRTI-triphosphate that has been added to the growing viral DNA chain[324,325]. Primer unblocking mutations are also referred to as thymidine analogue mutations (TAMs) because they are selected by the thymidine analogues zidovudine and stavudine[309]. The TAMs include M41L, D67N, K70R, L210W, T215F/Y and K219Q/E.

Mutations which occur in the connection (289–423) and RNAseH (424–560) domains can affect RTI susceptibility alone or in combination with mutations in the N terminal region. These mutations are located away from the binding sites of RTIs and therefore exclude any direct effects on their binding to RT.

RNase H activity is required to cleave the RNA moiety of RNA/DNA replication intermediates. Connection domain mutations reduces RNase H activity and increases resistance to NRTI through both RNase H-dependent and RNase H-independent mechanisms. In the RNAseH-dependant mechanism, the mutations reduce RNase H cleavage hence contributing to NRTI-resistance by providing more time for RT to carry out nucleotide excision and resume productive DNA synthesis. These mutations increase NRTI resistance synergistically with TAMs. In the RNase H-independent mechanism, connection domain mutations contribute to resistance by directly improving the ability of RT to carry out nucleotide excision. Examples of these mutations that confer resistance to NRTIs include: G335C, N348I, A360I/V, T369I/V, A376S, E399D and G333D/E (as reviewed by Delviks-Frankenberry et al., 2010)[326]. Interestingly, these mutations also increase resistance to NNRTIs[327,328].

Connection domain mutations slow down RNaseH cleavage and digestion of RNA genome template hence providing more time for NNRTI to dissociate from the RT. This results in the resumption of DNA synthesis thus enhancing NNRTI resistance[329,330]. Several resistance-conferring connection and RNASe H domain mutations have been described. D549N, Q475A, and Y501A mutants enhances resistance to nevirapine (NVP) and delavirdine (DLV) by reducing RNase H cleavage. Additionally, A400T mutation was found to change the conformation of the RNaseH primer grip region leading to slower degradation of the viral RNA genome thereby providing more time for dissociation of the bound NNRTI from

the stalled RT-template/primer complex, after which reverse transcription can resume resume[331].

The N348I mutation has been reported to confer resistance to zidovudine and nevirapine[332]. The acquisition of N348I in HIV-1 reverse transcriptase may provide a simple genetic pathway that allows the virus to select both TAMs and mutations that are antagonistic toward TAMs[333].

1.12.2.2   Resistance to NNRTIs

When mutations occur in the NNRTI binding pocket, the ability of NNRTIs to bind the enzyme is altered, thereby leading to NNRTI resistance. The NNRTIs have a low genetic barrier to resistance. For example, high-level resistance to nevirapine (NVP) generally requires one mutation, high-level resistance to efavirenz (EFV) generally requires one to two mutations, and high-level resistance to etravirine (ETR) requires two mutations[309].

The following are some common NNRTI resistance mutations: A98G, L100I, K101E, K103N, V106A/I, V108I, V179D Y181C, Y188C/H, G190A, P225H, M230L, P236L and Y318W[334,335].

1.12.3   Resistance to Integrase Inhibitors

The geometry of three residues stabilizes the catalytic core of the integrase enzyme. These are termed the 'DDE' triad (amino acid residues D64, D116, and E152)[336]. Mutations that occur in the catalytic core induces conformational changes in the core, hence increasing binding energy requirements of integrase strand transfer inhibitors (INSTIs) and leading to resistance. The N155H and QI48H mutations are typical examples.

Raltegravir failure is associated with integrase mutations in at least three distinct, but not exclusive, genetic pathways defined by two or more mutations including (i) a mutation at Q148H/K/R (up to a 25-fold decrease in susceptibility[337]), N155H (up to a 10-fold decrease in susceptibility[337]), or Y143R/H/C; and (ii) 1 or more additional minor mutations. Minor mutations described in the Q148H/K/R pathway include L74M plus E138A, E138K, or G140S[338]. The most common

mutational pattern in this pathway is Q148H plus G140S, which also confers the greatest loss of drug susceptibility (up to 100-fold decrease in susceptibility)[339]. High-level resistance to elvitegravir is associated with mutations at E92Q in combination with E138K, Q148K/R/H, or N155H, leading to a 150-fold loss of susceptibility. Resistance patterns involving Q148H/G140S and Q148R/G140S demonstrate resistance to both elvitegravir and raltegravir, suggesting cross-resistance is likely[340]. Resistance to elvitegravir could also be caused by Q148R + any of L74I/M, E138K/A/D/T, G140A/S, Y143H/R, E157Q, G163E/K/Q/R/S, G193E/R R236K, N155H, and S230R[339,341].

Dolutegravir exhibits a higher genetic barrier to resistance, attributed to its prolonged dissociation half-life from HIV-1 integrase-DNA complexes, and is considered a second-generation INSTI. The presence of Q148R with two or more of the following INSTI mutations (L74I/M, E138K/A/D/T, G140A/S, Y143H/R, E157Q, G163E/K/Q/R/S, G193E/R) is associated with a substantially lower response to dolutegravir[341]. Treatment-emergent resistance is uncommon with dolutegravir but has been associated with the development of INSTI mutations at R236K, N155H, and S230R[342]. [82]

Similar to dolutegravir, bictegravir has an extended dissociation half-life from HIV-1 integrase-DNA complexes and exhibits a higher barrier to resistance than first-generation INSTIs (raltegravir, elvitegravir). Limited information presently exists regarding the resistance profile of bictegravir; however, invitro, site-directed mutagenesis studies demonstrate that M50I, S153F, R236K, and M50I + R236K resulted in 1.3, 1.9, 2.2, and 2.9-fold reductions respectively in susceptibility to bictegravir in cell culture[343].

Mutations outside of the integrase gene have been shown to play a role in resistance to integrase inhibitors. For example, in the presence of DTG, mutations (E209K, A539V, and H641Y) were seen to arise in Env which helped the virus to circumvent restriction by providing an escape for the drug[344]. Additionally, using in vitro experiments, five nucleotide mutations, all located in the *Nef* region, of which four were clustered in the 3' polypurine tract (PPT) have

been shown confer a high level of resistance to dolutegravir, raltegravir, and elvitegravir[345].

1.12.4   Resistance to protease inhibitors

The initial belief was that resistance to protease inhibitors during treatment would be infrequent given the small size and the vital role HIV-1 protease plays in the viral life cycle. The protease gene has, however, been shown to have high plasticity; mutations have been detected in 49 out of the 99 amino acids, and more than 20 substitutions associated with resistance to PIs [346].

Resistance to protease inhibitors often arises as a result of a stepwise accumulation of primary (major) mutations and secondary (minor, compensatory or accessory) mutations [347]. Each inhibitor usually selects for primary signature mutations and a characteristic pattern of secondary mutations.

Multiple substitutions are therefore required for the development of complete PI resistance while maintaining active virus replication and maturation. Primary resistance mutations are generally located near the active site of the protease, at positions involved in inhibitor and substrate binding. These mutations often have a deleterious effect on the replication capacity of the resistant virus[348,349]. The emergence of secondary mutations in the protease alleviates the negative effect of the primary mutations. Secondary mutation amino acid changes are generally outside the substrate-binding cavity of the enzyme and promote adaptation to the primary changes observed in the protease and compensate at least partially the impairment of HIV-1 replication[350-353].

Lopinavir-boosted ritonavir (LPV/r) and Darunavir-boosted ritonavir (DRV/r) are two PIs that have the highest genetic barriers, requiring at least three mutations in LPV/r and more in DRV/r for high-level drug resistance to occur[354]. Consequently, the multi-step mechanism and high genetic barrier of these PIs results in a complex combination of mutational pathways leading to drug resistance[355].

1.13 Resistance mutations in protease

The Stanford Resistance Database describes PI resistance mutations at 35 of the 99 amino acids positions in protease. These are collated every year by the International AIDS Society (as shown in Figure 1.16). Mutations at 20 of these positions are accessory (secondary) mutations, while fifteen are described as major (primary) mutations. The major mutations confer high levels of phenotypic resistance or show evidence of a reduced response to treatment *in vivo* for at least one PI. These 15 positions are: 30, 32, 46, 47, 48, 50, 54, 58, 74, 76, 82, 83, 84, 88 and 90 [338]. Of these 30, 32, 48, 50, 82 and 84 are located in the substrate-binding cleft of protease and 46, 47 and 54 near the cleft [356,357]. These mutations confer resistance by enlarging the binding cleft, thus reducing the affinity of the PI [358]. Mutations at positions 76, 88 and 90 affect PI susceptibility indirectly and are not located at or near the substrate binding cleft[359]. L76V and N88S mutations, for example, affect the hydrogen bond network in the protease's active site, hence the indirect effect of these mutations on inhibitor binding [360].

In addition to major mutations, there are also a couple of accessory mutations found in protease which affect PI susceptibility primarily through compensatory mechanisms. These cause reduced PI susceptibility when found in combination with other PI resistance mutations. Some mutations are found in a significant proportion of treatment-naïve viruses, especially in non-B subtypes. These mutations are called polymorphic mutations. Mutations which are observed to arise only as a result of PI exposure are known as non-polymorphic mutations.

Several mutations are associated with atazanavir resistance. Their impacts differ, with I50L, I84V, and N88S having the greatest effect [361]. Mutations that are selected during unboosted atazanavir are not different from those selected during boosted atazanavir, but the relative frequency of mutations may differ [362]. Higher atazanavir levels obtained with ritonavir boosting increase the number of mutations required for loss of activity. The presence of M46I plus L76V might increase susceptibility to atazanavir when no other related mutations are present[363]. The I50L mutation is the most commonly described in vivo mutation associated with ATV, indeed it is the signature resistance mutation for ATV [364].

In *in vitro* passage experiments with ATV, there was the emergence of the following mutations in the resistant strains: N88S, I84V, I50L, M46I and V32I [365]. Phenotypically, an association has been established between L90M, I84V, V82A/F/S/T, G73C/S/T/A, A71V/I/T/L, G48V, M46I/L, L33I/F/V, L24I, K20R/M/I and L10I/V/F with reduced ATV susceptibility [364].

Virologic response to ritonavir-boosted darunavir correlates with baseline susceptibility and the presence of several specific PI resistance-associated mutations. The negative impact of the protease mutations I47V, I54M, T74P, and I84V and the positive impact of the protease mutation V82A on virologic response to ritonavir-boosted darunavir were shown independently in 2 data sets [366,367]. Some of these mutations appear to have a greater effect on susceptibility than others (eg, I50V vs V11I). The presence at baseline of 2 or more of the substitutions V11I, V32I, L33F, I47V, I50V, I54L/M, T74P, L76V, I84V, or L89V was associated with a decreased virologic response to ritonavir-boosted darunavir [368]. Darunavir exhibits a high genetic barrier to the development of drug-resistant viruses relative to other PIs [369,370].

Virologic response to ritonavir-boosted lopinavir is affected by the presence of 3 or more of the following amino acid substitutions in protease at baseline: L10F/I/R/V, K20M/N/R, L24I, L33F, M36I, I47V, G48V, I54L/T/V, V82A/C/F/S/T, and I84V. Also, the combination of 47A/V with V32I is associated with high-level resistance [356,363,371-374]. I50V is only occasionally selected *in vivo* but has a clear impact on susceptibility[375-377]. Subtype C patterns with M46L, I54V, L76V, and V82A are frequently observed in patients receiving ritonavir-boosted lopinavir. In some non-subtype-B HIV-1, D30N is selected less frequently than are other PI resistance-associated mutations.

Several major mutations arising from APV/r exposure have been described, including the following: L90M, I84V, V82A/T/S/F, L76V, I54A/T/A/L/M, I50V, I47V/A, I47V/A, M46I/L and V32[378]. Additionally, G73S and L10F/I/R/V accessory mutations have been observed. A study by Maguire and colleagues identified four

main pathways of APV/r resistance in PI-naïve patients (either I84V, I54L/M, I50V or V32I with I47V) [379]. Interestingly, the same pathways to resistance were observed in patients who are PI-experienced [380]. In *in vitro* passage experiments with APV/r, I50V mutation was observed, accompanied by L10F, M46I/L or I47V were identified as is the key resistance pathway [381,382]. This difference between *in vivo* and *in vitro* observations could have occurred because the most significant reduction in susceptibility has been conferred by I50V which also have the most significant impact on viral fitness.

The following major resistance mutations have been reported to arise from boosted Indinavir (IDV/r): M46I/L, V82A/F/T, I84V. Also described are the following accessory mutations: L10I/R/V, K20M/R, L24I, V32I, M36I, I54V, A71V/T, G73S/A, L76V, V77I and L90M[338,378]. The M46I/L and V82A/F/T mutations were identified as accounting for most treatment failures with IDV in clinical trial. However, in phenotypic studies, these mutations singly did not appear to confer reduced susceptibility [383]. Following therapy *in vivo*, changes have been observed at the following positions 46, 54, 71, 82, 89 and 90 [384].

The following major resistance mutations have been described for NFV: D30N and L90M. The following accessory mutations have also been described: L10F/I, M36I, M46I/L, A71V/T, V77I, V82A/F/T/S, I84V, N88D/S [378], [338]. Mutations at positions 48, 82, 84 and 90 have been associated with reduced rates of virological response in patients[385]. Also, the K20I mutation has been shown to reduced susceptibility to NFV when present in combination with other mutations [386]. L23I is a rare mutation positioned in the substrate-binding cleft which has been shown to cause low-level resistance to NFV[387]. The development of L10I, D30N, M36I, V77I, N88S/D and L90M in response to NFV/r-based therapy has also been reported [388].

The G48V and L90M mutations have been reported to confer resistance to boosted Saquinavir (SQV/r)[338], in addition to accessory mutations (L10I/R/V, L24I, I54V/L, I62V, A71V/T, G73S, V77I, V82A/F/T/S AND I84V [378]. Mutations G48V, V82A and L90M were observed following *in vitro* passage of virus with

SQV [389]. Mutations at position 73 – G73S/T/C/A – are important accessory mutations for SQV/r, particularly in combination with the L90M major resistance mutation. L90M is the most commonly occurring major resistance mutation on failure with SQV and has been reported to occur with I84V, conferring a significant reduction in susceptibility [390].

For TPV/r, major resistance mutations which have been described are: I47V, Q58E, T74P, V82L/T, N83D and I84V. The accessory mutations include; L10V, L33F, M36I/L/V, K43T, M46L, I54A/M/V, H69K and L89I/M/V [338]. The major mutations V32I, V82A and I84V, were most commonly observed following *in vitro* passage of virus with TPV/r [391]. As TPV/r is approved for use in previously PI-exposed patients, V82T develops only in the presence of the major mutation V82A at baseline. In patients who had wild-type virus at baseline, V82L was observed [392]. The best predictors of viral failure to TPV/r are; T74P, I47V, V82L/T, Q58E and N83D. In combination with other mutations, the L33F accessory mutation is the most commonly observed [393].

Atazanavir +/- ritonavir[18]

L10	G16	K20	L24	V32	L33	E34	M36	M46	G48	I50	F53	I54	D60	I62	I64	A71	G73	V82	I84	I85	N88	L90	I93
I	E	R	I	I	I	Q	I	I	V	**L**	L	L	E	V	L	V	C	A	**V**	V	**S**	M	L
F		M			F		L	L			Y	V			M	I	S	T				L	M
V		I			V		V					M			V	T	T	F					
C		T										T				L	A	I					
		V										A											

Darunavir/ ritonavir[19]

V11	V32	L33	I47	I50	I54	T74	L76	I84	L89
I	I	F	**V**	**V**	**M**	P	**V**	**V**	V
					L				

Lopinavir/ ritonavir[20]

L10	K20	L24	V32	L33	M46	I47	I50	F53	I54	L63	A71	G73	L76	V82	I84	L90
F	M	I	**I**	F	I	**V**	**V**	L	**V**	P	V	S	**V**	**A**	**V**	M
I	R				L	**A**			**L**		T			F		
R									**A**					T		
V									**M**					S		
									T							
									S							

Tipranavir/ ritonavir

L10	L33	M36	K43	M46	I47	I54	Q58	H69	T74	V82	N83	I84	L89
V	F	I	T	L	**V**	A	**E**	K	**P**	**L**	**D**	**V**	I
		L				M		R		**T**			M
		V				V							V

Fosamprenavir/ ritonavir[21]

L10	V32	M46	I47	I50	I54	G73	L76	V82	I84	L90
F	I	I	V	**V**	L	S	V	A	**V**	M
I		L			V			F		
R					M			S		
V								T		

Indinavir/ ritonavir[21]

L10	K20	L24	V32	M36	M46	I54	A71	G73	L76	V77	V82	I84	L90
I	M	I	I	I	**I**	V	V	S	V	I	**A**	**V**	M
R	R				**L**		T	A			**F**		
V											**T**		

Nelfinavir[21,22]

L10	D30	M36	M46	A71	V77	V82	I84	N88	L90
F	**N**	I	I	V	I	A	V	D	**M**
I			L	T		F		S	
						S			
						T			

Saquinavir/ ritonavir[21]

L10	L24	G48	I54	I62	A71	G73	V77	V82	I84	L90
I	I	**V**	V	V	V	S	I	A	V	**M**
R			L		T			F		
V								S		
								T		

Figure 1.16 Mutations in the protease gene associated with PI resistance

Diagram showing the PI resistance mutations within protease that have been documented to date. The mutations for each of the nine FDA approved PIs are shown, with major mutations shown in bold. Bars shaded in grey are antiretroviral drugs that are no longer recommended. Ritonavir is not listed separately, as it is currently used only at low doses as a pharmacologic booster of other PIs. Figure produced by the International AIDS Society (IAS)[338]

1.14 The Role of Gag Mutations in Resistance to Protease Inhibitors

Mutations in Gag cleavage sites may confer or contribute to resistance to PIs and may even emerge before mutations in protease[394]. A large proportion of virus samples from patients with confirmed virologic failure on a PI-containing regimen is not found to have PI resistance-associated mutations. Out of the total of 12 protease cleavage sites within the HIV genome, 5 of them are in the Gag gene. The gag gene has therefore been shown to play an important role in resistance to protease inhibitors[394-398].

Mutations in the gag gene which have previously been reported to be associated with PI-exposure or reduced PI-susceptibility (resistance) are presented in Table 1.3.

1.14.1 Gag cleavage site mutations (CSMs) and PI resistance

HIV-1 resistance mechanisms to PIs have been proposed to include the development of mutations in the Gag gene at protease cleavage sites. Gag mutations at cleavage sites are hypothesized to cause an interaction with the active site of protease, reducing the binding affinity of PIs and thereby allowing the enzyme to resume its proteolytic function[196,399,400].

Gag cleavage site mutations could lead to reduced susceptibility to PIs in a variety of ways. The mutations could improve interactions between the substrate and the mutated enzyme and correspondingly increase cleavage. Alternatively, Gag cleavage site mutations would improve the efficiency of ribosomal frameshift at the Gag-Pol junction, thus increasing the amounts of protease[401].

It is thought that Gag cleavage site mutations increase PI resistance when they occur with primary drug resistance mutations in protease. Some common Gag cleavage site mutations that have been reported include: V128I (MA/CA); S373P, I376V (p2/NC); A431V and K436R (NC/p1) and P453A(p1/p6)[395]. Mutations at Gag cleavage sites confers resistance, independent of compensating for reduced fitness via several possible mechanisms. The kinetics of cleavage may be improved by Gag mutations beyond the level of compensating for fitness, and even in the absence of PIs. An alternative mechanism could be that with the

accumulation of mutations in the cleavage site, the substrate is able to more effectively compete with PIs for the active site of the protease enzyme[395].

Following in vitro passage of the virus with protease inhibitors, the C terminus of Gag (NC/p1/p6 region) was identified as harbouring most of the CSMs[402,403]. The same was observed *in vivo* among patients undergoing PI treatment, some of whom were failing therapy[404,405]. In *in vitro* studies following passage with PIs, the most frequently-observed cleavage site mutation was the A431V mutation which occurred in the NC/p1 cleavage site[402,403]. The same was observed *in vivo* in patients undergoing therapy with SQV, NFV, RTV and IDV[197,404]. In vitro phenotypic studies showed that viruses harbouring the A431V had reduced susceptibility to all protease inhibitors (except darunavir)[196,197].

Amino acid changes have been reported at positions 436 and 437 in the NC/p1 cleavage site following *in vitro* and *in vivo* exposure to PIs[384,404,406]. In the absence of protease mutations, susceptibility to PIs is directly affected by these two mutations[197]. *In vitro* passage with a PI produced a number of mutants exhibiting substitutions in the Gag p2 spacer peptide – K436E and I437T/V – and none in protease. Synthesis of a recombinant virus with wild-type protease showed that the presence of the mutant Gag alone significantly increased the EC_{50} of the recombinant virus to various PIs. Another study examining A431V and I437V, showed that the reversion of the Gag CSM to wild type had a significant effect on PI resistance, but a far more modest effect on RC, indicating a direct effect on the resistance of these mutations independent of replication capacity (RC) compensation[196,400].

Following exposure to PIs both *in vitro* and *in vivo,* mutations at Gag positions 449, 452 and 453 in the p1/p6 cleavage site have been observed. The L449F/V/P was more commonly observed *in vivo* following exposure to SQV, NFV, APV, IDV, ATV and RTV[379,384,394,407,406]. In *in vitro* passages with LPV, APV and some experimental PIs, the L449F have been selected[402,403]. The presence of L449P was negatively associated with virological response in the ANRS 127 trial[407]. There have also been mutations at positions 452 and 453 of Gag *in vivo* following exposure to a number of PIs[379,408]. In addition, changes at these positions have

also been observed *in vitro* following PI exposure, P452K, P453L and P453T, although these have been shown to only convey resistance when found with resistance mutations in protease[409].

In the N terminus of Gag, mutation at position 128 at the MA/CA cleavage site have been reported to affect susceptibility to PIs. Mutation at this position has been associated with both *in vivo* and *in vitro* PI-exposure and is also associated with increased of virological failure of regimens containing darunavir [407,410]. Also, the Y132F mutation at the MA/CA cleavage site has also been associated with resistance to protease inhibitors – LPV and ATV[411].

Gag cleavage site mutations present at baseline have been shown to have an effect on the outcome of subsequent PI-based therapy. In the ANRS trial, mutations at Gag position 128 (near the MA/CA cleavage site) was negatively associated with response to treatment. In the POWER trials, there was the emergence of V128I in more that 10% of the patients who experienced virological rebound[407,412]. Analyses of data from the POWER trials showed a number of other mutations were also associated with reduced response to DRV-based regimens: E428G, S451T and R452S[412]. A surprising observation from this trial was the apparent paradox where the A431V was associated with positive outcomes with DRV-based treatments, despite several studies establishing its role in reduced PI susceptibility *in vitro* as well as its association with poor outcomes in several clinical trials[413,414]. Results from the NARVAL trial have reported I437V as being associated with reduced virological response to several PIs and with virological failure in darunavir-containing regimens[415].

1.14.2   Non-Cleavage Site Mutations in Gag and PI Resistance

Studies have shown that there is a minimal association of Gag cleavage site mutations with primary resistance mutations and that Gag non-cleavage mutations play a role in response to protease inhibitors[416-418]. The most common Gag non-cleavage site mutations characterized (mainly in the MA and CA regions) include T242N/S(198); T427D/N, E46V/K, Q369L[419]; R76K, Y79F, and T81A[417,420]. A longitudinal study involving four patients from Spain who had been

treated with PIs for nine years, the co-evolution of MA and CA non-cleavage site mutations was strongly associated with protease mutations[421] implying the role these mutations may have in impacting PI-based therapy.

Our lab previously used single cycle *in vitro* assays and identified changes within matrix (MA) which directly affects susceptibility to protease inhibitors[420]. Specifically, three amino acids in MA were shown to be responsible for decreased susceptibility these being: R76K, Y79F and T81A[417]. These mutations are found in a predicted alpha helix region of MA, and it is thought that the changes may cause the loss of a hydrogen bond resulting in increased flexibility of the region. It has been suggested that this alteration in the conformation of the Gag protein may enhance the efficiency of cleavage[417]. Other changes in MA have been identified following *in vitro* exposure to APV, E12K and L75R, and an experimental PI, G62R and K112E[410,418,422]. These mutations may affect Gag multimerisation during virion assembly, explaining their involvement in PI resistance.

Following *in vitro* exposure to protease inhibitors, the M200I and H219Q mutations in the capsid (CA) have been shown to emerge[410,418,422]. Capsid mutations have not yet been associated with treatment failure *in vivo*[394]. The binding of CypA to Gag is affected by the H219Q/R mutation this mutation may increase replication efficiency by reducing the requirement of CypA for replication. In a study using CRF01_AE viruses, it was shown that the presence of lysine residue at position 165 of Gag can directly lead to a reduction in susceptibility to protease inhibitors[423]. Although several changes have been reported in other regions of Gag, the role of these changes remain largely unclear. During PI therapy, it has been observed that mutations accumulate at positions 369-371 within the p2 spacer peptide *in vivo*. These mutations have not been shown to be associated with treatment failure, neither have they been shown to emerge following in vitro passage with PIs[424]. Additionally, there have not been any changes in *in vitro* susceptibility as a result of these mutations.

Similarly, although amino acid changes have been observed in the NC domain, the role of these changes remains largely unknown.

The I389T and I401V have been associated with treatment failure although yet again, the mechanism by which they affect treatment outcome is unknown[424]. Also, following *in vitro* passage with PIs, V390A/D, I401T and R409K mutations have been observed to emerge, they have not been shown to confer changes in susceptibility *in vitro*[410,418,422].

1.15 Resistance Mutations and Viral Fitness

Because of the need to achieve maximum affinity, protease inhibitors were designed to occupy more space within the active site than the natural substrate[425]. Most major PI resistance mutations confer resistance to PIs by enlarging the active site of protease. This enlargement reduces the affinity of the drug to the protease enzyme. It has been suggested that when the active site is enlarged, the binding of PIs is affected rather than the substrate to protease binding[358]. Despite this, major PI resistance mutations reduce the rate of Gag and Gag-Pol processing, resulting in a reduction of viral fitness[426]. Usually, major resistance mutations arise first, and with continuous exposure to PIs, secondary mutations begin to accumulate. The occurrence of different PI resistance mutations affects viral fitness to different extents, as a result, measuring the effect of single mutations on fitness is difficult since resistance mutations are often found in combination. Gag mutations could also act as secondary or compensatory mutations and restore viral fitness[395]. One example of compensatory Gag mutations is a study involving the PI darunavir (DRV), the observation of Gag mutations H219Q and I223V appeared to compensate for viral fitness in the presence of primary protease mutations[427].

The p6 domain in the C-terminal of Gag has also been implicated in conferring resistance to antiretroviral drugs in a similar mechanism by restoring viral fitness. An example is a study that examined the association of Gag p6 mutations P5L/T (HXB2 Gag 453) and K27Q/N (HXB2 Gag 475) in enhancing p6 function in

packaging *pol* molecules and reducing the potency of NRTI-based therapy against HIV CRF02_AG and subtypes G[47]. Further evidence of the role of Gag p6 in resistance and restoration of viral fitness is provided by studies that found amino acid motif duplications and insertions in the p6 gene. The most common of these motifs in p6 that provide an escape mechanism for increased incorporation of RT molecules include PTAPP duplications[428,429].Other studies have described a PYxE insertion in Gag p6 of HIV-1 subtype C that enhances viral replication and is associated with PI-based treatment failure[430,431].

When present in clinical isolates or molecular clones, the following major PI resistance mutations have been associated with a significant reduction in viral fitness: D30N, M46I/L, G48V, I50L, I54V, V82A/T, I84V, N88D/S and L90M [432]. A number of studies have examined the exact role of resistance mutations on viral fitness *in vitro*, both independently and when present in combination. For example, major mutations V82T, I84V, M36I and I54V were observed following *in vitro* passage of virus with RTV and a significant reduction in fitness was observed. However, continued passage led to the appearance of A71V and K20R, which increased the fitness of the virus towards that of wild-type[353]. Mutations I84V and V82A independently confer a reduction in fitness to around 80-90% that of wild-type, but the mutation L10I has been shown to compensate for this reduction in fitness when present with either change[433,434]. L10I, alongside L63P, has also been shown to compensate for the reduced fitness conferred by the major mutation L90M[435].

A second mechanism of compensation for reduced fitness is via mutations in Gag; both at the cleavage sites and at other sites distant from the cleavage sites. Cleavage site mutations found in concert with protease mutations are thought to alter the tertiary structure of the Gag to make it fit more tightly within the active site of mutant protease, thus compensating for the reduced RC caused by mutant protease. The protease resistance mutation V82A creates a gap between the protease active site and the PI, simultaneously reducing the affinity of Gag to protease and to the PI. A413V in Gag has been shown to create a protrusion into

the substrate binding pocket, compensating at least partially for the reduction in affinity to protease[176]. In addition, L449F is associated with increased fitness when present with major PI resistance mutations[379].

Table 1.3: Gag mutations which have previously been associated with PI resistance or exposure

Gag Mutations	Location within Gag	References
E12K	MA	418,422
G62R	MA	436
L75R	MA	418,422
R76K	MA	417
Y79F	MA	417
T81A	MA	417
K95R	MA	437
K112E	MA	410
G123E	MA	422
V128A/I/T/del	MA	407,410,412,438
165	CA	423,439
M200I	CA	410
H219Q	CA	418,422
R286K	CA	437
A360V	CA	406
V362I	CA	391,440
L363M/F/C/N/Y	p2	436
S368C/N	p2	404
Q369H	p2	441
T371del	p2	441
S373P	p2	404,414,441
A374P/S	p2	414
T375N/S	p2	406,414
I376V	p2	404,406
R380K	NC	424
G381S	NC	406
I389T	NC	424

V390D/A	NC	418,422
I401T/V	NC	424
R409K	NC	410,418,422,436
K415R	NC	442
E428G	NC	412
Q430R	NC	391
A431V	NC	394,405,406,443
K436E/R	p1	358,400,443,444
I437V/T	p1	196,358,394,405,406
L449F/P	p6	379,384,394,403,424,443
S451N/T/G/R	p6	412,442,443
P452S/K	p6	409,412,443,444
T456S	p6	445
E468K	p6	418,422
Q474L	p6	424
R479	p6	446
A487S	p6	424
P497L	p6	424

1.16 Association between protease and Gag mutations

In addition to the emergence of mutations in the protease gene, exposure to protease inhibitors also leads to the emergence of mutations in Gag. This phenomenon provides a compelling evidence as to the co-evolution of Gag-protease in response to selective pressure. Gag mutations have been shown to emerge to compensate for reduced fitness conferred by mutations in protease. Strong associations between specific mutations in Gag and protease that develop together following treatment has been reported in several studies, as reviewed by Clavel and Mammano[395]. The selection of mutations at specific positions in Gag-protease suggests that these coevolving mutational changes occur to

maintain structural integrity during Gag cleavage[447]. It is yet to be fully elucidated how this co-evolution confers fitness advantages to the virus.

The A431V and I437V mutations in Gag are associated with V82A in protease. Also, the A431V mutation is associated with M46I and changes at position 54 of protease[408,444]. Mutations at Gag positions L449 and P453 have been linked to the I50V resistance mutation in protease and shown to confer reduced susceptibility to APV and improve viral fitness *in vitro*[379]. The pre-existence of Gag L449F mutation has been suggested to favour the emergence of I84V in the protease gene. Additionally, the Gag L449F mutation is associated with I84V and D30N mutations in the protease gene. It has also been demonstrated that P453L Gag mutation emerges with the I84V mutation in protease[408]. Additionally, the P453L Gag mutation has been associated with the N88D and in protease mutations. The presence of P453L has been hypothesised to restore the hydrogen bond that occurs between D30 of protease and position 452 of Gag that is disrupted by the D30N mutation[448].

Specific interactions with I50V/A71V protease are observed to be lost or formed in response to coevolution mutations in the p1-p6 substrate cleavage site. Particularly, mutations in the p1-p6 cleavage site are statistically associated with I50V protease mutation in the viral sequences retrieved from patients[449]. In previous studies, utilizing co-variation analyses, mutations at Gag 437 in combination with D30N/N88D protease mutations resulted in viruses with reduced sensitivity to APV, ATV, IDV, LPV, NFV, RTV, and SQV. Gag R452 was also observed to mutate in an associated manner with D30N/N88D. In addition, mutations within p1-p6 are also associated with the N88D and N88D/L90M protease mutations and decreased sensitivity to APV, IDV, NFV, SQV, and TPV was observed in viruses with a Gag 449 mutation in combination with N88D[443].

1.17   Transmission and persistence of PI resistance

When a patient on therapy develops drug resistance (known as acquired drug resistance, ADR), the resistant viruses could be transmitted to a new individual,

(a phenomenon known as transmitted drug resistance (TDR)). TDR can either be horizontal transmission via heterosexual/ homosexual intercourse or be vertical via mother to child transmission (MTCT). It may be hard to detect owing to the sensitivity limit of standard genotypic assays and the possibility of reversion of drug resistance virus to wild- type forms [450].

In a study by Chin-Hong and colleagues in the SCOPE study, it was reported that patients with acquired drug resistance (ADR) continued to engage in high-risk sexual behaviour. In the study, up to 27% reported engaging in unprotected sexual intercourse with a partner of either negative or unknown HIV status [451].

In countries with widespread use of ART, the acquisition of resistant viruses at new diagnoses is an important consideration. When resistant viruses are transmitted to a new individual, the treatment options available to this newly infected person become further limited. In addition, infection with resistant virus have a negative impact on response to treatment and time to first virological failure is reduced significantly[452]. Rates of transmitted drug resistance (TDR) are normally monitored by looking at the prevalence of drug resistance in treatment-naïve, recently infected individuals.

Generally, resistant viruses have been shown to have lower replicative capacities than the wild type viruses. Some earlier assumptions were that in the absence of selective drug pressure in the newly infected person, transmitted resistant viruses would revert to wildtype. The disappearance of PI DRMs by reversion to wildtype over time has been reported in a couple of studies[311]. Conversely, resistant viruses have been shown to persist over time, despite the absence of drug pressure and the reduced replicative capacity [453]. In the UK, and other resource-rich settings, genotypic resistance test is undertaken for all newly diagnosed persons. The results of these tests are used as a determinant for the drug regimen to be administered to these new patients. In low-and-middle-income countries, however, genotypic resistance testing is not routinely available prior to commencement of treatment.

As well as the role for the persistence of drug resistance in transmission, it is also an essential consideration in the context of treatment interruptions. If a patient

develops resistance to PIs during therapy, then the persistence or reversion of the resistance mutations once PI treatment has been stopped has important implications for future treatment options. If resistance mutations persist over time, then PI therapy cannot be re-introduced successfully; however, if the mutations disappear and revert to wild type, it may be possible to use the same PI successfully[454]. It has also been suggested that due to compensatory fixation, protease inhibitor drug resistance mutations may not revert to wildtype in the new host. The concept of compensatory fixation argues that replicative capacity is lost when one PI DRM is reverted to wild type, this effectively blocks the route of evolution back to wild type[455].

1.18 Methods for testing protease inhibitor resistance

1.18.1 HIV-1 Genotypic Resistance Testing Methods

In clinical practice, the presence or absence of known mutations within the viral sequence can be used to predict the susceptibility or resistance of a virus. To predict this, patient-derived viruses are PCR-amplified (usually the *pol* gene), followed by DNA sequencing and analyses of sequence data. This can either be done using commercial test kits, or alternatively, using in-house assays. There are several online genotypic interpretation tools and algorithms which can be used to translate specific mutation patterns into predicted susceptibility levels. The most common of these online algorithms which have found widespread use in clinical practice are; Stanford HIVdb, ANRS and Rega.

1.18.2 Phenotypic Resistance Assays

In phenotypic assays, *in vitro* susceptibility of a virus or enzyme is measured directly and not based on a prediction. Phenotypic assays could be enzyme-susceptibility-based or replication-based. In both replication and enzyme-based assays, the concentration of drug required to inhibit 50% (EC_{50}) or 90% (EC_{90}) of replication or enzyme activity is determined. The value obtained is compared to that of a designated wildtype virus.

It is technically challenging to isolate and use live patient viruses. This challenge is overcome by the use of recombinant viruses, where the performance of the assay is not affected by factors such as virus tropism. In the commercial Phenosense assay, resistance test vectors are constructed with part of the *pol* gene (N-terminal region) that contains known PI and RTI resistance mutations derived from patient virus. The HIV env gene is however substituted with a luciferase reporter gene. The vectors are co-transfected into cells with a plasmid encoding an MLV-env, resulting in the production of pseudovirions[456]. The infectivity of these pseudoviruses (which are only capable of a single-round of replication) in the presence of drug is tested. The amount of luciferase activity is a measure of the infectivity of the virus[457].

There are two methods through which the mosaic virus (which contains patient-derived sequences) can be phenotypically tested for susceptibility to drugs, as described by Garcia-Diaz, 2012[458].

(i)   Single-cycle phenotypic assay: this method is based on a single round of infection. Here, replication-defective retroviral vectors (RTVs) are created. After PCR amplification, patient-derived gene sequences of interest are inserted into an HIV expression vector lacking the region of interest from the patient. Single cycle phenotypic assays can either be utilized in a 2-plasmid or a 3-plasmid transfection assay system.

Petropolous and colleagues developed a novel phenotypic drug susceptibility assay which employs a 2-plasmid transfection system[457]. In this assay, amplified protease and RT gene segments from patient plasma samples are inserted into an indicator gene viral vector by using suitable restriction sites. To monitor virus replication, a luciferase indicator gene cassette is inserted into a deleted region of the *env* gene, preventing HIV-1 envelope protein expression. Pseudotyped virus particles are produced by co-transfecting cells with RTV DNA and a plasmid that expresses the envelope proteins of amphotropic murine leukemia virus (MLV). Following transfection, virus particles are harvested and are used to infect fresh target cells. The ability of virus particles to complete a single round of replication is assessed by measuring luciferase activity in target cells.

The three-plasmid system employs; (a) A vector which expresses vesicular stomatitis G protein (VSV-g) which provides the envelope; (b) a reporter vector (such as pCSFLW or pCSGW) which expresses luciferase or GFP, as well as the HIV packaging sequence and (c) the resistance test vector which contains the patient-derived sequences. These plasmids are used to transfect appropriate cell lines, and the pseudoviruses produced are exposed to increasing concentrations of drug. Luciferase production is quantified and is indicative of the amount of pseudotyped virus produced. Drugs that inhibit virus replication reduces luciferase activity in a dose-dependent manner, providing a quantitative measure of drug susceptibility.

Single-cycle phenotypic assays offer several advantages. The assays are highly efficient and reproducible and do not rely on the relatively inefficient and random process of homologous recombination to incorporate patient-derived protease and RT amplification products into HIV-1 vectors containing protease and RT deletions. Also, transfection of HEK-293 cells with a standard amount of retroviral DNA is a rapid method of producing consistent high-titre virus stocks and does not require virus outgrowth from patient samples or recombinant virus outgrowth following homologous recombination

Furthermore, since retroviral vectors are restricted to a single round of replication, a single cycle system is advantageous when studying the effect of a specific mutation, as the format limits the opportunity for "in vitro" selection of genetically diverse virus subpopulations, which may not accurately reflect the effect of the mutation of interest.

Lastly, the single-replication-cycle assay and the large dynamic range of the luciferase reaction also eliminate the need to determine virus titres prior to the infection of target cells. In addition to reducing the turnaround time, configuration of the assay in a single-cycle format with HIV-1 vectors containing an indicator gene (RTVs) provides increased sensitivity and reproducibility over those of existing assays that require multiple replication cycles and that are performed with indicator cell lines.

In terms of laboratory infrastructure, a BSL-3 is not required for the single replication assays as the assays can be carried out safely in a BSL-2 lab.

The single cycle assays do however have some disadvantages. These assays cannot fully mimic the conditions in a human host as infected cell types and microenvironments vary and can have different selective constraints[459]. Also, variations in the late stages of the virus life cycle (such as cell-to-cell spread) are not encompassed in a single cycle assay hence constituting a disadvantage as significant differences in viral fitness and susceptibility has been reported when single- and multicycle assays were compared[460]. Single-cycle assays only include the stages of the viral life cycle up to the transcription of viral genes, meaning that the efficiency of viral assembly is not represented.

(ii)   Multiple cycle phenotypic assay: This method involves the use of replication-competent viruses which undergo multiple cycles (rounds) of infection. In this assay, the HIV genomic region of interest (from the patient) is inserted into an HIV-1 molecular clone (usually HXB2 or pNL4-3) which does not possess that region. Recombinant viral particles which are capable of multiple rounds of replication are hence produced. The virus is grown in the presence of increasing concentrations of drugs.

The reporter gene, encoding yellow fluorescence protein (YFP) is inserted in the virus genome and used to infect highly permissive MT-4 T-cell lines and infectivity determined by YFP fluorescence. Alternatively, viruses are produced in producer cell lines and used to infect Rev-CEM reporter T cells as targets. These cells express green fluorescent protein (GFP) in the presence of HIV early proteins Tat and Rev. The proportion of GFP-positive Rev-CEM cells after infection with the virus is proportional to the infectivity of the virus[214].

The multiple cycle assay has the advantage that it more closely mimics the conditions that the virus experiences "in vivo". It is however more laborious and requires biosafety level 3 laboratory infrastructure.

1.18.3   Viral Fitness Assays

Viral fitness can be assessed using either *in vivo* or *in vitro* methods (as reviewed by Garcia, 2012)[458]. In vivo methods mimic the natural setting in the natural host, viral fitness is assessed by comparing the amount of wildtype and mutant viruses.

These methods examine the viral kinetics in the host environment such as in blood plasma. The existence of different microenvironments and different cell types within the host means there is a variation in selective pressure. This makes it difficult to extrapolate the findings within one microenvironment (such as blood plasma) with another microenvironment (such as CNS).

In vitro methods employ HIV-1 isolates or more frequently recombinant viruses and can be very useful as models for determining the effect of drug-resistant variants on replication in a fixed environment. These methods can be broadly classified into two; growth kinetics (also called mono-infection) assays and growth competition assays. In growth kinetics assays, the replicative capacity of different HIV isolates, or recombinant viruses are individually tested. The production of specific viral proteins or viral enzymes are measured. Replicative capacity of viruses can also be measured in mono-infection assays by using a reporter genes such as GFP or luciferase via single cycle infectivity assays [461]. In this method, luciferase production of the mutant virus is compared with that of the wildtype reference virus. One disadvantage of single cycle assays is that not all stages of the viral life cycle are included. Here, the assay only includes the life cycle up to transcription of viral genes. This means it does not measure the efficiency of subsequent life-cycle stages such as assembly, budding/release and maturation.

In growth competition experiments, two phenotypically distinguishable viruses are mixed at similar or different proportions, and the outgrowth of one of the populations is measured [462,463]. By doing this, the fitness of both viral strains can be directly compared as two viral populations in cell culture compete with each other until one outgrows the other one. In general, cells are infected with the mixture of viruses and after several passages, the proportion of both viruses is measured and compared with their proportions in the initial mixture[462,464].
Overall, growth competition experiments are more accurate and sensitive for the determination of small differences in fitness than mono-infection assays. The single-cycle assay offers a fast and reproducible method to measure the

replicative capacity of the mutant virus that can be compared and expressed as a percentage of that observed for a WT reference strain. As it is a mono-infection assay, it cannot accurately determine small differences in replicative capacity. However, it can be of use for the characterization of novel mutations since growth competition experiments increase the potential for mutations to occur in the different passages and divert form the population of interest (as reviewed by Garcia, 2012)[458].

1.19 WHO recommendations for Antiretroviral Treatment (ART) in Low- and- Middle-income Countries

Antiretroviral therapy (ART) has substantially decreased HIV morbidity and mortality in high-income as well as low- and middle-income countries (LMICs). Several randomized trials have demonstrated benefits from starting ART regardless of CD4 count[239,465,466]. The World Health Organization (WHO) adopted a "treat all" strategy in 2015. Significant attention has been focused on rapidly initiating ART, reflected in the 2017 WHO guidelines, which recommend that ART be initiated within 7 days of HIV diagnosis and on the same day whenever possible[467-469].

To counter the emergence of drug resistance, patients are typically treated with cocktails of three drugs simultaneously, with at least two independent mechanisms of action[470]. Such regimens have proved highly effective, extending life expectancy and slowing spread of the AIDS epidemic where treatment is available. Patients on combination antiretroviral therapy (cART) often show little or no clinical evidence of viral infection; however, the treatment is not a cure[471].

Until mid-2018, the preferred initial ART recommended in the WHO guidelines for HIV-1 infection consisted of two nucleoside reverse-transcriptase inhibitors (NRTIs) combined with a non-nucleoside reverse-transcriptase inhibitor (NNRTI), namely efavirenz at a dose of 600 mg daily (known as EFV600)[472]. A first-line regimen of tenofovir disoproxil fumarate (TDF) with either lamivudine (3TC) or

emtricitabine (FTC) plus efavirenz (EFV) for HIV-1 treatment was recommended because it can be safely used in pregnancy and during tuberculosis treatment[473].

After failure of a first line NRTI-based regimen, a boosted PI plus 2 NRTIs is the strategy preferred for second-line ART. As the first ritonavir-boosted PI, LPV/r has been widely used as a standard third drug in second-line regimens[474]. Boosted PI options are recommended as part of second-line regimens because of their safety and efficacy as indicated by systematic reviews and meta-analyses[475,476]. Additionally, PIs have high genetic barrier to development of drug resistance and are widely suggested as antiretroviral regimens for treatment-experienced patients with documented resistance to (N)NRTIs[477].

At the commencement of this project, the ART guidelines for initiating ART in treatment naive adults in Nigeria was TDF+3TC (or FTC) +EFV as fixed dose combinations. When an individual fails first-line regimen, recommended second-line was boosted-PI and an NRTI. If TDF was used in first line therapy, then either AZT+3TC+LPV/r or AZT + 3TC+ATV/r was administered. If AZT was used in first line therapy, then either TDF+3TC+ATV/r or TDF+3TC+LPV was used[478]. These recommendations have now been replaced by the 2019 guidelines. The major change in the 2019 guidelines is the introduction of the integrase strand transfer inhibitors (INSTI), dolutegravir (DTG) as the preferred first line regime. In these guidelines, recommended first line regimen is: TDF+3TC (or FTC) +DTG or TDF+3TC+EFV (400 mg). After failure to first line treatment, preferred second-line regimens consist of AZT+3TC+ATV/r (or LPV/r) or AZT+3TC+DRV/r[479].

1.20   Drug resistance in non-B HIV-1 subtypes

Generally, antiretroviral drugs are as effective in non-subtype B, group M HIV-1 viruses as they are in the subtype-B viruses[480,481]. Mutations that cause resistance in subtype B viruses also cause resistance in each of the other subtypes. However, subtypes vary in their propensities to develop specific mutations as a result of three potential factors: (i) inter-subtype differences in codon usage; (ii) inter-subtype amino acid differences that result in subtle

structural differences in the targets of therapy; and (iii) inter-subtype differences in the sequence context surrounding a nucleotide at which a substitution results in a drug resistance mutation[309].

The propensity for subtype C viruses to develop V106M during NNRTI treatment – rather than V106A, which is more commonly observed in subtype B viruses – results from the fact that V106 is encoded by GTA in subtype B viruses and GTG in subtype C viruses. A single G-to-A transition (the most common reverse transcriptase error) at the first position of codon 106 in subtype C viruses results in V106M, which confers high-level efavirenz and nevirapine resistance. In contrast, in subtype B viruses, V106M requires two nucleotide substitutions (GTA-ATG) and therefore occurs uncommonly[482,483]. A similar phenomenon has been observed at protease codon 82 in subtype G viruses, which are much more likely than subtype B viruses to develop the poorly characterized mutation V82M[484].

Differences in the amino acids between subtypes can create subtle differences in the structural micro-environment that may predispose HIV-1 to different mutations under similar selective pressure. For example, subtype B-infected patients receiving nelfinavir are more likely to develop D30N than are those with viruses belonging to subtypes C, F, G and CRF01_AE, which are more likely to develop L90M or N88S (as reviewed by Tang & Shafer, 2012)[309]. In vitro studies have suggested that inter-subtype differences in the consensus amino acid at position 89 with leucine (L) in subtype B and methionine (M) in most other subtypes are responsible for the different patterns of observed resistance[485].

Several lines of evidence suggest that patients infected with subtype C viruses who are treated with stavudine/lamivudine/efavirenz or stavudine/lamivudine/nevirapine may be more prone than patients infected with subtype B viruses to develop the NRTI-resistance mutation, K65R (as reviewed by Tang & Shafer, 2012)[309]. Elegant biochemical research suggests that the unique sequence context in the region of K65R, specifically – a span of five consecutive adenosines preceding the adenosine at the second position in the

K65 codon – renders it more likely to be mutated during reverse transcription[486,487].

The majority of non-B HIV-1 subtype isolates possess wild-type susceptibilities similar to those of subtype B wild-type isolates. Compared to B subtypes, diminished susceptibilities among wild-type isolates have been found for CRF02_AG recombinant viruses in different studies in regard to ATV and NFV [488]. Previous study from our lab also found a decreased susceptibility to PIs by CRF02_AG and subtype G viruses in comparison to B subtypes[489].

1.21   Project overview

Despite the incredible success of AIDS treatment during the last years with current ART therapies, resistance mutations, and the accumulation of severe side effects is an enormous challenge that continually needs to be addressed[490]. Most HIV-1-infected individuals with virological failure on a boosted protease inhibitor (PI) regimen do not develop PI-resistance protease mutations[491].

There is currently a huge amount of evidence to show that Gag can directly confer PI resistance in the absence of known resistance mutations in protease [196,197,417,420]. In addition, studies from our lab have shown that the inclusion of co-evolved Gag alongside protease in phenotypic assays can affect the susceptibility of the virus [397,420]. Our lab has also reported an association between reduced PI susceptibility at baseline in the absence of known resistance mutations and subsequent virological failure on lopinavir/ritonavir monotherapy in patients harbouring subtype CRF02_AG/G viruses[489]. It, however, remains unknown if baseline susceptibility can be used to predict virological outcomes on second-line ART in real-world settings.

In the first part of this thesis, we set out to test the hypothesis that baseline susceptibility, as determined by a full-length cognate Gag – protease assay, is correlated with eventual second-line treatment outcome. This was achieved through the use of single-round infection phenotypic drug assays, utilizing patient samples from Nigeria, West Africa. Baseline samples from patients who achieved virological suppression were matched with baseline samples from patients who eventually failed second-line treatment in a case-control manner. The aim was to

use genotypic and phenotypic data to establish if we can predict virological outcomes on PI-based regimens.

In the second part, we investigated the role of amino acid mutations in the matrix domain of Gag in conferring reduced susceptibility to protease inhibitors. We employed a series of site-directed mutagenesis and single-round infection assays for phenotypic drug susceptibility to study the effects of different mutations. We also used full-length, replication-competent viruses in multiple round of infection assays in order to elucidate the mechanism of reduced susceptibility due to amino acid changes occurring in HIV-1 matrix.

CHAPTER 2:   Materials and Methods

2.1   General Molecular Biology Techniques

2.1.1   RNA extraction

HIV-1 viral RNA was manually extracted from plasma samples using Qiagen QIAamp Viral RNA Kit (Qiagen). The method is based on the nucleic acid extraction protocol developed by Boom and colleagues [492] which employed the lysing and nuclease-inactivating properties of the chaotropic agent guanidinium thiocyanate together with the nucleic acid-binding properties of silica particles in the presence of this agent. Up to one millilitre of plasma was centrifugated (20,000 x g) for 1 hour at 4°C to concentrate the virus. The supernatant was then removed, and the pellet re-suspended to a final volume of 140μl. The re-suspended pellet was employed for nucleic acid extraction using the Qiagen QIAamp Viral RNA Kit. Nucleic acid extraction was performed following the manufacturer's instructions.

2.1.2   cDNA synthesis

10 μl of extracted RNA was denatured for 10 minutes at 80°C prior to the cDNA synthesis reaction. cDNA synthesis was carried out using the antisense primers ProOutR or KVL065 [493] RNAseOut (Invitrogen) and SuperScript III enzyme (Invitrogen) in a 20 μl reaction at 53° for 60 minutes, followed by an RT inactivation step at 70°C for 15 minutes. Resultant cDNA was either used immediately for nested PCR (as detailed below), or stored at -80°C.

2.1.3   Gag-protease Amplification PCR

PCR was performed using the Phusion High Fidelity PCR system (New England Biolabs). Master mix was prepared containing 5x HF Phusion buffer, 200μM dNTPs, 0.5μM of forward and reverse primers (BKTO3 and KVL065) and 2.5 Units of Platinum Taq HF Taq DNA polymerase. 1μl of cDNA was added to 24μl of master mix in a PCR tube, transferred to a thermocycler and cycled as below:

Temp (°C)	Time	Cycles
98	30 sec	x1
98	10 sec	
53	30 sec	
72	90 sec	x30
72	10 mins	x1
4	Hold	∞

2.1.4 Second Round (Nested PCR)

Master mix was prepared as above. GagNot+ and ProXhoR2 primers were used as sense and anti-sense primers respectively. In PCR tubes, 24μl of the master mix was added to 1μl of 1st round PCR product. PCR was carried out under the following cycling conditions:

Temp (°C)	Time	Cycles
98	30 sec	x1
98	10 sec	
50*	30 sec	
72	90 sec	x30
72	10 mins	x1
4	hold	∞

* The annealing temperature was reduced to 50°C because the GagNot+ and ProXhoR2 primers contain mismatched bases. The NotI and XhoI restriction sites were added using these primers. DNA primers used for PCR and sequencing are listed in table 2.1 below

Table 2.1: Primers used for cDNA synthesis, nested PCR and sequence analysis

Primer	Sequence (5' – 3')	Application
ProOutR	TTGGGCCATCCATTCCTGG	cDNA synthesis
KVL065[1]	TCCTAATTGAACYTCCCARAARTCYTGAGTTC	1st round PCR
BKTO3	CGCAGGACTCGGCTTGC	1st round PCR
GagNot+	GCG***GCGGCCGC***AAGGAGAGAGATGGGTGCG	Nested PCR/ Sequencing
ProXhoR2	CTGGTACAGT***CTCGAG***RGGACTRATKGG	Nested PCR
Gag1F	CATTATCAGAAGGAGCCACC	Sequencing
Gag1.5R	TCTATCCCATTCTGCAGC	Sequencing
Gag2F	ATGATGACAGCATGTCAGGG	Sequencing

The areas of primers designed with mismatches to introduce restriction enzyme sites are underlined

2.1.5 Agarose gel electrophoresis

Agarose gel electrophoresis was used to confirm the correct size of the PCR product. To make 1% agarose, 1.0 g of agarose was dissolved in 100 ml TAE buffer, mixed and dissolved in a microwave. Once cooled, 10mg/ml ethidium bromide was added. PCR products were mixed with 5x loading dye and loaded onto the gel with a DNA mass ladder. Gels were run for approximately 45 minutes at 100 volts, depending on the size of the band expected. Gel bands were visualised and photographed using a UV transilluminator (UVP).

2.1.6 Purification of PCR products

PCR products were visualised using ethidium bromide staining and agarose gel electrophoresis. Products of the expected size were either excised and purified using the QIAQuick Gel Extraction Kit (Qiagen) or directly purified from the PCR mix using the QIAQuick PCR Purification kit (Qiagen) if the PCR product was going to be subsequently sequenced and cloned. Both purification methods were performed in accordance with the manufacturer's instructions.

Briefly, the PCR product was added to the buffer containing the chaotropic agent guanidine thiocyanate and bound to the silica membrane in the QIAquick spin column through centrifugation for 1 minute at 13,000rpm. Impurities and contaminants were removed through washing with an ethanol containing buffer and DNA was eluted into a low salt and pH containing buffer. If gel excision was employed, the gel slice was previously dissolved in 3 volumes of buffer QG and subsequently the DNA purified as indicated above for a PCR product.

2.1.7 A-tailing

Where cloning into the intermediate vector (pGEM) was performed (as detailed in 2.1.8.1), the addition of 3' A-overhangs to the PCR product from the NEB Phusion High Fidelity PCR system was required. PCR product from the Phusion High Fidelity reaction was purified using the QIAQuick Gel Extraction Kit (Qiagen) as detailed in 2.1.6 above and incubated with *Taq* polymerase and 0.1 μM dNTPs, to add 3' A overhangs. Cycling conditions were: 3 cycles of 53°C for 30 seconds and 72°C for 4 minutes, followed by 72° for 7 minutes. The product was then purified (see 2.1.6) and is ready for cloning into pGEM intermediate vector.

2.1.8 Cloning

2.1.8.1 Cloning using pGEM vector

Cloning of Gag-protease from clinical samples was carried out using an intermediate vector – pGEM-T Easy (Promega) – to enable clonal sequencing. After A-tailing and purification, PCR products were then TA cloned into the pGEM-T vector.
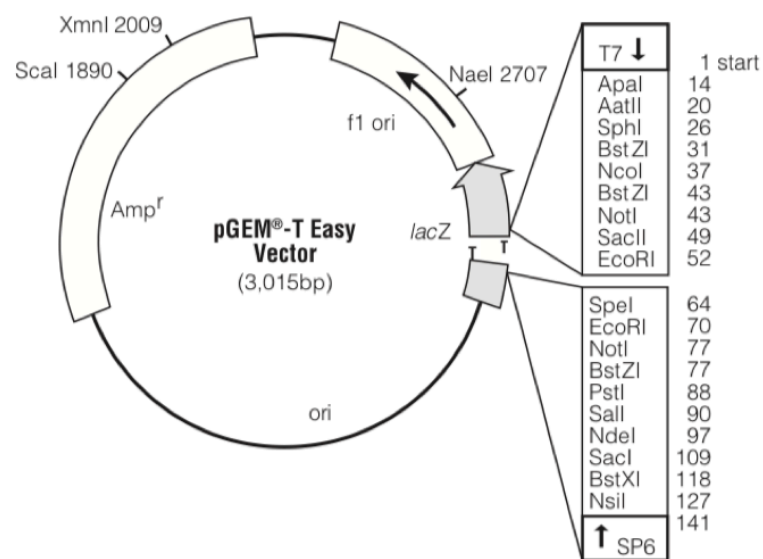
Figure 2.1: pGEM®-T Easy Vector Map and Sequence Reference Points
Seen within the lacZ gene are the T-overhangs and a number of restriction sites.

2.1.8.2   **Ligation**

Ligation of PCR products into pGEM vector proceeded at 4°C overnight in 10 μl total volume as shown below:

	Sample reaction (μl)	Vector only control (μl)
Mastermix:		
2 x Ligation Buffer	5	5
pGEM	1	1
T4 Ligase	1	1
Total mastermix	7	7
Samples and control:		
PCR product	3	-
H_2O	-	3
Total	10	10

2.1.8.3   **Transformation of bacteria**

2 μl of ligation product was incubated with 50 μl of JM109 competent cells on ice for 30 minutes. Cells were heat shocked at 42°C for 45 seconds, followed by incubation for 2 minutes on ice. Cells were incubated, shaking, at 37°C for 1 hour with 450 μl Super Optimal Broth (SOC). Varying volumes of cells were plated out onto 10cm agar plates with 100 µg/ml ampicillin and incubated overnight at 37°C. Colour screening plates were prepared with the addition of Blue-Gal and IPTG (Invitrogen).

2.1.8.4   Colony PCR

To determine which bacteria colonies were positive for *Gag-protease*, colony PCR was carried out. Master mix was prepared as below:

	Volume (uL)
Buffer	4.0
dNTPs	0.5
GagNot+	1.0
Gag1.5R	1.0
Taq polymerase	0.25
Nuclease free Water	13.25
Total	20.0

A pipette tip was used to pick up single bacterial colony, which was the swivelled in the master mix.

Cycling condition was as follows:

Temp (°C)	Time	Cycles
94	2 mins	x1
94	30 seconds	x20
53	30 seconds	
72	45 seconds	
72	2 mins	x1
4	Hold	∞

Colony PCR products were run on 0.8% agarose gel to select bacterial colonies with the right band size of 1.8kb (having incorporated the patient *Gag-pro*. Following selection as necessary, colonies were seeded into 4 ml of LB with 100 µg/ml ampicillin and incubated, shaking at 37°C overnight. Where required, glycerol stocks of transformed bacteria were created by mixing 1 ml of bacterial culture with 1ml of 50% glycerol solution (Sigma) in cryotubes and stored at -80°C.

2.1.8.5 Minipreps

Double stranded plasmid DNA was extracted from bacterial cells using Qiagen Miniprep Kit (Qiagen), as per manufacturer's protocol. Briefly, a single transformed E. coli colony was inoculated into 4 ml of LB broth containing 50 mg/ml ampicillin and incubated overnight at 37°C in an orbital shaker.

The overnight bacterial culture was spun at 6,800 x g for 3 minutes at room temperature. The supernatant was discarded. Bacterial cell pellets were resuspended and lysed under alkaline conditions. The lysate was then neutralised with acetic acid and bound to the silica membrane of the QIAprep spin column through centrifugation for 1 minute at 17,900 x g. Remaining impurities were washed away using an ethanol-based buffer. The plasmid DNA was then eluted under low salt conditions into RNase-free water or elution buffer (10 mM Tris-Cl, pH 8.5).

2.1.8.6 DNA quantification

The quality and quantity of plasmid DNA extracted was assessed by UV-Vis-Spectophotometry using 1 μl of mini-prep and a Nanodrop 3100 Spectrophotometer (NanoDrop). DNA was measured at a wavelength of 260nm.

2.1.8.7 DNA sequencing

DNA sequencing was carried out in-house by an external company (Genewiz, UK) using sanger methodology. Primers employed for *Gag-pro* sequencing are shown in Table 1.0. Primers were employed in different combinations depending on the specific sample. Briefly, purified PCR products or plasmids identified as containing the PCR insert were diluted to a concentration of around 20 ng/μl and sequenced using the BigDye Sequencing mix v3.1. Sequencing reaction contained 8 μl of PCR product or plasmid, 0.5 μM of the selected primer and nuclease-free water to a final volume of 20 μl. Sequencing PCR conditions were as follows, 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes and a hold at 4°C. Sequencing reactions were purified by precipitating the DNA with 52 μl of a mix containing 50 μl of 100% ethanol (EtOH) and 2 μl of 3M sodium acetate followed by a washing step with 150 μl of 70% EtOH. Purified

sequencing reactions were then run on a 3730-Avant Genetic Analyzer (Applied Biosystems, UK).

2.1.8.8 Sequence analysis

Sequencing data were aligned and analysed using DNA dynamo software (BlueTractor Software). Sequences were trimmed and manually checked for any discrepancies between overlapping primers. This was done to check if pGEM cloning was successful (i.e full length Gag-pro from the patient has been successfully cloned into pGEM).

2.1.8.9 Transfer from pGEM to the HIV expression vector P8.9NSX+

Patient's *Gag-pro* sequences were transferred from pGEM vector to the HIV expressing vector (P8.9NSX+) by a restriction digest (using 2 enzymes NotI and XhoI) to cut Gag-pro out of pGEM for ligation into p8.9NSX+).

Buffer	2ul
NotI	0.5ul
XhoI	0.5ul
Plasmid (1ug)	~4uL
H_2O	To 20uL

This was followed by a 37°C incubation for 2 hours, then enzyme deactivation for at 65°C for 5 minutes. Products were resolved on 1% agarose gel for about 1 hour. DNA fragments were excised from the agarose gel and purified as described in section 2.1.6 above.

2.1.8.10 Creating p8.9NSX+ ΔGagPro and ligation of 'empty' vector with patient-derived Gag-Pro fragments

Gag-Pro fragment of the P8.9NSX+ vector was deleted through restriction digest thereby creating an 'empty' vector. 'Empty' p8.9NSX+ vector, was prepared by restriction digest as below for 2 hours.

Buffer 3.1	4 μl
NotI	1.5 μl
XhoI	1.5 μl
p 8.9 = 3 ug	μl
H$_2$O to make up to 40	μl
Total	40 μl

This was followed by dephosphorylation of the vector at 37°C for 30 minutes.

Buffer	5 μl
Antarctic phosphatase	2 μl
Empty vector	40 μl
H$_2$O	3 μl
Total	50 μl

The vector was subsequently resolved on a 0.8% gel for 2 hours, followed by gel-extraction of the DNA as previously described. The eluate now contains the 'empty' vector (p8.9NSXdelGagPro) into which patient-derived Gag-Pro can be cloned in.

The 'insert' DNA was ligated into the 'empty' vector in a 5:1 ratio in a 20μl reaction consisting of 2μl buffer, X μl of 'empty' vector, Y μl insert (as calculated), 1μL enzyme and Z uL H$_2$0 and incubated at room temperature for 1 hour. This was followed by transformation into HB101 cells, and plasmid DNA extraction (miniprep).

2.1.8.11 Triple Restriction Digest

To confirm successful cloning, a triple restriction digest was carried out using plasmid DNA after miniprep. 2μl of each plasmid was added to a tube containing appropriate volumes/units of restriction enzymes and using p8.9NSX+ as negative control and a p8.9GagPro as positive control. Master mix volumes are as shown below:

Buffer	2.0μl
NotI	0.5 μl
PvuI	0.5 μl
EcoRI	0.5 μl
H$_2$0	14.5 μl

This was incubated at 37°C for 1 hour then visualized on a gel. Positive plasmids should be 2 bands only because PvuI restriction site is not present in insert DNA but present in vector. p8.9NSX and thus should generate 3 bands.

2.1.9   Sequence analysis

Plasmid DNA samples were sent to Genewiz laboratories for sequencing (as previously described).

Sequencing data were aligned and analysed using DNAdynamo software (BlueTractor Software) as previously described in section 2.1.8.8.

Alignment of multiple sequences for further analyses, for example clonal sequences within a patient, was carried out using Molecular Evolutionary Genetic Analysis (MEGA) v7.0 software (Tamura et al., 2011). Virus subtype was determined using the REGA HIV-1 Subtyping tool provided by Stanford University (http://dbpartners.stanford.edu/RegaSubtyping/)[494]. Analyses for PI resistance mutations within protease was carried out using the Stanford HIV Drug Resistance Database (http://hivdb.stanford.edu/) which reports major and minor resistance mutations and other polymorphisms[495] (Liu and Shafer, 2006). Reference to the IAS-USA list of drug resistance mutations was also made[496]. Gag changes were identified manually by reference to mutations in Gag previously associated with PI resistance or exposure in different published work.

2.1.10   Site-directed mutagenesis

Site directed mutagenesis was carried out using Quickchange Lightning Site Directed Mutagenesis Kit (Agilent) as per manufacturer's protocol to insert

desired mutations. All primer combinations were designed specifically to incorporate the desired mutation.

Reactions included 25 ng of DNA template and 125 ng of each primer – sense and antisense. Typical cycling conditions were as follows:

1	(x1)	95ºC	2 minutes
2	(x18)	95°C	20 seconds
		60°C	10 seconds
		68°C	6 minutes 15 seconds
			(30 seconds per kb of template DNA)
3	(x1)	68°C	5 minutes

In order to degrade the parental DNA plasmid, the PCR product was incubated at 37ºC for 5 minutes with the restriction enzyme Dpn1. 50 μl of XL1-blue super competent cells were then transformed with 2 μl of Dpn-digested DNA. Transformation was carried out by incubating cells and ligation reaction for 30 minutes on ice followed by heat shock at 42ºC for 30 seconds and cooled the reaction on ice for 2 minutes. 500 μl of NZY+ broth was added to the cells and incubated in a 37ºC orbital shaker for 1 hour. Cells were then plated onto LB agar plates containing 50 mg/ml ampicillin and incubated overnight at 37ºC. The following days, a number of colonies were selected for screening for the presence of the correct mutation. Plasmid DNA was extracted with QIAPrep Spin Miniprep Kit (Qiagen) as described in section 2.1.8.5 and the presence of the correct mutations was confirmed by sequencing the full Gag-Pro region using Sanger sequencing as previously described in section 2.1.8.7.

Table 2.2 below is the list of primers used for the site directed mutagenesis experiments in Chapter 6.

Table 2.2: List of Primers designed and used for Site Directed Mutagenesis

Primer Name	Primer sequence (5'- 3')	Orientation	Description
25c2M46M	CCAGGAAGATGGAAACCAAAAATGATAGGGGG	Forward	V→M at protease position 46 in the susceptible clone
25c2M46M	CCCCCTATCATTTTTGGTTTCCATCTTCCTGG	Reverse	V→M at protease position 46 in the susceptible clone
25c8M46V	CCAGGAAGATGGAAACCAAAAGTGATAGGGGG	Forward	M→V protease mutation in the resistant clone
25c8M46V	CCCCCTATCACTTTTGGTTTCCATCTTCCTGG	Reverse	M→V protease mutation in the resistant clone
>25c8Mut4	GCAGCAGCTGCCGCAGAGAGCAGCAGCAGCCAAAATTACCCT	Forward	To insert S at Gag 126 of the resistant clone
>25c8Mut4	AGGGTAATTTTGGCTGCTGCTGCTCTCTGCGGCAGCTGCTGC	Reverse	To insert S at Gag 126 of the resistant clone
>25c8Mut6	GCAGCTGCCGCAGGGAGCAGCAGCCAC	Forward	To create 123 (E→G) in the resistant clone
>25c8Mut6	GTGGCTGCTGCTCCCTGCGGCAGCTGC	Reverse	To create 123 (E→G) in the resistant clone
25c8Mut5	GGCAGCAGCTGCCGCAGAGAGCAGCCACAGCCAAAATTACCCT	Forward	To insert His at Gag 127 of the resistant clone
25c8Mut5	AGGGTAATTTTGGCTGTGGCTGCTCTCTGCGGCAGCTGCTGCC	Reverse	To insert Hist at Gag 127 of the resistant clone
>25c8Mut7	GGCAGCAGCTGCCACAGAGAGCAGCAGCCACAGCCAAAATT	Forward	To create 122 (A→T) in the resistant clone
>25c8Mut7	AATTTTGGCTGTGGCTGCTGCTCTCTGTGGCAGCTGCTGCC	Reverse	To create 122 (A→T) in the resistant clone
>P8.9V128Del	GGACACAGCAATCAGAGCCAAAATTACCC	Forward	To delete V at Gag 128 of subtype-B virus
>P8.9V128Del	GGGTAATTTTGGCTCTGATTGCTGTGTCC	Reverse	To delete V at Gag 128 of subtype-B virus

R9BalCRF02AG Res_126S127H _Insertion	GCAGAGAGCAGCAGCCACAGCCAAAATTACCCTATAG	Forward	Gibson assembly oligos for the insertion of S and H at positions 126 and 127 of the patient-derived resistant clone
R9BalCRF02AG Res_126S127H _Insertion	CTATAGGGTAATTTTGGCTGTGGCTGCTGCTCTCTGC	Reverse	Gibson assembly oligos for the insertion of Ser and His at positions 126 and 127 of the patient-derived resistant clone
>R9BalCRF02A GRes_A122T_ E123G	GCAGCTGCCACAGGGAGCAGCAGC	Forward	Gibson assembly oligos for the reversion of A→T and E→G at positions 122 and 123 of the patient-derived resistant clone
>R9BalCRF02G Resistant_A122 T_E123G	GCTGCTGCTCCCTGTGGCAGCTGC	Reverse	Gibson assembly oligos for the reversion of A→T and E→G at positions 122 and 123 of the patient-derived resistant clone
>R9BalMut _126to128del	GCTGACACAGGAAACAACAGCCAAAATTACCCTATAGTGCAGAAC	Forward	To delete amino acid residues from Gag 126 to 128 of R9-BaL vector
>R9BalMut _126to128del	GTTCTGCACTATAGGGTAATTTTGGCTGTTGTTTCCTGTGTCAGC	Reverse	To delete amino acid residues from Gag 126 to 128 of R9-BaL vector
>R9bal_Mut_12 2A123E	GCAGCTGACGCAGAAACAACAGCCAGGTC	Forward	To create a T122A and G123E mutations in the R9-BaL vector
>R9bal_Mut_12 2A123E	GACCTGGCTGTTGTTTCTGCGTCAGCTGC	Reverse	To create a T122A and G123E mutations in the R9-BaL vector

>T122AG123E _126127DEL	GCAGCAGCTGACGCAGAACACAGCGTCAGCCAAAATTACCC	Forward	To create T122A and G123E mutations in R9-BaL vector
>T122AG123E _126127DEL	GGGTAATTTTGGCTGACGCTGTGTTCTGCGTCAGCTGCTGC	Reverse	To create T122A and G123E mutations in R9-BaL vector
>p8.9Del126127	GCAGCTGACACAGGACACAGCGTCAGCCAAAATTACCC	Forward	To delete amino residues at Gag positions126 and 127 of p8.9NSX vector
>p8.9Del126127	GGGTAATTTTGGCTGACGCTGTGTCCTGTGTCAGCTGC	Reverse	To delete amino residues at Gag positions126 and 127 of p8.9NSX vector
>p8.9T122AG12 3E	GCACAGCAAGCAGCAGCTGACGCAGAACACAGCAATCAGGTCA GCC	Forward	To create a T122A and G123E mutations in the p8.9NSX vector
>p8.9T122AG12 3E	GGCTGACCTGATTGCTGTGTTCTGCGTCAGCTGCTGCTTGCTG TGC	Reverse	To create a T122A and G123E mutations in the p8.9NSX vector
>25c2Mut2-1FWD	CCCCTTCCCCGAAGCAGGAGCTGGAACCGGGGGAC	Forward	To delete the amino acid E and L in the p6 of the resistant clone
>25c2Mut2-1FWD	GTCCCCCGGTTCCAGCTCCTGCTTCGGGGAAGGGG	Reverse	To delete the amino acid E and L in the p6 of the resistant clone
>25c2Mut2-2FWD	CCCCGAAGCAGGAGCTGCGGGAGGAACCGGGGGACAAGGGA C	Forward	To delete the amino acid R and E in the p6 of the resistant clone
>25c2Mut2-2REV	GTCCCTTGTCCCCCGGTTCCTCCCGCAGCTCCTGCTTCGGGG	Reverse	To delete the amino acid R and E in the p6 of the resistant clone

>25c8Mut2-1FWD	CCCCTTCTCCGAAGCAGGAACTGGAACCGAGGGACAAG	Forward	To Insert the amino acid E and L in the p6 of the susceptible clone
>25c8Mut2-1REV	CTTGTCCCTCGGTTCCAGTTCCTGCTTCGGAGAAGGGG	Reverse	To insert the amino acid E and L in the p6 of the susceptible clone
>25c8Mut2-2FWD	CCCCTTCTCCGAAGCAGGAACCGAGGGACAAGGGAC	Forward	To Insert the amino acid R and E in the p6 of the susceptible clone
>25c8Mut2-2REV	GTCCCTTGTCCCTCGGTTCCTGCTTCGGAGAAGGGG	Reverse	To insert the amino acid R and E in the p6 of the susceptible clone

2.2 General tissue culture techniques

All cells and pseudo-virus cultures were grown in humidified 37°C incubators with 5% CO_2 in varying volumes and passaged as required.

