Identification of mutational pathways to Tenofovir resistance in subtype C isolates using a Bayesian Network.

By

Ntombikhona F Maphumulo

(Student number: 212561310)

Submitted in fulfilment of requirements of the degree Master of Medical Science in Virology, School of Laboratory Medicine and Medical Science, Faculty of Health Sciences, University of KwaZulu-Natal.

2016

Supervisor: Dr. Michelle L. Gordon

Preface

The experimental work described in this dissertation was carried out in HIV-1 Pathogenesis Program at the Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, from April 2014 to October 2015 under the supervision of Dr Michelle Gordon.

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

Signed: _____ Date : 12 May 2017
Ntombikhona Maphumulo (Student)

Signed: _____
Dr Michelle Gordon (Supervisor)

Declaration

I, Ntombikhona Fortunate Maphumulo, declare that:

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

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

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

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

a. Their words have been re-written but the general information attributed to them has been referenced;

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

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

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

Signed: _____ Date : 12 May 2017

Ntombikhona Maphumulo (Student)

Signed: _____ Date: 12 May 2017

Dr Michelle Gordon (Supervisor)

Ethical Approval

Ethical approval for this study was obtained from the Biomedical Research Ethics Committee of the Nelson R. Mandela School of Medicine, University of KwaZulu-Natal (BE024/16).

Presentations

Maphumulo NF and Gordon ML. *Mutational Associated with Tenofovir Resistance in Subtype C Isolates.* University of Kwa-Zulu Natal, Durban, South Africa, School of Laboratory Medicine and Medical Sciences Research Day, 5 August 2016. (Oral poster presentation).

Maphumulo NF and Gordon ML. Mutational Associated with Tenofovir Resistance in Subtype C Isolates. *HIV-1 Research for Prevention 2016: AIDS Vaccine, Microbicide and ARV-based Prevention Science (HIV-1R4P),* Chicago, USA from 17 – 21 October 2016. (Poster presentation).

Acknowledgements

First and foremost, To God Almighty, it only your grace that's has lead me thus far.
I wish to express my sincere gratitude to my supervisor Dr Michelle Gordon, for the many opportunities she has given me, for her guidance, support, encouragement and teachings, throughout my degree.

To the director of the HIV Pathogenesis Program, Professor Thumbi N'dungu, thank you for giving me the opportunity to further my studies.
The University of Kwa-Zulu Natal's College of Health Sciences, and the HIV Pathogenesis Program are acknowledged for their financial support.

To my mother (Virginia ma-Ngcobo Maphumulo), thank you for all you have instilled in me, my every achievement and success is yours.
To my family, thank you for your support and confidence in me, your love and smiles kept me going.

To you, my friend, my love (Sinaye Ngcapu), the priceless gift of your love, care, and encouragement has seen me through each and every accomplishment, I am truly blessed to have you in my life. My dearest angel (Sinesihle Maphumulo), mummy loves you.

Table of Contents

List of Figures

List of table

Abbreviations

3D	Three Dimensional
3TC	Lamivudine
ABC	Abacavir
AIDS	Acquired Immunodeficiency Syndrome
ARVs	Antiretrovirals
ATP	Adenosine Triphosphate
ATV	Atazanavir
AZT	Zidovudine
BN	Bayesian Network
cDNA	Complementary DNA
CRFs	Circulating Recombinant Forms
d4T	Stavudine
ddC	Zalcitabine
ddI	Didanosine
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleoside Triphosphate
dsDNA	Double Stranded Deoxyribonucleic Acid
EFV	Efavirenz
ETR	Etravirine
Env	Envelope Protein
FTC	Emtricitibine
Gag	Group-Specific Antigen
gp120	Glycoprotein 120
HAART	Highly Active Antiretroviral Therapy

HIV-1	Human Immunodeficiency Syndrome
HIV-1	Human Immunodeficiency Syndrome Type 1
HIV-1-2	Human Immunodeficiency Syndrome Type II
LPV/r	Lopinavir boosted with Ritonavir
mRNA	Messenger Ribonucleic Acid
NLS	Nuclear Localization Sequence
NES	Nuclear Export Sequence
Nef	Negative Regulatory Factor
NRTIs	Nucleotide Reverse Transcriptase Inhibitors
NNRTIs	Non-nucleoside Reverse Transcriptase Inhibitors
NVP	Navarapine
gp41	Glycoprotein 41
P24	p24 Capsid protein
PDB	Protein Data Bank
PI	Protease Inhibitors
Pol	Polymerase
PPi	Pyrophosphate
Rev	Regulator of Virion
RNA	Ribonucleic Acid
RNase H	Ribonuclease H
RPV	Rilpivirine
RT	Reverse Transcriptase
ssRNA	Single Stranded Ribonucleic Acid
TAMs	Thymidine Analogue mutations

Tat	Trans-activator of Transcription
TDF	Tenofovir
vif	Viral Infectivity Factor
vpr	Viral Protein R
vpu	Viral Protein U
α	Alpha
β	Beta
&	and
-OH	Hydroxyl group

Chapter 1: Literature Review

1. Introduction

South Africa is currently one of the countries most affected by the Acquired Immunodeficiency Syndrome (AIDS) epidemic in the world (UNAIDS, 2013). In 2013, approximately 6.3 million (6 – 6.5 million) people were living with Human Immunodeficiency Syndrome – type 1 (HIV-1), with over 200 000 AIDS related deaths occurring annually (UNAIDS, 2013). The subtype-C epidemic is the largest worldwide and is also the major circulating HIV-1 subtype in South Africa (Department of Health, 2014). In 2016, the South African national treatment rollout program increased access to antiretroviral treatment for people living with HIV-1, with approximately 3.4 million people (with CD4 count under 500 cells/µl) receiving highly active antiretroviral therapy (HAART) in the public sector alone (UNAIDS, 2016). The South African national treatment program includes 2 nucleoside reverse transcriptase inhibitors (NRTIs) {AZT/TDF and 3TC or FTC if there are contradictions to 3TC} and 1 non-nucleoside reverse transcriptase inhibitors (NNRTI) {EFV or NVP for pregnant women} provided as a fixed-dose combination in the first line regimen and 2 NRTIs (AZT/TDF and 3TC or FTC and 1 PI (LPV/r or ATV/r) in the second line regime (Department of Health, 2014).

However, the gains (including, improved prognosis of HIV-1 infected people and quality of life) due to the increased access to HAART risk being eroded by the emergence of drug resistance (Ives et al., 2001). Drug resistance is attributed to the high replication capacity of HIV-1, with 10^{10} viral particles being produced daily in people living with HIV-1 (Ives et al., 2001, Preston et al., 1988). In addition, reverse transcriptase (RT), that helps HIV-1 to replicate, lacks proofreading capabilities resulting in the incorporation of the wrong nucleotide at least once per replication cycle (Preston et al., 1988). Consequently, this results in large variation in new particles produced (Roberts et al., 1988). In the development of drug resistance, antiretroviral drugs (ARVs) apply selective pressure on these naturally mutating viruses (Coffin, 1995). In the presence of incomplete viral suppression, variant with a competitive advantage are selected and these drug resistant mutants become the dominant strains thereafter (Coffin, 1995). Drug resistance is the leading factor for treatment failure. Other factors include a low genetic barrier to resistance of certain drugs, poor adherence, and inadequate drug plasma concentrations and the presence of pre-existing drug resistant mutants (Coffin, 1995).

Studies have well characterized the development of nucleoside reverse transcriptase inhibitor resistance, with AZT/d4T associated thymidine analogue resistance mutations (TAMs) being the most studied. There are two TAM pathways which show evidence of appearing in an ordered sequence; namely:

TAM-1 (41L, 210W, 215Y) and TAM-2 (67N, 70R, 215F, 219E/Q) (Cozzi-Lepri et al., 2005). TDF is one of the newer NRTIs replacing d4T in the South African national treatment plan in 2010 (Department of Health, 2014). Subtype B studies have reported a TDF resistance mutation pathway that includes K65R and nucleoside analogue mutations: Q151M, A62V, V75I, F77L and F116Y (Miller, 2004, Garforth et al., 2014b). However, the interaction between TDF associated mutations and other NRTI and NNRTI resistance mutations in subtype C sequences remains unclear.

There have been recent reports of rapid selection of K65R mutation in subtype C viruses in TDF-treated patients (Skhosana et al., 2015, Sunpath et al., 2012a). This has important public health implications in limited resource countries, such as South Africa, as resistance to the first line regimen might lead to switching to a more costly 2nd line regimen. In addition, the drug resistant viruses can be transmitted to newly infected individuals and reduce patient's treatment options when they start treatment. It is therefore important to monitor the development of drug resistance and cross-resistance in patients on TDF containing regimen to ensure optimal viral suppression and care of HIV-1 infected patients.

1.1 Human Immunodeficiency Virus

Human immunodeficiency virus (HIV-1) was discovered in 1983 and is known to be the leading cause of Acquired Immunodeficiency Syndrome (AIDS) (Barré-Sinoussi et al., 2013, Mauck and Straten, 2008). HIV-1 belongs to a group of retroviruses called Retroviridae (Engelman and Cherepanov, 2012b, Mauck and Straten, 2008). HIV-1 has a different structure from the other members of the family. It is roughly spherical, somewhat pleomorphic and measures around 120nm in diameter. HIV-1 consists of two single stranded RNA molecules wrapped in a conical capsid comprising of the p24 viral protein (Figure 1.1) (Inhibitors, 2002).

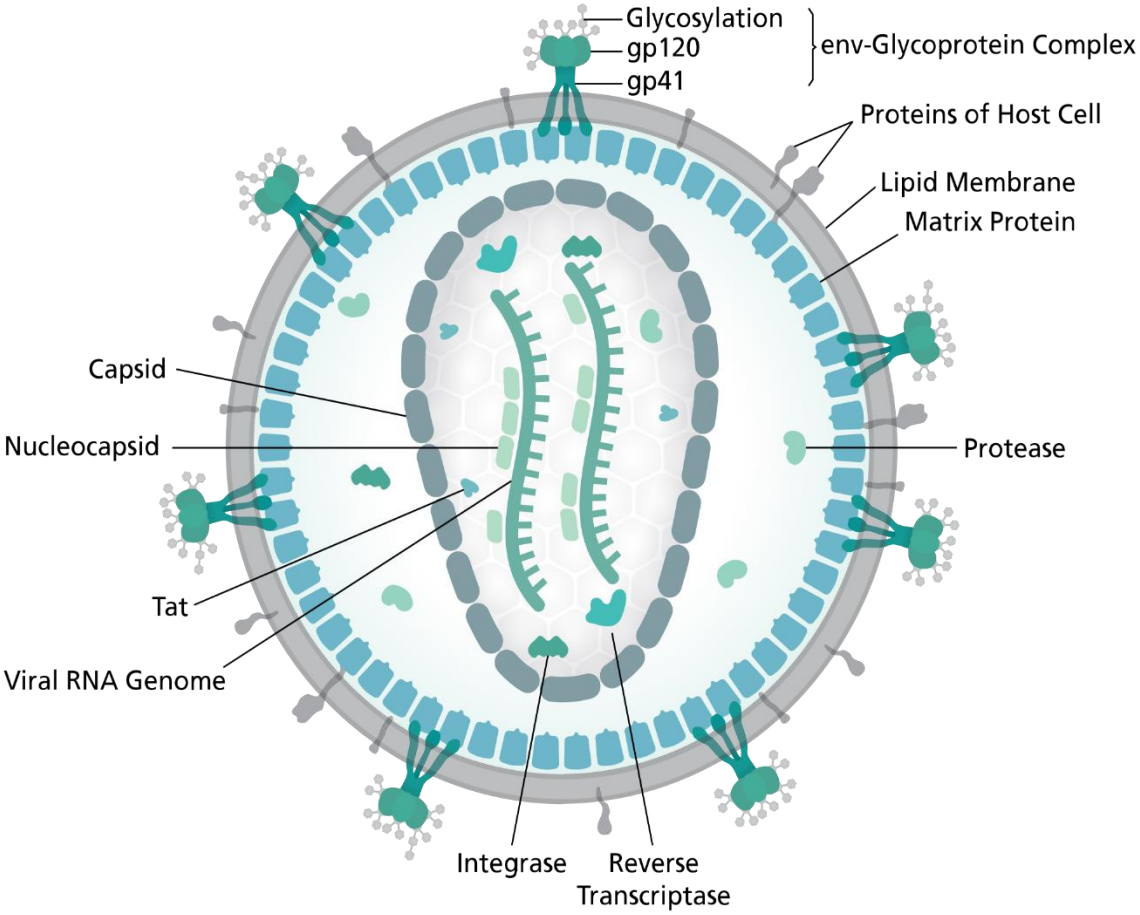

Figure 1.1: General features of HIV-1 virion. Adapted from (Freed, 1998).

The full HIV-1 genome is encoded on one long strand that contains a RNA genome that encodes 9 genes (Figure1.2). Within these nine genes, there are three major genes called gag (group-specific antigen), pol (polymerase) and env (envelope glycoprotein) (Freed, 2001)). Gag is a genomic region that codes for four structural proteins, which are matrix (p17), capsid (p24), nucleocapsid (p7) and p6

(Freed, 1998). The matrix is important for the incorporation of viral surface glycoproteins into virion and early post entry events (Dorfman et al., 1994, Freed, 1998). The capsid plays a crucial role in virus assembly and maturation (Freed, 1998)). The nucleocapsid is required for the encapsidation of the viral RNA, membrane binding, pre-integration and transcription (Accola et al., 1998, Freed, 2001). P6 is needed for the release of assembled virion from the cell surface (Kondo and Göttlinger, 1996). The Pol region codes for 3 enzymes which are protease (p11), reverse transcriptase (p66/p51) and Integrase (p32). The Env viral glycoprotein codes for surface GP120 and transmembrane GP41(Frankel and Young, 1998). The other six genes are vif, vpr, rev, vpu, tat and nef (Frankel and Young, 1998). Tat and rev are two regulatory genes for HIV-1 gene expression. Tat (p16 and p14) consists of 80 to 101 amino acids which are needed for transcription of the viral genome (Kuiken et al., 2008). Rev (p19) is required for export of RNA from the nucleus to the cytoplasm. It consists of 116 amino acids which are nuclear localization sequence (NLS) and nuclear export sequence (NES) (Freed, 2001, Pond et al., 2009). Vif, vpr,vpu and nef are not required for virus replication. However, they provide different effect in disease induction and the spread of virus(Frankel and Young, 1998, Mauck and Straten, 2008).

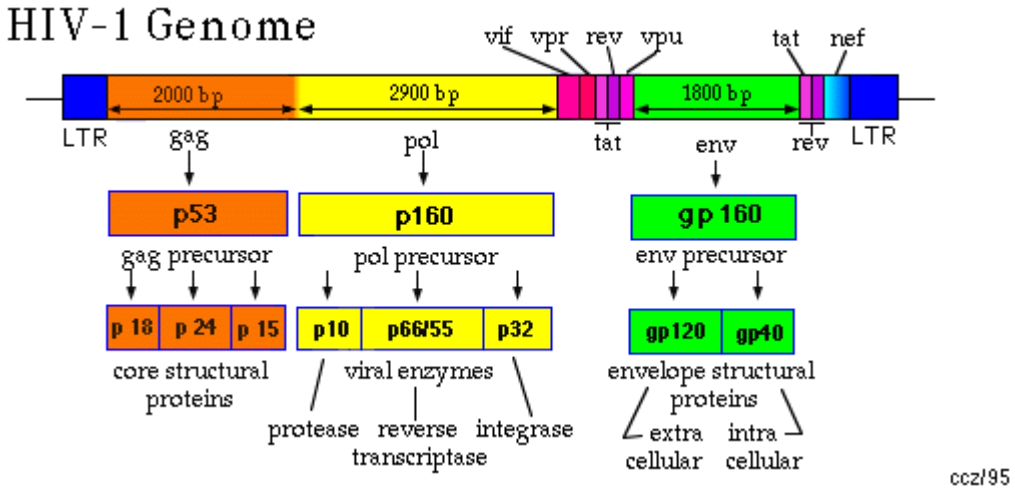

Figure1.2: HIV-1 full genome with three major genes that codes for different proteins and enzymes. Adapted from Jeff Huckaby et al., 1998

1.2 Classification of HIV-1 Strains

HIV-1 is divided into two related lentiviruses, HIV-1 and HIV-2 (Nyamweya et al., 2013). HIV-1 is divided into 4 major phylogenic groups: M, N O and P. Group M which is referred to as the main or major group, is responsible for the majority of the global HIV-1 epidemic (Kostrikis et al., 1995). Group M is also subdivided into 9 subtypes (A, B, C, D, F, G, H, J and K). Furthermore, closely related subtype can combine to form CRFs (circulating recombinant forms) (Hemelaar, 2012). Subtype B is dominant in America, Europe and Australia(Buonaguro et al., 2007), whereas subtype C is dominant in South Africa, horn of Africa and India (Lole et al., 1999).

1.3 HIV-1 Life Cycle

During the HIV-1 replication cycle, gp120 binds to a CD4 receptor on the surface of the target cell. Upon binding, the gp120 undergoes a conformational change resulting in binding to the co-receptors (Melikyan et al., 2000). After binding to receptors, HIV-1 envelope and the CD4 cell membrane fuses, and the virion single-stranded ribonucleic acid (ssRNA) is used as a template to synthesize double-stranded deoxyribonucleic acid (dsDNA) (Harrich and Hooker, 2002). Integrase then carries the DNA into the nuclease and integrates it into the host DNA (Figure 1.3) (Freed, 2001). RNA polymerase transcribes the integrated viral DNA into mRNA. After transcription, the mRNA is spliced and exported from the nucleus into the cytoplasm. In the cytoplasm, pieces of mRNA are translated into elongated chains of viral proteins and enzymes. On the cell surface, the viral RNA assembles with newly synthesized proteins and forms an immature virus (Engelman and Cherepanov, 2012a). Viral protease then cleaves the poly-proteins within the virion for the production of mature infectious virus particles (Marsden and Zack, 2013).

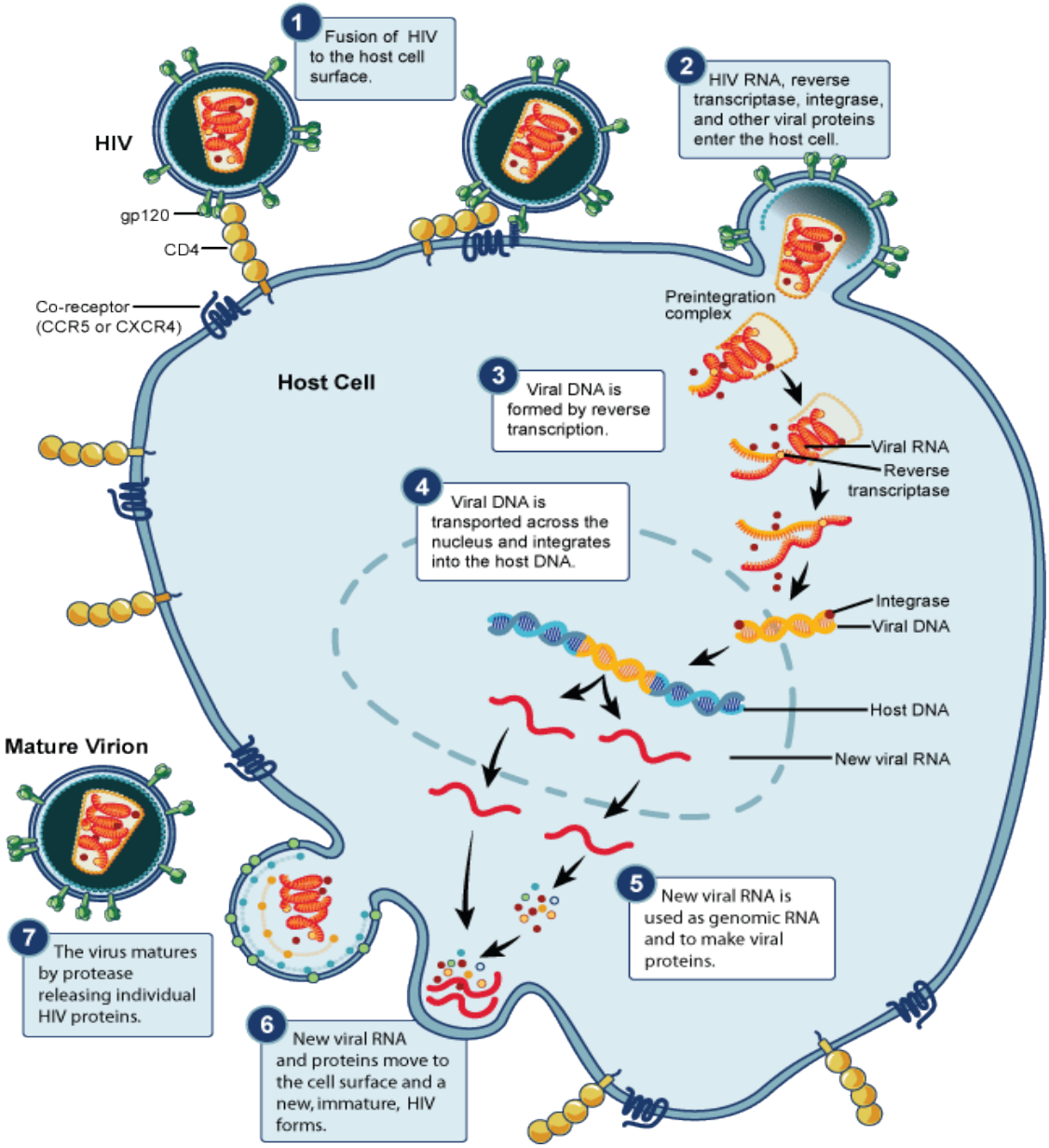

Figure 1.3: Diagrammatic representation of the HIV-1 life cycle. HIV-1 infects the CD4 T cells using gp120 of the envelope, the viral RNA is then released inside the host cells. HIV-1 RNA is reverse transcribed into cDNA which is then integrated into the host's DNA. Viral DNA is then synthesised into new viral RNA used to make viral proteins followed by protein translation inside the cytoplasm and a new, immature, HIV-1 virus is released and infects more cells (Monini et al., 2004).

1.4 HIV-1 Reverse Transcriptase

Reverse transcriptase is a viral enzyme encoded within the pol region of the HIV-1 genome. It mediates the conversion of the RNA genome to double-stranded DNA during reverse transcription. This enzyme is a heterodimer made up of a p66 and p51 subunit. HIV-1 RT performs multiple enzyme activities: RNA-dependent DNA polymerase, DNA-dependent DNA polymerase and RNAse H activities. The p66 subunit consists of 560 amino acid while p51 consists of only 440 amino acids and lacks the

Ribonuclease H (RNase H) domain that is formed by the additional 120 residues. However, both subunits contain similar subdomains in their polymerase domain which are described with reference to a human right hand (Hu and Hughes, 2012) namely the Finger, Thumb, Palm and Connection subdomains (Temiz and Bahar, 2002). The p66 polymerase domain has the active site in the palm subdomain that contains a catalytic triad in position D110, D185 and D186. P66 plays a catalytic role while p51 plays a structural role (Shafer, 2002)

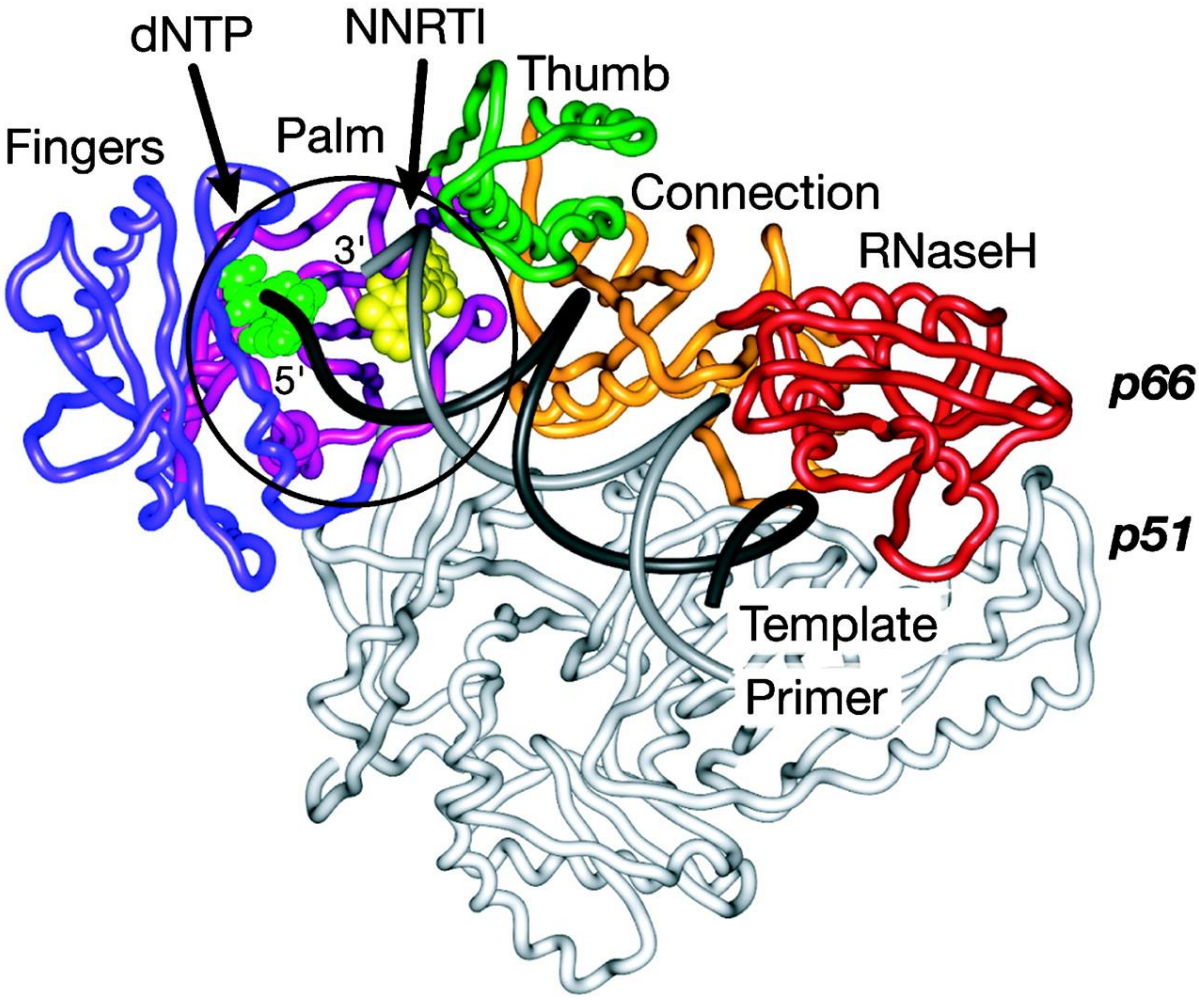

Figure 1.4: Reverse Transcriptase structure with two subunits p66 and p51. Adapted from (Huang et al., 1998).

1.5 Antiretroviral Therapy Drug Classes

Antiretroviral therapy has greatly altered the management of HIV-1 infection globally and has led to the reduction of mortality and morbidity rates in many countries that have successfully implemented a national ARV rollout programme. There are currently 6 different antiretroviral drug classes used in HIV-1 therapy (Shafer, 2012). These include nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), fusion inhibitors and inhibitors of co-receptors and integrase inhibitors. The efficacies of these drugs are limited by HIV-1 drug resistance, which is usually caused by mutations in protease and RT (Shafer, 2002).

1.6 The Nucleoside Reverse Transcriptase Inhibitors (NRTIs) mechanism of action

The NRTIs are used in the first-line regimens of most countries worldwide, these include: Abacavir (ABC), Didanosine (ddI), Zalcitabine (ddC), Stavudine (d4T), Zidovudine (AZT), Tenofovir (TDF), Emtricitibine (FTC), and Lamivudine (3TC) (AIDSinfo, 2013). NRTIs inhibit HIV-1 replication by acting as DNA chain terminators during reverse transcription of the HIV-1 single-stranded RNA genome into complimentary DNA (cDNA). NRTIs are first phosphorylated to an active triphosphate form of the drug that competes with the cells naturally produced deoxynucleoside triphosphate (dNTP) pools (Shafer, 2002). Once incorporated into a growing cDNA strand, it acts as a chain terminator by blocking further insertion of another dNTP molecule, as NRTIs lack a 3’ hydroxyl group that is crucial for the formation of a phosphodiester bond during cDNA synthesis by the HIV-1 RT enzyme (Das and Arnold, 2013).

1.6.1 Molecular mechanisms of NRTI resistance

Molecular resistance with NRTIs is the result of the accumulation of mutations in HIV-1 *pol* gene (Johnson et al., 2003). These mutations can be classified as discriminatory or primer unblocking mutations (Tang and Shafer, 2012). Discriminatory mutations have been shown to decrease the binding of the NRTIs to the RT enzyme, allowing the RT enzyme to discriminate against NRTI incorporation onto the growing DNA chain during polymerisation (Whitcomb et al., 2003). These mutations are also linked to reduced catalytic activity of the HIV-1 RT enzyme while the insertion of dNTPs remains unchanged, causing the virus to continue replicating (Whitcomb et al., 2003). The other mechanism involve the phosphorolytic activities, whereby after NRTI incorporation into the viral DNA, chain terminating NRTI is removed from the 3’- terminus of the primer. However, this may be affected by

NRTI mutation that enhance primer unblocking activities e.g. TAMs (M41L, D67N, K70R, L210W, T215F/Y and K219Q/E (Marcelin et al., 2006), which promote the hydrolytic removal of NRTIs, leaving a free 3'-OH group to continue DNA synthesis. During this process, they also unblock the primer stimulating further extension of viral DNA (Whitcomb et al., 2003).