2.2.1 Thawing Cell Lines

HEK-293T cells were removed from liquid nitrogen (or -80 freezer) and thawed rapidly at 37°C. Cells were added to 10 ml of pre-heated DMEM media (Gibco) supplemented with 100 U/ml penicillin, 100 μg/ml streptomycin (Gibco) and 10% FCS (Biosera, UK). The cells were subsequently pelleted at 325g for 5 minutes, washed once in 10 ml of DMEM media and re-suspended in 15 ml of DMEM media in 10 cm dishes. The following day the media was replaced with fresh media.

2.2.2 Cell Lines Passage

HEK-293T and TZM-bl cells were maintained in DMEM media (Gibco) supplemented with 10% FCS (Biosera, UK), 100 U/ml of penicillin and 100 μg/ml streptomycin (Gibco). Cells were washed with phosphate buffered saline (PBS) [137 mM NaCl, 2.7 mM KCl, 4.3 mM Na2H(PO)4, 1.4 mM KH2(PO)4], incubated with 2 ml of trypsin-EDTA (Gibco) until the cells were detached from the dish. Cells were then pelleted at 325g for 5 minutes, the trypsin removed, and the cells re-suspended in fresh DMEM media. Cells were split 1:4 to 1:8, depending on the cell density and rate of growth, two or three times a week and grown in 5% CO_2 at 37°C.

2.2.3 Cell line Freezing

HEK-293 and TZM-bl cells were centrifugated at 325g for 5 minutes and re-suspended at $1x10^7$ cells/ml in 40% DMEM media, 50% FCS and 10% dimethyl sulphoxide (DMSO, Sigma, UK). Cells were then aliquoted into cryovials (Nunc, USA) and gradually cooled to -80°C in an isopropanol-containing cryo-container (Nalgene, USA) before being transferred to liquid nitrogen or -80˚C freezer.

2.2.4 Generation of pseudotyped Viruses

Pseudotyped viruses were produced by transient transfection of HEK 293-T cells with three plasmids: RTV containing patient' related Gag and protease sequences; pCSFLW containing the HIV packaging sequencing and the luciferase encoding gene and pMDG expressing vesicular stomatitis G (VSV-g) protein.

Transfection of 293T cells for PI susceptibility assay was typically carried out in tissue culture 6 well plates. Briefly, HEK 293-T cells were plated with 1.25×10^6 cells and incubated at 37 for 5 hours. 6 μl of FuGENE-6 (Promega) was added to 70 μl of Opti-MEM medium (Invitrogen). 500ng of pCSFLW, 300ng of pMDG and 300ng of p8.9NSX+ HIV Gag-pol expression vector were made up to 10 μl of TE buffer and added to the Fu-GENE-6 and Opti-MEM mixture. The transfection mixture was incubated for 15 minutes at room temperature before being added dropwise to the sub-confluent HEK 293-T cells in 1.5 ml of fresh DMEM medium (Invitrogen) supplemented with 10% FCS (Biosera, UK), 100 U/ml penicillin and 100 μg/ml streptomycin (Invitrogen). HEK 293-T cells and transfection mixture were incubated overnight at 37°C and 5% CO2. The following day, the cell culture medium was replaced for fresh medium. The pseudovirus containing supernatants were either directly employed in drug susceptibility assays or harvested at 48 and 72 hours, filtered with a 0.45 μm filter to eliminate cell debris and stored at -80°C in 1ml aliquots for subsequent applications.

Transfection conditions used for other plate sizes for both the PI susceptibility and single- round infectivity assays are detailed in table 2.0

Table 2.3 Transfection conditions

Plate Size	Number of Cells	Volume of Fugene (μl)	Amount of DNA (μg)	Harvest Volume (ml)
24 well plate	2.5×10^5	1.2	0.22	N/A
6 well plate	1.25×10^6	6	1.1	8
6 cm plate	3×10^6	6	1.1	22
10cm plate	1.25×10^7	18	3.5	N/A

2.2.5 Protease Inhibitor Susceptibility Assay

PI susceptibility was determined using a previously described phenotypic drug susceptibility assay[397,420]. Pseudovirus stocks used for PI susceptibility testing were obtained by co-transfecting HEK 293 T cells with RTV, pCSFLW and PMDG plasmids as described above. The cells were trypsinized approximately 16 hours after transfection and distributed into 96-well plates containing serial dilutions spanning and empirical determined range for each PI (between 1000 nM -0.005 nM). The top concentration of drug used for each PI can be seen in table 2.8. The 96 well plate layout for pseudovirion production containing three-fold serial dilutions of PI, is shown in figure 2.

Table 2.4 Top drug concentrations used in phenotypic drug susceptibility assay

PI	Top concentration (nM)
ATV	100
DRV	200
LPV	500
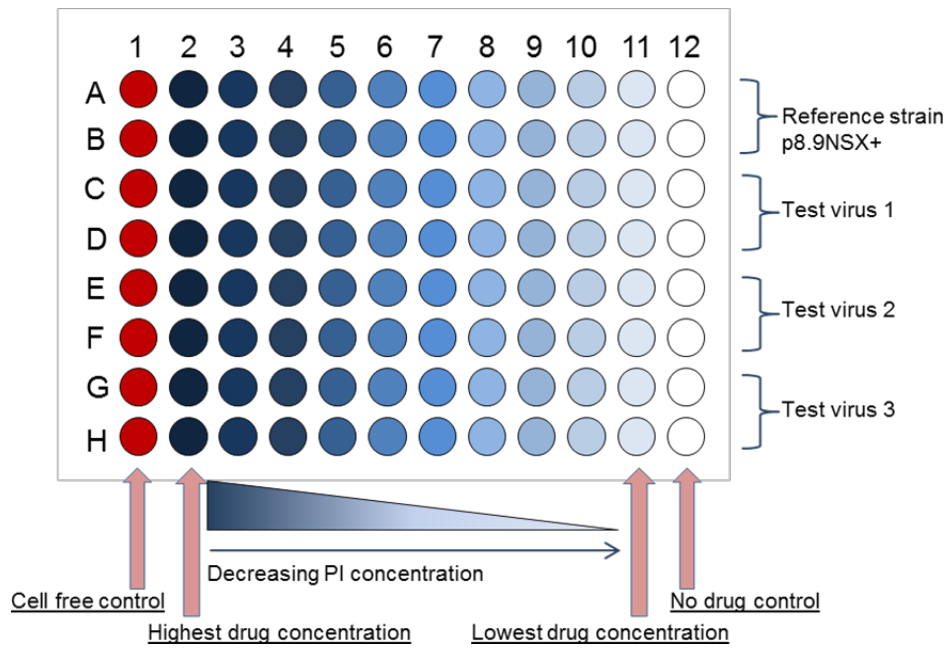
Figure 2.2 Layout of 96-well plate for PI susceptibility assays
The first row of the plate (1) was a no-cell control used to calculate the background levels of luciferase and row 12 a no-drug control, used to control for variation in transfection efficiency and replicative capacity between viruses. (Adapted with permission from Katherine Sutherland[456])

150 μl of cells were added to each well, containing 150 μl of media with the required concentration of PI. After 24 hours, media containing pseudovirions were harvested and used to infect fresh 293T or TZM-bl cells in 96 well plates (layout as shown in figure 2.2). Infectivity was then measured 48 hours after virus infection of the target cells, using SteadyGlo luciferase substrate (Promega) and a Glomax luminometer to determine luciferase activity levels.

Data were analyzed by plotting the percent inhibition of luciferase activity versus $\log_{10}$ drug concentration. The percent inhibition was derived as follows: [1-(luciferase activity in the presence of drug-background)/ (luciferase activity in the absence of drug-background)] x 100. Mean percent inhibition for each drug concentration was determined from independent replicates and the standard deviation calculated. Inhibition curves, defined by the four- parametric sigmoidal function f(x) = a- [b/ (1+(x/c)d], were fitted to the data by nonlinear least- squares and used to calculate the drug concentration required to inhibit virus replication by 50% (EC_{50}). The fold difference (FC) in drug susceptibility is determined by comparing the EC_{50} for the tested virus to the EC_{50} of the WT reference virus (P8.9NSX+) which contains the PR and RT sequences of the HXB2 strain of HIV-1. All analysis was performed employing GraphPad Prism version 8. Assays were performed either in duplicate or in triplicate for each virus construct.
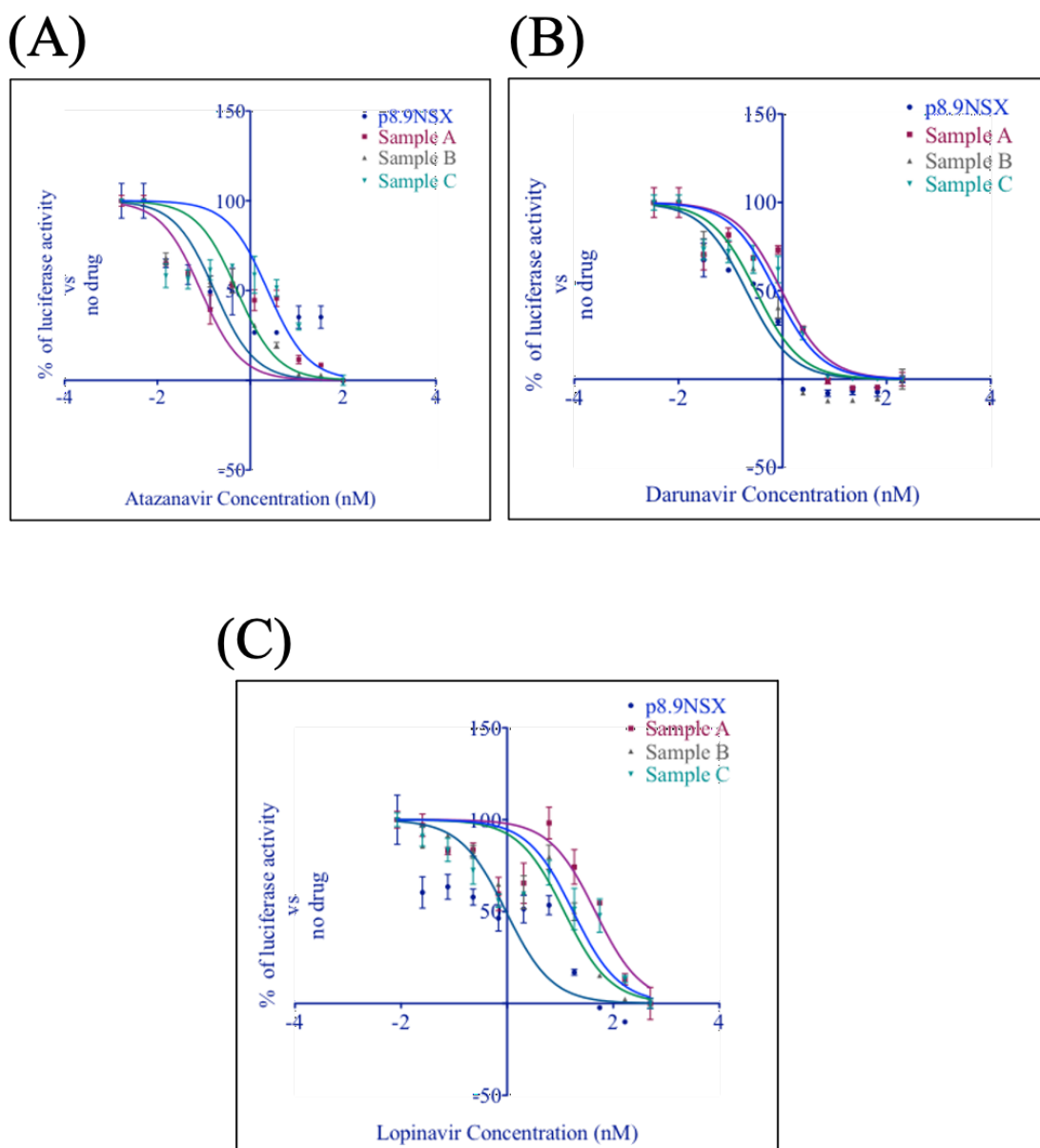
Figure 2.3: Representative titration curve for phenotypic protease inhibitor susceptibility assays

Pseudotyped viruses were produced by co-transfection of 293T cells with HIV Gag-Pol expression vector P8.9NSX+, VSV-g expression vector pMDG and pCSFLW. Eighteen hours post-transfection, cells were incubated with serial dilutions of PI. The infectivity of pseudovirions produced in the presence of each PI concentration was tested. Typical curves for the three PIs ATV (A), DRV (B) and LPV (C) are shown.

2.2.6 Single-round infectivity assay

Single-round infectivity was determined as previously described by titration of serial pseudovirion dilutions incubated with 293T cells in the absence of drug.[397,420] Pseudovirus was harvested 48 hours post-transfection and passed through a 0.45 μm filter to remove cellular debris. 100 μl virus was used immediately for titration, and the remaining volume stored at -80°C until required. Titration was carried out in 96 well, white plates with each virus in a single row as shown in figure 2.3. Virus was plated using two-fold dilutions across the plate, resulting in a range from 50 μl volume of virus in the first column to 0.025 μl of virus in the twelfth. A blank row was included in each run to measure background luminescence. 48 hours post-titration, infectivity was measured using SteadyGlo luciferase substrate (Promega) and a Glomax luminometer.
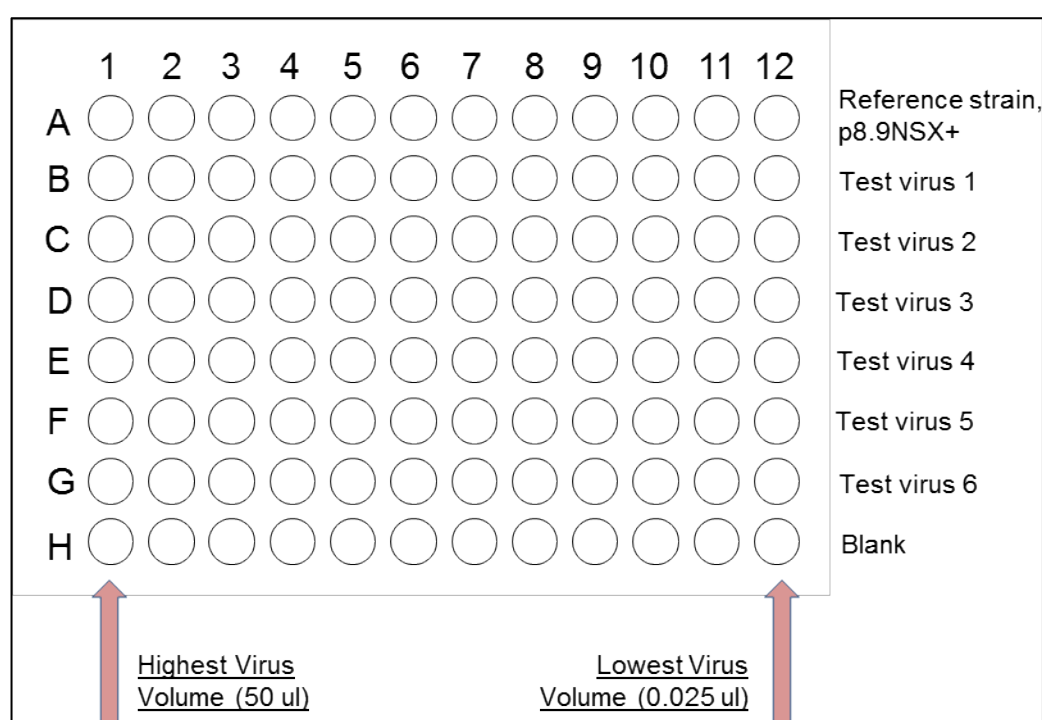
Figure 2.3 Layout of 96-well plate for single-round infectivity assay

2.2.7 Lentivector concentration

Lentiviral vectors were concentrated by ultracentrifugation in a Sorvall Discovery (Hitachi) at 23,000 rpm for 2 hours at 4°C under vacuum conditions through a 20% sucrose cushion. The pellet was then resuspended in DMEM + 10% FCS and stored at -80°C.

2.3 Statistical analyses

Statistical analyses were performed using formulae in Microsoft Excel or GraphPad prism software. In most cases, unpaired t-tests were used to compare the means of two populations, with $P < 0.05$ considered statistically significant.

Differences in PI susceptibility were compared using the Mann-Whitney U rank sum tests (GraphPad Software Inc., La Jolla, CA, USA). Replication capacity and fold-changes in EC_{50} were also compared using 2-way ANOVA with Holm-Šídák's multiple-comparisons. A $p < 0.05$ was considered to be statistically significant. When groups differed significantly, a Bonferroni's multiple comparison post-test was performed to make two by two comparisons

2.4 Western Blotting

2.4.1 Preparation of Laemlli Buffer

To make 10 ml of 4x stock of Laemmli reducing buffer, the following recipe was used: 2.0 ml 1M Tris-HCl pH 6.8., 0.8 g SDS, 4.0 ml 100% glycerol, 2 ml β-mercaptoethanol, and 8mg bromophenol Blue.

2.4.2 Sodium dodecyl sulphate - polyacrylamide gel electrophoresis (SDS-PAGE) and Immunoblotting

For producer cell lysates, plates containing the cells were placed on ice, Laemlli buffer (1x) was added to the cells and cell lysates were scrapped into Eppendorf tubes, still on ice. This was followed by boiling at 100°C for 10 minutes.

For harvested virus supernatant, 1 volume of Laemlli buffer was added to 4 volumes of virus sup, followed by boiling at 100°C for 10 minutes. Cell lysates

and virus supernatant were briefly centrifuged and either stored at -20°C or used immediately.

Prepared samples were loaded onto pre-cast gels submerged in 1x running buffer (prepared in-house at 10x; 144g glycine, 30g Tris base, up to 1L H_2O, pH8.3). Twenty microlitres ($20\mu L$) of sample was loaded per well, alongside $5\mu L$ SDS-Page ruler plus pre-stained protein ladder (Fermentas). Samples were run at 80V for 25mins to get through the stacking gel and 120V for 1 hour, or until appropriate separation of proteins was achieved.

2.4.3 Transfer and Blocking

After PAGE, proteins were transferred to a Hybond nitrocellulose or PVDF membrane (Amersham biosciences) in transfer buffer (25 mM Tris-HCl, 250 mM glycine, 20 % (v/v) methanol) using a semi-dry transfer system (Biorad). After transfer, membranes were blocked for non-specific antibody binding by incubation on a rocking platform for 1 hour at room temperature in 5 % (w/v) milk proteins + 0.01 % (v/v) Tween-20 in PBS (PBST) or TBS depending on downstream processes.

2.4.4 Visualization of proteins

The membranes were then incubated on a rocking platform overnight at 4°C with primary antibody (Ab) diluted in 5 % (w/v) milk proteins in PBST or TBST. This was followed by a series of wash steps in PBST or TBST, and incubation in secondary antibody diluted in 5 % (w/v) milk proteins at room temperature for 1 hour. This was followed by washing in PBST and the addition of substrate (Amersham™ ECL™ Prime western blotting detection system) to the membranes and detection of HIV-1 proteins using a transilluminator (Alpha Nanotech) or Chemidoc (Biorad).

CHAPTER 3: Genotypic comparison of baseline viruses from patients who succeeded on, and those who failed Protease inhibitor-based second-line treatment

3.1 Introduction

HIV-1 treatment in Low- and Middle-Income Countries (LMICs) follow the WHO guidelines[474] which recommends the use of two nucleoside reverse-transcriptase inhibitors (NRTIs) plus a non-nucleoside reverse-transcriptase inhibitor (NNRTI) or an integrase inhibitor (INSTI) as first line treatment in adults. Upon failing first line treatment, the guideline recommends a switch to second line, which should consist of two nucleoside reverse-transcriptase inhibitors (NRTIs) plus a ritonavir-boosted protease inhibitor (PI).

Prevalence of virological failure for first-line antiretroviral therapy can be as high as 30%, [497] with high-level resistance to NNRTI, tenofovir and cytosine analogues common in resource-limited settings and compounded by prior undisclosed ART.[498,499] Second-line ART recommended by WHO comprises a ritonavir-boosted PI and two NRTIs, commonly lopinavir or atazanavir.[474] PIs are the second- and last-line therapy for the majority of HIV-infected patients worldwide as access to third-line therapy is still limited[500]. Virological failure with PIs as second-line therapy occurs in around 20% of individuals[501-503]. In contrast to first-line therapy, with which >80% develop drug resistance mutations, only around 10%–20% develop major resistance mutations to PIs by week 48,[403,501,502,504] but this proportion increases over time[500].

It is known that proteins such as Gag and Env can affect susceptibility to PIs even in the absence of known major resistance mutations in the protease gene[197,355,394,397,505]. There has been increasing evidence on the role of HIV-1 Gag mutations in resistance to protease inhibitors mainly in subtype B viruses[394,397,407,420,506-508]. Mutations and substitutions occurring in the Gag

cleavage sites of HIV-1 isolated from ART-treated patients have been identified as playing roles in resistance to PIs. However, PI-resistant HIV-1 variants lacking cleavage site amino acid substitutions have been observed[395,506].

There are limited data on changes in Gag following treatment failure with PIs in the non-B subtypes that dominate low- and middle-income countries.[394,416,439,509-511]. In around 15% of patients failing boosted PI (bPI) without major protease mutations, a decrease in phenotypic susceptibility to the drug appears to occur when *gag-protease* is phenotyped[507,512,513]. Therefore, it is conceivable that underlying phenotypic susceptibility resulting from variation in genes such as Gag and env might impact clinical responses to PI.

It was previously shown that Gag-Protease-derived phenotypic susceptibility differed between CRF02_AG and subtype G-infected patients who went on to successfully suppress viral replication versus those who experienced virological failure (VF) of lopinavir/ritonavir monotherapy as first-line treatment in a clinical trial.[505] In order to determine the relevance of this finding for real-world settings in the context of combination NRTI therapy plus ritonavir-boosted PI (lopinavir or atazanavir) as used in Nigeria, this study analysed the relationship between PI susceptibility and the outcome of PI-based second-line ART in Nigeria, where subtypes CRF02_AG and G dominate the epidemic[514].

Unlike in previous studies where 'baseline' was described as treatment naïve, this current study utilized samples from patients who have failed 1st line ART and subsequently switched to 2nd-line PI-based ART. 'Baseline' in this chapter therefore refers to patient samples taken after first line ART, but prior to the administration of PI. In other words, 'baseline' means PI-naïve, but not ART naïve. In this chapter, patients who achieve virological suppression are denoted 'Baseline successes' while patients who did not achieve virological suppression are denoted 'Baseline failures.

3.1.1 Chapter Aims:

1) Amplify full-length Gag-protease from baseline samples of 'Baseline successes', matched to the 'Baseline failure' patients employing the assays described in chapter 2.

2) Analyse genotypic changes for amino acid positions previously described to be associated with PI exposure or PI resistance in viruses derived from patient pairs

3) Determine a putative significant association between protease mutations and Gag mutations that may identify novel pathways of PI resistance.

3.2 Study participants

This study involved retrospectively testing samples from patients attending for HIV care at University of Abuja Teaching Hospital (UATH).

A case-control analysis was used mainly to determine the viral factors associated with second-line PI treatment failure. The statistical analysis took into account matching by using a conditional logistic regression with a Cox model.

3.2.1 Selection criteria:

A. Cases designated as "Baseline Failures"

1. Age >15 years

2. Received a first-line ART regimen of 2 NRTIs and 1 NNRTI, then a second-line regimen of 2 NRTIs and a PI (lopinavir, darunavir or atazanavir)

3. Baseline stored plasma sample obtained before starting second-line therapy with a viral load > 1,000 copies/mL

4. Subsequent virological failure on second-line regimen (viral load at least 1,000 copies/mL) at least six months after second line (2L) initiation

B. Controls, designated as "Baseline Successes"

1. Age >15 years

2. Received a first-line ART regimen of 2 NRTIs and 1 NNRTI, then a second-line regimen of 2 NRTIs and a PI (lopinavir, darunavir or atazanavir)

3. Baseline stored plasma sample obtained before starting second-line therapy with a viral load > 1,000 copies/mL

4. Confirmed virological suppression on second-line regimen (to viral load <400 copies/mL) on at least one blood test following second-line initiation

5. Matched with the "failures" on subtype, sex, CD4 at initiation, duration of second-line ART

Patient samples were carefully selected to include only samples without any major PI mutations using the Stanford HIV resistance database algorithm. Also, only HIV-1 subtypes G and CRF02_AG were included in the viral genetic and phenotypic analyses. Baseline (pre-PI) plasma samples from these matched pairs were retrospectively retrieved. Viral loads and CD4 counts were measured at accredited facilities in Abuja, Nigeria.

3.3   Methods

3.3.1 Whole genome sequencing and generation of consensus sequences

Whole HIV-1 genome NGS was undertaken in collaboration with the Breuer lab at UCL and Kate El Bouzidi (Wellcome Trust Clinical Research Training Fellow). Briefly, following viral RNA extraction (as described in chapter 2, section 2.1.1) library preparation and target enrichment was carried out using S*ureSelect* kit. NGS data was generated using the Illumina Miseq sequencer. Visual inspection/quality control of fastq files was performed using Fastqc, followed by the use of *Trimgalore* to remove poor quality reads. These reads were then compared to reference panel of 170 HIV subtypes/CRFs from LANL database. The top hits generated were filed. The best reference (CRF02_AG or G subtypes) was selected and reads mapped to that reference using Burrows-Wheeler aligner (BWA), followed by the use of Samtools to convert *sam* files to *bam* file format. Duplicate reads were removed using *picard*. Consensus sequences were then generated with *bcftools*. Consensus *Gag* and *pol* sequences were generated. Consensus sequences were manually edited and trimmed in MEGA[515] or DNA Dynamo (https://www.bluetractorsoftware.com) software.

The NGS consensus sequences of the patients generated were aligned to HXB2 strain (which is used as reference throughout this thesis).

3.3.2 Classification of protease and Reverse RT mutations

We assessed Protease and Reverse Transcriptase (RT) resistance mutations according to the 2018 list of mutations of the International AIDS Society[496]. PI mutations were classified into three groups: major resistance mutations, accessory mutations or "other" mutations. NRTI and NNRTI mutations were classified into two groups: resistance mutations and 'other' mutations according to the Stanford HIV Drug Resistance Database:

(https://hivdb.stanford.edu/hivdb/by-sequences/).

3.3.3 Classification of Gag Mutations

We analysed mutations in the entire Gag protein. Gag sequences were aligned with the reference sequence HXB2 and subtyped by submitting the sequence to Rega subtyping tools available online at:

http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/.

Mutations were defined as any change relative to the HXB2 reference sequence and divided into those seen in cleavage sites (CSMs) and those seen outside cleavage sites (non-CSMs). Each cleavage site consisted of the five amino acids on both sides of the cleavage bond. P5 to P1 and P5' to P1' for residues on the N and C terminal sides of the target, respectively (Figure 4.2).

MA
1
MGARASVLSGG ELDRWEKIRL RPGGKKKYKL KHIVWASREL ERFAVNPGLL ETSEGCRQIL GQLQPSLQTG

CA
128 - 137 ⇩
SEELRSLYNT VATLYCVHQR IEIKDTKEAL DKIEEEQNKS KKKAQQAAAD TGHSNQVSQN YPIVQNIQGQ

MVHQAISPRT LNAWVKVVEE KAFSPEVIPM FSALSEGATP QDLNTMLNTV GGHQAAMQML KETINEEAAE

WDRVHPVHAG PIAPGQMREP RGSDIAGTTS TLQEQIGWMT NNPPIPVGEI YKRWIILGLN KIVRMYSPTS

ILDIRQGPKE PFRDYVDRFY KTLRAEQASQ EVKNWMTETL LVQNANPDCK TILKALGPAA TLEEMMTACQ

p2   NC
359 – 368 ⇩   ⇩ 373 - 381
GVGGPGHKAR VLAEAMSQVT NSATIMMQRG NFRNQRKIVK CFNCGKEGHT ARNCRAPRKK GCWKCGKEGH

p1   p6
428 – 437 ⇩   ⇩ 444 – 453   487
QMKDCTERQA NFLGKIWPSY KGRPGN FLQS RPEPTAPPEE SFRSGVETTT PPQKQEPIDK ELYPLT

Figure 3.1: Gag HXB2 Sequence

The letters indicate the reference (HXB2) Gag amino acid sequence. All amino acids are indicated by their one letter code. The number position of the first and last amino acid of the Gag polyprotein is indicated. The beginning and end of each individual protein is indicated by arrows. Cleavage site sequence and positions are indicated in red. The arrows indicate positions of the cleavage bond.

3.4 Results

Originally, after a rigorous process of sample identification and screening, fifteen pairs (30 samples) were identified for this study. All thirty samples were PCR-amplified. However, after using different cloning techniques, we were only able to successfully clone 6 pairs (12 samples). The results presented in this chapter and in chapter 4 are from these 12 patient samples.

3.4.1 Clinical and virological information for 'Baseline successes' vs 'Baseline failures'

Patients were broadly classified as 'Baseline successes' – defined as those who suppressed viral replication for 48 weeks after switching to second-line PI-based treatment or 'failures' – defined as patients who experienced virological failure on second-line PI treatment (as shown in Table 3.1) 'Successes' were matched to the 'failures' using the following baseline criteria: viral subtype, CD4 count, age, sex and protease inhibitor used as second line treatment. These matched pairs based on these criteria are presented in table 3.2.

The median (IQR) CD4$^+$ count of 'Baseline successes' vs 'Baseline failures' was 108 (46 - 198) and 174 (119 - 217) cells/mm^3 respectively while the median (IQR) log$_{10}$ viral load across the two groups was 4.60 (4.4 – 5.5) and 4.75 (4.3 – 5.4) copies/mL respectively. Median age (IQR) of the patients across the two groups was 35 (32 - 44) and 38 (31 - 41) years respectively with all patients being female. As can be seen from table 3.2, all pairs had a baseline CD4 count <200 cells/mm^3 except sample pair 4. All patient pairs had a lopinavir-based second-line treatment except for patient pair 5 who received atazanavir-based second line therapy.

Table 3.1: Clinical and virological information for patients studied

Treatment Outcome	Patient Number	Subtype	Baseline (before commencement of 2nd line PI treatment)				After ≥ 48 weeks on PI-based 2nd line treatment	
			Log_{10} Viral Load		CD4+T cell count		Log_{10} Viral Load	
			copies/ml	Median (IQR)	cells/mL	Median (IQR)	copies/ml	Median (IQR)
'Baseline Successes'	#54	CRF02_AG	5.7	4.60 (4.4 – 5.5)	29	108 (46 - 198)	<20	
	#20	G	4.6		145		<20	
	#17	G	4.5		52		<20	
	#40	CRF02_AG	5.4		331		<20	
	#5	CRF02_AG	4.2		71		<20	
	#21	CRF02_AG	4.6		153		<20	
'Baseline Failures'	#36	CRF02_AG	5.1	4.75 (4.3 – 5.4)	162	174 (119 - 217)	3.8	4.7 (4.1 -5.0)
	#11	G	4.3		203		5.1	
	#43	G	5.4		27		4.9	
	#13	CRF02_AG	4.4		259		4.5	
	#58	CRF02_AG	5.4		149		4.2	
	#9	CRF02_AG	4.3		185		4.8	

Table 3.2: Clinical data for matched patient pairs comprising virological Baseline successes and Baseline failures

Sample Pair	Age (years)		Sex		Baseline* CD4 count (cells/mm³)		Second Line PI used		Baseline* Viral load (Copies of HIV-1 RNA/mL)	
	success	failure	success	failure	success	failure	success	failure	success	failure
1	33	45	female	female	<200	<200	LPV	LPV	503 951	140 991
2	43	36	female	female	<200	<200	LPV	LPV	39 844	20 178
3	27	26	female	female	<200	<200	LPV	LPV	32 284	271 974
4	47	39	female	female	200 - 499	200 -499	LPV	LPV	228 083	24 693
5	35	40	female	female	<200	<200	ATV	ATV	14 487	274 504
6	34	33	female	female	<200	<200	LPV	LPV	39 929	18 056

LPV, lopinavir; ATV, atazanavir; * Baseline refers to pre-initiation of second line therapy

3.4.2   NRTI and NNRTI Resistance mutations

Tables 3.1 and 3.2 show that of 6 pairs, 2 were infected with subtype G HIV-1 and 4 with CRF02_AG. Age of participants ranged from 26-47 and all were female. All but one pair had CD4 counts at 'baseline' of below 200 cells/mm³ indicating profound immune suppression that is typical of patients failing first line in sub-Saharan Africa. Viral loads ranged from 14,500 copies/ml to nearly 504,000 copies/ml. All but one pair was treated with LPV/r and the remaining pair was treated with ATV/r. In terms of the NRTI component of second line, all pairs but one was treated with TDF/FTC and one pair received AZT/3TC, likely due to use of TDF/FTC in the first line regimen.

Given that the patients in this study have all been on first line NRTI and NNRTI regimens (as shown in table 3.3), as at the time of sample collection, the *pol* NGS consensus sequences were submitted to the HIVdb for the assessment of protease and RT mutations. The NRTI and NNRTI resistance mutations in *pol* detected prior to second-line initiation at 20% frequency or above are shown in Table 3.3.

We observed multi-drug resistance in most individuals. All 12 individuals had high-level NNRTI resistance with mutations including **K103N**, the most commonly observed mutation following efavirenz failure[516]. Seven of the twelve individuals had accumulated more than one major NNRTI resistance mutation, including **Y181C, K101E** and **G190A** in four patients each.

All patients had lamivudine resistance [**M184V/I** in reverse transcriptase (RT)] and 7/12 (58.3%) had at least moderate resistance to tenofovir (3 with **K65R**, 3 with **K70E** and 1 with three thymidine analogue mutations including **M41L, L210W and T215Y**). Importantly, resistance to tenofovir and lamivudine was similar in 'Baseline success' and 'Baseline failure' groups, and therefore the NRTI activity of second line regimens was similar across the two.

The presence of **M41L, L210W** and **T215Y** (selected by thymidine analogues such as zidovudine and stavudine) from the TAM 1 pathway together confers intermediate resistance to tenofovir[517]. Interestingly, two individuals in the 'Baseline success' group had multiple mutations from the TAM 2 pathway, including **D67N** and **K70R**.

Table 3.3: NRTI and NNRTI mutations observed at first-line failure, prior to initiation of second-line PI-based ART

		NRTI mutations	NNRTI mutations	Baseline* VL (copies of HIV-1 RNA/mL)	HIV-1 subtype	2L backbone
Pair 1	success	M41L, L74LI, M184V, L210W, T215F	K101E, E138Q, G190A	503 951	CRF02_AG	TDF/FTC
	failure	M184V	K103N	140 991	CRF02_AG	TDF/FTC
Pair 2	success	E44D, D67N, T69D, K70R, M184V, T215Y	K101E, K103N	39 844	G	TDF/FTC
	failure	M184V	K101E, G190A	20 178	G	TDF/FTC
Pair 3	success	K70E, M184V	A98G, Y181C	32 284	G	TDF/FTC
	failure	K70E, M184V	Y181C, G190A, H221Y	271 974	G	TDF/FTC
Pair 4	success	K70E, Y115F, M184V	K103N	228 083	CRF02_AG	AZT/3TC
	failure	K65R, M184V	K101E, V108I, Y181C, G190A	24 693	CRF02_AG	AZT/3TC
Pair 5	success	D67N, K70R, M184V, T215F, K219E	Y188C	14 487	G	TDF/FTC
	failure	K70R, M184V, K219Q	K103N, Y318F	274 504	G	TDF/FTC
Pair 6	success	K65R, M184I	K103N, Y181C	39 929	CRF02_AG	TDF/FTC
	failure	K65R, M184I	K103N, Y181C	18 056	CRF02_AG	TDF/FTC

AZT, zidovudine; 3TC, lamivudine; FTC, emtricitabine; TDF, tenofovir disoproxil fumarate
*Baseline refers to pre-initiation of second-line therapy

3.4.3 Protease Mutations (and Polymorphisms)

The NGS consensus protease sequences were submitted to the Stanford HIVdb. Across both the 'Baseline success' and 'Baseline failure' groups, no major protease inhibitor resistance associated mutations were detected. Although the protease was much conserved, there were 24 variable sites among the 'Baseline success' vs 20 variable sites in the 'Baseline failure' samples. Table 3.4 and figure 3.3 shows the protease polymorphisms detected in all samples. While table 3.3 shows the pairwise occurrence of the polymorphisms, figure 3.3 shows the difference in occurrence of the polymorphisms when the two groups are considered all together. Across all the protease sequences from the patients in both groups, the HIVdb predicted a susceptibility to the following PIs: atazanavir/r (ATV/r), darunavir/r (DRV/r), fosamprenavir/r (FPV/r), indinavir/r (IDV/r), lopinavir/r (LPV/r), nelfinavir (NFV), saquinavir/r (SQV/r), and tipranavir/r (TPV/r). The occurrence of protease polymorphisms was dispersed in both groups. The **I13V, M36I** and **L89M** 'other' mutations were found in all 12 samples, irrespective of treatment outcome group. Although the protease I13V is a consensus amino acid in CRF02_AG and G subtypes, and are shown to be present in all 12 patient samples, it has been shown to contribute to reduction in PI susceptibility, in the absence of major resistance mutations in subtype-B viruses[507]

The **M36I** in protease occurring alongside **E12K** in Gag, have been described to affect PI susceptibility in subtype B viruses, but are consensus amino acids in CRF02_AG and G subtypes[392,418]. The **L10I/V** mutation was observed in one sample (in the 'Baseline success' group). It is a polymorphic, PI-selected accessory mutation that increases the replication of viruses with other PI-resistance mutations. As for the **K20I** that was seen in all-but one sample, it is the consensus amino acid in subtype G and CRF02_AG. In subtypes B and C, K20I is a PI-selected accessory mutation that reduces NFV susceptibility. One sample had the **K20V** which has been described as rare, relatively non-polymorphic PI-selected mutation. **V82I** is a highly polymorphic mutation that is not selected by PIs. It is the consensus amino acid in subtype G viruses.

Table 3.4: Protease polymorphisms observed at first-line failure, prior to initiation of second-line PI-based ART

		Protease polymorphisms
Pair 1	Success (#54)	I13V, K14R, K20I, E34Q, E35D, M36I, P39Q, R41K, L63P, C67E, H69K, V77I, L89M
	Failure (#36)	I13V, K14R, G16E, K20I, E35Q, M36I, R41K, R57K, H69K, L89M
Pair 2	Success (#20)	L10I, I13V, K20I, M36I, R41K, L63A, C67S, H69Q, L89M
	Failure (#11)	I13V, K14R, I15V, K20I, E35Q, M36I, R41K, R57K, C67E, H69K, I72V, V82I, L89M
Pair 3	Success (#17)	I13V, K14R, I15V, L19I, K20I, E35Q, M36I, R41K, R57K, C67E, H 69K, V82I, L89M
	Failure (#43)	I13V, K14R, K20I, M36I, R41K, C67E, H69R, K70R, V82I, L89M
Pair 4	Success (#40)	I13V, G16E, L19I, K20I, E35D, M36I, R41K, I64M, H69K, K70R, L89M
	Failure (#13)	I13V, K20I, M36I, L63P, G68E, H69K, L89M
Pair 5	Success (#5)	I13V, K20I, M36I, R41K, I64M, H69K, K70R, L89M
	Failure (#58)	I13V, K20I, M36I, R41K, L63P, C67S, H69K, L89M
Pair 6	Success (#21)	I13V, K14R, G17E, K20V, E34Q, M36I, N37S, R41K, L63P, H69K, K70R, V77I, L89M
	Failure (#9)	I13V, K14M, L19P, K20I, M36I, N37D, L63P, H69K, L89M
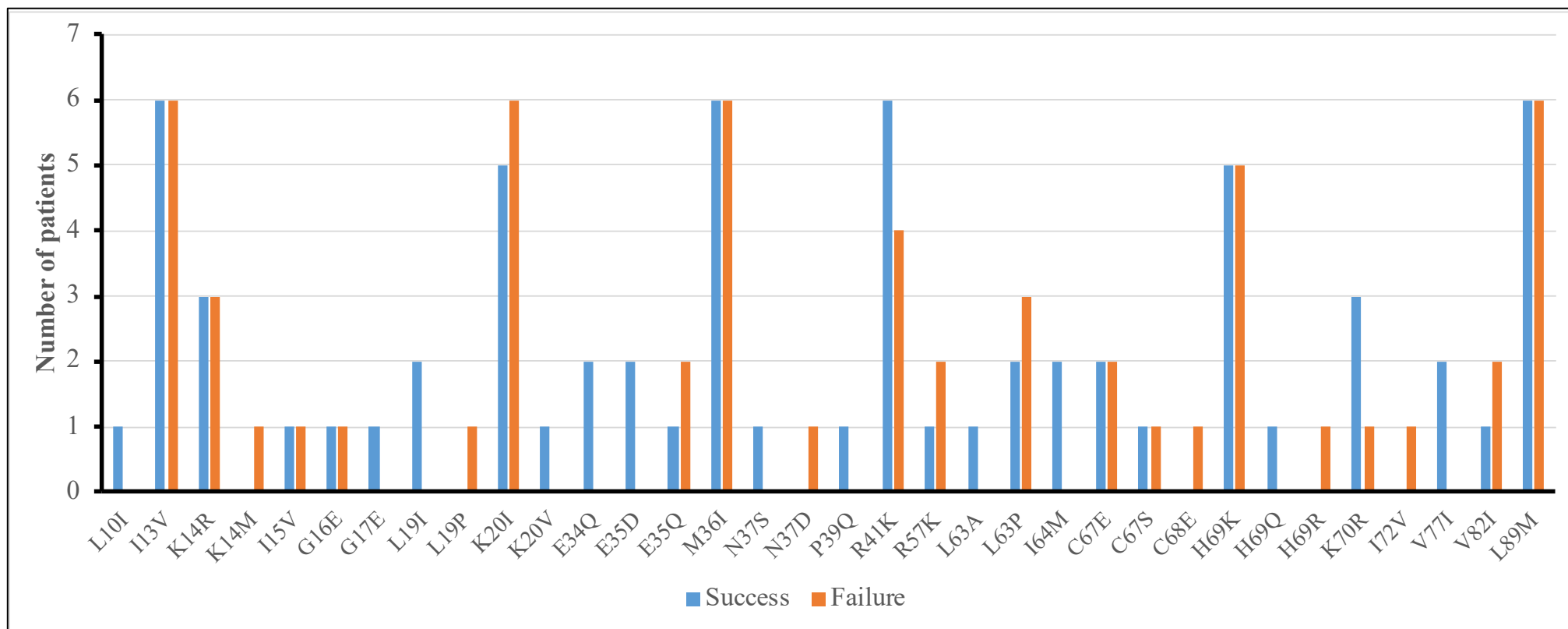
Fig 3.3: Protease polymorphisms observed at first-line failure, prior to initiation of second-line PI-based ART

There were no major resistance mutations in the protease gene. A number of protease polymorphisms were observed to be distributed throughout the protease gene. These polymorphisms were observed in both the 'Baseline Success' and 'Baseline failure' groups.

3.4.4 Gag variability at Cleavage and non-cleavage site positions

First, we examined the 487 amino acid positions setting up the sequence of Gag protein from both the 'Baseline success' and 'Baseline failure' groups. The sequences were compared to the reference sequence HXB2. Non-cleavage site amino acid positions (n=437) and cleavage site amino positions (n=50) were examined separately. The five Gag cleavage sites (p17/p24; p24/p2; p2/p7; p7/p1; and p1/p6) each consist of 10 amino acids.

Full-length Gag and protease sequences of from both groups of patients are shown in figures 3.4 and 3.5 with the 10 amino acids of each of the Gag CS shown in red boxes.

Sequence variability between the patient samples and the HXB2 reference strain was high and dispersed across the entire Gag protein, with the CA (p24) being the most conserved region and the p2 region being the least conserved. This is in agreement with a similar finding by Li and colleagues who examined Gag sequences. In their study, the capsid protein (29.4%) contained the lowest number of polymorphic positions followed by nucleocapsid (42.5%), matrix (59.9%), and p6 (65.6%)[442].

It has previously been reported that some amino acid mutations in Gag, which confer PI resistance in subtype B viruses, occur as natural polymorphisms in non-B[518,519] (such as CRF02_AG and G) subtypes. Previous studies has also identified Gag mutations and polymorphisms which are associated with PI exposure or PI treatment (as reviewed by Fun and colleagues)[394]. We generated a list of these previously reported, PI-conferring mutations (table 1.9, chapter 1), and compared the patient-derived sequences (aligned to HXB2 reference). The consensus amino acid sequences of these sites are shown in red colour in figures 3.3a and 3.3b.

3.4.5 Cleavage Site Polymorphisms of 'Baseline success' and 'Baseline failure' groups

In the MA/CA cleavage site, the V128del was seen in 4/6 samples from the 'Baseline success' group and 2/6 samples from the 'Baseline failure' group. The **Y32F** was seen in 1/6 samples from the 'Baseline success' group, but none from the 'Baseline failure' group. One patient sample in the 'Baseline failure' group had a **V135L**, with this not seen in any of the samples from the 'Baseline success' group.

The CA/p2 cleavage site was conserved between the two groups, except for the **V362I** that was observed in 1/6 samples in each of the two patient groups.

The p2/NC cleavage site was the most variable of all the sites. At amino acid position **S373, P/T** polymorphisms were observed in 1/6 samples in the 'Baseline success' group and in none of the 'Baseline failure' group samples. The **S373A** was observed in 3 samples each from both groups. There was the occurrence of **S373Q** in 3/6 samples in the 'Baseline failure' group, but not in any from the 'Baseline success' group. In the NC/p1 cleavage site, the **K436R** polymorphism was observed in 1/6 'Baseline failure' and 1/6 'Baseline success' samples. Two other polymorphisms (**I437L** and **I437V**) were seen in the 'Baseline success' group, but not in the 'Baseline failure' group.

In the p1/p6 domain of Gag, **F448P** occurred in 1/6 patients from the 'Baseline success' group only, **L449P** occurred in 4/6 and 3/6 patients from the 'Baseline success' vs 'Baseline failure' groups respectively. Occurring in 1/6 patients from both groups was the **S451N**, while the **R452L** was observed in 1/6 patients from the 'Baseline success' group and **P453L** in 1/6 patients from the 'Baseline failure' group.

3.4.6 Non-Cleavage Site Mutations (Polymorphisms) of 'Baseline success' and 'Baseline failure' groups

Several polymorphisms were seen at non-cleavage sites in across both 'Baseline success' and 'Baseline failure' groups. These were diverse and can be seen

across the entire Gag protein. Several non-cleavage site mutations have previously been linked to protease inhibitor exposure, or failure, with most of these studies carried out in B-subtypes. These include; **E12K/N/R, G62E/Q/T, L75I/F, R76K, Y79F, T81A, K112Q** in the matrix domain. Others are the **H219Q** (in the capsid), **Q369K, V370A, T371H/S/Q/N/del** (in the p2), **I389T/N/del, V390I, I401V/L** (in the NC**), E468G, Q474P** (in the p6). As well as others that have not been previously described and may be the consensus in CRF02_AG and G subtypes.

Gag

```
HXB2         1 MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI
54.Success   1 ............K..S.............R...L..........L.......SA...Q.L
36.Failure   1 ............K..H.............R...L..........L.......A...Q.L
20.Success   1 ............K..A.....................L.H.....A...V..
11.Failure   1 ............K..A.........Q.RI.L..........L.......SA...Q..
17.Success   1 ............K..S.............R...L..........L.......E...Q..
43.Failure   1 ............K..S.......R.............M...L.RD....A...M..
40.Success   1 ......I.R....A..R...........R...L..........L......T...Q.L
13.Failure   1 ............K..S.............R...L..........L.......A...Q.A
5.Success    1 ............K..Q.............R...L..........L.......SA...Q.L
58.Failure   1 ............R..A....................L.D....A...A.L
21.Success   1 ............K..A.......M.....R...L..........L..S....AG..QEL
9.Failure    1 ........T..N..A.............R...L..........L.......G...Q.L
consensus    1 ******.*.**.***.**.*****.*.*.*..**.******.****.*..***...**...

HXB2        61 LGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAA
54.Success  61 ME..SN.K......K..F..L...W.......D.......E.V..A.KR...PT.....
36.Failure  61 IE...TT.K.......F.AI...W............V.K...QNT..-..
20.Success  61 MQ..K.A....T...K..F.........K..V........EV.KM.---.S.T....R
11.Failure  61 MQ....A.K..T..IK............GV........E..KI.K..QQEX.K.E.
17.Success  61 M.....A.K..T..IK.........K..V........EV.KI.KN.QQET.K..-
43.Failure  61 ME.I..A.R..T......F............V........EV.KKRKD.QQ.T....K
40.Success  61 ME...S..K...IK.....I...W...KK.D.........V..V....Q.T....
13.Failure  61 IE....A.R......K..F..I...W...S.........I...N.Q.T..T..
5.Success   61 ME.I..A.K...IK..F..L...W.......R.........AK....A.T....
58.Failure  61 MT....A.N..T...K..F..L..........V.......EEV.KR.KN.Q..T....K
21.Success  61 ME....A.K.......F..I...W....V.........V..A....Q.T..T..
9.Failure   61 ME....A.R.....F......I...W...KK.K.Q.............V.K...PQT.....
consensus   61 ..*....*.**.**..**.*..***.***..*..******...*..........*....

                                MA (p17)/ CA (p24)
HXB2       121 DTGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGAT
54.Success 121 A..S.-+-...........A.......M.........I..........T.......
36.Failure 120 AAXN.K..............A..X.....S.........I..........D...
20.Success 118 EE.N..P.............A.......T...................
11.Failure 121 SK.N.S..............A......SL.........I..........T.....
17.Success 120 -..T.S..............A......SL.........I..........T.....
43.Failure 121 .E.NN.S.............A......P...................
40.Success 121 A..I.--+............A....T..S.....AI..R..........T.....
13.Failure 121 A..SG-+-.......L..A....TY.SM.......I..G..........A....
5.Success  121 AA.S.-+-....F.......A......SM.........I...........
58.Failure 121 .E.K.SP.............A..............................N...
21.Success 121 AAAAAGSS............A....T..SM.........I..........T.....
9.Failure  121 A..S.SH-............A.......V.........I..........T.....
consensus  121 ........*.***.**.**..****..*...*********..**..*.*******.**..***

HXB2       181 PQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTT
54.Success 178 .....M...I............D...........T...Q....P...............
36.Failure 180 .....M...I............D.......D...T......VP...............
20.Success 178 ......................D...D.......I..PQ....P...I..........
11.Failure 181 ......................D...D.......L..RQ...LP...I...S.G.....
17.Success 179 ......................D...........L..PQ........I..........
43.Failure 181 ......................D...........L..PQ...FP...I.D.T.......
40.Success 178 ....A...I............D...........T...P...............
13.Failure 178 .....M...I............D...........T...P...............
5.Success  178 .....M...I............D...........L...P...............
58.Failure 181 ......................D...........L..RQ...LP...I...S.G....A.
21.Success 181 .....M...I............D...........T......AP...............
9.Failure  180 .....M...I............D...........T......VP...............
consensus  181 *****.***.***********.**.***.***.***.**.*.***..***.*.*.*.****.*
```

157

```
HXB2         241 STLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
54.Success   238 .N.........S...................V..............V................
36.Failure   240 ...........S...V......R................V.....................
20.Success   238 .......R...S.......D..................V.....K................
11.Failure   241 .......T...S.........................V.....K................
17.Success   239 .N.....A...S.......................C.V.....K................
43.Failure   241 .N.....R...S.........................V.....................
40.Success   238 ...........S...................V......V.....................
13.Failure   238 .........I.S...V.....................V.....................
5.Success    238 .N.........S.......D.................V.....K................
58.Failure   241 .......T...S.........................V.....K................
21.Success   241 ...........S...................V......V.....................
9.Failure    240 ..I........S...V........V.....X.....V.....K.........E.....
consensus    241 *..****.*.*.***.***.**.***.******.***.*.*****.********.*****
                                                                            CA
HXB2         301 YKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKA
54.Success   298 F.......T.D.......................R....G..................
36.Failure   300 F.......CT..........M..........S..R....G................S...
20.Success   298 F.......T....G...D...................G..................
11.Failure   301 F.......T....G...D...............R...AG..................
17.Success   299 F...........G...D...................TG.S.............S...
43.Failure   301 F.......T....G...D.................G.N..................
40.Success   298 F.......T...............S..R....G................S...
13.Failure   298 F.......T.D.......................R....G................S...
5.Success    298 F.......T....G...................R....G.S..................
58.Failure   301 F.......T....G...D...................G..................
21.Success   301 F.......T..........M............S....G..................
9.Failure    300 F.......T...............S..R....G................S...
consensus    301 .*.*******..*.**.***.**.*********.**.***..*.**************.**
                 (p24) /p2 (SP1)    p2 (SP1) / NC (p7)
HXB2         361 RVLAEAMSQVT--NSATIMMQRGNFRNQRKIVKCFNCGKEGHTARNCRAPRKKGCWKCGK|
54.Success   358 ...L......H--Q.NI-.I..K...KG.-RRI..............I..............
36.Failure   360 ...L.....AQ--QTNI-.I...G.-RTI.............L......K..........
20.Success   358 ...L.....ASGAAA.A..I..K...KG..RTI.............L......K.........S
11.Failure   361 ...I.........--.ASAA.I..KS..KGP.RNI.............L.............
17.Success   359 ...I.......A.SASA.I-...KS...GP.RNI.............L.............
43.Failure   361 .I......AN--TA.A...KS..KG.RTI.............L.............
40.Success   358 ........Q--QTNV-.I..K...G.-RTI.............L......K....R.....
13.Failure   358 ........KAQ--Q.NV-...K...G.GR.I.............L......K....R......R
5.Success    358 ........Q--QPNI-.I....G.-.TI.............L......K.........R
58.Failure   361 ........ASGAAA.A...K...KG.RTI.............L.............
21.Success   361 .I......AQ--QANIM.I...G.-RTI.............V...K...R......
9.Failure    360 ........Q--QTNI-.I..K...G.-.TI.............L.............R......
consensus    361 *.*.*****.*...........*.*..**.......*********.****.***..*******.
                 NC (p7) /p1 (SP2)    p1 (SP2) /p6
HXB2         419 EGHQMKDC-TERQANFLGKIWPSYKGRPGNFLQSRPEPTAPPEESFRSGVETTTPPQKQE
54.Success   414 ........-...........N.........P..........A...GIX-.EV...P...
36.Failure   416 ........-...........S.........P..........A..CGM.-.EIAS.P...
20.Success   418 ........-.........RL..N.........N.........A..LGV.-.EI...PR..
11.Failure   419 .......-...........N........N.L........A...GF.-.EI..S....
17.Success   418 ....RE.-...........N........P..L......A...GF.-.DIA.SP...
43.Failure   419 .....E.-...........N.........A..G--.EIA.SPR..
40.Success   414 ....R..-...........N........P..........A..LGM.-.EIP.SPQ..
13.Failure   415 ........-...........S.........P..........A..GM.-.EI..SPQ..
5.Success    414 ........-........V..S.........P..........A..CGM.-.EI.S.P...
58.Failure   421 ........-..........R..N........P..........A..LGVR-.EIA.SP...
21.Success   418 ....R..T...........N.........P..........A..GV.-.EIASS....
9.Failure    416 ........-...........N.........P..........A..LGME-GE...SP...
consensus    421 *****..*.*.*********..***.*.******.*.*..******.**............**

HXB2         478 PIDK-ELY-PLT-
54.Success   472 .R..-G..P...X
36.Failure   474 .R.R-G..P...X
20.Success   476 .KE.-...-..A-
11.Failure   477 .RE.-.TP-..A-
17.Success   476 .RE.-.TP-..A-
43.Failure   475 .KE.-.T.-...X
40.Success   472 .R..-G..P...X
13.Failure   473 .R..-...P..AX
5.Success    472 .R..-...-...X
58.Failure   479 .KE.E...-..--
21.Success   477 .R..-GI.P..A-
9.Failure    474 .G..-G..P...X
consensus    481 *.......**..
```

```
 Protease

HXB2         1 PQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
54.Success   1 ..I.........VR.....I.............QDIN.Q.K...................
36.Failure   1 ..I.........VR.E...I.............QIN...K...............K...
20.Success   1 ..I......I..V.....I.............IN...K...................
11.Failure   1 ..I.........VRV....I.............QIN...K...............K...
17.Success   1 ..I.........VRV...II.............QIN...K...............K...
43.Failure   1 ..I.........VR.....I.............IN...K...................
40.Success   1 ..I.........V..E..II.............DIN...K...................
13.Failure   1 ..I.........V......I.............IN.....................
5.Success    1 ..I.........V.....I.............IN...K...................
58.Failure   1 ..I.........V.....I.............IN...K...................
21.Success   1 ..I.........VR..E..V.............Q.I....K...................
9.Failure    1 ..I.........VM....PI.............ID.....................
consensus    1 **.******.**.....*..*************....*.*.***************.***

HXB2        61 QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
54.Success  61 ..P...E.K.......I...........M..........
36.Failure  61 ........K...................M..........
20.Success  61 ..A...S.Q...................M..........
11.Failure  61 ......E.K..V.........I......M..........
17.Success  61 ......E.K...........I......M..........
43.Failure  61 ......E.RR..........I......M..........
40.Success  61 ...M....KR..................M..........
13.Failure  61 ..P....EK...................M..........
5._Success  61 ...M....KR..................M..........
58.Failure  61 ..P...S.K...................M..........
21._Success 61 ..P.....KR......I...........M..........
9.Failure   61 ..P.....K...................M..........
consensus   61 **..**....*.****.****.******.**********
```

Figure 3.5: Full length Gag and protease NGS consensus sequences of 'Baseline Success' and 'Baseline failure' groups

Multiple sequence alignment of test viruses Gag and protease amino acid sequences against reference HXB2. Dots (.) indicate identity and where residues differ from reference, they are denoted. Amino acid deletions are denoted by a dash (-). The 10 Amino acids comprising each of the cleavage sites are shown in the red boxes, with the cleavage site shown as green dashed lines. Amino acids previously reported as associated with PI exposure or resistance are highlighted in <mark>yellow</mark>.
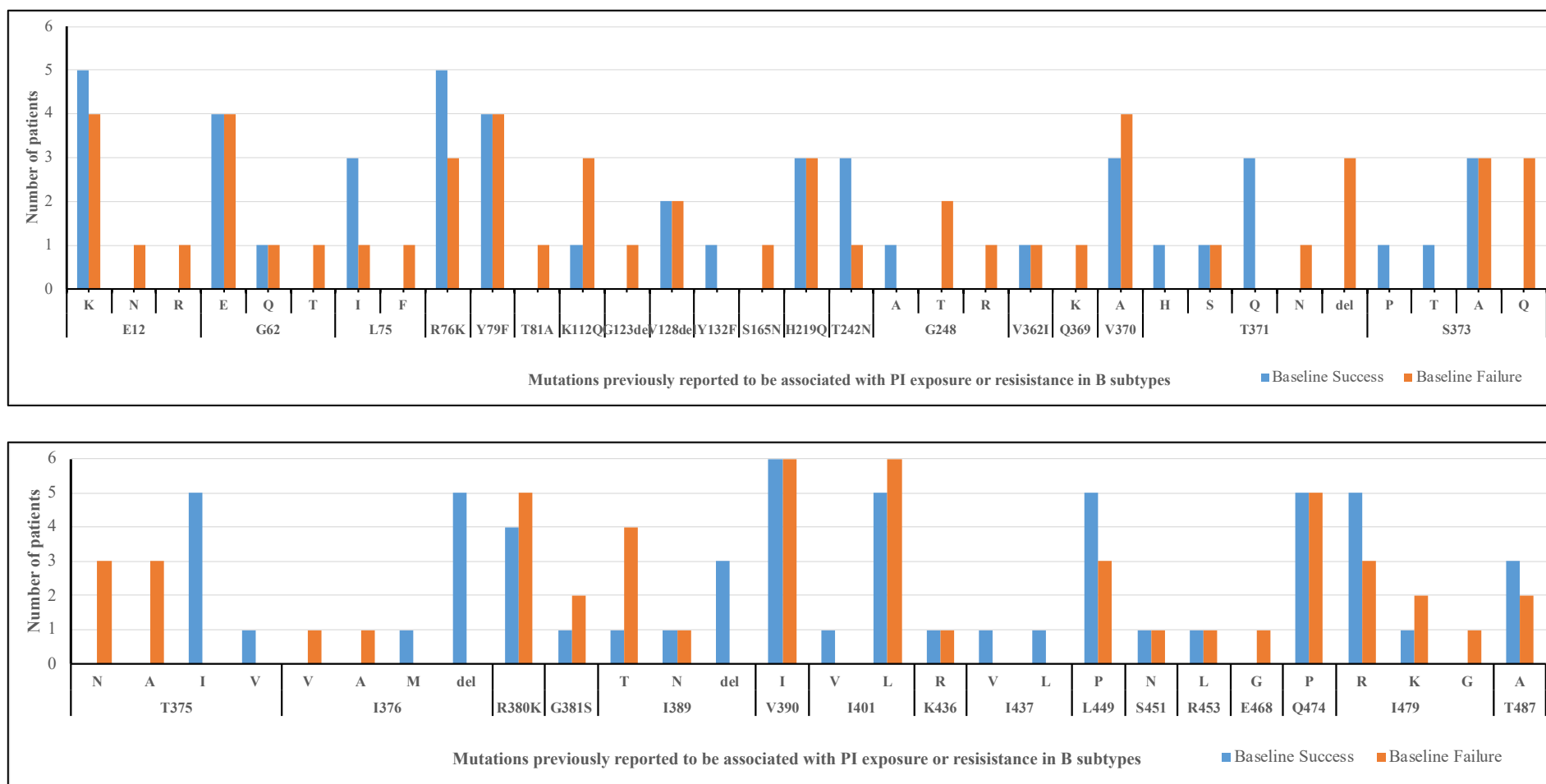
Figure 3.6: Gag Amino acid residues at sites associated with PI-exposure or resistance.

Sites previously associated with PI exposure or failure between the two groups ('Baseline successes' in blue and 'Baseline failures' in orange were examined. Amino acid residues in some of these sites have been previously associated with PI-resistance or exposure in B-subtype viruses

3.4.7 Previously reported Gag polymorphisms/mutations observed in 'Baseline failure' group, but not in 'Baseline success' group

With a view of seeking a relationship between subsequent 2L treatment and outcome, we next examined both groups for the presence of Gag polymorphisms or mutations occurring uniquely in the 'Baseline failure' group but not in the 'Baseline success'. This comparison of the polymorphisms at these sites between 'Baseline success' and 'Baseline failure' groups is shown in figure 3.6 above.

We observed the following changes in Gag among the 'Baseline failure' group only: In the MA, there were 6 polymorphisms (**E12N, E12R, G62T, L75F, T181A** and **G123E**), 2 polymorphisms in CA (**S165N, G248T, G248R**). In the p2 domain, there were 8 polymorphisms (**Q369K, T371N, T371del, S373Q, T375N, T375A, I376V, I376A**) seen in the 'Baseline failure' samples, which were not present in the 'Baseline success' group. In the p6 domain, there were 2 Gag polymorphisms (**E468G** and **I479G**). A summary of these is shown in table 3.5 on the next page.

Table 3.5: Amino acid residues of 'Baseline failure' group only

Gag sites which have previously been reported to be associated with PI resistance or exposure were examined. The residues in this table were only observed amongst the 'Baseline failure', but not the 'Baseline success' group.

Gag domain	Polymorphism (frequency)
MA (p17)	E12N (1/6); E12R (1/6); G62T (1/6); L75F (1/6); T181A (1/6); G123E (1/6)
CA (p24)	S165N (1/6), G248T (2/6); G248R (1/6)
p2 (SP1)	Q369K (1/6), T371N (1/6), T371del (4/6), S373Q (3/6), T375N (3/6), T375A (3/6), I376V (1/6), I376A (1/6)
NC (p7)	None
p1 (SP2)	None
p6	E468G (1/6); I479G (1/6)

3.5 Discussion

In this chapter, patients were selected based on eventual 2L, PI-based treatment outcome. They were divided into two groups: those who went on to suppress viral replication successfully (VL <400cp/mL) and those who did not achieve virological suppression after >48 weeks on PI-based treatment (VL >1,000 cp/mL).

Patients were matched on characteristics such as age, sex, CD4 count, HIV-1 subtype, and RTI drug backbone used during 2L treatment. We collected baseline plasma samples (pre-PI switch) from patients and performed a comparison of NGS consensus sequences from 'Baseline success' and 'Baseline failure' groups in order to assess differences in the prevalence and patterns of mutations in the Gag and protease proteins. We specifically sought for polymorphisms and mutations in both cleavage and non-cleavage site residues of Gag that have previously been reported to be significantly associated with PI-exposure or treatment failure. Previous studies have suggested that there is evidence regarding the effect of Gag cleavage site mutations present at baseline on the outcome of subsequent PI-based therapy[398,407,520].

Neither major nor accessory PI resistance mutations were identified in the consensus viral sequences of the protease of any of the patients. However, several polymorphisms were identified by the Stanford HIVdb to be present in all the viruses across the groups; these were: **I13V**, **M36I**, **L89M** and **K20I** (occurring as **K20V** in one patient from the 'Baseline success' group).

Although the **I13V** polymorphism in protease appeared to be the consensus in the subtypes studied, there has been evidence to show that this polymorphism, occurring with two others (L63P and A71T) convey reduced susceptibility to PIs in the absence of major PI resistance mutations in-vivo[392,521-523], in-vitro[524,525] or from statistical analysis prediction[371].

From previous studies, the **K20I** is the consensus amino acid in subtype G and CRF02_AG viruses. However, in most other subtypes, it is a PI-selected mutation associated with reduced NFV susceptibility[393,526]. The **K20V** polymorphism has almost only been exclusively observed in subtype G viruses. K20M/V are rare,

relatively non-polymorphic PI-selected mutations that have not been well studied[527].

The **M36I** is the consensus amino acid in most of the non-B subtypes. It occurs in about 15% of PI-naive and 35% of PI-experienced individuals with subtype B viruses[378]. In subtype B viruses, M36I increases the replication fitness of viruses with PI-resistance mutations[353,528] and has been reported to reduce in vitro susceptibility to saquinavir[529].

The **L89M** is the consensus in CRF02_AG and G viruses. In subtype F viruses however, it has been reported as having a high genetic barrier to the accumulation of the L90M resistance mutation and can function as a resistance mutation, depending on the presence of other polymorphisms in the subtype F protease backbone[530]. When the **L89M** occurs alongside a **G48T** mutation in protease, it has been shown to confer resistance to other PIs such as amprenavir (APV), indinavir (IDV), ritonavir (RTV), nelfinavir (NFV) and saquinavir (SQV)[531,532]. In our samples however, we did not observe the G48T polymorphism.

Our analyses identified a number of amino acid changes in Gag and polymorphisms in protease that were only present in viruses from the 'Baseline failure' group, correlating with treatment outcome. In protease, the following changes (relative to HXB2 reference) were only identified in viruses from the 'Baseline failure' group: **K14M, L19P, N37D, C68E, H69R** and **I72V**. These 6 protease polymorphisms have not previously been associated with PI failure and the most probable association with subsequent virological failure on 2L, PI-based treatment in this cohort may just be a matter of chance, given the small sample size.

When we examined Gag sites previously reported to be associated with PI failure or exposure, polymorphisms were observed in both cleavage and non-cleavage sites. At the MA-CA cleavage site, the V128 del mutation was observed in patients across both groups. It has been associated with PI exposure *in vivo* and *in vitro* and associated with an increased risk of virological failure following administration of DRV-containing regimens[407,410]. Mutations at position 128 near

the MA/CA cleavage site have been negatively associated with viral response in the ANRS 127 trial and V128I was observed in >10% patients experiencing virological rebound in the POWER 1, 2 and 3 clinical trials[407,520]. In the MA-CA cleavage site, one sample in the 'Baseline success' group was shown to have the **Y132F** mutation. This mutation has been reported to be associated with PI exposure[406,533]. Also, one sample from the 'Baseline failure' group had a **V135L** at the MA-CA cleavage site, which has not been described elsewhere.

The CA-p2 cleavage site, was generally conserved across both groups. The only mutation seen here is the V362I which has previously been reported to associated with *in vitro* resistance to the PI tipranavir[391], and also associated with resistance to the maturation inhibitor, bevirimat[440,534].

In both B and non-B subtypes, the p2/NC cleavage site has been reported as showing the most significant amino acid variations as this region is highly polymorphic[404,414,535], with a statistically significant association between these mutations and the development of high-level PI cross-resistance[404]. The presence of more than two substitutions in p2/NC site at baseline has previously been shown to be significantly associated with virological failure on PI-based monotherapy[398]. Given that Non-B subtype isolates are more likely to harbour more than two substitutions in this specific site, virological failure would therefore significantly more frequent in non-B subtypes. In all the 12 samples tested and across both patient groups, multiple variations in the p2/NC cleavage site was observed. These include: **S373A/T/P, A374N/T/S, I375N/A/I/V, I376A/V/M, M377I, R380K** and **G381S**. A lot of these mutations have previously been associated with *in vivo/in vitro* PI exposure or failure.

The NC/p1 cleavage was conserved, with only the K436R mutation seen in 2 samples, one each from each of the two groups. Additionally, The I437V and I437L was observed in two samples in the 'Baseline success' group. These mutations at Gag 436 and 437 have also been previously shown to be associated with PI exposure *in vivo*[197,384,404,406,414,443,444,536], *in vitro*[197,436] or with PI failure[197,400,443]. **I437V** was associated with reduced virological responses to a

number of PIs in the NARVAL trial and with virological failure in DRV-based regimens[415].

At the p1/p6 cleavage site, a number of mutations were also observed in the consensus sequences of the samples. Previously associated with PI exposure *in-vivo*[379,384,406-408,424,443,444,536-539], *in-vitro*[402,403,409,540] or PI resistance[443] were the **L449P** (observed in 4 sequences from the 'Baseline success' group and in 3 sequences from the 'Baseline failure' group). The **L449F** has been shown to have effects *in vitro* on both viral fitness and phenotypic resistance to PIs[379]. It was also shown in the ANRS 127 trial that the presence of **L449P** was negatively associated with virological response[407]. Mutations at Gag position 452 (seen as **R452L** in 1 sample from the 'Baseline success' group) has been reported in the POWER trials to be associated with reduced response to DRV-based regimens[520]. The **P453L** (seen in 1 sample from the 'Baseline failure' group) has been identified as a natural polymorphism but associated with protease mutations[379,408].

In addition to the cleavage site mutations mentioned above, non-cleavage site mutations which have also been previously found to correlate with PI exposure or resistance were observed in the patient samples across the two groups. These include; **E12K/N/R, G62E/Q/T, L75I/F, R76K, Y79F, T81A, K112Q** in the matrix domain. Others are the **H219Q** in the capsid, **Q369K, V370A, T371H/S/Q/N/del** in the p2, **I389T/N/del, V390I, I401V/L** in the NC, and **E468G, Q474P** in the p6.

The Gag **E12K** mutation has previously been reported following *in vitro* PI exposure in B-subtypes[418], it however, appears to be the consensus sequence in CRF02_AG and G subtypes. The **R76K** was present in 5 and 4 samples from the 'Baseline success' and 'Baseline failure' groups respectively. This mutation is the consensus sequence in the CRF02_AG reference virus (L39106.ibNG). The **Y79F** was seen in 4 patient consensus sequences from each group while, the **T81A** was only seen in 1 patient sequence who subsequently failed 2L PI-based treatment. Together, these three mutations in MA (**R76K, Y79F and T81A**) have been shown to contribute to the reduced susceptibility to the PI, lopinavir[417]. The observed mutations in CA, NC and p6 domains have all been previously

associated with PI exposure in B subtypes. For example, **I389T** and **I401V** have been associated with treatment failure, however the mechanism by which they affect treatment outcome is unknown[424]. The p2 mutations observed in the patient sequences across both groups (**Q369K, V370A, T371H/S/Q/N/del**) in addition to being associated with PI exposure *in vivo*[424,441], have also been associated with maturation inhibitor resistance[300,440,534,541-543].