1.7 The Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) mechanism of action

NNRTIs are non-competitive inhibitors that bind to hydrophobic pocket close to the active site in the enzyme. The binding to the less well-conserved site leads to conformational changes that interfere with the formation of the binding pocket leading to a reduced RT activity (Shafer, 2002). There are four NNRTIs in clinical use: Navarapine (NVP), Efavirenz (EFV), Etravirine (ETR) and Rilpivirine (RPV) (AIDSinfo, 2013). A major limitation of NNRTIs is their relatively low genetic barrier, where only a single point mutation is required to cause resistance (Tang and Shafer, 2012). High level resistance to NNRTIs may results from single mutation (Shafer, 2002). K101E/P, K103N and E138K are present at the rim of pocket and rest other NNRTI resistance mutations are present within the NNRTI binding pocket (Ren and Stammers, 2008). A previous study suggests that there are broad classes of NNRTI resistance mechanisms: loss of an aromatic ring, steric hindrance and alteration of hydrophobic interactions (Das and Arnold, 2013).

1.8 South African treatment programme and TDF

In 2010, the South African national antiretroviral treatment program adopted a policy to phase out the more toxic d4T and replace it with TDF as part of the first-line regimen in South Africa (Department of Health, 2014). It is a widely prescribed antiretroviral that blocks RT activity (Menéndez-Arias, 2010). It is usually administered in combination with other NRTIs and NNRTIs (Nelson et al., 2007). The K65R mutation has been reported to decrease efficacy of TDF by four-fold (Margot et al., 2003)

1.9 Resistance to TDF

Initially, low frquency (2.1%) of resistance was reported in subtype B TDF-treated patients (Margot et al., 2003). In vitro studies identified K65R as the only mutation associated with TDF resistance. However, this single mutation causes high level TDF resistance (Wensing et al., 2015). Subtype B study has reported the TDF-resistance pathway including association of K65R with S68G, L100I and Y181C

(Theys et al., 2009). K65R still remain the main cause of TDF resistance, and is known to be preferentially selected in subtype C (Brenner and Coutsinos, 2009a). Studies reported a rapid selection of K65R in subtype C TDF-treated patient (Skhosana et al., 2015, Sunpath et al., 2012b). This has important public health implications in limited resource countries, such as South Africa, as resistance to first line regimens might lead to switching to more costly 2nd line regimens. In addition, the drug resistant virus can be transmitted to newly infected individuals and reduce their options when they start treatment. It is therefore important to monitor the development of drug resistance and cross-resistance in patients on TDF containing regimens to ensure optimal viral suppression and care of HIV-1 infected patients.

1.10 K65R mechanism of resistance

The K65R mutation is the substitution of lysine (K) to arginine (R) amino acids at residue 65 in the catalytic site of HIV-1 RT. K6R is one of the discrimination mutations that confers intermediate resistance to ABC, ddI, d4T, 3TC and FTC; causes high resistance to TDF and increase susceptibility to AZT (Garforth et al., 2014a). The Lysine (K) 65 which forms a part of nucleotide binding protein is located between the β3 and β4 sheets that form the flexible loop in the finger region of HIV-1 RT structure (Garforth et al., 2014a). For the incoming dNTP or NRTI-TP to be incorporated into a growing DNA strand, the finger subdomain of HIV-1 RT-structure folds towards the primer- template and the polymerase active site in the palm subdomain. Subsequent to this a salt bridge is formed between codon K65 and the γ phosphate. In the presence of K65R the dNTP can still bind, but alters the nucleotide binding specificity (Huang et al., 1998, Sarafianos et al., 1999). The mechanism for rapid selection may be explained by the nucleotide sequence dissimilarities that exist in the subtype specific template (Invernizzi et al., 2009). Subtype C sequences have unique nucleotide polymorphisms in position 64 (AAA) 65 (AAG) and 66 (AAG) when compared to non-subtype C sequences, including subtype B (Invernizzi et al., 2009). In addition, Subtype C have the homopolymeric stretch that starts from codon 63 and end at codon 65 and followed by a C base (Brenner and Coutsinos, 2009b). During the synthesise of dsDNA from ssDNA template, the enzyme approaches the homopolymeric end at codon 65 and this leads to misalignment and slippage. Therefore, the tamplate folds onto itself, covering the final T and exposing C at codon 65. The incoming nucleotide provides the stability of the misaligned template through hydrogen bond, resulting in correct binding of G to C on the misaligned template. Thereafter, the primer and the template strands are re-aligned and this leads to the binding of G to T and the re-exposed of C on the aligned template and a second G being incorporated to bind with C base. Finally, DNA synthesis continues normally (Coutsinos et al., 2011).

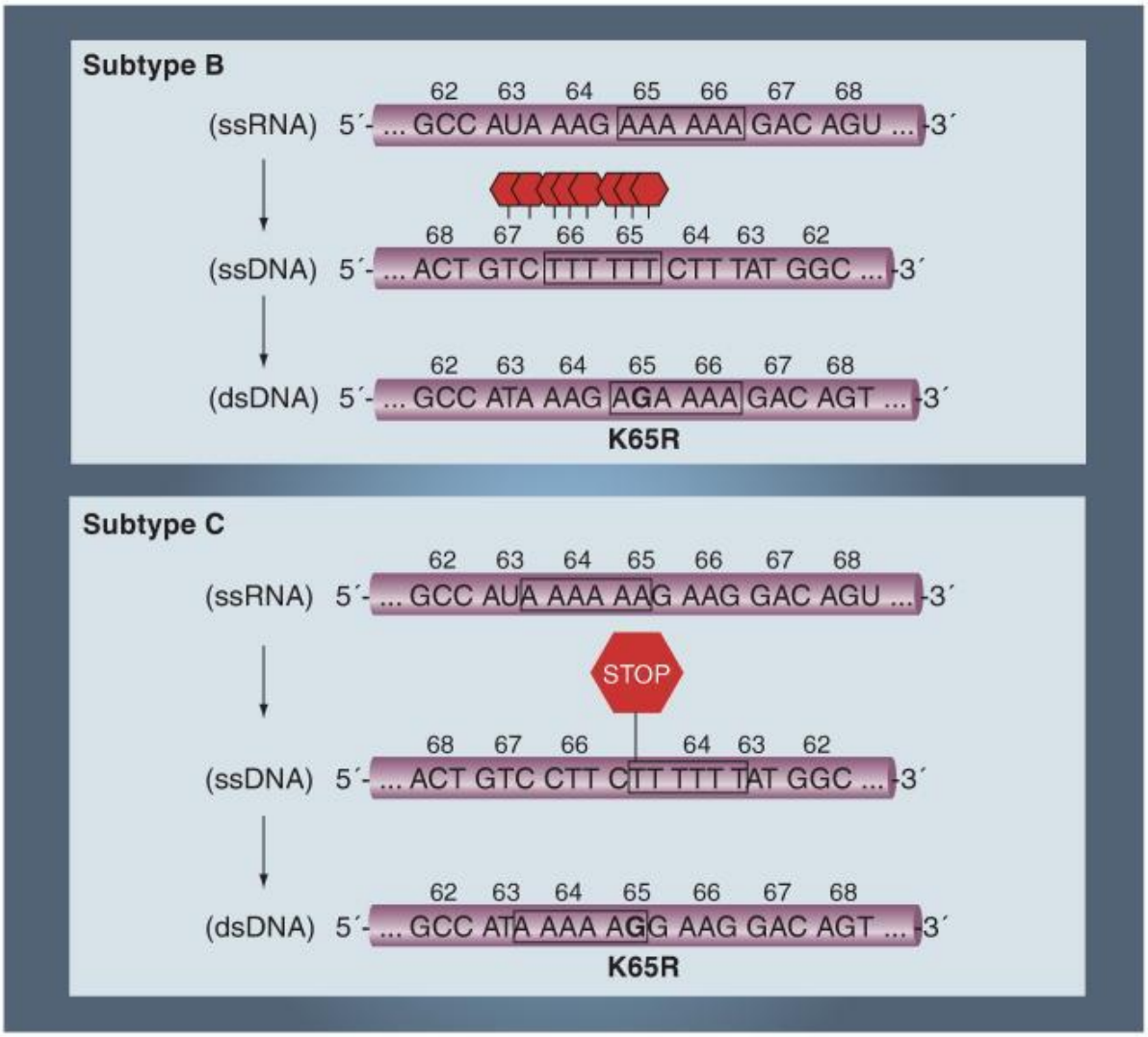

Figure 1.5: Increased selection of K65R in subtype C. The selection of K65R is initiated by the poly-adenine stretches at position 63 and 65. The stop occur in poly stretches in position 65 (Brenner and Coutsinos, 2009b).

1.10.1 K65R mutation in subtype C patients

In-vitro studies have shown an increased risk in developing the K65R mutation in subtype C infections compared to subtype B (Brenner and Coutsinos, 2009a). Data regarding the prevalence of the K65R mutation in TDF exposed patients are mixed, with two studies (Skhosana et al., 2015, Sunpath et al., 2012a) showing a prevalence of more than 65% while two studies (Van Zyl et al., 2013a, Hoffmann et al., 2013b) show K65R prevalence of 46% and 23% respectively. Furthermore, it has been found that the K65R mutation leads to cross-resistance to the majority of NRTIs {including abacavir (ABC), emtricitabine (FTC), lamivudine (3TC), stavudine (d4T)} (Johnson et al., 2011). This could lead to more failures to the first line regimen in South Africa, since K65R causes resistance to TDF, leaving 2

drugs active (3TC/FTC and EFV). Since the emergence of K65R has been reported in patients receiving EFV or NVP (von Wyl et al., 2008), this implies that all active drugs in the first line regimen will eventually fail.

1.10.2 Antagonistic relationship between K65R and other NRTI-resistance mutation

Specific combinations of NRTIs cause HIV-1 to evolve in a pathway of resistance mediated either by TAMs or K65R but not both (White et al., 2006b). The mechanism by which TAMs (41L, 67N, 70R, 210W, 215Y/F, 219E/Q) confer resistance is by facilitating the entry and binding of either pyrophosphate (PPi) or adenosine triphosphate (ATP) to the NRTI that has been incorporated at the 3' end of the terminated DNA chain (Arion et al., 1998, Boyer et al., 2001, Meyer et al., 1998). The triphosphate intermediate allows the inhibitor to be excised from the DNA chain (Arion et al., 1998, Bishop et al., 2008, Meyer et al., 1998), thereby allowing DNA synthesis to resume. The ability of TAMs to efficiently excise NRTIs can become increasingly impaired with the accumulation of other mutations (Clavel and Hance, 2004), such as the discriminatory M184V mutation (Gu et al., 1995). M184V decreases the binding of 3TC and FTC to the RT enzyme by interacting with the oxathiolane ring, and therefore increases sensitivity to AZT by reducing the removal of the chain-terminating nucleoside (Huang et al., 1998). This can be explained by the fitness constraints imposed by M184V which diminishes the ability of RT to effectively remove the drug (Gu et al., 1995). Therefore M184V has an antagonistic effect on TAMS, Similarly, K65R and TAMs are also mutually antagonistic and rarely seen in the same viral genome (Gu et al., 1995).

K65R further increase drug resistance to ddI, ABC, TDF, 3TC by a discrimination mechanism (Boyer et al., 1994, Deval et al., 2004). This mutation is able to disrupt the hydrogen bonding that involves the 3'-OH of the incoming deoxynucleotide (dNTP) (Boyer et al., 1994, Deval et al., 2004). It favours the direct contact of the natural dNTPs and indirectly negatively influences the binding of NRTI analogue by repositioning the template/primer binding site (Boyer et al., 1994, Deval et al., 2004).

1.10.3 Effect of K65R and other mutations

As already stated, K65R and M184V increases discrimination between the natural nucleotide and the NRTIs and impairs its incorporation (Deval et al., 2004). While K65R disrupts the interaction with the γ phosphate of the incoming nucleotide, M184V interacts with the oxathiolane ring, allowing the discrimination between the incoming dNTPs and the NRTI triphosphates (Huang et al., 1998). M184V

is selected in the presence of 3TC and FTC. K65R and M184V single mutants are known to have intermediate resistance to ddC, ddI, ABC, 3TC; but the combination of both K65R and M184V is reported to increase resistance to these drugs. In contrast, when M184V occurs on the same genome as K65R, their combination may have a synergistic resistance effect on 3TC (Gu et al., 1995). Also, this combination was reported to retain susceptibility to TDF (Das et al., 2009, Deval et al., 2004, Parikh et al., 2006). The presence of M184V together with K65R increases the binding of the natural substrate, i.e. the dNTP, compared to the inhibitor (White et al., 2002).

Y115F is selected in patients taking ABC (Miller et al., 2000). However, this mutation has also been reported to be selected together with K65R in patients taking TDF and never exposed to ABC (Skhosana et al., 2015). The combination of these two mutations had previously been found to cause resistance to TDF (Stone et al., 2004).

1.11 Computational methods of HIV-1 drug resistance

Recently, various computational methods have become an important part of drug resistance surveillance, and can offer detailed information about the interaction of mutations (Kirchmair et al., 2011). The Bayesian method uses the data to refine prior knowledge into posterior knowledge which is expressed as a probability distribution of models (Deforche et al., 2006). Also various modelling methods such as 3D homology modelling, molecular docking and molecular dynamics, are used to understand the implications of drug resistance in RT (Das et al., 2009, Huang et al., 1998).

1.11.1 Bayesian Network

A Bayesian Network (BN) is a probabilistic model represented by direct acyclic graphs that describes statistical correlation between multiple variables (Pearl, 1998). A BN is learned from data by searching for the most credible network structure that explains casual and cause-effective relationships from data using a minimum number of arcs (direct connections between variables), and eventually produces an acyclic causality graph (Deforche et al., 2006). BNs can also be learned from the observation of variables (nodes), by searching for the network with the maximum number of correlations in the data using a minimum number of arcs (Heckerman et al., 1999). The acyclic structures of a BN are also capable of representing relationships between the variables through direct conditional or unconditional dependencies (Figure 1.6) (Myllymäki et al., 2002). Dependencies are represented by the presence of an arc from one variable to another, showing dependencies between all variables in the data (Myllymäki et al., 2002). Conditional dependency is represented by the lack of an arc, while unconditional

dependency is represented by the arc (Deforche et al., 2006) For example, if a dependency model has four variables A, B, C and D, it could be analyzed as follows (Myllymäki et al., 2002):

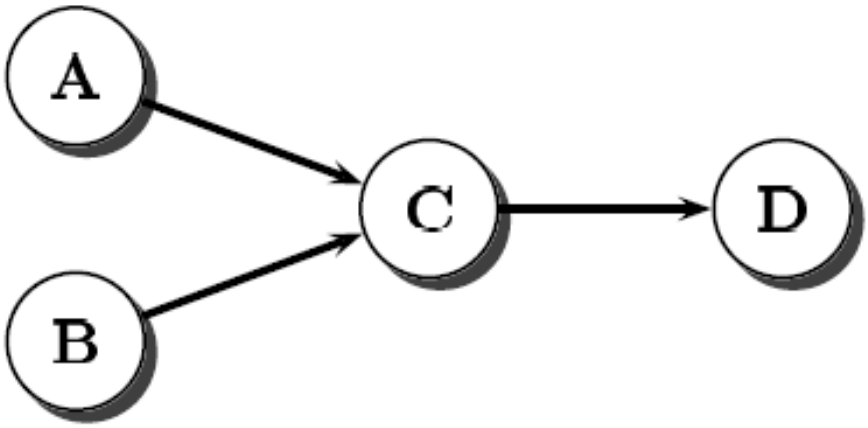

Figure 1.6: A Bayesian Network representing conditional and unconditional dependencies(Myllymäki et al., 2002).

- "A and B are conditionally dependent on each other, whether values of C and D are known".
- "A and C are unconditionally dependent on each other even whether values of both B and D are known or not",
- "B and C are unconditionally dependent on each other even whether values of both A and D are known or not",
- "C and D are unconditionally dependent on each other even whether values of A and B are known or not".

After the most credible BN network structure has been found, assessments of the network robustness are searched using bootstrapping (Deforche et al., 2008). One-hundred replicates of non-parametric bootstraps are performed to derive network robustness by the presence or absence of a particular arc (Deforche et al., 2006). According to (Myllymäki et al., 2002, Pearl, 1998), the existence and thickness of arcs showing a direct influence amongst the corresponding variables are relative to bootstrap values and their importance are coloured according to the arc weight. For example in Figure 1.7, the black arc indicates direct influence between resistance mutations, while association between background

polymorphisms is shown in green arcs and the blue arcs indicate an influence from background polymorphisms on drug resistance associated mutations. Combination between the NNRTI and NRTI associated mutations are shown in grey and purple arcs, which show a direct dependency between treatments reflecting preference in treatment combinations. Dotted arcs represent a bootstrap of 35% or more and solid arcs represent bootstraps over 65% (Deforche et al., 2006). The quality of the candidate BN may be assessed using the posterior Bayesian probability score of the network, that is often used for model selection; log p(D,S^h)= log p(S^h) + log p (D\$S^h$ShSh). The parameters are defined as D= a data set of cases; S^h= the hypothesis corresponding to network structure S and p= probability(Heckerman et al., 1995).

BNs have been successfully used to investigate the interactions of mutations that confer resistance to NRTIs and NNRTIs (Deforche et al., 2008, Theys et al., 2009). BNs were also used to gain insight into the pathway to nelfinavir resistance, by identifying the role of mutations selected during treatment (Deforche et al., 2006).

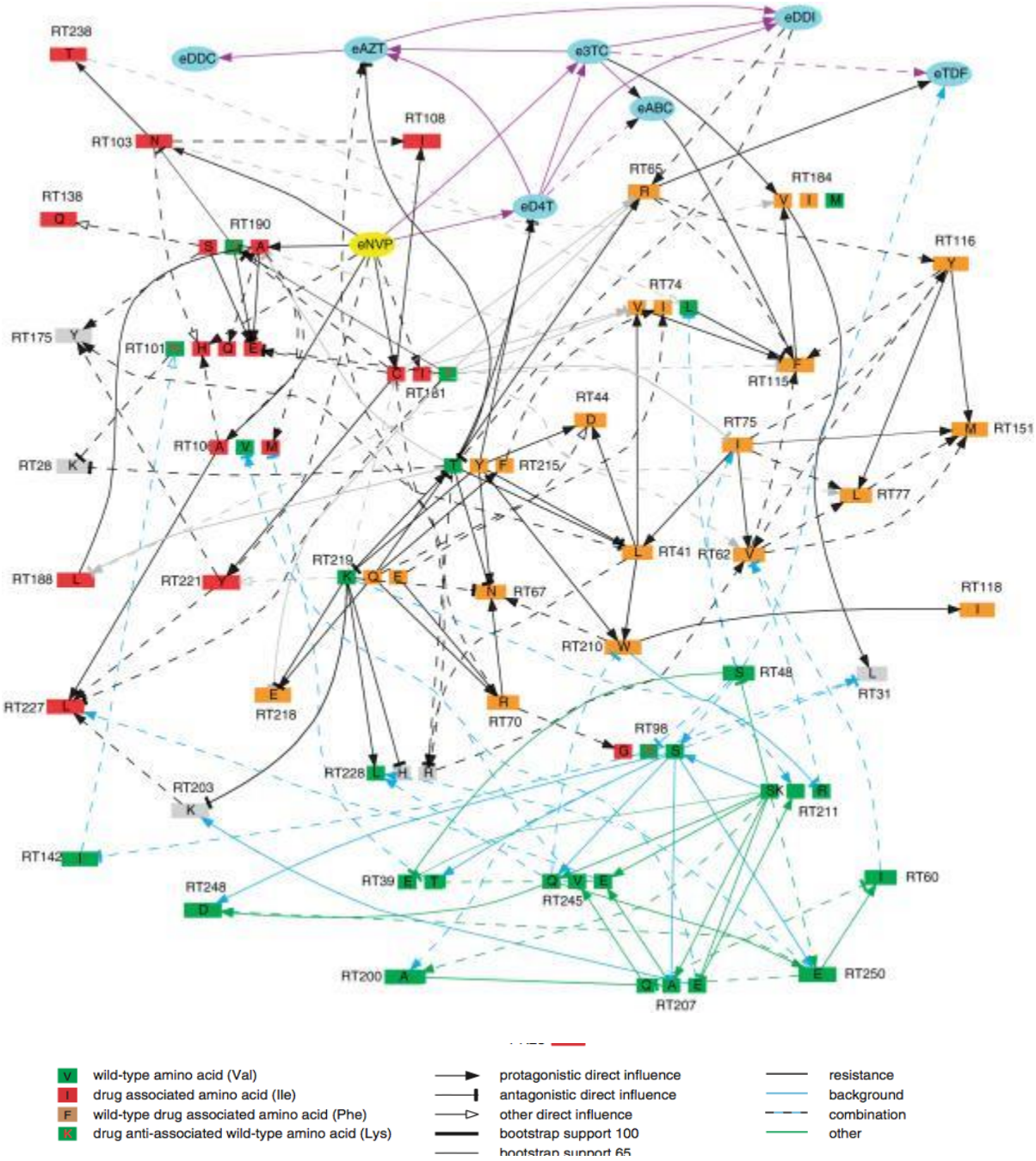

Figure 1.7: Annotated nelfinavir experience Bayesian network expressing direct association between nelfinavir-associated mutations, polymorphisms and nelfinavir treatment (eNFV) (Deforche et al., 2008).

1.11.2 3D Homology Modelling using SWISS-MODEL

Homology modelling is a computational technique, used to determine the three dimensional (3D) structure of proteins. High resolution protein structures that are available in public databases are used to construct a model of a protein of similar, but unknown structure (Tastan Bishop et al., 2008). This provides an understanding of the biological behaviour and biochemical function of uncharacterised sequences, brought about by the recognition of homology between protein sequences and known structures (Shi et al., 2001). SWISS-MODEL is a widely used automated homology technique for protein modelling. It is designed to work with only the amino acid sequence of the submitted target protein and uses a known characterized protein structure found in the Protein Data Bank (PDB) as the template (Schwede et al., 2003).

Homology modelling can be divided into four steps, as shown in Figure 1.8: (1) Template selection: identification of the correct templates from the PDB, which are necessarily similar to the target sequence to be modelled. (2) Alignment: creating an alignment of the target sequence with the template structures; (3) Model building;construction of the 3D structure using specific modelling software, eg modeller v9. Templates are weighted by their similarity with the target sequences, and (4) Evaluation:evaluate the model (Tastan Bishop et al., 2008).

1.11.3 Viewing Structures

UCSF CHIMERA is used for visualization of the structural modelling results (Pettersen et al., 2004). CHEMLAB molecular viewer (Lanaro, 2002) or PyMOL molecular graphics system 2002 (DeLano, 2002) are amongst the many that can be used.

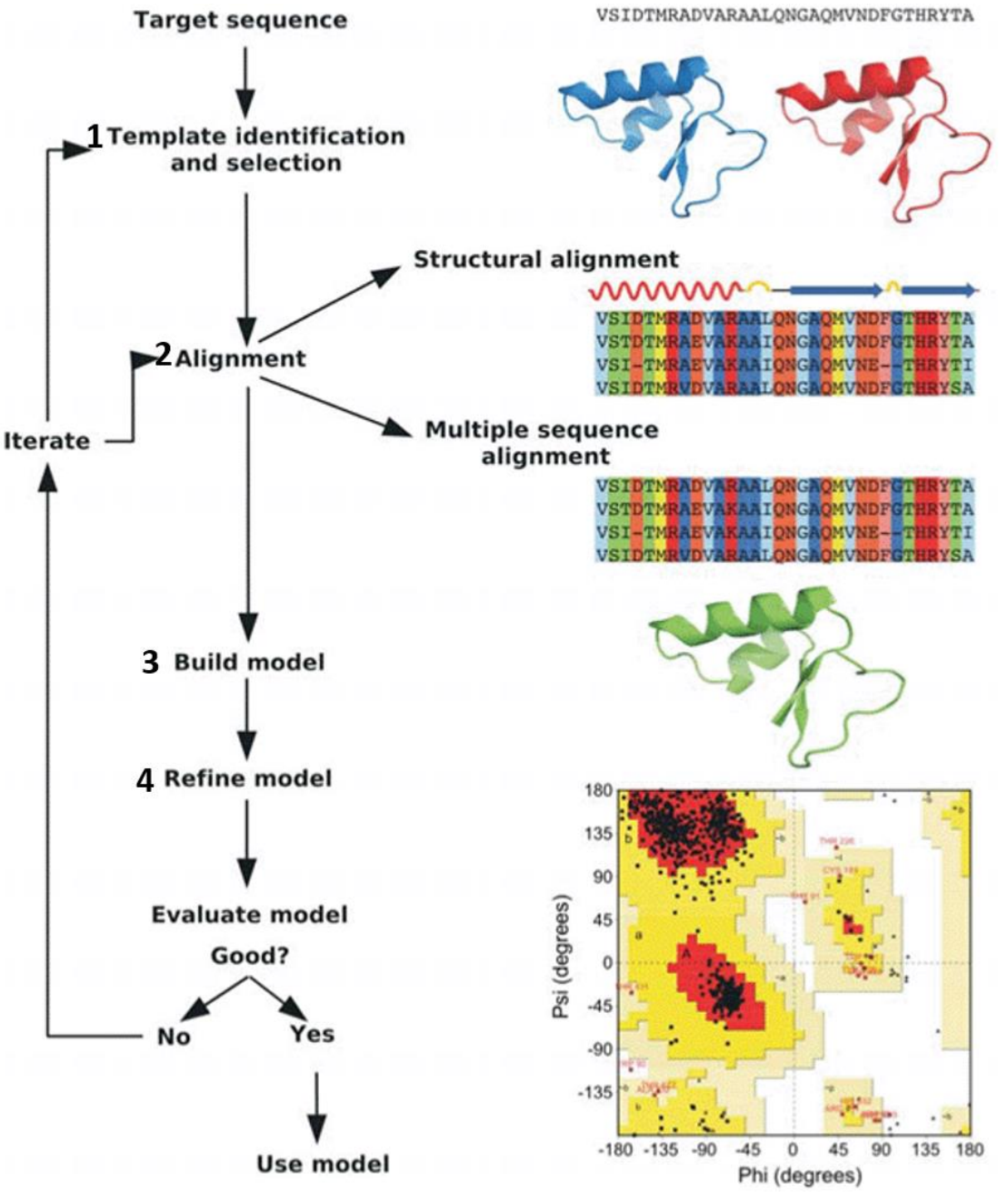

Figure 1.8: Illustration of the homology modelling. The steps involve the alignment of homologous sequences and adjustment of the alignment using known motifs and conserved features from secondary structure information. This is followed by submission to SWISSPROT for homology modelling and viewed using the Deep View software.

1.12 Project Aims and objectives

Aim

a) To investigate mutational pathways to Tenofovir resistance using a Bayesian Network.

Objectives

a) To download subtype B and C sequences from treatment-naïve, TDF exposed and tenofovir-naïve isolates(d4T)
b) To determine the frequency of Tenofovir resistance mutations in both subtype B and C sequences in naïve vs treated patients.
c) To investigate the interaction of K65R mutation and other NRTI and NNRTI mutations in TDF resistance in subtype B and C using Bayesian Network model.
d) To demonstrate the interaction of K65R and other mutations using 3D homology modelling

Hypothesis

- We hypothesize a Tenofovir-associated resistance pathway unique to subtype C.

Chapter 2: Prevalence of Tenofovir Resistance Mutations in both Subtype B and C Sequences.

2.1 Introduction

In 2012, the South African national treatment rollout program increased access to antiretroviral treatment for people living with HIV-1 infected, with more than 2,6 million people receiving highly active antiretroviral therapy (HAART) in the public sector alone (UNAIDS, 2013, Department of Health, 2014). South African national treatment guidelines includes a first-line regimen of 2 NRTIs and 1 non-nucleoside reverse transcriptase inhibitors (NNRTIs) and a second-line regimen of 2 NRTIs and 1 protease inhibitors (Department of Health, 2014).

Tenofovir is a widely-prescribed antiretroviral drug and a preferred component of first-line regimens during antiretroviral therapy world-wide (Menéndez-Arias, 2010). However, the efficacy of this drug is limited by the emergence of the K65R mutation (Department of Health, 2014, Sunpath et al., 2012). Several subtype C studies showed high rates of K65R selection in TDF exposed patients, with two studies showing a prevalence of more than 65% (Skhosana et al., 2015, Sunpath et al., 2012a) while two studies showed a prevalence of 46% and 23% respectively (Van Zyl et al., 2013a, Hoffmann et al., 2013b). K65R was significantly associated with A62V and S68G mutations, which in combination increased viral fitness (Svarovskaia et al., 2008). A previous subtype B study reported a TDF mutation pattern that included the TAMs (M41L and L210W) in patients who were previously exposed to AZT and d4T (Antinori et al., 2006). Furthermore, Eteibag reported 12% of TDF patients with the K65R that developed TAMs (Etiebet et al., 2013).Since TDF was introduced in South Africa in 2010, increasing reports of drug resistance make it necessary to monitor the TDF resistance patterns that develop under this drug, especially with it's prophylactic use as a microbicide gel (Karim et al., 2010).

The aim of this chapter was to determine the frequency of TDF-associated mutations in sequences from TDF exposed versus TDF-naïve isolates. In addition, the frequency of TDF-associated mutations found in patients treated with d4T were investigated.