In this chapter, we aimed to analyse consensus Gag and protease sequences in order to determine significant associations between baseline Gag mutations and treatment outcome, possibly also to identify novel pathways of PI resistance. Our analyses identified a number of amino acid changes in Gag and polymorphisms in protease that were only present in viruses from the 'Baseline failure' group, correlating with treatment outcome. In protease, the following changes (relative to HXB2 reference) were only identified in viruses from the 'Baseline failure' group: **K14M, L19P, N37D, C68E, H69R** and **I72V**. Also, the following Gag amino acid mutations (also relative to HXB2 reference) were only identified in viruses from the 'Baseline failure' group: **E12N/R, G62T, L75F, T81A, G123E, S165N, G248T/R, Q369K, T371N, T371del, S373Q, T375N, T375A, I376V, I376A, E468G** and **I479G**. These were not seen in any sequences in the paired 'Baseline success' group. These 5 protease polymorphisms and 19 Gag mutations may be associated with treatment outcome, given that they were only seen in the 'Baseline failure' group. However, it is well documented that significant levels of variability in Gag between patients infected with the same subtype exist, therefore it is possible that these amino acid positions were only seen in the treatment failure group solely by chance.

We conclude that mutations previously shown to be selected under PI –selective pressure in subtype B HIV-1 can occur as natural polymorphisms in PI-naïve CRF02_AG and G viruses. We however cannot yet conclude on the particular Gag polymorphisms or mutations that we can correlate with treatment outcome. In order to elucidate further any relationship between baseline virogical sequence data and subsequent treatment outcome, we carried out phenotypic drug (PI)

susceptibility assays on the samples. These assays are described, and results presented in the next chapter.

CHAPTER 4: Phenotypic comparison of baseline viruses from patients who succeeded on, and those who failed Protease inhibitor-based second-line treatment

4.1   Introduction

Our goal is to determine whether baseline drug resistance assays could help to predict treatment failure with 2L protease inhibitor-based ART in a Nigerian cohort infected with HIV-1 CRF02_AG and G subtypes using carefully selected patients from a large treatment program. We wanted to assess the feasibility of using baseline drug resistance data as predictors of response to antiretroviral therapy in a clinical setting and use the information as a prognostic indicator of therapy response and as a treatment management tool.

In chapter 3, we examined full-length Gag and protease NGS consensus sequences of patients in a case-control ('Baseline failures' vs 'Baseline success') manner. We wanted to explore the role of Gag mutations (polymorphisms) which could be driving PI failure. We examined both cleavage and non-cleavage sites of patient-derived consensus sequences of both groups and could not conclude on mutations or polymorphisms unique to either of the two groups.

In this chapter, we cloned full-length patient-derived*, Gag-protease* sequences from chapter 3 into a vector and carried out the clonal analysis of the sequences. We employed *in vitro* single cycle recombinant assays to characterize the viruses in terms of drug susceptibility and replicative capacity (RC). Results are expressed as the n-fold difference (FC) in the EC_{50} compared to the wild-type (WT) reference virus (as reviewed by Sebastian and Faruki) [544].

In this chapter, the single-cycle recombination assay system developed and previously described by Petropoulos[457] was modified and used in a three vector assay system to characterize drug susceptibility and replicative capacity (RC) of

virus variants from chapter 3. This assay (and its modified versions) have been used in a number of peer-reviewed, publications[397,417,420,505,507,508,545,546]. Additionally, the assay also controlled for variation in infectivity and transfection efficiency between viruses as each EC_{50} value was normalised using a no-drug control luciferase reading for each virus. The single-round replication assay is, therefore, suitable for use in this chapter. Multiple rounds of infection experiments were carried out subsequently and would be discussed in chapter six.

4.2   Chapter Aims

1) Determine PI susceptibility and single-round infectivity of representative viral variants (clones) from each patient
2) Compare the PI susceptibility and single-round infectivity of baseline variants from 'Baseline successes' and 'Baseline failure' patients

4.3   Methods

4.3.1 Patient samples

The 'Baseline success' and 'Baseline failure' samples discussed in chapter 3 were used for the assays in this chapter.

4.3.2 Generation of resistance test vectors

In order to phenotypically assess the PI susceptibility of patient-derived full-length Gag-protease viruses, it was necessary to determine a method by which the *quasi-species* within a patient could be represented in the assay. One option was to test the phenotypic susceptibility of all the variants (clones) in the pool. We, however, did not do that because that would complicate the investigation of specific regions or changes contributing to any observed changes in susceptibility.

We decided, therefore, to select the viral clone, which most resemble the NGS-derived consensus to represent the viruses derived from each patient in the phenotypic susceptibility assay.

HIV-1 RNA was extracted, followed by cDNA synthesis and nested PCR to amplify full-length Gag-protease. The PCR product was cloned into *pGEM* intermediate vector and transformed into JM109 competent cells. Colony PCR and restriction digestion were used to identify clones containing full-length, patient Gag-protease sequences. These positively identified colonies were cultured, and plasmid DNA purified (as detailed in chapter 2, section 2.1.8). The purified plasmid DNA was then sequenced using Sanger sequencing. Full-length *Gag-protease* from the selected variants was cloned from the *pGEM* intermediate vector into the reference strain p8.9NSX+, using the *NotI* and *XhoI* restriction sites.

4.3.4 Drug susceptibility testing

PI susceptibility testing was performed, as indicated in chapter 2, section 2.2.5. Briefly, pseudovirus stocks were produced by co-transfecting confluent human embryonic kidney 293 (HEK293T) cells with a resistance test vector (RTV) DNA plasmid containing patient-derived HIV sequences (obtained as described in section 4.3.2 above); pMDG (encoding the vesicular stomatitis virus G protein) and pCSFLW (encoding the firefly luciferase and the HIV packaging sequence). Cells were harvested 16 hours after transfection and seeded in the presence of different PI concentrations. Three second-generation and widely used PIs were measured: atazanavir (ATV) and lopinavir (LPV), and darunavir (DRV. Pseudovirus stocks produced in the presence of PIs were harvested approximately 24 hours later and used to infect fresh target HEK293T or TZM-bl cells. Infection was monitored by measuring luciferase production in infected target cells 48 hours post-infection and compared to a control in the absence of the drug. The EC_{50} was calculated by plotting the percentage luciferase inhibition vs log_{10} drug concentration and using GraphPad Prism v8.0 (GraphPad Software Inc., La Jolla, CA, USA) to fit the inhibition curve by nonlinear least-squares analysis. Results were expressed as fold difference (FC) in the EC_{50} compared to the wild-type subtype B HIV-1 reference (P8.9NSX+). Experiments were done in duplicate and the calculated EC_{50} represented the mean of at least two independent experiments.

4.3.5 Replicative capacity testing

Pseudoviruses stocks were produced as described in 4.3.3 above, by transfecting HEK-293T cells with the three plasmids (retroviral vector, pMDG, and pCSFLW). Replicative capacity of these viruses was assessed by comparing the luciferase activity of recombinant virus with that of the WT subtype B control virus in the absence of drug.

4.4 Validating the Phenotypic susceptibility assay

Having selected the viral clones from each patient to be utilized for phenotypic susceptibility testing as described in 4.3.3 above, the next step was to carry out drug susceptibility assays.

Although the phenotypic assay methods employed have previously been validated and published in peer-reviewed journals, we wanted to re-validate the susceptibility assay to determine if it is 'fit for purpose' for our study. First, CRF02_AG and G virus clones from two PI-treated Nigerian patients (#37 and #57) with protease inhibitor resistance-associated mutations (as predicted by the Stanford HIVdb Genotypic Resistance Interpretation Algorithm) in the protease gene was phenotypically assayed. Sample #37 had the following major PI mutations in the protease gene: M46I, I54V and L76V while sample #57 had M46I, I54V, L76V, V82A and K43T. A CRF02_AG virus sample (#12) with no protease inhibitor resistance-associated mutations were assayed alongside to serve as a control.

As can be seen from figure 4.1 (below), the viral clones with major protease mutations were highly resistant to lopinavir, with fold difference in EC_{50} values of 258-fold and 480-fold for #37 and #57 respectively.

The observed phenotypic susceptibility pattern of these viruses correlated well with the HIVdb Genotypic Resistance Interpretation Algorithm PI resistance prediction (Table 4.1, see below). Therefore, giving us confidence that the choice of assay method is appropriate.
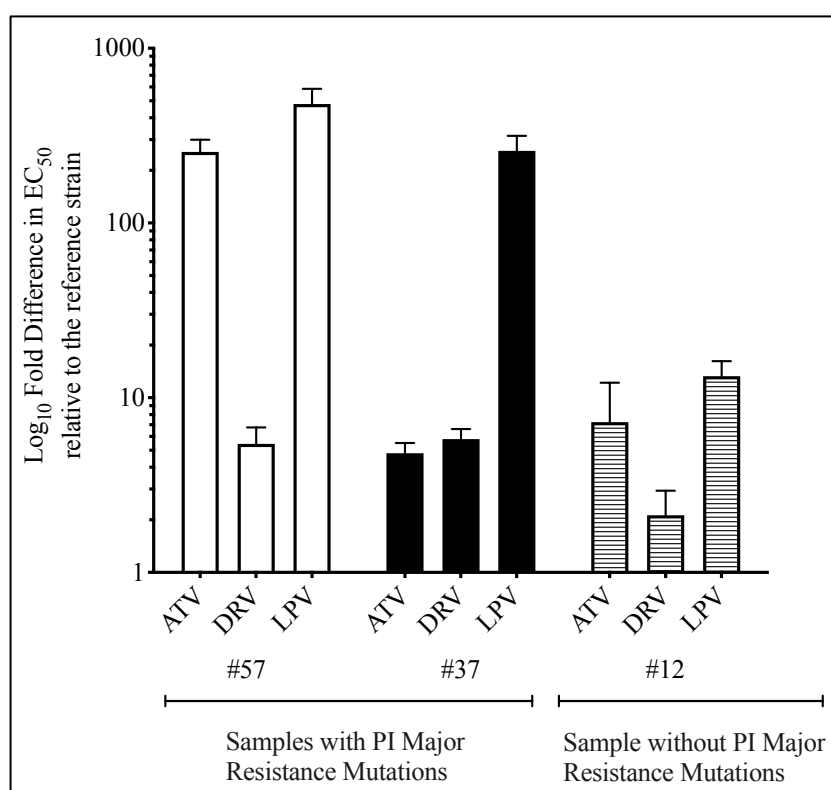
Figure 4.1: Validating phenotypic assay method
Phenotypic susceptibility testing of two viral constructs (#37 and #57) with major PI resistance mutations in protease and one viral construct with no major PI-associated mutations in protease. PI susceptibility was determined using EC_{50} data for pseudovirions produced by co-transfection of 293T cells, derived from full-length Gag-protease of patients. Fold change in EC_{50} relative to p8.9NSX+ reference strain is presented. Each bar represents the mean of two independent experiments performed in duplicate. Error bars represent SEM. LPV= lopinavir; ATV = atazanavir; DRV = darunavir.

Table 4.1 Comparison of HIVdb protease inhibitor susceptibility predictions with phenotypic assay results

Sample ID (PI Resistance mutations)		ATV	DRV	LPV
#37 (M46I, I54V and L76V)	HIVdb Predicted resistance	Low-Level Resistance	Low-Level Resistance	High-Level Resistance
	Phenotypic drug susceptibility assay	Borderline Susceptible (FC = 4)	Reduced Susceptibility (FC = 5)	Extremely low susceptibility **(FC = 258)**
#57 (M46I, I54V, L76V, V82A and K43T*)	HIVdb Predicted resistance	High-Level Resistance	Low-Level Resistance	High-Level Resistance
	Phenotypic drug susceptibility assay	Very low susceptibility **(FC = 250)**	Reduced susceptibility (FC = 5.8)	Extremely low susceptibility **(FC = 480)**
#12 (None)	HIVdb Predicted resistance	Susceptible	Susceptible	Susceptible
	Phenotypic drug susceptibility assay	Reduced susceptibility (FC = 7.2)	Susceptible (FC = 2.1)	Reduced susceptibility (FC = 13.2)

A fold change in IC50 of >4 is described as reduced susceptibility in the phenotypic assay

FC, Fold difference EC_{50}. ATV, atazanavir DRV, darunavir LPV, lopinavir

* PI-accessory mutation

Additionally, we wanted to ensure that any phenotypic changes observed is from the *Gag-protease* of the patient-derived virus and not from the vector. To achieve this, we performed phenotypic susceptibility testing of the viral constructs against the reverse transcriptase inhibitor zidovudine AZT. Rationale was that since the RT of all constructs is the same, there will be no significant difference in phenotypic susceptibility between the constructs.

As seen in figure 4.4, the FC in EC_{50} for all constructs was less than 2-fold, for AZT, suggestive of high susceptibility to RT inhibitors. We were therefore confident to proceed with the phenotypic susceptibility assays as any changes seen can be attributed to full length Gag-protease from the patient and not from the RT.
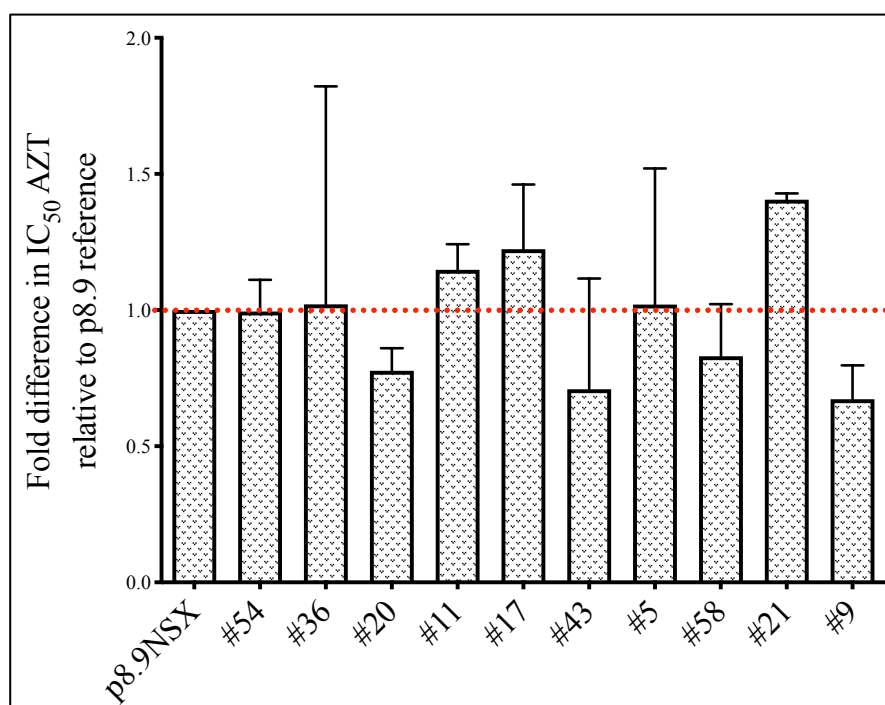
Figure 4.2: RTI Drug susceptibility of Gag-protease constructs
Mean fold difference in EC_{50} for the panel of viruses against zidovudine, a Nucleoside reverse transcriptase inhibitor. Susceptibility was determined using EC50 data for pseudovirions produced by co-transfection of 293T cells, derived from full-length *Gag-protease* of patients. All the samples were the same in the RT region. Each bar represents the mean of three independent experiments performed in duplicate. Error bars represent SEM. AZT= zidovudine

Finally, we wanted to determine if the amount of input DNA used for transfection has any effect on the phenotypic assay results. This was investigated by using twice the amount of DNA for transfection and proceeding with the experiment as usual. We used four CRF02_AG retroviral vectors (VF1A, VF1B, VF1C, VF1D) for this experiment. There was no difference between the 300ng DNA input and the 600ng DNA input for each of the four viral constructs tested (as shown in figure 4.3).
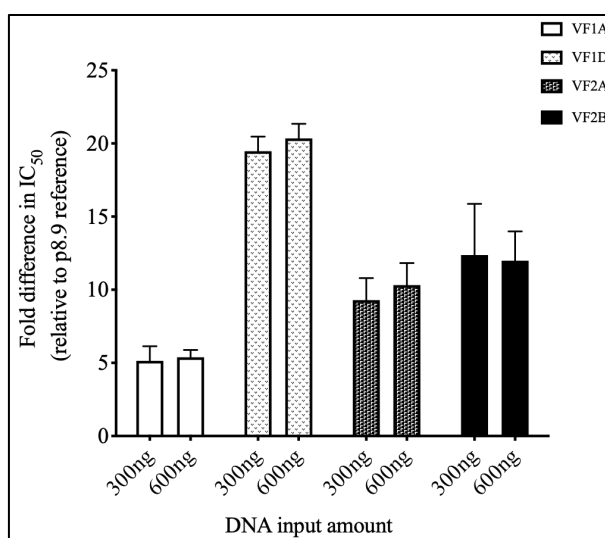
Figure 4.3: Comparison of the effect of amount of input DNA on susceptibility results
We investigated the effect (if any) of the amount of input DNA used for transfection. Doubling the amount of DNA did not have a significant effect on fold difference in EC_{50} of lopinavir. Results shown are the mean of 3 independent experiments carried out in duplicates. Error bars represent Standard error of the mean.

4.3.2.1   Naming viral variants (clones)

The nomenclature employed for naming the viral variants (clones) in this chapter was; Sample ID.Subtype_Colony Number (for example, 40.AG_c8 would mean the 8th colony (viral variant) of sample 40, which is a CRF02_AG virus.
In order to identify protease mutations, the protease sequences were submitted to the Stanford HIVdb Resistance Algorithm to be analysed and also compared to the IAS list of mutations[495,496]. Gag mutations and polymorphisms of importance were identified by manually comparing gag sequences to previously published mutations reported to be associated with both *in vivo* and *in vitro* PI resistance and/or PI exposure (Table 1.5, chapter 1).

4.4   Results

4.4.1   Selection of Clonal variants for phenotyping

Phylogenetic analysis by maximum likelihood method was used to reconstruct phylogenetic trees of clonal variants from each patient. Also included was the the NGS consensus *Gag-protease* sequence. Nucleotide sequences were aligned in MEGA X [547] using the MUSCLE algorithm and imported into the PHYLIP program for phylogeny construction using the Maximum Likelihood method under the GTR model of nucleotide substitution. Maximum likelihood trees were constructed using PhyML 3.0 software with the confidence of the tree tested using 1000 bootstrap replicates. Phylogenetic trees were viewed using FigTree v1.4.4 and MEGA X software.

From each patient, one viral variant (clone) identical, or most similar to the NGS-derived consensus amino acid sequence was selected. The most similar variant was defined as the variant with the least number of non-conserved amino acid changes from the NGS-derived consensus sequence. The selected variant (clone) - marked in red lettering in Figure 4.4 was used for the drug susceptibility and replicative capacity assays described in the following sections.
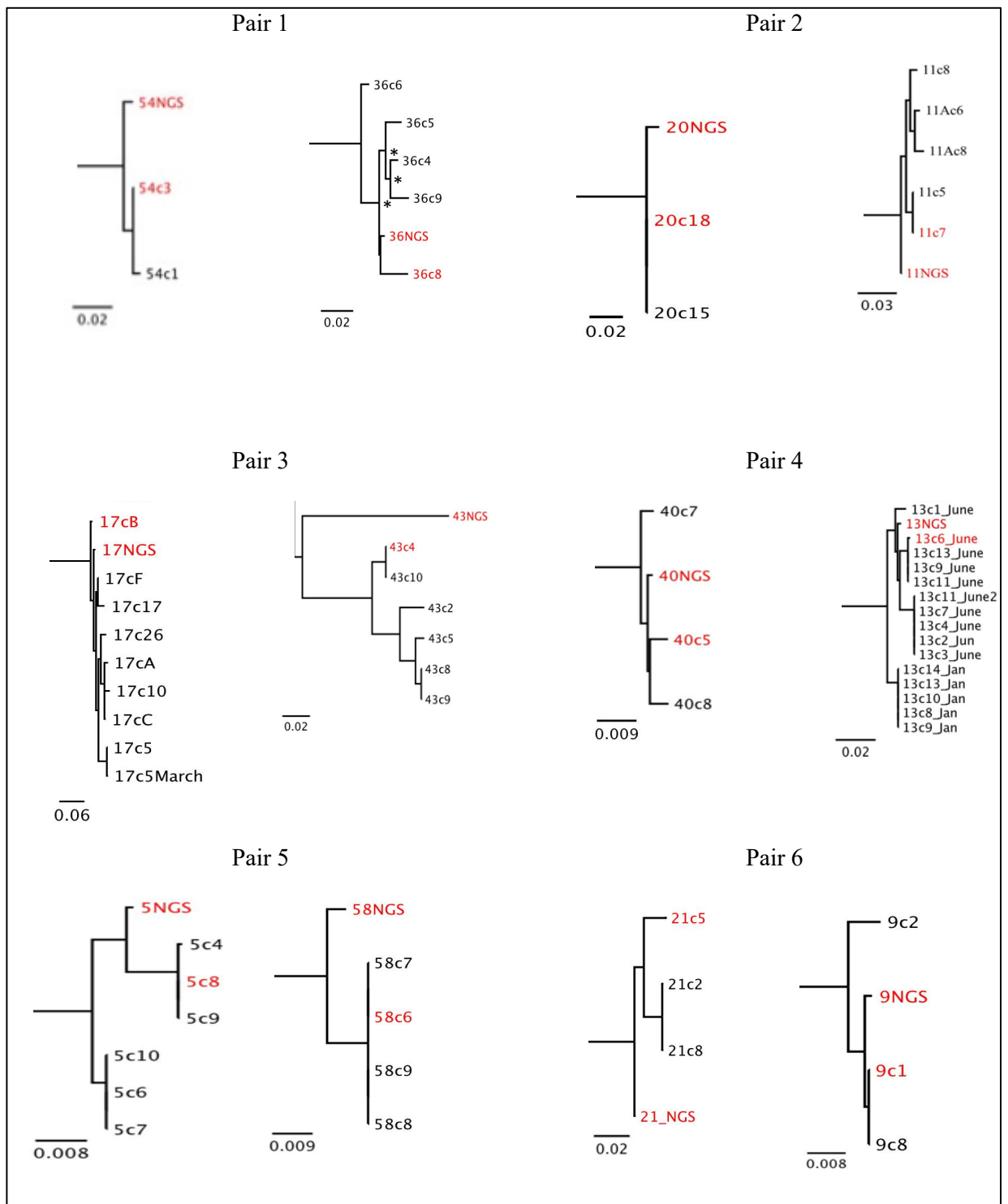
Figure 4.4: Phylogenetic trees of patient-derived clones
Phylogeny construction using the Maximum Likelihood method under the GTR model of amino acid substitution. Maximum likelihood trees were constructed using PhyML 3.0 software with the confidence of the tree tested using 1000 bootstrap replicates. Clonal variants most similar to the NGS consensus, which were selected for phenotypic assays are shown in red.

4.4.1 Large variation in Gag amino acid sequences of patient-derived clonal variants

Once we had selected the patient-derived viral clones to be used for phenotypic drug assays, we performed multiple sequence alignment of Gag and protease sequences (Figure 4.5) to determine if there are any amino mutations (or polymorphisms) that are particularly present in one group, but not the other.

First, we studied the cleavage sites for any previously reported resistance-conferring mutations. Within the MA/CA cleavage site, one sample from each of the groups had an amino acid mutation. These are Y132F (sample #5) from the 'Baseline Success' group and V135L (sample #13) from the 'Baseline Failure' group. At the CA/p2 cleavage site, there was a V362I (sample #21) mutation in one sample from the 'Baseline Success' group, while there was no mutation at this position in any sample from the 'Baseline Failure' group. The p2/NC cleavage site showed the most variability, with 22 variable sites in the 'Baseline Success' group, and 20 variable sites in the 'Baseline Failure' group (as shown in Figures 4.5 and 4.6). Within the NC/p1 cleavage site, there were three (3) and one (1) site variabilities in the 'Baseline Success' and 'Baseline Failure' groups respectively. Finally, at the p1/p6 cleavage site, there were 7 and 6 variable sites between the 'Baseline Success' and 'Baseline Failure' groups respectively. Taken together, there were no cleavage site mutations which are peculiar to one group, and not the other group.

At the non-cleavage sites, the variability across the entire gene was almost equally distributed across both the 'Baseline Success' and 'Baseline failure' groups as seen in Figure 4.5 and Figure 4.6. Additionally, when we examined sites that have previously been reported as resistance-conferring or associated with PI-exposure (highlighted in yellow in Figure 4.5 and Figure 4.6), we also note that these were similar between both groups.

Gag

```
HXB2          1 MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI
54.AG_c3      1 ...........K..S..............R...L.............L.......SA...Q.L
20.G_c18      1 ...........K..A................................L.H.....A...V..
17.G_cB       1 ...........K..S..............R...L.............L.......E...Q..
40.AG_c5      1 ......I.R....A..R..........R...L.............L.......T...Q.L
5.AG_c8       1 ...........K..Q..............R...L.............L.......SA...Q.L
21.AG_c5      1 ...........K..A........M.....R...L............L..S....AG..QEL
consensus     1 ******.*.**.**.**.*****.*****.***.***********.*..***...**...

HXB2         61 LGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAA
54.AG_c3     61 ME...ST.K.........F..L...W......D........E.V..A.KR...PT.....
20.G_c18     61 MQ..K.A....T...K..F............K..V........EV.KM.---.S.T....R
17.G_cB      61 M.....A....T..IK..............K...V........EV.KI.K..QQET.K..-
40.AG_c5     61 ME...S..K.....IK....I...W...KK.D.........V..V.....Q.T.....
5.AG_c8      61 ME.I..A.K.....IK..F..L...W.......R...........AK....A.T.....
21.AG_c5     61 ME....A.K........F..I...W......A........V..A.....Q.T.....
consensus    61 ..*....*.**.**..**.**.***.***..*...******...*..........*.**.

HXB2        121 DTGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGAT
54.AG_c3    121 A..S.---.........A........M..............I.............T......
20.G_c18    118 EE.N..P.........A........T..................................
17.G_cB     120 -..T.S..........A......SL.........I.............T......
40.AG_c5    121 A..I.---.........A....T..S...........AI..R...........T......
5.AG_c8     121 AA.S.---...F.....A....SM..............I..................
21.AG_c5    121 AA.S.---.........A....T..SM...........I.............T......
consensus   121 ..*.*...***.*****.****.**...**********..**.**********.*******

HXB2        181 PQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTT
54.AG_c3    178 .....M...I.............D...........T...Q....P...............
20.G_c18    178 ....................D...D.......I..PQ....P...I..............
17.G_cB     179 ....................D...........L..PQ........I..............
40.AG_c5    178 .....A...I.............D.........T.......P..................
5.AG_c8     178 .....M...I.............D.........L.......P..................
21.AG_c5    178 .....M...I.............D...........T.......AP...I...........
consensus   181 *****.***.************.***.*******.**..***..***.************

HXB2        241 STLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
54.AG_c3    238 .N.........S................V.............V..................
20.G_c18    238 .......R...S.......D................V.....K..............
17.G_cB     239 .N.....A...S...................C.V.....K..............
40.AG_c5    238 ...........S................V.............V..............
5.AG_c8     238 .N.........S.......D..............V.....K..............
21.AG_c5    238 ...........S................V.............V..............
consensus   241 *.*****.***.*******.******.**********.*.*****.*************

HXB2        301 YKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKA
54.AG_c3    298 F........T.D....................R....G...................
20.G_c18    298 F.......T....G...D..................G...................
17.G_cB     299 F............G...D..................TG.S..............S...
40.AG_c5    298 F.......T.................S..R...G..............S...
5.AG_c8     298 F.......T...G...............R....G.S..................
21.AG_c5    298 F.......T.........M.........S.......G...................
consensus   301 .********.*.**.***.**.*********.**.***..*.************.***
```

```
HXB2        361 RVLAEAMSQVT--NSATI-MMQRGNFRNQRKIVKCFNCGKEGHTARNCRAPRKKGCWKCG
54.AG_c3    358 ..........H---QSN.-...K...KG..R-I.........L................
20.G_c18    358 ........ASGAAA.A.-...K...KG..RTI.........L....K...........
17.G_cB     359 ........A.-SA..A.-...KS...GP.RNI.........L................
40.AG_c5    358 .........Q---QTNV-...K...KG..T-I.........L....K....R......
5.AG_c8     358 .........Q---QPN.-..I....G.KT-I.........L....K...........
21.AG_c5    358 .I.......AQ---Q.N.M........G..T-I.........V....K...R......
consensus   361 *.*******.........*.*..**......**********.****.***..******

HXB2        418 KEGHQMKDCTERQANFLGKIWPSYKGRPGNFLQSRPEPTAPPEESFRSGVETTTPPQKQE
54.AG_c3    413 ........................N.......P..........A...GIA-.EI...P...
20.G_c18    417 ................RL...N.........N.........A..LGV.-.EI...PR..
17.G_cB     417 ......E..............N.......P...L......A...GF.-.DIA.SP...
40.AG_c5    413 .....R..............N.......P..........A..LGM.-.EIP.SPQ..
5.AG_c8     413 R..................V...S.......P..........A..CGM.-.EI.S.P...
21.AG_c5    414 ......R..............N.......P..........A...GV.-.EIASS....
consensus   421 .*****..**********..***.*******.*.*.******.**.....*.......**

HXB2        478 PIDKELY-PL--
54.AG_c3    472 .R..G..P..TX
20.G_c18    476 .KE....-..A-
17.G_cB     476 .RE..TP-..A-
40.AG_c5    472 .R..G..P..TX
5.AG_c8     472 .R.....-..T-
21.AG_c5    473 .R..GI.P..AX
consensus   481 *..*....**..
```

Protease

```
HXB2          1 PQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
54.AG_c3      1 ..I.........VR.....I.............QDIN.Q.K.................
20.G_c18      1 ..I......I..V......I............IN...K...................
17.G_cB       1 ..I........VRV...II.............QIN...K.................K...
40.AG_c5      1 ..I.........V..E..II.............DIN...K.................
5.AG_c8       1 ..I.........V......I.............DIN...K.................
21.AG_c5      1 ..I.........VR..E..V............Q.I....K.................
consensus     1 **.******.**.....*..*************....*.*.***************.***

HXB2         61 QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
54.AG_c3     61 ..P...E.K......I..........M.........
20.G_c18     61 ..A...S.Q................M.........
17.G_cB      61 ......E.K...........I......M.........
40.AG_c5     61 ...M....KR...............M.........
5.AG_c8      61 ..FM....KR...............M.........
21.AG_c5     61 ........KR...I..I..........M.........
consensus    61 **..**.*..***.**.****.******.**********
```

Figure 4.5: Sequence Alignment of Clones from the 'Baseline success' group

Gag and Protease sequence alignment of the patient-derived virus of each patient most similar to the NGS consensus sequence in the 'Baseline Success' group showing all amino acid variations relative to HXB2. Sites previously associated with PI failure or PI exposure are highlighted in yellow. The 5 Gag cleavage sites are shown in red boxes. The clonal variants shown in this figure were used for drug susceptibility and replicative capacity assays.

Gag

```
HXB2          1 MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI
36.AG_c8      1 ...............Q..............R...L.............L.......A...Q.L
11.G_c7       1 ..........K..A............Q.RI..L.............L.......A...Q..
43.G_c4       1 ..........K..A..........R.R.....I...K..D.............A...Q..
13.AG_c6      1 ..........K..S.............R...L.............L..S....A...Q.A
58.G_c6       1 .........R..A.........................L..D....A...A.L
9.AG_c1       1 ........T..N..A............R...L.............L.......G...Q.L
consensus     1 ********.**.**.**********.*.*..**..***.**.***.**.****.***.*.

HXB2         61 LGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAA
36.AG_c8     61 IE...TT.K.........F.AIV..W......D.............V.K.G.QDT.....
11.G_c7      61 MQ....A.K..T..IK.................GV.........E..KI.K..QQEN.K.EV
43.G_c4      61 MK....A.K..T..VK..F.............DV.........M....QQET.K..T
13.AG_c6     61 IE....A.R......K..F..I...W...S.............I...N.Q.T.....
58.G_c6      61 MT....A.N..T...K..F..L.........V.......EEV.KR.KN.Q..T....K
9.AG_c1      61 ME....A.R.....F......I...W...KK.K.Q.........V.K...PQT.....
consensus    61 ..***..*.**.**..**.*...**.***..*...*****...*..*.......*.*..

HXB2        121 DTGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGAT
36.AG_c8    121 .---.K...........A.......S..............I................D...
11.G_c7     121 SK.N.S..........A.......SL..............I.............T......
43.G_c4     121 ...N.R..........A.......SM..............I....................
13.AG_c6    121 A..GG---......L..A....TY.SM..............I...G...........A....
58.G_c6     121 .E.K.SP........A.........................N................
9.AG_c1     121 A..S.SH-.........A.......V..............I.............T......
consensus   121 ........******.**.****..*..**********.***.*.*******.**..***

HXB2        181 PQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTT
36.AG_c8    178 .....M...I............D.......D...T..P....VP................
11.G_c7     181 .................D...D.......L..RQ...LP...I...S.G......
43.G_c4     181 .....M...I...........D.........L..AQ...FP................
13.AG_c6    178 .....M...I..............D........T.......P................
58.G_c6     181 ....................D...........L..RQ...LP...I...S.G....A.
9.AG_c1     180 .....M...I...........D............T.......VP................
consensus   181 *****.***.************.***.***.***.**..***..***.***.*.****.*

HXB2        241 STLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
36.AG_c8    238 ...........S...V.....R................V..................
11.G_c7     241 .......T...S.........................V.....K.............
43.G_c4     241 ......V...S..........R..............V.....K.............
13.AG_c6    238 ........I.S...V......................V..................
58.G_c6     241 .......T...S.........................V.....K.............
9.AG_c1     240 ..I.......S...V.........V............V.....K........E.....
consensus   241 **.***..*.*.***.******.***.************.*****.********.*****

HXB2        301 YKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKA
36.AG_c8    298 F.......CT...........M.........S..R....G.................S...
11.G_c7     301 F........T....G...D.............R...AG...................
43.G_c4     301 F........T....G.............S..R...SG...................
13.AG_c6    298 F........T.D...................R....G.................S...
58.G_c6     301 F........T....G...D............R....G...................
9.AG_c1     300 F........T....................S..R....G.................S...
consensus   301 .*******..*.**.***.**.*********.**.***..****************.***
```

```
HXB2        361 RVLAEAMSQVTN--SATIMMQRGNFRNQRKIVKCFNCGKEGHTARNCRAPRKKGCWKCGK
36.AG_c8    358 .........AQQ---TN.........G.-RTI..........L....K............
11.G_c7     361 ............-A..AM...KS..KGP.RNI........L..................
43.G_c4     361 ........QQ---SN....K....G.-RTI.........V...................
13.AG_c6    358 ........KAQQ---SNV...K....G.GR.I..........L....K....R......R
58.G_c6     361 ........ASGAAA.A....K...KG..RTI.........L.................
9.AG_c1     360 ..........QQ---TNV...K....G.-.TI..........L.........R......
consensus   361 ********..........***..**.......**********.****.****.******.

HXB2        419 EGHQMKDCT-ERQANFLGKIWPSYKGRPGNFLQSRPEPTAPPE-----ESFRSGVETTTP
36.AG_c8    414 .........-.............S.......P....K.....A-----..CGM.-.EIAS
11.G_c7     420 .........-.............N.........N.L......A-----...GF.-.EI..
43.G_c4     417 .........N.............H.......P...........PTAPPAESFGVG.EIAS
13.AG_c6    415 .........-.............S.......P..........A-----...GM.-.EI..
58.G_c6     421 .........-.......R....S...................A-----..LGVR-.EMA.
9.AG_c1     416 .........-.............C.......P..........A-----..LGME-GE...
consensus   421 *********.********.****.*******.*.*..*****.................

HXB2        473 PQKQEPIDKE-LYPLT--
36.AG_c8    467 .P....RE..-...PLAX
11.G_c7     473 S.....RE..-TP..A--
43.G_c4     477 .P..G.R...P...PL--
13.AG_c6    468 SPQ...R...-...PLAX
58.G_c6     474 SP....KE..E.....--
9.AG_c1     469 SP....G..G-...PLTX
consensus   481 ...*.*..*....*....
```

Protease

```
HXB2        1  PQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
36.AG_c8    1  ..I.........V..E...I...............QIN...K...............K...
11.G_c7     1  ..I.........VRV.E..I...............QIN...K...............K...
43.G_c4     1  ..I.........V......I...............KIN...K..................
13.AG_c6    1  ..I.........V......I...............IN.....................
58.G_c6     1  ..I.........V......I...............IN...K...................
9.AG_c1     1  ..I.........VM....PI...............ID.......................
consensus   1  **.*********.....*..*************...***.***************.***

HXB2        61 QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
36.AG_c8    61 ........K....................M..........
11.G_c7     61 ......E.K..V.........I......M..........
43.G_c4     61 .......K....................M..........
13.AG_c6    61 ..P....EK...................M..........
58.G_c6     61 ..P...S.K...................M..........
9.AG_c1     61 ..P.....K...................M..........
consensus   61 **.***...**.*********.******.**********
```

Figure 4.6: Sequence Alignment of Clones from the 'Baseline failure' group

Gag and Protease sequence alignment of the patient-derived virus of each patient most similar to the NGS consensus sequence in the 'Baseline failure' group showing all amino acid variations relative to HXB2. Sites previously associated with PI failure or PI exposure are <mark>highlighted in yellow</mark>. The 5 Gag cleavage sites are shown in red boxes. The clonal variants shown in this figure were used for drug susceptibility and replicative capacity assays.

4.5   Full-length Gag-Protease viruses have a wide range of susceptibilities to PI

The 12 selected viral clones from the six pairs were tested against the following PIs: atazanavir (ATV), darunavir (DRV) and lopinavir (LPV). Fold difference relative to subtype B is shown in Figure 4.7. We observed a wide range of PI susceptibilities of the full-length Gag-protease constructs.

Susceptibility to ATV was widely varied with the most susceptible virus showing FC in EC_{50} of 1.3 and the least susceptible virus showing a 26-fold difference in EC_{50} of ATV. Of the 12 virus constructs tested phenotypically, eight of the viruses had greater than 2-fold difference in EC_{50} of ATV with four virus constructs showing less than 2-fold difference. When we examined susceptibility to LPV, all 12 viruses had FC > 2, with one virus having a fold difference > 10. Overall, the viruses showed the most susceptibility to DRV where nine (of the twelve) viral constructs showed FC < 2.

The viral clone from sample #43 (from the baseline failure group) was the least susceptible of all the viruses tested. This clone showed reduced susceptibility to all three PIs tested; 26-fold, 4-fold and 10.7-fold difference in EC_{50} ATV, DRV and LPV, respectively. Conversely, the viral clone from patient #58 (from the baseline failure group) showed the most susceptibility to all PIs tested; 1.3-fold, 1.1-fold and 2-fold differences in EC_{50} ATV, DRV and LPV, respectively (Figure 4.7).

4.6 Comparison of Phenotypic susceptibilities of 'Baseline success' vs 'Baseline failure' pairs

In patient pair 1 (#54 and #36), there was a 5-fold difference in EC_{50} of atazanavir relative to the assay reference strain in the virus from the 'Baseline success'. This is interpreted as the patient-derived virus is 5-times less susceptible to ATV than the reference strain. For the virus from the 'failure,' fold difference in EC_{50} of atazanavir was 2-fold, meaning the virus is only 2-times less susceptible to ATV. In this pair therefore, the 'Baseline success' virus is more resistant to ATV in comparison to the 'Baseline failure' virus. When we consider the second drug lopinavir, the fold difference EC_{50} was 8-fold for the 'Baseline success' and 7-fold for the 'Baseline failure'. This implies a similar susceptibility response in both viruses from this group. For the third PI darunavir, the fold difference in EC_{50} for the 'Baseline success' was 2-fold whereas for the failure, the FC EC_{50} was about 1 which is similar to that of the reference strain (Figure 3.20A).

In patient pair 2 (#20 and #11), susceptibility levels to ATV and DRV were similar to that of the reference strain. This was the case for both the 'Baseline success' and the 'Baseline failure' viruses. The fold difference in EC_{50} of lopinavir was 5-fold for the 'Baseline success' and 4-fold for the 'failure samples (figure 3.15). In summary there was no difference in EC_{50} fold difference between the 'Baseline success' and the 'Baseline failure' for all the three PIs tested. (see figure 3.20B).

For pair 3 (#17 and #43), the 'Baseline failure' patient displayed a significantly reduced susceptibility to ATV – having a FC in EC_{50} of 26. This is in contrast with the virus from the 'Baseline success' patient which only showed a FC EC_{50} of 3 to ATV. At baseline therefore, the virus from the 'Baseline success' patient was more susceptible to ATV than the virus from the 'Baseline failure' patient. This same susceptibility pattern was observed with LPV where he 'Baseline failure' virus was less susceptible (FC= 10) than the 'Baseline success' virus (FC = 3). When susceptibility to DRV was assessed, the virus from the 'Baseline failure'

was less susceptible (FC = 4) than the virus from the 'Baseline success' (FC = 1 which is similar to the FC of the reference strain) as seen in figure 3.20C.

In patient pair 4 (#40 and #13), the susceptibility to DRV and ATV in both groups was similar to that of the assay reference strain where the virus from 'Baseline success' had a FC EC_{50} of 1.5 and 1.1 to ATV and DRV respectively. In comparison, the virus from the 'Baseline failure' patient only had very slight difference in EC_{50} FC to ATV (2.4) and DRV (1.3). The viruses in this pair also had similar susceptibility response to LPV (FC EC_{50} of 3 and 4.1 for the 'Baseline success' and 'Baseline failure' viruses respectively). Overall, both the 'Baseline success' and 'Baseline failure' viruses in this pair showed similar susceptibility patterns to all three PIs as seen in figure 3.20D.

For patient pair 5 (#5 and #58), the virus from the 'Baseline failure' had fold difference EC_{50} similar to the assay reference strain for all the three PIs tested hence were susceptible to the three drugs. In contrast, the virus from the 'Baseline success' patient showed reduced susceptibility to all three drugs with a 14-fold, and 8-fold difference to ATV and LPV and a borderline susceptibility to DRV (FC=2.4) as shown in figure 3.20E.

In patient pair 6 (#21 and #9), the viruses showed reduced susceptibility to ATV where there was 5-fold difference in EC_{50} in the 'Baseline success' and a 4-fold difference in EC_{50} in the 'Baseline failure' viruses.
There was a marginally reduced susceptibility to LPV in both the 'Baseline success' (FC= 3.5) and 'Baseline failure' (FC = 2.6). The viruses from both patients in this pair were fully susceptible to DRV with the 'Baseline failure' patient showing hyper-susceptibility (FC=0.6) as seen in figure 3.20F.
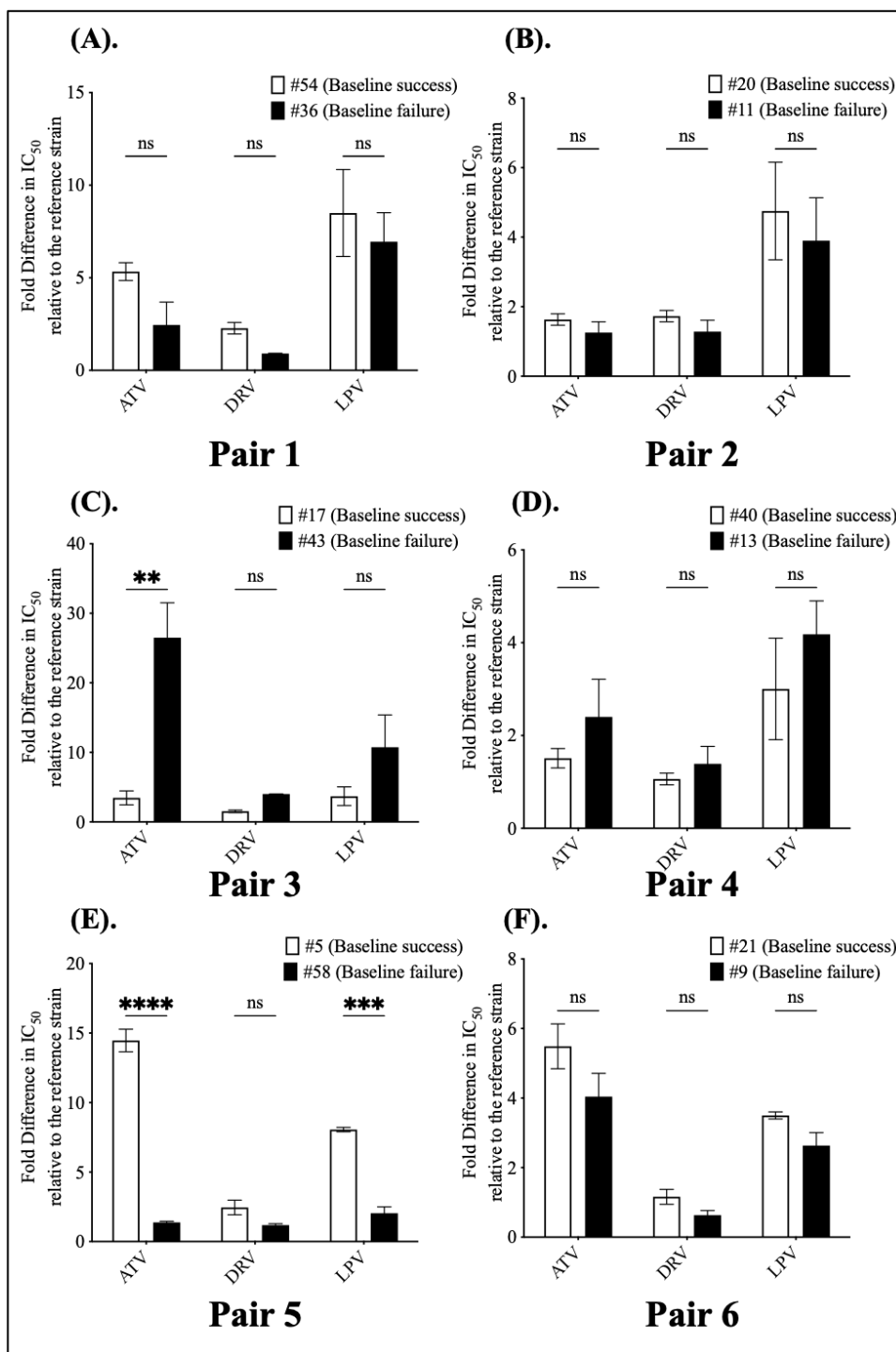
Figure 4.7: PI susceptibility of viruses from all patients

PI susceptibility was determined using IC_{50} data for pseudovirions produced by co-transfection of 293T cells, derived from full-length Gag-protease of patients prior to initiation of second-line boosted PI treatment who either did (Success) or did not (Failure) suppress viral replication after 48 weeks. Each bar represents the mean of three independent experiments performed in duplicate. Error bars represent SEM. Data were analyzed using 2-way ANOVA with Holm-Šídák's multiple-comparisons tests. Asterisks (*) represent statistical significance: * ($p<0.05$), ** ($p<0.01$), *** ($p<0.001$), **** ($p<0.0001$)

 LPV= lopinavir; ATV = atazanavir; DRV = darunavir.

4.6 Association between PI susceptibility at baseline and therapy failure

In section 4.5 above, we determined the PI susceptibility of viral variants derived from 'Baseline successes' and 'Baseline failure' groups. Next, we performed a comparison of the two patient groups. The goal was to identify any correlation between baseline PI susceptibility and subsequent treatment outcome. The susceptibility of baseline viral variants derived from patients in both treatment outcome groups for all three PIs is displayed in figure 4.8 below.

For the PI ATV, 3/6 of the viral variants from the 'Baseline success' group and 2/6 from the 'Baseline failure' group displayed EC_{50} values greater than the cut off for reduced susceptibility (≥ 4). Viruses across both groups were generally susceptible to DRV, with only 1 (#43) in the 'Baseline failure' having a fold difference in EC_{50} of 4.0. For LPV, 3/6 of the virus variants from both groups showed a FC≥ 4, implying reduced susceptibility. We used the previously determined significance cut-off value in phenotypic assays of $\geq$ four-fold[365].

Taken together, the median (IQR) lopinavir fold difference (FC) was 4.04 (2.49–7.89) for 'Baseline failure' and 4.13 (3.14–8.17) for 'Baseline success' (P=0.94, paired t test). The median (IQR) atazanavir FC was 2.43 (1.35–9.66) for 'Baseline failures' and 4.39 (1.60–7.73) for 'Baseline success' (P=0.47). The median (IQR) darunavir FC was 1.23 (0.84 – 2.05) for 'Baseline failures' and 1.53 (1.14– 2.32) for 'Baseline success' (P=0.47) as described in Figure 4.8. Overall, there was no difference in phenotypic baseline susceptibility of 'successes' vs 'failures' for all the 3 PI drugs tested.

Figure 4.8: Comparison of PI susceptibility of viral variants from 'Baseline success' and 'Baseline failure' patients

The phenotypic susceptibility of VSV-g pseudotype viruses derived at baseline from 6 cases ('failures') and 6 controls ('successes') was determined measuring luciferase activity. Susceptibility to three PIs was measured. Data are expressed as a fold-difference in comparison to p8.9NSX+. Each data point is the mean of at least two independent experiments; hairs represent mean and SD. LPV = lopinavir; ATV = atazanavir; DRV = darunavir. The difference in susceptibility between 'successes' and 'failures' was not statistically significant for any of the three PIs.

4.7   Association between single-round infectivity and PI susceptibility

We next analysed the relationship between single-round infectivity and fold difference in EC_{50} to the three PIs (lopinavir, darunavir and atazanavir) in all viruses tested. Previous work using a full-length replication competent virus system had suggested a relationship between replicative capacity (RC) and PI susceptibility and so we wanted to know if there was any such relationship in our single round system. For lopinavir, there was a weak-moderate correlation between these parameters (correlation coefficient 0.32, Figure 4.9 (C)). These data point to a potential link between replication efficiency and the LPV susceptibility. For darunavir and atazanavir, there was no linear relationship between FC and RC (correlation coefficient -0.15 and 0.05 respectively). Importantly the relationship between RC and FC did not appear related to treatment outcome as indicated by the lack of segregation of successes and failures in the plots below.

(A)   (B)

(C)

Figure 4.9: Scatter plot of FC over RC in EC$_{50}$ of three PIs
The relationship between replicative capacity and fold difference in EC$_{50}$ of Atazanavir (A), darunavir (B) and lopinavir (C) are shown. The EC$_{50}$ is relative to a subtype B reference strain, replicative capacity was measured in a single round of infection (also relative to a subtype B reference strain which is represented by 100%)

4.8 Association between single-round infectivity and treatment outcome

It is well established that patients with high viral load have lower probability of suppressing viral load clinically[548-550], suggesting that 'fitter' virus may contribute to treatment failure. We therefore also measured the single-round replication efficiency of patient-derived Gag-protease-containing pseudoviruses derived from both patient groups. The infectivity of viruses from each patient showed a wide but expected variation (ranging from 53-211%) relative to the assay reference strain.

Mean relative infectivity relative to a subtype B reference strain was 117.7% (SD = 65) for the successes and 105.8% (SD = 56) for failures (P=0.93 by Mann–Whitney U-test) as shown in figure 4.10 (A). There was however no clear association with treatment outcome even when each pair was analyzed separately (Figure 4.10 B).
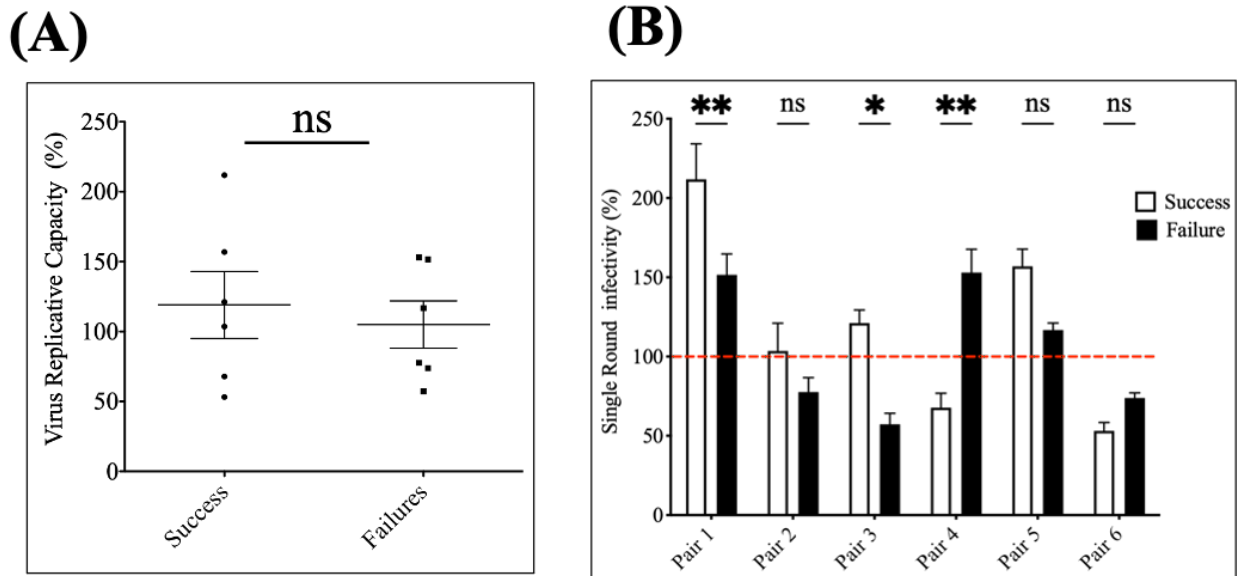
Figure 4.10: Single round infectivity at baseline does not correlate with treatment outcome

Mean replication efficiency of all 'successes' vs 'failures' (A) and (B) relative infectivity of each of the six patient pairs. The single-round infectivity of pseudovirions produced from co-transfection of 293T cells was determined by titration of viruses using a single-cycle of infection. Luciferase activity was measured using SteadyGlo and expressed relative to that of the assay reference strain, p8.9NSX+ (100%). Error bars are standard error of the means of three independent experiments, each performed in duplicate.

Data in (A) were analyzed using the Mann–Whitney U-test while (B) was analysed using 2-way ANOVA with Holm-Šídák's multiple-comparisons tests. Asterisks (*) represent statistical significance: * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$)

4.9 Subtype AG/G Gag-protease sequences from success and failure have little impact on RT susceptibility in a recombinant virus expressing a subtype B RT

As a control for any potential impact of replication capacity on drug susceptibility, we performed phenotypic susceptibility testing for the panel of viral constructs against the RT inhibitors zidovudine (AZT), tenofovir disoproxil fumarate (TDF), nevirapine (NVP) and lamivudine (3TC). In order to do this - for each vector, pseudovirions were produced by co-transfection in 293T cells. The pseudovirions were used to infect fresh 293T target cells in the presence of titrated NRTI drugs AZT, TDF, 3TC, or NNRTI drug (NVP) for 48 hours. The percentage of viral inhibition was calculated from luciferase readout. For these drugs, the fold differences relative to the subtype B reference were less than 2.5-fold across all constructs tested. Figure 4.11 (A) shows limited variability in FC EC_{50} for RT inhibitors, in stark contrast to the considerable variability in fold difference EC50 for LPV (mean FC 10) in the viruses tested. For the NRTI drug (AZT), there was a no correlation between these parameters (correlation coefficient -0.13, Figure 4.11 (B).

(A)   (B)

Figure 4.11: Susceptibility to RTIs
For each vector, pseudovirions were produced by co-transfection in 293T cells. The pseudovirions were used to infect fresh 293T target cells in the presence of titrated NRTI drugs AZT, TDF, 3TC, or NNRTI drug (NVP) for 48 hours. The percentage of viral inhibition was calculated from luciferase readout. Data are expressed as a fold difference in comparison to that of the assay reference strain. Data shows mean of three independent experiments performed in biological replicate. Hairs represent SD.

4.11 Discussion

Major mutations in protease are known to be associated with PI failure, and a number of algorithms can be used to score susceptibility based on protease mutations, for example, the Stanford database (https://hivdb.stanford.edu). No such scoring system exists for gag or env, and mutations in these genes play controversial roles in PI susceptibility. Furthermore, the HIV-1 Gag protein of non-B subtypes is highly polymorphic, resulting in epistatic interactions with protease and effects on protease inhibitor susceptibility. Full-length Gag alongside its co-evolved protease is therefore unsurprisingly more useful than protease alone in determining PI susceptibility using phenotypic assays[397]. Furthermore, an association between pre-ART baseline phenotypic susceptibility to PI and outcome of first-line ritonavir-boosted lopinavir monotherapy in a clinical trial has been reported by our lab[512]. Our goal was to examine whether testing phenotypic PI susceptibility of full-length Gag-protease from Nigerian patients with subtype CRF02_AG and G HIV-1 before initiation of 2nd line therapy, in the absence of PI resistance mutations could differentiate those individuals who were likely to suppress versus those who did not suppress.

In both success and failure groups, there was a high prevalence of amino acids previously associated with PI exposure in subtype B viruses. Many of these amino acids were around Gag cleavage sites and previously described to mutate along with major protease mutations.

However, in our patients these Gag amino acids did not occur with major DRM in protease. These findings are consistent with the known high degree of polymorphism in Gag in non-B subtypes and may explain the higher EC_{50} seen in subtype AG and G as compared to B across this and other previous studies[398,489].

Among the six sample pairs analysed genotypically in this study, the Gag A374S mutation was only found in clonal variants from the 'Baseline failure' group. Although this position was variable across clones (A374N/P/del), only failure

variants had an A➔S mutation in this position. Position 374 is in the p2/NC cleavage site and is highly variable between patients even within the same subtype. The presence of more than 2 mutations in the p2/NC cleavage site at baseline has been reported to predict virological failure. In the samples analysed in this study only three (#20, #11 and #58) out of 12 samples had the amino acid Alanine conserved at this position. There were more than two mutations at this site in both groups. Amino-acid residues G, T, N, P and S at position A374 in the Gag reading frame have been shown to predict virological failure[398]. These substitutions are usually more likely in non-B subtypes.

The PTAP motif duplication in p6 of Gag was only observed in the 'Baseline failure' clone of sample pair 3. This was not seen in the NGS consensus sequence of viruses from this patient. It was only observed in the clonal sequences. This may have occurred as minority variants or may well be artefacts during the cloning process. The PTAP motif plays an essential role in viral budding by recruiting the ESCRT III complex. Its duplication has been shown to confer a selective advantage in viral replication by increasing Gag processing efficiency in the context of protease inhibitor treatment, thereby enhancing the drug resistance of the virus[431,551,552].

Among all virus clones phenotypically assayed in this study, the virus from patient #43 where the PTAP motif duplication was present showed the least susceptibility to ATV (fold difference in EC_{50} of 26-fold, relative to the subtype -B reference). Interestingly, the least susceptibility to DRV was also observed in this virus. The fold difference in EC_{50} of darunavir was 4-fold which is high, when placed in context of the high barrier to resistance of darunavir. Also, this virus was had the highest fold difference in EC_{50} to lopinavir (FC = 10). There was however no effect on the RC of the virus (in the absence of drug). This virus had a relative infectivity of 57% as against the 121% of the 'Baseline success' variant in the pair relative to the subtype B reference strain.

The R57K protease polymorphism was observed in only one of the 'Baseline success' clonal variants (clone #17.G_cB). However, two out of 6 patients in the 'Baseline failure' group (36.AG_c8 and 11.G_c7) have clonal variants exhibiting this mutation. Baseline R57K has previously been independently associated with a higher frequency of subsequent virological failure on the protease inhibitor NFV[553,554].

In this cohort, there was no significant difference in mean EC_{50} FC of viruses with the R57K vs those without this polymorphism. When the R57K polymorphism occurs with another protease polymorphism (E35D), it leads to an alteration of the protease protein dynamics and the conformational landscape[554]. Interestingly, the E35D protease polymorphism was observed in viral variants from three out of 6 'Baseline success' patients. It was not present in any of the viral variants from the 'Baseline failure' group rather, three out of six 'Baseline failure' patient viruses had E35Q/K. Based on these observations it is possible that E35D confers a degree of reduced susceptibility to PI in some genetic backgrounds but not others.

Protease polymorphism L63P was observed in viruses from three out of six 'Baseline failure' group patients. Viruses from patients in the 'Baseline success' group however had L63A (1 out of 6), L63P (1 out of 6). Although L63P does not appear to confer protease-inhibitor resistance by itself, it has been shown to provide some replication benefit in subtype-B viruses in the context of major protease mutations.

Although emergence of M36I has been shown to be associated with treatment failure in subtype B, M36I is the consensus amino acid in subtype AG and G. Genotypic analysis of protease protein revealed the M36I protease polymorphism in all patients irrespective of eventual treatment outcome.

Among all the patient pairs tested phenotypically, patient pair 3 was the only pair to show higher resistance at baseline in the patient with subsequent treatment failure. Within this pair, LPV FC relative to subtype B was 3 vs 10 for 'Baseline

success' and 'Baseline failure', respectively. It is worth noting that only two out of 12 viruses tested had baseline susceptibility to lopinavir similar to the assay reference strain. The other 10 viruses were significantly more resistant to the PI, consistent with previous data[507,509]. In this study the only patient with FC >10 for lopinavir (patient #43) failed treatment with this drug. Interestingly, our lab previously showed that 2/2 patients with FC>10 prior to PI monotherapy went on to virological failure [489]. Taken together, we propose that a lopinavir threshold FC of 10 is relevant.

We also showed here that infection efficiency over a single round was weakly correlated with lopinavir susceptibility prior to initiation of the boosted protease inhibitor. The same relationship did not hold when we tested patient isolates against NRTI drugs. We have previously reported similar findings in replication-competent subtype C viruses that contained patient derived Gag and partial protease genes [505] These data suggest that increased replicative capacity and resistance to PI might involve an overlapping mechanism.

Overall, Gag-protease mediated phenotypic susceptibility to PIs was not statistically different between the 'Baseline failures' (cases) and 'Baseline successes' (controls) for any of the PIs tested: lopinavir, atazanavir or darunavir. In summary, our experiments do not support a role for *in vitro Gag-protease* phenotypic susceptibility testing for PI in predicting virological failure of second-line ART in Nigeria.

The negative result could be due to the influence of adherence, in that second-line therapy is used in patients for whom first-line therapy has failed, usually as the result of incomplete adherence. Therefore, the patient group was enriched for poor adherers, which could have overcome the effects of differences in susceptibility.

We conclude that baseline PI susceptibility alone is not adequate to predict treatment outcomes, given that both 'Baseline success' and 'Baseline failure'

groups had similar Gag-protease PI-susceptibility patterns. However, given that all patients in the 'Baseline failure' group failed PI treatment in the absence of major protease mutations, we are still left with the question of what the determinants of treatment failure are. We performed a longitudinal study where we obtained the patient samples at the failure timepoint and performed genotypic and phenotypic experiments to elucidate determinants of failure. This longitudinal study is presented in the next chapter.

CHAPTER 5: Genetic evolution of HIV-1 Gag and protease in patients failing second-line, ritonavir-boosted protease inhibitor-based antiretroviral regimen

5.1   Introduction

As antiretroviral therapy (ART) scale-up progresses globally in the absence of universal viral load monitoring, significant numbers of persons living with HIV (PLWH) are experiencing virological failure (VF) with emergent drug resistance[497,498,555]. In addition, pre-treatment drug resistance (PDR) has been rising over the past decade[516,556,557]. Although integrase inhibitors are now recommended by WHO in regions where PDR exceeds 10% [474,558] second-line ART in low and middle-income countries (LMIC) is likely to remain dependent on boosted protease inhibitors (PI), specifically lopinavir/ritonavir or atazanavir/ ritonavir.

Studies demonstrate that the detection of major canonical protease mutations is around 20% in PLWH treated with PI- containing combination ART[501,503] raising the question of how virological failure occurs in the remaining cases. Inadequate adherence to medication has been implicated[559-561], but determinants of susceptibility outside the protease gene have also been considered[394]. Interestingly, although PI monotherapy can be effective in some populations in clinical practice[562], this is associated with a higher prevalence of major PI resistance mutations at VF as compared to PI combined with 2 two nucleoside reverse transcriptase inhibitors (NRTI)[563,564].

The HIV-1 envelope (Env) has been reported in two studies to impact PI susceptibility[355,565], with a number of reports of diverse env sequence changes during PI failure[416,566]. HIV-1 Gag is highly polymorphic across subtypes, and existing literature reports diverse mutations occurring both within and outside cleavage sites following treatment with older PI such as indinavir, saquinavir and nelfinavir in subtype B infections[394,395,414,416,566,567]. Although there has been very

limited evidence on the role of HIV-1 *Gag* in susceptibility to modern boosted protease inhibitors such as lopinavir/ritonavir used in second line ART in non-B subtypes, we and others have reported that around 1 in 6 individuals infected with non-subtype B HIV who fail modern PI have *Gag* encoded reduced phenotypic susceptibility to PI [197,396,512,513,568], though specific amino acid determinants have remained elusive.

Cleavage site mutations are thought to partially restore efficient cleavage by protease in the presence of bound drug[196,406]. The mechanism for non-cleavage site mutations may include allosteric changes in protease-Gag interactions that influence the efficiency by which protease locates cleavage sites through dynamic intermolecular interactions in the presence of drug [569,570]. For example, our group previously reported the emergence of T81A in *Gag* that appeared to correlate with reduced susceptibility to the modern PI lopinavir in a subtype AG infected individual in France[512]. This mutation was predicted to impact intermolecular interactions between Gag and protease by Deshmukh and colleagues using nuclear magnetic resonance (NMR)[570].

In this chapter, we studied co-evolved, full-length Gag-protease (both genotypically and phenotypically) to investigate the determinants of treatment failure in non-B (CRF02_AG and G) HIV-1 subtypes in Nigerian patient viruses at baseline (BL) and time of virological treatment failure (VF).
First, we assessed the emergence of HIV-1 Gag and protease genes during the failure to the second line, boosted PI-based regimen. Patients with matched pre-PI and virological failure (VF) plasma samples were selected. Next, we amplified full-length Gag and protease at two time points and compared the sequences obtained and noted any amino acid substitutions which emerged at VF, but which were not present at BL. This was followed by clonal analysis of viral variants from each patient at both time points and the use of the FUBAR online tool to identify positive selection sites in Gag and protease. One representative viral clone from each patient at both timepoints was chosen and we employed single cycle

recombinant assays for in vitro phenotypic characterization in terms of drug susceptibility and replicative capacity (RC)/relative infectivity.

This chapter has the following aims:

1) To amplify full-length Gag-protease from paired plasma samples taken at baseline and at the time of treatment failure from Nigerian patients infected by CRF02_AG and subtype G viruses.

2) To perform clonal analysis of full-length Gag-protease of viruses derived from each time point for each patient to identify amino acid changes previously reported to be involved in or associated with PI resistance

3) To use bioinformatic methods to examine sequences for evidence of positive selection and evolution.

4. To phenotypically investigate determinants of failure to protease inhibitors.

5.2.1 Patient population: clinical and virological information of patients experiencing virological failure on LPV/r selected for this study

The patients selected for this study all have the following similar characteristics: First, all were patients who have failed 1st-line ARV treatment (consisting of 2 NRTIs and 1 NNRTI) – as described in section 3.2.1 of chapter three. Secondly, all participants switched to 2nd-line, PI-based treatment had at least one major RT, drug resistance mutation (the only exception being patient 1 who had no NRTI nor NNRTI mutations at both BL and VF timepoints. The NRTI and NNRTI drug resistance mutations are shown in Table 5.1

We were able to amplify and clone six patient pairs for this study successfully. Patient-1 was a 45-year-old, female patient. Baseline (prior to 2L switch) CD4 count was162 cells/mm^3 and baseline viral load of 140,991 cp/ml. After thirty-four (34) months on 2nd line ART (LPV/r + TDF+FTC), viral load was 6,193 cp/ml.

Patient-2 was a 36-year-old, female patient with baseline CD4 count of 145 cells/mm^3 and baseline viral load of 20,178 cp/ml. After forty-two (42) months on 2nd line ART (LPV/r + AZT+3TC), viral load was 74,224 cp/ml.

Patient-3 was a 39-year-old, female patient with baseline viral load of 24,693 cp/ml. After thirty-six (36) months on 2nd line ART (LPV/r + AZT+3TC), viral load was 32,683cp/ml and CD4 count was 259 cells/mm^3.

Patient-4 was a 40-year-old, female patient whose baseline viral load was 274,504 cp/ml. After thirty-one (31) months on 2nd line ART (LPV/r + TDF+FTC), viral load was 32,683cp/ml and CD4 count was 149 cells/mm^3.

Patient-5 was a 33-year-old, female patient with a baseline viral load of 18,056 cp/ml. After sixty-four (64) months on 2nd line ART (LPV/r + TDF+FTC), viral load was 66,277 cp/ml and CD4 count was 185 cells/mm^3.

Patient-6 was a 38-year-old, male patient. Baseline viral load was 531,626 cp/ml. After thirty-nine (39) months on 2nd line ART (LPV/r + TDF+FTC), viral load was 2,945 cp/ml and CD4 count was 39 cells/mm^3.

These patient characteristics are summarised in table 5.1

Table 5.1: Patient Characteristics and Clinical data of 6 patients at baseline (BL) and at virological failure (VF) timepoints

Patient ID	Sex	Age at 2L Start	Subtype	BL Viral Load (cp/mL)	BL NRTI Mutations	BL NNRTI Mutations	VF Viral Load (cp/mL)	VF NRTI Mutations	VF NNRTI Mutations	PI	2L backbone	Failure Time (months)
Patient-1	F	45	CRF02_AG	140,991	None	None	6,193	None	None	LPV/r	TDF/FTC	34
Patient-2	F	36	G	20,178	M184V	K103E, G190A	117,942	**T69insertion**	None	LPV/r	ABC/AZT/3TC	42
Patient-3	F	39	CRF02_AG	24,693	K65R, M184V	K101E, V108I, Y181C, G190A	32,683	K65R, M184V, **K219E**	K101E, V108I, G190A	LPV/r	TDF/FTC	36
Patient-4	F	40	G	274,504	K70R, M184V, K219Q	K103N, Y318F	16,304	K70R, M184V	**L100I**, K103N, Y318F	ATV/r	TDF/FTC	31
Patient-5	F	33	CRF02_AG	18,056	K65R, M184I	K103N, Y181C	66,277	K65R, M184I	K103N, Y181C	LPV/r	TDF/FTC	64
Patient-6	M	38	CRF02_AG	531,626	M41L, M184V, L210W, T215Y	A98G, G190A, Y318F	2,945	M41L, M184V, L210W, T215Y	A98G, G190A, Y318F	LPV/r	TDF/FTC	39

Emerging NRTI/NNRTI mutations are shown in **bold** lettering.
AZT, zidovudine; 3TC, lamivudine; FTC, emtricitabine; TDF, tenofovir disoproxil fumarate; LPV/r, ritonavir-boosted lopinavir; ATV/r, ritonavir-boosted atazanavir.

5.2.2 Definition of baseline and virological failure for patient population

In this study, baseline (BL) is defined as the timepoint when patient has failed 1st line ART, prior to switching to PI-based second line treatment. In other words, BL samples were collected from PI-naïve but not ART-naïve patients.

Virological failure (VF) was defined as viral load >1,000 copies /ml. As described in chapter 1, viral load testing was usually performed at least 30 days after patient has shown immunological (CD4) decline and received intensive adherence counselling.

5.3   Methods

A detailed description of the laboratory methods is described in chapter 2 and briefly described below.

5.3.1   Amplification of full-length Gag-protease genes

HIV-1 RNA was manually extracted from archived plasma samples using the QIAamp viral RNA extraction kit. Full-length Gag-protease was amplified and either cloned into an intermediate vector (pGEM) or cloned directly into a subtype B-based (p8.9NSX+) vector. Clonal sequencing of up to 10 plasmids (where possible) was performed by standard Sanger sequencing. Sequences were manually edited using DNA dynamo software (http://www.bluetractorsoftware.co.uk) and aligned to the HXB2 reference in MEGA v7.0 software. Protease sequences were analysed for PI resistance mutations using the Stanford Resistance Database (https://hivdb.stanford.edu) while Gag sequences were analysed manually. By drawing phylogenetic trees, we selected the viral variant that most closely represented the consensus (obtained via next-generation sequencing) which was taken forward for phenotypic PI drug susceptibility and infectivity assays.