2.2 Methods

2.2.1 Study design

This was a retrospective, descriptive study. All subtype B and C naïve sequences were downloaded from the Los Alamos HIV-1 database (http://www.HIV-1.lanl.gov),while subtype B and C treated isolates were downloaded from the Stanford HIV-1 database (http://HIV-1db.stanford.edu/), Briefly, the Genotype-Rx application was used to search for any subtype C (or B) isolate from a patient receiving TDF. Similarly, those receiving d4T were queried and downloaded. Treatment history was also included. Duplicate sequences were removed using Elimdupes available at (http:// www.HIV-1.lanl.gov).

2.2.2 Sequence analysis

Multiple sequence alignments were generated using the ClustalX program (http://www.clustal.org/) (Larkin et al., 2007) and manually edited using BIOEDIT (Ibis Biosciences, An Abbott Company, CA. USA). Sequences were also analysed using the RegaDB sequence analysis tools to identify the drug resistance and natural occurring mutations in the RT region (https://**rega**.kuleuven.be/cev/**regadb**).

2.2.3 Statistical analysis

All statistics was performed using GraphPad Prism® 5 software (Graphpad© Software, Inc) and IBM SPSS Statistics 21(SPSS inc., Chicago, Illinois,USA). Chi-square analysis (Fisher's exact test) with a two tailed p-value was performed using Graphpad Prism® 5 to calculate and compare the frequency of mutations in subtype B and C viruses from: (1) TDF-treated patients vs naïve patients; (2) d4T-treated vs naïve patients. Spearman's rank correlation (nonparametric correlation) in SPSS was used to explore the association between K65R and other NRTI and NNRTI mutations occurring in the presence of TDF. A p-value of <0.05 was considered statistically significant.

2.3 Results

A total of 4313 subtype C isolates were downloaded. These included 618 isolates from TDF-treated patients (irrespective of prior treatment with d4T), 2597 d4T experienced and 1098 naïve sequences. After removing those that were previously treated with d4T, 475 TDF only treated patients were analysed. In addition, 2585 TDF-treated, 4030 d4T experienced and 1630 treatment naïve subtype B sequences were downloaded and analysed. Duplicates were removed in both TDF and d4T treated sequences; the final analysis was done using 2585 TDF-treated and 4030 d4T experience (Figure 2.1).

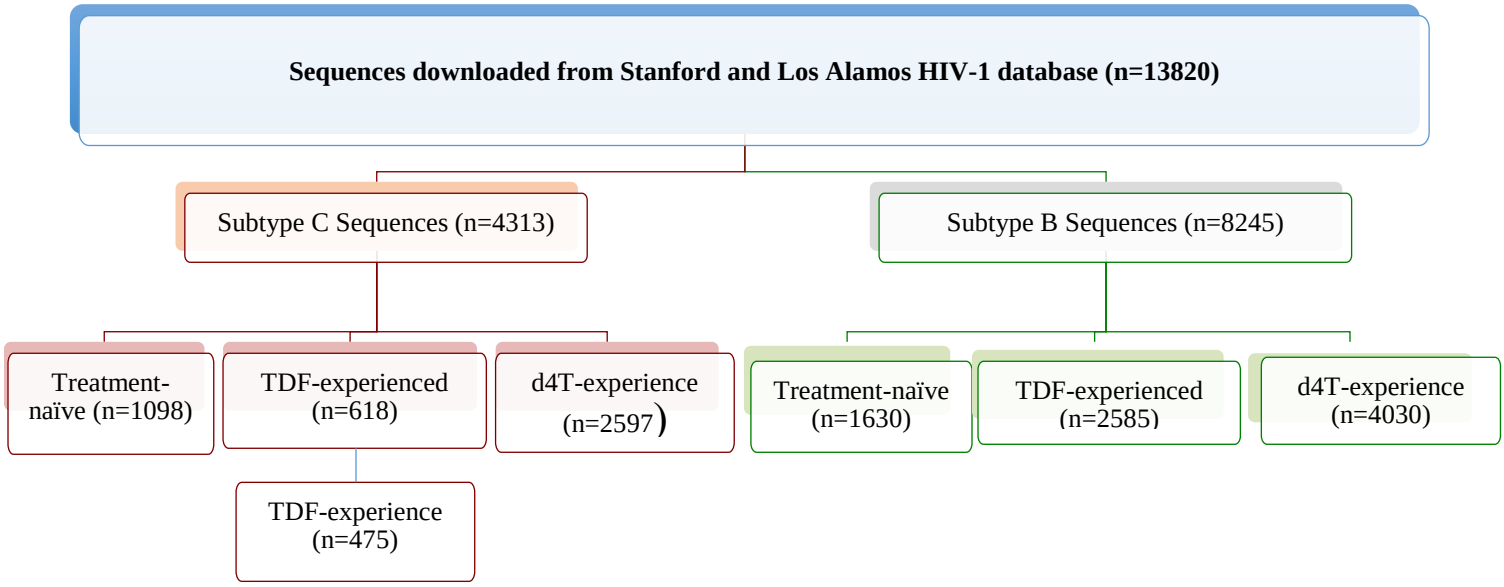

Figure 2.1: Number of downloaded sequences from both subtype B and C

2.3.1 Subtype C TDF-treated vs naïve

Initially, the frequency of mutations in sequences from any patient treated with TDF (irrespective of prior treatment with d4T) was compared to sequences from treatment naïve patients (as shown in Figure 2.2). Significant increases in frequency (p-value <0.05) are indicated in bold. Interestingly, six of these mutations (E36A, E44D, S48T, A158S, K166R and L214F) showed a significant decrease in frequency (*p*-value <0.05). TAMs were detected in the treated isolates, which may be because some sequences were from patients who took d4T prior to TDF. The most frequent drug resistance mutations were M184V (47%), followed by K65R (24%) and two NNRTIs: K103NRS (35%) and V106M (30%), respectively.

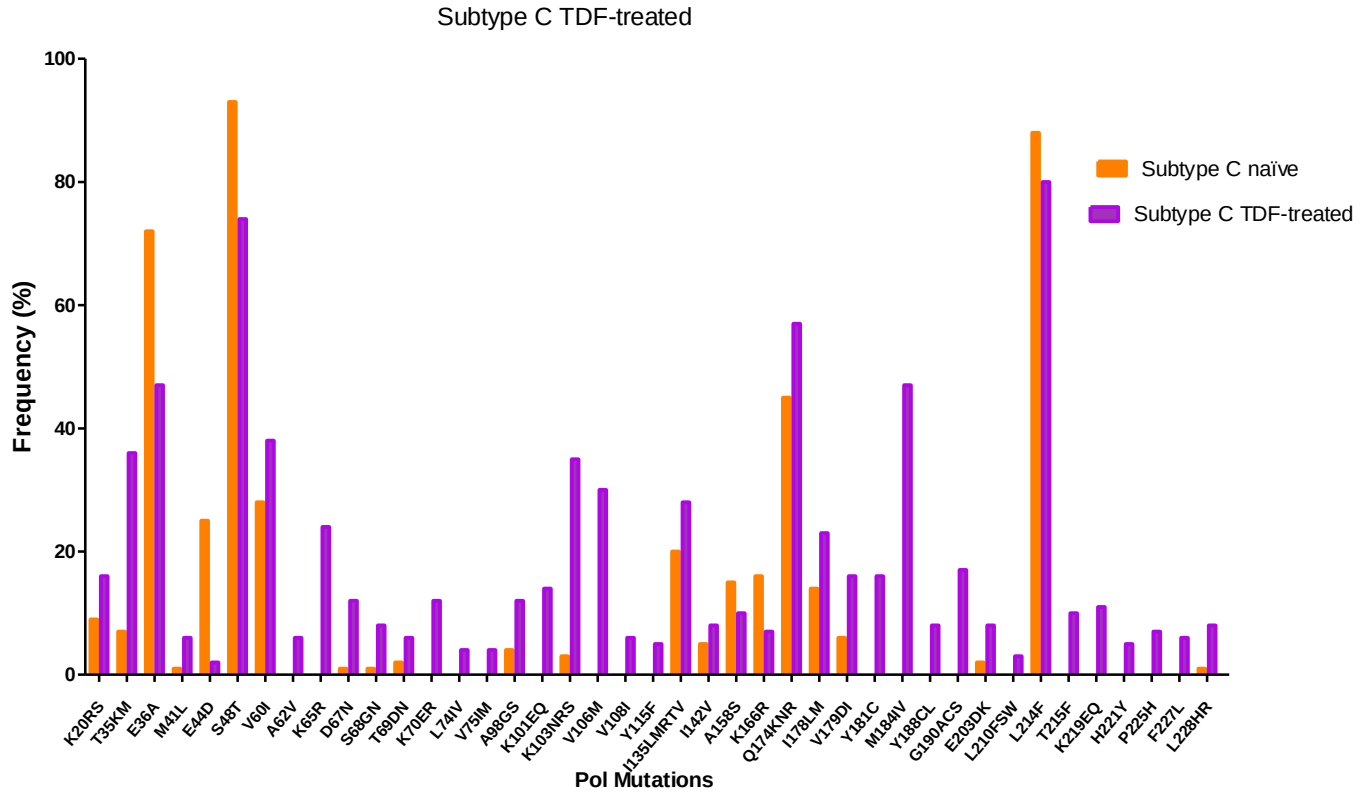

Figure 2.2: Significant prevalence of subtype C TDF-treated vs subtype C naïve

TDF-treated sequences were then grouped according to their treatment history: 1) TDF-treated without a history of d4T treatment (Figure 2.2A); 2) d4T-treated prior to TDF (Figure 2.2B). This was to determine whether the presence of TAMs prevented the emergence of K65R in patients switched to TDF. While the frequency of K65R was higher (31%) in those treated with TDF only, K65R was still prevalent (17%) in those switched from d4T to TDF. K46Q was found only in the TDF-treated group in 2% (8/475) of sequences and all sequences that developed K46Q also developed K65R. As expected, the frequency of TAMs were much higher in the d4T treated group compared to the TDF group, as d4T is known to select for TAMs.