5.3.2   PI susceptibility and infectivity assays

HEK-293T cells were co-transfected with a Gag-Pol protein expression vector (p8.9NSX+) containing cloned patient-derived Gag-protease sequences, pMDG expressing vesicular stomatitis virus envelope glycoprotein (VSV-g), and

pCSFLW (expressing the firefly luciferase reporter gene with HIV-1 packaging signal). Transfected cells were seeded with serial dilutions of protease inhibitors (ATV, DRV and LPV) and harvested pseudovirions were used to infect fresh 293T cells. To determine strain infectivity, transfected cells were seeded in the absence of drug. Infectivity was monitored by measuring luciferase activity 48-hours after infection. Results derived from at least two independent experiments (each in duplicate) were analysed. The EC50 was calculated using GraphPad Prism 8.0 (GraphPad Software Inc., La Jolla, CA, USA). Susceptibility was expressed as a fold difference in EC_{50} compared with the subtype B reference strain (p8.9NSX+). The replicative capacity of these viruses was assessed by comparing the luciferase activity of the recombinant virus with that of the WT subtype-B control virus in the absence of the drug. Equal amounts of input plasmid DNA were used, and it has previously been shown that percentage infectivity correlates well with infectivity/ng p24 in this system.

5.4   Results

5.4.1 Clonal analysis of Gag-protease from study patients who experienced treatment failure

Clonal variants were sequenced using Sanger sequencing (as previously described in chapter 2). The consensus sequence of the clones were compared between baseline and VF timepoints.

A total of 5 viral variants (clones) were derived from Patient-1 at the baseline time point, while four viral variants were derived at failure. A number of Gag mutations were only observed in the VF variants but not in the baseline variants. These include previously described mutations associated with PI failure or exposure (E12K) and other mutations which have not been described previously (E93D, D177E, and E464G).   Also, the previously described R76K mutation[417] was observed in all failure variants and 2/5 of the baseline variants. Of interest, T81A was present in 4/5 viral variants from this patient at baseline but had reverted to WT (subtype-B consensus) amino acid by the time of treatment failure in all 4

viral variants. Importantly, all the mutations described above were non-cleavage site mutations. These are shown in Table S-1

A total of 6 viral variants (clones) were derived from Patient-2 at the baseline timepoint while 3 viral variants were derived at failure. Seven sites previously associated with PI exposure or resistance mutated between the baseline and VF timepoints and are shown in Table S-2. At Gag position 62, all six baseline variants occurred as 62Q with all three variants at the VF timepoint observed to be 62K. At position 75, all baseline variants had the amino acid I while all variants at VF had the amino acid, L. Interestingly, the R76K mutation occurred in all baseline variants but was substituted and occurred as 76Q in all VF variants. At position 79, all baseline variants occurred as Y79 but mutated to Y79H in all VF variants. Additionally, the G→E mutation occurred at position 123, with all baseline variants having the amino acid, G at this position while all VF variants had the amino acid E. Other mutations occurred at positions 128 and 219 which have been previously associated with reduced susceptibility to PIs.

A total of 10 viral variants (clones) were derived from Patient-3 at the baseline timepoint while eight viral variants were derived at failure. A relative homogeneity was observed in the clonal sequences from this patient at both time points (Table S-3). Several amino acid substitutions occurring at the time of virological failure were not present at baseline, and interestingly, none of the substitutions occurred at sites previously known to be associated with PI-exposure or failure. In this patient sample, the mutations occurring at virological failure occurred as follows: In MA, positions 15 (S→M), 49 (G→S), 69 (R→K), 90 (S→R) and 124 (S→G). In the CA, Q311L occurred in all viral variants at failure. Others are an R→K mutation at position 418 (NC), and P453S and P458S in the p6 domain.

A total of 4 viral variants (clones) were derived from Patient-4 at the baseline time point, while five viral variants were derived at failure. Two mutations, (E→K and G→R) at positions 12 and 62 respectively have previously been reported to be associated with PI failure or exposure. In virus variants from this patient, we

observed an R12K and T62A mutations emerging at failure. Other observed mutations which emerged during failure occurred at Gag positions 15 (A→S), 58 (A→V), Q90H and S111H in the MA domain. Additionally, 215 (L→A), 223 (L→V), 248 (T→A), I256V occurred in the CA, 401 (L→I) in the NC and 470 (I→M) in the p6 domain. These are shown in supplementary Table S-4

A total of four viral variants (clones) were derived from Patient-5 at the baseline timepoint, while three viral variants were derived at failure. There was the emergence of several mutations in the MA at VF timepoint. The following MA mutations emerged at failure: 12 (N→K), R20Q, 62 (E→K), 75 (F→L), 86 (W→C), 90 (K→E), 95 (Q→R), 107 (V→I), S111N, 113 (P→Q), T122A, G123E, 126 (S→del), and 127 (H→del). In the NC domain, two mutations emerged: 387 (K→R) and R406K. There was one emerging mutation in the p1 domain: 441 (N→S) and one mutation in p6: 467 (G→E). These mutations are summarized in Table S-5.

A total of 6 viral variants (clones) were derived from Patient-6 at the baseline time point, while seven viral variants were derived at failure. The following emerging matrix (MA) mutations were observed: K28R, 34 (I→L), 49 (S→G), 61 (I/M →I), 63 (H→Q), 66 (T/R→T), K110N and A119T. Others are 385 (G→S) and 473 (G/S→P) in NC and p6, respectively. These are shown in Table S-6.

5.4.2 Phylogenetic and genetic distance analysis

Our original plan was to employ phylogenetic and genetic distance analyses to study the evolution of the viral variants within each patient. We wanted to perform ancestral reconstruction, which would enable us to identify amino acid changes correlating with treatment failure. We would then also be able to compute Mean Pairwise Genetic Distances (MPWGD) between the amino acid sequences of the viral variants. This would enable us to compare genetic diversity at baseline and virological failure.

To achieve this, we would ideally have at least ten (10) viral clones from each patient in order for increased accuracy of the phylogenetic analyses. This was

however not possible as we were only able to have up to 10 (or more) viral clones from only the baseline timepoint of Patient 3. We therefore only used the positive selection analysis to identify sites under selective drug pressure.

5.4.3 Positive selection analysis

For identification of particular sites under diversifying/purifying selection, the FUBAR (Fast Unbiased Bayesien AppRoximation) method[571] was used. FUBAR is a web-based Datamonkey software package which was used to analyse the intra-patient selection pressures on the HIV-1 Gag and protease gene in the patient samples. The software calculates the ratio (ω) of the rate of non-synonymous substitutions per non-synonymous site (dN) over the rate of synonymous substitutions per synonymous site (dS). If dN/dS is equal to 1 then no selection is inferred. If dN/dS is <1, the number of dS is greater than dN and therefore negative selection is inferred. If, however dN/dS is >1, the number of dN is greater than dS and so positive selection is in play[572]. FUBAR detects selection under a model which allows substitution rate variations from site to site. It calculates the mean posterior distribution of synonymous (α) and non-synonymous (β) substitution rates.

The viral variants from each patient were subjected to selective pressure testing to identify evidence of positive selection as a result of PI exposure. All viral variants from a single patient, derived from both baseline and failure samples, were aligned in MEGA software using the MUSCLE algorithm and nucleotides 1450-1 removed to ensure Gag and protease were in a single reading frame. Aligned nucleotide sequences were submitted to data monkey in FASTA format. The sites identified by the FUBAR analysis as undergoing positive selection within each patient are shown in Table 5.2 along with the amino acid present at that position in each of the viral variants.

At least one amino acid position was identified as undergoing positive selection in 5 patients (except in Patient-6 where no sites were identified). In Patient-1, four sites were identified in Gag and none in protease. Five (5) sites were identified in Patient-2 (three Gag, and two protease). One and five Gag sites were identified

in Patient-3 and Patient-4, respectively. The highest number of positively selected sites was identified in Patient-5, which had six Gag and one protease sites.

Some of the sites which showed evidence of positive selection are sites that have previously been associated with PI-exposure or PI resistance, whereas others have not been reported to correlate with PI exposure. The only positively selected sites which occurred in more than one patient was at Gag positions 15 (in Patient-3 and Patient-4). Interestingly, this site has not been associated with PI exposure or failure previously.

Gag positions 76, 81and 479 in Patient-1 were positively selected. All three sites have previously been associated with both PI-exposure and PI-resistance *in vitro*[417,573]. Other previously selected sites were at Gag positions 62 (in Patient-3) and 441(in Patient-5). Gag positions 62 and 441 have been reported to emerge upon in vitro exposure to PIs[436,574,575].

Other positively selected sites in Gag which have not been previously reported were; positions 124 (in Patient-1), 248 and 280 (Patient-2), 69, 403 and 443 (Patient-4) and at positions 66, 90, 107, 113 and 387 (Patient-5).

In the protease gene, three positions were positively selected; positions 15 and 67 (in Patient-2) and position 14 (in Patient-5). These positions were neither major nor accessory PI mutation sites associated with drug resistance.

Importantly, we note that most of the amino acid positions identified by positive selection analyses were located within the p17 (MA) subunit of Gag, where there was a total of 11 sites. No site was selected in the p2 sub-unit, while two positively selected sites were in the p24 (CA), p7 (NC) and p1 domains, with one site in the p6 domain. These are graphically summarized in Figure 5.1

Table 5.2 Positively selected sites in Gag

		Baseline										VF									
	Pos. site	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
Patient 1	**76**	K	R	R	R	K						K	K	K	K						
	81	T	A	A	A	A						T	T	T	T						
	124	N	N	S	D	N						N	N	N	N						
	479	G	K	R	R	K						R	R	R	R						
Patient 2	115	A	A	A	A	A	A					K	K	K							
	248	T	T	T	T	T	T					A	A	A							
	280	V	V	V	V	V	V					V	V	V							
	Pro (15)	R	R	R	R	R	R					K	K	K							
	Pro (67)	E	E	E	E	E	E					C	C	C							
Patient 3	15	S	S	S	S	S	S	S	S	S	S	M	M	M	M	M	M	M	M		
Patient 4	15	A	A	A								S	S	S	S	S					
	62	T	T	T	T							A	A	A	A	A					
	69	N	N	N	N							K	K	K	K	K					
	403	L	L	L	L							L	L	L	L	L					
	443	S	S	S	S							Q	Q	Q	Q	Q					
Patient 5	66	P	P	P	P							S	S	S							
	90	K	K	K	K							K	E	E							
	107	V	V	V	V							I	I	I							
	113	P	P	P	P							Q	Q	Q							
	387	K	K	K	K							R	R	R							
	441	C	N	N	N							S	S	S							
	Pro (14)	M	M	M	M							K	K	K							
Patient 6	None																				

Positively selected sites are shown, along with the amino acid residue present in each viral variant at that site. Positively selected sites are numbered according to HXB2 numbering. Sites in red have previously been associated with PI resistance or exposure

Pro = protease
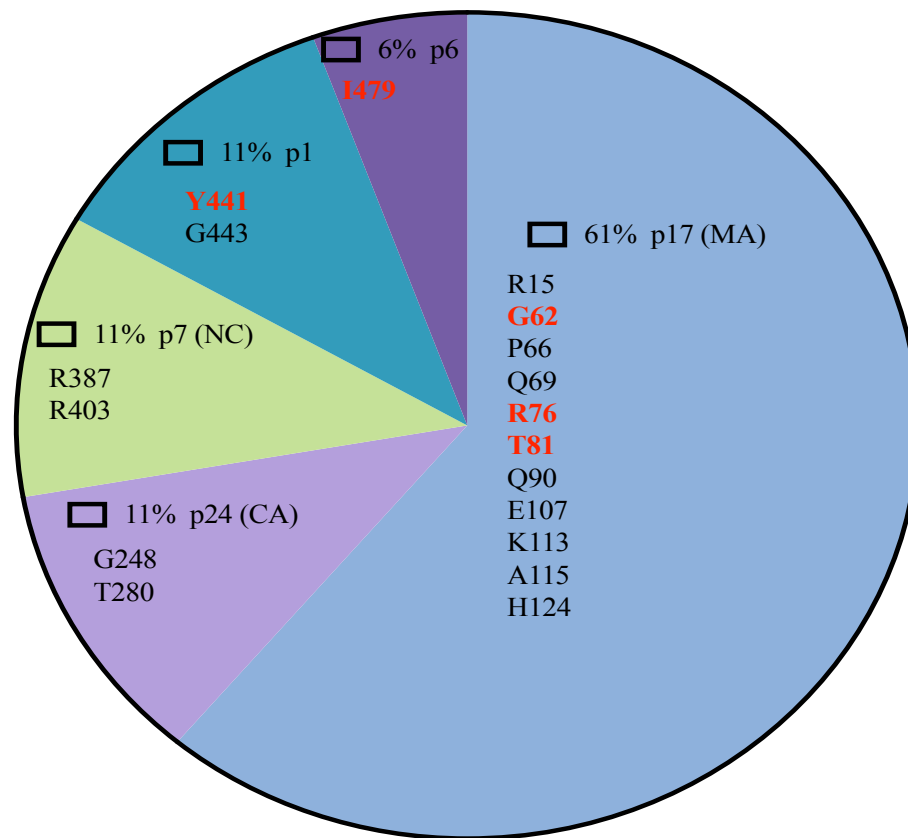
Figure 5.1: Positively selected sites by FUBAR
FUBAR online algorithm identified these 17 gag sites as undergoing positive selection due to drug pressure.

5.4.4 Genotypic characteristics of clones selected for phenotypic drug susceptibility assays

A summary of protease and Gag mutations at baseline and at virological failure is presented in Figures 6.1 to 6.4

First, we note that several Gag mutations previously reported in HIV-1 subtype-B viruses as arising due to PI exposure, or PI resistance exist as natural polymorphisms in CRF02_AG and G subtypes and can be found even prior to any exposure to PI. In the p17 (MA), the R76K and **Y79F** mutations which were previously described as conferring PI resistance in subtype-B viruses[417] were present at baseline in viruses from Patient-1, Patient-2, Patient-3, Patient-4 and Patient-6. Additionally, Gag **E12K** mutations (**E12R** in Patient-4) were present at baseline in all viruses from patients (except Patient-1). At the p17/p24 cleavage site, the deletion of valine at Gag position 128 is associated with PI-exposure in vivo, in-vitro and PI-resistance[407,410,412,438]. This deletion (**V128del**) was observed at baseline timepoint in viruses from patients 3, 5 and 6, only emerging at failure timepoint in patients 2 and 4 viruses. Additionally, the **Y132F** p17/p24 cleavage site mutation, which is associated with PI resistance *in vivo*[406,537] was observed at baseline in the virus from Patient-6.

The only emerging p24/p2 cleavage site mutation was the **V362I** in the virus from Patient-4, which was not present at baseline time point. This mutation has been associated with PI resistance in subtype-B viruses[391]. In the Gag p2 sub-domain, the **T371Q/N/del** mutation was present in all patient viruses at baseline (except Patient-2) and at virological failure timepoint. In subtype-B viruses, this mutation was previously associated with *in vivo* PI exposure. At the p2/p7 cleavage site, the **S373T/A/Q** (previously shown to emerge due to PI-exposure *in vivo* [404,414,441]) occurred at baseline in virus patients 1, 4, 5 and 6.

In the p7 (NC) region, a deletion of valine occurred at Gag position 90 at both baseline and failure timepoints in viruses from Patient-1, Patient-5 and Patient-6 while in viruses from Patient-2, Patient-3 and Patient-4, it was a **V390I**. Amino acid substitutions at this position have been shown to associated with *in vitro* PI-exposure[418,422]. Additionally, mutation at Gag position 401, which is associated

with PI exposure *in vivo* and *in vitro* was present in all patient viruses at both baseline and failure timepoints. At the p7/p1 cleavage site, the emergence of **I437L** mutation was only observed at failure (in patients 2 and 4). Also present at the p7/p1 cleavage site was the K436R which emerged at failure in viruses from Patient-2 but was present at both timepoints in Patient-4 viruses, and not observed in any other virus from the other patients. These p7/p1 cleavage site Gag mutations (at positions 436 and 437) have been reported by several studies as associated with PI-exposure in *vivo*, *in vitro* and PI-resistance[197,384,400,404,406,414,436,443,444,536]. Within the p1 domain, amino acid changes at Gag position 441 was variable was observed in viruses from both time points in all patients, except in Patient-1 where there was no amino acid change. The **Y441S** mutation in Gag was recently identified as being associated with virological failure[574]. In the six patients from this study, viruses from Patient-2, Patient-4, and Patient-5 had Y441N at baseline time point, virus from Patient-6 had **Y441H** mutation while virus from Patient-3 had **Y441S**, with no mutation in patient-1 virus. At virological failure timepoint, however, there was the emergence at Gag 441 of N→Q (Patient-2), N→S (Patient-4 and Patient-5) while there was no amino acid change in viruses from Patient-1, Patient-3 and Patient-6 at this position.

At the p1/p6 cleavage site, the **L449P** was the most frequent mutation which occurred in both baseline and failure timepoints in viruses from Patient-1, Patient-3, Patient-5 and Patient-6. This mutation has been shown in several studies[379,384,394,403,424,443] to play a significant role in PI resistance. Additionally, at this cleavage site was the **S451N** mutation which occurred in both baseline and failure timepoints of Patient-2 while it only emerged at failure in Patient-4. This mutation (**S451N**) is associated with *in vivo* exposure to PI[412,443]. The third mutation observed was the P453S which emerged at failure in Patient-3 and which has been associated with PI-failure[379,443].

In the Gag p6 region, mutations at position 474 occurred in viruses from Patient-2, Patient-4, Patient-5 and Patient-6. In Patient-1 and Patient-2, there was a deletion at position 474 at baseline, but this reverted to the wild type during virological failure. In Patient-4, there was a Q74 (P→Q) mutation between the

baseline and failure timepoints while in Patient-5 and Patient-6, there was no change in the amino acid at this position although the amino acids differed from the subtype-B reference (**Q474P** and **Q474de**l respectively). Amino acid mutation **Q474L** has previously been reported to emerge after *in vivo* PI exposure[424]. Additionally, we observe changes in some amino acids between baseline and failure timepoints. These included F463 (W→C), R464 (464E→G), S464 (K→M) and I479(K→R) in Patient-1. In Patient-2, the changes were T470 (I→A), P473 (S→K), I479(R→L) and D480 (E→T). In Patient-4, the change between baseline and failure timepoints was T470 (I→M). Of these changes, only the I479 mutation has previously been associated with PI resistance[573].

In the protease gene, none of the viruses from the patients at both timepoints had any PI major resistance mutations. There was, however, the emergence of the PI accessory resistance mutation (**M46V**) at the time of virological failure in Patient-5. Other protease polymorphisms which only emerged at virological failure included the K14R (in patients 1, 2 and 6) and **R57K** (in Patient-3).

			Gag											Protease
			p17	p24			p2	p7	p1	p6			

Patient	Time	ARV	p17 (1–127)	p17/p24 (128–137)	p24 (138–358)	p24/p2 (359–368)	p2 (369–372)	p2/p7 (373–382)	p7 (383–427)	p7/p1 (428–437)	p1 (438–443)	p1/p6 (444–453)	p6 (454–500)	Protease (1–99)
Pt. 1	BL		**R15H**, G62E†, Y79F§,ß	-		-	T371Q§	S373T§, T375N§	V390Δ†, T401L§,†	-		L449P§,†,ß	**F463W**, **R464E**,E468K† **S465K**, Q474Δ§, **I479K**§	V3I, I13V, G16E, K20I, E35Q, M36I, S37R, R41K, R57K, H69K, L89M,
	VF	LPV/r	E12K†, **R15L**, G62E, R76K, Y79F, T81A, E93D, T122A	-		-	T371Q§	S373T§, T375N§	V390Δ, T401L§,†	-		L449P§,†,ß	**F463C**, **R464G**, **S465M**, **I479R**§	V3I, I13V, **K14R**, G16E, K20I, E35Q, M36I, R41K, R57K, H69K, L89M,
Pt. 2	BL		E12K, **R15A**, **K28Q**, **K30R**, **R58Q**, **G62Q†**, **R76K**§,ß, **E93G**, K112E†, **K114E**, **A115N**, **Q117K**, **A119E**, **A120V**, **D121S**, H124N, N126S	-	**I138A**, H219Q†, **G248T**, **N252S**, **P339A**	-	372↓A	**T375A**§, I376M§, R380K, G381S§	I389N§, **V390I†**, T401L§,†		**Y441N**ß	S451N§	**T470I**, **P473S**, Q474Δ§, **I479R**§, **D480E**	V3I, I13V, I15V, G17E, K20I, E35Q, M36I, S37N, R41K, R57K, C67E, H69K, I72V, V82I, L89M,
	VF	LPV/r	V7I, E12K†, S9R, **R15T**, **K28H**, **K30M**, **R58K**, **G62K†**, Q65H, **R76Q**§,ß, Y79H§,ß, Q90K, R91G, **E93K**, K95Q, K112E†, **K114Δ**, **A115Δ**, Q116Δ, **Q117Δ**, A118Δ, **A119Δ**, **A120Δ**, **D121T**, G123E†, H124D, S125D, **N126K**, Q127K	V128I§,†,ß	**I138L**, Q182S, T186S, **G248A**, **N252H**, E260D, T303A, **P339S**, G357S	-	V370A§, T371N§, N372Δ	A374T§, **T375N**§, G381S§	I389F§, **V390I†**, T401L§,†	K436R§,†,ß, I437L§,†,ß	**Y441Q**§	S451N§	G466Δ, **T470A**, **P473K**§, Q476R, **I479L**§, **D480T**	V3I, **K14R**, I15V, L19I, M36I, R41K, L63P, H69K, V77I, L89M,
Pt. 3	BL		E12K†, **R15S**, G62E†, R76K§,ß, Y79F§,ß, **Q69R**, **Q90S**, K113Q, **H124S**	V128Δ§,†,ß		-	Q369K§, T371Q§	A374Δ§, T375N§, I376V§, R380K	V390I†, T401L§,†	-	Y441S§	L449P§,†,ß		V3I, I13V, K20I, M36I, S37N, L63P, C67S, G68E, H69K, L89M,
	VF	LPV/r	E12K†, **R15M**, G49S, G62E†, **Q69K**, R76K§,ß, Y79F§,ß, **Q90R**, K113Q, **H124G**	V128Δ§,†,ß	Q311L	-	Q369K§, T371Q§	A374Δ§, T375N§, I376V§, R380K	V390I†, T401L§,†, K418R	-	Y441S§	L449P§,†,ß, P453S§,†,ß	P458S	V3I, I13V, K20I, M36I, S37N, R57K, L63P, G68E, H69K, L89M,
Pt. 4	BL		**E12R†**, **R15A**, **R58A**, **G62T†**, R76K§,ß, Y79F§,ß, K112Q†, **Q127P**	-	E203D, V215L, H219Q, **I223L**, **G248T**, R286K, N315G	-	T371Δ§	S373A§, T375A§, R380K,	I389F§, **V390I†**, T401L§,†	K436R§,†,ß	**Y441N**ß	-	**T470I**, Q474P§	V3I, I13V, L19I, K20I, M36I, S37N, R41K, L63P, H69K, L89M,
	VF	LPV/r	**E12K†**, **R15S**, **R58V**, **G62A†**, R76Q§,ß, Y79H§,ß, Q90H, K95Q, S111H, K112Q†, G123?, **Q127?**	V128I§,†,ß	E203D, V215A, **I223V**, **G248A**, I256V, N315N V362I	V362I†	T371Δ§	S373S§, A374T§, T375N§, G381S§	I389F§, **V390I†**, T401L§,†	K436R§,†,ß, I437L§,†,ß	**Y441S**ß	S451N§	**T470M**, Q474Δ§, T487S§	V3I, I5V, K20I, M36I, L63P, H69K, V77I, L89M, I93L
Pt. 5	BL		E12N†, G62E†, K95Q, **E107V**, **K113D**	V128Δ§,†,ß	E203D, V215T, R286K, **T280V**	-	T371Q§	S373T§, A374Δ§, T375N§, R380K	V390Δ†, T401L§,†	-	**Y441N**ß	L449P§,†,ß	Q474P§	V3I, **K14M**, L19P, K20I, M36I, S37D, L63P, H69K, L89M,
	VF	LPV/r	E12N†, G62E†, K95Q, **E107I**, **K113Q**	V128Δ§,†,ß	E203D, V215T, R286K, **T280I**	-	T371Q§	S373T§, A374Δ§, T375N§, I376V§, R380K	V390Δ†, T401L§,†	-	**Y441S**ß	L449P§,†,ß	Q474P§	V3I, **K14M**, L19P, K20I, M36I, S37D, M46V, L63P, H69K, L89M,
Pt. 6	BL		E12K†, G62E†, R76K§,ß———	V128Δ§,†,ß, Y132F§	E203D, V215T	-	T371Δ§	S373Q§, A374Δ§, T375N§, I376V§, R380K	**N385G**, V390Δ†, T401L§,†	-	**Y441H**ß	L449P§,†,ß	Q474Δ§	V3I, K20I, M36I, S37D, R41K, H69K, K70R, L89M,
	VF	LPV/r	E12K†, K28R, G62E†, R76K§,ß, E93D, A119T	V128Δ§,†,ß, Y132F§	E203D, V215T	-	T371Δ§	S373Q§, A374Δ§, T375N§, I376V§, M377R, R380K	**N385S**, V390Δ†, T401L§,†, R409K†	-	**Y441H**ß	L449P§,†,ß	Q474Δ§	V3I, **K14R**, K20I, M36I, S37D, R41K, H69K, K70R, L89M,

(Figure legend on next page)

Figure 5.2: Gag and protease clonal plasma sequences

Differences in gag and protease sequences between Baseline (BL) and virological failure (VF) timepoints.

†= associated with exposure to PIs invitro;

§=associated with exposure to PIs in vivo;

ß=associated with PI resistance;

  = amino acid insertion

del= amino acid deletion

Bold fonts = emerging mutations

5.4.5 Phenotypic PI susceptibility of Gag-protease derived from patients

In order to explain the mechanisms contributing to treatment failure in the absence of known PI resistance mutations, we determined the phenotypic PI susceptibility and single-round infectivity in comparison to reference strain p8.9NSX+ for viral variants representative of each patient sample at baseline (BL) and failure (VF).

As described in section 5.2, resistance test vectors (RTVs) were produced by cloning full length, patient-derived Gag-protease sequences into a subtype B-based (p8.9NSX+) vector. We then tested the susceptibility of the RTVs to ATV, DRV and LPV.

In interpreting the phenotypic assay results, we utilized the previously determined significance cut-off value in phenotypic assays of $\geq$ four-fold[365]. We, therefore, defined reduced susceptibility as $\geq$ a 4-fold increase in EC_{50} of the patient-derived virus in comparison with the assay reference and viruses with <4-fold difference in EC_{50} are considered susceptible.

In Patient-1 at baseline (BL), the fold difference (FC) in EC_{50} was 2.5 for ATV, 0.9 for DRV, and 7 for LPV. At the time of virological failure (VF), the FC increased slightly for all PIs tested. The 2.5 to 5.3-fold increase for ATV implied that the virus became less susceptible at VF timepoint. Although FC in EC_{50} of DRV significantly increased from 0.9-fold to 2.3, this increase was still less than the threshold of $\geq$4-fold, implying that at VF, the patient virus was still susceptible to DRV. At baseline, susceptibility to LPV was inherently reduced (FC=7), and this reduced LPV susceptibility did not change at VF.

In Patient-2 at BL, susceptibility to ATV and DRV were similar to that of the reference strain (FC: ATV=1.3-fold, DRV=1.3-fold). At VF however, there were slight increases in FC with ATV and DRV showing 3.2 and 2.1-fold differences respectively. This increase was not enough to confer reduced susceptibility using a threshold of $\geq$4-fold. The virus from this patient showed borderline reduced susceptibility to LPV at BL, with a further reduction in susceptibility at the time of virological failure (a fold-difference increase from 3.9 at BL and 4.8 at VF).

For all the three PIs tested, viruses from Patient-3 became more susceptible to the drugs at failure timepoint than they were at baseline. There was generally a

reduction in EC_{50} FC (2.4 to 0.6-fold, 1.4 to 0.8-fold and 4.2 to 1.2-fold for ATV, DRV and LPV respectively) at virological failure.

In Patient-4 at BL, susceptibility to ATV and DRV was similar to that of the reference strain (FC: ATV=1.4-fold, DRV=1.2-fold) and a 2.0-FC for LPV. At the time of virological failure, there was no change in FC of DRV and LPV, with only ATV showing a very slight decrease in susceptibility (1.4 to 2.3-fold difference).

In Patient-5 at baseline (BL), the FC was 4.0 for ATV, 0.6 for DRV and 2.6 for LPV. At the time of virological failure (VF), the FC significantly increased for all PI tested from 4.0 to 23.4-fold for ATV, from 0.6-fold to 4.3-fold for DRV and from 2.6-fold to 10.7-fold for LPV. In this patient, the virus was fully susceptible to all three PIs at baseline; however, at VF, susceptibility to all three drugs was significantly reduced.

Fold difference in EC_{50} of DRV remained unchanged (FC = 2.3) between BL and VF in Patient-6. This implies that susceptibility of the virus to DRV in this patient did not decrease over time. The virus became slightly more susceptible to ATV at the time of VF where the FC reduced from 18-fold t baseline to 15-fold at time of failure. Against LPV, although the virus already exhibited decreased susceptibility at baseline (FC=4.2), susceptibility was further significantly reduced (FC = 7.9) at the time of virological failure.

We next aggregated and compared the FC in EC_{50} of all viruses at baseline time point, versus at virological failure as shown in Figure 5.4. The general trend was towards reduced susceptibility at failure in comparison to baseline. However, the reduced PI susceptibility was not significantly associated with treatment failure using Mann-Whitney U rank sum tests at the P = 0.05 significance level (ATV p=0.59; DRV p=0.24; LPV p=0.48).
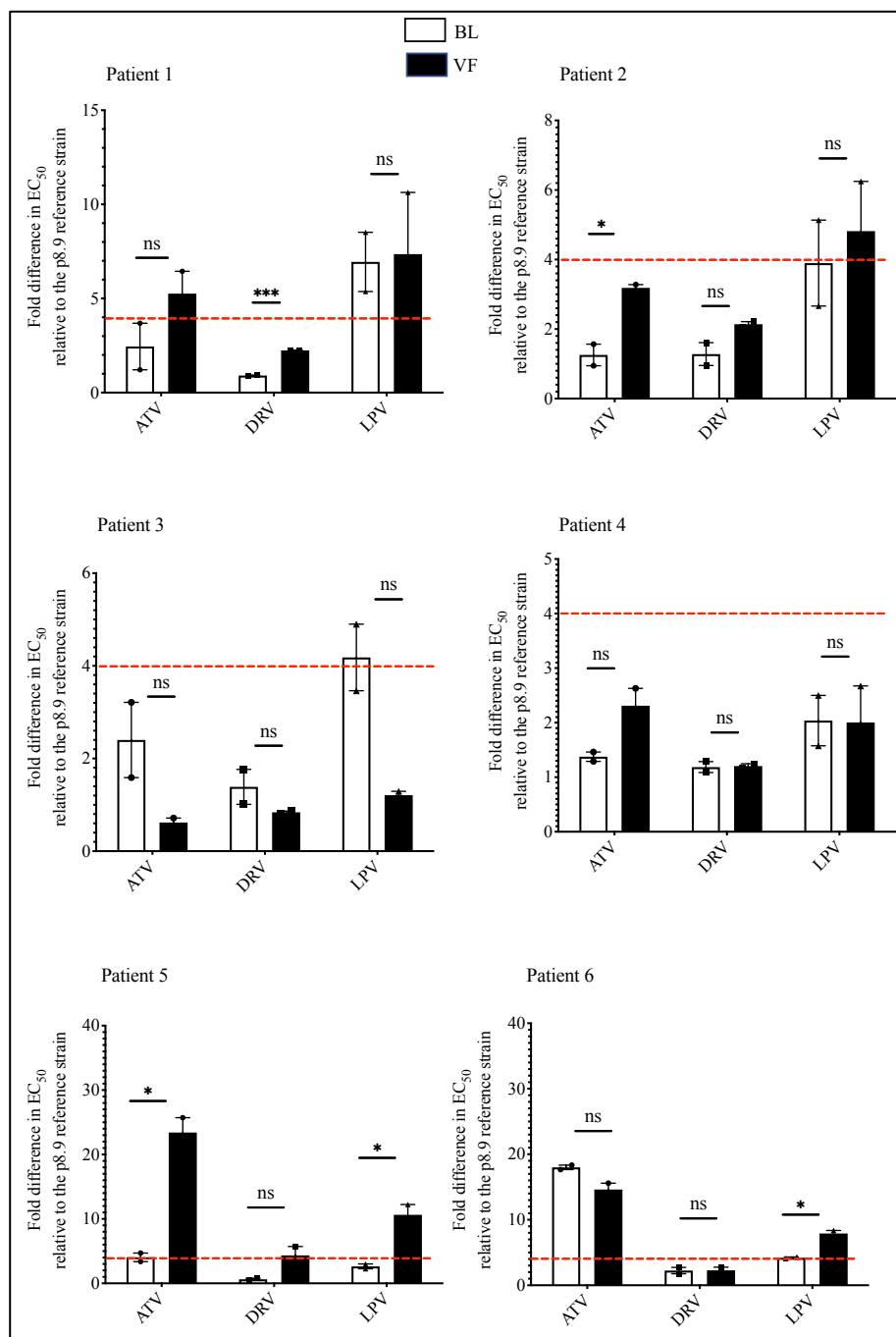
5.4.6 Single-round infectivity of Gag-protease derived from study patients

We next compared the infectivity of the resistance test vector obtained from each of the patients at baseline versus at virological failure. When compared to the p8.9NSX reference strain, virus infectivity in 2 patients (Patients 1 & 4) at baseline was greater than infectivity at virological failure. Conversely, at failure, the viruses in 2 patients infected more efficiently than at baseline, as shown in Figure 5.5. Baseline virus from Patient-1 showed $152 \pm 13.2\%$ infectivity relative to wild-type virus. At the time of virological failure, the virus replicated less efficiently with relative infectivity of $66 \pm 15\%$ ($p=0.001$, unpaired t-test).

There was no significant difference in relative infectivity of viruses from Patient-3; with baseline and VF relative infectivity of $153 \pm 14.6\%$ and $143 \pm 16.5\%$ respectively ($p=0.81$, unpaired t-test). Baseline infectivity in Patient 4-derived virus showed higher relative infectivity at baseline ($117 \pm 4.4\%$) compared to infectivity at failure timepoint, which was $69 \pm 8.4\%$ ($p=0.006$, unpaired t-test). A similar trend was seen in the virus from Patient-6, where the relative infectivity at baseline and failure was $143 \pm 16.3\%$ and $128 \pm 19.5\%$ ($p=0.81$, unpaired t-test) respectively.

There was a slight increase in relative infectivity of the virus from Patient-2 at failure ($116 \pm 15.9\%$) compared to baseline infectivity of $78 \pm 9.0\%$ although the difference was not statistically significant ($p=0.18$, unpaired t-test). The virus derived from Patient-5 had a significantly higher relative infectivity at failure ($106 \pm 4.2\%$) than at baseline with relative infectivity of $74 \pm 3.2\%$ ($p=0.009$, unpaired t-test).

Next, we compared the relative infectivity of all viruses at baseline timepoint (130%) to viruses at failure timepoint (111%). There was no significant difference in relative infectivity of the viruses at baseline compared to failure timepoints ($p=0.2403$, Mann Whitney test).
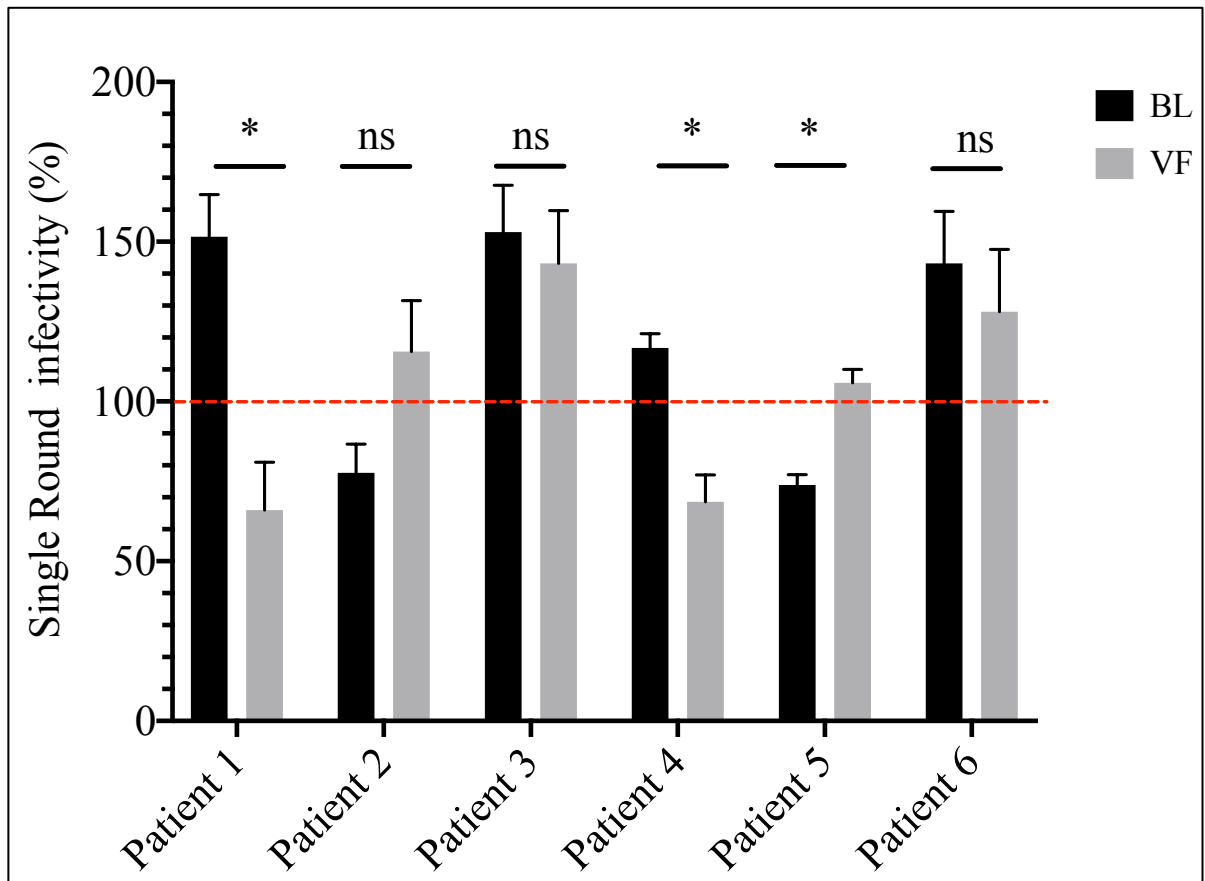
Figure 5.5. Single-round infectivity of patient-derived viruses from baseline vs virological failure timepoint

The single-round infectivity of pseudovirions produced from co-transfection of 293T cells was determined by titration of viruses using a single-cycle of infection. Luciferase activity was measured using SteadyGlo and expressed as a fold difference in comparison with that of the assay reference strain, p8.9NSX+ (set at 100%), red dashed lines. The single-round infectivity of baseline viral variant (black bars) and virological failure timepoint (white bars) is shown. (A) The mean and standard error for each timepoint. Data was analysed using 2-way ANOVA with Holm-Šídák's multiple-comparisons tests. Asterisks (*) represent statistical significance: * ($p<0.05$), ** ($p<0.01$), *** ($p<0.001$), **** ($p<0.0001$).

5.5   Discussion

Our lab has previously reported the importance of full-length Gag and its co-evolved protease in phenotypic susceptibility assays[397]. In this chapter, we aimed to utilize phenotypic PI susceptibility of full-length Gag-protease from patients who have failed PI therapy but with no drug resistance-associated mutations in protease. The objective was to examine if this could shed some light on the causes of treatment failure. Given that we had access to longitudinal samples from patients (at least one sample taken before PI treatment and a second sample at the time of treatment failure) were able to compare PI susceptibility and single round infectivity and amino acid sequences of viruses from at the different time points, from a single patient.

Classically, patients fail ART as a result of the accumulation of resistance mutations at, or near the active site of the drug. For patients treated with ritonavir-boosted protease inhibitors; however, treatment failure could arise in the absence of major resistance-associated mutations in the protease gene that may account for clinical resistance[477,564,576,577].

Although mutations at non-canonical sites (such as Gag and env) have been shown to affect PI susceptibility, the evidence continues to mount, while the entire picture is yet unresolved. Further complicating this is the subtype-specific differences in the amino acid composition at some of the critical PI resistance-conferring sites whereby, resistance-conferring Gag mutations in subtype B viruses occur as consensus amino acids in PI-naïve Subtype AG/G viruses.

In this chapter, we sought to gain deeper insights into the Gag determinants of treatment failure in subtypes AG/G-infected individuals failing 2nd line, boosted protease inhibitor-based ART from Nigeria. Through genotypic assessment of full-length Gag-protease sequences of samples collected longitudinally from six patients, we sought to examine amino acid changes which emerge during treatment, then correlate such changes with treatment outcome.

Given a large number of polymorphisms in the Gag gene, we systematically and considered any observed mutations and polymorphisms in this manner;

a. Mutations (polymorphisms) previously reported as being associated with PI-exposure or PI-resistance.

b. Mutations occurring at cleavage sites and non-cleavage site mutations

c. Mutations that emerged at VF which were not present at BL and which have not been previously described.

First, we note that the presence of several amino acids which have previously been associated with PI failure or PI exposure in subtype-B viruses occur either as natural polymorphisms or consensus amino acid sequence in PI-naïve, AG/G viruses. In the MA domain, for example, mutations at Gag positions 12, 62, 76, 79 have previously been associated with PI failure or exposure in subtype-B viruses, variations, however, occurred at these sites at baseline, before any PI treatment in our study samples. In the CA domain, the only mutation which has been associated with PI exposure in-vivo [410,418,422] was the H219Q which was present at baseline in Patient-2 but not detected at virological failure timepoint. In the p2, mutations previously associated with PI exposure (Q369K, T371Q/del/N) were identified, however, these mutations were not newly emergent as they were present at the baseline time point, pre-PI treatment. This was similar across the Gag gene where mutations previously associated with PI exposure or resistance did not selectively emerge at failure timepoint but were present at baseline. This suggests and adds to the already existing evidence that drug resistance-causing mutations in subtype-B viruses may occur as natural polymorphisms in non-B subtypes. Because baseline timepoint samples were collected after 1st line failure on NNRTI and NRTI regimens, we are unable to ascertain the amino acid sequences before the commencement of any ART at all.

Much evidence already exists on the role of Gag cleavage site mutations in resistance to protease inhibitor[395,406,414,443]. Gag mutations occurring at protease cleavage sites are thought to confer resistance to PIs by reducing the binding affinity of PIs to the active site of the protease, thereby allowing the enzyme to continue its proteolytic function[196,399,400,455]. Alternatively, PI resistance driven by Gag mutations may involve the development of Gag mutations that act as

secondary or compensatory mutations and restore viral fitness[395]. Within our samples, at the MA/CA cleavage site, the Y132F mutation occurred at both BL and VF timepoints in patient 6. The deletion of the valine residue at Gag position 128 is associated with PI-failure and PI-resistance in subtype-B viruses[394,407,410,412,438]. In our samples, the V128del was seen in both BL and VF viruses from Patients 3, 5 and 6. This is not surprising as V128del is the consensus amino acid at that position in CRF02_AG viruses. In patients 2 and 4 (subtype G), there was the emergence of V128I at VF timepoint, which was not present at baseline. This emergence may have played a role in the slight decrease in susceptibility to all three PIs tested (in Patient-2) and to ATV in Patient-4.

At the MA/p2 cleavage site, only in one patient was there the emergence of a cleavage site mutation, this was the emergence of V362I at the VF timepoint of Patient-4. The V362I mutation is associated with in vitro exposure to PIs[391] and has been shown to confer resistance to the maturation inhibitor, bevirimat[534,578]. Although there is a slight decrease in susceptibility to ATV in Patient-4, this was below the four-fold cut off mark.

At the NC(p7)/p1 cleavage site, we see the emergence of K436R and I437L in patient-2 and patient-4. Interestingly, these two patients harboured subtype G viruses. The K436R mutation is associated with PI exposure in-vivo, in-vitro and PI-resistance[197,400,443,444] while the I437L mutation is associated with PI exposure and/or resistance[384,404,414,436,536].

In the p1/p6 cleavage site, the L449P mutation was present in the BL and VF viruses from all CRF02_AG viruses (Patients 1, 3, 5 and 6) but not in the subtype G viruses. Although L449P mutation has been reported to be associated with PI exposure and resistance[407-409,538-540], its occurrence here in both BL and VF samples of all CRF02_AG viruses could be a subtype-specific phenomenon. In the subtype G viruses, we also observed the emergence of S451N at VF in Patient-4 while for patient 2, the S451N mutation was present even at baseline. The occurrence of S451T/G/R is associated with PI-exposure in-vivo in subtype B viruses[412,443] as well as in non-B subtypes[509,579]. The emergence of S451N at

VF may be as a result of drug pressure or could be due to other factors such as host factors.

Equally crucial to the emergence of PI resistance is the role of non-cleavage site mutations in Gag. For example, a study on the effect of various substitutions on the development of HIV-1 resistance to APV identified the emergence of following non-cleavage site mutations L75R (MA), H219Q (CA), V390D/A and R403K (NC) and E468K (p6) after in-vitro exposure to APV[422]. This study (by Gatanaga and colleagues) concluded that mutations at both cleavage sites and non-cleavage sites were essential for the efficient replication of APV-resistant HIV-1. Additionally, non-cleavage site mutations have been demonstrated to be important for the recovery of fitness in PI-resistant viruses[424]. Our lab has also previously demonstrated the role of non-cleavage site mutations in resistance to PIs by demonstrating that domains in Gag beyond its cleavage sites can have both resistance and fitness effects in viruses bearing primary mutations in the protease[417,420]. More recently, the predominance of non-cleavage site mutations in Gag during Gag-protease co-evolution was reported by Codoner and colleagues where they showed that the co-evolution of MA and CA non-cleavage site mutations was strongly associated with protease mutations[421]. Predominant changes in non-cleavage sites were also reported in a recent study by Blanch-Lombarte and colleagues[437].

In this present study, non-cleavage site mutations were more predominant than cleavage site mutations. A lot of these changes we observed have not been reported previously, and we observed non-cleavage site amino acid changes between BL and VF. At Gag position 15, between baseline and virological failure timepoints, we observed the emergence of histidine to leucine in Patient-1, alanine to threonine in Patient-2, serine to methionine in Patient-3 and alanine to serine in Patient-4. Other non-cleavage site mutations observed included E203D and V215L/A/T present at both time points of Patients 4, 5 and 6. Similarly, the T401L and Y441N/SQ were present at both time points in all six patients.

Mutations selected under PI–selective pressure in subtype B HIV-1 can occur as natural polymorphism in non-B subtypes and may contribute to decreased susceptibility to PIs of non-B subtype HIV-1 viruses [580-583]. Positive selection analysis using the FUBAR model provided evidence of the evolution of the virus under PI pressure, and a number of amino acid positions undergoing positive selection in both Gag and protease were identified. We identified eighteen (18) unique sites in Gag (11in MA, 2 each in CA, NC, p1 and 1 site in the p6 domain). There was no positively selected site in the p2 domain. Additionally, FUBAR identified three (3) unique sites in protease as undergoing positive selection. Interestingly, there was no overlap in the selected sites between the patients – except at Gag R15, which overlapped between Patient-3 and Patient -4. This lack of overlap could be as a result of differences in the ART regimen or the differences in the duration on 2nd line ART or even differences in the selective pressure exerted on the virus by host-specific factors. Of the 18 amino acid sites identified as undergoing positive selection, only 5 (28%) occurred at previously described PI-resistance or exposure sites (Gag positions 62, 76, 81, 441 and 479), with the remaining 13 (72%) occurring at positions which have not been previously described. Finding a large number of positively selected amino acids at previously undescribed positions could suggest that the viruses are being subjected to other forms of pressure apart from that exerted by drugs. Additionally, the selection of these sites could indicate a possible role in drug resistance in AG/G viruses since most of the studies in which PI-resistance or PI-exposure Gag sites were described were carried out on subtype B viruses. It is also possible that given the few numbers of clonal variants in some patient samples (3 clones were the least), we may have obtained more positively selected sites, or there may have been some difference in the data.

We found a high number of polymorphisms and mutations at both baseline and failure timepoints. Again, this vast number of polymorphisms could be explained by subtype-differences as CRF02_AG, and G subtypes have previously been reported to be highly polymorphic in the gag gene. Another possible explanation could be that the viruses are under pressure from the host immune system. Early HIV-1 intra-host evolution studies found that cellular immune pressure was a

dominant selective force in viral evolution, accounting for up to 50% of the intra-host amino-acid sequence diversity selected throughout infection in some cases[584-586].

The following Gag mutations/polymorphisms: K28Q/R (Patient-2, BL; Patient-6, VF), I138L (Patient-2, VF), G357S (Patient-2, VF) and V362I (Patient-4, VF) occurred within well-defined CD8+ T-cell epitopes[587]. A previous study showed that Gag variants located in immunodominant CD8+ epitopes in P17, such as K28Q, I34L or Y79F, failed to be killed by CD8+ T cell due to an impaired antigen processing and presentation[588]. From these studies, we know that host immune response influences the evolution of Gag and specifically by adaptive CD8+ T-cell selective pressure, and this phenomenon may be at play in some of our samples.

Phenotypic susceptibility testing of the viruses showed a general trend in which baseline samples were more susceptible than samples taken at virological failure in 4 out of 6 patients (Patients 1, 2, 4 and 5) with the reverse trend observed in 2 patients (Patients 3 and 6). Comparing baseline and virological failure sequences would have enabled us to identify mutations or amino acid substitutions which arose during treatment. While we were able to identify these changes intra-patient, reaching a definite conclusion about the role of individual substitutions/mutations was complicated based on the fact that there were no mutations which selectively appeared at failure, but not at baseline across all patients. Therefore, whereas we can speculate on the possible role of amino acid mutations in viruses from a patient, we are not able to infer the effect of the same amino acid mutations in the next patient, given that viruses from the next patient may harbour completely different mutations. When we compared the single-round infectivity between BL and VF groups, there was no significant difference between the two groups.

In this chapter, we demonstrated variation at baseline in Gag of CRF02_AG and G viruses at sites of interest for PI resistance and evolution at these sites during PI therapy. However, we could not establish the emergence of clear genetic determinants of therapy failure across the six patients studied. Admittedly, the

sample size may be too small for us to establish a clear general relationship between failure and emergence of particular Gag mutations. Another possibility is that given the large number of Cytotoxic T-lymphocytes (CTL) epitopes present in Gag, the HLA type of the patient would affect the development of resistance mutations in Gag. Additionally, other regions outside of Gag and protease could be playing a role in PI resistance, for example, Rabi and colleagues previously established that HIV-1 envelope mutations could confer PI resistance[355]. Finally, poor adherence to PI therapy could be playing a role in second-line PI failure in these patients.

Using the previously determined significance cut-off value in phenotypic assays of ≥ four-fold[365], virus from Patient-1 showed decreased susceptibility to ATV and no change in susceptibility to LPV. In Patient-2 and Patient-6, it was a slight increase in resistance to LPV. Viruses from Patient-5 showed a significant increase in resistance to all three PIs between BL and VF timepoints.

In the next chapter, we focused on Patient-5 to study the contribution of the emerging Gag mutations on drug susceptibility in order to clarify the potential role of Gag mutations in the failure of protease inhibitors.

CHAPTER 6: In Vivo Emergence of a Novel Protease Inhibitor Resistance Signature in HIV-1 Matrix

6.1   Introduction

The best documented resistance-associated substrate mutations are located in, or near, the cleavage sites in the NC/SP2/p6 region of Gag. These mutations improve interactions between the substrate and the mutated enzyme and correspondingly increase cleavage[395]. Although generally less-studied, studies from our lab, and others have shown that mutations in the matrix region of Gag could *be* sufficient to rescue replication capacity and also confer resistance to protease inhibitors[417,420,421,512,589]. Utilizing Coevolution Analyses for Protein Sequences (CAPS) to identify correlated Gag residues involved in the evolution of the protease and resistance to PIs, Codoñer and colleagues identified the matrix protein as the major contributor to protease evolution under selective pressure from PIs, with 50% of the 38 Gag residues located in the p17 viral matrix[421]. Mutations in the matrix proteins have been shown to contribute to the continued PI therapy failure, with these matrix mutations playing a major role in the network of strongly correlated mutations in Gag, as well as between Gag and protease[589]. In a study by Flynn and colleagues, it was suggested that matrix and p1/p6 mutations form the core of a network of strongly correlated Gag mutations and contribute to recurrent treatment failure.

In chapter five, Patient-5 had a significant difference in PI susceptibility between baseline and failure time points for all three PIs tested (Figure 5.5 and Table 5.2). This failure to protease inhibitors occurred in the absence of any major mutations in protease. This phenomenon is not entirely surprising as mutations outside of protease (such as Gag and Env) have been shown to confer PI resistance. In this chapter, we explored samples from this patient in-depth to elucidate the determinants of resistance to protease inhibitors, especially lopinavir, given that this was the drug regimen the patient was on, and failed treatment.

After phenotypically establishing PI failure, we examined and compared sequences, identified distinct mutations which may be responsible, followed by

site-directed mutagenesis and phenotypic drug resistance assays to establish the roles of various amino acid mutations either occurring singly or in synergy with other mutations. We then attempted to establish the mechanism by which these MA mutations cause reduced susceptibility to protease inhibitors.

6.2 Methods

6.2.1 Next-generation sequencing

Manual nucleic acid extraction was done using the QIAamp Viral RNA mini kit, Qiagen (Hilden, Germany) with a plasma input volume of 0.5-1.5 mL. The first strand of cDNA was synthesised using SuperScript IV reverse transcriptase, Invitrogen, (Waltham, MA, USA), followed by NEBNext second-strand cDNA synthesis E6111, New England Biolabs GmbH, (Frankfurt, Germany). Sample libraries were prepared as per the SureSelectXT automated target enrichment protocol, Agilent Technologies (Santa Clara, CA, USA) with in-house HIV baits. Whole-genome deep sequencing was performed using Illumina Miseq platform (San Diego, CA, USA). Trimmed reads were then compared to a reference panel of 170 HIV subtypes/CRFs from the Los Alamos database (https://www.hiv.lanl.gov), and the best match was used for reference mapping. Duplicate reads were removed from the BAM files. Mutations were included if they were present at over 2% frequency within the read mixture at that position, with a minimum read depth of 100. An in-house custom script was used to identify SNPs at each position by BLAST analysis of individual HIV *pol* against the HXB2 reference genome.

6.2.2 Haplotype Reconstruction and Phylogenetics

Whole-genome haplotype reconstruction was performed using a newly developed maximum-likelihood method, HaROLD (Haplotype assignment of virus NGS data using co-variation of variant frequencies)[590]. SNPs were assigned to each haplotype so that the frequency of a variant at any time point was represented by the sum of the frequencies of the haplotypes containing that variant. Time-dependent frequencies for longitudinal haplotypes were optimised by maximizing the log-likelihood, which was calculated by summing over all

possible assignments of variants to haplotypes. Haplotypes were then reconstructed based on posterior probabilities. The calculations were repeated with a range of possible haplotype numbers, and the optimal number of haplotypes was determined by the resulting value of the log-likelihood. After constructing haplotypes, a refinement process remapped reads from BAM files to the constructed haplotypes. Haplotypes were also combined or divided according to AIC scores, in order to give the most accurate representation of viral populations. Phylogenetic trees of constructed haplotypes were constructed using RAxML-NG using the GTR model and 1000 bootstraps.

6.2.3 Site-Directed Mutagenesis

Full details of the site-directed mutagenesis process are as described in section 2.1.10 of chapter 2. The primers designed and synthesised for this the SDM in this chapter are listed in Table 2.1, Chapter 2.

6.2.4 Western blotting

Details of the western blotting method was presented in chapter 2, section 2.4. Briefly, equal amounts of each of the viral clone plasmid were used to transfect 293T cells, in addition to a VSV-G plasmid and reporter genome expressing plasmid. Each of the pseudovirions was produced in the absence and presence of a range of concentrations of LPV, which had been added at 16 hours following transfection. At forty-eight (48) hours post-transfection, the culture supernatant was harvested and passed through a 0.45-μm pore-size filter to remove cellular debris. The filtrate was centrifuged at 14,000rpm for 90 minutes to pellet virions. The pelleted virions were lysed in Laemmli reducing buffer (1M Tris-HCl pH 6.8, SDS, 100% glycerol, β-mercaptoethanol and bromophenol Blue). Virus supernatant and cell lysates were subjected to electrophoresis using 4–12% Bis-Tris Protein Gels (Thermo Fisher Scientific). This was followed by electroblotting onto PVDF or nitrocellulose membranes. The HIV-1 Gag proteins were visualized by a trans-Illuminator (Alpha Innotech or ChemiDoc) using anti-p24 Gag antibody.

6.2.5 Multiple cycle replication assay

A detailed description of this assay is described in chapter 2, section 2.4. Briefly, $20\mu l$ of WT (R9-BaL) and Mutant (R9-BaL with the five amino acid changes in MA) virus prep was used to infect 1.5×10^6 of SupT1-CCR5 suspension cells in 2mL of media per well and incubated at 37°C for 2 hours. The supernatant was discarded, followed by low centrifugation (800xg) for 10 minutes. The cell pellets were resuspended in RPMI media and used to infect 4×10^6 SupT1-CCR5 cells which have been pre-treated with (or without) varying concentrations of lopinavir. The supernatant was harvested on day 2, 4, 7, 9 and 11. The harvested supernatant was used to infect fresh TZMbl cells to assay for infectivity which was based on the Tat-dependent upregulation of LTR-driven firefly luciferase expression upon HIV-1 infection of TZMbl cells. Luciferase Assay Reagent was added, and the luminescence was measured using GloMax 96 Microplate Luminometer (Promega).

On day 11 when the final virus supernatant was harvested, the HIV-1 infected SupT1-CCR5 cells were washed with PBS and fixed with 3% paraformaldehyde (PFA), permeabilized in BD Perm Buffer (BD Biosciences). This was followed by staining with anti-HIV p24 monoclonal antibody conjugated to Fluorescein isothiocyanate [HIV-1 p24 (24–4) FITC] to detect intracellular Gag. The acquisition was performed using a BD X-205 Laser Fortessa machine, and data were analysed using FlowJo software.

6.3 Results

The VF1 sample was taken 41 months after PI initiation when the viral load was 241,894 copies/ml while the VF2 sample was taken 64 months after PI initiation when the viral load was 66,277 copies/ml. We successfully phenotyped four clones from baseline, four clones from VF1 and two clones from VF2. The phenotypic drug susceptibility results are shown in figure 6.1 and the full-length Gag-protease genotypic sequences of the clones are shown in the multiple sequence alignment in figure 6.2.

All four clones from baseline (BL) showed similar LPV susceptibility to the reference strain. Two VF1 clones (clone 2 & clone 3) and one VF2 clone (clone 2) showed significant decrease in LPV susceptibility. Although both clones at VF2 are above the threshold for susceptibility (FC>4), VF2Clone1 was about 4x more susceptible as compared with VF2Clone2. This finding suggests a mixture of 'susceptible' and 'resistant' viruses at the failure time points (VF1 and VF2) (Figure 6.2a)

Next, we carried out multiple sequence alignment of the clonal variants at the three time-points in order to establish whether any mutations correlated with decreased susceptibility to LPV. Interestingly, we identified amino acid mutations which only occurred in the 'resistant' viral clones, but not in the 'susceptible' clones, this would suggest a correlation with reduced LPV susceptibility.

In the MA domain of the VF1 and VF2 'resistant' clones, we identified Threonine to Alanine mutation at Gag position 122 (T122A), a Glycine to Glutamic acid mutation at Gag position 123 (G123E) and the deletion of Serine and Histidine residues at Gag positions 126 and 127 respectively. Additionally, the 'resistant' viral clone at VF2 also had an insertion of 4 amino acids: Glutamic acid-Leucine-Arginine-Glutamic acid (E-L-R-E) between positions 475 and 476 (based on HXB2) in Gag p6 domain. These mutations were only observed in the 'resistant' but not in the 'susceptible' viral clones. We also noted the occurrence of Methionine to Valine mutation at position 46 of protease (M46V) in the VF2 'susceptible' clone (figure 6.2b). The M46I/L are nonpolymorphic mutations selected primarily by IDV, NFV, FPV, ATV and LPV[591-594] and are associated with reduced susceptibility to ATV, FPV, IDV, LPV and NFV[371,376,393,595,596]. The M46V which was observed in VF2 clone1 has been described by the Stanford HIVdb as a rare nonpolymorphic PI-selected mutation that has not been well studied.

Of the two distinct VF2 virus clones isolated and successfully phenotyped, there was a 4-fold difference in LPV susceptibility between the two clones. This difference in susceptibility between the two clones, as well as the distinct mutations observed in the genotypic sequences granted us the opportunity to map determinants of PI susceptibility using these two clones identified at failure. (Figure 6.1)
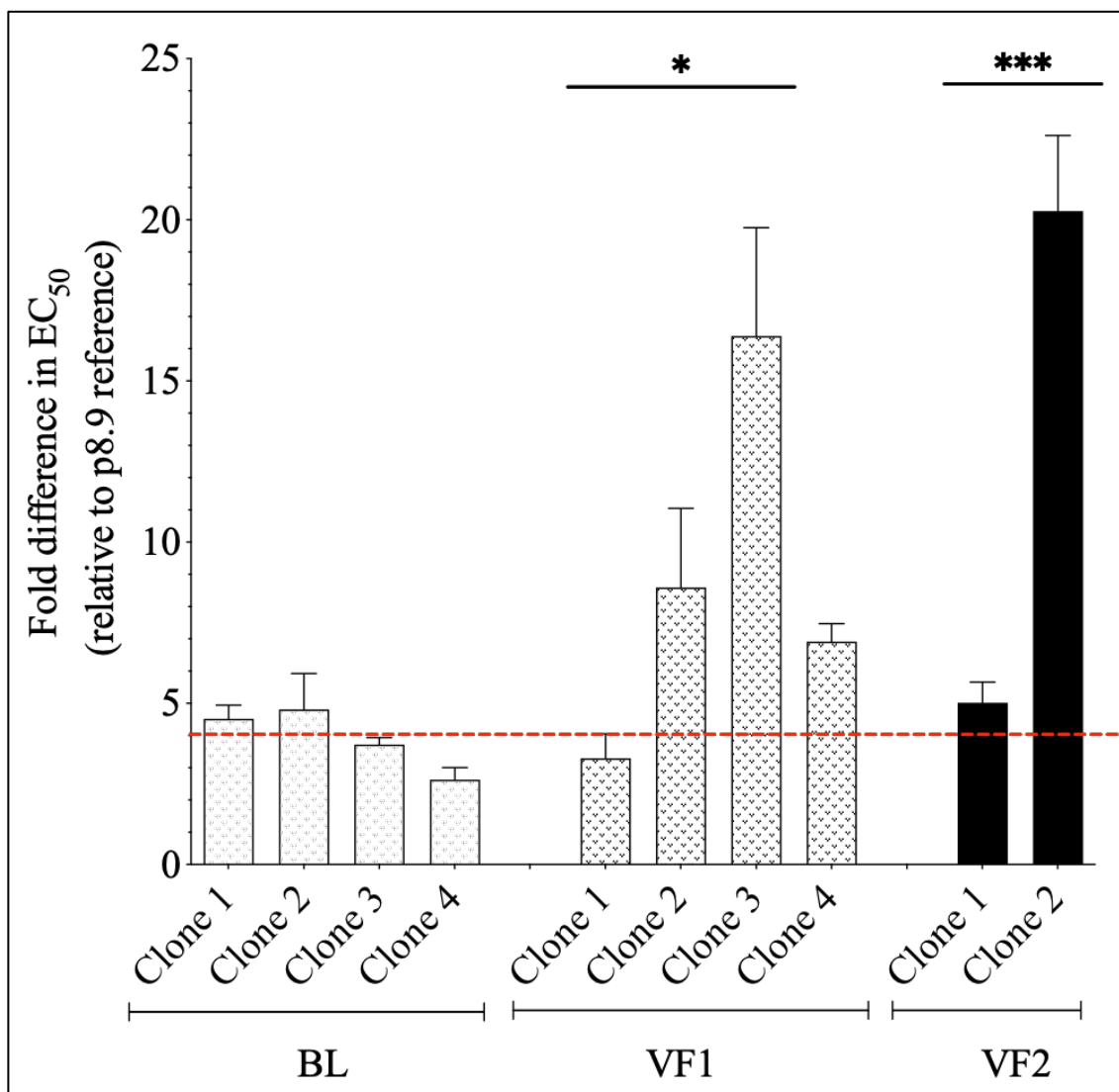
Figure 6.1: Variation in LPV phenotypic susceptibility
Full-length Gag-protease was amplified from plasma samples before commencement of PI-based ART (baseline) – and at two timepoints after failing PI-based ART (VF1 & VF2) and cloned into p8.9NSX+. VSV-g pseudotyped viruses encoding luciferase were produced by co-transfection in 293T cells. PI susceptibility of pseudovirions derived from each patient was determined using a single replication-cycle drug susceptibility assay as measured by luciferase activity. Data displayed are fold difference in EC_{50} values of PIs in comparison to that of the assay reference strain, p8.9NSX+. Error bars represent the standard error of the mean of three independent experiments performed in duplicate. White bars represent clones from baseline; grey bars are clones virologic failure timepoint 1; black bars are clones from virological failure timepoint 2. BL and VF1 data was analysed using 1-way ANOVA with Tukey's multiple comparisons tests while VF2 data was analysed using Wilcoxon test of data. Asterisks (*) represent statistical significance: * ($p<0.05$), ** ($p<0.01$), *** ($p<0.001$), **** ($p<0.0001$). The red dashed line shows the previously reported cut off for a significant reduction in susceptibility of greater than four-fold in comparison with the assay reference strain.

```
HXB2          1 MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI
BL_Clone_1    1 ........T..N..A.............R...L.............L.......G...Q.L
BL_Clone_2    1 ........T..N..A.............R...L.............L.......G...Q.L
BL_Clone_3    1 ........T..N..A.............R...L.............L.......G...Q.L
BL_Clone_4    1 ........T..N..A.............R...L.............L.......G...Q.L
VF1_Clone_1   1 ........T..N..A.............R...L.............L.......G...Q.L
VF1_Clone_2   1 ........T..N..A.............R...L.............L.......G...Q.L
VF1_Clone_3   1 ........T..N..A.............R...L.............L.......G...Q.L
VF1_Clone_4   1 ........T..N..A.............R...L.............L.......G...Q.L
VF2_Clone_1   1 ...........N..A.............R...L.............L.......G...Q.L
VF2_Clone_2   1 ........T..K..A...Q.........R.................L.......G...Q.L
VF2_Clone_3   1 ........T..K..A...Q.........R.................L.......G...Q.L
consensus     1 ********.**.**.****.*********.***.***********.*******.***.*.

HXB2         61 LGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAA
BL_Clone_1   61 ME....A.R.....F......I...W...KK.K.Q...........V.K...PQT.....
BL_Clone_2   61 ME....A.R.....F......I...W...KK.K.Q...........V.K...PQT.....
BL_Clone_3   61 ME....A.R.....F......I...W...KK.K.Q...........V.K...PQT.....
BL_Clone_4   61 ME....A.R.....F......I...W...KK.K.Q...........V.K...PQT.....
VF1_Clone_1  61 ME...SA.R...........I...W....K.K.Q...........M.K...PQT.....
VF1_Clone_2  61 ME...SA.R...........I...W....K.K.Q...........M.K...PQT.....
VF1_Clone_3  61 ME...SA.R...........L...C...EK...R............I...N.QQT.....
VF1_Clone_4  61 ME...SA.R...........I...W....K.K.Q...........V.K...PQT.....
VF2_Clone_1  61 ME....A.R.....F......I...W...KK.K.Q...........V.K...PQT.....
VF2_Clone_2  61 MK....A.R...........L...C...EK...R............I...N.QQT.....
VF2_Clone_3  61 MK....A.R...........L...C...EK...R............I...N.QQT.....
consensus    61 ..***..*.*****.******.***.***..*.*.***********.*.*.*...*****

HXB2        121 DTGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGAT
BL_Clone_1  121 A..S.SH-.........A........V.............I............T.......
BL_Clone_2  121 A..S.SH-.........A........V.............I............T.......
BL_Clone_3  121 A..S.SH-.........A........V.............I............T.......
BL_Clone_4  121 A..S.SH-.........A........V.............I............T.......
VF1_Clone_1 121 A..S.SH-.........A........V.............I............T.......
VF1_Clone_2 121 A..S.---.........A........V.............I............T.......
VF1_Clone_3 121 AAES.---.........A........V.............I............T.......
VF1_Clone_4 121 A..S.SH-.........A........V.............I............T.......
VF2_Clone_1 121 A..S.SH-.........A........V.............I............T.......
VF2_Clone_2 121 AAES.---.........A........V.............I............T.......
VF2_Clone_3 121 AAES.---.........A........V.............I............T.......
consensus   121 ....*...*********.*******.***********.**************.*******

HXB2        181 PQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTT
BL_Clone_1  180 .....M...I...........D...........T.......VP.............
BL_Clone_2  180 .....M...I...........D...........T.......VP.............
BL_Clone_3  180 .....M...I...........D...........T.......VP.............
BL_Clone_4  180 .....M...I...........D...........T.......VP.............
VF1_Clone_1 180 .....M...I...........D...........T.......VP.............
VF1_Clone_2 178 .....M...I...........D...........T.......VP.............
VF1_Clone_3 178 .....M...I...........D...........T.......VP.............
VF1_Clone_4 180 .....M...I...........D...........T.......VP.............
VF2_Clone_1 180 .....M...I...........D...........T.......VP.............
VF2_Clone_2 178 .....M...I...........D...........T.......VP.............
VF2_Clone_3 178 .....M...I...........D...........T.......VP.............
consensus   181 *****.***.***********.***********.*******..***************

HXB2        241 STLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
BL_Clone_1  240 ..I........S...V.........V............V.....K........E.....
BL_Clone_2  240 ..I........S...V.........V............V.....K........E.....
BL_Clone_3  240 ..I........S...V.........V............V.....K........E.....
BL_Clone_4  240 ..I........S...V.........V............V.....K........E.....
VF1_Clone_1 240 ..I........S...V.........V............V.....K........E.....
VF1_Clone_2 238 ..I........S...V.........V............V.....K........E.....
VF1_Clone_3 238 ..I........S...V.........V............V.....K........E.....
VF1_Clone_4 240 ..I........S...V.........V............V.....K........E.....
VF2_Clone_1 240 ..I........S...V.........V............I.....K........E.....
VF2_Clone_2 238 ..I........S...V.........V............V.....K........E.....
VF2_Clone_3 238 ..I........S...V.........V............V.....K........E.....
consensus   241 **.********.***.**********.***********.*****.********.*****
```

```
HXB2        301 YKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKA
BL_Clone_1  300 F........T........................S..R....G................S...
BL_Clone_2  300 F........T........................S..R....G................S...
BL_Clone_3  300 F........T........................S..R....G................S...
BL_Clone_4  300 F........T........................S..R....G................S...
VF1_Clone_1 300 F........T........................S..R....G................S...
VF1_Clone_2 298 F........T........................S..R....G................S...
VF1_Clone_3 298 F........T........................S..R....G................S...
VF1_Clone_4 300 F........T........................S..R....G................S...
VF2_Clone_1 300 F........T........................S..R....G................S...
VF2_Clone_2 298 F........T........................S..R....G................S...
VF2_Clone_3 298 F........T........................S..R....G................S...
consensus   301 .********.********************.**.****.***************.***

HXB2        361 RVLAEAMSQVTNSATIMMQRGNFRNQRKIVKCFNCGKEGHTARNCRAPRKKGCWKCGKEG
BL_Clone_1  360 ..........QQT-N....K....G.KT.-..........L.........R.........
BL_Clone_2  360 ..........QQT-NV...K....G.KT.-..........L.........R.........
BL_Clone_3  360 ..........QQT-N....K....G.KT.-..........L.........R.........
BL_Clone_4  360 ..........QQT-N....K....G.KT.-..........L.........R.........
VF1_Clone_1 360 ..........QQT-NV...K....G..T.-..........L.........R.........
VF1_Clone_2 358 ..........QQT-N....K....G..T.-..........L....K....R.........
VF1_Clone_3 358 ..........QQT-N....K....G.KT.-..........L.........R.........
VF1_Clone_4 360 ..........QQT-NV...K....G..T.-..........L.........R.........
VF2_Clone_1 360 ..........QQT-NV...K....G..T.-..........L.........R.........
VF2_Clone_2 358 ..........QQT-N....K....G..T.-..........L....K....R.........
VF2_Clone_3 358 ..........QQT-N....K....G..T.-..........L....K....R.........
consensus   361 **********......***.****.*...*.**********.****.****.*********

HXB2        421 HQMKDCTERQANFLGKIWPSYKGRPGNFLQSRPEPTAPPEESFRSGVETTTPPQK---QE
BL_Clone_1  418 ......................N.......P..........A..LGMEG...PS.KQ----.
BL_Clone_2  418 ......................C.......P..........A..LGMEG...PS.KQ----.
BL_Clone_3  418 ......................N.......P..........A..LGMEG...PS.KQ----.
BL_Clone_4  418 ......................N.......P..........A..LGMEG...PS.KQ----.
VF1_Clone_1 418 ......................N.......P..........A..LGMEG...PS.KQ----.
VF1_Clone_2 416 ......................N.......P..........A..LGMEG...PS.KQ----.
VF1_Clone_3 416 ......................N.......P..........A..LGMEG...PA.KQ----.
VF1_Clone_4 418 ......................S.......P..........A..LGMEG...PA.KQ----.
VF2_Clone_1 418 ......................S.......P..........A..LGMEG...PS.KQ----.
VF2_Clone_2 416 ......................S.......P..........A...GM.E.I.PS.KQELRE.
VF2_Clone_3 416 ......................S.......P..........A...GM.E.I.PS.KQ----.
consensus   421 ******************.*******.**********.**......*.*...*......*

HXB2        478 PIDKELYPL---
BL_Clone_1  474 .G..G...PLT-
BL_Clone_2  474 .G..G...PLT-
BL_Clone_3  474 .G..G...PLT-
BL_Clone_4  474 .G..G...PLT-
VF1_Clone_1 474 .G..G...PLT-
VF1_Clone_2 472 .G..G...PLTX
VF1_Clone_3 472 .G..G...PLTX
VF1_Clone_4 474 .G..G...PLT-
VF2_Clone_1 474 .G..G...PLT-
VF2_Clone_2 476 .R..G...PLT-
VF2_Clone_3 472 .G..G...PLTX
consensus   481 *.**.***....
```

Figure 6.2 (a): Full length Gag Multiple Sequence Alignment of all clonal variants from patient 5 at the three timepoints.

Sequence alignment of the clonal variants derived from each of the three timepoints. Red boxes denote the amino acid positions of interest which appeared to vary between clonal variants and which was further investigated for their roles in phenotypic susceptibility tests.

```
HXB2            1 PQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
BL_Clone_1      1 ..I.........VM....PI...............ID....................K...
BL_Clone_2      1 ..I.........VM....PI...............ID........................
BL_Clone_3      1 ..I.........VM....PI...............ID....................K...
BL_Clone_4      1 ..I.........VM....PI...............ID....................K...
VF1_Clone_1     1 ..I.........VM....PI...............ID........................
VF1_Clone_2     1 ..I......V..VM....PI...............ID........................
VF1_Clone_3     1 ..I.........VM....PI...............ID........................
VF1_Clone_4     1 ..I........AVI....PI...............ID........................
VF2_Clone_1     1 ..I.........VM....PI...............ID........V...............
VF2_Clone_2     1 ..I.........V.....PI...............ID........................
VF2_Clone_3     1 ..I.........VM....PI...............ID........................
consensus       1 **.******.*...****..**************..********.**********.***

HXB2           61 QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
BL_Clone_1     61 ..P.....K....................M.........
BL_Clone_2     61 ..P.....K....................M.........
BL_Clone_3     61 ..P.....K....................M.........
BL_Clone_4     61 ..P.....K....................M.........
VF1_Clone_1    61 ..P.....K....................M.........
VF1_Clone_2    61 ..P.....K....................M.........
VF1_Clone_3    61 ..P.....K....................M.........
VF1_Clone_4    61 ..P.....K....................M.........
VF2_Clone_1    61 ..P.....K....................M.........
VF2_Clone_2    61 ..P.....K....................M.........
VF2_Clone_3    61 ..P.....K....................M.........
consensus      61 **.*****.******************.**********
```

Figure 6.2 (b): Protease Multiple Sequence Alignment of all clonal variants from patient 5 at the three timepoints.
The occurrence of M46V accessory mutation in VF2_Clone 1 is highlighted in yellow

6.3.1 The M46V accessory mutation in protease played no role in reduced susceptibility

Our interest is in resistance to protease inhibitors. Hence any major or accessory resistance mutations in protease were noted. In all the clones at different time points, there were no PI major resistance mutations in protease. There was only one PI accessory resistance mutation (M46V) which was seen in the more 'susceptible' clone at virological failure timepoint 2 (VF2clone1). Additionally, the L10V mutation developed in one of the VF1 clones (VF1clone2). L10V is a polymorphic accessory PI-selected mutation that either reduces PI susceptibility or increases the replication of viruses containing PI-resistance mutations [352,597]. All the other polymorphisms such as I13V, K14M, K20I, M36I (as shown in figure 6.2b) are consensus amino acid in most of the non-B subtypes and were not explored further.

Given that the M46V accessory mutation has not been well studied and was the only PI accessory resistance mutation seen, we used SDM to revert valine to methionine at protease position 46 in the VF2clone1 virus, and conversely mutated methionine to valine in the VF2clone2 virus and tested the retroviral vectors phenotypically. The result is presented in figure 6.3, and it shows that the M46V did not affect the susceptibility of this virus.

Having established that the only accessory protease mutation played little role in PI resistance, we hypothesized that mutations in Gag are driving the resistance observed in some of the clonal variants.
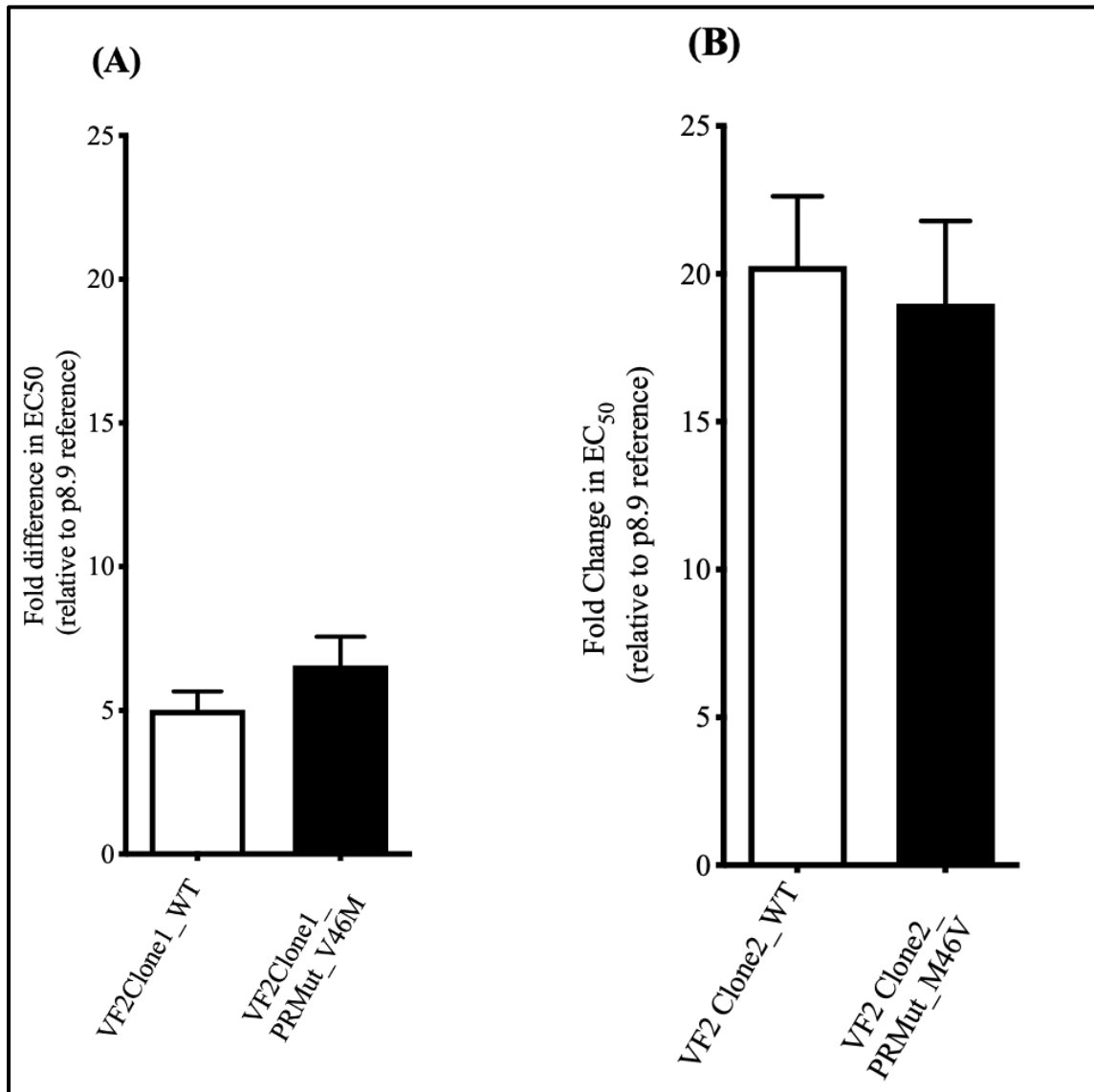
Figure 6.3   Minor protease mutation (M46V) did not affect LPV susceptibility

(A) M46V accessory mutation in protease was reverted in the more 'susceptible (and called VF2Clone1_PRMut_V46M)' clone (B) M46V protease minor mutation was introduced in the resistant clone (and called VF2Clone2_PRMut_M46V). Pseudovirions were produced by co-transfection in 293T cells and their PI susceptibility tested using a single replication-cycle assay by measuring luciferase activity. Data shows fold difference in EC50 between the WT and mutant clones. Error bars represent the standard error of the mean of at least three independent experiments carried out in duplicate. Reverting or creating the M46V accessory mutation did not have an effect on LPV susceptibility.

241

6.3.2 The G123E MA mutation alone confers a significant reduction in LPV susceptibility

We next sought to establish the effect of each of the four amino acid changes occurring in the 'resistant' viral clone (VF2clone2). Four different mutant viruses were created with single amino acid changes at Gag: 122 (A→T), 123 (E→G), 126S insertion and 127H insertion. These four individual mutants were phenotypically assayed. Results of the phenotypic drug susceptibility testing of these mutants are shown in figure 6.4. Reverting glutamic acid to glycine at Gag position 123 (E→G) had the biggest impact on LPV susceptibility. However, none of the mutation reversions alone increased LPV susceptibility as shown in Figure 6.4.

Figure 6.4: Role of single amino acid changes on LPV susceptibility
(A) shows the matrix amino acid residues at the positions being investigated. Amino acid residues shown in red letterings are the mutants created by site directed mutagenesis. (B) shows the fold difference in EC_{50} for LPV of each of the site-directed mutants determined using a single round infectivity assay utilizing luciferase as a read out. Δ denotes amino acid deletion. Error bars represent standard error of mean of two independent experiments carried out in duplicates.

6.3.3 The insertion of four amino acids (E-L-R-E) in p6 domain of Gag did not affect susceptibility

Next, we sought to determine the role of the four amino acid insertion in the p6 domain. Using standard site directed mutagenesis techniques, amino acids E, L, R and E were sequentially inserted into the 'susceptible' clone between positions 477 and 478 (HXB2 475 and 476) in the p6 domain (and called VF2clone1_SuscMut2). Conversely, E, L, R and E residues were deleted in 'resistant' clone from the same location (and called VF2clone1_ResMut2. There was no significant change in phenotypic susceptibility to LPV as a result of the ELRE insertion (in the 'susceptible' clone) or deletion (in the 'resistant' clone (Figure 6.5).

6.3.4 Deletion of S126 and H127 residues near the MA/CA cleavage site confer reductions in LPV susceptibility

In section 6.3.2 above, none of the amino acid mutations alone completely restored LPV susceptibility. We, therefore, introduced the mutations in combinations, rather than singly.

First, we explored the possible role of Serine and Histidine deletion at positions 126 and 127.

Using site-directed mutagenesis, Serine (Gag position 126) and Histidine (Gag position 127) residues were deleted in the 'susceptible' clone (and called VF2clone1_SuscMut1). Conversely, Serine and Histidine residues were inserted in the 'resistant' clone (and called VF2clone2_ResMut1).

Deletion of Serine 126 and Histidine 127 in the 'susceptible' virus, led to a significant decrease in LPV susceptibility of the VF2clone1_SuscMut1 mutant virus (Figure 6.5). Conversely, the insertion of Serine and Histidine residues in the 'resistant' virus increased susceptibility of the VF2clone2_ResMut1 mutant (Figure 6.5). However, the changes at positions 126 and 127 did not completely account for the differences in LPV susceptibility when compared to the WT clones.

6.3.5 The matrix deletions at S126 and H127 act synergistically with T122A and G123E mutations to confer reduced susceptibility to lopinavir

Next, we introduced two more mutations in the mutants above by introducing a T122A and G123E mutations in VF2clone1_SuscMut1 (we called this mutant VF2clone1_SuscMut3) and conversely reverting the same amino acids in ResMut1 (and called this mutant VF2clone2_ResMut3).

With these set of mutagenesis, the mutant of the 'susceptible' clone was now genetically similar to the WT 'resistant' virus; conversely, the mutant of the 'resistant' clone was now genetically similar to the WT 'susceptible' virus. We then carried out phenotypic drug susceptibility of these mutants.

A combination of S126del, H127del and T122A, G123E mutations in the 'susceptible' virus led to a 4x decrease in susceptibility to LPV (FC EC_{50} from 5.3 to 22.7), figure 6.6b. Conversely, 126S, 127H insertions, and 122 (A→T), 123 (E→G) in the 'resistant' virus led to a three-fold decrease in resistance as shown in figure 6.6C.

The phenotype was completely restored upon the mutation (or reversion) of the four amino acids. We, therefore, reached the preliminary conclusion that these four Gag amino acid combinations act in synergy to confer resistance to the protease inhibitor, LPV.

A (1)

	HXB2 Pos	120	121	122	123	124	125	126	127	128	129	130	131	132	//	473	474	475					477	478	479	480	481	482	483
Susceptible' virus	VF2clone1_WT	A	A	T	G	S	S	S	H	-	S	Q	N	Y		P	K	Q	-	-	-	-	E	P	R	D	K	G	L
	VF2clone1_SuscMut1	A	A	T	G	S	S	-	-	-	S	Q	N	Y		P	K	Q	-	-	-	-	E	P	R	D	K	G	L
	VF2clone1_SuscMut2	A	A	T	G	S	S	-	-	-	S	Q	N	Y		P	K	Q	E	L	R	E	E	P	R	D	K	G	L
	VF2clone1_SuscMut3	A	A	A	E	S	S	-	-	-	S	Q	N	Y		P	K	Q	E	L	R	E	E	P	R	D	K	G	L

A (2)

	HXB2 Pos	120	121	122	123	124	125	126	127	128	129	130	131	132	//	473	474	475					477	478	479	480	481	482	483
Resistant' virus	VF2clone2_WT	A	A	A	E	S	S	-	-	-	S	Q	N	Y		P	K	Q	E	L	R	E	E	P	R	D	K	G	L
	VF2clone2_SuscMut1	A	A	A	E	S	S	S	H	-	S	Q	N	Y		P	K	Q	E	L	R	E	E	P	R	D	K	G	L
	VF2clone2_SuscMut2	A	A	A	E	S	S	S	H	-	S	Q	N	Y		P	K	Q	-	-	-	-	E	P	R	D	K	G	L
	VF2clone2_SuscMut3	A	A	T	G	S	S	S	H	-	S	Q	N	Y		P	K	Q	-	-	-	-	E	P	R	D	K	G	L

Figure 6.5: Gag 126del and 127del mutations occurring with T122A and G123E confers resistance to inhibitor lopinavir in the absence of any major protease mutations and with the insertion of four amino acids (ELRE) in p6 not playing a significant role.

(A) Sequences of the viral clones showing the amino acid changes (in red) introduced using standard site directed mutagenesis. (A1) shows the sequential site-directed mutation amino acid residues in the 'susceptible' clone, to mimic the 'resistant' clone while (A2) shows the sequential site-directed mutation amino acid residues in the 'resistant' clone to mimic the 'susceptible' clone – with amino acids introduced or deleted via SDM shown in red letterings. (B) and (C) Full-length Gag-protease with indicated mutations was amplified from plasma samples and cloned into p8.9NSX+. VSV-g pseudotyped viruses encoding luciferase were produced by co-transfection in 293T cells. PI susceptibility of pseudovirions derived from each patient was determined using a single replication-cycle drug susceptibility assay as measured by luciferase activity. Data displayed are fold difference in EC_{50} values of LPV in comparison to that of the assay reference strain, p8.9NSX+. Lopinavir susceptibility decreases with sequential mutation of amino acid residues in the 'susceptible' clone (B) while susceptibility increases with sequential mutation of amino acid residues in the 'resistant' clone (C). Error bars represent the standard error of the mean of three independent experiments performed in duplicate.

6.3.3 Phenotypic susceptibility to Darunavir was unaffected by the four amino acid mutations

Having established the role of these four amino acid mutation signatures in reduced susceptibility to LPV, we also tested the effect on susceptibility to the second-generation PI darunavir (DRV) using the viral mutants and the results are shown in Figure 6.7. The four amino acid mutations did not appear to have any significant impact on DRV susceptibility.

								MA														p6							
HXB2 Pos	120	121	122	123	124	125	126	127	128	129	130	131	132	473	474	475					477	478	479	480	481	482	483
VF2clone1_WT	A	A	T	G	S	S	S	H	-	S	Q	N	Y	P	K	Q	-	-	-	-	E	P	R	D	K	G	L
VF2clone1_SuscMut1	A	A	T	G	S	S	**-**	**-**	**-**	S	Q	N	Y	P	K	Q	-	-	-	-	E	P	R	D	K	G	L
VF2clone1_SuscMut2	A	A	T	G	S	S	-	-	-	S	Q	N	Y	P	K	Q	**E**	**L**	**R**	**E**	E	P	R	D	K	G	L
VF2clone1_SuscMut3	A	A	**A**	**E**	S	S	-	-	-	S	Q	N	Y	P	K	Q	E	L	R	E	E	P	R	D	K	G	L

								MA														p6							
HXB2 Pos	120	121	122	123	124	125	126	127	128	129	130	131	132	473	474	475					477	478	479	480	481	482	483
VF2clone2_WT	A	A	A	E	S	S	-	-	-	S	Q	N	Y	P	K	Q	E	L	R	E	E	P	R	D	K	G	L
VF2clone2_SuscMut1	A	A	A	E	S	S	**S**	**H**	-	S	Q	N	Y	P	K	Q	E	L	R	E	E	P	R	D	K	G	L
VF2clone2_SuscMut2	A	A	A	E	S	S	S	H	-	S	Q	N	Y	P	K	Q	**-**	**-**	**-**	**-**	E	P	R	D	K	G	L
VF2clone2_SuscMut3	A	A	**T**	**G**	S	S	S	H	-	S	Q	N	Y	P	K	Q	-	-	-	-	E	P	R	D	K	G	L

Figure 6.6: Gag 126del and 127del mutations occurring with T122A and G123E did not have a significant effect on susceptibility to darunavir

A) Sequences of the viral clones showing the amino acid changes (in red) introduced using standard site-directed mutagenesis. (A1) shows the sequential site-directed mutation amino acid residues in the 'susceptible' clone, to mimic the 'resistant' clone. (A2) shows the sequential site-directed mutation amino acid residues in the 'resistant' clone to mimic the 'susceptible' clone. (B) and (C) Full-length Gag-protease with indicated mutations was amplified from plasma samples and cloned into p8.9NSX+. VSV-g pseudotyped viruses encoding luciferase were produced by co-transfection in 293T cells. PI susceptibility of pseudovirions derived from each patient was determined using a single replication-cycle drug susceptibility assay as measured by luciferase activity. Data displayed are fold difference in EC_{50} values of DRV in comparison to that of the assay reference strain, p8.9NSX+. Sequential mutation of the 'susceptible' viral clone and sequential reversion of amino acid residues in the 'resistant' viral clone did not have any effect on darunavir susceptibility. Error bars represent the standard error of the mean of three independent experiments performed in duplicate.

6.3.4 Introducing the four MA mutant signature amino acids into HIV-1 subtype B reduced susceptibility to ATV and LPV, but not DRV

We next tested whether the four amino acid signatures: T122A/ G123E/ S126del, H127del could confer resistance to PIs in a different subtype context. We chose the reference p8.9NSX subtype B virus and created the amino acid deletions at Gag positions 126 and 127 as well as the adjacent T122A and G123E mutations. Also, we deleted V128 given that subtype CRF02_AG consensus contains this deletion as compared to subtype B. The phenotypic susceptibility assay results of these mutations are shown in Figure 6.7.

Interestingly, and contrary to our observation in the CRF02_AG virus, the deletions of Asparagine and Glutamine residues from positions 126 and 127 respectively appeared to make the subtype-B virus slightly more susceptible to all three PIs. The combination of T122A and G123E, however, conferred a 2.6-fold reduction in LPV susceptibility, 2.9-fold reduction in ATV susceptibility, with no change in DRV susceptibility.

However, the five mutations (T122A/ G123E/ N126del/ Q127del/V128del) reduced susceptibility to LPV and ATV by 3.4-fold and 3.1-fold, respectively, with minimal effect on DRV. This would suggest that these MA mutations, occurring in synergy, are effective in a divergent HIV-1 subtype (Figure 6.7).
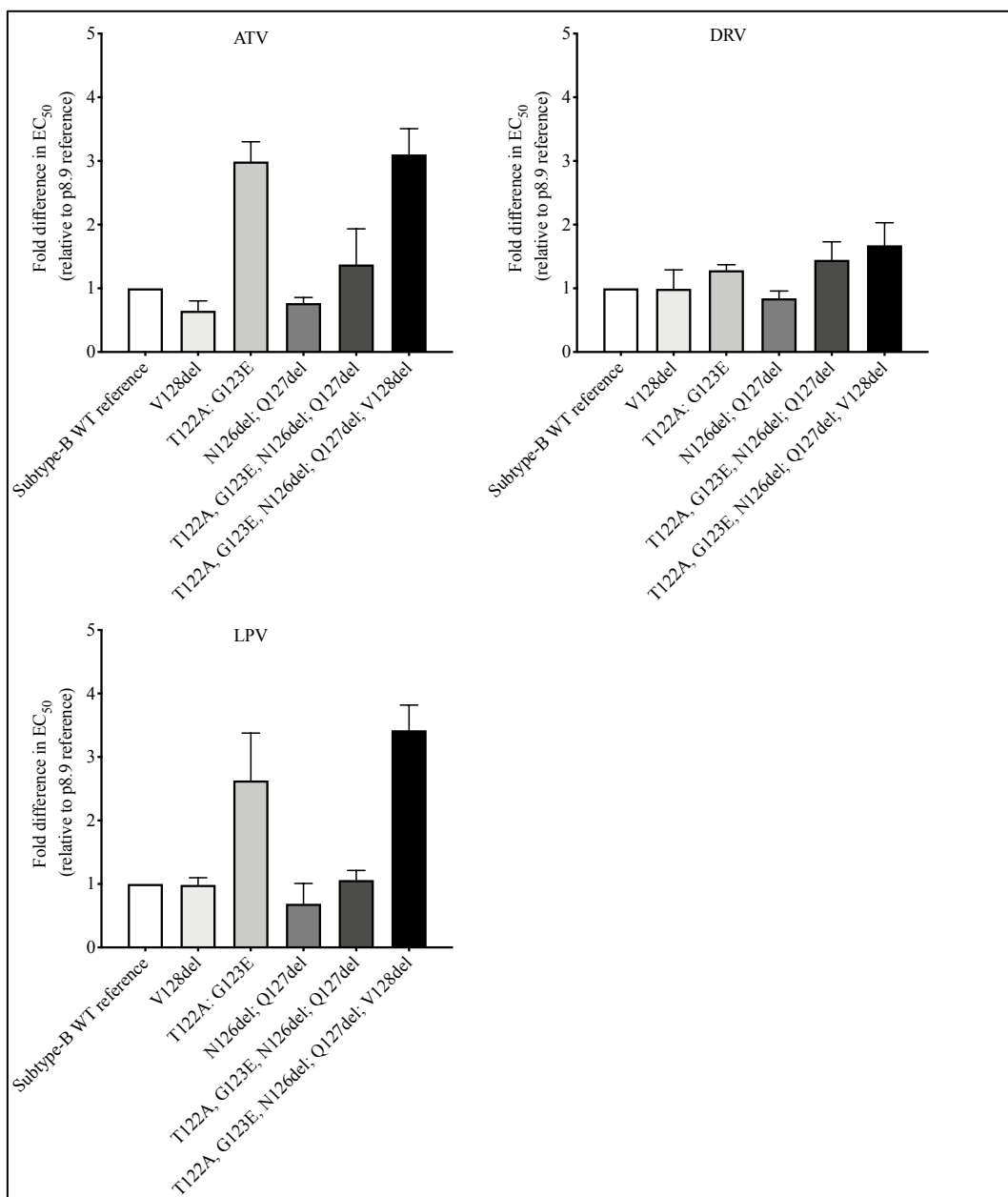
Figure 6.7: The four amino acid MA mutant signature introduced into subtype B reduces LPV and ATV susceptibility

Site directed mutants were generated in the subtype B reference strain used in our assays. The V128del was also added as this deletion is present in HIV-1 CRF02_AG. Data displayed are fold difference in EC_{50} values in comparison to that of the assay reference strain, p8.9NSX. Susceptibility to atazanavir (A) and lopinavir (C) reduced by up to 3x with the introduction of the mutations. Susceptibility to darunavir (B) was largely unaffected by the amino acid mutations. Error bars represent the standard error of the mean of three independent experiments performed in duplicate.

6.3.5 The resistance signature arises from a minority viral population detected at baseline

We proceeded to ask the question of when the resistance emerged. NGS analysis at whole-genome level was undertaken for all 3-time points, and table 6.2 shows variant frequencies at sites in Gag and Pol associated with drug exposure. Of note, we observed loss of mutations to lamivudine (**M184I**), tenofovir (**K65R**) and efavirenz (**K103N**) between baseline and VF1. The individual was prescribed lamivudine, zidovudine and lopinavir/ritonavir for second-line and the resistance data indicate lack of drug pressure from lamivudine.

The NGS showed that T122A/ G123E were present at low abundance before initiation of PI (approx. 5% of reads, Table 6.2). The proportion of T122A/ G123E increased at VF1 to 13%. These mutations were observed at an increased frequency at VF2 both by target enriched NGS and also direct Gag-Pro PCR from plasma, but NGS also showed the emergence of lamivudine resistance mutant M184V, suggesting improved adherence to lamivudine between VF1 and VF2.

We next generated whole-genome haplotypes for each time point using NGS data in order to firstly establish the phylogenetic relationships between viruses with differing PI resistance-associated mutations, and also to determine the co-receptor usage of virus haplotypes as this might provide clues as to origins of virus variants (Figure 6.8). For geno2pheno co-receptor usage prediction, we used the optimal thresholds of 5% based on clinical data from the MOTIVATE study[598]. In addition, to geno-2-pheno, we also used the WebPSSM algorithm (https://indra.mullins.microbiol.washington.edu/webpssm/).

All inferred haplotypes were predicted to use CCR5 with FPR of >20%, and no CXCR4 using viruses were predicted in either of the two algorithms used.

None of the four Gag-pro clones from baseline (before initiation of PI) contained any of the four amino acid changes T122A/ G123E/ S126del/ H127del, consistent with NGS data showing these variants were present at <5% (table 6.2). Clones from the intermediate time point (VF1) clustered with the VF2 clones rather than with the baseline clones (Figure 6.9). Overall, there was excellent concordance

between the inferred whole genome haplotypes and Gag-pro clones, though there appeared to be greater diversity in haplotypes.

In vitro phenotypic drug susceptibility of cloned sequences revealed both sensitive and resistant viruses at VF1 as well as VF2 (Figure 6.9), with the resistant clones from VF1 and VF2 clustering together and sharing the 4 amino acid resistance associated signature S126del, H127del and T122A, G123E. As expected, the susceptible clones from VF1 and VF2 time points also clustered with each other in a distinct part of the tree.

Table 6.2: NGS variant derived data for three time points during LPV treatment

.

Gene	Mutation	Baseline (0 months) 405,158 reads	VF1 (41 months) 250,932 reads	VF2 (64 months) 604,157 reads
Gag	E12K	5%	21%	42%
Gag	R76K	0	0	2%
Gag	Y79F	0	0	2%
Gag	T122A	4.8%	13%	26%
Gag	G123E	5.0%	13%	26%
Gag	V128del	100%	100%	100%
Gag	V370A	5%	0	1%
Gag	S373T	97%	98%	100%
Gag	R409K	3%	0	1%
Gag	S451T	100%	1.7%	0
RT	K65R	98%	0	0
RT	K103N	94%	0	0
RT	E138K	0	0	4%
RT	Y181C	100%	0	0
RT	M184I	100%	1.2%	21%
RT	M184V	0	0	79%

The table shows the percentage of reads encoding resistance-associated mutations in RT and Gag mutations known to be associated with protease inhibitor exposure from prior reports
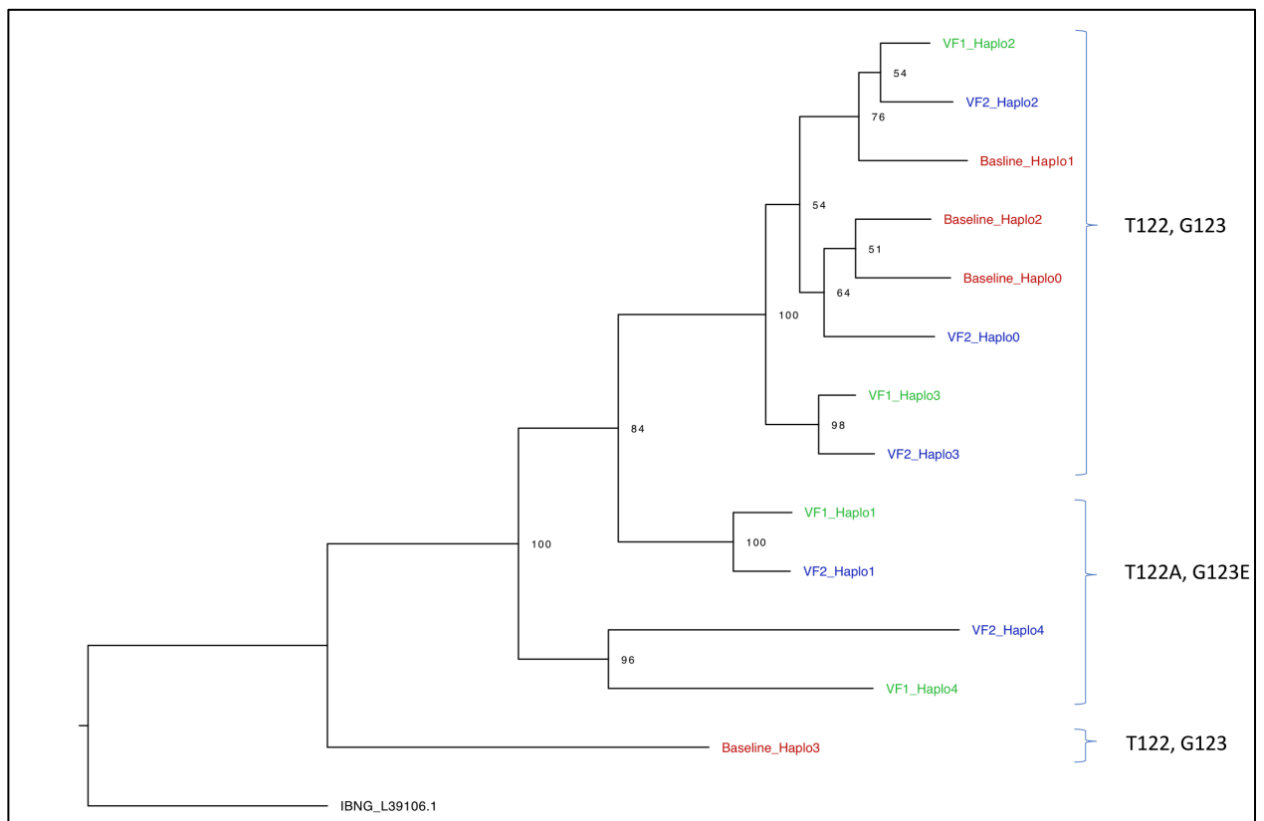
Figure 6.8: Whole-genome HIV haplotype reconstruction
Haplotype reconstruction was performed using target enriched NGS Illumina MiSeq data from each time point (baseline, VF1 and VF2), with maximum likelihood analysis and bootstrap support indicated using 1000 replicates. Labelled on the right side are the amino acids at positions Gag 122 and 123. Haplotypes: BL=red; VF1=green; VF2=blue
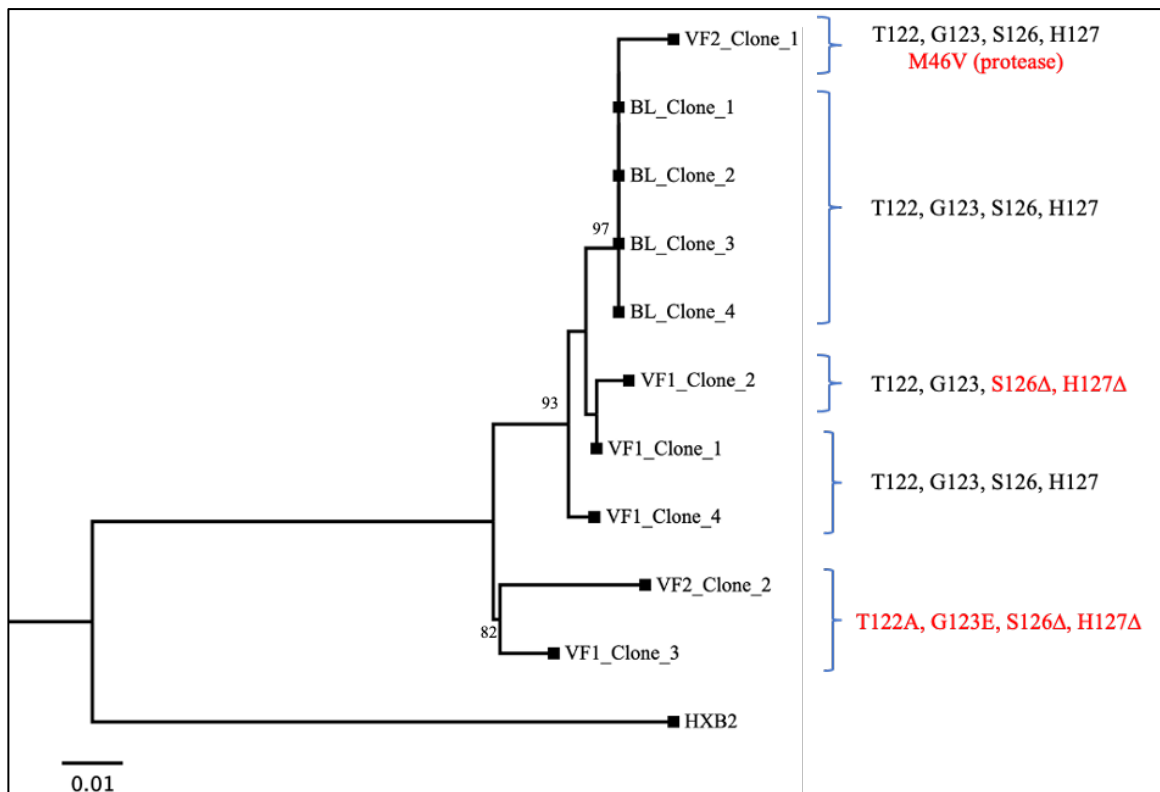
Figure 6.9: Phylogenetic relationships between viral clones isolated at the three different timepoints

Sequences isolated at baseline (pre-PI) and at two failure time points (VF1 and VF2). Illustrated is a maximum likelihood tree with bootstraps indicated. Outlier indicated is HXB2, a subtype B virus.

6.3.6 Persistence of both resistant and susceptible viruses can be explained by replication capacity

A surrogate for fitness in our assay is single round infectivity (measured in RLU) in the absence of drug, which is given a value of 100% for our reference subtype B virus. We measured the single-round infectivity (replication capacity, RC) of clones bearing patient-derived Gag-protease sequences from each time point. Interestingly, resistant clones (VF1clone_3 and VF2clone_2) had around 2-fold lower RC than susceptible viruses (VF2clone_1 and VF1clone_1), regardless of whether they were isolated from VF1 or VF2 (Figure 6.10). The mixture of sensitive and resistant strains could be due to incomplete drug adherence and therefore, variable drug pressure, or with the compartmentalisation of virus sequences in anatomical areas with differing drug levels.
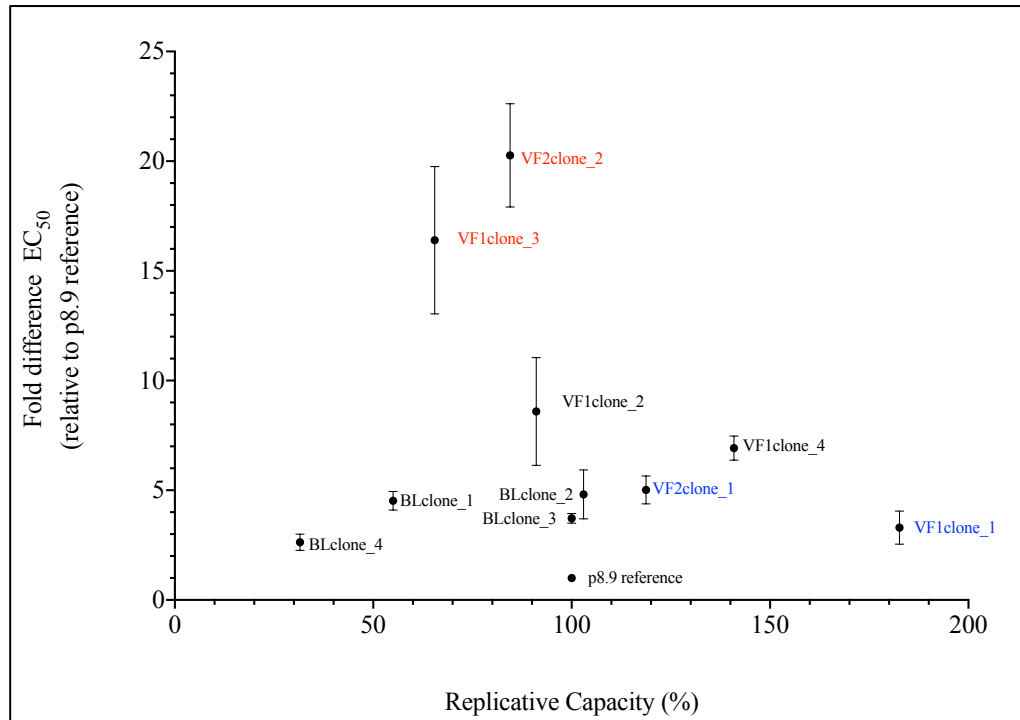
Figure 6.10: Relationship between single round infectivity (RC) and LPV susceptibility of patient derived viruses

The single-round infectivity of pseudovirions produced from co-transfection of 293T cells was determined by titration of viruses using a single-cycle of infection. Luciferase activity was measured using SteadyGlo and expressed as a fold difference in comparison with that of the assay reference strain, p8.9NSX+. Susceptible clones (shown in blue) had about 2-fold higher replicative capacity compared to the resistant clones (shown in red). Error bars represent the standard error of the mean of three independent experiments performed in duplicate.

6.3.7 The signature mutations are not common in CRF02_AG and G subtype viruses

In order to ascertain how novel, the occurrence of these mutations is, we obtained all the HIV-1 Gag sequences up to 2018 from the Los Alamos database (https://www.hiv.lanl.gov/cgi-bin/NEWALIGN/align.cgi). We carried out multiple sequence alignment and submitted 125 CRF02_AG Gag sequences from both naïve and treatment-experienced patients into The ConSurf Server, a server designed for the identification of functional regions in proteins (https://consurf.tau.ac.il)[599]. The amino acid variability in CRF02_AG viruses in the positions of interest is shown in Table 6.1.

The results show that Gag position 122 is relatively conserved in this subtype with the T→A mutation occurring in 11.2% of all sequences and with 83% of sequences having the amino acid threonine (T) at that position. Position 123 is more conserved in comparison as 93.6% of all sequences had Glycine (G) in that position, with the G→E mutation accounting for only 2.4% in all sequences. At Gag position 126, 78.4% of all sequences had Serine (S) residues with Cysteine (C) residue in 14.4% of all sequences. Interestingly, 100% of all CRF02_AG viruses had a deletion at Gag position 127. The subtype-B consensus amino acid residue at that position is Glutamine (Q).

In all the viruses from the database, none had the combination of all four amino mutations occurring together, hence we think our observation is novel.

Figure 6.11 shows the overall CRF02_AG variability at that region of the Gag protein (near the MA/CA cleavage site.

Table 6.3: Amino Acid variability in CRF02_AG viruses at Gag positions 122, 123, 126 and 127

Gag pos	A (Ala)	C (Cys)	D (Asp)	E (Glu)	G (Gly)	I (Ile)	L (Leu)	N (Asn)	P (Pro)	Q (Gln)	S (Ser)	T (Thr)	Val (Val)	Other (X)	Max AA (%)
122	**11.2**	-	-	0.8	-	0.8	0.8	-	0.8	-	0.8	83.2	0.8	0.8	T 83.2
123	1.6	-	0.8	**2.4**	93.6	-	-	-	-	-	0.8	-	-	0.8	G 93.6
126	-	14.4	-	-	4.8	-	-	1.6	-	-	78.4	0.8	-	-	S 78.4
127	-	-	-	-	-	-	-	-	-	-	-	-	-	100*	Q 100*

The percentage of occurrence of T122A and G123E are shown in red letterings.
*There was a 100% deletion of amino acid residue at Gag position 127.

Sequence alignment (HIV-1 CRF02_AG reference panel). Column headers (amino acid positions):

Q Q A A A D - - | T G | H S | N Q | V S Q N Y P I V Q N I Q G Q M V H Q A I S P R T L N A

#	Name
1	B.FR.83.HXB2 LAI IIIB BRU.K03455
2	02 AG.CH.09.pBV32.KJ689270
3	02 AG.CH.10.pBV17.KJ689256
4	02 AG.CH.11.pBV10.KJ689250
5	02 AG.CH.11.pBV13.KJ689253
6	02 AG.CI.16.DE00216CI019.MH078536
7	02 AG.CI.x.CI20.L11755
8	02 AG.CI.x.CI51.L11758
9	02 AG.CI.x.CI59.L11759
10	02 AG.CI.x.IC144.L11767
11	02 AG.CM.00.LA10GuVa.KU168265
12	02 AG.CM.01.01CM 0002BBY.AY371122
13	02 AG.CM.01.01CM 0005BBY.AY371123
14	02 AG.CM.01.01CM 0008BBY.AY371124
15	02 AG.CM.01.01CM 0074NY.AY371131
16	02 AG.CM.01.01CM 0131NY.AY371137
17	02 AG.CM.01.01CM 0925MO.AY371134
18	02 AG.CM.01.01CM 1237NG.AY371136
19	02 AG.CM.01.01CM 1475MV.AY371138
20	02 AG.CM.01.01CM 4410HAL.AY371142
21	02 AG.CM.02.02CM 0013BBY.AY371125
22	02 AG.CM.02.02CM 0014BBY.AY371126
23	02 AG.CM.02.02CM 0015BBY.AY371127
24	02 AG.CM.02.02CM 1669LE.AY371139
25	02 AG.CM.02.02CM 1677LE.AY371140
26	02 AG.CM.02.02CM 1901LE.AY371146
27	02 AG.CM.02.02CM 1970LE.AY371128
28	02 AG.CM.02.02CM 2162SA.AY371129
29	02 AG.CM.02.02CM 2348SA.AY371130
30	02 AG.CM.02.02CM 4082STN.AY371141
31	02 AG.CM.04.459 16.MH705136
32	02 AG.CM.07.BS81.JX244948
33	02 AG.CM.07.CM100 17.KU168310
34	02 AG.CM.08.BP00079 LH01.JN687693
35	02 AG.CM.08.DE00208CM001.JX140646
36	02 AG.CM.08.DE00208CM004.JX140647
37	02 AG.CM.10.DE00210CM019.KU749419
38	02 AG.CM.96.96CM25.AF212284
39	02 AG.CM.97.97CM21.AF212288
40	02 AG.CM.97.CM52885.AF377954
41	02 AG.CY.05.CY048.FJ388902
42	02 AG.CY.07.CY206.JF683758
43	02 AG.CY.08.CY240.JF683786
44	02 AG.CY.09.CY247.JF683792
45	02 AG.CY.09.CY252.JF683795
46	02 AG.CY.09.CY256.JF683799
47	02 AG.DE.04.SA.JQ982505
48	02 AG.DJ.91.DJ258 DJ2581.L11763
49	02 AG.ES.05.05SP20 270892.EU342801
50	02 AG.ES.06.06SP7 306121.EU342791
51	02 AG.ES.06.P1261.EU786671
52	02 AG.ES.06.P1423.EU884501
53	02 AG.ES.07.07SP52 367860.EU342810
54	02 AG.ES.07.07SP54 371139.EU342811
55	02 AG.ES.07.07SP57 372125.EU342813
56	02 AG.ES.99.R74 s nt0545 9481.MF157740
57	02 AG.FR.02.AAS638.MH351678
58	02 AG.GB.13.13592 1 37.3.MF109380
59	02 AG.GB.13.13592 1 38.3.MF109381
60	02 AG.GB.13.13774 1 57.3.MF109482
61	02 AG.GB.13.13774 1 78.3.MF109500
62	02 AG.GB.13.15171 1 26.3.MF109657
63	02 AG.GB.13.15228 1 52.3.MF109690
64	02 AG.GB.13.15228 1 64.3.MF109702
65	02 AG.GB.13.15228 1 65.3.MF109703
66	02 AG.GB.13.15228 1 71.3.MF109709
67	02 AG.GB.13.15228 1 83.3.MF109721
68	02 AG.GB.14.13612 1 7.4.MF109432
69	02 AG.GB.14.13659 1 50.3.MF109439
70	02 AG.GB.14.13659 1 60.3.MF109448
71	02 AG.GB.14.14592 1 65.3.MF109559
72	02 AG.GH.03.03GH181AG.AB286856
73	02 AG.GH.03.03GH182AG.AB286857
74	02 AG.GH.03.03GH189AG.AB286862
75	02 AG.GH.03.GHNJ185.AB231895
76	02 AG.GH.90.90GH01.AF212276
77	02 AG.GH.93.93GH05.AF212277
78	02 AG.GH.94.94GH10.AF212280
79	02 AG.GH.95.95GH14.AF212282
80	02 AG.GH.96.96GH16.AF212283
81	02 AG.GH.97.97GH17.AF212285
82	02 AG.GH.97.97GH23.AF212286

Figure 6.11: Multiple Sequence alignment of CRF02_AG Gag sequences obtained from the LANL database

Gag positions of interest are shown in red boxes. None of the four amino acid mutation combination was observed in all CRF02_AG Gag sequences downloaded from the los Alamos database.

6.3.8 Analysis of p17/p24 cleavage site to Explain difference in PI susceptibility

Having established the role of the four Gag amino acid mutations in LPV susceptibility, our next step was to understand the mechanism through which resistance is conferred.

We hypothesised that MA/CA cleavage is impaired as a consequence of the mutations occurring near the MA/CA cleavage site hence there would be a difference between susceptible and resistant clones in the presence of LPV.

To test this hypothesis, we employed western blot analysis where Gag cleavage patterns were examined using the supernatants and cellular extracts of 293T cells transfected with each plasmid in the presence and absence of increasing concentrations of LPV (Figure 6.12). We probed with a polyclonal p24 antibody and as expected there was incomplete cleavage of p24-p2 at higher LPV doses in both virus containing supernatants and the cell extracts. This was consistent with previous data by Dam and co-workers [196].

We calculated ratios of p24/p41 to specifically probe the p17/p24 cleavage site in the vicinity of the four amino acid signatures. Although Gag cleavage processing appeared to be more efficient in the resistant viral clone (at no drug, 10nM, and 30nM) the ratios did not indicate that resistant viruses were able to maintain cleavage at the p17/p24 site more efficiently than the sensitive virus in presence of LPV.
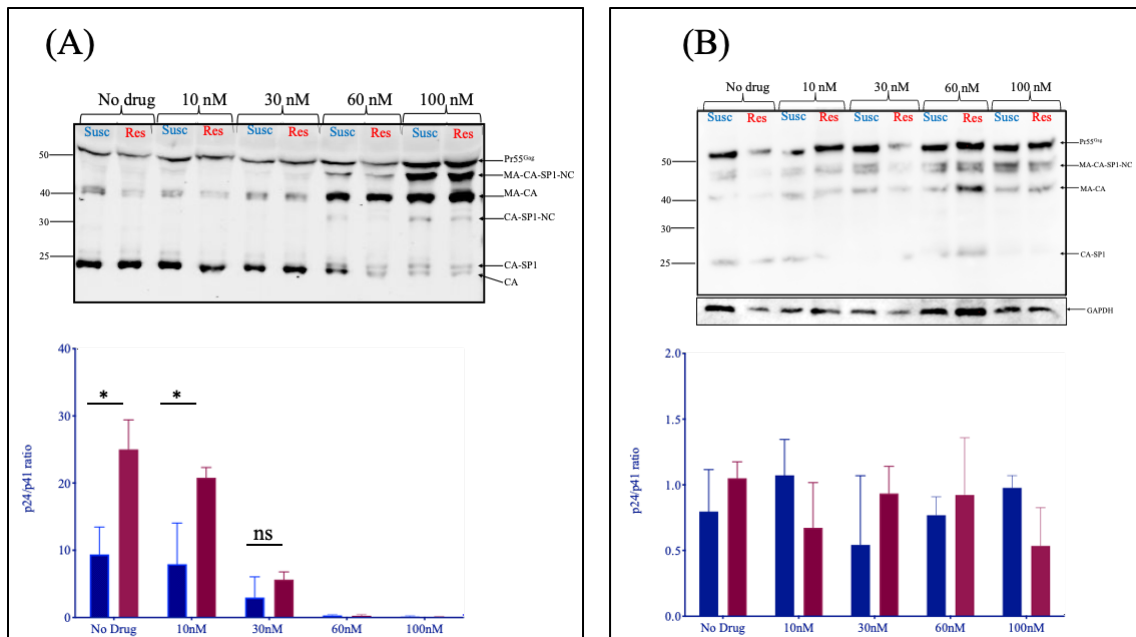
Figure 6.12: HIV-1 Gag cleavage efficiency in resistant (Res) versus susceptible (Susc) isolates.

Representative Western blot of (A) virus-containing supernatant and (B) cell lysates at increasing drug doses, using a p24 antibody. Mass (in kilodaltons) is indicated on the left. MA, matrix (p17); CA, capsid (p24); NC, nucleocapsid; SP1, spacer peptide 1. Bar graphs show ratios of p24/p41 at increasing drug doses. Data are means and standard deviations from 2 independent experiments. In each pair, the blue bars represent the 'susceptible' clone, and the red bar represents resistant virus clone.

6.3.8 Multiple cycle of infection

To further understand the mechanism of resistance due to the four MA amino acid mutations, we wished to undertake a replication assay without and with LPV titration on the 'susceptible' and the 'resistant' virus. A full length, replication-competent HIV-1 vector was required for this experiment. R9-Bal, a full-length, HIV-1 molecular clone R9 with PCR-cloned macrophage-tropic CCR5 using Env gene from HIV-1 Ba-L was utilized. Using the Gibson assembly cloning protocol, we deleted the entire Gag and protease encoding sequences from the vector and successfully cloned the full-length, patient-derived Gag-pro sequences from the 'susceptible' and 'resistant' clones into the R9-Bal vector with cloning success confirmed by sanger sequencing. The chimeric plasmids produced, however, failed to infect TZM-bl target cells after several attempts.

We, therefore, opted for the use of site-directed mutagenesis to introduce the 5 mutations (4 amino acid changes, and V128del) in the R9-Bal plasmid. SDM success was confirmed by Sanger sequencing. This gave rise to 2 plasmids; R9-Bal_WT ('susceptible') and R9-Bal_Mutant ('resistant'). These plasmids were utilized for the multi-round, replication assay (as described in section 6.2.3 above).

In the absence of any drug, at days 2, 4 and 7, there was no difference in infectivity between the wild-type and the mutant viruses. However, there was a 3x difference in infectivity on days 9 and 11 between the wildtype and mutant viruses, with the WT virus showing 3x more infection than the mutant virus as shown in Figure 6.14 (A).

At the end of the experiment, the cells were harvested, pelleted and used for p24 intracellular Gag staining (as described in Chapter 2, section 2.2.9). There was about 2-fold higher percentage infection in the cells infected by the mutant virus as compared to the wild type infected cells, Figure 6.14 (B). This was in contrast to our observation with the infectivity by virus supernatants where the WT virus appeared to infect better than the mutant. This observation may suggest that WT viruses were released more efficiently than the mutants from the infected cells.

Our next step was to repeat the same experiment, at reduced concentrations of LPV (0.25nM, 0.5nM, 1nM, 1.5nM… etc). We hypothesized that the 'resistant' virus (Mutant) would infect more efficiently than the wild-type virus in the presence of the drug at a lower concentration.

This experiment was started and was ongoing but shutdown of the lab facilities due to the COVID-19 pandemic meant that the experiment was prematurely aborted. We however would continue the experiment as soon as lab facilities are re-opened, but that may be outside the time limit of my PhD.
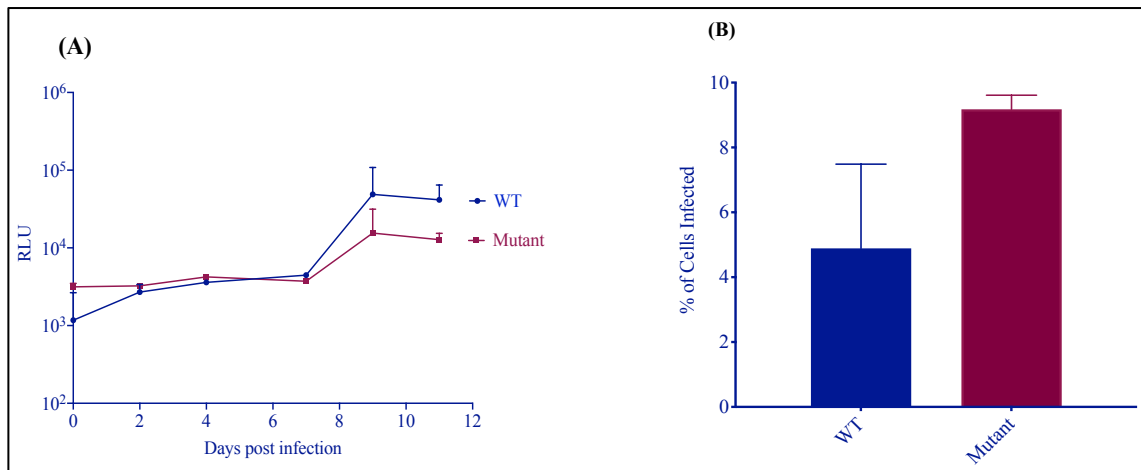
Figure 6.14: Relative infectivity of WT and mutant viruses after 11 days multiple rounds of infection

The relative infectivity of WT and mutant viruses over 11 days is shown. (A) shows a 3x higher infectivity of the wildtype virus at days 9 and 11 (B). Intracellular Gag p24 staining with the mutant virus showing a 2-fold higher infectivity (red bar) compared the WT virus (blue bar).

6.4   Discussion

Based on NMR and X-ray crystallography studies, p17 comprises five major alpha helices connected primarily by short loops [570,600]. The C terminus of matrix is predicted to be disordered, which has hampered efforts to examine the structural characteristics of this region. One study suggested that deletions at 125 and 126 would stabilise p17[601], indicating that despite disorder, changes in the region might lead to significant changes in stability and therefore possibly altered effects of protease inhibition on cleavage.

In our study on CRF02_AG and subtype G clinical isolates from a Nigerian cohort, we demonstrated the role of p17 amino acid mutations occurring near the p17/p24 cleavage site in contributing to PI resistance. The double deletion of Ser and His at Gag 126 and 127 respectively had a modest impact on in-vitro phenotypic PI-susceptibility. When this deletion occurred together alongside T122A and G123E, we observed a 4-5-fold decrease in susceptibility to lopinavir. The four-mutation combination was also able to confer similar resistance to a subtype B virus, indicating that it may play a role across subtypes.

We aimed to understand the mechanism at play in the T122A, G123E, S126del, H127del phenotype. Western blotting of virus-containing supernatants from producer cells, or the cells themselves, did not reveal significant differences in cleavage at the p17-p24 cleavage site with or without the drug. Gag cleavage patterns at the MA/CA cleavage site (as obtained from p24/p41 ratios) were similar between the susceptible and resistant viral clones. Therefore, the rescue of infectivity in the presence of the drug is not explained by a change in the detective cleavage pattern of Gag caused by inhibitor. This suggests that resistance is not mediated by rescuing wild type Gag cleavage patterns but rather by tolerating the changes that the inhibitor drives. This, in turn, suggests that the defect in cleavage may be indirectly related to the mechanism of inhibition and inhibition of infection may be mediated by more subtle effects than simply defects in overall cleavage levels. This more complex mechanism may be particularly crucial at drug concentrations at which defects in Gag cleavage measured by western blot are not apparent.

Also, it is possible that the Gag polyprotein processing, although functionally significant is not easily detected in western blot analysis or that these amino acid mutations can facilitate other functions of Gag (such as budding, packaging, assembly) and these are not easily detected by western blotting – possibly because polyprotein cleavage is not required. It is even possible that these mutations alter some functions of Gag that are yet unknown.

While writing up this thesis, a new study pre-print was published on July 5, 2020 in BioRxiv archives which provided great insights into the possible mechanism of PI resistance arising from non-cleavage site MA and CA mutations. This study by Samsudin and colleagues[602] used multiscale modelling and simulations of the complete oligomeric structure of the largest Gag proteolytic product in its viral membrane-bound state. It revealed how non-cleavage site mutations (such as the four amino acid mutations from our study) could directly interact with cleavage site residues to affect their local environment, facilitated by conformational changes upon lipid interaction.

For virus assembly to occur, HIV-1 Gag is directed to the plasma membrane by Gag polyprotein. The MA domain of the virus drives this process, and it occurs via a phosphatidylinositol-4,5-bisphosphate (PIP2)–dependent myristyl switch mechanism[603,604]. Samsudin and colleagues established that mutations in the MA domain led to the enrichment of PIP2 lipids, and this enrichment was facilitated by the creation of novel PIP2 binding sites[602].

These simulations, therefore, suggested that non-cleavage site mutations enhance interactions between the MA domain and PIP2 lipids and may consequently improve membrane targeting of HIV-1 Gag during viral assembly. Overall, their data suggested that mutations occurring far from the protease binding site, can physically interact with the cleavage site residues, either within the same Gag subunit or with a neighbouring subunit, as the linker region between MA and CA domain contracts following the first proteolytic cleavage. Additionally, the group studied the mature CA hexamer and found that some Gag mutations such as H219Q can modulate the recruitment of cyclophilin A (CypA) into the mature virion, as well as stabilise the oligomerisation of the viral core.

The G123E mutation was reported to arise when viruses were propagated with investigational protease inhibitors KNI-272 and UIC-94003 [422]. Gag G123E was found to potentially interact with protease by NMR[570], providing a potential mechanism for its effect.

Interestingly and importantly, to our own study and findings, Samsudin *et al.,* provided the mechanism by which G123E contributes to PI resistance. Through the use of contact analysis between the MA/CA cleavage site residues and Gag position 123 in a WT (G123) and mutant (E123) proteins, the residue at position 123 was shown to make contact primarily with the N-terminal portion of the cleavage site from the same Gag subunit. Both WT (glycine) and mutant (glutamate) residues showed a similar percentage of contact throughout the simulations. When the CG simulations were "back-mapped" and transformed into atomic resolution, atomistic simulations of a single MA-CA-SP1 sub-unit showed that glutamate residue at position 123 (mutant) interacted primarily with the cleavage site Y132 and also contacted residues N131 and Q130. In contrast, a glycine residue at this position (WT) interacted primarily with V128 and S129. Given the change in the overall size and charge of the residue in the WT and mutants (from small and neutral to large and acidic), the G123E mutation alters the accessibility and electrostatic properties in the vicinity of the cleavage site and therefore was expected to directly interfere with proteolysis. This provides another possible mechanism through which G123E mutation confers resistance to PIs.

Although our present study also implicates G123E in reduced PI susceptibility, we show here that the combination of mutations that were observed in the patient was needed for maximal effect.

We used NGS to explore the dynamics of the emergence of Gag amino acid changes during ongoing viremia under PI treatment. We were able to detect both T122A and G123 at low abundance at baseline, prior to PI exposure. Importantly, PCR from plasma RNA using Gag-protease specific primers did not amplify any sequences with these changes at baseline, highlighting an important contribution of NGS to the study of drug resistance. Whole-genome reconstruction enabled us to infer phylogenetic trees and confirm findings that resistance-conferring

mutations occurred at both time points in phylogenetically related sequences. All virus haplotypes were predicted to be CCR5 using and therefore sensitive to the CCR5 antagonist maraviroc.

In future work, it would be exciting and essential to know whether Gag mutations are capable of facilitating the emergence of major protease mutations in prolonged culture conditions under suboptimal drug pressure. This could potentially explain why the prevalence of major protease mutations increases over time during PI exposure in clinical studies [500]. Next one could perform population dynamics simulations to incorporate RC and susceptibility data in order to model the proportion of resistant and susceptible viruses over time, and possibly, therefore, predict the emergence of major mutations in protease.

Our data are limited by the small sample size, lack of availability of plasma drug level measurements, and by the use of standard clonal approaches as opposed to single genome sequencing and amplification. Nonetheless, we hypothesise that the four amino acid HIV-1 Gag signature is a contributory factor in PI failure in PLWH from Nigeria.

As we move towards next-generation sequencing, this work highlights the limitations of current genotyping methods to infer PI susceptibility and supports sequencing outside protease to broaden the evidence base for the clinical management of patients who experience VF on PIs without major protease mutations. The work may ultimately also help to identify individuals with lower PI susceptibility before treatment with this class of drugs.

CHAPTER 7: General Discussion and Future Work

7.1   Final Discussion

Since the first AIDS cases were reported in the United States in June 1981, the number of cases and deaths among persons with AIDS increased rapidly during the 1980s followed by substantial declines in new cases and deaths in the late 1990s. The prognosis of the disease was significantly improved with the introduction of highly active antiretroviral therapy (HAART). The success of antiretroviral therapy has been attributed to massive rollouts in the most affected parts of the world, especially sub-Saharan Africa. According to WHO/UNAIDS 2020 data, by the end of 2019, out of the estimated 38 million people living with HIV/AIDS, 25.4 million (67%) are accessing treatment.

HIV drug resistance has been reported to increase significantly in several low/middle-income countries in the last few years[605]. Apart from the resistance acquired by an individual over the course of treatment, there has been a steady rise in transmitted drug resistance (TDR). This means newly infected individuals harbour viruses which have already developed resistance in the former host. The spread of TDR can substantially reduce therapeutic choices of ART and increase the chance of virological treatment failure among newly-diagnosed patients on treatment[606]. Consequently, further accumulation of drug-resistant mutations may lead to exhaustion in available drug options; this, in turn, would lead to poorer prognosis for infected persons and transmission of TDR strains to the wider community[605-607].

As per WHO recommendations for LMICs, boosted protease inhibitors are used as the third drug in second-line cART after an individual has failed first-line treatment. PIs are a highly potent class of inhibitors which have reportedly led to significantly fewer patients experiencing virological failure, compared to the use of NNRTIs as the third agent[477]. However, failure during treatment with PIs remains a significant issue. More worryingly, virological failure in patients receiving PI-based therapy often occurs in the absence of major PI resistance

mutations in protease[502,504,608,609], but yet the exact determinants of treatment failure on PIs remain mostly unknown.

Several studies have been conducted to ascertain the virological basis of failure on PIs without major mutations in protease. These studies have shown that mutations in Gag can directly contribute to PI resistance[196,417,437,439,507]. In phenotypic assays, our lab has also shown that when co-evolved Gag alongside its cognate protease is included, it has a direct effect on susceptibility to PIs [397,417]. Furthermore, previous studies, including from our lab, have provided some evidence as to the difference in susceptibility to PIs between subtype-B viruses and CRF02_AG and G sutptypes[398,456,507]. Importantly, studies from our lab on patients on PI-monotherapy had hypothesized that reduced baseline PI susceptibility renders patients more vulnerable to virological rebound, especially when their adherence is not optimal[456].

Earlier studies on the role of the Gag in protease inhibitor susceptibility focused on mutations and polymorphisms occurring at the protease cleavage sites[379,384,400,403,404,406,414,610-612]. Other, and more recent studies have implicated mutations outside of cleavage sites in reduced PI-susceptibility[417,419-424,613].

7.1.1 Summary of findings

The primary goal of this present study was to investigate the role of Gag in protease inhibitor susceptibility amongst West African HIV-1 subtypes. In previous studies from our lab using viruses from patients enrolled in the MONARK trial, it was found that an association exists between reduced PI susceptibility and single-round infectivity at baseline and subsequent virological failure on LPV/r[456]. In addition to that, CRF02_AG and subtype G viruses were found to be substantially less susceptible to lopinavir than subtype B viruses [489]. The potential implication of those findings was that it might be possible to utilize baseline phenotyping to predict eventual PI-based, second-line treatment outcomes.

In the first part of my PhD studies, we utilized archived baseline (pre-PI) plasma samples from a well-defined Nigerian cohort who subsequently failed to achieve virological suppression on boosted PI-based, second-line treatment. We

matched these with baseline samples from patients who subsequently achieved virological suppression in a case-control manner. First, we obtained NGS consensus viral sequences at (20% threshold) from the patient pairs. We analysed the full-length Gag-Pro sequences. Our goal for this analysis was to identify any amino acid residues which are predominant in one group, but not the other. When we examined gag sequences from the two groups and compared these sequences with the subtype-B reference strain (HXB2), we note in both groups the presence of amino acid residues at positions which have previously been reported to be associated with PI exposure or resistance. However, given that these samples were obtained prior to any bPI treatment, our finding in this study adds to the mounting evidence that amino acid mutations which confer resistance to subtype-B viruses may exist as natural polymorphisms in CRF02_AG and G viruses. In the protease gene, we report protease polymorphisms at six sites which were only identified in viruses from patients who subsequently failed bPI treatment. These polymorphisms were: **K14M**, **L19P**, **N37D**, **C68E**, **H69R** and **I72V**. In the gag gene, our current study identified nineteen Gag residues which were only observed in viruses from patients who subsequently failed bPI treatment. These residues are **E12N/R**, **G62T**, **L75F**, **T81A**, **G123E**, **S165N**, **G248T/R**, **Q369K**, **T371N**, **T371del**, **S373Q**, **T375N**, **T375A**, **I376V**, **I376A**, **E468G** and **I479G**. Interestingly, mutations at the position where these Gag residues occurred have previously been reported to be associated with PI exposure or failure[394], but here, they existed as polymorphisms even without PI pressure. Owing to small sample numbers in our study, we could not conclude on the particular Gag polymorphisms or mutations that we can correlate with treatment outcome within our study.

Next, we carried out phenotypic drug susceptibility assays on viruses from these two patient groups. Here, we utilized single-cycle infectivity assays to study drug-response of retroviral vectors (RTVs) bearing patient-derived full-length gag-pro sequences. We tested these RTVs against three second generation PIs: ATV, DRV and LPV. When we aggregated the phenotypic drug assay data, there was no significant difference in the median fold difference in EC_{50} of all three PIs tested between the two groups. Additionally, viruses from both groups showed

similar replication capacity. We however found a moderate correlation between single-round replication efficiency and fold difference in EC_{50} of lopinavir in all viruses tested. Taken together, we found no support for our hypothesis in a role for *in vitro* Gag-protease phenotypic susceptibility testing for PI in predicting virological failure of second line ART.

In the second part of this study, we sought to gain more insight into the role of Gag in patients experiencing PI failure in the absence of protease inhibitor resistance associated mutations in the protease gene. We utilized longitudinal samples from patients who have failed second-line treatment to elucidate the role of Gag amino acid mutations in PI susceptibility. We obtained plasma samples before bPI treatment (baseline (BL)) and samples during bPI treatment, at virological failure (VF) for the same patients. With this longitudinal sampling, we were able to examine the emergence of mutations. We first compared gag and protease sequences of clonal variants at BL from each patient and compared these baseline sequences with sequences of viruses obtained from the same patient at VF. Using an online tool - Fast Unbiased Bayesien AppRoximation (FUBAR), we identified 18 sites in gag which showed evidence of positive selection as a result of PI exposure, with 11 of these sites occurring in the MA domain. Only 5 of these gag positions (62, 76, 81, 441 and 479) have previously been reported as associated with PI resistance or exposure. In the protease gene, 3 sites (14, 15 and 67) were identified as showing evidence of positive selection as a result of PI exposure. We could not establish the emergence of clear genetic determinants of therapy failure across the patients we studied. When we aggregated the phenotypic drug resistance data, there was generally a trend across all three PIs where VF samples were less susceptible than BL samples but this difference was not statistically significant.

In one particular patient, (Patient-5), the phenotypic drug assay results showed reduced PI susceptibility to all three PIs tested. In this patient, at VF, we obtained two distinct viral clones with one clone being less susceptible than the other. Examination of the viral clonal variants from this patient revealed the emergence

of 'unique' non-cleavage site amino acid mutation signature in the MA domain of the less-susceptible viral clone. Additionally, the same less-susceptible clone also had the insertion of four amino acids (E-L-R-E) in the p6 domain. Amino acid insertions in the p6 domain have previously been shown to play a role in resistance to PIs by subtype C viruses[428,431]. In this present study, we found no role for this amino acid insertion in the p6 with respect to treatment failure. We however reported, for the first time, the emergence of **T122A**, **G123E**, **S126del** and **H127del** signature which led to a significant reduction in susceptibility to LPV and ATV. As a proof-of-principle, we introduced these mutations in a subtype-B virus by site-directed mutagenesis, and the results were consistent with our findings in the CRF02_AG virus. Among these mutations, only the G123E has previously been reported as emerging during *in vitro* passage studies [422]. Using our replication capacity data, we explained the persistence of both resistant and susceptible viruses. We reported the resistant clones having 2-fold lower replication capacity compared with the susceptible viral clones.

To understand the origin of for the variant with reduced susceptibility, we utilized the NGS data, which allowed us to examine the viruses at <2% of variant frequency. We traced the emergence of the resistant virus from baseline to virological failure and found that the resistance signature arose from a minority viral population at baseline, with T122A/ G123E present at low abundance (5%) before initiation of PI treatment. Utilizing the NGS data, we used haplotype reconstruction to establish the phylogenetic relationship between the 'resistant' and the 'susceptible' clones obtained from this patient. When all CRF02_AG sequences from the Los Alamos database up to 2018 were aligned, none was shown to harbour these signatures. Hence our conclusion that finding these mutations was novel. To explain the mechanism leading to a difference in susceptibility of the viral clones, we analysed p17/p24 cleavage using western blotting. We showed that high doses of LPV, leads to incomplete cleavage of p24-p2. We showed also that Gag cleavage processing was more efficient in the 'resistant' viral clone. When we utilized full-length virus for multiple rounds of infection, we showed a 3x difference in infectivity on days 9 and 11 between the

wildtype ('susceptible') and mutant ("resistant') viruses, with the WT virus showing 3x more infection than the mutant virus – in the absence of drug.

7.2   Study Limitations and Future Work

The relatively small sample size was the most significant limitation of our study. Initial sample size calculation indicated 20 samples from cases and 20 samples from matched controls are required for 80% power to detect a 1.5-fold difference in phenotypic susceptibility. Although we successfully PCR-amplified up to 17 sample pairs, several cloning strategies and failed attempts limited our eventual sample size.

In the future, more samples would be used for this analysis. It is worth stating here that the findings from this study have served as strong support and providing preliminary data for a 5-year US-National Institute of Allergy and Infectious Diseases (NIAID)-NIH grant (R01AI147331) to our collaborators at the Institute of Human Virology Nigeria. This grant titled: Impact of non-B HIV-1 Subtypes on second-line Protease Inhibitor Regimens in Africa (INSPIRE) would expand on the findings in this project on a larger scale.

Our assay system did not incorporate the native HIV-1 gp160 envelope, instead, it used a VSV-g envelope glycoprotein. While Gag alone can generate virus-like particles (VLPs), for a particle to be infectious requires the expression of the viral enzymes, regulatory proteins, and the envelope (Env) glycoprotein. Given that MA interacts with Env for the incorporation of Env in the assembling virion and the correct positioning of Env in the mature plasma membrane, the use of VSV-g in our assay is a potential limitation. Studies have demonstrated the PIs block viral entry and that this block is only observed when an HIV-1 envelope is used and is not found when VSV-g and MLV pseudotyped viruses are used[355,614]. Additionally, Mueller and colleagues reported a difference in cell entry pathway between VSV-g and native HIV virions. They hypothesized that this difference could affect downstream pathways such as the transport of the capsid across the

cytoplasm, uncoating and nuclear entry, which have been shown to be affected by the presence of protease inhibitors [615].

To overcome this limitation, future work would be carried out using an assay system which incorporates the native HIV-1 gp120 envelope. It will be interesting to carry out these experiments using full-length, replication-competent viruses in a BSL-3 lab.