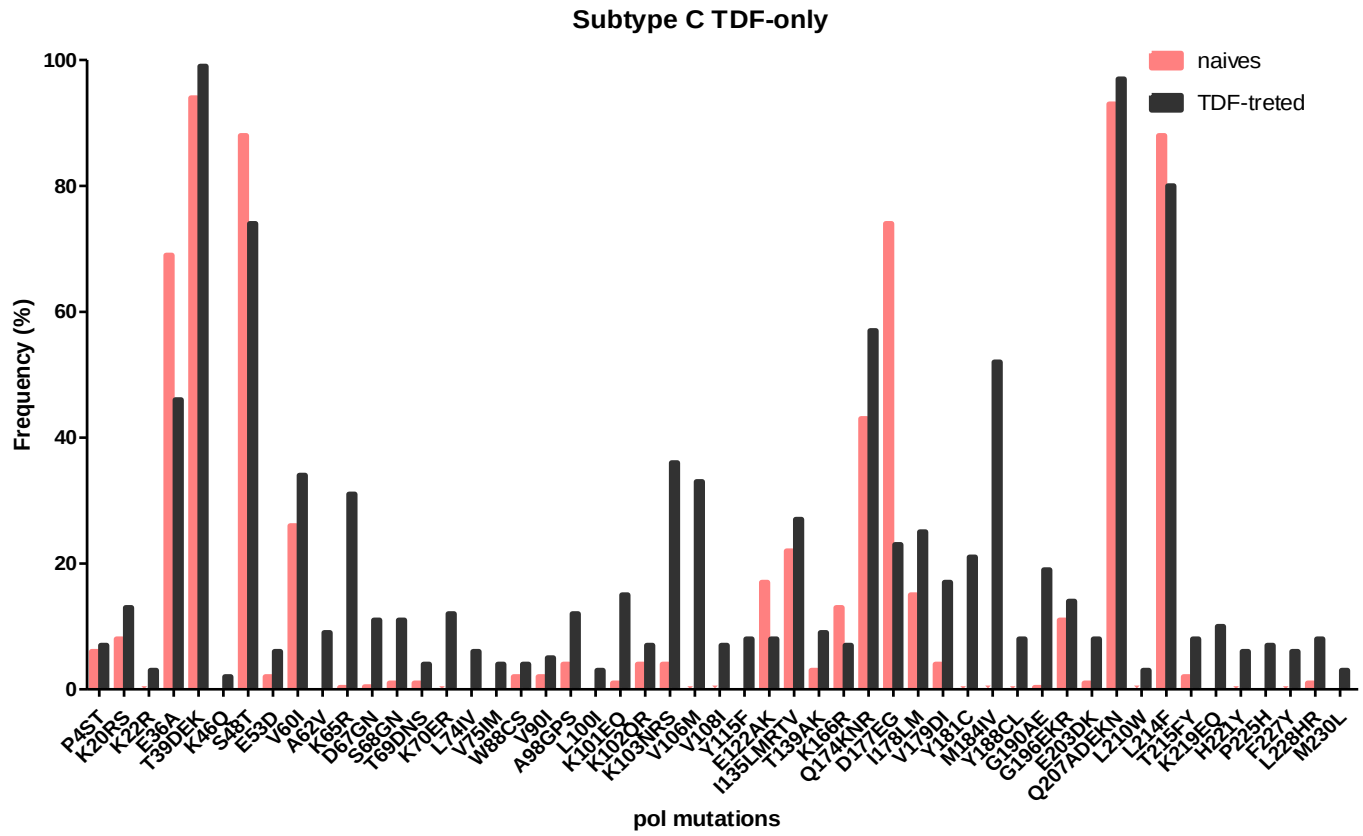

Figure 2.2A: Significant mutations of TDF-treated not prior exposed to d4T

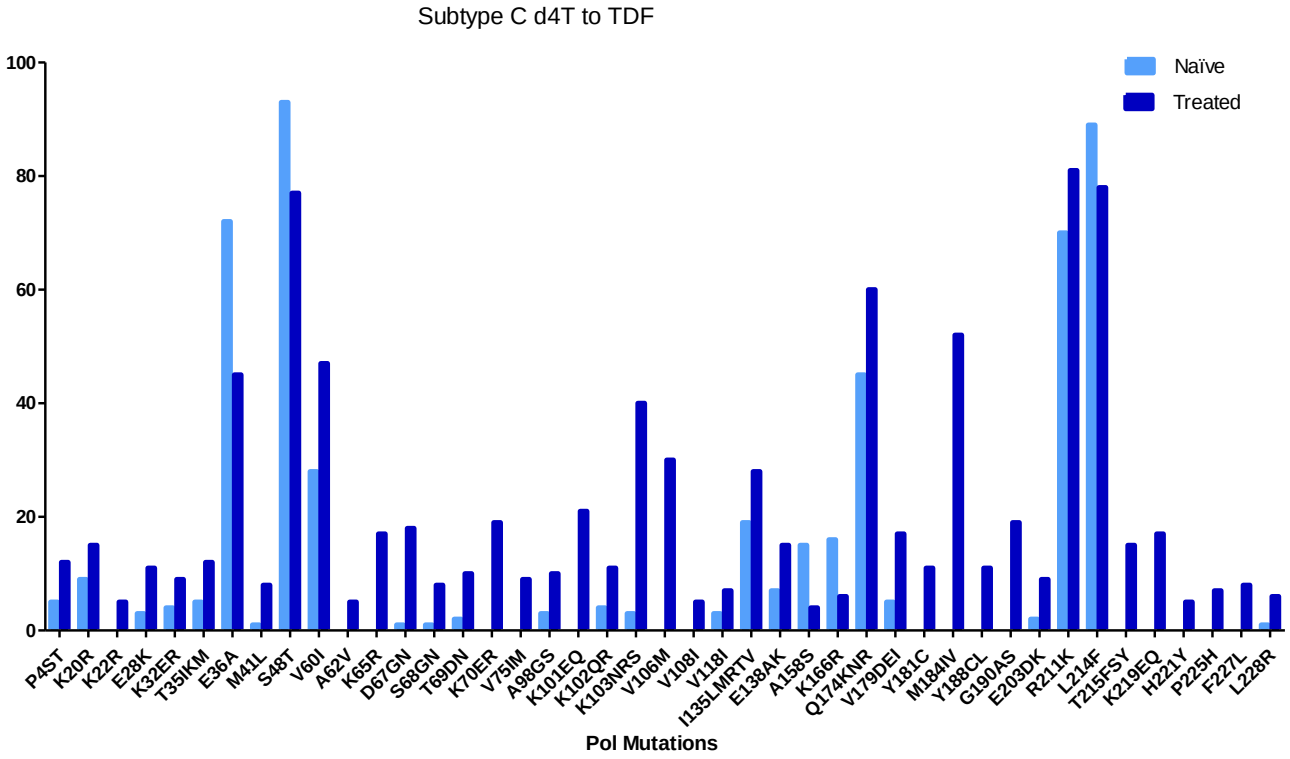

Figure 2.2B: Significant mutations of patients treated with d4T prior to TDF

2.3.2 Subtype C TDF vs d4T isolates

Sequences from d4T treated patients (without exposure to TDF) were also analysed to determine whether K65R is selected for by d4T. K65R was found in 5% of d4T-treated and 31% of TDF-treated sequences. S68N was higher in TDF (5%) than in d4T-treated (0.4%) sequences. A62V was also higher in TDF (9%) versus d4T-treated (4%) sequences. K70E which has been previously reported to be selected by TDF was 4% in TDF-treated and 1% in d4T-treated sequences. Y115F was found in 8% of TDF-treated, while in d4T was 1%. M184V, which is associated with 3TC resistance, had a higher frequency (64%) in d4T-treated compared to 46% in TDF-treated. Conversely, M184I was higher in TDF-treated (7%) compared to d4T-treated (2%), although the overall frequency was lower than M184V. As expected, TAMs had a higher frequency in d4T-treated sequences: D67N (17%), T215F/Y (14%), K70R (12%) and M41L (8%) versus D67N (11%), T215F/Y (8%), K70R (6%) and M41L(4%). Four NNRTI mutations (V106M (33%), Y181C (21%), G190A (19%) and V179D (17%) respectively) were higher in TDF-treated, while 2 NNRTI mutations (K103NS (44%) and K101E/Q (15%) were higher in d4T-treated.

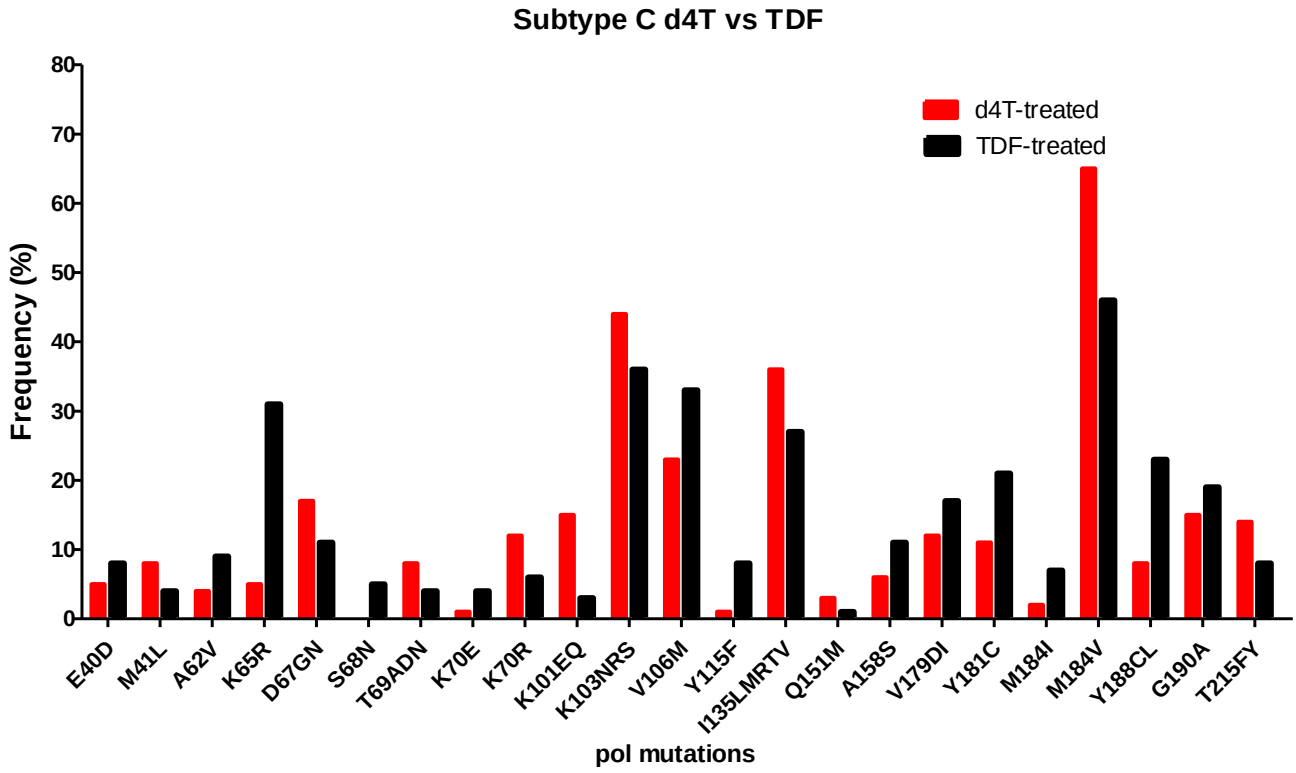

Figure 2.3: Significant frequencies of d4T-treated vs TDF treated in subtype C.

2.3.3 Subtype B TDF treated vs treated-naïve

The frequency of mutations in Subtype B viruses from patients treated with TDF were also investigated. 4215 subtype B sequences were downloaded. Of these 2585 were from TDF-treated, and 1630 from drug-naïve isolates; these were downloaded from the Stanford and Los Alamos HIV-1 databases respectively. Significantly different mutations are shown in the graph below (Figure 2.4). Most mutations showed a significant increased after TDF-treated. However, five mutations significantly decreased (K11R/T, E169D, K173E/N/R, Q207E/K and R211Q/R) (Figure: 13). K65R frequency was 10%, while Y115F was 9%. Interestingly, the frequency of TAMs also showed a significant increased (M41L (28%), K70R (15%), L210W (20%), T215F/Y (38%), K219Q/E (22%). Frequency of M184I and M184V was 4% and 38% respectively. Ten NNRTI mutations showed significantly increased frequencies: K103N (30%), A98G (16%), G190A/S (14%), K101E (13%), V179D (13%), V108I (7%), V106M (6%), E138A (6%), Y181C (6%) and L100I (5%) respectively.

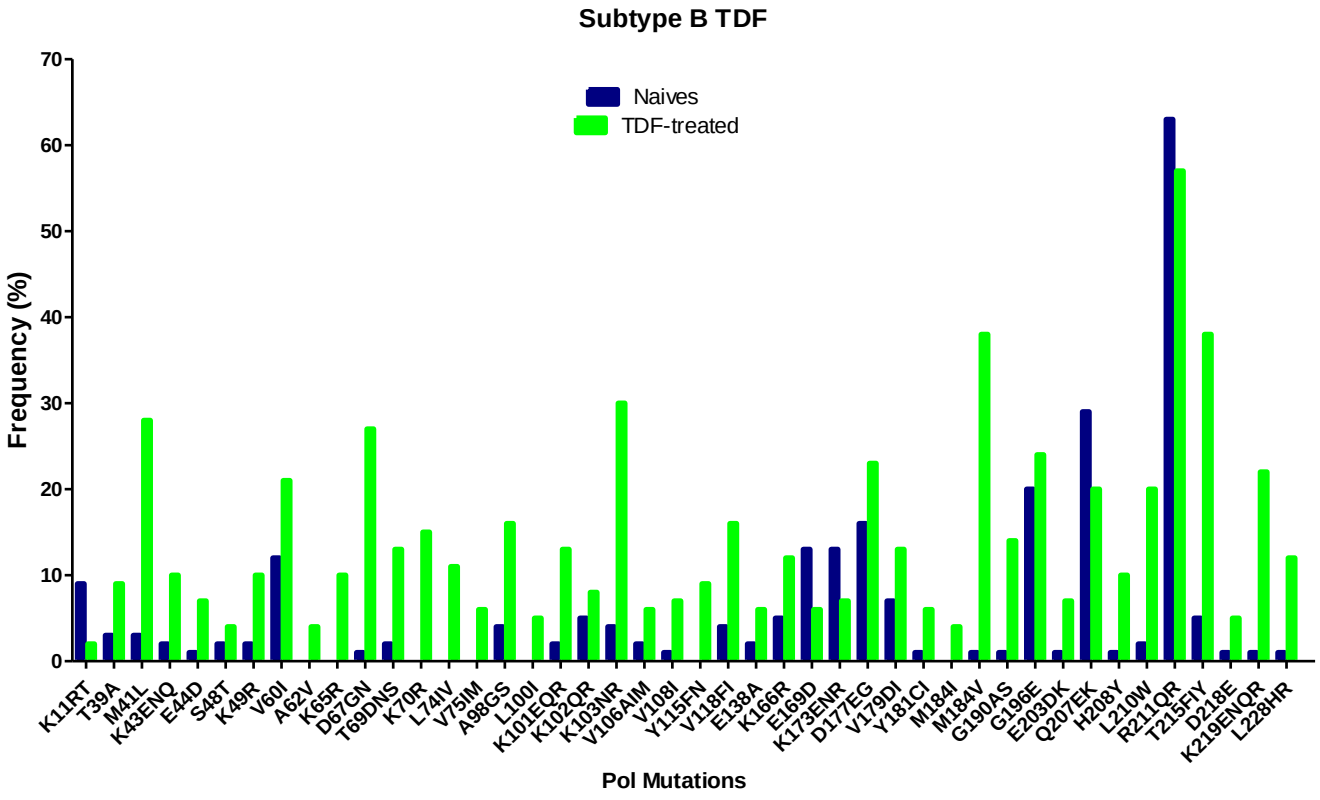

Figure 2.4: Prevalence showing significant different of subtype B TDF-treated vs Subtype B naïve-treated

2.3.4 Subtype B vs subtype C TDF-treated isolate

The mutations significantly associated with TDF treatment for both subtypes were plotted on a graph (Figure 2.5) to show the differences between the two subtypes. Of note, was the frequency of K65R in subtype C (31%) versus subtype B (10%). S68G which was previously reported to associate with K65R was higher in subtype B (7%) compared to subtype C (5%), while S68N was higher in subtype C (5%) versus subtype B (1%). Y115F which is known to be associated with K65R was higher in subtype C (8%) compared to B (2%). K70E which is selected by TDF was also higher in subtype C (4%) versus subtype B (1%). A62V was higher in subtype C (9%) compared to subtype B (4%). The frequency of TAMs were higher in subtype B compared to C.

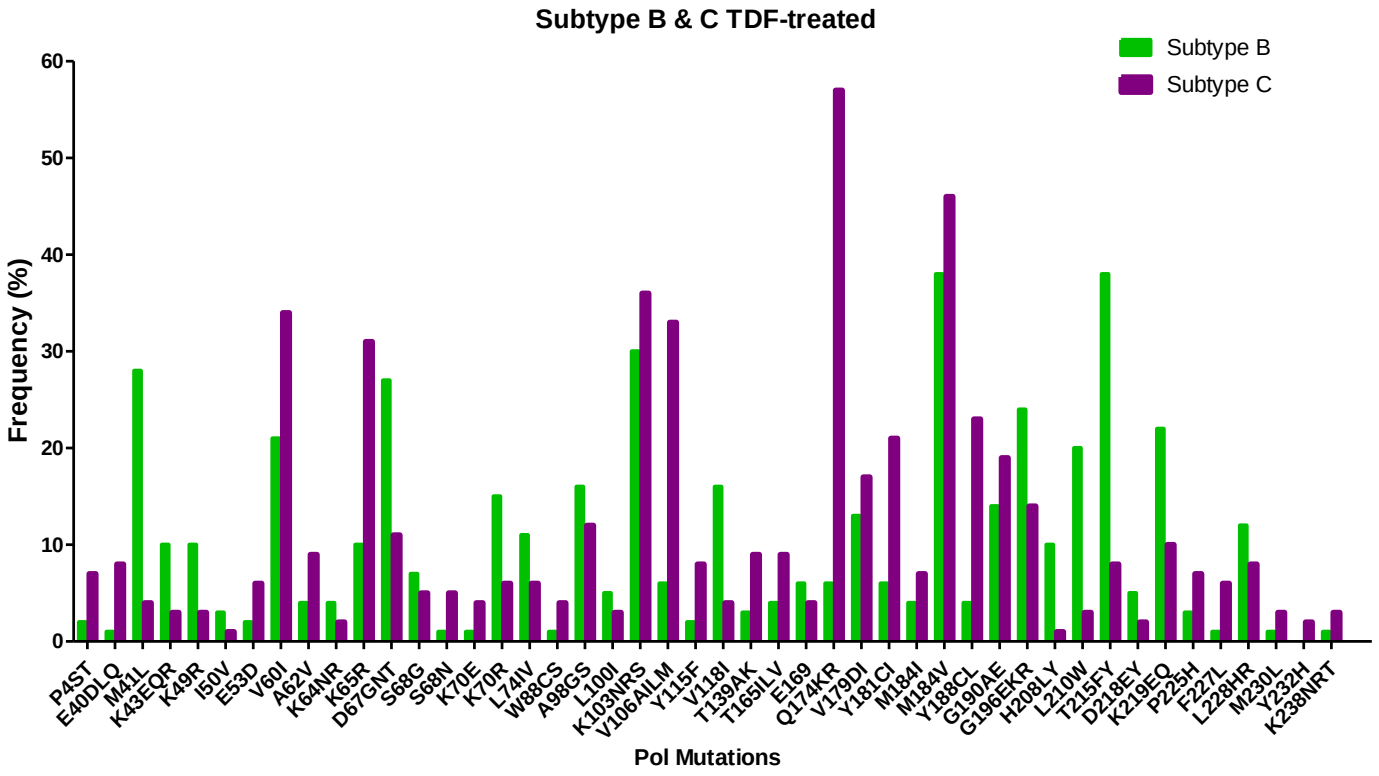

Figure 2.5: Graph showing the difference between subtype B TDF-treated and subtype C TDF-treated frequencies

2.3.5 K65R correlation with other NRTIs and NNRTIs

From those that received TDF irrespective of whether they were switched from d4T, K65R positively correlated (p=0.000) with the following NRTI mutations: A62V, S68G, S68N, Y115F and M184I. The frequency of these mutations in all sequences were 16%, 15%, 10%, 20% and 9% respectively. NNRTI mutations that correlated with K65R (p=0.000) were: V106M, Y181C, G190E and M230L, and were present at a frequency of 47%, 39%, 8% and 9% respectively. K65R also negatively correlated with the TAMs (D67N, K70R, L210W, T215F and T215Y), although they were found in combination in some sequences: D67N (5%), K70R (1.5%), L210W (0%), and T215F/Y (1%).