In the case of viral evolution where we utilized samples from baseline vs virological failure timepoints, intra-host virus recombination could occur. Studies have shown that recombination affects virus evolution[616-618]. Our assay system did not consider this intra-host recombination. The use of single genome analysis (SGA) in the future could overcome this limitation. SGA is a terminal dilution PCR assay that results in the amplification of a PCR product from a single RNA genome. In the SGA protocol, dilutions of cDNA to a single template precludes recombination between different viral genomes during PCR. This is advantageous because it permits the investigation of linkages between mutations on a single viral genome. The use of SGA in the future would allow us to investigate linkage of drug resistance mutations throughout the HIV-1 pol gene.

In order to gain deeper insights into the mechanism through which the four amino acid mutations in MA confer PI resistance, we intended to study full length, replication-competent chimeric viruses bearing the patient-derived Gag-protease sequences. Although we successfully cloned CRF02_AG patient viruses into the replication-competent (R9-BaL) vector, the resultant retroviral vectors failed to be infectious after several attempts. This could be due to incompatibility between the subtypes, given that R9-BaL is a subtype-B virus. We however created the mutations in a R9-Bal itself. Our aim was to carry out multi-round, spreading infection assays in the absence, and presence of various concentrations of PI. This experiment was however not completed as a result of the shutting down of the BSL3 facility due to the 2020 Covid-19 pandemic. In future work, these experiments would be performed to provide more insight into the mechanisms of resistance of these four amino acid mutations. Additionally, to overcome the

possible issue of subtype incompatibility, we plan to use a CRF02_AG-based infectious molecular clone (BD6-15 CRF02_AG). This vector has previously been characterised and described by Tebit and colleagues as "the first non-subtype B infectious molecular clone of a fast replicating, high producer, X4-tropic primary HIV-1 isolate" [619].

In addition to these multi-round infection experiments using full-length HIV-1, future work would involve the use of multiscale modelling and simulations, as described by Samsudin et al, 2020[602]. Although we do not yet have the expertise for these experiments in our lab, we have begun a collaboration with the lab of Samuel Gan at the Antibody & Product Development Lab, Agency of Science, Technology and Research (A*STAR), Singapore.

First, we would use available structures of HIV-1 Gag domains to build an integrative model of the complete oligomeric Gag polyprotein cleavage product (MA-CA-SP1) in its multimeric state and bound to a viral membrane model. To study the potential interactions between these four amino acid mutation sites and the nearby MA/CA cleavage site residues, we would use the Martini forcefield of the Gag variants ('susceptible' and 'resistant' clones) and subject the variants to coarse-grained (CG) molecular dynamics (MD) simulations. The entire process would be supported by careful calibration against atomic-resolution sampling. This would enable us to elucidate on the mechanism through which these four MA amino acid mutations confer reduced drug susceptibility.

Furthermore, it would be exciting and essential to know whether Gag mutations are capable of facilitating the emergence of major protease mutations in prolonged culture conditions under suboptimal drug pressure. This could potentially explain why the prevalence of major protease mutations increases over time during PI exposure in clinical studies[500]. We would perform population dynamics simulations to incorporate replicative capacity and susceptibility data in order to model the proportion of resistant and susceptible viruses over time, and possibly, therefore, predict the emergence of major mutations in protease. To achieve this, we would use full-length viruses carrying (or not carrying) the four amino acid mutations in multiple rounds of replication, spreading infection

assays over weeks (or months). At various time points, we would carry out population sequencing and study the sequences for the emergence of any major protease inhibitor resistance mutations in the protease gene.

7.2 Final remarks

My PhD thesis identified previously undescribed amino acid mutation signatures in HIV-1 MA, which reduces susceptibility to the protease inhibitors, lopinavir and atazanavir in the absence of major drug resistance-associated mutations in the protease gene.

This present study provided further evidence on the role of Gag in protease inhibitor susceptibility. It also contributed to the knowledge that naturally occurring polymorphisms in CRF02_AG and G subtypes may be responsible for the inherently reduced susceptibility to some protease inhibitors. With recent advances in NGS technologies, and the use of this technology for whole-genome sequencing of HIV, we recommend that more phenotypic methods be developed that would include full-length Gag and protease which would serve to complement the massive amount of genotypic data from NGS to determine the susceptibility of clinical isolates to protease inhibitors. Having shown that Gag plays a vital role in drug resistance, we would recommend that future novel antiretroviral therapeutics should take into consideration the role of gag gene mutations which may potentially serve as an escape mechanism for the virus. In designing these new therapeutic agents, it would be essential to consider subtype-specific polymorphisms which may impact on treatment.

We hope that the findings in this thesis will be used to further the agenda for the need for more research into the determinants of HIV drug resistance, in West Africa in particular, and sub-Saharan Africa in general.

References

1.   CDC. Pneumocystis pneumonia Los Angeles. *MMWR* 1981; **30**: 250-2.

2.   CDC. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR Morbidity and mortality weekly report* 1981; **30**(25): 305.

3.   Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR Morb Mortal Wkly Rep* 1981; **30**(25): 305-8.

4.   Gottlieb MS, Schroff R, Schanker HM, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med* 1981; **305**(24): 1425-31.

5.   Masur H, Michelis MA, Greene JB, et al. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. *N Engl J Med* 1981; **305**(24): 1431-8.

6.   Update on acquired immune deficiency syndrome (AIDS) among patients with hemophilia A. *MMWR Morb Mortal Wkly Rep* 1982; **31**(48): 644-6, 52.

7.   Opportunistic infections and Kaposi's sarcoma among Haitians in the United States. *MMWR Morb Mortal Wkly Rep* 1982; **31**(26): 353-4, 60-1.

8.   Follow-up on Kaposi's sarcoma and Pneumocystis pneumonia. *MMWR Morb Mortal Wkly Rep* 1981; **30**(33): 409-10.

9.   Vieira J, Frank E, Spira TJ, Landesman SH. Acquired immune deficiency in Haitians: opportunistic infections in previously healthy Haitian immigrants. *N Engl J Med* 1983; **308**(3): 125-9.

10.   CDC. Update on acquired immune deficiency syndrome (AIDS)--United States. *MMWR Morbidity and mortality weekly report* 1982; **31**(37): 507.

11.   Barré-Sinoussi F, Chermann JC, Rey F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983; **220**(4599): 868-71.

12.   Gallo RC, Salahuddin SZ, Popovic M, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* 1984; **224**(4648): 500-3.

13.   Ratner L, Gallo RC, Wong-Staal F. HTLV-III, LAV, ARV are variants of same AIDS virus. *Nature* 1985; **313**(6004): 636-7.

14.   Coffin J, Haase A, Levy JA, et al. What to call the AIDS virus? *Nature* 1986; **321**(6065): 10.

15. UNAIDS. Global HIV/AIDS statistics–2019 fact sheet. https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf (Accessed 4th August, 2020). 2020.

16. FMoH. National HIV/AIDS and Reproductive Health and Serological Survey, 2012 (NARHS Plus). Federal Ministry of Health Abuja, Nigeria; 2013.

17. UNAIDS. Global HIV/AIDS statistics–2019|Country fact sheet|Nigeria https://www.unaids.org/en/regionscountries/countries/nigeria (Accessed 31st July 2020). 2019.

18. NACA. Nigeria National Agency for the Control of AIDS. NAIIS National Summary Sheet (Available online at: https://docs.google.com/viewerng/viewer?url=https://naca.gov.ng/wp-content/uploads/2019/03/NAIIS-PA-NATIONAL-FACTSHEET-FINAL.pdf&hl=en) (Accessed 21st July 2020). 2019.

19. King A, Adams M, Carstens E, Lefkowitz E. International Committee on Taxonomy of Viruses: Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses. Academic Press; 2012.

20. Baltimore D. Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 1970; **226**(5252): 1209-11.

21. Freed E, Martin M. Fields Virology. Philadelphia: Lippincott Williams and Wilkins; 2001.

22. Coffin JM, Hughes SH, Varmus HE. Retroviral Virions and Genomes--Retroviruses: Cold Spring Harbor Laboratory Press; 1997.

23. Mizutani S, Boettiger D, Temin HM. A DNA-dependent DNA polymerase and a DNA endonuclease in virions of Rous sarcoma virus. *Nature* 1970; **228**(5270): 424-7.

24. Narayan O. Lentiviruses are etiological agents of chronic diseases in animals and acquired immunodeficiency syndrome in humans. *Canadian Journal of Veterinary Research* 1990; **54**(1): 42.

25. Narayan O, Clements JE. Biology and pathogenesis of lentiviruses. *Journal of General Virology* 1989; **70**(7): 1617-39.

26. Cheevers WP, McGuire TC. Equine infectious anemia virus: immunopathogenesis and persistence. *Reviews of infectious diseases* 1985; **7**(1): 83-8.

27. Gao F, Bailes E, Robertson DL, et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* 1999; **397**(6718): 436-41.

28. Plantier J-C, Leoz M, Dickerson JE, et al. A new human immunodeficiency virus derived from gorillas. *Nature medicine* 2009; **15**(8): 871-2.

29. Van Heuverswyn F, Li Y, Neel C, et al. SIV infection in wild gorillas. *Nature* 2006; **444**(7116): 164-.

30. Vallari A, Bodelle P, Ngansop C, et al. Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low. *AIDS research and human retroviruses* 2010; **26**(1): 109-15.

31. Vallari A, Holzmayer V, Harris B, et al. Confirmation of putative HIV-1 group P in Cameroon. *Journal of virology* 2011; **85**(3): 1403-7.

32. Brennan CA, Bodelle P, Coffey R, et al. The prevalence of diverse HIV-1 strains was stable in Cameroonian blood donors from 1996 to 2004. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2008; **49**(4): 432-9.

33. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. *British medical bulletin* 2001; **58**(1): 19-42.

34. Peeters M, Aghokeng A, Delaporte E. Genetic diversity among human immunodeficiency virus-1 non-B subtypes in viral load and drug resistance assays. *Clinical microbiology and infection* 2010; **16**(10): 1525-31.

35. Hemelaar J, Elangovan R, Yun J, et al. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *The Lancet infectious diseases* 2019; **19**(2): 143-55.

36. Robertson DL, Anderson J, Bradac J, et al. HIV-1 nomenclature proposal. *Science* 2000; **288**(5463): 55-.

37. Désiré N, Cerutti L, Le Hingrat Q, et al. Characterization update of HIV-1 M subtypes diversity and proposal for subtypes A and D sub-subtypes reclassification. *Retrovirology* 2018; **15**(1): 80.

38. Peeters M. HIV Sequence Compendium, chapter Recombinant HIV sequences: their role in the global epidemic. *Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico* 2000: 139-54.

39. Carr JK, Foley BT, Leitner T, Salminen M, Korber B, McCutchan F. Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic. *Human retroviruses and AIDS* 1998; **111**(10).

40. Carr JK, Torimiro JN, Wolfe ND, et al. The AG recombinant IbNG and novel strains of group M HIV-1 are common in Cameroon. *Virology* 2001; **286**(1): 168-81.

41.   NDEMBI N, ABRAHA A, PILCH H, et al. Molecular characterization of HIV-1 and HIV-2 in Yaounde, Cameroon: Evidence of major drug resistance mutations in newly diagnosed non-B infected patients. *Journal of clinical microbiology* 2007.

42.   Torimiro JN, D'Arrigo R, Takou D, et al. Human immunodeficiency virus type 1 intersubtype recombinants predominate in the AIDS epidemic in Cameroon. *The new microbiologica* 2009; **32**(4): 325.

43.   Fischetti L, Opare-Sem O, Candotti D, Lee H, Allain J-P. Higher viral load may explain the dominance of CRF02_AG in the molecular epidemiology of HIV in Ghana. *Aids* 2004; **18**(8): 1208-10.

44.   Fischetti L, Opare-Sem O, Candotti D, Sarkodie F, Lee H, Allain JP. Molecular epidemiology of HIV in Ghana: dominance of CRF02_AG. *Journal of medical virology* 2004; **73**(2): 158-66.

45.   Sarr AD, Eisen G, Guèye-Ndiaye A, et al. Viral dynamics of primary HIV-1 infection in Senegal, West Africa. *The Journal of infectious diseases* 2005; **191**(9): 1460-7.

46.   Njai HF, Gali Y, Vanham G, et al. The predominance of Human Immunodeficiency Virus type 1 (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its replicative fitness. *Retrovirology* 2006; **3**(1): 1-11.

47.   Ojesina AI, Sankalé J-L, Odaibo G, et al. Subtype-specific patterns in HIV Type 1 reverse transcriptase and protease in Oyo State, Nigeria: implications for drug resistance and host response. *AIDS Research & Human Retroviruses* 2006; **22**(8): 770-9.

48.   Agwale SM, Zeh C, Robbins KE, et al. Molecular surveillance of HIV-1 field strains in Nigeria in preparation for vaccine trials1Use of trade names is for identification only and does not constitute endorsement by the US Department of Health and Human Service, or the Centers for Disease Control and Prevention.1. *Vaccine* 2002; **20**(16): 2131-9.

49.   Odaibo G, Olaleye D, Heyndrickx L, Vereecken K, Houwer K, Jassens W. Mother-to-child transmission of different HIV-1 subtypes among ARV naïve infected pregnant women in Nigeria. *Revista do Instituto de Medicina Tropical de São Paulo* 2006; **48**(2): 77-80.

50.   Ajoge HO, Gordon ML, de Oliveira T, et al. Genetic Characteristics, Coreceptor Usage Potential and Evolution of Nigerian HIV-1 Subtype G and CRF02_AG Isolates. *PloS one* 2011; **6**(3): e17865.

51.   McCutchan FE, Carr JK, Bajani M, et al. Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology* 1999; **254**(2): 226-34.

52.   Delatorre E, Mir D, Bello G. Spatiotemporal Dynamics of the HIV-1 Subtype G Epidemic in West and Central Africa. *PloS one* 2014; **9**(6): e98908.

53.   Kiwanuka N, Laeyendecker O, Quinn TC, et al. HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda. *AIDS (London, England)* 2009; **23**(18): 2479.

54.   Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine* 2011; **1**(1).

55.   De Cock KM, Adjorlolo G, Ekpini E, et al. Epidemiology and transmission of HIV-2: why there is no HIV-2 pandemic. *Jama* 1993; **270**(17): 2083-6.

56.   Ibe S, Yokomaku Y, Shiino T, et al. HIV-2 CRF01_AB: first circulating recombinant form of HIV-2. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2010; **54**(3): 241-7.

57.   Melhuish A, Lewthwaite P. Natural history of HIV and AIDS. *Medicine* 2018; **46**(6): 356-61.

58.   Chen P, Chen BK, Mosoian A, et al. Virological synapses allow HIV-1 uptake and gene expression in renal tubular epithelial cells. *Journal of the American Society of Nephrology* 2011; **22**(3): 496-507.

59.   Liu Y, Liu H, Kim BO, et al. CD4-independent infection of astrocytes by human immunodeficiency virus type 1: requirement for the human mannose receptor. *Journal of virology* 2004; **78**(8): 4120-33.

60.   Sabin CA, Lundgren JD. The natural history of HIV infection. *Current Opinion in HIV and AIDS* 2013; **8**(4): 311.

61.   McMichael AJ, Borrow P, Tomaras GD, Goonetilleke N, Haynes BF. The immune response during acute HIV-1 infection: clues for vaccine development. *Nature Reviews Immunology* 2010; **10**(1): 11-23.

62.   Stacey AR, Norris PJ, Qin L, et al. Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *Journal of virology* 2009; **83**(8): 3719-33.

63.   Reitz MS, Gallo RC. Human immunodeficiency viruses. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, GL Mandell, JE Bennett and R Dolin, Eds* 2010: 2323-33.

64.   Mehandru S, Poles MA, Tenner-Racz K, et al. Mechanisms of gastrointestinal CD4+ T-cell depletion during acute and early human immunodeficiency virus type 1 infection. *Journal of virology* 2007; **81**(2): 599-612.

65.	Organization WH. Interim WHO clinical staging of HVI/AIDS and HIV/AIDS case definitions for surveillance: African Region: World Health Organization, 2005.

66.	Pornillos O, Ganser-Pornillos BK. Maturation of retroviruses. *Current Opinion in Virology* 2019; **36**: 47-55.

67.	Zhu P, Liu J, Bess J, et al. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* 2006; **441**(7095): 847-52.

68.	Kaushik R, Ratner L. Role of human immunodeficiency virus type 1 matrix phosphorylation in an early postentry step of virus replication. *Journal of virology* 2004; **78**(5): 2319-26.

69.	Sierra S, Kupfer B, Kaiser R. Basics of the virology of HIV-1 and its replication. *Journal of Clinical Virology* 2005; **34**(4): 233-44.

70.	de Oliveira T, Engelbrecht S, Janse van Rensburg E, et al. Variability at Human Immunodeficiency Virus Type 1 Subtype C Protease Cleavage Sites: an Indication of Viral Fitness? *Journal of Virology* 2003; **77**(17): 9422-30.

71.	Laskey SB, Siliciano RF. A mechanistic theory to explain the efficacy of antiretroviral therapy. *Nature Reviews Microbiology* 2014; **12**(11): 772-80.

72.	LANL. Landmarks of the HIV-1 genome, HXB2: HIV-1 Gene Map. 24th January 2017 2017. https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html (accessed 18th August 2020).

73.	Guenzel CA, Hérate C, Benichou S. HIV-1 Vpr-a still "enigmatic multitasker". *Front Microbiol* 2014; **5**: 127.

74.	Heinzinger NK, Bukrinsky MI, Haggerty SA, et al. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proceedings of the National Academy of Sciences of the United States of America* 1994; **91**(15): 7311-5.

75.	Vanitharani R, Mahalingam S, Rafaeli Y, et al. HIV-1 Vpr transactivates LTR-directed expression through sequences present within -278 to -176 and increases virus replication in vitro. *Virology* 2001; **289**(2): 334-42.

76.	Karn J, Stoltzfus CM. Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. *Cold Spring Harbor Perspectives in Medicine* 2012; **2**(2).

77.	Malim MH, Bieniasz PD. HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harbor Perspectives in Medicine* 2012; **2**(5).

78.   Landi A, Iannucci V, Nuffel AV, Meuwissen P, Verhasselt B. One protein to rule them all: modulation of cell surface receptors and molecules by HIV Nef. *Curr HIV Res* 2011; **9**(7): 496-504.

79.   Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 2008; **451**(7177): 425-30.

80.   Kmiec D, Iyer SS, Stürzel CM, Sauter D, Hahn BH, Kirchhoff F. Vpu-Mediated Counteraction of Tetherin Is a Major Determinant of HIV-1 Interferon Resistance. *mBio* 2016; **7**(4).

81.   Margottin F, Bour SP, Durand H, et al. A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. *Mol Cell* 1998; **1**(4): 565-74.

82.   Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. *Science* 1996; **271**(5255): 1582-6.

83.   Clapham PR, McKnight Á. Cell surface receptors, virus entry and tropism of primate lentiviruses. *J Gen Virol* 2002; **83**(Pt 8): 1809-29.

84.   Colman PM, Lawrence MC. The structural biology of type I viral membrane fusion. *Nat Rev Mol Cell Biol* 2003; **4**(4): 309-19.

85.   Markosyan RM, Cohen FS, Melikyan GB. HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation. *Mol Biol Cell* 2003; **14**(3): 926-38.

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

87.   Mangasarian A, Foti M, Aiken C, Chin D, Carpentier JL, Trono D. The HIV-1 Nef protein acts as a connector with sorting pathways in the Golgi and at the plasma membrane. *Immunity* 1997; **6**(1): 67-77.

88.   Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 Entry Cofactor: Functional cDNA Cloning of a Seven-Transmembrane, G Protein-Coupled Receptor. *Science* 1996; **272**(5263): 872-7.

89.   Mild M, Esbjörnsson J, Fenyö EM, Medstrand P. Frequent intrapatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch. *J Virol* 2007; **81**(7): 3369-76.

90.   Troyer RM, Collins KR, Abraha A, et al. Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J Virol* 2005; **79**(14): 9006-18.

91.   van Opijnen T, Berkhout B. The host environment drives HIV-1 fitness. *Rev Med Virol* 2005; **15**(4): 219-33.

92.   Engelman A, Cherepanov P. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature Reviews Microbiology* 2012; **10**(4): 279-90.

93.   Checkley MA, Luttge BG, Freed EO. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. *Journal of Molecular Biology* 2011; **410**(4): 582-608.

94.   Hallenberger S, Moulard M, Sordel M, Klenk HD, Garten W. The role of eukaryotic subtilisin-like endoproteases for the activation of human immunodeficiency virus glycoproteins in natural host cells. *Journal of virology* 1997; **71**(2): 1036-45.

95.   Poon AFY, Lewis FI, Pond SLK, Frost SDW. Evolutionary interactions between N-linked glycosylation sites in the HIV-1 envelope. *PLoS computational biology* 2007; **3**(1): e11-e.

96.   Tsibris AM, Kuritzkes DR. Chemokine antagonists as therapeutics: focus on HIV-1. *Annu Rev Med* 2007; **58**: 445-59.

97.   Wilen CB, Tilton JC, Doms RW. HIV: cell binding and entry. *Cold Spring Harbor perspectives in medicine* 2012; **2**(8): a006866.

98.   Pantophlet R, Burton DR. GP120: Target for Neutralizing HIV-1 Antibodies. *Annual Review of Immunology* 2006; **24**(1): 739-69.

99.   Lehmann M, Nikolic DS, Piguet V. How HIV-1 takes advantage of the cytoskeleton during replication and cell-to-cell transmission. *Viruses* 2011; **3**(9): 1757-76.

100.   Yamashita M, Emerman M. Retroviral infection of non-dividing cells: old and new perspectives. *Virology* 2006; **344**(1): 88-93.

101.   Fassati A. Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus Research* 2012; **170**(1): 15-24.

102.   Hilditch L, Towers GJ. A model for cofactor use during HIV-1 reverse transcription and nuclear entry. *Curr Opin Virol* 2014; **4**(100): 32-6.

103.   Burdick RC, Li C, Munshi M, et al. HIV-1 uncoats in the nucleus near sites of integration. *Proceedings of the National Academy of Sciences of the United States of America* 2020; **117**(10): 5486-93.

104.   Rasaiyaah J, Tan CP, Fletcher AJ, et al. HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature* 2013; **503**(7476): 402-5.

105.   Sluis-Cremer N, Arion D, Parniak M. Molecular mechanisms of HIV-1 resistance to nucleoside reverse transcriptase inhibitors (NRTIs). *Cellular and Molecular Life Sciences CMLS* 2000; **57**(10): 1408-22.

106.   Sarafianos SG, Marchand B, Das K, et al. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol* 2009; **385**(3): 693-713.

107.   Götte M, Li X, Wainberg MA. HIV-1 reverse transcription: a brief overview focused on structure-function relationships among molecules involved in initiation of the reaction. *Arch Biochem Biophys* 1999; **365**(2): 199-210.

108.   Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 1998; **282**(5394): 1669-75.

109.   Chen Y, Balakrishnan M, Roques BP, Fay PJ, Bambara RA. Mechanism of minus strand strong stop transfer in HIV-1 reverse transcription. *J Biol Chem* 2003; **278**(10): 8006-17.

110.   Trono D. Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses. *J Virol* 1992; **66**(8): 4893-900.

111.   Hu WS, Temin HM. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proceedings of the National Academy of Sciences of the United States of America* 1990; **87**(4): 1556-60.

112.   Hu WS, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* 2012; **2**(10).

113.   Charneau P, Alizon M, Clavel F. A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *J Virol* 1992; **66**(5): 2814-20.

114.   Smith JS, Roth MJ. Specificity of human immunodeficiency virus-1 reverse transcriptase-associated ribonuclease H in removal of the minus-strand primer, tRNA(Lys3). *J Biol Chem* 1992; **267**(21): 15071-9.

115.   Di Nunzio F. New insights in the role of nucleoporins: A bridge leading to concerted steps from HIV-1 nuclear entry until integration. *Virus Research* 2013; **178**(2): 187-96.

116.   Farnet CM, Haseltine WA. Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *Journal of virology* 1991; **65**(4): 1910-5.

117.   Craigie R. HIV integrase, a brief overview from chemistry to therapeutics. *Journal of Biological Chemistry* 2001; **276**(26): 23213-6.

118.   Mbisa JL, Martin SA, Cane PA. Patterns of resistance development with integrase inhibitors in HIV. *Infect Drug Resist* 2011; **4**: 65-76.

119.   McColl DJ, Chen X. Strand transfer inhibitors of HIV-1 integrase: bringing IN a new era of antiretroviral therapy. *Antiviral research* 2010; **85**(1): 101-18.

120.   Engelman A, Hickman AB, Craigie R. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *Journal of virology* 1994; **68**(9): 5911-7.

121.   Lataillade M, Kozal MJ. The hunt for HIV-1 integrase inhibitors. *AIDS Patient Care & STDs* 2006; **20**(7): 489-501.

122.   Delelis O, Carayon K, Saïb A, Deprez E, Mouscadet J-F. Integrase and integration: biochemical activities of HIV-1 integrase. *Retrovirology* 2008; **5**(1): 1-13.

123.   Krishnan L, Li X, Naraharisetty HL, Hare S, Cherepanov P, Engelman A. Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proceedings of the National Academy of Sciences* 2010; **107**(36): 15910-5.

124.   Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 2010; **464**(7286): 232-6.

125.   Li L, Olvera JM, Yoder KE, et al. Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *Embo j* 2001; **20**(12): 3272-81.

126.   Kilzer JM, Stracker T, Beitzel B, Meek K, Weitzman M, Bushman FD. Roles of host cell factors in circularization of retroviral dna. *Virology* 2003; **314**(1): 460-7.

127.   Cullen BR, Lomedico PT, Ju G. Transcriptional interference in avian retroviruses—implications for the promoter insertion model of leukaemogenesis. *Nature* 1984; **307**(5948): 241-5.

128.   Klaver B, Berkhout B. Comparison of 5'and 3'long terminal repeat promoter function in human immunodeficiency virus. *Journal of virology* 1994; **68**(6): 3830-40.

129.   Pereira LA, Bentley K, Peeters A, Churchill MJ, Deacon NJ. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res* 2000; **28**(3): 663-8.

130.   Brady J, Kashanchi F. Tat gets the "green" light on transcription initiation. *Retrovirology* 2005; **2**: 69.

131.   Berkhout B, Silverman RH, Jeang KT. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell* 1989; **59**(2): 273-82.

132.   Raha T, Cheng SW, Green MR. HIV-1 Tat stimulates transcription complex assembly through recruitment of TBP in the absence of TAFs. *PLoS Biol* 2005; **3**(2): e44.

133.   Wei P, Garber ME, Fang SM, Fischer WH, Jones KA. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* 1998; **92**(4): 451-62.

134.   Yamaguchi Y, Takagi T, Wada T, et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* 1999; **97**(1): 41-51.

135.   Khan H. Investigating antagonism of innate immunity by HIV-1 accessory protein Vpr. London, United Kingdom: University College London; 2019.

136.   Ho CK, Shuman S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* 1999; **3**(3): 405-11.

137.   Cherrington J, Ganem D. Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences. *Embo j* 1992; **11**(4): 1513-24.

138.   Ashe MP, Pearson LH, Proudfoot NJ. The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *Embo j* 1997; **16**(18): 5752-63.

139.   Purcell DF, Martin MA. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol* 1993; **67**(11): 6365-78.

140.   Bolinger C, Boris-Lawrie K. Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. *Retrovirology* 2009; **6**: 8.

141.   de Breyne S, Chamond N, Décimo D, et al. In vitro studies reveal that different modes of initiation on HIV-1 mRNA have different levels of requirement for eukaryotic initiation factor 4F. *Febs j* 2012; **279**(17): 3098-111.

142.   Inamdar K, Floderer C, Favard C, Muriaux D. Monitoring HIV-1 Assembly in Living Cells: Insights from Dynamic and Single Molecule Microscopy. *Viruses* 2019; **11**(1).

143.   Goodenow MM, Rose SL, Tuttle DL, Sleasman JW. HIV-1 fitness and macrophages. *Journal of leukocyte biology* 2003; **74**(5): 657-66.

144.   Perno CF, Svicher V, Schols D, Pollicita M, Balzarini J, Aquaro S. Therapeutic strategies towards HIV-1 infection in macrophages. *Antiviral research* 2006; **71**(2-3): 293-300.

145.   Briggs JAG, Kräusslich H-G. The Molecular Architecture of HIV. *Journal of Molecular Biology* 2011; **410**(4): 491-500.

146.   Bukrinskaya A. HIV-1 assembly and maturation. *Archives of virology* 2004; **149**(6): 1067-82.

147.   Ono A, Freed EO. Cell-type-dependent targeting of human immunodeficiency virus type 1 assembly to the plasma membrane and the multivesicular body. *Journal of virology* 2004; **78**(3): 1552-63.

148.   Ganser-Pornillos BK, Yeager M, Sundquist WI. The structural biology of HIV assembly. *Current opinion in structural biology* 2008; **18**(2): 203-17.

149.   Derdowski A, Ding L, Spearman P. A novel fluorescence resonance energy transfer assay demonstrates that the human immunodeficiency virus type 1 Pr55Gag I domain mediates Gag-Gag interactions. *Journal of virology* 2004; **78**(3): 1230-42.

150.   Martin-Serrano J, Neil SJ. Host factors involved in retroviral budding and release. *Nature Reviews Microbiology* 2011; **9**(7): 519-31.

151.   Kutluay SB, Bieniasz PD. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLoS pathogens* 2010; **6**(11): e1001200.

152.   Moore MD, Hu W-S. HIV-1 RNA dimerization: It takes two to tango. *AIDS reviews* 2009; **11**(2): 91.

153.   Sundquist WI, Kräusslich H-G. HIV-1 assembly, budding, and maturation. *Cold Spring Harbor perspectives in medicine* 2012; **2**(7): a006924.

154.   Davis MR, Jiang J, Zhou J, Freed EO, Aiken C. A mutation in the human immunodeficiency virus type 1 Gag protein destabilizes the interaction of the envelope protein subunits gp120 and gp41. *Journal of virology* 2006; **80**(5): 2405-17.

155.   Chojnacki J, Staudt T, Glass B, et al. Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science* 2012; **338**(6106): 524-8.

156.   Yang Y, Qu N, Tan J, Rushdi MN, Krueger CJ, Chen AK. Roles of Gag-RNA interactions in HIV-1 virus assembly deciphered by single-molecule localization microscopy. *Proceedings of the National Academy of Sciences* 2018; **115**(26): 6721-6.

157.   Yu FH, Huang KJ, Wang CT. HIV-1 Mutant Assembly, Processing and Infectivity Expresses Pol Independent of Gag. *Viruses* 2020; **12**(1).

158.   Meng B, Lever AM. Wrapping up the bad news: HIV assembly and release. *Retrovirology* 2013; **10**: 5.

159.   Hanson PI, Shim S, Merrill SA. Cell biology of the ESCRT machinery. *Current opinion in cell biology* 2009; **21**(4): 568-74.

160.   Hurley JH, Emr SD. The ESCRT complexes: structure and mechanism of a membrane-trafficking network. *Annu Rev Biophys Biomol Struct* 2006; **35**: 277-98.

161.   Wollert T, Hurley JH. Molecular mechanism of multivesicular body biogenesis by ESCRT complexes. *Nature* 2010; **464**(7290): 864-9.

162.   Meng B, Ip NCY, Abbink TEM, Kenyon JC, Lever AML. ESCRT-II functions by linking to ESCRT-I in human immunodeficiency virus-1 budding. *Cellular microbiology* 2020; **22**(5): e13161-e.

163.   Hurley JH, Hanson PI. Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nat Rev Mol Cell Biol* 2010; **11**(8): 556-66.

164.   Martin-Serrano J, Zang T, Bieniasz PD. HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nat Med* 2001; **7**(12): 1313-9.

165.   Fujii K, Munshi UM, Ablan SD, et al. Functional role of Alix in HIV-1 replication. *Virology* 2009; **391**(2): 284-92.

166.   Hanson PI, Roth R, Lin Y, Heuser JE. Plasma membrane deformation by circular arrays of ESCRT-III protein filaments. *J Cell Biol* 2008; **180**(2): 389-402.

167.   Raiborg C, Stenmark H. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* 2009; **458**(7237): 445-52.

168.   Freed EO. HIV-1 assembly, release and maturation. *Nature Reviews Microbiology* 2015; **13**(8): 484-96.

169.   Kaplan A, Zack J, Knigge M, et al. Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles. *Journal of virology* 1993; **67**(7): 4050-5.

170.   Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 1988; **331**(6153): 280-3.

171.   Louis JM, Aniana A, Weber IT, Sayer JM. Inhibition of autoprocessing of natural variants and multidrug resistant mutant precursors of HIV-1 protease by clinical inhibitors. *Proceedings of the National Academy of Sciences* 2011; **108**(22): 9072-7.

172.   Louis JM, Ishima R, Torchia DA, Weber IT. HIV-1 protease: structure, dynamics, and inhibition. *Advances in pharmacology* 2007; **55**: 261-98.

173.	Collins JR, Burt SK, Erickson J. Activated dynamics of flap opening in HIV-1 protease. Aspartic Proteinases: Springer; 1995: 455-60.

174.	Daniels SI, Davis DA, Soule EE, et al. The initial step in human immunodeficiency virus type 1 GagProPol processing can be regulated by reversible oxidation. *PloS one* 2010; **5**(10): e13595.

175.	Davis DA, Soule EE, Davidoff KS, Daniels SI, Naiman NE, Yarchoan R. Activity of human immunodeficiency virus type 1 protease inhibitors against the initial autocleavage in Gag-Pol polyprotein processing. *Antimicrobial agents and chemotherapy* 2012; **56**(7): 3620-8.

176.	Prabu-Jeyabalan M, Nalivaika EA, King NM, Schiffer CA. Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *Journal of virology* 2004; **78**(22): 12446-54.

177.	Venkatakrishnan B, Palii M-L, Agbandje-McKenna M, McKenna R. Mining the Protein Data Bank to Differentiate Error from Structural Variation in Clustered Static Structures: An Examination of HIV Protease. *Viruses* 2012; **4**(3): 348-62.

178.	Freed EO. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* 1998; **251**(1): 1-15.

179.	Scarlata S, Carter C. Role of HIV-1 Gag domains in viral assembly. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 2003; **1614**(1): 62-72.

180.	Bell NM, Lever AM. HIV Gag polyprotein: processing and early viral particle assembly. *Trends in microbiology* 2013; **21**(3): 136-44.

181.	Waheed AA, Freed EO. HIV type 1 Gag as a target for antiviral therapy. *AIDS research and human retroviruses* 2012; **28**(1): 54-75.

182.	Anderson-Daniels J, Singh PK, Sowd GA, Li W, Engelman AN, Aiken C. Dominant Negative MA-CA Fusion Protein Is Incorporated into HIV-1 Cores and Inhibits Nuclear Entry of Viral Preintegration Complexes. *J Virol* 2019; **93**(21).

183.	Freed EO. HIV-1 replication. *Somatic cell and molecular genetics* 2001; **26**(1-6): 13-33.

184.	Tedbury PR, Novikova M, Alfadhli A, et al. HIV-1 Matrix Trimerization-Impaired Mutants Are Rescued by Matrix Substitutions That Enhance Envelope Glycoprotein Incorporation. *J Virol* 2019; **94**(1).

185.	Kroupa T, Datta SA, Rein A. Distinct Contributions of Different Domains within the HIV-1 Gag Polyprotein to Specific and Nonspecific Interactions with RNA. *Viruses* 2020; **12**(4): 394.

186.   Francis AC, Melikyan GB. Single HIV-1 Imaging Reveals Progression of Infection through CA-Dependent Steps of Docking at the Nuclear Pore, Uncoating, and Nuclear Transport. *Cell Host Microbe* 2018; **23**(4): 536-48.e6.

187.   Ohi Y, Clever JL. Sequences in the 5′ and 3′ R elements of human immunodeficiency virus type 1 critical for efficient reverse transcription. *Journal of Virology* 2000; **74**(18): 8324-34.

188.   Harrison GP, Miele G, Hunter E, Lever AM. Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line. *Journal of virology* 1998; **72**(7): 5886-96.

189.   Freed EO. HIV-1 and the host cell: an intimate association. *Trends in microbiology* 2004; **12**(4): 170-7.

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

191.   Ono A, Waheed AA, Freed EO. Depletion of cellular cholesterol inhibits membrane binding and higher-order multimerization of human immunodeficiency virus type 1 Gag. *Virology* 2007; **360**(1): 27-35.

192.   Mattei S, Tan A, Glass B, Müller B, Kräusslich H-G, Briggs JAG. High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation. *Proceedings of the National Academy of Sciences* 2018; **115**(40): E9401-E10.

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

194.   Pettit SC, Henderson GJ, Schiffer CA, Swanstrom R. 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* 2002; **76**(20): 10226-33.

195.   Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* 2005; **2**(1): 66.

196.   Dam E, Quercia R, Glass B, et al. Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS pathogens* 2009; **5**(3): e1000345.

197.   Nijhuis M, van Maarseveen NM, Lastere S, et al. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med* 2007; **4**(1): e36.

198.   Pettit S, Simsic J, Loeb D, Everitt L, Hutchison C, Swanstrom R. Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. *Journal of Biological Chemistry* 1991; **266**(22): 14539-47.

199.   Kräusslich H-G, Ingraham RH, Skoog MT, Wimmer E, Pallai PV, Carter CA. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. *Proceedings of the National Academy of Sciences* 1989; **86**(3): 807-11.

200.   Pettit SC, Moody MD, Wehbie RS, et al. 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* 1994; **68**(12): 8017-27.

201.   Fields BN, Knipe DM, Howley PM, Everiss KD, Kung H-J. Fundamental virology: Lippincott-Raven Philadelphia^ ePA PA; 1996.

202.   Chen P, Hübner W, Spinelli MA, Chen BK. Predominant Mode of Human Immunodeficiency Virus Transfer between T Cells Is Mediated by Sustained Env-Dependent Neutralization-Resistant Virological Synapses. *Journal of Virology* 2007; **81**(22): 12582-95.

203.   Jolly C, Kashefi K, Hollinshead M, Sattentau QJ. HIV-1 cell to cell transfer across an Env-induced, actin-dependent synapse. *The Journal of experimental medicine* 2004; **199**(2): 283-93.

204.   Dimitrov DS, Willey RL, Sato H, Chang LJ, Blumenthal R, Martin MA. Quantitation of human immunodeficiency virus type 1 infection kinetics. *Journal of Virology* 1993; **67**(4): 2182-90.

205.   Sourisseau M, Sol-Foulon N, Porrot F, Blanchet F, Schwartz O. Inefficient human immunodeficiency virus replication in mobile lymphocytes. *Journal of Virology* 2007; **81**(2): 1000-12.

206.   Alvarez RA, Barría MI, Chen BK. Unique Features of HIV-1 Spread through T Cell Virological Synapses. *PLoS pathogens* 2014; **10**(12): e1004513.

207.   Piguet V, Sattentau Q. Dangerous liaisons at the virological synapse. *The Journal of clinical investigation* 2004; **114**(5): 605-10.

208.   Phillips DM. The role of cell-to-cell transmission in HIV infection. *AIDS* 1994; **8**(6): 719-32.

209.   Dimitrov D, Broder C, Berger E, Blumenthal R. Calcium ions are required for cell fusion mediated by the CD4-human immunodeficiency virus type 1 envelope glycoprotein interaction. *Journal of virology* 1993; **67**(3): 1647-52.

210.   Carr JM, Hocking H, Li P, Burrell CJ. Rapid and Efficient Cell-to-Cell Transmission of Human Immunodeficiency Virus Infection from Monocyte-

Derived Macrophages to Peripheral Blood Lymphocytes. *Virology* 1999; **265**(2): 319-29.

211.   Martin N, Sattentau Q. Cell-to-cell HIV-1 spread and its implications for immune evasion. *Current Opinion in HIV and AIDS* 2009; **4**(2): 143-9.

212.   Xie M, Leroy H, Mascarau R, et al. Cell-to-Cell Spreading of HIV-1 in Myeloid Target Cells Escapes SAMHD1 Restriction. *mBio* 2019; **10**(6): e02457-19.

213.   Pedro KD, Henderson AJ, Agosto LM. Mechanisms of HIV-1 cell-to-cell transmission and the establishment of the latent reservoir. *Virus Research* 2019; **265**: 115-21.

214.   Sigal A, Kim JT, Balazs AB, et al. Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. *Nature* 2011; **477**(7362): 95-8.

215.   Martin N, Welsch S, Jolly C, Briggs JA, Vaux D, Sattentau QJ. Virological synapse-mediated spread of human immunodeficiency virus type 1 between T cells is sensitive to entry inhibition. *Journal of virology* 2010; **84**(7): 3516-27.

216.   Titanji BK, Aasa-Chapman M, Pillay D, Jolly C. Protease inhibitors effectively block cell-to-cell spread of HIV-1 between T cells. *Retrovirology* 2013; **10**(1): 161.

217.   Bergantz L, Subra F, Deprez E, Delelis O, Richetta C. Interplay between intrinsic and innate immunity during hiv infection. *Cells* 2019; **8**(8): 922.

218.   Ding S, Pan Q, Liu S-L, Liang C. HIV-1 mutates to evade IFITM1 restriction. *Virology* 2014; **454-455**: 11-24.

219.   Smith SE, Weston S, Kellam P, Marsh M. IFITM proteins—cellular inhibitors of viral entry. *Current Opinion in Virology* 2014; **4**: 71-7.

220.   Ganser-Pornillos BK, Pornillos O. Restriction of HIV-1 and other retroviruses by TRIM5. *Nat Rev Microbiol* 2019; **17**(9): 546-56.

221.   Lahouassa H, Daddacha W, Hofmann H, et al. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nature immunology* 2012; **13**(3): 223-8.

222.   Kane M, Yadav SS, Bitzegeio J, et al. MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* 2013; **502**(7472): 563-6.

223.   Neil SJ. The antiviral activities of tetherin. Intrinsic Immunity: Springer; 2013: 67-104.

224.   Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 2008; **451**(7177): 425-30.

225. Zhu Y, Chen G, Lv F, et al. Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proceedings of the National Academy of Sciences* 2011; **108**(38): 15834-9.

226. Erazo A, Goff SP. Nuclear matrix protein Matrin 3 is a regulator of ZAP-mediated retroviral restriction. *Retrovirology* 2015; **12**(1): 57.

227. Li M, Kao E, Gao X, et al. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature* 2012; **491**(7422): 125-8.

228. Soliman M, Srikrishna G, Balagopal A. Mechanisms of HIV-1 control. *Current Hiv/aids Reports* 2017; **14**(3): 101-9.

229. Okumura A, Lu G, Pitha-Rowe I, Pitha PM. Innate antiviral response targets HIV-1 release by the induction of ubiquitin-like protein ISG15. *Proceedings of the National Academy of Sciences* 2006; **103**(5): 1440-5.

230. D'Urbano V, De Crignis E, Re MC. Host restriction factors and Human immunodeficiency Virus (HIV-1): a dynamic interplay involving all phases of the viral life cycle. *Current HIV research* 2018; **16**(3): 184-207.

231. Gupta RK, Abdul-Jawad S, McCoy LE, et al. HIV-1 remission following CCR5Δ32/Δ32 haematopoietic stem-cell transplantation. *Nature* 2019; **568**(7751): 244-8.

232. Gupta RK, Peppa D, Hill AL, et al. Evidence for HIV-1 cure after CCR5Δ32/Δ32 allogeneic haemopoietic stem-cell transplantation 30 months post analytical treatment interruption: a case report. *The Lancet HIV* 2020; **7**(5): e340-e7.

233. Hütter G, Nowak D, Mossner M, et al. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *New England Journal of Medicine* 2009; **360**(7): 692-8.

234. Hütter G, Thiel E. Allogeneic transplantation of CCR5-deficient progenitor cells in a patient with HIV infection: an update after 3 years and the search for patient no. 2. *Aids* 2011; **25**(2): 273-4.

235. Allers K, Hütter G, Hofmann J, et al. Evidence for the cure of HIV infection by CCR5Δ32/Δ32 stem cell transplantation. *Blood, The Journal of the American Society of Hematology* 2011; **117**(10): 2791-9.

236. Zhang X. Anti-retroviral drugs: current state and development in the next decade. *Acta Pharmaceutica Sinica B* 2018; **8**(2): 131-6.

237. Palella FJ, Delaney KM, Moorman AC, et al. Declining Morbidity and Mortality among Patients with Advanced Human Immunodeficiency Virus Infection. *New England Journal of Medicine* 1998; **338**(13): 853-60.

238. Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med* 2012; **2**(4): a007161.

239. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* 2011; **365**(6): 493-505.

240. Reynolds SJ, Makumbi F, Nakigozi G, et al. HIV-1 transmission among HIV-1 discordant couples before and after the introduction of antiretroviral therapy. *Aids* 2011; **25**(4): 473-7.

241. Donnell D, Baeten JM, Kiarie J, et al. Heterosexual HIV-1 transmission after initiation of antiretroviral therapy: a prospective cohort analysis. *Lancet* 2010; **375**(9731): 2092-8.

242. Eisinger RW, Dieffenbach CW, Fauci AS. HIV Viral Load and Transmissibility of HIV Infection: Undetectable Equals Untransmittable. *Jama* 2019; **321**(5): 451-2.

243. Njenda DT. From bedside to bench and back: future options for antiretroviral drugs in non-B HIV-1 subtypes. Stockholm, Sweden: Karolinska Institutet; 2020.

244. Hardy H, Skolnik PR. Enfuvirtide, a new fusion inhibitor for therapy of human immunodeficiency virus infection. *Pharmacotherapy* 2004; **24**(2): 198-211.

245. Manfredi R, Sabbatani S. A novel antiretroviral class (fusion inhibitors) in the management of HIV infection. Present features and future perspectives of enfuvirtide (T-20). *Curr Med Chem* 2006; **13**(20): 2369-84.

246. Chan DC, Fass D, Berger JM, Kim PS. Core Structure of gp41 from the HIV Envelope Glycoprotein. *Cell* 1997; **89**(2): 263-73.

247. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 1997; **387**(6631): 426-30.

248. Lieberman-Blum SS, Fung HB, Bandres JC. Maraviroc: A CCR5-receptor antagonist for the treatment of HIV-1 infection. *Clinical Therapeutics* 2008; **30**(7): 1228-50.

249. Briz V, Poveda E, Soriano V. HIV entry inhibitors: mechanisms of action and resistance pathways. *The Journal of antimicrobial chemotherapy* 2006; **57**(4): 619-27.

250. US-FDA. US FDA approves new HIV treatment for patients who have limited treatment options [media release]. 6 Mar 2018. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm599657.htm. 2018.

251. Markham A. Ibalizumab: First Global Approval. *Drugs* 2018; **78**(7): 781-5.

252.   US-FDA. US FDA. TROGARZO™ (ibalizumab-uiyk): US prescribing Information. 2018. www.accessdata.fda.gov/drugsatfda_docs/label/2018/761065lbl.pdf. (accessed 6th August 2020 2020).

253.   FDA-US. FDA Approves New HIV Treatment for Patients With Limited Treatment Options. 2020.

254.   FDA-US. Rukobia: Highlights of prescribing information. 2020. https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/212950s000lbl.pdf (accessed 6th August 2020.

255.   Kozal M, Aberg J, Pialoux G, et al. Fostemsavir in Adults with Multidrug-Resistant HIV-1 Infection. *New England Journal of Medicine* 2020; **382**(13): 1232-43.

256.   Langley DR, Roy Kimura S, Sivaprakasam P, et al. Homology models of the HIV-1 attachment inhibitor BMS-626529 bound to gp120 suggest a unique mechanism of action. *Proteins: Structure, Function, and Bioinformatics* 2015; **83**(2): 331-50.

257.   Lataillade M, Zhou N, Joshi SR, et al. Viral Drug Resistance Through 48 Weeks, in a Phase 2b, Randomized, Controlled Trial of the HIV-1 Attachment Inhibitor Prodrug, Fostemsavir. *Journal of acquired immune deficiency syndromes (1999)* 2018; **77**(3): 299-307.

258.   Pancera M, Lai YT, Bylund T, et al. Crystal structures of trimeric HIV envelope with entry inhibitors BMS-378806 and BMS-626529. *Nat Chem Biol* 2017; **13**(10): 1115-22.

259.   Cihlar T, Ray AS. Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Research* 2010; **85**(1): 39-58.

260.   Weller IV, Williams IG. ABC of AIDS. Antiretroviral drugs. *Bmj* 2001; **322**(7299): 1410-2.

261.   Patel PH, Zulfiqar H. Reverse Transcriptase Inhibitors In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan. 2020. https://www.ncbi.nlm.nih.gov/books/NBK551504/.

262.   Nicolas S-C, Temiz NA, Ivet B. Conformational Changes in HIV-1 Reverse Transcriptase Induced by Nonnucleoside Reverse Transcriptase Inhibitor Binding. *Current HIV Research* 2004; **2**(4): 323-32.

263.   Hazuda DJ, Felock P, Witmer M, et al. Inhibitors of Strand Transfer That Prevent Integration and Inhibit HIV-1 Replication in Cells. *Science* 2000; **287**(5453): 646-50.

264.   Trivedi J, Mahajan D, Jaffe RJ, Acharya A, Mitra D, Byrareddy SN. Recent Advances in the Development of Integrase Inhibitors for HIV Treatment. *Current HIV/AIDS Reports* 2020; **17**(1): 63-75.

265.   WHO. WHO Guidelines Approved by the Guidelines Review Committee. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a Public Health Approach. Geneva: World Health Organization

Copyright © World Health Organization 2016.; 2016.

266.   Hare S, Smith SJ, Métifiot M, et al. Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). *Mol Pharmacol* 2011; **80**(4): 565-72.

267.   Ghosh AK, Osswald HL, Prato G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *J Med Chem* 2016; **59**(11): 5172-208.

268.   Lefebvre E, Schiffer CA. Resilience to resistance of HIV-1 protease inhibitors: profile of darunavir. *AIDS reviews* 2008; **10**(3): 131-42.

269.   Adamson CS. Protease-Mediated Maturation of HIV: Inhibitors of Protease and the Maturation Process. *Mol Biol Int* 2012; **2012**: 604261.

270.   Geretti AM, Easterbrook P. Antiretroviral resistance in clinical practice. *Int J STD AIDS* 2001; **12**(3): 145-53.

271.   Craig J, Duncan I, Hockley D, Grief C, Roberts N, Mills J. Antiviral properties of Ro 31-8959, an inhibitor of human immunodeficiency virus (HIV) proteinase. *Antiviral research* 1991; **16**(4): 295-305.

272.   Kempf DJ, Marsh KC, Denissen JF, et al. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proceedings of the National Academy of Sciences* 1995; **92**(7): 2484-8.

273.   Turner SR, Strohbach JW, Tommasi RA, et al. Tipranavir (PNU-140690): a potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5, 6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *J Med Chem* 1998; **41**(18): 3467-76.

274.   FDA-US. Roche, Fortovase® (saquinavir) Silver Spring, MD. https://www.accessdata.fda.gov/drugsatfda_docs/label/2003/20828s015ppi.pdf (accessed 7th August 2020).

275.   Cameron DW, Japour AJ, Xu Y, et al. Ritonavir and saquinavir combination therapy for the treatment of HIV infection. *Aids* 1999; **13**(2): 213-24.

276.	Lv Z, Chu Y, Wang Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl)* 2015; **7**: 95-104.

277.	Kempf DJ, Marsh KC, Kumar G, et al. Pharmacokinetic enhancement of inhibitors of the human immunodeficiency virus protease by coadministration with ritonavir. *Antimicrobial agents and chemotherapy* 1997; **41**(3): 654-60.

278.	Zeldin RK, Petruschke RA. Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients. *The Journal of antimicrobial chemotherapy* 2004; **53**(1): 4-9.

279.	Sham HL, Kempf DJ, Molla A, et al. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrobial agents and chemotherapy* 1998; **42**(12): 3218-24.

280.	Clair MS, Millard J, Rooney J, et al. In vitro antiviral activity of 141W94 (VX-478) in combination with other antiretroviral agents. *Antiviral research* 1996; **29**(1): 53-6.

281.	Tie Y, Wang YF, Boross PI, et al. Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors. *Protein Science* 2012; **21**(3): 339-50.

282.	Eron J, Jr., Yeni P, Gathe J, Jr., et al. The KLEAN study of fosamprenavir-ritonavir versus lopinavir-ritonavir, each in combination with abacavir-lamivudine, for initial treatment of HIV infection over 48 weeks: a randomised non-inferiority trial. *Lancet* 2006; **368**(9534): 476-82.

283.	Becker S. Atazanavir: improving the HIV protease inhibitor class. *Expert review of anti-infective therapy* 2003; **1**(3): 403-13.

284.	Squires K, Lazzarin A, Gatell JM, et al. Comparison of once-daily atazanavir with efavirenz, each in combination with fixed-dose zidovudine and lamivudine, as initial therapy for patients infected with HIV. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2004; **36**(5): 1011-9.

285.	Murphy RL, Sanne I, Cahn P, et al. Dose-ranging, randomized, clinical trial of atazanavir with lamivudine and stavudine in antiretroviral-naive subjects: 48-week results. *Aids* 2003; **17**(18): 2603-14.

286.	Lefebvre E, Schiffer CA. Resilience to resistance of HIV-1 protease inhibitors: profile of darunavir. *AIDS reviews* 2008; **10**(3): 131.

287.	Tie Y, Boross PI, Wang Y-F, et al. High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *Journal of molecular biology* 2004; **338**(2): 341-52.

288.	Orkin C, DeJesus E, Khanlou H, et al. Final 192-week efficacy and safety of once-daily darunavir/ritonavir compared with lopinavir/ritonavir in HIV-1-

infected treatment-naive patients in the ARTEMIS trial. *HIV medicine* 2013; **14**(1): 49-59.

289. Lathouwers E, Wong EY, Luo D, Seyedkazemi S, De Meyer S, Brown K. HIV-1 resistance rarely observed in subjects using darunavir once-daily regimens across clinical studies. *HIV Clinical Trials* 2017; **18**(5-6): 196-204.

290. Huang L, Chen C. Autoprocessing of human immunodeficiency virus type 1 protease miniprecursor fusions in mammalian cells. *AIDS research and therapy* 2010; **7**(1): 27.

291. Huang D, Caflisch A. How does darunavir prevent HIV-1 protease dimerization? *Journal of chemical theory and computation* 2012; **8**(5): 1786-94.

292. Ghosh AK, Xia Z, Kovela S, et al. Potent HIV-1 Protease Inhibitors Containing Carboxylic and Boronic Acids: Effect on Enzyme Inhibition and Antiviral Activity and Protein-Ligand X-ray Structural Studies. *ChemMedChem* 2019; **14**(21): 1863-72.

293. Brower ET, Bacha UM, Kawasaki Y, Freire E. Inhibition of HIV-2 protease by HIV-1 protease inhibitors in clinical use. *Chemical biology & drug design* 2008; **71**(4): 298-305.

294. Urano E, Timilsina U, Kaplan JA, et al. Resistance to Second-Generation HIV-1 Maturation Inhibitors. *J Virol* 2019; **93**(6).

295. Zhou J, Yuan X, Dismuke D, et al. Small-molecule inhibition of human immunodeficiency virus type 1 replication by specific targeting of the final step of virion maturation. *Journal of virology* 2004; **78**(2): 922-9.

296. Li F, Goila-Gaur R, Salzwedel K, et al. PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. *Proceedings of the National Academy of Sciences* 2003; **100**(23): 13555-60.

297. Gupta S, Louis JM, Tycko R. Effects of an HIV-1 maturation inhibitor on the structure and dynamics of CA-SP1 junction helices in virus-like particles. *Proceedings of the National Academy of Sciences* 2020; **117**(19): 10286-93.

298. McCallister S, Lalezari J, Richmond G, et al. HIV-1 Gag polymorphisms determine treatment response to bevirimat (PA-457). *Antiviral therapy* 2008; **13**(Suppl 3): A10.

299. Wang D, Lu W, Li F. Pharmacological intervention of HIV-1 maturation. *Acta Pharm Sin B* 2015; **5**(6): 493-9.

300. Van Baelen K, Salzwedel K, Rondelez E, et al. Susceptibility of Human Immunodeficiency Virus Type 1 to the Maturation Inhibitor Bevirimat Is Modulated by Baseline Polymorphisms in Gag Spacer Peptide 1. *Antimicrobial agents and chemotherapy* 2009; **53**(5): 2185-8.

301.   Adamson CS, Sakalian M, Salzwedel K, Freed EO. Polymorphisms in Gag spacer peptide 1 confer varying levels of resistance to the HIV-1maturation inhibitor bevirimat. *Retrovirology* 2010; **7**(1): 36.

302.   Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. *J Virol* 2010; **84**(19): 9733-48.

303.   Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol* 2010; **84**(19): 9864-78.

304.   Mansky LM. Retrovirus mutation rates and their role in genetic variation. *J Gen Virol* 1998; **79 (Pt 6)**: 1337-45.

305.   Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science* 1988; **242**(4882): 1171-3.

306.   Ji JP, Loeb LA. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. *Biochemistry* 1992; **31**(4): 954-8.

307.   Cuevas JM, Geller R, Garijo R, López-Aldeguer J, Sanjuán R. Extremely High Mutation Rate of HIV-1 In Vivo. *PLOS Biology* 2015; **13**(9): e1002251.

308.   Menéndez-Arias L. Molecular basis of human immunodeficiency virus drug resistance: an update. *Antiviral research* 2010; **85**(1): 210-31.

309.   Tang MW, Shafer RW. HIV-1 antiretroviral resistance: scientific principles and clinical applications. *Drugs* 2012; **72**(9): e1-25.

310.   Shafer R. Genotypic Testing for HIV-1 Drug Resistance. July, 2019 2004. http://hivinsite.ucsf.edu/InSite?page=kb-03-02-07#S3.3X (accessed 18th August 2020).

311.   Brenner BG, Routy J-P, Petrella M, et al. Persistence and fitness of multidrug-resistant human immunodeficiency virus type 1 acquired in primary infection. *Journal of virology* 2002; **76**(4): 1753-61.

312.   Pao D, Andrady U, Clarke J, et al. Long-term persistence of primary genotypic resistance after HIV-1 seroconversion. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2004; **37**(5): 1570-3.

313.   Inzaule SC, Hamers RL, Bertagnolio S, Siedner MJ, Rinke de Wit TF, Gupta RK. Pretreatment HIV drug resistance in low-and middle-income countries. *Future Virology* 2019; **14**(6): 427-40.

314.   Blair HA. Ibalizumab: A Review in Multidrug-Resistant HIV-1 Infection. *Drugs* 2020; **80**(2): 189-96.

315.   Toma J, Weinheimer SP, Stawiski E, et al. Loss of asparagine-linked glycosylation sites in variable region 5 of human immunodeficiency virus type 1

envelope is associated with resistance to CD4 antibody ibalizumab. *Journal of virology* 2011; **85**(8): 3872-80.

316.   Pace CS, Fordyce MW, Franco D, Kao C-Y, Seaman MS, Ho DD. Anti-CD4 monoclonal antibody ibalizumab exhibits breadth and potency against HIV-1, with natural resistance mediated by the loss of a V5 glycan in envelope. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2013; **62**(1): 1-9.

317.   Pérez-Alvarez L, Carmona R, Ocampo A, et al. Long-term monitoring of genotypic and phenotypic resistance to T20 in treated patients infected with HIV-1. *Journal of medical virology* 2006; **78**(2): 141-7.

318.   Wei X, Decker JM, Liu H, et al. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrobial agents and chemotherapy* 2002; **46**(6): 1896-905.

319.   Ray N, Blackburn LA, Doms RW. HR-2 mutations in human immunodeficiency virus type 1 gp41 restore fusion kinetics delayed by HR-1 mutations that cause clinical resistance to enfuvirtide. *Journal of virology* 2009; **83**(7): 2989-95.

320.   Reeves JD, Gallo SA, Ahmad N, et al. Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics. *Proceedings of the National Academy of Sciences* 2002; **99**(25): 16249-54.

321.   Reeves JD, Miamidian JL, Biscone MJ, et al. Impact of mutations in the coreceptor binding site on human immunodeficiency virus type 1 fusion, infection, and entry inhibitor sensitivity. *Journal of virology* 2004; **78**(10): 5476-85.

322.   Xu L, Pozniak A, Wildfire A, et al. Emergence and evolution of enfuvirtide resistance following long-term therapy involves heptad repeat 2 mutations within gp41. *Antimicrobial agents and chemotherapy* 2005; **49**(3): 1113-9.

323.   Hughes A, Barber T, Nelson M. New treatment options for HIV salvage patients: An overview of second generation PIs, NNRTIs, integrase inhibitors and CCR5 antagonists. *Journal of Infection* 2008; **57**(1): 1-10.

324.   Clavel F, Hance AJ. HIV Drug Resistance. *New England Journal of Medicine* 2004; **350**(10): 1023-35.

325.   Kuritzkes D. Drug resistance. Navigating resistance pathways. *AIDS Read* 2002; **12**(9): 395-400, 7.

326.   Delviks-Frankenberry KA, Nikolenko GN, Pathak VK. The "connection" between HIV drug resistance and RNase H. *Viruses* 2010; **2**(7): 1476-503.

327.   Yap S-H, Sheen C-W, Fahey J, et al. N348I in the Connection Domain of HIV-1 Reverse Transcriptase Confers Zidovudine and Nevirapine Resistance. *PLOS Medicine* 2007; **4**(12): e335.

328.   Kemp SD, Shi C, Bloor S, Harrigan PR, Mellors JW, Larder BA. A Novel Polymorphism at Codon 333 of Human Immunodeficiency Virus Type 1 Reverse Transcriptase Can Facilitate Dual Resistance to Zidovudine and l-2′,3′-Dideoxy-3′-Thiacytidine. *Journal of Virology* 1998; **72**(6): 5093-8.

329.   Nikolenko GN, Delviks-Frankenberry KA, Pathak VK. A novel molecular mechanism of dual resistance to nucleoside and nonnucleoside reverse transcriptase inhibitors. *J Virol* 2010; **84**(10): 5238-49.

330.   Nikolenko GN, Palmer S, Maldarelli F, Mellors JW, Coffin JM, Pathak VK. Mechanism for nucleoside analog-mediated abrogation of HIV-1 replication: balance between RNase H activity and nucleotide excision. *Proceedings of the National Academy of Sciences of the United States of America* 2005; **102**(6): 2093-8.

331.   Wright DW, Deuzing IP, Flandre P, et al. A polymorphism at position 400 in the connection subdomain of HIV-1 reverse transcriptase affects sensitivity to NNRTIs and RNaseH activity. *PloS one* 2013; **8**(10): e74078.

332.   Yap S-H, Sheen C-W, Fahey J, et al. N348I in the connection domain of HIV-1 reverse transcriptase confers zidovudine and nevirapine resistance. *PLoS medicine* 2007; **4**(12): e335-e.

333.   Radzio J, Yap SH, Tachedjian G, Sluis-Cremer N. N348I in reverse transcriptase provides a genetic pathway for HIV-1 to select thymidine analogue mutations and mutations antagonistic to thymidine analogue mutations. *Aids* 2010; **24**(5): 659-67.

334.   Soriano V, Mendoza Cd. Genetic mechanisms of resistance to NRTI and NNRTI. *HIV Clinical Trials* 2002; **3**(3): 237-48.

335.   Seminari E, Castagna A, Lazzarin A. Etravirine for the treatment of HIV infection. *Expert Review of Anti-infective Therapy* 2008; **6**(4): 427-33.

336.   Pommier Y, Johnson AA, Marchand C. Integrase inhibitors to treat HIV/AIDS. *Nature Reviews Drug Discovery* 2005; **4**(3): 236-48.

337.   Hazuda D, Miller M, Nguyen B, Zhao J. Resistance to the HIV-integrase inhibitor raltegravir: analysis of protocol 005, a phase II study in patients with triple-class resistant HIV-1 infection. Antiviral therapy; 2007: INT MEDICAL PRESS LTD 2-4 IDOL LANE, LONDON EC3R 5DD, ENGLAND; 2007. p. S10-S.

338.   Wensing AM, Calvez V, Ceccherini-Silberstein F, et al. 2019 update of the drug resistance mutations in HIV-1. *Top Antivir Med* 2019; **27**(3): 111-21.

339.	Hirsch MS, Günthard HF, Schapiro JM, et al. Antiretroviral Drug Resistance Testing in Adult HIV-1 Infection: 2008 Recommendations of an International AIDS Society-USA Panel. *Clinical Infectious Diseases* 2008; **47**(2): 266-85.

340.	McColl D, Fransen S, Gupta S, et al. Resistance and cross-resistance to first generation integrase inhibitors: insights from a phase II study of elvitegravir (GS-9137). *Antiviral therapy* 2007; **12**(5): S11.

341.	Castagna A, Maggiolo F, Penco G, et al. Dolutegravir in Antiretroviral-Experienced Patients With Raltegravir- and/or Elvitegravir-Resistant HIV-1: 24-Week Results of the Phase III VIKING-3 Study. *The Journal of infectious diseases* 2014; **210**(3): 354-62.

342.	Kuritzkes DR. Resistance to Dolutegravir—A Chink in the Armor? *The Journal of infectious diseases* 2018; **218**(5): 673-5.

343.	Gilead-Sciences Incorporated. BIKTARVY: Highlights of prescribing information. 08/2019 2020. https://www.gilead.com/-/media/files/pdfs/medicines/hiv/biktarvy/biktarvy_pi.pdf (accessed 8th August 2020).

344.	Van Duyne R, Kuo LS, Pham P, Fujii K, Freed EO. 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 of the United States of America* 2019; **116**(18): 9040-9.

345.	Malet I, Subra F, Charpentier C, et al. Mutations Located outside the Integrase Gene Can Confer Resistance to HIV-1 Integrase Strand Transfer Inhibitors. *mBio* 2017; **8**(5).

346.	Shafer RW, Kantor R, Gonzales MJ. The genetic basis of HIV-1 resistance to reverse transcriptase and protease inhibitors. *AIDS reviews* 2000; **2**(4): 211.

347.	Molla A, Korneyeva M, Gao Q, et al. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nature medicine* 1996; **2**(7): 760-6.

348.	Nijhuis M, Deeks S, Boucher C. Implications of antiretroviral resistance on viral fitness. *Current opinion in infectious diseases* 2001; **14**(1): 23-8.

349.	Quiñones-Mateu ME, Arts EJ. HIV-1 fitness: implications for drug resistance, disease progression, and global epidemic evolution. *HIV sequence compendium* 2001; **2001**: 134-70.

350.	Eastman PS, Mittler J, Kelso R, et al. Genotypic changes in human immunodeficiency virus type 1 associated with loss of suppression of plasma viral RNA levels in subjects treated with ritonavir (Norvir) monotherapy. *Journal of virology* 1998; **72**(6): 5154-64.

351. Ho DD, Toyoshima T, Mo H, et al. Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C2-symmetric protease inhibitor. *Journal of Virology* 1994; **68**(3): 2016-20.

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

353. Nijhuis M, Schuurman R, de Jong D, et al. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *Aids* 1999; **13**(17): 2349-59.

354. Doherty KM, Nakka P, King BM, et al. A multifaceted analysis of HIV-1 protease multidrug resistance phenotypes. *BMC bioinformatics* 2011; **12**(1): 477.

355. Rabi SA, Laird GM, Durand CM, et al. Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. *J Clin Invest* 2013; **123**(9): 3848-60.

356. Kagan RM, Shenderovich MD, Heseltine PN, Ramnarayan K. Structural analysis of an HIV-1 protease I47A mutant resistant to the protease inhibitor lopinavir. *Protein Sci* 2005; **14**(7): 1870-8.

357. Wu J, Matunis MJ, Kraemer D, Blobel G, Coutavas E. Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *J Biol Chem* 1995; **270**(23): 14209-13.

358. Nijhuis M, van Maarseveen NM, Boucher CA. HIV protease resistance and viral fitness. *Current opinion in HIV and AIDS* 2007; **2**(2): 108-15.

359. Louis JM, Zhang Y, Sayer JM, Wang Y-F, Harrison RW, Weber IT. The L76V drug resistance mutation decreases the dimer stability and rate of autoprocessing of HIV-1 protease by reducing internal hydrophobic contacts. *Biochemistry* 2011; **50**(21): 4786-95.

360. Bastys T, Gapsys V, Walter H, et al. Non-active site mutants of HIV-1 protease influence resistance and sensitisation towards protease inhibitors. *Retrovirology* 2020; **17**(1): 13.

361. Colonno R, Rose R, McLaren C, Thiry A, Parkin N, Friborg J. Identification of I50L as the Signature Atazanavir (ATV)-Resistance Mutation in Treatment-Naive HIV-1-Infected Patients Receiving ATV-Containing Regimens. *The Journal of infectious diseases* 2004; **189**(10): 1802-10.

362. Guffanti M, Caumo A, Galli L, et al. Switching to unboosted atazanavir improves glucose tolerance in highly pretreated HIV-1 infected subjects. *European journal of endocrinology* 2007; **156**(4): 503-9.

363.   Young TP, Parkin NT, Stawiski E, et al. Prevalence, mutation patterns, and effects on protease inhibitor susceptibility of the L76V mutation in HIV-1 protease. *Antimicrobial agents and chemotherapy* 2010; **54**(11): 4903-6.

364.   Colonno RJ, Thiry A, Limoli K, Parkin N. Activities of atazanavir (BMS-232632) against a large panel of human immunodeficiency virus type 1 clinical isolates resistant to one or more approved protease inhibitors. *Antimicrobial agents and chemotherapy* 2003; **47**(4): 1324-33.

365.   Gong Y-F, Robinson BS, Rose RE, et al. In vitro resistance profile of the human immunodeficiency virus type 1 protease inhibitor BMS-232632. *Antimicrobial agents and chemotherapy* 2000; **44**(9): 2319-26.

366.   De Meyer S, Descamps D, Van Baelen B, et al. Confirmation of the negative impact of protease mutations I47V, I54M, T74P and I84V and the positive impact of protease mutation V82A on virological response to darunavir/ritonavir. *Antiviral therapy* 2009; **14**(Suppl 1): A147.

367.   Descamps D, Lambert-Niclot S, Marcelin A-G, et al. Mutations associated with virological response to darunavir/ritonavir in HIV-1-infected protease inhibitor-experienced patients. *Journal of antimicrobial chemotherapy* 2009; **63**(3): 585-92.

368.   Janssen, Therapeutic. Prezista® (Darunavir) Highlights of prescribing Information. 05/2019 2020. http://www.janssenlabels.com/package-insert/product-monograph/prescribing-information/PREZISTA-pi.pdf (accessed 9th August 2020).

369.   Koh Y, Matsumi S, Das D, et al. Potent inhibition of HIV-1 replication by novel non-peptidyl small molecule inhibitors of protease dimerization. *Journal of Biological Chemistry* 2007; **282**(39): 28709-20.

370.   Hayashi H, Takamune N, Nirasawa T, et al. Dimerization of HIV-1 protease occurs through two steps relating to the mechanism of protease dimerization inhibition by darunavir. *Proceedings of the National Academy of Sciences* 2014; **111**(33): 12234-9.

371.   Kempf DJ, Isaacson JD, King MS, et al. Identification of genotypic changes in human immunodeficiency virus protease that correlate with reduced susceptibility to the protease inhibitor lopinavir among viral isolates from protease inhibitor-experienced patients. *J Virol* 2001; **75**(16): 7462-9.

372.   Mo H, King MS, King K, Molla A, Brun S, Kempf DJ. Selection of resistance in protease inhibitor-experienced, human immunodeficiency virus type 1-infected subjects failing lopinavir- and ritonavir-based therapy: mutation patterns and baseline correlates. *J Virol* 2005; **79**(6): 3329-38.

373.   Masquelier B, Breilh D, Neau D, et al. Human immunodeficiency virus type 1 genotypic and pharmacokinetic determinants of the virological response

to lopinavir-ritonavir-containing therapy in protease inhibitor-experienced patients. *Antimicrobial agents and chemotherapy* 2002; **46**(9): 2926-32.

374.   AbbVie-Incorporated. KALETRA® (Lopinavir/Ritonavir): Highlights of prescribing information. 04/2020 2020. https://www.rxabbvie.com/pdf/kaletratabpi.pdf (accessed 9th August 2020).

375.   Hermans L, Steegen K, ter Heine R, et al. PI drug-level testing as screening tool for drug resistance in 2nd-line ART failure. *Top Antivir Med* 2019; **27**(1s): 169s.

376.   Rhee SY, Taylor J, Fessel WJ, et al. HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrobial agents and chemotherapy* 2010; **54**(10): 4253-61.

377.   Lam E, Parkin NT. Amprenavir resistance imparted by the I50V mutation in HIV-1 protease can be suppressed by the N88S mutation. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2003; **37**(9): 1273-4.

378.   Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 2003; **31**(1): 298-303.

379.   Maguire MF, Guinea R, Griffin P, et al. Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *Journal of virology* 2002; **76**(15): 7398-406.

380.   Ait-Khaled M, Rakik A, Griffin P, et al. HIV-1 reverse transcriptase and protease resistance mutations selected during 16-72 weeks of therapy in isolates from antiretroviral therapy-experienced patients receiving abacavir/efavirenz/amprenavir in the CNA2007 study. *Antiviral therapy* 2003; **8**(2): 111-20.

381.   Partaledis JA, Yamaguchi K, Tisdale M, et al. In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) isolates with reduced sensitivity to hydroxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. *J Virol* 1995; **69**(9): 5228-35.

382.   Tisdale M, Myers RE, Maschera B, Parry NR, Oliver NM, Blair ED. Cross-resistance analysis of human immunodeficiency virus type 1 variants individually selected for resistance to five different protease inhibitors. *Antimicrobial agents and chemotherapy* 1995; **39**(8): 1704-10.

383.   Condra JH, Holder DJ, Schleif WA, et al. Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor. *J Virol* 1996; **70**(12): 8270-6.

384.	Zhang Y-M, Imamichi H, Imamichi T, et al. Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *Journal of virology* 1997; **71**(9): 6662-70.

385.	Walmsley SL, Becker MI, Zhang M, Humar A, Harrigan PR. Predictors of virological response in HIV-infected patients to salvage antiretroviral therapy that includes nelfinavir. *Antiviral therapy* 2001; **6**(1): 47-54.

386.	Perno CF, Cozzi-Lepri A, Balotta C, et al. Secondary mutations in the protease region of human immunodeficiency virus and virologic failure in drug-naive patients treated with protease inhibitor-based therapy. *The Journal of infectious diseases* 2001; **184**(8): 983-91.

387.	Johnston E, Winters MA, Rhee S-Y, Merigan TC, Schiffer CA, Shafer RW. Association of a novel human immunodeficiency virus type 1 protease substrate cleft mutation, L23I, with protease inhibitor therapy and in vitro drug resistance. *Antimicrobial agents and chemotherapy* 2004; **48**(12): 4864-8.

388.	Pellegrin I, Breilh D, Montestruc F, et al. Virologic response to nelfinavir-based regimens: pharmacokinetics and drug resistance mutations (VIRAPHAR study). *Aids* 2002; **16**(10): 1331-40.

389.	Jacobsen H, Haenggi M, Ott M, et al. Reduced sensitivity to saquinavir: an update on genotyping from phase I/II trials. *Antiviral Res* 1996; **29**(1): 95-7.

390.	Rhee SY, Taylor J, Wadhera G, Ben-Hur A, Brutlag DL, Shafer RW. Genotypic predictors of human immunodeficiency virus type 1 drug resistance. *Proceedings of the National Academy of Sciences of the United States of America* 2006; **103**(46): 17355-60.

391.	Doyon L, Tremblay S, Bourgon L, Wardrop E, Cordingley MG. Selection and characterization of HIV-1 showing reduced susceptibility to the non-peptidic protease inhibitor tipranavir. *Antiviral research* 2005; **68**(1): 27-35.

392.	Baxter JD, Schapiro JM, Boucher CA, et al. Genotypic changes in human immunodeficiency virus type 1 protease associated with reduced susceptibility and virologic response to the protease inhibitor tipranavir. *J Virol* 2006; **80**(21): 10794-801.

393.	Vermeiren H, Van Craenenbroeck E, Alen P, Bacheler L, Picchio G, Lecocq P. Prediction of HIV-1 drug susceptibility phenotype from the viral genotype using linear regression modeling. *J Virol Methods* 2007; **145**(1): 47-55.

394.	Fun A, Wensing AM, Verheyen J, Nijhuis M. Human Immunodeficiency Virus Gag and protease: partners in resistance. *Retrovirology* 2012; **9**: 63.

395.	Clavel F, Mammano F. Role of Gag in HIV Resistance to Protease Inhibitors. *Viruses* 2010; **2**(7): 1411-26.

396. Sutherland KA, Parry CM, McCormick A, et al. Evidence for Reduced Drug Susceptibility without Emergence of Major Protease Mutations following Protease Inhibitor Monotherapy Failure in the SARA Trial. *PloS one* 2015; **10**(9): e0137834.

397. Gupta RK, Kohli A, McCormick AL, Towers GJ, Pillay D, Parry CM. Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays. *AIDS* 2010; **24**(11): 1651-5.

398. Ghosn J, Delaugerre C, Flandre P, et al. Polymorphism in Gag gene cleavage sites of HIV-1 non-B subtype and virological outcome of a first-line lopinavir/ritonavir single drug regimen. *PloS one* 2011; **6**(9): e24798.

399. Weber IT, Agniswamy J. HIV-1 protease: structural perspectives on drug resistance. *Viruses* 2009; **1**(3): 1110-36.

400. van Maarseveen NM, Andersson D, Lepsik M, et al. Modulation of HIV-1 Gag NC/p1 cleavage efficiency affects protease inhibitor resistance and viral replicative capacity. *Retrovirology* 2012; **9**: 29.

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

402. Carrillo A, Stewart KD, Sham HL, et al. In vitro selection and characterization of human immunodeficiency virus type 1 variants with increased resistance to ABT-378, a novel protease inhibitor. *Journal of Virology* 1998; **72**(9): 7532-41.

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

404. Cote HC, Brumme ZL, Harrigan PR. Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir. *J Virol* 2001; **75**(2): 589-94.

405. Zhang YM, Imamichi H, Imamichi T, et al. Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol* 1997; **71**(9): 6662-70.

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

407. Larrouy L, Chazallon C, Landman R, et al. Gag mutations can impact virological response to dual-boosted protease inhibitor combinations in

antiretroviral-naive HIV-infected patients. *Antimicrobial agents and chemotherapy* 2010; **54**(7): 2910-9.

408.   Bally F, Martinez R, Peters S, Sudre P, Telenti A. Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors. *AIDS research and human retroviruses* 2000; **16**(13): 1209-13.

409.   Yates P, Hazen R, Clair MS, Boone L, Tisdale M, Elston R. In vitro development of resistance to human immunodeficiency virus protease inhibitor GW640385. *Antimicrobial agents and chemotherapy* 2006; **50**(3): 1092-5.

410.   Callebaut C, Stray K, Tsai L, Xu L, Lee W, Cihlar T. In vitro HIV-1 resistance selection to GS-8374, novel phosphonate protease inhibitor: comparison with lopinavir, atazanavir and darunavir. Antiviral therapy; 2007: INT MEDICAL PRESS LTD 2-4 IDOL LANE, LONDON EC3R 5DD, ENGLAND; 2007. p. S18-S.

411.   Chang MW, Oliveira G, Yuan J, Okulicz JF, Levy S, Torbett BE. Rapid deep sequencing of patient-derived HIV with ion semiconductor technology. *Journal of Virological Methods* 2013; **189**(1): 232-4.

412.   Dierynck I, De Meyer S, CV K, et al. Impact of Gag cleavage site mutations on the virological response to darunavir/ritonavir in treatment-experienced patients in POWER 1, 2 and 3. *Antiviral therapy* 2007; **12**(5): S23.

413.   Larrouy L, Charpentier C, Landman R, et al. Dynamics of gag-pol minority viral populations in naive HIV-1-infected patients failing protease inhibitor regimen. *Aids* 2011; **25**(17): 2143-8.

414.   Malet I, Roquebert B, Dalban C, et al. Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. *The Journal of infection* 2007; **54**(4): 367-74.

415.   Meynard JL, Vray M, Morand-Joubert L, et al. Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. *Aids* 2002; **16**(5): 727-36.

416.   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. *Journal of the International AIDS Society* 2017; **20**(3).

417.   Parry CM, Kolli M, Myers RE, Cane PA, Schiffer C, Pillay D. Three residues in HIV-1 matrix contribute to protease inhibitor susceptibility and replication capacity. *Antimicrobial agents and chemotherapy* 2011; **55**(3): 1106-13.

418.   Aoki M, Venzon DJ, Koh Y, et al. Non-cleavage site gag mutations in amprenavir-resistant human immunodeficiency virus type 1 (HIV-1) predispose

HIV-1 to rapid acquisition of amprenavir resistance but delay development of resistance to other protease inhibitors. *J Virol* 2009; **83**(7): 3059-68.

419. Kletenkov K, Hoffmann D, Boni J, et al. Role of Gag mutations in PI resistance in the Swiss HIV cohort study: bystanders or contributors? *The Journal of antimicrobial chemotherapy* 2017; **72**(3): 866-75.

420. Parry CM, Kohli A, Boinett CJ, Towers GJ, McCormick AL, Pillay D. Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1. *J Virol* 2009; **83**(18): 9094-101.

421. Codoner FM, Pena R, Blanch-Lombarte O, et al. Gag-protease coevolution analyses define novel structural surfaces in the HIV-1 matrix and capsid involved in resistance to Protease Inhibitors. *Scientific reports* 2017; **7**(1): 3717.

422. Gatanaga H, Y. Suzuki, H. Tsang, K. Yoshimura, M. F. Kavlick, K. Nagashima, R. J. Gorelick, S. Mardy, C. Tang, M. F. Summers, and H. Mitsuya. 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**(Feb 22): 5952–61.

423. Kameoka M, Isarangkura-na-Ayuthaya P, Kameoka Y, et al. The role of lysine residue at amino acid position 165 of human immunodeficiency virus type 1 CRF01_AE Gag in reducing viral drug susceptibility to protease inhibitors. *Virology* 2010; **405**(1): 129-38.

424. Myint L, Matsuda M, Matsuda Z, et al. Gag non-cleavage site mutations contribute to full recovery of viral fitness in protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrobial agents and chemotherapy* 2004; **48**(2): 444-52.

425. Prabu-Jeyabalan M, Nalivaika EA, King NM, Schiffer CA. Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: structural insights for better antiviral therapy. *J Virol* 2003; **77**(2): 1306-15.

426. Zennou V, Mammano F, Paulous S, Mathez D, Clavel F. Loss of viral fitness associated with multiple Gag and Gag-Pol processing defects in human immunodeficiency virus type 1 variants selected for resistance to protease inhibitors in vivo. *J Virol* 1998; **72**(4): 3300-6.

427. Koh Y, Amano M, Towata T, et al. In vitro selection of highly darunavir-resistant and replication-competent HIV-1 variants by using a mixture of clinical HIV-1 isolates resistant to multiple conventional protease inhibitors. *J Virol* 2010; **84**(22): 11961-9.

428. Sharma S, Aralaguppe SG, Abrahams MR, et al. The PTAP sequence duplication in HIV-1 subtype C Gag p6 in drug-naive subjects of India and South Africa. *BMC Infect Dis* 2017; **17**(1): 95.

429.	Soriano V, de Mendoza C. Genetic mechanisms of resistance to NRTI and NNRTI. *HIV Clin Trials* 2002; **3**(3): 237-48.

430.	Aralaguppe SG, Winner D, Singh K, et al. Increased replication capacity following evolution of PYxE insertion in Gag-p6 is associated with enhanced virulence in HIV-1 subtype C from East Africa. *Journal of medical virology* 2017; **89**(1): 106-11.

431.	Neogi U, Rao SD, Bontell I, et al. Novel tetra-peptide insertion in Gag-p6 ALIX-binding motif in HIV-1 subtype C associated with protease inhibitor failure in Indian patients. *AIDS* 2014; **28**(15): 2319-22.

432.	De Luca A. The impact of resistance on viral fitness and its clinical implications. Antiretroviral resistance in clinical practice: Mediscript; 2006.

433.	Devereux HL, Emery VC, Johnson MA, Loveday C. Replicative fitness in vivo of HIV-1 variants with multiple drug resistance-associated mutations. *Journal of medical virology* 2001; **65**(2): 218-24.

434.	Menzo S, Monachetti A, Balotta C, et al. Processivity and drug-dependence of HIV-1 protease: determinants of viral fitness in variants resistant to protease inhibitors. *Aids* 2003; **17**(5): 663-71.

435.	Martinez-Picado J, Savara AV, Sutton L, Richard T. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *Journal of virology* 1999; **73**(5): 3744-52.

436.	Koh Y, Das D, Leschenko S, et al. GRL-02031, a novel nonpeptidic protease inhibitor (PI) containing a stereochemically defined fused cyclopentanyltetrahydrofuran potent against multi-PI-resistant human immunodeficiency virus type 1 in vitro. *Antimicrobial agents and chemotherapy* 2009; **53**(3): 997-1006.

437.	Blanch-Lombarte O, Santos JR, Peña R, et al. HIV-1 Gag mutations alone are sufficient to reduce darunavir susceptibility during virological failure to boosted PI therapy. *Journal of Antimicrobial Chemotherapy* 2020.

438.	Knops E, Daumer M, Awerkiew S, et al. Evolution of protease inhibitor resistance in the gag and pol genes of HIV subtype G isolates. *The Journal of antimicrobial chemotherapy* 2010; **65**(7): 1472-6.

439.	Jinnopat P, Isarangkura-na-ayuthaya P, Utachee P, et al. Impact of amino acid variations in Gag and protease of HIV type 1 CRF01_AE strains on drug susceptibility of virus to protease inhibitors. *Journal of acquired immune deficiency syndromes (1999)* 2009; **52**(3): 320-8.

440.	Margot NA, Gibbs CS, Miller MD. Phenotypic susceptibility to bevirimat in isolates from HIV-1-infected patients without prior exposure to bevirimat. *Antimicrobial agents and chemotherapy* 2010; **54**(6): 2345-53.

441.   Verheyen J, Verhofstede C, Knops E, et al. High prevalence of bevirimat resistance mutations in protease inhibitor-resistant HIV isolates. *Aids* 2010; **24**(5): 669-73.

442.   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**(1): 126.

443.   Kolli M, Stawiski E, Chappey C, Schiffer CA. Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. *J Virol* 2009; **83**(21): 11027-42.

444.   Verheyen J, Litau E, Sing T, et al. Compensatory mutations at the HIV cleavage sites p7/p1 and p1/p6-gag in therapy-naive and therapy-experienced patients. *Antiviral therapy* 2006; **11**(7): 879-87.

445.   Breuer S, Sepulveda H, Chen Y, Trotter J, Torbett BE. A cleavage enzyme-cytometric bead array provides biochemical profiling of resistance mutations in HIV-1 Gag and protease. *Biochemistry* 2011; **50**(20): 4371-81.

446.   McKinnon JE, Delgado R, Pulido F, Shao W, Arribas JR, Mellors JW. Single genome sequencing of HIV-1 gag and protease resistance mutations at virologic failure during the OK04 trial of simplified versus standard maintenance therapy. *Antiviral therapy* 2011; **16**(5): 725.

447.   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**(18): 3219-23.

448.   Shibata J, Sugiura W, Ode H, et al. Within-host co-evolution of Gag P453L and protease D30N/N88D demonstrates virological advantage in a highly protease inhibitor-exposed HIV-1 case. *Antiviral Res* 2011; **90**(1): 33-41.

449.   Özen A, Lin K-H, Kurt Yilmaz N, Schiffer CA. Structural basis and distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease. *Proceedings of the National Academy of Sciences* 2014; **111**(45): 15993-8.

450.   Steegen K, Bronze M, Papathanasopoulos MA, et al. Prevalence of Antiretroviral Drug Resistance in Patients Who Are Not Responding to Protease Inhibitor-Based Treatment: Results From the First National Survey in South Africa. *The Journal of infectious diseases* 2016; **214**(12): 1826-30.

451.   Chin-Hong PV, Deeks SG, Liegler T, et al. High-risk sexual behavior in adults with genotypically proven antiretroviral-resistant HIV infection. *Journal of acquired immune deficiency syndromes (1999)* 2005; **40**(4): 463-71.

452.   Little SJ, Holte S, Routy JP, et al. Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 2002; **347**(6): 385-94.

453.   Barbour JD, Hecht FM, Wrin T, et al. Persistence of primary drug resistance among recently HIV-1 infected adults. *Aids* 2004; **18**(12): 1683-9.

454.   Gianotti N, Lazzarin A. Sequencing antiretroviral drugs for long-lasting suppression of HIV replication. *New Microbiol* 2005; **28**(4): 281-97.

455.   van Maarseveen N, Boucher C. Resistance to protease inhibitors. In: Geretti AM, ed. Antiretroviral Resistance in Clinical Practice. London: Mediscript

Mediscript.; 2006.

456.   Sutherland K. Contribution of Gag and protease to variation in susceptibility to protease inhibitors between different strains of HIV-1 [PhD]. London: University College London; 2014.

457.   Petropoulos CJ, Parkin NT, Limoli KL, et al. A Novel Phenotypic Drug Susceptibility Assay for Human Immunodeficiency Virus Type 1. *Antimicrobial agents and chemotherapy* 2000; **44**(4): 920-8.

458.   Garcia-Diaz A. An investigation of the role of HIV-1 Gag mutations in failure of protease inhibitors [PhD Thesis]. London: University College London; 2012.

459.   Quiñones-Mateu ME, Arts EJ. Fitness of drug resistant HIV-1: methodology and clinical implications. *Drug Resistance Updates* 2002; **5**(6): 224-33.

460.   Dykes C, Wu H, Sims M, Holden-Wiltse J, Demeter LM. Human immunodeficiency virus type 1 protease inhibitor drug-resistant mutants give discordant results when compared in single-cycle and multiple-cycle fitness assays. *Journal of clinical microbiology* 2010; **48**(11): 4035-43.

461.   Dykes C, Demeter LM. Clinical significance of human immunodeficiency virus type 1 replication fitness. *Clin Microbiol Rev* 2007; **20**(4): 550-78.

462.   Domingo E, Holland JJ. RNA virus mutations and fitness for survival. *Annu Rev Microbiol* 1997; **51**: 151-78.

463.   Domingo E, Menéndez-Arias L, Holland JJ. RNA virus fitness. *Reviews in medical virology* 1997; **7**(2): 87-96.

464.   Holland JJ, De La Torre JC, Clarke D, Duarte E. Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. *Journal of Virology* 1991; **65**(6): 2960-7.

465.   Group TAS. A trial of early antiretrovirals and isoniazid preventive therapy in Africa. *New England Journal of Medicine* 2015; **373**(9): 808-22.

466.   Group ISS. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *New England Journal of Medicine* 2015; **373**(9): 795-807.

467.   Rosen S, Maskew M, Fox MP, et al. Initiating Antiretroviral Therapy for HIV at a Patient's First Clinic Visit: The RapIT Randomized Controlled Trial. *PLOS Medicine* 2016; **13**(5): e1002015.

468.   Koenig S, Dorvil N, Severe P, et al. Same-day HIV testing and antiretroviral therapy initiation results in higher rates of treatment initiation and retention in care. Journal of the International AIDS Society; 2016: JOHN WILEY & SONS LTD THE ATRIUM, SOUTHERN GATE, CHICHESTER PO19 8SQ, W …; 2016.

469.   Organization WH. Guidelines for managing advanced HIV disease and rapid initiation of antiretroviral therapy, July 2017. 2017.

470.   Arts E, Hazuda D. HIV-1 Antiretroviral Drug Therapy. Cold Spring Harbor perspectives in medicine. 2 (4). 2012.

471.   Tedbury PR, Freed EO. HIV-1 gag: an emerging target for antiretroviral therapy. *Current topics in microbiology and immunology* 2015; **389**: 171-201.

472.   WHO. Updated recommendations on first-line and second-line antiretroviral regimens and post-exposure prophylaxis and recommendations on early infant diagnosis of HIV: interim guidelines. Supplement to the 2016 consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection. 2018. https://apps.who.int/iris/bitstream/handle/10665/277395/WHO-CDS-HIV-18.51-eng.pdf?ua=1 (accessed 10th August 2020).

473.   Venter WDF, Moorhouse M, Sokhela S, et al. Dolutegravir plus Two Different Prodrugs of Tenofovir to Treat HIV. *New England Journal of Medicine* 2019; **381**(9): 803-15.

474.   WHO. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: Recommendations for a public health approach - Second edition. 2016. https://www.who.int/hiv/pub/arv/arv-2016/en/.

475.   Hermes A, Squires K, Fredrick L, et al. Meta-Analysis of the Safety, Tolerability, and Efficacy of Lopinavir/Ritonavir-Containing Antiretroviral Therapy in HIV-1–Infected Women. *HIV clinical trials* 2012; **13**(6): 308-23.

476.   Huang Y, Huang X, Luo Y, et al. Assessing the Efficacy of Lopinavir/Ritonavir-Based Preferred and Alternative Second-Line Regimens in HIV-Infected Patients: A Meta-Analysis of Key Evidence to Support WHO Recommendations. *Frontiers in Pharmacology* 2018; **9**(890).

477.   Gupta R, Hill A, Sawyer AW, Pillay D. Emergence of drug resistance in HIV type 1-infected patients after receipt of first-line highly active antiretroviral therapy: a systematic review of clinical trials. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2008; **47**(5): 712-22.

478.	FMoH, Nigeria. National guidelines for HIV prevention, treatment and care. Abuja: FMOH Abuja, Nigeria; 2016. p. 250.

479.	WHO. Update of recommendations on first-and second-line antiretroviral regimens. 2019 2019. https://apps.who.int/iris/bitstream/handle/10665/325892/WHO-CDS-HIV-19.15-eng.pdf?ua=1 (accessed 10th August 2020).

480.	Bannister WP, Ruiz L, Loveday C, et al. HIV-1 subtypes and response to combination antiretroviral therapy in Europe. *Antiviral therapy* 2006; **11**(6): 707-15.

481.	Geretti AM, Harrison L, Green H, et al. Effect of HIV-1 subtype on virologic and immunologic response to starting highly active antiretroviral therapy. *Clinical infectious diseases* 2009; **48**(9): 1296-305.

482.	Brenner B, Turner D, Oliveira M, et al. A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. *Aids* 2003; **17**(1): F1-F5.

483.	Grossman Z, Istomin V, Averbuch D, et al. Genetic variation at NNRTI resistance-associated positions in patients infected with HIV-1 subtype C. *Aids* 2004; **18**(6): 909-15.

484.	Palma AC, Covens K, Snoeck J, Vandamme A-M, Camacho RJ, Van Laethem K. HIV-1 protease mutation 82M contributes to phenotypic resistance to protease inhibitors in subtype G. *Journal of antimicrobial chemotherapy* 2012; **67**(5): 1075-9.

485.	Abecasis AB, Deforche K, Snoeck J, et al. Protease mutation M89I/V is linked to therapy failure in patients infected with the HIV-1 non-B subtypes C, F or G. *Aids* 2005; **19**(16): 1799-806.

486.	Coutsinos D, Invernizzi CF, Xu H, et al. Template usage is responsible for the preferential acquisition of the K65R reverse transcriptase mutation in subtype C variants of human immunodeficiency virus type 1. *Journal of virology* 2009; **83**(4): 2029-33.

487.	Invernizzi CF, Coutsinos D, Oliveira M, Moisi D, Brenner BG, Wainberg MA. Signature nucleotide polymorphisms at positions 64 and 65 in reverse transcriptase favor the selection of the K65R resistance mutation in HIV-1 subtype C. *The Journal of infectious diseases* 2009; **200**(8): 1202-6.

488.	Wainberg MA, Brenner BG. The impact of HIV genetic polymorphisms and subtype differences on the occurrence of resistance to antiretroviral drugs. *Molecular biology international* 2012; **2012**.

489.	Sutherland KA, Ghosn J, Gregson J, et al. HIV-1 subtype influences susceptibility and response to monotherapy with the protease inhibitor

lopinavir/ritonavir. *The Journal of antimicrobial chemotherapy* 2015; **70**(1): 243-8.

490.   Dick A, Cocklin S. Recent Advances in HIV-1 Gag Inhibitor Design and Development. *Molecules* 2020; **25**(7).

491.   Tzou PL, Rhee S-Y, Pond SLK, Manasa J, Shafer RW. Selection analyses of paired HIV-1 gag and gp41 sequences obtained before and after antiretroviral therapy. *Scientific Data* 2018; **5**(1): 180147.

492.   Boom R, Sol C, Wertheim-van Dillen P. Rapid purification of ribosomal RNAs from neutral agarose gels. *Nucleic Acids Res* 1990; **18**(8): 2195.

493.   Van Laethem K, Schrooten Y, Dedecker S, et al. A genotypic assay for the amplification and sequencing of gag and protease from diverse human immunodeficiency virus type 1 group M subtypes. *J Virol Methods* 2006; **132**(1-2): 181-6.

494.   de Oliveira T, Deforche K, Cassol S, et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 2005; **21**(19): 3797-800.

495.   Liu TF, Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2006; **42**(11): 1608-18.

496.   Wensing AM, Calvez V, Gunthard HF, et al. 2017 Update of the Drug Resistance Mutations in HIV-1. *Top Antivir Med* 2017; **24**(4): 132-3.

497.   Boender TS, Sigaloff KC, McMahon JH, et al. Long-term Virological Outcomes of First-Line Antiretroviral Therapy for HIV-1 in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2015; **61**(9): 1453-61.

498.   TenoRes Study G. Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study. *Lancet Infect Dis* 2016; **16**(5): 565-75.

499.   Gregson J, Kaleebu P, Marconi VC, et al. Occult HIV-1 drug resistance to thymidine analogues following failure of first-line tenofovir combined with a cytosine analogue and nevirapine or efavirenz in sub Saharan Africa: a retrospective multi-centre cohort study. *Lancet Infect Dis* 2017; **17**(3): 296-304.

500.   Rawizza HE, Chaplin B, Meloni ST, et al. Accumulation of protease mutations among patients failing second-line antiretroviral therapy and response to salvage therapy in Nigeria. *PloS one* 2013; **8**(9): e73582.

501.   Collier D, Iwuji C, Derache A, et al. Virological Outcomes of Second-line Protease Inhibitor-Based Treatment for Human Immunodeficiency Virus Type 1

in a High-Prevalence Rural South African Setting: A Competing-Risks Prospective Cohort Analysis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2017; **64**(8): 1006-16.

502. Hosseinipour MC, Gupta RK, Van Zyl G, Eron JJ, Nachega JB. Emergence of HIV drug resistance during first- and second-line antiretroviral therapy in resource-limited settings. *The Journal of infectious diseases* 2013; **207 Suppl 2**: S49-56.

503. Ajose O, Mookerjee S, Mills EJ, Boulle A, Ford N. Treatment outcomes of patients on second-line antiretroviral therapy in resource-limited settings: a systematic review and meta-analysis. *AIDS* 2012; **26**(8): 929-38.

504. Stockdale AJ, Saunders MJ, Boyd MA, et al. Effectiveness of Protease Inhibitor/Nucleos(t)ide Reverse Transcriptase Inhibitor-Based Second-line Antiretroviral Therapy for the Treatment of Human Immunodeficiency Virus Type 1 Infection in Sub-Saharan Africa: A Systematic Review and Meta-analysis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2018; **66**(12): 1846-57.

505. Sutherland KA, Collier DA, Claiborne DT, et al. Wide variation in susceptibility of transmitted/founder HIV-1 subtype C Isolates to protease inhibitors and association with in vitro replication efficiency. *Scientific reports* 2016; **6**: 38153.

506. Stray KM, Callebaut C, Glass B, et al. Mutations in multiple domains of Gag drive the emergence of in vitro resistance to the phosphonate-containing HIV-1 protease inhibitor GS-8374. *J Virol* 2013; **87**(1): 454-63.

507. Sutherland KA, Mbisa JL, Cane PA, Pillay D, Parry CM. Contribution of Gag and protease to variation in susceptibility to protease inhibitors between different strains of subtype B human immunodeficiency virus type 1. *J Gen Virol* 2014; **95**(Pt 1): 190-200.

508. Sutherland KA, Parry CM, McCormick A, et al. Evidence for Reduced Drug Susceptibility without Emergence of Major Protease Mutations following Protease Inhibitor Monotherapy Failure in the SARA Trial. *PloS one* 2015; **10**(9): e0137834.

509. Sutherland KA, Goodall RL, McCormick A, et al. Gag-Protease Sequence Evolution Following Protease Inhibitor Monotherapy Treatment Failure in HIV-1 Viruses Circulating in East Africa. *AIDS research and human retroviruses* 2015; **31**(10): 1032-7.

510. Giandhari J, Basson AE, Coovadia A, et al. Genetic Changes in HIV-1 Gag-Protease Associated with Protease Inhibitor-Based Therapy Failure in Pediatric Patients. *AIDS research and human retroviruses* 2015; **31**(8): 776-82.

511. Li G, Verheyen J, Theys K, Piampongsant S, Van Laethem K, Vandamme AM. HIV-1 Gag C-terminal amino acid substitutions emerging under

selective pressure of protease inhibitors in patient populations infected with different HIV-1 subtypes. *Retrovirology* 2014; **11**: 79.

512.   Sutherland KA, Mbisa JL, Ghosn J, et al. Phenotypic characterization of virological failure following lopinavir/ritonavir monotherapy using full-length Gag-protease genes. *The Journal of antimicrobial chemotherapy* 2014; **69**(12): 3340-8.

513.   Giandhari J, Basson AE, Sutherland K, et al. Contribution of Gag and Protease to HIV-1 Phenotypic Drug Resistance in Pediatric Patients Failing Protease Inhibitor-Based Therapy. *Antimicrobial agents and chemotherapy* 2016; **60**(4): 2248-56.

514.   Chaplin B, Akanmu AS, Inzaule SC, et al. Association between HIV-1 subtype and drug resistance in Nigerian infants. *The Journal of antimicrobial chemotherapy* 2019; **74**(1): 172-6.

515.   Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 2018; **35**(6): 1547-9.

516.   Gupta RK, Gregson J, Parkin N, et al. HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. *Lancet Infect Dis* 2018; **18**(3): 346-55.

517.   Miller MD, Margot N, Lu B, et al. Genotypic and phenotypic predictors of the magnitude of response to tenofovir disoproxil fumarate treatment in antiretroviral-experienced patients. *The Journal of infectious diseases* 2004; **189**(5): 837-46.

518.   Ghosn J, Delaugerre C, Flandre P, et al. Polymorphism in Gag Gene Cleavage Sites of HIV-1 Non-B Subtype and Virological Outcome of a First-Line Lopinavir/Ritonavir Single Drug Regimen. *PloS one* 2011; **6**(9): e24798.

519.   Wainberg MA, Brenner BG. The Impact of HIV Genetic Polymorphisms and Subtype Differences on the Occurrence of Resistance to Antiretroviral Drugs. *Molecular Biology International* 2012; **2012**: 256982.

520.   Dierynck I, De Wit M, Gustin E, et al. Binding kinetics of darunavir to human immunodeficiency virus type 1 protease explain the potent antiviral activity and high genetic barrier. *J Virol* 2007; **81**(24): 13845-51.

521.   Johnson M, Grinsztejn B, Rodriguez C, et al. Atazanavir plus ritonavir or saquinavir, and lopinavir/ritonavir in patients experiencing multiple virological failures. *Aids* 2005; **19**(7): 685-94.

522.   Pellegrin I, Breilh D, Ragnaud JM, et al. Virological responses to atazanavir-ritonavir-based regimens: resistance-substitutions score and pharmacokinetic parameters (Reyaphar study). *Antivir Ther* 2006; **11**(4): 421-9.

523.   Pellegrin I, Wittkop L, Joubert LM, et al. Virological response to darunavir/ritonavir-based regimens in antiretroviral-experienced patients (PREDIZISTA study). *Antiviral therapy* 2008; **13**(2): 271-9.

524.   Lisovsky I, Schader SM, Martinez-Cajas JL, Oliveira M, Moisi D, Wainberg MA. HIV-1 protease codon 36 polymorphisms and differential development of resistance to nelfinavir, lopinavir, and atazanavir in different HIV-1 subtypes. *Antimicrobial agents and chemotherapy* 2010; **54**(7): 2878-85.

525.   Vergne L, Stuyver L, Van Houtte M, Butel C, Delaporte E, Peeters M. Natural polymorphism in protease and reverse transcriptase genes and in vitro antiretroviral drug susceptibilies of non-B HIV-1 strains from treatment-naive patients. *J Clin Virol* 2006; **36**(1): 43-9.

526.   van Westen GJ, Hendriks A, Wegner JK, Ijzerman AP, van Vlijmen HW, Bender A. Significantly improved HIV inhibitor efficacy prediction employing proteochemometric models generated from antivirogram data. *PLoS computational biology* 2013; **9**(2): e1002899.

527.   Shahriar R, Rhee SY, Liu TF, et al. Nonpolymorphic human immunodeficiency virus type 1 protease and reverse transcriptase treatment-selected mutations. *Antimicrobial agents and chemotherapy* 2009; **53**(11): 4869-78.

528.   Clemente JC, Hemrajani R, Blum LE, Goodenow MM, Dunn BM. Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N. *Biochemistry* 2003; **42**(51): 15029-35.

529.   Holguin A, Paxinos E, Hertogs K, Womac C, Soriano V. Impact of frequent natural polymorphisms at the protease gene on the in vitro susceptibility to protease inhibitors in HIV-1 non-B subtypes. *J Clin Virol* 2004; **31**(3): 215-20.

530.   Calazans A, Brindeiro R, Brindeiro P, et al. Low accumulation of L90M in protease from subtype F HIV-1 with resistance to protease inhibitors is caused by the L89M polymorphism. *The Journal of infectious diseases* 2005; **191**(11): 1961-70.

531.   Goldfarb NE, Ohanessian M, Biswas S, et al. Defective hydrophobic sliding mechanism and active site expansion in HIV-1 protease drug resistant variant Gly48Thr/Leu89Met: mechanisms for the loss of saquinavir binding potency. *Biochemistry* 2015; **54**(2): 422-33.

532.   Wang RG, Zhang HX, Zheng QC. Revealing the binding and drug resistance mechanism of amprenavir, indinavir, ritonavir, and nelfinavir complexed with HIV-1 protease due to double mutations G48T/L89M by molecular dynamics simulations and free energy analyses. *Phys Chem Chem Phys* 2020; **22**(8): 4464-80.

533.	Lambert-Niclot S, Flandre P, Malet I, et al. Impact of gag mutations on selection of darunavir resistance mutations in HIV-1 protease. *The Journal of antimicrobial chemotherapy* 2008; **62**(5): 905-8.

534.	Fun A, van Maarseveen NM, Pokorna J, et al. HIV-1 protease inhibitor mutations affect the development of HIV-1 resistance to the maturation inhibitor bevirimat. *Retrovirology* 2011; **8**: 70.

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

536.	Knops E, Brakier-Gingras L, Schülter E, Pfister H, Kaiser R, Verheyen J. Mutational patterns in the frameshift-regulating site of HIV-1 selected by protease inhibitors. *Medical microbiology and immunology* 2012; **201**(2): 213-8.

537.	Roquebert B, Malet I, Wirden M, et al. Role of HIV-1 minority populations on resistance mutational pattern evolution and susceptibility to protease inhibitors. *Aids* 2006; **20**(2): 287-9.

538.	Mo H, Parkin N, Stewart KD, et al. Identification and structural characterization of I84C and I84A mutations that are associated with high-level resistance to human immunodeficiency virus protease inhibitors and impair viral replication. *Antimicrobial agents and chemotherapy* 2007; **51**(2): 732-5.

539.	Resch W, Parkin N, Watkins T, Harris J, Swanstrom R. Evolution of human immunodeficiency virus type 1 protease genotypes and phenotypes in vivo under selective pressure of the protease inhibitor ritonavir. *Journal of virology* 2005; **79**(16): 10638-49.

540.	Prado JG, Wrin T, Beauchaine J, et al. Amprenavir-resistant HIV-1 exhibits lopinavir cross-resistance and reduced replication capacity. *Aids* 2002; **16**(7): 1009-17.

541.	Salzwedel K, Reddick M, Matalland C, et al. Role of Gag polymorphisms in HIV-1 sensitivity to the maturation inhibitor bevirimat. ANTIVIRAL THERAPY; 2008: INT MEDICAL PRESS LTD 2-4 IDOL LANE, LONDON EC3R 5DD, ENGLAND; 2008. p. A31-A.

542.	Li F, Zoumplis D, Matallana C, et al. Determinants of activity of the HIV-1 maturation inhibitor PA-457. *Virology* 2006; **356**(1-2): 217-24.

543.	Adamson CS, Sakalian M, Salzwedel K, Freed EO. Polymorphisms in Gag spacer peptide 1 confer varying levels of resistance to the HIV-1maturation inhibitor bevirimat. *Retrovirology* 2010; **7**(1): 36.

544.	Sebastian J, Faruki H. Update on HIV resistance and resistance testing. *Med Res Rev* 2004; **24**(1): 115-25.

545.	Koning FA, Castro H, Dunn D, et al. Subtype-specific differences in the development of accessory mutations associated with high-level resistance to

HIV-1 nucleoside reverse transcriptase inhibitors. *The Journal of antimicrobial chemotherapy* 2013; **68**(6): 1220-36.

546.   Datir R, El Bouzidi K, Dakum P, Ndembi N, Gupta RK. Baseline PI susceptibility by HIV-1 Gag-protease phenotyping and subsequent virological suppression with PI-based second-line ART in Nigeria. *Journal of Antimicrobial Chemotherapy* 2019; **74**(5): 1402-7.

547.   Stecher G, Tamura K, Kumar S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution* 2020; **37**(4): 1237-9.

548.   Organization WH. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach: World Health Organization; 2016.

549.   HIV/AIDS JUNPo, HIV/Aids JUNPo. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. *Geneva: Unaids* 2014.

550.   Phillips AN, Staszewski S, Weber R, et al. HIV Viral Load Response to Antiretroviral Therapy According to the Baseline CD4 Cell Count and Viral Load. *JAMA* 2001; **286**(20): 2560-7.

551.   Martins AN, Waheed AA, Ablan SD, et al. Elucidation of the Molecular Mechanism Driving Duplication of the HIV-1 PTAP Late Domain. *J Virol* 2016; **90**(2): 768-79.

552.   Sharma S, Arunachalam PS, Menon M, et al. PTAP motif duplication in the p6 Gag protein confers a replication advantage on HIV-1 subtype C. *J Biol Chem* 2018; **293**(30): 11687-708.

553.   Masquelier B, Droz C, Dary M, et al. R57K polymorphism in the human immunodeficiency virus type 1 protease as predictor of early virological failure in a cohort of antiretroviral-naive patients treated mostly with a nelfinavir-containing regimen. *Antimicrob Agents Chemother* 2003; **47**(11): 3623-6.

554.   Liu Z, Huang X, Hu L, et al. Effects of Hinge-region Natural Polymorphisms on Human Immunodeficiency Virus-Type 1 Protease Structure, Dynamics, and Drug Pressure Evolution. *J Biol Chem* 2016; **291**(43): 22741-56.

555.   Gupta RK, Hill A, Sawyer AW, et al. Virological monitoring and resistance to first-line highly active antiretroviral therapy in adults infected with HIV-1 treated under WHO guidelines: a systematic review and meta-analysis. *The Lancet infectious diseases* 2009; **9**(7): 409-17.

556.   Gupta RK, Jordan MR, Sultan BJ, et al. Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. *The Lancet* 2012; **380**(9849): 1250-8.

557.   Rhee S-Y, Blanco JL, Jordan MR, et al. Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient-and sequence-level meta-analysis. *PLoS medicine* 2015; **12**(4).

558.   WHO. Guidelines on the Public Health Response to Pretreatment HIV Drug Resistance. 2017. https://www.who.int/hiv/pub/guidelines/hivdr-guidelines-2017/en/.

559.   Court R, Gordon M, Cohen K, et al. Random lopinavir concentrations predict resistance on lopinavir-based antiretroviral therapy. *International journal of antimicrobial agents* 2016; **48**(2): 158-62.

560.   Cohen K, Stewart A, Kengne AP, et al. A clinical prediction rule for protease inhibitor resistance in patients failing second-line antiretroviral therapy. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2019; **80**(3): 325-9.

561.   Van Zyl GU, Van Mens TE, McIlleron H, et al. Low lopinavir plasma or hair concentrations explain second line protease inhibitor failures in a resource-limited setting. *Journal of acquired immune deficiency syndromes (1999)* 2011; **56**(4): 333.

562.   El Bouzidi K, Collier D, Nastouli E, Copas AJ, Miller RF, Gupta RK. Virological efficacy of PI monotherapy for HIV-1 in clinical practice. *Journal of Antimicrobial Chemotherapy* 2016; **71**(11): 3228-34.

563.   Thompson JA, Kityo C, Dunn D, et al. Evolution of Protease Inhibitor Resistance in Human Immunodeficiency Virus Type 1 Infected Patients Failing Protease Inhibitor Monotherapy as Second-line Therapy in Low-income Countries: An Observational Analysis Within the EARNEST Randomized Trial. *Clinical infectious diseases* 2018; **68**(7): 1184-92.

564.   Delaugerre C, Flandre P, Chaix ML, et al. Protease inhibitor resistance analysis in the MONARK trial comparing first-line lopinavir-ritonavir monotherapy to lopinavir-ritonavir plus zidovudine and lamivudine triple therapy. *Antimicrobial agents and chemotherapy* 2009; **53**(7): 2934-9.

565.   Van Duyne R, Kuo LS, Pham P, Fujii K, Freed EO. 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* 2019; **116**(18): 9040-9.

566.   Manasa J, Varghese V, Pond SLK, et al. Evolution of gag and gp41 in patients receiving ritonavir-boosted protease inhibitors. *Scientific reports* 2017; **7**(1): 1-11.

567.   Stray KM, Callebaut C, Glass B, et al. Mutations in multiple domains of Gag drive the emergence of in vitro resistance to the phosphonate-containing HIV-1 protease inhibitor GS-8374. *Journal of virology* 2013; **87**(1): 454-63.

568.   Sutherland KA, Goodall RL, McCormick A, et al. Gag-Protease Sequence Evolution Following Protease Inhibitor Monotherapy Treatment Failure in HIV-1 Viruses Circulating in East Africa. *AIDS research and human retroviruses* 2015; **31**(10): 1032-7.

569.   Descours B, Cribier A, Chable-Bessia C, et al. SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4+ T-cells. *Retrovirology* 2012; **9**(1): 87.

570.   Deshmukh L, Louis JM, Ghirlando R, Clore GM. Transient HIV-1 Gag-protease interactions revealed by paramagnetic NMR suggest origins of compensatory drug resistance mutations. *Proceedings of the National Academy of Sciences of the United States of America* 2016; **113**(44): 12456-61.

571.   Murrell B, Moola S, Mabona A, et al. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology and Evolution* 2013; **30**(5): 1196-205.

572.   Lemey P, Salemi M, Vandamme A-M. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing: Cambridge University Press; 2009.

573.   McKinnon JE, Delgado R, Pulido F, Shao W, Arribas JR, Mellors JW. Single genome sequencing of HIV-1 gag and protease resistance mutations at virologic failure during the OK04 trial of simplified versus standard maintenance therapy. *Antiviral therapy* 2011; **16**(5): 725-32.

574.   Perrier M, Castain L, Regad L, et al. HIV-1 protease, Gag and gp41 baseline substitutions associated with virological response to a PI-based regimen. *Journal of Antimicrobial Chemotherapy* 2019; **74**(6): 1679-92.

575.   Castain L, Perrier M, Charpentier C, et al. New mechanisms of resistance in virological failure to protease inhibitors: selection of non-described protease, Gag and Gp41 mutations. *Journal of Antimicrobial Chemotherapy* 2019; **74**(7): 2019-23.

576.   Kempf DJ, King MS, Bernstein B, et al. Incidence of resistance in a double-blind study comparing lopinavir/ritonavir plus stavudine and lamivudine to nelfinavir plus stavudine and lamivudine. *The Journal of infectious diseases* 2004; **189**(1): 51-60.

577.   Lathouwers E, De Meyer S, Dierynck I, et al. Virological characterization of patients failing darunavir/ritonavir or lopinavir/ritonavir treatment in the ARTEMIS study: 96-week analysis. *Antiviral therapy* 2011; **16**(1): 99.

578.   Margot NA, Gibbs CS, Miller MD. Phenotypic susceptibility to bevirimat in isolates from HIV-1-infected patients without prior exposure to bevirimat. *Antimicrobial agents and chemotherapy* 2010; **54**(6): 2345-53.

579. Liu Y, Rao U, McClure J, et al. Impact of mutations in highly conserved amino acids of the HIV-1 Gag-p24 and Env-gp120 proteins on viral replication in different genetic backgrounds. *PloS one* 2014; **9**(4): e94240.

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

581. Kantor R, Katzenstein DA, Efron B, et al. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. *PLoS Med* 2005; **2**(4): e112.

582. Sylla M, Chamberland A, Boileau C, et al. Short communication Characterization of drug resistance in antiretroviral-treated patients infected with HIV-1 CRF02_AG and AGK subtypes in Mali and Burkina Faso. *Antiviral therapy* 2008; **13**: 141-8.

583. Kinomoto M, Appiah-Opong R, Brandful J, et al. HIV-1 proteases from drug-naive West African patients are differentially less susceptible to protease inhibitors. *Clinical infectious diseases* 2005; **41**(2): 243-51.

584. Allen TM, Altfeld M, Geer SC, et al. Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *Journal of virology* 2005; **79**(21): 13239-49.

585. O'Connor DH, Allen TM, Vogel TU, et al. Acute phase cytotoxic T lymphocyte escape is a hallmark of simian immunodeficiency virus infection. *Nature medicine* 2002; **8**(5): 493-9.

586. Jones NA, Wei X, Flower DR, et al. Determinants of human immunodeficiency virus type 1 escape from the primary CD8+ cytotoxic T lymphocyte response. *The Journal of experimental medicine* 2004; **200**(10): 1243-56.

587. Frahm N, Baker B, Brander C. Identification and optimal definition of HIV-derived cytotoxic T lymphocyte (CTL) epitopes for the study of CTL escape, functional avidity and viral evolution. *HIV molecular immunology* 2008; **2008**: 3-24.

588. Yokomaku Y, Miura H, Tomiyama H, et al. Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection. *Journal of virology* 2004; **78**(3): 1324-32.

589. Flynn WF, Chang MW, Tan Z, et al. Deep sequencing of protease inhibitor resistant HIV patient isolates reveals patterns of correlated mutations in Gag and protease. *PLoS computational biology* 2015; **11**(4): e1004249.

590.   Goldstein RA, Tamuri AU, Roy S, Breuer J. Haplotype assignment of virus NGS data using co-variation of variant frequencies. *bioRxiv* 2018: 444877.

591.   Barber TJ, Harrison L, Asboe D, et al. Frequency and patterns of protease gene resistance mutations in HIV-infected patients treated with lopinavir/ritonavir as their first protease inhibitor. *The Journal of antimicrobial chemotherapy* 2012; **67**(4): 995-1000.

592.   Santos JR, Llibre JM, Imaz A, et al. Mutations in the protease gene associated with virological failure to lopinavir/ritonavir-containing regimens. *The Journal of antimicrobial chemotherapy* 2012; **67**(6): 1462-9.

593.   Van Zyl GU, Liu TF, Claassen M, et al. Trends in Genotypic HIV-1 Antiretroviral Resistance between 2006 and 2012 in South African Patients Receiving First- and Second-Line Antiretroviral Treatment Regimens. *PloS one* 2013; **8**(6): e67188.

594.   Dolling DI, Dunn DT, Sutherland KA, et al. Low frequency of genotypic resistance in HIV-1-infected patients failing an atazanavir-containing regimen: a clinical cohort study. *The Journal of antimicrobial chemotherapy* 2013; **68**(10): 2339-43.

595.   Dandache S, Sévigny G, Yelle J, et al. In vitro antiviral activity and cross-resistance profile of PL-100, a novel protease inhibitor of human immunodeficiency virus type 1. *Antimicrobial agents and chemotherapy* 2007; **51**(11): 4036-43.

596.   Henderson GJ, Lee SK, Irlbeck DM, et al. Interplay between single resistance-associated mutations in the HIV-1 protease and viral infectivity, protease activity, and inhibitor sensitivity. *Antimicrobial agents and chemotherapy* 2012; **56**(2): 623-33.

597.   Chang MW, Torbett BE. Accessory mutations maintain stability in drug-resistant HIV-1 protease. *J Mol Biol* 2011; **410**(4): 756-60.

598.   van Lelyveld SF, Wensing AM, Hoepelman AI. The MOTIVATE trials: maraviroc therapy in antiretroviral treatment-experienced HIV-1-infected patients. *Expert Rev Anti Infect Ther* 2012; **10**(11): 1241-7.

599.   Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* 2016; **44**(W1): W344-W50.

600.   Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proceedings of the National Academy of Sciences* 1996; **93**(7): 3099-104.

601. Dolcetti R, Giagulli C, He W, et al. Role of HIV-1 matrix protein p17 variants in lymphoma pathogenesis. *Proceedings of the National Academy of Sciences* 2015; **112**(46): 14331-6.

602. Samsudin F, Gan SK-E, Bond PJ. The Structural Basis for Gag Non-Cleavage Site Mutations in Determining HIV-1 Viral Fitness. *bioRxiv* 2020.

603. Saad JS, Miller J, Tai J, Kim A, Ghanam RH, Summers MF. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proceedings of the National Academy of Sciences* 2006; **103**(30): 11364-9.

604. Saad JS, Kim A, Ghanam RH, et al. Mutations that mimic phosphorylation of the HIV-1 matrix protein do not perturb the myristyl switch. *Protein Science* 2007; **16**(8): 1793-7.

605. Pham QD, Wilson DP, Law MG, Kelleher AD, Zhang L. Global burden of transmitted HIV drug resistance and HIV-exposure categories: a systematic review and meta-analysis. *AIDS* 2014; **28**(18): 2751-62.

606. Günthard HF, Calvez V, Paredes R, et al. Human Immunodeficiency Virus Drug Resistance: 2018 Recommendations of the International Antiviral Society-USA Panel. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2019; **68**(2): 177-87.

607. Cambiano V, Bertagnolio S, Jordan MR, Lundgren JD, Phillips A. Transmission of drug resistant HIV and its potential impact on mortality and treatment outcomes in resource-limited settings. *The Journal of infectious diseases* 2013; **207**(suppl_2): S57-S62.

608. HIV Glasgow 2018, 28-31 October 2018, Glasgow, UK. *Journal of the International AIDS Society* 2018; **21 Suppl 8**(Suppl Suppl 8): e25187-e.

609. Levison JH, Orrell C, Gallien S, et al. Virologic failure of protease inhibitor-based second-line antiretroviral therapy without resistance in a large HIV treatment program in South Africa. *PloS one* 2012; **7**(3): e32144.

610. Brumme ZL, Chan KJ, Dong WW, et al. Prevalence and clinical implications of insertions in the HIV-1 p6Gag N-terminal region in drug-naive individuals initiating antiretroviral therapy. *Antiviral therapy* 2003; **8**(2): 91-6.

611. Banke S, Lillemark MR, Gerstoft J, Obel N, Jørgensen LB. Positive selection pressure introduces secondary mutations at Gag cleavage sites in human immunodeficiency virus type 1 harboring major protease resistance mutations. *J Virol* 2009; **83**(17): 8916-24.

612. Brann TW, Dewar RL, Jiang MK, et al. Functional correlation between a novel amino acid insertion at codon 19 in the protease of human immunodeficiency virus type 1 and polymorphism in the p1/p6 Gag cleavage site in drug resistance and replication fitness. *J Virol* 2006; **80**(12): 6136-45.

613.   Callebaut C, Stray K, Tsai L, et al. In vitro characterization of GS-8374, a novel phosphonate-containing inhibitor of HIV-1 protease with a favorable resistance profile. *Antimicrobial agents and chemotherapy* 2011; **55**(4): 1366-76.

614.   Siliciano JD, Siliciano RF. Recent trends in HIV-1 drug resistance. *Current opinion in virology* 2013; **3**(5): 487-94.

615.   Müller B, Anders M, Akiyama H, et al. HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 infectivity. *Journal of Biological Chemistry* 2009; **284**(43): 29692-703.

616.   Kellam P, Larder BA. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. *Journal of virology* 1995; **69**(2): 669-74.

617.   Brown RJ, Peters PJ, Caron C, et al. Intercompartmental recombination of HIV-1 contributes to env intrahost diversity and modulates viral tropism and sensitivity to entry inhibitors. *Journal of virology* 2011; **85**(12): 6024-37.

618.   Mild M, Esbjörnsson J, Fenyö EM, Medstrand P. Frequent intrapatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch. *Journal of Virology* 2007; **81**(7): 3369-76.

619.   Tebit DM, Zekeng L, Kaptué L, et al. Construction and characterization of an HIV-1 group O infectious molecular clone and analysis of vpr-and nef-negative derivatives. *Virology* 2004; **326**(2): 329-39.

Appendices

Appendix I: Amino acid positions within Gag previously associated with PI resistance or exposure that demonstrate intra-patient variability within patient

Table S-1: Amino acid positions within Gag previously associated with PI resistance or exposure that demonstrate intra-patient variability within patient

	Viral variants from Patient 1								
	Baseline					VF			
Reported Mutation	1	2	3	4	5	1	2	3	4
E12K	E	E	E	E	E	K	K	K	K
R76K	K	R	R	R	K	K	K	K	K
T81A	T	A	A	A	A	T	T	T	T
**E93	E	E	E	E	E	D	D	D	D
**E177	D	D	D	D	D	E	E	E	E
**R464	E	E	G	E	E	G	G	G	G

** These are not sites previously reported to be associated with PI exposure or resistance but were included because amino acid substitutions were only exclusively seen in viral clones from one time point, but not the other.

Table S-2: Amino acid positions within Gag previously associated with PI resistance or exposure that demonstrate intra-patient variability within patient

	Viral variants from patient 2								
	Baseline						VF		
Reported Mutation	1	2	3	4	5	6	1	2	3
G62R	Q	Q	Q	Q	Q	Q	K	K	K
L75R	I	I	I	I	I	I	L	L	L
R76K	K	K	K	K	K	K	Q	Q	Q
Y79F	Y	Y	Y	Y	Y	Y	H	H	H
**E93	G	G	G	G	G	G	K	K	K
**A115	A	A	A	A	A	A	K	K	K
G123E	G	G	G	G	G	G	E	E	E
V128A/I/T/del	V	V	V	V	V	V	I	I	I
H219Q	Q	Q	Q	Q	Q	Q	H	H	H

** These are not sites previously reported to be associated with PI exposure or resistance but were included because amino acid substitutions were only exclusive seen in viral clones from one time point, but not the other.

Table S-3: Amino acid positions within Gag previously associated with PI resistance or exposure that demonstrate intra-patient variability within patient

Reported Mutation	Viral variants from Patient 3																	
	Baseline										VF							
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8
R15	S	S	S	S	S	S	S	S	S	S	M	M	M	M	M	M	M	M
G49	G	G	G	G	G	G	G	G	G	G	S	S	S	S	S	S	S	S
Q69	R	R	R	R	K	R	K	K	R	K	K	K	K	K	K	K	K	K
Q90	S	S	S	S	S	S	S	S	S	S	R	R	R	R	R	R	R	R
H124	S	S	S	S	S	S	S	S	S	S	G	G	G	G	G	G	G	G
Q311	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	L	L	L	L	L	L	L	L
K418	R	R	R	R	R	R	R	R	R	R	K	K	K	K	K	K	K	K
P453	P	P	P	P	P	P	P	P	P	P	S	S	S	S	S	S	S	S
P458	P	P	P	P	P	P	P	P	P	P	S	S	S	S	S	S	S	S

** These are not sites previously reported to be associated with PI exposure or resistance but were included because amino acid substitutions were only exclusive seen in viral clones from one time point, but not the other.

Table S-4: Amino acid positions within Gag previously associated with PI resistance or exposure that demonstrate intra-patient variability within patient

Reported Mutation	Viral variants from Patient 4								
	Baseline				VF				
	1	2	3	4	1	2	3	4	5
E12K	R	R	R	R	K	K	K	K	K
R15	A	A	A	A	S	S	S	S	S
R58	A	A	A	A	V	V	V	V	V
G62R	T	T	T	T	A	A	A	A	A
Q90	Q	Q	Q	Q	H	H	H	H	H
S111	S	S	S	S	H	H	H	H	H
V215	L	L	L	L	A	A	A	A	A
I223	L	L	L	L	V	V	V	V	V
G248	T	T	T	T	A	A	A	A	A
I256	I	I	I	I	V	V	V	V	V
T401	L	L	L	L	I	I	I	I	I
T470	I	I	I	I	M	M	M	M	M

** These are not sites previously reported to be associated with PI exposure or resistance but were included because amino acid substitutions were only exclusive seen in viral clones from one time point, but not the other.

Table S-5: Amino acid positions within Gag previously associated with PI resistance or exposure that demonstrate intra-patient variability within patient

	Viral variants from Patient 5						
	Baseline				VF		
Reported Mutation	1	2	3	4	1	2	3
E12K	N	N	N	N	N	K	K
R20	R	R	R	R	R	Q	Q
G62R	E	E	E	E	E	K	K
L75R	F	F	F	F	F	L	L
Y86	W	W	W	W	W	C	C
Q90	K	K	K	K	K	E	E
K95	Q	Q	Q	Q	Q	R	R
E107	V	V	V	V	I	I	I
S111	S	S	S	S	S	N	N
K113	P	P	P	P	Q	Q	Q
T122	T	T	T	T	T	A	A
G123E	G	G	G	G	E	E	E
N126	S	S	S	S	S	-	-
Q127	H	H	H	H	H	-	-
R387	K	K	K	K	R	R	R
R406	R	R	R	R	R	K	K
Y441	N	C	N	N	S	S	S
V467	G	G	G	G	G	E	E

** These are not sites previously reported to be associated with PI exposure or resistance but were included because amino acid substitutions were only exclusive seen in viral clones from one time point, but not the other.

Table S-6: Amino acid positions within Gag previously associated with PI resistance or exposure that demonstrate intra-patient variability within patient

	Viral variants from Patient 6											
Reported Mutation	Baseline					VF						
	1	2	3	4	5	1	2	3	4	5	6	7
K28	K	K	K	K	K	R	R	R	R	R	R	R
I34	L	L	L	L	L	I	I	I	I	I	I	I
G49	S	S	S	S	S	G	G	G	G	G	G	G
L61	I	I	M	M	M	I	I	I	I	I	I	I
Q63	H	H	H	H	H	Q	Q	Q	Q	Q	Q	Q
P66	T	T	R	R	R	T	T	T	T	T	T	T
K110	K	K	N	N	N	K	K	K	K	K	K	K
A119	A	A	A	A	A	T	T	T	T	T	T	T
N385	G	G	G	G	G	S	S	S	S	S	S	S
P473	G	G	S	S	S	P	P	P	P	P	P	P

** These are not sites previously reported to be associated with PI exposure or resistance but were included because amino acid substitutions were only exclusive seen in viral clones from one time point, but not the other.

Appendix II: Publication arising from Chapters 3 and 4 of this thesis

Journal of Antimicrobial Chemotherapy

Baseline PI susceptibility by HIV-1 Gag-protease phenotyping and subsequent virological suppression with PI-based second-line ART in Nigeria

R. Datir[1], K. El Bouzidi[1], P. Dakum[2], N. Ndembi[2]† and R. K. Gupta[1,3]*†

[1]Division of Infection and Immunity, University College London, London, UK; [2]Institute of Human Virology, Abuja, Nigeria; [3]Africa Health Research Institute, Durban, South Africa

*Corresponding author. Cruciform Building UCL, 90 Gower St, London WC1E 6BT, UK. Tel: +44 7500792984; E-mail: rebmrag@ucl.ac.uk
†These authors contributed equally.

Objectives: Previous work showed that *gag-protease*-derived phenotypic susceptibility to PIs differed between HIV-1 subtype CRF02_AG/subtype G-infected patients who went on to successfully suppress viral replication versus those who experienced virological failure of lopinavir/ritonavir monotherapy as first-line treatment in a clinical trial. We analysed the relationship between PI susceptibility and outcome of second-line ART in Nigeria, where subtypes CRF02_AG/G dominate the epidemic.

Methods: Individuals who experienced second-line failure with ritonavir-boosted PI-based ART were matched (by subtype, sex, age, viral load, duration of treatment and baseline CD4 count) to those who achieved virological response ('successes'). Successes were defined by viral load <400 copies of HIV-1 RNA/mL by week 48. Full-length Gag-protease was amplified from patient samples for *in vitro* phenotypic susceptibility testing, with PI susceptibility expressed as IC_{50} fold change (FC) relative to a subtype B reference strain.

Results: The median (IQR) lopinavir IC_{50} FC was 4.04 (2.49–7.89) for virological failures and 4.13 (3.14–8.17) for virological successes ($P = 0.94$). One patient had an FC >10 for lopinavir at baseline and experienced subsequent virological failure with ritonavir-boosted lopinavir as the PI. There was no statistically significant difference in single-round replication efficiency between the two groups ($P = 0.93$). There was a moderate correlation between single-round replication efficiency and FC for lopinavir (correlation coefficient 0.32).

Conclusions: We found no impact of baseline HIV-1 Gag-protease-derived phenotypic susceptibility on outcomes of PI-based second-line ART in Nigeria.

Introduction

Prevalence of virological failure for first-line antiretroviral therapy can be as high as 30%,[1] with high-level resistance to NNRTI, tenofovir and cytosine analogues common in resource-limited settings and compounded by prior undisclosed ART.[2,3] Second-line ART recommended by WHO comprises a ritonavir-boosted PI and two NRTIs, commonly lopinavir or atazanavir.[4] PIs are the second- and last-line therapy for the majority of HIV-infected patients worldwide as access to third-line therapy is still limited.[5] Virological failure with PIs as second-line therapy occurs in around 20% of individuals.[6–8] In contrast to first-line therapy, with which >80% develop drug resistance mutations, only around 10%–20% develop major resistance mutations to PIs by week 48,[6,7,9,10] and this proportion increases over time.[5]

It is known that proteins such as Gag and Env can affect susceptibility to PIs even in the absence of known major resistance mutations in the *protease* gene.[11–15] There are limited data on changes in *gag* following treatment failure with PIs in the non-B subtypes that dominate low- and middle-income countries.[15–20] It appears that in around 15% of patients failing boosted PI (bPI) without major protease mutations, a decrease in phenotypic susceptibility to the drug appears to occur when *gag-protease* is phenotyped.[21–23] Therefore it is conceivable that underlying phenotypic susceptibility resulting from variation in genes such as *gag* and *env* might impact clinical responses to PI.

We previously showed that *gag-protease*-derived phenotypic susceptibility differed between CRF02_AG and subtype G-infected patients who went on to successfully suppress viral replication

1402

JAC

versus those who experienced virological failure (VF) of lopinavir/ritonavir monotherapy as first-line treatment in a clinical trial.[12] In order to determine the relevance of this finding for real-world settings in the context of tenofovir disoproxil fumarate + lamivudine or zidovudine + lamivudine with ritonavir-boosted PI (lopinavir or atazanavir) we analysed the relationship between PI susceptibility and the outcome of PI-based second-line ART in Nigeria, where subtypes CRF02_AG and G dominate the epidemic.[24]

Patients and methods

Study participants

This study involved retrospectively testing samples from patients attending for HIV care at University of Abuja Teaching Hospital (UATH) who experienced second-line failure (HIV-1 RNA >1000 copies/mL after >6 months on treatment) on a lopinavir/ritonavir- or atazanavir/ritonavir-containing regimen, without any major PI mutations, who were selected as 'cases'. They were matched to 'controls', who had achieved virological suppression lasting up to 12 months (HIV-1 RNA <400 copies/mL) with a similar age, sex, baseline CD4 count and duration of treatment. Baseline (pre-PI) plasma samples from these matched pairs were retrospectively retrieved.

Amplification of full-length gag-protease genes

HIV-1 RNA was manually extracted from archived plasma samples using the QIAamp viral RNA extraction kit. Using previously described techniques,[11,25] full-length *gag-protease* was amplified and cloned into a subtype B-based (p8.9NSX+) vector. Clonal sequencing of up to 10 plasmids (where possible) was performed by standard Sanger sequencing. The variant that most closely represented the consensus (obtained via next-generation sequencing as previously described[6]) was taken forward for phenotypic testing. Sequences were manually analysed using DNA dynamo software (http://www.bluetractorsoftware.co.uk) and MEGA v7.0 software.[26] Protease sequences were analysed for PI resistance mutations using the Stanford Resistance Database (https://hivdb.stanford.edu).

PI susceptibility and infectivity assays

PI susceptibility and viral infectivity were determined using a previously described single assay. Briefly, 293T cells were co-transfected with a Gag-Pol protein expression vector (p8.9NSX+) containing cloned patient-derived *gag-protease* sequences, pMDG expressing vesicular stomatitis virus envelope glycoprotein (VSV-g), and pCSFLW (expressing the firefly luciferase reporter gene with HIV-1 packaging signal).

PI drug susceptibility testing was carried out as previously described.[25] Transfected cells were seeded with serial dilutions of lopinavir and harvested pseudovirions were used to infect fresh 293T cells. To determine strain infectivity, transfected cells were seeded in the absence of drug. Infectivity was monitored by measuring luciferase activity 48 h after infection. Results derived from at least two independent experiments (each in duplicate) were analysed. The IC_{50} was calculated using GraphPad Prism 5 (GraphPad Software Inc., La Jolla, CA, USA). Susceptibility was expressed as a fold change in IC_{50} compared with the subtype B reference strain (p8.9NSX+). Replicative capacity of these viruses was assessed by comparing the luciferase activity of recombinant virus with that of the WT subtype B control virus in the absence of drug. Equal amounts of input plasmid DNA were used, and it has previously been shown that percentage infectivity correlates well with infectivity/ng p24 in this system.[25] The PI drugs used in this study were obtained from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH.

Ethics

Informed consent was obtained from all subjects and ethics approval for virological testing was obtained from the National Research Ethics Committee of Nigeria (NHREC/01/01/2007).

Statistical analysis

Differences in PI susceptibility were compared with the Wilcoxon rank-sum test (GraphPad Software Inc., La Jolla, CA, USA), which is robust to data that are not normally distributed.

Results

Six matched pairs of patients were included. Table 1 contains clinical and laboratory data on cases, who experienced virological failure (duration), and controls, who suppressed viral replication for 48 weeks. Of note, all pairs but one had a CD4 count <200 cells/mm^3. All but one pair was treated with lopinavir-based ART (atazanavir was used in one pair). Table 2 shows NRTI and NNRTI resistance mutations detected prior to second-line initiation. All patients had lamivudine resistance [M184V/I in reverse transcriptase (RT)] and 7/12 (58.3%) had at least moderate resistance to tenofovir (3 with K65R, 3 with K70E and 1 with three thymidine analogue mutations including M41L, L210W and T215Y). All 12 individuals had high-level NNRTI resistance. Two pairs were infected with subtype G viruses and four pairs with CRF02_AG viruses (Table 2). No major mutations in protease were observed in the patients. We analysed sequences for mutations in Gag in cases and controls associated with PI susceptibility or exposure (Table 3).

The median (IQR) lopinavir fold change (FC) was 4.04 (2.49–7.89) for virological failures and 4.13 (3.14–8.17) for virological successes ($P = 0.94$), as described in Figure 1(a). The median (IQR) atazanavir FC was 2.43 (1.35–9.66) for virological failures and 4.39 (1.60–7.73) for successes ($P = 0.47$). The median (IQR) darunavir FC was 1.234 (0.84–2.05) for virological failures and 1.529 (1.14–2.319) for successes ($P = 0.47$).

One patient had an FC >10 for lopinavir at baseline and experienced subsequent virological failure on boosted lopinavir as the PI. We also measured the single-round replication efficiency of patient-derived *gag-protease*-containing pseudoviruses derived prior to initiation of second-line boosted PI treatment from patients who either did (success) or did not (failure) suppress viral replication after 48 weeks (Figure 1b). Mean replication efficiency relative to a subtype B reference strain was 117.7% for the successes and 105.8% for failures. There was no statistically significant difference in replication efficiency between the two groups ($P = 0.93$ by Mann–Whitney *U*-test).

Finally, we analysed the relationship between single-round replication efficiency and FC to lopinavir in all viruses tested. There was a moderate correlation between these parameters (correlation coefficient 0.32, Figure 2). When a single outlier was excluded from analysis (FC 10.7 with replication efficiency 50.0%), the correlation coefficient increased to 0.78.

Discussion

Given the contribution of the highly polymorphic Gag protein and resulting epistatic interactions to PI susceptibility, we hypothesized that patients would respond differently to these drugs, particularly

1403

Table 1. Clinical data for matched patient pairs comprising virological successes and failures

Sample pair	Age (years)		Sex		Baseline[a] CD4 count (cells/mm³)		Second-line PI used		Baseline[a] viral load (copies of HIV-1 RNA/mL)	
	success	failure	success	failure	success	failure	success	failure	success	failure
1	33	45	female	female	<200	<200	LPV	LPV	503 951	140 991
2	43	36	female	female	<200	<200	LPV	LPV	39 844	20 178
3	27	26	female	female	<200	<200	LPV	LPV	32 284	271 974
4	47	39	female	female	200–499	200–499	LPV	LPV	228 083	24 693
5	35	40	female	female	<200	<200	ATV	ATV	14 487	274 504
6	34	33	female	female	<200	<200	LPV	LPV	39 929	18 056

LPV, lopinavir; ATV, atazanavir.
[a]Baseline refers to pre-initiation of second-line therapy.

Table 2. NRTI and NNRTI mutations observed at first-line failure, prior to initiation of second-line PI-based ART

		NRTI mutations	NNRTI mutations	Baseline[a] VL (copies of HIV-1 RNA/mL)		HIV-1 subtype		2L backbone	
				success	failure	success	failure	success	failure
Pair 1	success	M41L, L74LI, M184V, L210W, T215F	K101E, E138Q, G190A	503 951	140 991	CRF02_AG	CRF02_AG	TDF/FTC	TDF/FTC
	failure	M184V	K103N						
Pair 2	success	E44D, D67N, T69D, K70R, M184V, T215Y	K101E, K103N	39 844	20 178	CRF02_AG	CRF02_AG	TDF/FTC	TDF/FTC
	failure	M184V	K101E, G190A						
Pair 3	success	K70E, M184V	A98G, Y181C	32 284	271 974	G	G	TDF/FTC	TDF/FTC
	failure	K70E, M184V	Y181C, G190A, H221Y						
Pair 4	success	K70E, Y115F, M184V	K103N	228 083	24 693	CRF02_AG	CRF02_AG	AZT/3TC	AZT/3TC
	failure	K65R, M184V	K101E, V108I, Y181C, G190A						
Pair 5	success	D67N, K70R, M184V, T215F, K219E	Y188C	14 487	274 504	G	G	TDF/FTC	TDF/FTC
	failure	K70R, M184V, K219Q	K103N, Y318F						
Pair 6	success	K65R, M184I	K103N, Y181C	39 929	18 056	CRF02_AG	CRF02_AG	TDF/FTC	TDF/FTC
	failure	K65R, M184I	K103N, Y181C						

AZT, zidovudine; 3TC, lamivudine; FTC, emtricitabine; TDF, tenofovir disoproxil fumarate.
[a]Baseline refers to pre-initiation of second-line therapy.

in the context of extensive NRTI resistance. We previously reported an association between susceptibility to PI and outcome of first-line ritonavir-boosted lopinavir monotherapy in a clinical trial. Here we performed a similar study in patients about to start second-line combination ART, including ritonavir-boosted lopinavir or atazanavir as well as two NRTIs. We found the difference in phenotypic drug susceptibility (assessed by FC relative to a subtype B reference) was not statistically different between the virological failures (cases) and virological successes (controls) for any of the PIs tested: lopinavir, atazanavir or darunavir.

This negative result could be due to the influence of adherence, in that second-line therapy is used in patients for whom first-line therapy has failed, usually as the result of incomplete adherence.

Therefore, the patient group was enriched for poor adherers, which could have overcome the effects of small differences in susceptibility.

Interestingly, we previously showed that 2/2 patients with FC >10 prior to PI monotherapy went on to virological failure.[27] In this study the only patient with FC >10 for lopinavir failed treatment with this drug. Further work needs to be undertaken to explore whether a threshold FC of 10 in our assay is relevant in larger datasets.

We also showed here that replication efficiency over a single round was correlated with lopinavir susceptibility prior to initiation of the bPI. We have previously reported similar findings in replication-competent subtype C viruses that contained patient-derived *gag* and partial *protease* genes.[12] These data suggest that
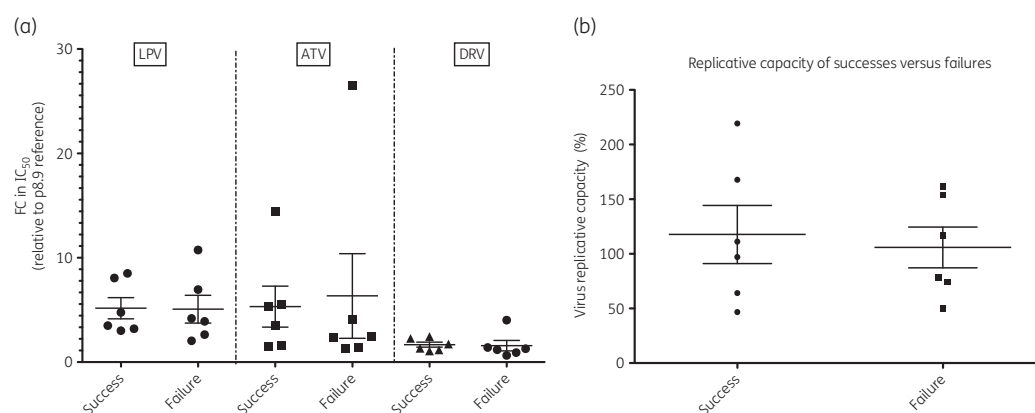
Figure 1. (a) PI susceptibility relative to a subtype B reference strain, expressed as FC in IC_{50}, and (b) single-round replication efficiency (relative to a subtype B reference strain) of patient-derived *gag-protease*-containing pseudoviruses derived from patients prior to initiation of second-line boosted PI treatment who either did (Success) or did not (Failure) suppress viral replication after 48 weeks. Each data point is the mean of at least two independent experiments and hairs represent mean and SD. LPV, lopinavir; ATV, atazanavir; DRV, darunavir.
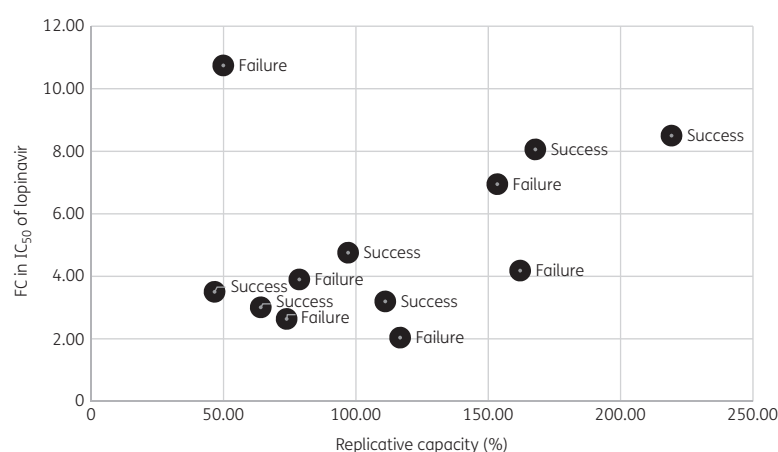
Figure 2. Scatter plot of FC in IC_{50} of lopinavir relative to a subtype B reference strain, versus replicative capacity as measured over a single round of replication (also relative to a subtype B reference strain, which is represented by 100%).

increased replicative capacity and resistance to PI might involve an overlapping mechanism.