Sequences from TDF treated patients were then separated according to their treatment history: those that only took TDF and those that took d4T prior to TDF.

TDF-treated only: 147/475(31%) isolates harboured the K65R mutation. Eighteen mutations positively correlated with K65R, of these 5 were NRTI (A62V [22%], S68G [12%], S68N [16%],Y115F [21%] and M184I [11%]) and 8 were NNRTIs mutations. It also negatively correlated with TAMs (K70R, T215F/Y and K219Q) and 2 NNRTIs (A98S and P225H).

D4T to TDF treated: 25/140 (18%) isolates harboured the K65R mutations. Significant (p=0.000) correlations were: A62V (20%), S68G (24%), S68N/K (12%) and Y115F (20%), as well as V106M (56%), Y181C (32%) and M230L (8%). Interestingly, the negative correlation with the TAMs was not seen in this group. Negative correlations were found between K65R and I135T (p=0.038) and K65R and R211K.

Table 1: subtype C sequences, correlation and association of K65R with other mutation from 3 groups of TDF-treated

TDF-treated patients (overall)				TDF-treated patients only				TDF-treated patients (d4T to TDF)		
Mutations	P-value	Percentage		Mutations	P-value	Percentage		Mutations	P-value	Percentage
A62V	0.000**	16%		K46Q	0.007**	4%		E28V	0.027*	4%
D67N	0.004**	5%		A62V	0.000**	22%		A62V	0.000**	20%
S68G	0.000**	15%		S68G	0.000**	12%		S68G	0.000**	24%
S68N	0.000**	10%		S68N	0.000**	16%		S68K	0.024*	4%
K70R	0.004**	1.50%		K70R	0.002**	1%		S68N	0.018*	8%
V75I	0.002**	5%		K70T	0.034*	1%		L74I	0.024*	4%
L100I	0.004**	5%		N88S	0.003**	3%		V75I	0.001**	12%
K101E	0.004**	15%		A98S	0.007**	3%		A98S	0.020*	16%
K101Q	0.031*	5%		L100I	0.000**	7%		L100I	0.024*	12%
K102R	0.000**	11%		K101E	0.000**	16%		K101R	0.024*	4%
V106M	0.000**	47%		K102R	0.017*	7%		K102R	0.000**	20%
V108I	0.035*	10%		V106M	0.000**	46%		K103H	0.024*	4%
Y115F	0.000**	20%		Y115F	0.000**	21%		V106M	0.000**	56%
I142V	0.019*	2%		E122P	0.034*	1%		L109M	0.024*	4%
Q174K	0.049*	47%		Q174K	0.021*	54%		Y115F	0.000**	20%
V179D	0.001**	15%		Y181C	0.000**	39%		I135T	0.038*	0%
V179G	0.010*	1.50%		M184I	0.035*	11%		E138Q	0.024*	4%
Y181C	0.000**	39%		G190A	0.017*	19%		K173L	0.024*	4%
M184I	0.038*	9%		G190E	0.001**	6%		K173R	0.024*	4%
G190E	0.000**	8%		G190S	0.005**	3%		Y181C	0.000**	32%
G190S	0.006**	5%		T215F	0.006**	1%		Y188H	0.001**	8%
L210W	0.036*	0%		T215Y	0.030*	0%		G190Q	0.024*	4%
T215F	0.006**	0%		K219Q	0.014*	1%		R211D	0.025*	4%
T215Y	0.027*	1%		P225H	0.006**	1%		R211K	0.009**	60%
H221Y	0.003**	8%		M230L	0.000**	8%		F227A	0.025*	4%
P225H	0.015*	1.50%						M230L	0.001**	8%
M230L	0.000**	9%								

2.3.6 K70E correlation with other NRTIs and NNRTIs

K70E was investigated as an alternate pathway to TDF resistance. We found that 21/475 (4%) isolates harboured the K70E mutation. Only three of these isolates also harboured the K65R mutation; however this negative correlation with K65R was not statistically significant (p=0.09). K70E positively correlated with 5 NRTIs [D67N(p=0.000) V75I/S/T(p=0.009), M184V(p=0.001), E203K(p=0.014) and L228I/R(p=0.002)]. Correlation table in appendix A. K70E was found together with D67N in 52% of the sequences and with M184V in 81% of sequences. This is in contrast to the M184I variant that correlated with K65R, as shown above (Table 2). In addition, K70E correlated with 7 NNRTIs (V90I,

A98G, K101A/D/R, K103S, V106M, G190A/E and K238T) and 3 novel mutations (E122P, T139K and G196R).

2.3.7 TDF- treated pattern

We manually investigated mutational patterns in those that were only TDF-treated. K65R in combination with Y115F was the most dominant pattern observed, followed by K65R with S68N and K65R with S68G. The combination of K65R and Y115F was seen in 15% (22/147) of the sequences, followed by S68G in 9% (13/147) of the sequences, while a novel combination with S68N occurred in 14% (21/147) sequences. K65R+M184I was found in 10% (14/147) of the sequences The triple combination of K65R, S68G + Y115F mutations occurred in 3% (4/147) followed by K65R, S68N+Y115F; and K65R, L74V+Y115F mutations occurred in 2% (3/147) of the sequences. Interestingly, we also found mutations from the TAM II pathway in the TDF treated group (D67N, K70R and K219Q).

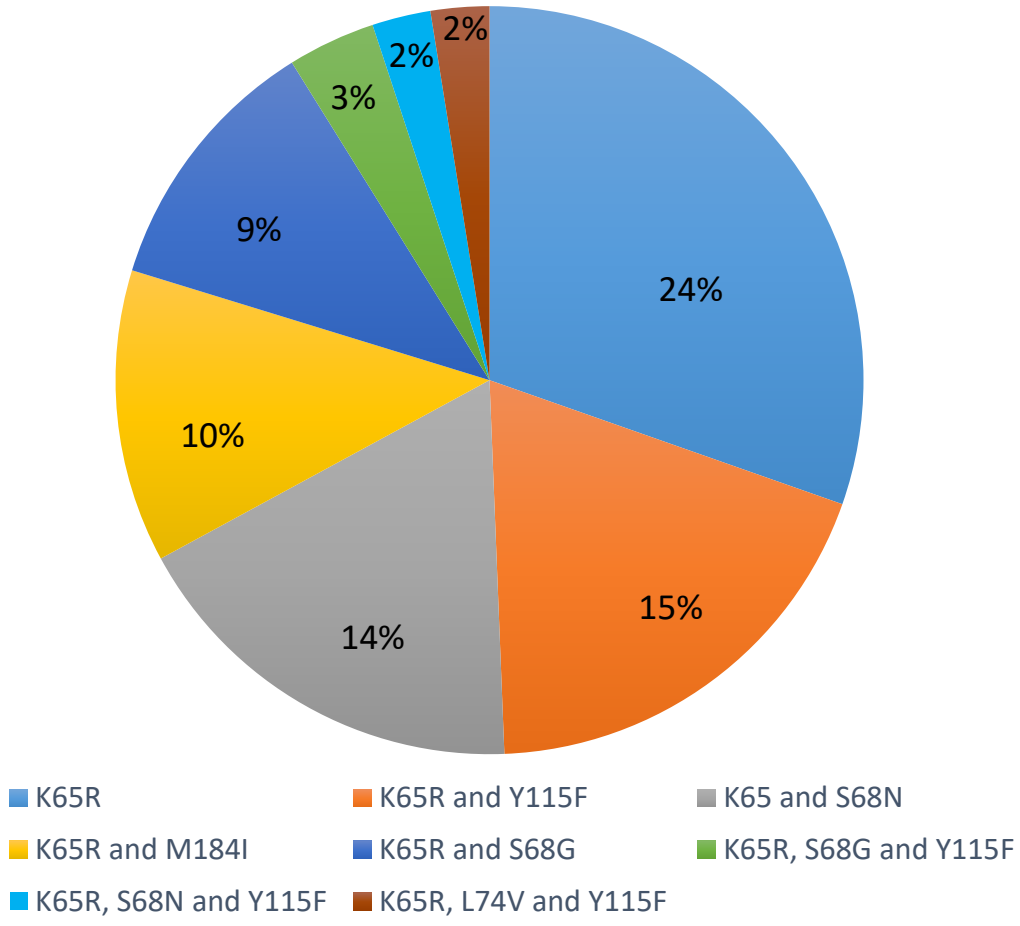

Figure 2.6: Pie chart showing the TDF-treated mutational pattern.

2.4 Discussion

In subtype C, K65R was found in 31% of TDF-treated and in 5% of d4T- treated sequences. This is in agreement with other studies that reported a high prevalence of K65R in subtype C (Skhosana et al., 2015, Sunpath et al., 2012b). This was expected since TDF is known to select for K65R both in *vivo* and *in vitro*, while d4T selects for TAMs (Doualla-Bell et al., 2006, Kuritzkes et al., 2004, Sunpath et al., 2012b). However, K65R has also been found in d4T treated patients 5%, suggesting that K65R is also selected in d4T-treated isolates. Those that were TDF-treated also showed a negative correlation between the TAMs and K65R, which is in agreement with other studies that reported an antagonistic effect between K65R and TAMs. However, 2% of the TDF treated sequences also harbored mutations from the TAM 2 pathway. This was found in patients who took TDF with no prior history of d4T or AZT, which are the drugs known to select for TAMs.

K46R was frequently selected in TDF-treated isolates and also positively correlated with K65R in subtype C. Also, K46Q was found only in sequences containing K65R. On the other hand, this was not the same with subtype B, therefore K46Q a may be associated with K65R in TDF-treated only in subtype C.

K65R was found at a lower frequency (10%) in subtype B TDF-treated sequences. It is known that there is dissimilarity of the nucleotide sequence at codon K65, with nucleotide sequences AAA, AAG and AAG at position 64, 65 and 66 in subtype C and AAA-AGA nucleotide sequences in subtype B (Brenner and Coutsinos, 2009a, Garforth et al., 2014a).

Y115F, which is known to be associated with K65R, was found in both C (6%) and B (2%) TDF treated isolates. The frequency was significantly higher ($p<0.0001$) in subtype C, suggesting that Y115F is more frequently selected under TDF pressure in subtype C. Initially, it was thought that Y115F was only selected in patients receiving ABC, however in a study by Skhosana et al, however 13.8% of TDF treated patients harboured Y115F and none of these patients were exposed in ABC (Miller et al., 2000, Skhosana et al., 2015). Also, Y115F was reported to be selected after the development of K65R, and their combination is known to decrease TDF susceptibility (Stone et al., 2004, Das and Arnold, 2013).

K70E was seen in 4% of TDF-treated sequences and has been previously reported to be selected in TDF-treated isolates (Skhosana et al., 2015); this was higher than in subtype B (1%). Furthermore, we also found negative correlation between K65R and K70E, which have been previously reported to have an antagonistic effect of K70E with K65R (Sluis-Cremer et al., 2007). It has been suggested that K70E is an alternative to the K65R pathway to TDF resistance (Kagan et al., 2007) While K65R and K70E have not previously been found together on the same viral genome, we did find that K65R and K70E

were detected in 3 sequences; however these were bulk sequences. K70E also negatively correlated with S68N and Y181C, which were positively correlated with K65R. Interestingly, K70E significantly correlate with one of the TAMs (D67N).

Our data shows a significant correlation of K65R with the NNRTI mutations. It has been previously reported that K65R is highly associated with the use of EFV and NVP. NNRTI mutations were also strongly associated with TDF treatment: V106M, V179D, Y181C and G190A ($p<0.0001$). The association of K65R and Y181C and the pattern that include K65R, Y181C and G190 has also been previously reported in patient taking a combination therapy (Theys et al., 2009, von Wyl et al., 2008). Interestingly, K101E/Q and K103NS were associated with d4T treatment. This may be due to the positioning of the mutations within the NNRTI binding pocket; K101E and K103NS are located in the rim of the pocket entrance while other NNRTIs mutations are located inside the pocket (Das and Arnold, 2013).

Other combinations with K65R that were found included S68G and L74V. These have been previously reported in subtype B sequences (Laura et al., 2008, Trotta et al., 2006). We also identified a combination of K65R with M184I. M184I is normally reported to change to M184V over time and its less fit than M184V (Hu and Kuritzkes, 2011). This is because HIV-1 is more prone to the change from G to A which results to M184I mutation. However, their enzymatic efficiency is less than that of M184V (Frost et al., 2000).

Furthermore, we found a novel combination of K65R and S68N. The frequent combination of K65R with Y115F and S68N seen only in subtype C suggests that this combination may contribute to TDF-resistance.

Chapter 3 The interaction of K65R (and K70E) with other NRTI and NNRTI mutations in TDF exposed, using Bayesian network analysis and 3D homology modelling

3.1 Introduction

The high frequency of K65R in TDF-treated patients has been noted in several studies on subtype C (Hoffmann et al., 2013a, Skhosana et al., 2015, Van Zyl et al., 2013b).The increase of K65R in subtype C can be explained by the homopolymeric nature of nucleotide sequences at codon 64 and 65 (AAA-AAG) in the RNA template (Coutsinos et al., 2009). Others have also reported that K65R can be selected by NNRTI mutations (Theys et al., 2009). Furthermore, K70E has also been shown to play a role in TDF resistance. In a study by Skhosana et al, a small number of TDF-treated developed K70E (Skhosana et al., 2015). This was also seen in our analysis as described in the previous chapter. Interestingly, K65R and K70E have been reported to be antagonistic and are rarely found on the same viral genome (Kagan et al., 2007, Skhosana et al., 2015). Both K65 and K70 residues are located within the β3-β4 loop found in the finger subdomain of HIV-1 RT (Sluis-Cremer et al., 2007).