Limitations

Limitations of our study include the relatively small sample size, the inclusion of more than one subtype and the possibility of viral recombination through our PCR and cloning strategy. In addition, the process of mapping next-generation sequencing reads to a consensus reference sequence to generate a patient consensus can introduce biases against variation, which may affect the identification of novel drug resistance mutations. Finally, our assay system did not incorporate the native gp160 envelope.

Despite introduction of second-generation integrase inhibitors such as dolutegravir as first-line therapy in areas where pre-treatment resistance is >10%,[28,29] bPI will still be used as second-line therapy for those who fail dolutegravir-based first-line

337

Table 3. Variation in Gag cleavage and non-cleavage sites and protease

| Pair | HIV-1 subtype | Treatment outcome | Gag cleavage sites | | | | | Gag non-cleavage sites |
			MA/CA 128–137 VSQNY/ PIVQN	CA/p2 359–368 KARVL/ AEAMS	p2/NC 373–382 SATIM/ MQRGN	NC/p1 428–437 ERQAN/ FLGKI	p1/p6 444–453 RPGNF/ LQSRP	
Pair 1	CRF02_AG	success	—–/—/——	——/—/——	-+N-/—/——	——/—/——	——–/—/P——	E12K, R76K, R79F, H219Q, T242N, R409K, T487S
		failure	—–/—/——	——/—/——	T+N-/—/——	——/—/——	——–/—/P——	R76K, R79F, T81A, G248A, V370A, R409K, E468K, T487S
Pair 2	G	success	—–/—/——	——/—/——	A-A-/—/——	——/—/——	——–/—/-/N-	E12K, R76K, R79F, G123E, H219Q, G248A, V370A, R409K, T487S
Pair 3	G	failure	——–/—/——	——/—/——	—+/-S-	——/—/——	——–/-/N-L	E12K, G62R, R76K, H219Q, V370A, R409K, T487S
	G	success	——–/—/——	——/—/——	-A-/-KS-	——/—/——	——–/-/N-L	E12K, G62R, R76K, H219Q, T242N, G248A, V370A, R409K, T487S
		failure	——–/—/——	——/—/——	-+NV-/-K-	——/—/——	——–/—/P——	E12K, R76K, R79F, H219Q, T242N, V370A, T371del, R409K, T456S, T487S
Pair 4	CRF02_AG	success	+—/—/——	——/—/——	T+NV-/-K-	——/—/——	——–/—/P——	R76K, R409K, T487S
		failure	+—/—/——	——/—/——	-+NV-/-K-	——/—/——	——–/—/P——	E12K, R76K, R79F, G248A, V370A, R409K, T487S
Pair 5	G	success	—–/—/——	——/—/——	A-A-/—/——	——/-/R-	——–/—/P——	E12K, R76K, R79F, G123E, V370A, R409K, T487S
		failure	+-F/—/——	——/—/——	QPN-/1——	——/-/V	——–/—/P——	E12K, R76K, R79F, S165N, H219Q, V370A, R409K, T487S
Pair 6	CRF02_AG	success	S—/—/——	—/1-/——	ANI-/—/——	——/—/——	——–/—/P——	E12K, R76K, R79F, V370A, R409K, T487S
		failure	+—/—/——	——/—/——	T+NV-/-K-	——/—/——	——–/—/P——	E12K, G248A, V370A, R409K, S451T, T487S

The consensus clonal sequence at each of the Gag cleavage sites is shown for the six pairs using HXB2 numbering. Deletions are represented by +.

regimens. Therefore, research into determinants of responses to PI in non-B subtypes is as important as ever.

Transparency declarations

None to declare.

References

1 Boender TS, Sigaloff KC, McMahon JH *et al.* Long-term virological outcomes of first-line antiretroviral therapy for HIV-1 in low- and middle-income countries: a systematic review and meta-analysis. *Clin Infect Dis* 2015; **61**: 1453–61.

2 TenoRes Study Group. Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study. *Lancet Infect Dis* 2016; **16**: 565–75.

3 Gregson J, Kaleebu P, Marconi VC *et al.* Occult HIV-1 drug resistance to thymidine analogues following failure of first-line tenofovir combined with a cytosine analogue and nevirapine or efavirenz in sub Saharan Africa: a retrospective multi-centre cohort study. *Lancet Infect Dis* 2016; **16**: 565–75.

4 WHO. *Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a Public Health Approach—Second Edition.* http://www.who.int/hiv/pub/arv/arv-2016/en.

5 Rawizza HE, Chaplin B, Meloni ST *et al.* Accumulation of protease mutations among patients failing second-line antiretroviral therapy and response to salvage therapy in Nigeria. *PLoS One* 2013; **8**: e73582.

6 Collier D, Iwuji C, Derache A *et al.* Virological outcomes of second-line protease inhibitor-based treatment for human immunodeficiency virus type 1 in a high-prevalence rural South African setting: a competing-risks prospective cohort analysis. *Clin Infect Dis* 2017; **64**: 1006–16.

7 Hosseinipour MC, Gupta RK, Van Zyl G *et al.* Emergence of HIV drug resistance during first- and second-line antiretroviral therapy in resource-limited settings. *J Infect Dis* 2013; **207** Suppl 2: S49–56.

8 Ajose O, Mookerjee S, Mills EJ *et al.* Treatment outcomes of patients on second-line antiretroviral therapy in resource-limited settings: a systematic review and meta-analysis. *AIDS* 2012; **26**: 929–38.

9 Stockdale AJ, Saunders MJ, Boyd MA *et al.* Effectiveness of protease inhibitor/nucleos(t)ide reverse transcriptase inhibitor-based second-line antiretroviral therapy for the treatment of human immunodeficiency virus type 1 infection in sub-Saharan Africa: a systematic review and meta-analysis. *Clin Infect Dis* 2018; **66**: 1846–57.

10 Doyon L, Croteau G, Thibeault D *et al.* Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol* 1996; **70**: 3763–9.

11 Gupta RK, Kohli A, McCormick AL *et al.* Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays. *AIDS* 2010; **24**: 1651–5.

12 Sutherland KA, Collier DA, Claiborne DT *et al.* Wide variation in susceptibility of transmitted/founder HIV-1 subtype C isolates to protease inhibitors and association with in vitro replication efficiency. *Sci Rep* 2016; **6**: 38153.

13 Rabi SA, Laird GM, Durand CM *et al*. Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. *J Clin Invest* 2013; **123**: 3848–60.

14 Nijhuis M, van Maarseveen NM, Lastere S *et al*. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med* 2007; **4**: e36.

15 Fun A, Wensing AM, Verheyen J *et al*. Human immunodeficiency virus Gag and protease: partners in resistance. *Retrovirology* 2012; **9**: 63.

16 Sutherland KA, Goodall RL, McCormick A *et al*. Gag-protease sequence evolution following protease inhibitor monotherapy treatment failure in HIV-1 viruses circulating in East Africa. *AIDS Res Hum Retroviruses* 2015; **31**: 1032–7.

17 Giandhari J, Basson AE, Coovadia A *et al*. Genetic changes in HIV-1 Gag-protease associated with protease inhibitor-based therapy failure in paediatric patients. *AIDS Res Hum Retroviruses* 2015; **31**: 776–82.

18 Li G, Verheyen J, Theys K *et al*. HIV-1 Gag C-terminal amino acid substitutions emerging under selective pressure of protease inhibitors in patient populations infected with different HIV-1 subtypes. *Retrovirology* 2014; **11**: 79.

19 Jinnopat P, Isarangkura-na-ayuthaya P, Utachee P *et al*. Impact of amino acid variations in Gag and protease of HIV type 1 CRF01_AE strains on drug susceptibility of virus to protease inhibitors. *J Acquir Immune Defic Syndr* 2009; **52**: 320–8.

20 Coetzer M, Ledingham L, Diero L *et al*. Gp41 and Gag amino acids linked to HIV-1 protease inhibitor-based second-line failure in HIV-1 subtype A from Western Kenya. *J Intern AIDS Soc* 2017; **20**: e25024.

21 Sutherland KA, Mbisa JL, Cane PA *et al*. Contribution of Gag and protease to variation in susceptibility to protease inhibitors between different strains of subtype B HIV-1. *J Gen Virol* 2014; **95**: 190–200.

22 Sutherland KA, Mbisa JL, Ghosn J *et al*. Phenotypic characterization of virological failure following lopinavir/ritonavir monotherapy using full-length gag-protease genes. *J Antimicrob Chemother* 2014; **69**: 3340–8.

23 Giandhari J, Basson AE, Sutherland K *et al*. Contribution of Gag and protease to HIV-1 phenotypic drug resistance in pediatric patients failing protease inhibitor-based therapy. *Antimicrob Agents Chemother* 2016; **60**: 2248–56.

24 Chaplin B, Akanmu AS, Inzaule SC *et al*. Association between HIV-1 subtype and drug resistance in Nigerian infants. *J Antimicrob Chemother* 2019; **74**: 172–6.

25 Parry CM, Kohli A, Boinett CJ *et al*. Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1. *J Virol* 2009; **83**: 9094–101.

26 Tamura K, Stecher G, Peterson D *et al*. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013; **30**: 2725–9.

27 Sutherland KA, Ghosn J, Gregson J *et al*. HIV-1 subtype influences susceptibility and response to monotherapy with the protease inhibitor lopinavir/ritonavir. *J Antimicrob Chemother* 2015; **70**: 243–8.

28 WHO. *Guidelines on the Public Health Response to Pretreatment HIV Drug Resistance*. http://who.int/hiv/pub/guidelines/hivdr-guidelines-2017/.

29 Gupta RK, Gregson J, Parkin N *et al*. HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. *Lancet Infect Dis* 2018; **18**: 346–55.

1407

Appendix III: Publication arising from Chapters 5 and 6 of this thesis

In Vivo Emergence of a Novel Protease Inhibitor Resistance Signature in HIV-1 Matrix

Rawlings Datir,[a] Steven Kemp,[a] Kate El Bouzidi,[a] Petra Mlchocova,[b] Richard Goldstein,[a] Judy Breuer,[a] Greg J. Towers,[a] Clare Jolly,[a] Miguel E. Quiñones-Mateu,[c] Patrick S. Dakum,[d,e] Nicaise Ndembi,[d,e] Ravindra K. Gupta[b,f]

[a]University College London, London, United Kingdom
[b]Department of Medicine, University of Cambridge, Cambridge, United Kingdom
[c]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand
[d]Institute for Human Virology, Abuja, Nigeria
[e]Institute of Human Virology, University of Maryland School of Medicine, Baltimore, Maryland, USA
[f]Africa Health Research Institute, Durban, South Africa

Nicaise Ndembi and Ravindra K. Gupta contributed equally.

ABSTRACT Protease inhibitors (PIs) are the second- and last-line therapy for the majority of HIV-infected patients worldwide. Only around 20% of individuals who fail PI regimens develop major resistance mutations in protease. We sought to explore the role of mutations in *gag-pro* genotypic and phenotypic changes in viruses from six Nigerian patients who failed PI-based regimens without known drug resistance-associated protease mutations in order to identify novel determinants of PI resistance. Target enrichment and next-generation sequencing (NGS) with the Illumina MiSeq system were followed by haplotype reconstruction. Full-length Gag-protease gene regions were amplified from baseline (pre-PI) and virologic failure (VF) samples, sequenced, and used to construct *gag-pro*-pseudotyped viruses. Phylogenetic analysis was performed using maximum-likelihood methods. Susceptibility to lopinavir (LPV) and darunavir (DRV) was measured using a single-cycle replication assay. Western blotting was used to analyze Gag cleavage. In one of six participants (subtype CRF02_AG), we found 4-fold-lower LPV susceptibility in viral clones during failure of second-line treatment. A combination of four mutations (S126del, H127del, T122A, and G123E) in the p17 matrix of baseline virus generated a similar 4-fold decrease in susceptibility to LPV but not darunavir. These four amino acid changes were also able to confer LPV resistance to a subtype B Gag-protease backbone. Western blotting demonstrated significant Gag cleavage differences between sensitive and resistant isolates in the presence of drug. Resistant viruses had around 2-fold-lower infectivity than sensitive clones in the absence of drug. NGS combined with haplotype reconstruction revealed that resistant, less fit clones emerged from a minority population at baseline and thereafter persisted alongside sensitive fitter viruses. We used a multipronged genotypic and phenotypic approach to document emergence and temporal dynamics of a novel protease inhibitor resistance signature in HIV-1 matrix, revealing the interplay between Gag-associated resistance and fitness.

KEYWORDS HIV, resistance, protease, drug, Africa, antiretroviral, Gag, antiretroviral resistance, human immunodeficiency virus, protease inhibitors, proteases

As global scale-up of antiretroviral therapy (ART) progresses in the absence of universal viral load monitoring, significant numbers of persons living with HIV (PLWH) are experiencing virological failure (VF) with emergent drug resistance (1–3). In addition, pretreatment drug resistance (PDR) has been rising over the past decade (4–6). Although integrase inhibitors are now recommended by WHO in regions where

PDR exceeds 10% (7, 8), second-line ART in low- and middle-income countries (LMIC) is likely to remain dependent on boosted protease inhibitors (PI), specifically lopinavir/ritonavir or atazanavir/ritonavir.

Studies demonstrate that the detection of major canonical protease mutations (9) is around 20% in PLWH treated with PI-containing combination ART (10, 11), raising the question of how virologic failure occurs in the remaining cases. Inadequate adherence to medication has been implicated (12–14), and the contribution of minor protease mutations has been explored (15). Determinants of susceptibility outside the protease gene have also been considered (16). Interestingly, although PI monotherapy can be effective in some populations in clinical practice (17), this is associated with a higher prevalence of major PI resistance mutations at VF than PI combined with 2 two nucleoside reverse transcriptase inhibitors (NRTI) (18, 19).

The HIV-1 envelope (Env) has been reported in two studies to impact PI susceptibility (20, 21), with a number of reports of diverse *env* sequence changes during PI failure (22, 23). Gag is highly polymorphic across HIV-1 subtypes, and existing literature reports diverse mutations occurring both within and outside cleavage sites following treatment with older PIs, such as indinavir, saquinavir, and nelfinavir, in subtype B infections (16, 22–26). Although there is very limited information on the role of HIV-1 *gag* in susceptibility to modern boosted protease inhibitors, such as lopinavir/ritonavir, used in second-line ART for non-B subtypes, we and others have reported that around 1 in 6 individuals infected with non-subtype B HIV who fail modern PI have *gag*-encoded reduced phenotypic susceptibility to PI (27–31), though specific amino acid determinants have remained elusive.

Cleavage site mutations are thought to partially restore efficient cleavage by protease in the presence of bound drug (32, 33). The mechanism for non-cleavage site mutations may include allosteric changes in protease-Gag interactions that influence the efficiency by which protease locates cleavage sites through dynamic intermolecular interactions in the presence of drug (34, 35). For example, our group previously reported the emergence of T81A in Gag that appeared to correlate with reduced susceptibility to the modern PI lopinavir in a subtype AG-infected individual in France (28). This mutation was predicted to impact intermolecular interactions between Gag and protease by Deshmukh and colleagues using nuclear magnetic resonance (NMR) (35).

Here, we sought to explore the role of mutations in *gag*-encoded determinants of reduced PI susceptibility in non-subtype B HIV-1 and to elucidate their evolution in PLWH in Nigeria.

RESULTS

Phenotypic drug susceptibility following PI failure. Participant characteristics of the six HIV-infected individuals failing PI-based second-line ART are shown in Table 1. Three were infected with CRF02_AG recombinant strain and three with subtype G HIV strains. Next-generation sequencing (NGS) of *gag* and *pol* was used to generate consensus sequences for the six patients at two time points: before PI treatment (baseline) and at virologic failure (VF). A significant number of amino acid changes occurred between time points in each individual, with most occurring in the matrix (p17) domain of Gag. Phenotypic PI susceptibility testing was performed on plasma-derived clones obtained at the same time points.

Participant 6 had a significant difference in PI susceptibility between baseline and failure time points (Fig. 1 and Fig. S1). At VF, the difference in the 50% inhibitory concentration (IC$_{50}$) for lopinavir (LPV), expressed as fold change (FC) compared to the subtype B reference, was 20.3 compared to 5.2 prior to initiation of LPV treatment. We phenotyped four clones from baseline, all with similar LPV susceptibility. Baseline genotype (pre-PI) indicated that the individual had developed extensive resistance to first-line ART, with the nucleoside reverse transcriptase inhibitor (NRTI) mutations K65R and M184I conferring high-level tenofovir and lamivudine resistance, respectively, as well as K103N and Y181C conferring resistance to nonnucleoside reverse transcriptase

TABLE 1 Participant and virus characteristics

Patient (subtype)	Time point	Viral load (copies/ml)	Time between baseline and VF sample (mo)
Patient 1 (CRF02_AG)	Baseline	140,991	34
	VF	6,193	
Patient 2 (G)	Baseline	20,178	42
	VF	117,942	
Patient 3 (CRF02_AG)	Baseline	271,974	50
	VF	74,224	
Patient 4 (CRF02_AG)	Baseline	24,693	36
	VF	32,683	
Patient 5 (G)	Baseline	274,504	31
	VF	16,304	
Patient 6 (CRF02_AG)	Baseline	18,056	64
	VF	66,277	

inhibitors (NNRTI). The co-occurrence of the latter two NNRTI mutations suggests that the individual may have been pretreated with first-line ART containing nevirapine or have received single-dose nevirapine for prevention of mother-to-child transmission (36).

We further explored virus from this participant in order to elucidate determinants of resistance. Sequence alignment of full-length *gag* and protease genes from sensitive and resistant clones revealed 19 amino acid changes in matrix (MA), one change each in capsid (CA), p2, and nucleocapsid (NC), and an insertion of four amino acids (E, L, R, and E) at *gag* position 477 in the p6 region of the resistant clone (Fig. 1). In protease, there was an M46V mutation in the resistant virus that was found to have no impact on LPV susceptibility (Fig. S2).

Interestingly, the VF sample was taken 64 months after PI initiation, when the viral load was 66,277 copies/ml, and within this plasma sample two distinct virus clones were isolated (Fig. 1B, hatched and black bars). There was a 4- to 5-fold difference in LPV susceptibility between the two clones, suggesting a mixture of susceptible and "resistant" viruses at the failure time point (Fig. 1). We proceeded to map determinants of susceptibility using these two clones identified at failure. First, we sought to determine the role of a four-amino-acid insertion in the p6 domain. Using standard site-directed mutagenesis techniques, amino acids E, L, R, and E were inserted into a susceptible clone at position 477 in the p6 domain (Fig. S3). Conversely, E, L, R, and E residues were deleted in the less susceptible clone from the same location. There was no significant change in susceptibility to LPV as a result of the ELRE insertion (Fig. S3).

Matrix deletion of S126 and H127 confers reductions in LPV susceptibility. Given that the greatest number of changes occurred in the MA region, we sought to explore a possible role for MA amino acid changes in PI susceptibility. First, sequence changes occurring near the MA/CA cleavage site (within 10 amino acids) were considered. We noted that the more resistant virus had a deletion of Gag positions 126 and 127 as well as adjacent T122A and G123E mutations. Using site-directed mutagenesis, serine (Gag position 126) and histidine (Gag position 127) residues were deleted in the susceptible clone. Conversely, serine and histidine residues were inserted in the less susceptible clone. Deletion of Ser-126 and His-127 in the susceptible virus led to a significant decrease in LPV susceptibility for the mutant virus (Fig. 2). Conversely, the insertion of Ser and His residues in the resistant virus increased susceptibility of the mutant (Fig. 2). However, the changes at positions 126 and 127 did not completely account for the differences in LPV susceptibility.

The matrix deletions of S126 and H127 act synergistically with T122A and G123E in Gag. A combination of S126del, H127del, and the T122A and G123E muta-

(A)

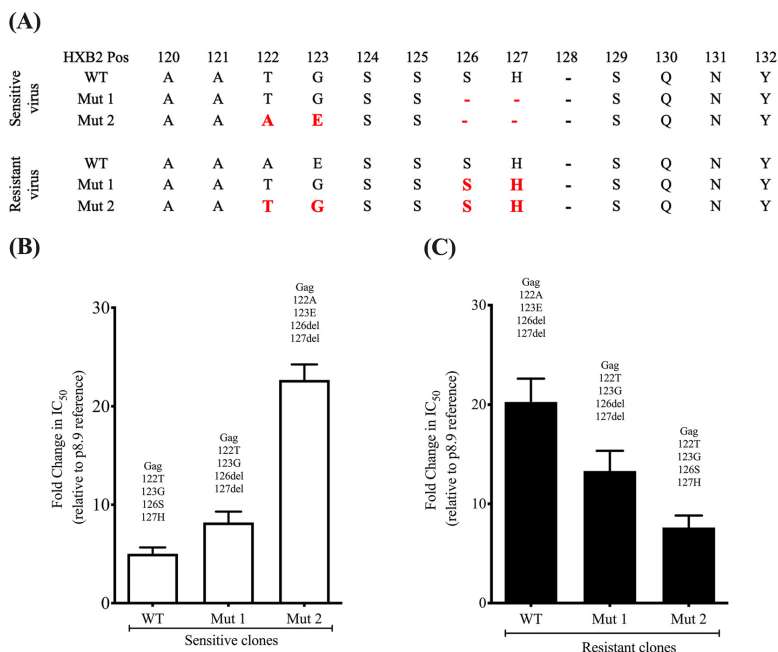|	9	12	20	34	62	75	82	86	90	93	95	107	109	111	113	122	123	126	127	280	376	406	463	466	467	469	477	478	479	480	483
	MA																			CA	p2	NC	p6								
HXB2 position	9	12	20	34	62	75	82	86	90	93	95	107	109	111	113	122	123	126	127	280	376	406	463	466	467	469	477	478	479	480	483
Baseline Clone	S	N	R	L	E	F	I	W	K	K	Q	V	K	S	P	T	G	S	H	V	I	R	L	E	G	T	-	-	-	-	G
VF Clone 1	S	N	R	L	E	F	I	W	K	K	Q	V	K	S	P	T	G	S	H	I	V	R	L	E	G	T	-	-	-	-	G
VF Clone 2	T	E	Q	I	K	L	L	C	E	E	R	I	N	N	Q	A	E	-	-	V	I	K	F	G	E	I	E	L	R	E	R

(B)
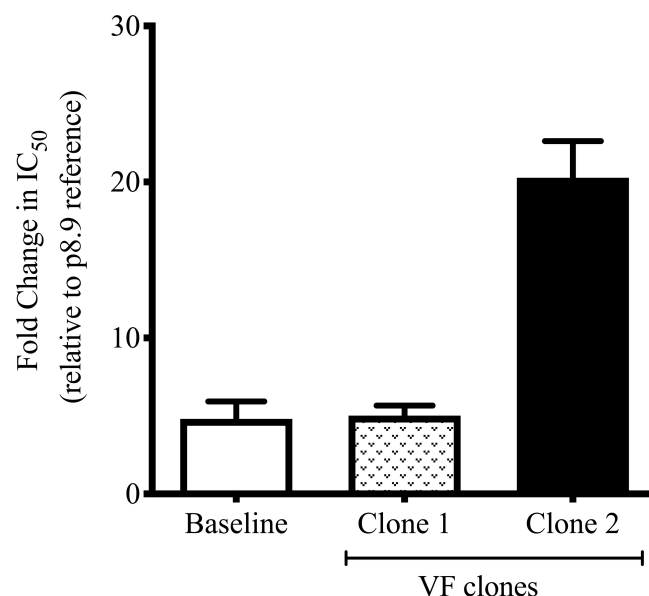
FIG 1 Variation in phenotypic PI susceptibility of full-length Gag-protease from HIV-1 infected patient at different time points. (A) Sequences of the viral clones showing the amino acid changes in the MA, CA, P2, NC, p1, and p6 regions of Gag between baseline (pre-PI treatment) and viral failure (during PI treatment). (B) Full-length Gag-protease sequence was amplified from plasma samples and cloned into p8.9NSX+. VSV-G pseudotyped viruses encoding luciferase were produced by cotransfection in 293T cells. The PI susceptibility of pseudovirions derived from each patient was measured by luciferase activity, as determined using a single-replication-cycle drug susceptibility assay. Data are fold differences in IC$_{50}$s of LPV in comparison to that for the assay reference strain, p8.9NSX. Error bars represent standard errors of the means from at least three independent experiments performed in duplicate.

tions in the susceptible virus led to a 4-fold decrease in susceptibility to LPV (FC in IC$_{50}$ from 5.3 to 22.7) (Fig. 2 and Table S2). Conversely, S126ins, H127ins, and the A122T and E123G substitutions in the LPV-resistant virus led to a 3-fold decrease in resistance (Fig. 2). We also tested the effect of the four-amino-acid signature on susceptibility to the second-generation PI darunavir (DRV) and found no significant impact (Fig. S4).

We sought to establish the effect of each of the four amino acid changes occurring alone. Using the resistant viral clone, four different mutant viruses were created with single amino acid changes in Gag: A122T, E123G, S126ins, and H127ins. Results of the phenotypic drug susceptibility testing of these mutants showed that only E123G appeared to increase susceptibility (Fig. S5), and the combination of the four amino acids had the greatest impact on LPV susceptibility.

We next tested whether the four-amino-acid signature T122A/G123E/S126del/H127del could confer LPV resistance in a different subtype context. We chose the

(A)

	HXB2 Pos	120	121	122	123	124	125	126	127	128	129	130	131	132
Sensitive virus	WT	A	A	T	G	S	S	S	H	-	S	Q	N	Y
	Mut 1	A	A	T	G	S	S	-	-	-	S	Q	N	Y
	Mut 2	A	A	**A**	**E**	S	S	-	-	-	S	Q	N	Y
Resistant virus	WT	A	A	A	E	S	S	S	H	-	S	Q	N	Y
	Mut 1	A	A	T	G	S	S	**S**	**H**	-	S	Q	N	Y
	Mut 2	A	A	**T**	**G**	S	S	**S**	**H**	-	S	Q	N	Y

(B)   **(C)**

FIG 2 Gag 126del and 127del mutations occurring with T122A and G123E confer resistance to the protease inhibitor lopinavir in the absence of any major protease mutations. (A) Sequences of the viral clones showing the amino acid changes (in red) introduced using standard site-directed mutagenesis techniques. (B and C) Full-length Gag-protease with the indicated mutations was amplified from plasma samples and cloned into p8.9NSX+. VSV-G-pseudotyped viruses encoding luciferase were produced by cotransfection in 293T cells. PI susceptibility of pseudovirions derived from each patient was measured by luciferase activity, as determined using a single-replication-cycle drug susceptibility assay. Data are fold differences in IC_{50}s of LPV in comparison to that for the assay reference strain, p8.9NSX. Error bars represent standard errors of the means from at least three independent experiments performed in duplicate.

reference p8.9NSX subtype B virus and made the amino acid deletions at Gag positions 126 and 127 as well as the adjacent T122A and G123E mutations. In addition, we added a V128 deletion, given that the subtype CRF02_AG consensus contains this deletion compared to subtype B. The five mutations (T122A/G123E/S126del/H127del/V128del) reduced susceptibility to LPV more than 3-fold, indicating that they are effective in a divergent subtype (Fig. 3).

Matrix/capsid (p17-p24) cleavage and differential PI susceptibility. We hypothesized that the efficiency of MA/CA cleavage of HIV-1 polyproteins would differ between the susceptible and resistant clones in the presence of LPV. To test this hypothesis, we employed Western blot analysis. Gag cleavage patterns were examined using the supernatants and cellular extracts of 293T cells transfected with each plasmid in the presence and absence of increasing concentrations of LPV (Fig. 4). We probed with a polyclonal p24 antibody, and as expected, there was incomplete cleavage of p24-p2 at higher LPV doses in the virus-containing supernatants and the cell extracts, consistent with previous data (32). We calculated p24/p41 ratios to specifically probe the p17/p24 cleavage site in the vicinity of the four-amino-acid signature. We found that the resistant virus cleaved p17/p24 more efficiently in the absence of drug and up to 30 nM LPV.

The resistance signature arises from a minority viral population detected at baseline. We proceeded to investigate when resistance emerged. Given the lengthy

344
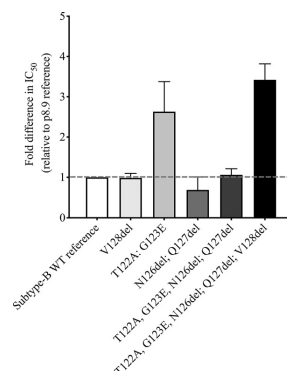
FIG 3 The four-amino-acid MA mutant signature can be introduced into subtype B to reduce PI susceptibility. Site-directed mutations were generated in the subtype B reference strain used in our assays. V128del was also added, as this deletion is present in HIV-1 CRF02_AG. Data are fold differences in IC$_{50}$s of LPV in comparison to that for the assay reference strain, p8.9NSX. Error bars represent standard errors of the means from at least two independent experiments performed in duplicate.

period of over 5 years between the two samples, we ideally needed a sample from an intermediate time point. We were able to identify a plasma sample from a patient on second-line therapy from 41 months, with a VL of 241,894 copies/ml. We refer to the 41-month time point as VF1 and the original 64-month time point as VF2. NGS analysis at the whole-genome level was undertaken for all 3 time points, and Table 2 shows variant frequencies at sites in Gag and Pol associated with drug exposure. Of note, we observed loss of mutations affecting susceptibility to lamivudine (M184I), tenofovir
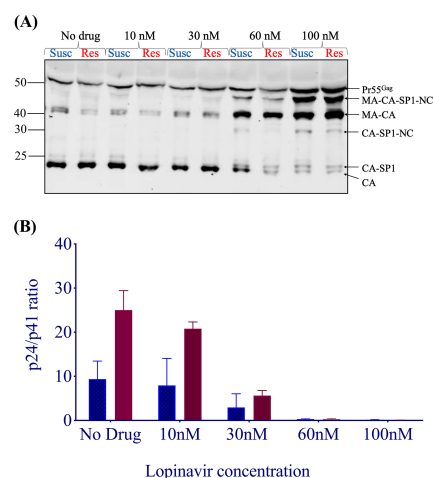
FIG 4 HIV-1 Gag cleavage efficiency in resistant (Res) versus susceptible (Susc) isolates. (A) Representative Western blot of virus-containing supernatant at increasing drug doses, using a p24 antibody. Mass (in kilodaltons) is indicated on the left. MA, matrix (p17); CA, capsid (p24); NC, nucleocapsid; SP1, spacer peptide 1. (B) Ratios of p24/p41 at increasing drug doses. Data are means and standard deviations from 2 independent experiments. In each pair, the left bar represents the wild type and the right bar represents resistant virus.

345

TABLE 2 NGS variant-derived data for three time points during LPV treatment

Gene	Mutation[a]	% of reads encoding mutation at[b]:		
		Baseline (0 mo) (405,158)	VF1 (41 mo) (250,932)	VF2 (64 mo) (604,157)
Gag	E12K	5	21	42
	R76K	0	0	2
	Y79F	0	0	2
	T122A	4.8	13	26
	G123E	5.0	13	26
	V128del	100	100	100
	V370A	5	0	1
	S373T	97	98	100
	R409K	3	0	1
	S451T	100	1.7	0
RT	K65R	98	0	0
	K103N	94	0	0
	E138K	0	0	4
	Y181C	100	0	0
	M184I	100	1.2	21
	M184V	0	0	79

[a]Gag mutations known to be associated with protease inhibitor exposure from prior reports and resistance-associated mutations in RT.
[b]Numbers in parentheses are total numbers of reads.

(K65R), and efavirenz (K103N) between baseline and VF1. The individual was prescribed lamivudine, zidovudine, and lopinavir/ritonavir for second-line therapy, and the resistance data indicate lack of drug pressure from lamivudine.

The NGS showed that T122A and G123E were present at low abundance before initiation of PI (approximately 5% of reads) (Table 2). The proportion of T122A/G123E increased at VF1 to 13%. These mutations were observed at increased frequency at VF2 both by target-enriched NGS and also direct *gag-pro* PCR from plasma, but NGS also showed emergence of the lamivudine resistance mutation M184V, suggesting improved adherence to the lamivudine regimen between VF1 and VF2.

We next generated whole-genome haplotypes for each time point using NGS data in order to first establish the phylogenetic relationships between viruses with differing PI resistance-associated mutations, and also to determine the coreceptor usage of virus haplotypes, as this might provide clues to the origins of virus variants (Fig. 5). All inferred haplotypes were predicted to use CCR5 with a false-positive rate (FPR) of <5%, and no CXCR4-using viruses were predicted in either of the two algorithms used.

We proceeded to clone sequences from plasma at VF1 in addition to those previously cloned from VF2 and inferred phylogenetic trees. None of the four *gag-pro* clones from baseline (before initiation of PI) contained any of the four amino acid changes (T122A, G123E, S126del, and H127del), consistent with NGS data showing that these variants were present at <5% (Table 2). Clones from the intermediate time point VF1 clustered with the VF2 clones rather than with the baseline clones (Fig. 6). Overall, there was excellent concordance between the inferred whole-genome haplotypes and *gag-pro* clones, though there appeared to be greater diversity in haplotypes. *In vitro* phenotypic drug susceptibility of cloned sequences revealed both sensitive and resistant viruses at VF1 as well as VF2 (Fig. 6), with the resistant clones from VF1 and VF2 clustering together and sharing the 4-amino-acid resistance-associated signature S126del/H127del/T122A/G123E. As expected, the susceptible clones from VF1 and VF2 also clustered with each other in a distinct part of the tree.

Persistence of both resistant and susceptible viruses can be explained by replication capacity. A surrogate for fitness in our assay is single-round infectivity (measured in relative light units [RLU]) in the absence of drug, which is given a value of 100% for our reference subtype B virus. We measured the single-round infectivity (replication capacity [RC]) of clones bearing patient-derived *gag-pro* sequences from each time point. Interestingly, resistant clones had a lower RC than susceptible viruses

346
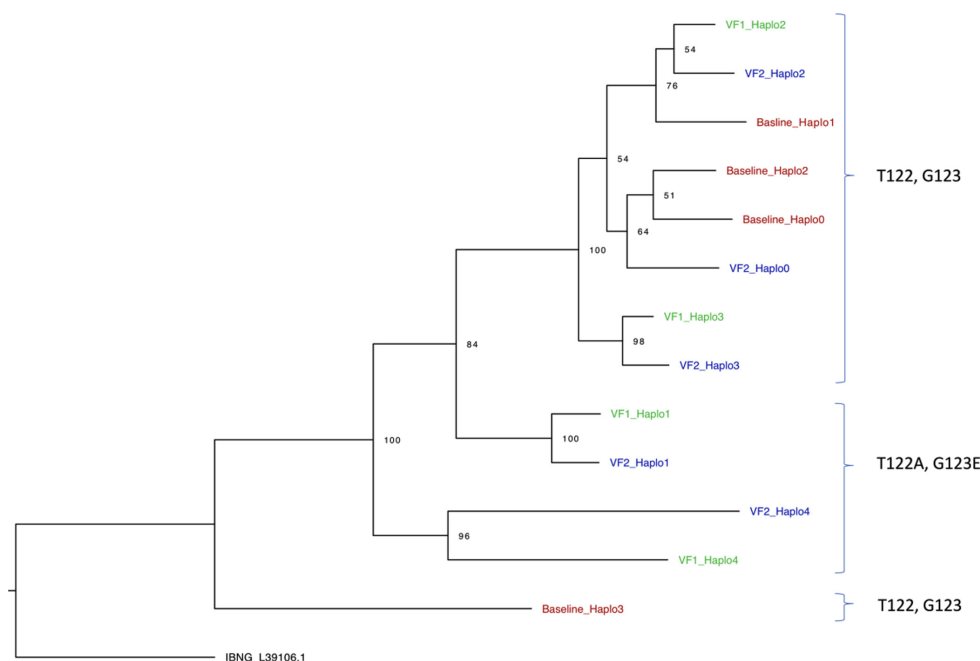
FIG 5 Whole-genome HIV haplotype reconstruction using target-enriched NGS Illumina MiSeq data from each time point (baseline, VF1, and VF2), with maximum-likelihood analysis and bootstrap support indicated using 1,000 replicates. Labels on the right are the amino acids at Gag positions 122 and 123.

(around 1.5-fold), regardless of whether they were isolated from VF1 or VF2 (Fig. 7). We also tested full-length replication-competent virus bearing the 4-amino-acid signature with the wild type over multiple rounds of replication and found a similar difference in RC (Fig. 7). The mixture of sensitive and resistant strains is consistent with incomplete drug adherence and therefore variable drug pressure, or alternatively with compartmentalization of virus sequences in anatomical areas with different drug levels.

DISCUSSION

Based on NMR and X-ray crystallography studies, p17 comprises five major alpha helices connected primarily by short loops (35, 37). The C terminus of matrix is predicted to be disordered, which has hampered efforts to characterize the structural characteristics of this region. One study suggested that deletions at 125 and 126 would stabilize p17 (38), indicating that despite disorder, changes in the region might lead to significant changes in stability and therefore possibly altered effects of protease inhibition on cleavage.

In this study on CRF02_AG and subtype G clinical isolates from a Nigerian cohort, we demonstrated the role of p17 amino acid mutations occurring near the p17/p24 cleavage site in PI resistance. The double deletion of Ser and His at Gag positions 126 and 127, respectively, had a modest impact on *in vitro* phenotypic PI susceptibility. When this deletion occurred alongside T122A and G123E, we observed a 4- to 5-fold decrease in susceptibility to lopinavir. The four-mutation combination was also able to confer similar resistance to a subtype B virus, indicating that it may emerge across subtypes.

We aimed to understand the mechanism at play in the T122A/G123E/S126del/

347
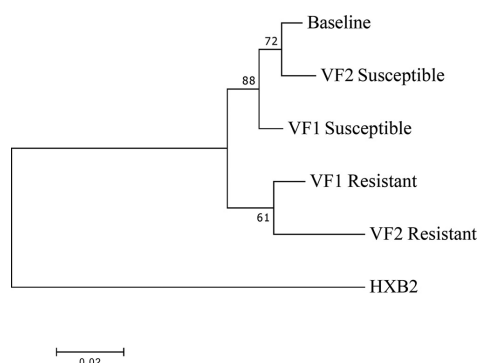
FIG 6 Phylogenetic relationships between viral Gag-protease plasma-derived sequences isolated at baseline (pre-PI) and at two failure time points (VF1 and VF2). The maximum-likelihood tree has bootstrap support indicated at the nodes. The outlier is HXB2, a subtype B virus. VF1, viral failure 1 at 41 months after initiation of protease inhibitor therapy; VF1, viral failure 2 at 64 months after initiation of protease inhibitor therapy.

H127del phenotype. Western blotting of virus-containing supernatants from producer cells revealed significant differences in cleavage without drug at the p17/p24 cleavage site, with the resistant clone demonstrating more efficient cleavage. In the presence of drug, Gag cleavage at the MA/CA cleavage site (as obtained from p24/p41 ratios) was also more efficient in the resistant viral clone. Therefore, rescue of infectivity in the presence of drug *in vivo* might be explained by inherently more efficient or kinetically favorable cleavage.

G123E was reported to arise when viruses were propagated with investigational protease inhibitors KNI-272 and UIC-94003 (39). Gag G123E was found to potentially interact with protease by NMR (35), providing a potential mechanism for its effect. This was more recently corroborated by Samsudin and colleagues (40), using multiscale modeling and simulations to reveal how non-cleavage site mutations can directly interact with cleavage site residues to affect their local environment. Through the use of contact analysis between the MA/CA cleavage site residues and Gag position 123 in wild-type (WT) (G123) and mutant (E123) proteins, the residue at position 123 was shown to make contact primarily with the N-terminal portion of the cleavage site from the same Gag subunit. Both WT (glycine) and mutant (glutamate) residues showed a similar percentage of contact over the course of the simulations. When the CG simulations were "back-mapped" and transformed into atomic resolution, atomistic simulations of a single MA-CA-SP1 subunit showed that the glutamate (mutant) residue at position 123, but not the glycine (wild-type) residue, interacted primarily with the cleavage site Y132 and also contacted residues N131 and Q130. Given the change in the overall size and charge of the residue in the WT and mutants (from small and neutral to large and acidic), the G123E mutation alters the accessibility and electrostatic properties in the vicinity of the cleavage site and therefore was expected to directly interfere with proteolysis. Although our present study implicates G123E in reduced PI susceptibility, we show here that the combination of mutations that was observed in the patient was needed for maximal effect.

We next used NGS to explore the dynamics of emergence of Gag amino acid changes during ongoing viremia under PI treatment. We were able to detect both T122A and G123 at low abundance at baseline, prior to PI exposure. Importantly, PCR from plasma RNA using *gag-pro*-specific primers did not amplify any sequences with these changes at baseline, highlighting an important contribution of NGS to the study of drug resistance. Whole-genome reconstruction enabled us to infer phylogenetic

348

(A)

(B)
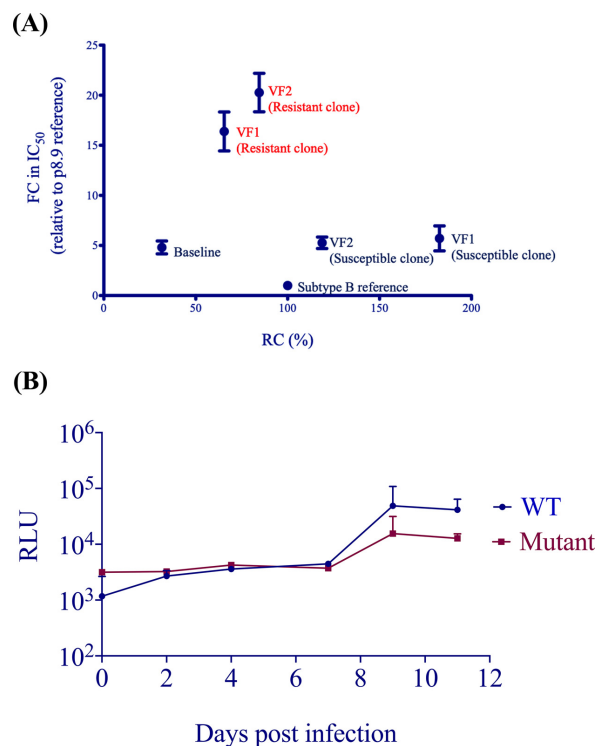
FIG 7 (A) Relationship between single-round infectivity (RC) and LPV susceptibility (FC in IC$_{50}$ compared to subtype B reference) for single-round VSV-G-pseudotyped viruses bearing patient-derived Gag-protease gene sequences. Error bars represent the standard error of the mean of at least two independent experiments performed in duplicate. (B) Comparison of replication capacity over multiple rounds of infection for wild-type Ba-L versus mutant bearing the 4-amino-acid Gag matrix signature T122A/G123E/126del/127del. Error bars represent the standard errors of the means for technical replicates. Data are representative of two independent experiments.

trees and confirm findings that resistance-conferring mutations occurred at both time points in phylogenetically related sequences. All virus haplotypes were predicted to use CCR5 and therefore to be sensitive to the CCR5 antagonist maraviroc. Intriguingly, we found that resistant viruses had lower replication efficiency than the wild type in both single-round and multiround infections when there was no drug present. These experiments support a model where the composition of viral quasispecies under nonsuppressive ART depends on drug levels and inherent differences in replication dynamics conferred by relatively small numbers of amino acids.

Our study provides further information on the role of Gag in resistance to protease inhibitors. Given that failure of treatment with protease inhibitors arose with no major mutations in protease, the Gag protein itself could be a target for the development of future therapeutics. Presently, there are no FDA-approved antiretroviral drugs that target HIV-1 Gag. A number of studies have attempted to establish Gag as a target. The design and development of drugs that target Gag could be approached in four broad ways, as reviewed by Su and colleagues (41). The first approach would involve screening and targeting of druggable allosteric sites present in Gag. A second approach is the

349

identification of novel Gag mutations and the use of models to preemptively design Gag inhibitors. The third approach is the use of synergistic drugs to target multiple sites by chemically joining different potential Gag inhibitors to function as dual or triple inhibitors. The fourth approach is to design inhibitors to disrupt the conformational transition of Gag during viral maturation (41).

The novel amino acid signatures that arose *in vivo* during treatment in the present study occurred in the matrix domain and at the non-cleavage site of Gag. The matrix (MA) domain of HIV-1 Gag plays critical roles in virus assembly by targeting the Gag precursor to the plasma membrane and directing the incorporation of the viral envelope (Env) glycoprotein into virions (42). A class of negatively charged lipids known as phosphoinositides, such as phosphatidylinositol-(4,5)-bisphosphate [$PI(4,5)P_2$], play an important role in the association of HIV-1 Gag with the plasma membrane. To target the MA of Gag, small molecules could be synthesized either to bind to the $PI(4,5)P_2$-binding cleft, thereby competing for an MA-P1(4,5)P_2 association, or to target the hydrophobic groove in the globular core matrix, which would then dysregulate the myristyl switch mechanism and block the association of Gag with the cell membrane, thereby disrupting virus assembly and release (43). This mechanism could have been involved in the reduction in susceptibility by our four-amino-acid resistance signature in matrix. The use of a small-molecule approach was also adopted by Machara et al. and led to the identification of two arylquinazolines which inhibited HIV-1 capsid assembly by binding to the C-terminal domain of capsid and blocking viral replication (44). Additionally, inhibitors could be synthesized to destabilize Gag assembly, thus slowing the viral maturation process (41).

In future work, it would be interesting and important to know whether Gag mutations are capable of facilitating emergence of major protease mutations in prolonged culture conditions under suboptimal drug pressure. This could potentially explain why prevalence of major protease mutations increases over time during PI exposure in clinical studies (45). Next, one could perform population dynamics simulations to incorporate RC and susceptibility data in order to model the proportion of resistant and susceptible viruses over time and possibly therefore predict emergence of major mutations in the protease gene.

Our data are limited by the small sample size, the lack of availability of plasma drug level measurements, and the use of standard clonal approaches as opposed to single-genome sequencing and amplification. This meant that we were not able to assess the contribution of minority variant populations to susceptibility. Some experiments were done in duplicate rather than triplicate. Nonetheless, we hypothesize that the four-amino-acid HIV-1 Gag signature is a contributory factor in PI failure in this PLWH from Nigeria.

As we move toward next-generation sequencing, this work highlights the limitations of current genotyping methods to infer PI susceptibility and supports sequencing outside protease to broaden the evidence base for the clinical management of patients who experience VF on PIs without major protease mutations. The work may ultimately also help to identify define individuals with lower PI susceptibility before treatment with this class of drugs.

MATERIALS AND METHODS

Study participants. We identified six individuals on second-line, protease inhibitor-based ART who experienced virological failure without major protease mutations from a PEPFAR-funded treatment cohort in Nigeria and who had samples collected at at least two time points (before second-line treatment initiation and following second-line virologic failure). Having previously reported that baseline phenotypic susceptibility was not associated with subsequent virologic "failure" (46) in this cohort, here we sought to explore changes over time in phenotypic susceptibility that could be associated with changes in HIV-1 Gag and protease genes.

Next-generation sequencing. Manual nucleic acid extraction was done using the QIAamp viral RNA minikit (Qiagen, Hilden, Germany) with a plasma input volume of 0.5 to 1.5 ml. The first strand of cDNA was synthesized using SuperScript IV reverse transcriptase (Invitrogen, Waltham, MA, USA), followed by NEBNext second-strand cDNA synthesis (E6111; New England Biolabs GmbH, Frankfurt, Germany). Sample libraries were prepared as per the SureSelect^XT automated target enrichment protocol (Agilent Technologies, Santa Clara, CA, USA) with in-house HIV baits. Whole-genome deep sequencing was

350

performed using the Illumina MiSeq platform (Illumina, San Diego, CA, USA). Trimmed reads were then compared to a reference panel of 170 HIV subtypes/CRFs (circulating recombinant forms) from the Los Alamos database (https://www.hiv.lanl.gov), and the best match was used for reference mapping. Duplicate reads were removed from the BAM files, and a consensus sequence was generated using a 50% threshold. Mutations were included if they were present at a frequency greater than 2% within the read mixture at that position, with a minimum read depth of 100. An in-house custom script was used to identify single nucleotide polymorphisms (SNPs) at each position by BLAST analysis of individual HIV *pol* genes against the HXB2 reference genome.

Haplotype reconstruction and phylogenetics. Whole-genome haplotype reconstruction was performed using a newly developed maximum-likelihood method, HaROLD (haplotype assignment of virus NGS data using covariation of variant frequencies [47]). SNPs were assigned to each haplotype so that the frequency of a variant at any time point was represented by the sum of the frequencies of the haplotypes containing that variant. Time-dependent frequencies for longitudinal haplotypes were optimized by maximizing the log likelihood, which was calculated by summing over all possible assignments of variants to haplotypes. Haplotypes were then reconstructed based on posterior probabilities. The calculations were repeated with a range of possible haplotype numbers, and the optimal number of haplotypes was determined by the resulting value of the log likelihood. After construction of haplotypes, a refinement process remapped reads from BAM files to the constructed haplotypes. Haplotypes were also combined or divided according to Akaike information criterion (AIC) scores, in order to give the most accurate representation of viral populations. Phylogenetic trees of constructed haplotypes were constructed using RAxML-NG using the general time-reversible (GTR) model and 1,000 bootstraps.

Coreceptor usage. CCR5/CXCR4 usage was predicted using *env* sequences with the online tools Geno2Pheno (https://www.geno2pheno.org) and WebPSSM (https://indra.mullins.microbiol.washington.edu/webpssm/).

Amplification of full-length Gag-protease genes. We amplified from plasma taken before PI initiation and from a failure time point for each individual. NGS was used to obtain a consensus whole-genome sequence for each of the 12 samples. Full-length Gag-protease gene sequences were obtained from plasma by standard PCR; HIV-1 RNA was extracted from plasma samples using the QIAamp viral RNA extraction kit. Using previously described techniques (48, 49), the full-length Gag-protease sequence was amplified and cloned into a subtype B-based (p8.9NSX+) vector. Clonal sequencing of up to 10 plasmids was performed by standard Sanger sequencing. The variant that most closely represented the next-generation sequencing-derived consensus was taken forward for phenotypic testing. Sequences were manually analyzed using DNADynamo software (http://www.bluetractorsoftware.co.uk). Protease sequences were analyzed for PI resistance mutations using the Stanford Resistance Database (https://hivdb.stanford.edu). Phylogenetic analysis was performed using maximum-likelihood methods in MEGA v7.0 (50). Bootstrapping was performed as previously described (28).

Site-directed mutagenesis. Site-directed mutagenesis was carried out using the QuikChange kit (Stratagene) according to the manufacturer's instructions. Mutagenesis was verified by Sanger sequencing.

PI susceptibility and infectivity assays. PI susceptibility and viral infectivity were determined using a previously described single assay. Briefly, 293T cells were cotransfected with a Gag-Pol protein expression vector (p8.9NSX) containing cloned patient-derived full-length Gag-protease sequences, pMDG (which expresses vesicular stomatitis virus envelope glycoprotein [VSV-G]), and pCSFLW (which expresses the firefly luciferase reporter gene with the HIV-1 packaging signal) as previously described. PI drug susceptibility testing was carried out as previously described (48). Transfected cells were seeded with serial dilutions of lopinavir, and harvested pseudovirions were used to infect fresh 293T cells. To determine strain infectivity, virus was produced in the absence of drug.

Infectivity was monitored by measuring luciferase activity 48 h after infection. Results derived from at least two independent experiments (each in duplicate) were analyzed. The IC_{50} was calculated using GraphPad Prism 5 (GraphPad Software Inc., La Jolla, CA, USA). Susceptibility was expressed as fold change in IC_{50} compared to that of the subtype B reference plasmid p8.9NSX. Replicative capacity of these viruses was assessed by comparing the luciferase activity of recombinant virus with that of the WT subtype B control virus in the absence of drug. Equal amounts of input plasmid DNA were used, and it has previously been shown that percentage infectivity correlates well with infectivity per nanogram of p24 in this system (48). Differences in PI susceptibility were compared with the paired *t* test.

Multiround infectivity assay. WT (R9-BaL) and mutant (R9-BaL with the 5 amino acid changes in MA) virus preparations were used to infect 1.5×10^6 of SupT1-CCR5 suspension cells in 2 ml of medium per well and incubated at 37°C for 2 h, followed by low-speed centrifugation ($800 \times g$) for 10 min. The supernatant was discarded, and the cell pellets were resuspended in RPMI medium and used to infect 4×10^6 SupT1-CCR5 cells. Infectious virion supernatant was harvested on days 2, 4, 7, 9, and 11. The harvested virion supernatant was used to infect fresh TZM-bl cells to assay for infectivity, which was based on the Tat-dependent upregulation of long terminal repeat (LTR)-driven firefly luciferase expression upon HIV-1 infection of TZM-bl cells. Luciferase assay reagent was added, and the luminescence was measured using a GloMax 96 microplate luminometer (Promega).

The PI drugs used in this study were obtained from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH.

Western blot analysis. Using a previously described method (51), equal amounts of each of the viral clone plasmid were used to transfect 293T cells, in addition to a VSV-G plasmid and reporter genome-

351

expressing plasmid. Each of the pseudovirions was produced in the absence and presence of a range of concentrations of LPV, added 16 h following transfection.

Forty-eight hours after transfection with the plasmid preparations, the culture supernatant was harvested and passed through a 0.45-μm-pore-size filter to remove cellular debris. The filtrate was centrifuged at 14,000 rpm for 90 min to pellet virions. The pelleted virions were lysed in Laemmli reducing buffer (1 M Tris-HCl [pH 6.8], SDS, 100% glycerol, β-mercaptoethanol, and bromophenol blue). Cell lysates were subjected to electrophoresis on SDS–4 to 12% bis-Tris protein gels (Thermo Fisher Scientific) under reducing conditions. This was followed by electroblotting onto polyvinylidene difluoride (PVDF) membranes. The HIV-1 Gag proteins were visualized by a transilluminator (Alpha Innotech) using anti-p24 Gag antibody.

Ethics. Informed consent was obtained from all participants, and ethics approval for virological testing was obtained from the Nigeria National Research Ethics Committee of Nigeria (NHREC/01/01/ 2007). Ethical approval was also obtained from the ethics board of University College London, United Kingdom.

Data availability. Sequences are available from GenBank under accession numbers MW125626 to MW125640.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, DOCX file, 0.1 MB.
FIG S2, DOCX file, 0.1 MB.
FIG S3, DOCX file, 0.1 MB.
FIG S4, DOCX file, 0.4 MB.
FIG S5, DOCX file, 0.3 MB.
TABLE S1, DOCX file, 0.3 MB.
TABLE S2, DOCX file, 0.01 MB.

REFERENCES

1. Gupta RK, Hill A, Sawyer AW, Cozzi-Lepri A, von Wyl V, Yerly S, Lima VD, Gunthard HF, Gilks C, Pillay D. 2009. Virological monitoring and resistance to first-line highly active antiretroviral therapy in adults infected with HIV-1 treated under WHO guidelines: a systematic review and meta-analysis. Lancet Infect Dis 9:409–417. https://doi.org/10.1016/S1473-3099(09) 70136-7.

2. TenoRes Study Group. 2016. Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study. Lancet Infect Dis 16:565–575. https://doi.org/10.1016/S1473-3099(15)00536-8.

3. Boender TS, Sigaloff KC, McMahon JH, Kiertiburanakul S, Jordan MR, Barcarolo J, Ford N, Rinke de Wit TF, Bertagnolio S. 2015. Long-term virological outcomes of first-line antiretroviral therapy for HIV-1 in low- and middle-income countries: a systematic review and meta-analysis. Clin Infect Dis 61:1453–1461. https://doi.org/10.1093/cid/civ556.

4. Gupta RK, Jordan MR, Sultan BJ, Hill A, Davis DH, Gregson J, Sawyer AW, Hamers RL, Ndembi N, Pillay D, Bertagnolio S. 2012. Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. Lancet 380: 1250–1258. https://doi.org/10.1016/S0140-6736(12)61038-1.

5. Gupta RK, Gregson J, Parkin N, Haile-Selassie H, Tanuri A, Andrade Forero L, Kaleebu P, Watera C, Aghokeng A, Mutenda N, Dzangare J, Hone S, Hang ZZ, Garcia J, Garcia Z, Marchorro P, Beteta E, Giron A, Hamers R, Inzaule S, Frenkel LM, Chung MH, de Oliveira T, Pillay D, Naidoo K, Kharsany A, Kugathasan R, Cutino T, Hunt G, Avila Rios S, Doherty M, Jordan MR, Bertagnolio S. 2018. HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. Lancet Infect Dis 18: 346–355. https://doi.org/10.1016/S1473-3099(17)30702-8.

6. Rhee SY, Blanco JL, Jordan MR, Taylor J, Lemey P, Varghese V, Hamers RL, Bertagnolio S, de Wit TF, Aghokeng AF, Albert J, Avi R, Avila-Rios S, Bessong PO, Brooks JI, Boucher CA, Brumme ZL, Busch MP, Bussmann H, Chaix ML, Chin BS, D'Aquin TT, De Gascun CF, Derache A, Descamps D, Deshpande AK, Djoko CF, Eshleman SH, Fleury H, Frange P, Fujisaki S, Harrigan PR, Hattori J, Holguin A, Hunt GM, Ichimura H, Kaleebu P, Katzenstein D, Kiertiburanakul S, Kim JH, Kim SS, Li Y, Lutsar I, Morris L, Ndembi N, Ng KP, Paranjape RS, Peeters M, Poljak M, Price MA, et al. 2015. Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient- and sequence-level meta-analysis. PLoS Med 12: e1001810. https://doi.org/10.1371/journal.pmed.1001810.

7. WHO. 2017. HIV drug resistance report. http://www.who.int/hiv/pub/ drugresistance/hivdr-report-2017/en/. Accessed 1 July 2020.

352

8. WHO. 2017. Consolidated guidelines on person-centred HIV patient monitoring and case surveillance. http://www.who.int/hiv/pub/guidelines/person-centred-hiv-monitoring-guidelines/en/. Accessed 1 July 2020.

9. Collier DA, Monit C, Gupta RK. 2019. The impact of HIV-1 drug escape on the global treatment landscape. Cell Host Microbe 26:48–60. https://doi.org/10.1016/j.chom.2019.06.010.

10. Collier D, Iwuji C, Derache A, de Oliveira T, Okesola N, Calmy A, Dabis F, Pillay D, Gupta RK, French National Agency for AIDS and Viral Hepatitis Research 12249 Treatment as Prevention Study Group. 2017. Virological outcomes of second-line protease inhibitor-based treatment for human immunodeficiency virus type 1 in a high-prevalence rural South African setting: a competing-risks prospective cohort analysis. Clin Infect Dis 64:1006–1016. https://doi.org/10.1093/cid/cix015.

11. Ajose O, Mookerjee S, Mills EJ, Boulle A, Ford N. 2012. Treatment outcomes of patients on second-line antiretroviral therapy in resource-limited settings: a systematic review and meta-analysis. AIDS 26:929–938. https://doi.org/10.1097/QAD.0b013e328351f5b2.

12. Court R, Gordon M, Cohen K, Stewart A, Gosnell B, Wiesner L, Maartens G. 2016. Random lopinavir concentrations predict resistance on lopinavir-based antiretroviral therapy. Int J Antimicrob Agents 48:158–162. https://doi.org/10.1016/j.ijantimicag.2016.04.030.

13. Cohen K, Stewart A, Kengne AP, Leisegang R, Coetsee M, Maharaj S, Dunn L, Hislop M, van Zyl G, Meintjes G, Maartens G. 2019. A clinical prediction rule for protease inhibitor resistance in patients failing second-line antiretroviral therapy. J Acquir Immune Defic Syndr 80:325–329. https://doi.org/10.1097/QAI.0000000000001923.

14. van Zyl GU, van Mens TE, McIlleron H, Zeier M, Nachega JB, Decloedt E, Malavazzi C, Smith P, Huang Y, van der Merwe L, Gandhi M, Maartens G. 2011. Low lopinavir plasma or hair concentrations explain second-line protease inhibitor failures in a resource-limited setting. J Acquir Immune Defic Syndr 56:333–339. https://doi.org/10.1097/QAI.0b013e31820dc0cc.

15. Perno CF, Cozzi-Lepri A, Forbici F, Bertoli A, Violin M, Stella Mura M, Cadeo G, Orani A, Chirianni A, De Stefano C, Balotta C, d'Arminio Monforte A, Italian Cohort Naive Antiretrovirals Study Group. 2004. Minor mutations in HIV protease at baseline and appearance of primary mutation 90M in patients for whom their first protease-inhibitor antiretroviral regimens failed. J Infect Dis 189:1983–1987. https://doi.org/10.1086/386307.

16. Fun A, Wensing AM, Verheyen J, Nijhuis M. 2012. Human immunodeficiency virus Gag and protease: partners in resistance. Retrovirology 9:63. https://doi.org/10.1186/1742-4690-9-63.

17. El Bouzidi K, Collier D, Nastouli E, Copas AJ, Miller RF, Gupta RK. 2016. Virological efficacy of PI monotherapy for HIV-1 in clinical practice. J Antimicrob Chemother 71:3228–3234. https://doi.org/10.1093/jac/dkw265.

18. Thompson JA, Kityo C, Dunn D, Hoppe A, Ndashimye E, Hakim J, Kambugu A, van Oosterhout JJ, Arribas J, Mugyenyi P, Walker AS, Paton NI. 2019. Evolution of protease inhibitor resistance in human immunodeficiency virus type 1 infected patients failing protease inhibitor monotherapy as second-line therapy in low-income countries: an observational analysis within the EARNEST Randomized Trial. Clin Infect Dis 68:1184–1192. https://doi.org/10.1093/cid/ciy589.

19. Delaugerre C, Flandre P, Chaix ML, Ghosn J, Raffi Fois, Dellamonica Pierre, Jaeger H, Schürmann D, Cohen-Codar I, Ngo Van P, Norton M, Taburet AM, Delfraissy JF, Rouzioux C. 2009. Protease inhibitor resistance analysis in the MONARK trial comparing first-line lopinavir-ritonavir monotherapy versus lopinavir-ritonavir plus zidovudine and lamivudine triple therapy. Antimicrob Agents Chemother 53:2934–2939. https://doi.org/10.1128/AAC.01643-08.

20. Van Duyne R, Kuo LS, Pham P, Fujii K, Freed EO. 2019. Mutations in the HIV-1 envelope glycoprotein can broadly rescue blocks at multiple steps in the virus replication cycle. Proc Natl Acad Sci U S A 116:9040–9049. https://doi.org/10.1073/pnas.1820333116.

21. Rabi SA, Laird GM, Durand CM, Laskey S, Shan L, Bailey JR, Chioma S, Moore RD, Siliciano RF. 2013. Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. J Clin Invest 123:3848–3860. https://doi.org/10.1172/JCI67399.

22. Manasa J, Varghese V, Pond SLK, Rhee SY, Tzou PL, Fessel WJ, Jang KS, White E, Rognvaldsson T, Katzenstein DA, Shafer RW. 2017. Evolution of gag and gp41 in patients receiving ritonavir-boosted protease inhibitors. Sci Rep 7:11559. https://doi.org/10.1038/s41598-017-11893-8.

23. 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. J Int AIDS Soc 20:e25024. https://doi.org/10.1002/jia2.25024.

24. Clavel F, Mammano F. 2010. Role of Gag in HIV resistance to protease inhibitors. Viruses 2:1411–1426. https://doi.org/10.3390/v2071411.

25. Stray KM, Callebaut C, Glass B, Tsai L, Xu L, Muller B, Krausslich HG, Cihlar T. 2013. Mutations in multiple domains of Gag drive the emergence of in vitro resistance to the phosphonate-containing HIV-1 protease inhibitor GS-8374. J Virol 87:454–463. https://doi.org/10.1128/JVI.01211-12.

26. Malet I, Roquebert B, Dalban C, Wirden M, Amellal B, Agher R, Simon A, Katlama C, Costagliola D, Calvez V, Marcelin AG. 2007. Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. J Infect 54:367–374. https://doi.org/10.1016/j.jinf.2006.06.012.

27. Giandhari J, Basson AE, Sutherland K, Parry CM, Cane PA, 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. Antimicrob Agents Chemother 60:2248–2256. https://doi.org/10.1128/AAC.02682-15.

28. Sutherland KA, Mbisa JL, Ghosn J, Chaix ML, Cohen-Codar I, Hue S, Delfraissy JF, Delaugerre C, Gupta RK. 2014. Phenotypic characterization of virological failure following lopinavir/ritonavir monotherapy using full-length Gag-protease genes. J Antimicrob Chemother 69:3340–3348. https://doi.org/10.1093/jac/dku296.

29. Sutherland KA, Parry CM, McCormick A, Kapaata A, Lyagoba F, Kaleebu P, Gilks CF, Goodall R, Spyer M, Kityo C, Pillay D, Gupta RK, Group DV, DART Virology Group. 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. https://doi.org/10.1371/journal.pone.0137834.

30. Sutherland KA, Goodall RL, McCormick A, Kapaata A, Lyagoba F, Kaleebu P, Thiltgen G, Gilks CF, Spyer M, Kityo C, Pillay D, Dunn D, Gupta RK, Group DV, Team DT, DART Trial Team. 2015. Gag-protease sequence evolution following protease inhibitor monotherapy treatment failure in HIV-1 viruses circulating in East Africa. AIDS Res Hum Retroviruses 31:1032–1037. https://doi.org/10.1089/aid.2015.0138.

31. Nijhuis M, van Maarseveen NM, Lastere S, Schipper P, Coakley E, Glass B, Rovenska M, de Jong D, Chappey C, Goedegebuure IW, Heilek-Snyder G, Dulude D, Cammack N, Brakier-Gingras L, Konvalinka J, Parkin N, Krausslich HG, Brun-Vezinet F, Boucher CA. 2007. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. PLoS Med 4:e36. https://doi.org/10.1371/journal.pmed.0040036.

32. Dam E, Quercia R, Glass B, Descamps D, Launay O, Duval X, Krausslich HG, Hance AJ, Clavel F, ANRS 109 Study Group. 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. https://doi.org/10.1371/journal.ppat.1000345.

33. Mammano F, Petit C, Clavel F. 1998. Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. J Virol 72:7632–7637. https://doi.org/10.1128/JVI.72.9.7632-7637.1998.

34. Descours B, Cribier A, Chable-Bessia C, Ayinde D, Rice G, Crow Y, Yatim A, Schwartz O, Laguette N, Benkirane M. 2012. SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4(+) T-cells. Retrovirology 9:87. https://doi.org/10.1186/1742-4690-9-87.

35. Deshmukh L, Louis JM, Ghirlando R, Clore GM. 2016. Transient HIV-1 Gag-protease interactions revealed by paramagnetic NMR suggest origins of compensatory drug resistance mutations. Proc Natl Acad Sci U S A 113:12456–12461. https://doi.org/10.1073/pnas.1615342113.

36. Gregson J, Kaleebu P, Marconi VC, van Vuuren C, Ndembi N, Hamers RL, Kanki P, Hoffmann CJ, Lockman S, Pillay D, de Oliveira T, Clumeck N, Hunt G, Kerschberger B, Shafer RW, Yang C, Raizes E, Kantor R, Gupta RK. 2017. Occult HIV-1 drug resistance to thymidine analogues following failure of first-line tenofovir combined with a cytosine analogue and nevirapine or efavirenz in sub Saharan Africa: a retrospective multicentre cohort study. Lancet Infect Dis 17:296–304. https://doi.org/10.1016/S1473-3099(16)30469-8.

37. Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI. 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. Proc Natl Acad Sci U S A 93:3099–3104. https://doi.org/10.1073/pnas.93.7.3099.

38. Dolcetti R, Giagulli C, He W, Selleri M, Caccuri F, Eyzaguirre LM, Mazzuca P, Corbellini S, Campilongo F, Marsico S, Giombini E, Muraro E, Rozera G, De Paoli P, Carbone A, Capobianchi MR, Ippolito G, Fiorentini S, Blattner WA, Lu W, Gallo RC, Caruso A. 2015. Role of HIV-1 matrix protein p17

353

variants in lymphoma pathogenesis. Proc Natl Acad Sci U S A 112: 14331–14336. https://doi.org/10.1073/pnas.1514748112.

39. Gatanaga H, Suzuki Y, Tsang H, Yoshimura K, Kavlick MF, Nagashima K, Gorelick RJ, Mardy S, Tang C, Summers MF, and, Mitsuya H. 2002. Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. J Biol Chem 277:5952–5961. https://doi.org/10.1074/jbc.M108005200.

40. Samsudin F, Gan SK-E, Bond PJ. 2020. The structural basis for Gag non-cleavage site mutations in determining HIV-1 viral fitness. bioRxiv https://doi.org/10.1101/2020.07.05.188326.

41. Su CT-T, Koh DW-S, Gan SK-E. 2019. Reviewing HIV-1 Gag mutations in protease inhibitors resistance: insights for possible novel Gag inhibitor designs. Molecules 24:3243. https://doi.org/10.3390/molecules24183243.

42. Tedbury PR, Novikova M, Alfadhli A, Hikichi Y, Kagiampakis I, KewalRamani VN, Barklis E, Freed EO. 2019. HIV-1 matrix trimerization-impaired mutants are rescued by matrix substitutions that enhance envelope glycoprotein incorporation. J Virol 94:e01526-19. https://doi.org/10.1128/JVI.01526-19.

43. Waheed AA, Freed EO. 2012. HIV type 1 Gag as a target for antiviral therapy. AIDS Res Hum Retroviruses 28:54–75. https://doi.org/10.1089/AID.2011.0230.

44. Machara A, Lux V, Kožíšek M, Grantz Šašková K, Štěpánek O, Kotora M, Parkan K, Pávová M, Glass B, Sehr P, Lewis J, Müller B, Kräusslich HG, Konvalinka J. 2016. Specific inhibitors of HIV capsid assembly binding to the C-terminal domain of the capsid protein: evaluation of 2-arylquinazolines as potential antiviral compounds. J Med Chem 59:545–558. https://doi.org/10.1021/acs.jmedchem.5b01089.

45. Rawizza HE, Chaplin B, Meloni ST, Darin KM, Olaitan O, Scarsi KK, Onwuamah CK, Audu RA, Chebu PR, Imade GE, Okonkwo P, Kanki PJ. 2013. Accumulation of protease mutations among patients failing second-line antiretroviral therapy and response to salvage therapy in Nigeria. PLoS One 8:e73582. https://doi.org/10.1371/journal.pone.0073582.

46. Datir R, El Bouzidi K, Dakum P, Ndembi N, Gupta RK. 2019. Baseline PI susceptibility by HIV-1 Gag-protease phenotyping and subsequent virological suppression with PI-based second-line ART in Nigeria. J Antimicrob Chemother 74:1402–1407. https://doi.org/10.1093/jac/dkz005.

47. Goldstein RA, Tamuri AU, Roy S, Breuer J. 2018. Haplotype assignment of virus NGS data using co-variation of variant frequencies. bioRxiv https://doi.org/10.1101/444877.

48. Parry CM, Kohli A, Boinett CJ, Towers GJ, McCormick AL, Pillay D. 2009. Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1. J Virol 83:9094–9101. https://doi.org/10.1128/JVI.02356-08.

49. Gupta RK, Kohli A, McCormick AL, Towers GJ, Pillay D, Parry CM. 2010. Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays. AIDS 24:1651–1655. https://doi.org/10.1097/qad.0b013e3283398216.

50. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33: 1870–1874. https://doi.org/10.1093/molbev/msw054.

51. Gupta RK, Mlcochova P, Pelchen-Matthews A, Petit SJ, Mattiuzzo G, Pillay D, Takeuchi Y, Marsh M, Towers GJ. 2009. Simian immunodeficiency virus envelope glycoprotein counteracts tetherin/BST-2/CD317 by intracellular sequestration. Proc Natl Acad Sci U S A 106:20889–20894. https://doi.org/10.1073/pnas.0907075106.

354