BNs have been successfully used to investigate the interactions of mutations that confer resistance to protease inhibitors (PIs), NRTIs and NNRTIs (Deforche et al., 2008, Theys et al., 2009). In particular, BN analysis has previously been used to identify positive associations of K65R with S68G, L100I and Y181C, as well as an antagonistic association with T215Y in HIV-1 subtype B (Theys et al., 2009).

Homology modelling has also been used to investigate the interaction of K65R with other resistance mutations e.g: K65R with M184V or Q151M and the negative association of K65R with L74V and most TAMs (Das et al., 2009). This is a computational technique, used to determine the three dimensional (3D) structure of proteins(Bishop et al., 2008). 3D protein structures provide valuable insights into the molecular basis of protein function (Schwede et al., 2003).

In this chapter, we investigated the possibility of a Tenofovir-associated resistance pathway unique to subtype C. We address the following objectives set out in the thesis:
a) To investigate the interaction of K65R with other NRTI and NNRTI mutations in subtype B and C using a Bayesian Network model.
b) To demonstrate the interaction of K65R with other resistance mutations using 3D homology modelling.

3.2 Methodology

3.2.1 Bayesian network

Bayesian network learning was done using the B-Course software programme available online (http:/www.b-course). Briefly, the dataset for BN analysis was prepared as a simple text file (tab delimited format) with the top row containing the name of the variables (mutations) and subsequent rows listing the cases (samples) to be analysed. The maximum size allowed for the dataset in B-Course is one megabyte; it is preferable to use a smaller dataset, since the larger the dataset, the longer it takes for the file to upload and be analysed. B-Course has 2 trail options: Dependency modelling and Classification. For our analysis, the Dependency model was used. The final network was provided as a picture (PNG).

The strength of the arcs were taken from the final report provided for the network. The network arcs explain the dependency of the variables, and describe the effect of removing the arc on the probability of the model. Arcs with support of one billionth probability were considered robust (Table 2).The arc were coloured according to their estimated probability. The arc direction reflect the causal influence indicating the dependency of arcs (appendix B). Networks were constructed to investigate the direct interaction of mutations with K65R, and K70E in viral sequences obtained from TDF-experience patients.

Table.2: The table below lists the different types of arcs that can be found in dependency models.

	The black arcs so 'strong' that removing any single one of them would cause the probability of the model to go down to less than one billionth of the probability of original model.
	The purple arcs are middle strong, so that removing any of them would results in a model with probability less than one millionth of that of the original model.
	Removing blue arcs from the model would decrease the probability of the model to less than one thousandth of the probability of the original model (exact ratio listed).
	The red represents an antagonistic effect.

3.2.2 3D homology modelling

3D homology modelling was done using SWISS-MODEL (https://swissmodel.expasy.org). Amino acids were uploaded to the SWISS-MODEL server. A template search was performed for models that best matched the query sequence with the highest percentage identity. The template was selected and the model was constructed on the SWISS-MODEL server. Structures were viewed and analysed in Chimera v1.8.1 (Petterson et al., 2004).

3.3 Results

3.3.1 Bayesian Network

3.3.1.1 Subtype C TDF-resistance associated pathway

The variables used in the data set for Bayesian network learning included mutations that were found to be associated with TDF-treatment and d4T-treatment. The dataset included all significant TDF-resistance mutations. As seen in Figure15, the network showed a direct association between TDF and K65Rwith a high probability of 1billionth (strong bootstrap) and K70E, although this was only weakly supported (1 thousandth probability). Therefore, this suggested two separate TDF resistance pathways: one that included K65R and another that included K70E. The network also showed that there was no direct association between d4T and K65R. Surprisingly, d4T did not show any direct influence on the TAMs.

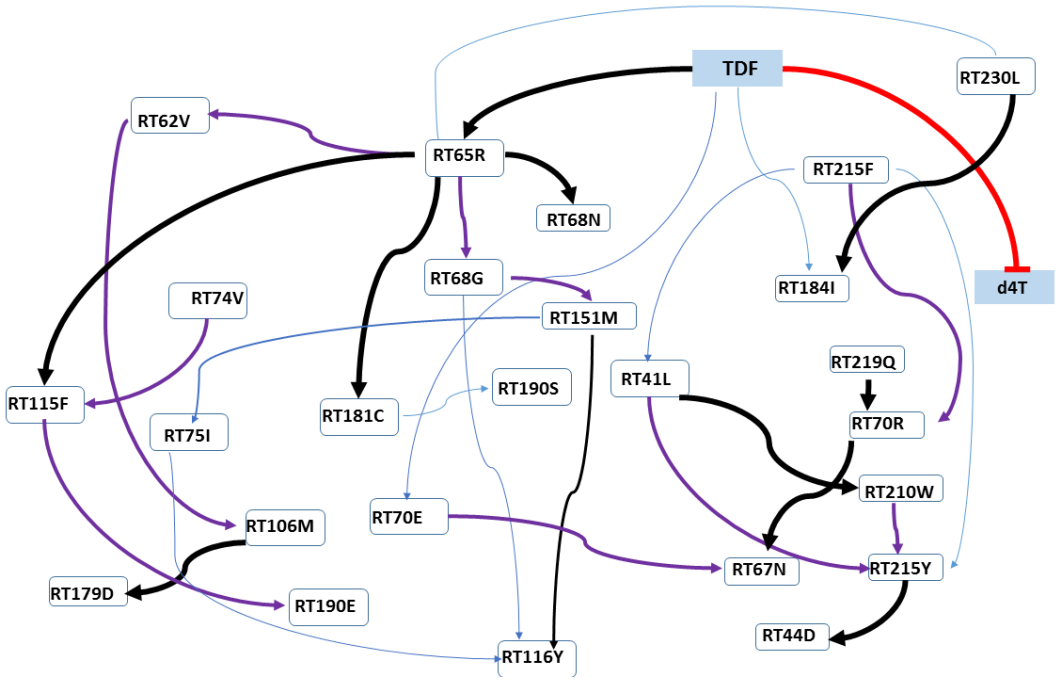

Figure 3.1: Bayesian Network learned from subtype C showing TDF-treated and d4T-treated mutations. Nodes represent mutations and exposure to treatment. Arc colour represents the probability support, removing any one of them would cause the probability of the model to go down by: one billionth probability (Black), or one millionth probability (purple), or one thousandth probability (blue), or antagonistic influence (red).

3.3.1.2 K65R pathway

A separate network was investigated using only those mutations that correlated with K65R. This included a total of 29 variables. The BN analysis revealed a strongly supported association of K65R withY115F andY181C. The Y181C pathway was further associated with other NNRTI mutations: G190A/S, V108I and P225H. Lesser supported associations were with A62V, S68N, V106M, and M230L. The association with T215F indicated a strong antagonism towards K65R, which is in agreement with several other reports (Theys et al., 20069). The T215F pathway was associated with the TAMs.

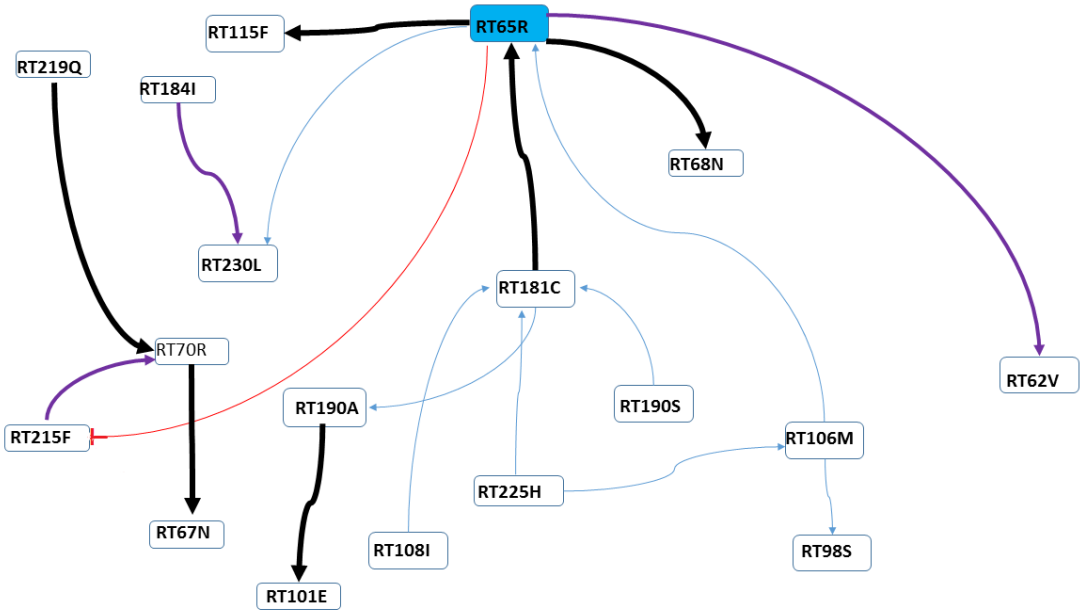

Figure 3.2: Bayesian Network expressing direct association between K65R and other mutations in TDF-treated subtype C infected patients. The nodes represent all amino acid changes that were significantly associated with K65R Arc colour represents the probability support: one billionth probability (Black), or one millionth probability (purple), or one thousandth probability (blue), or antagonistic influence (red).

3.3.1.3 K70E pathway

The dataset included all mutations that correlated with K70E. The Bayesian Network showed a direct association of K65R with 1 NRTI (D67N) and 3 NNRTIs (98G, 101A 103S and 106M). The Network also revealed a strong association between K70E and D67N, while D67N showed an interaction with M184V. M184V is associated with both A62V and G190A.

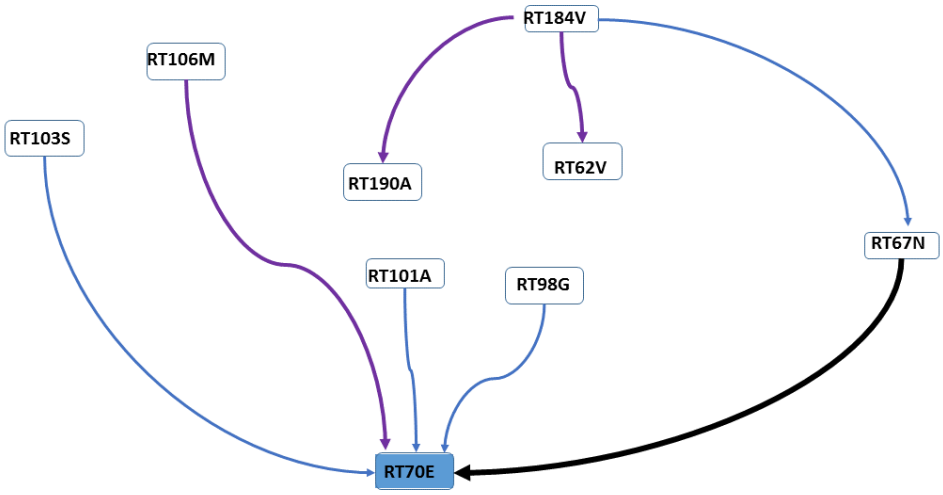

Figure 3.3: Bayesian Network expressing direct association between K70E and other mutations in TDF-treated subtype C infected patients. The nodes represent all amino acid changes that were significantly associated with K70E in TDF-treated. Arc colour represents the probability support: one billionth probability (Black), or one millionth probability (purple), or one thousandth probability (blue), or antagonistic influence (red).

3.3.1.4 Subtype B TDF-resistance associated pathway

A network was investigated using only mutations that correlated with K65R in subtype B. This included a total of 24 variables. Similar to Theys network, our subtype B network reveal the positive association of K65R with A62V, S68G, L100I, Y181C. It also showed an antagonistic association with T215Y.K65R pathway includes a direct positive association of 1 billion probability with Y181C.1millionth probability with A62V and 1 thousand probability with S68G and L100I. A62V and S68G are minor mutations that interact direct with a major mutation (K65R) with no other latent interaction.

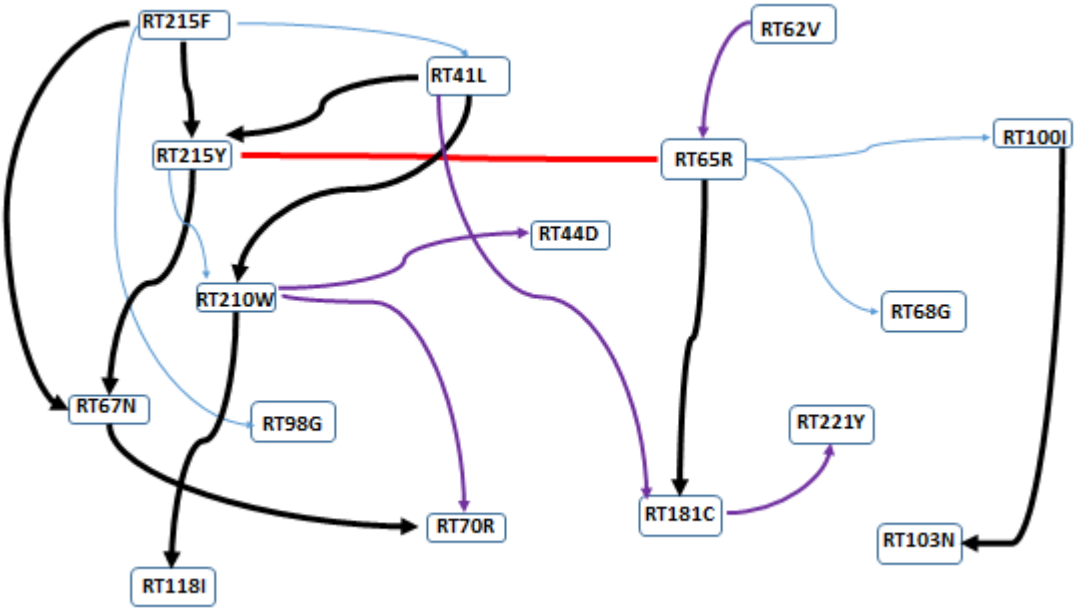

Figure 3.4: Bayesian Network expressing direct association between K65R and other mutations in TDF-treated subtype B infected patients. The nodes represent all amino acid changes that were significantly associated with K65R**.** Arc colour represents the probability support: one billionth probability (Black), or one millionth probability (purple), or one thousandth probability (blue), or an antagonistic influence (red).

3.3.2 3D homology modelling

Molecular modelling was used to visualize the positioning of the RT mutations that were found to be associated with TDF-resistance and their effect on the structure of RT. The positions of A62V, K65R D67N, S68G/N and K70E are shown on Figure 3.5. K65R, D67N, S68G/N and K70E are located in the β3- β4 loop within the finger sub-domain, while A62V, also located in the finger subdomain, and is found outside of theβ3- β4 loop. Interestingly, other mutations (Y115F and Y181C) that were also associated with TDF-resistance are located within the palm.

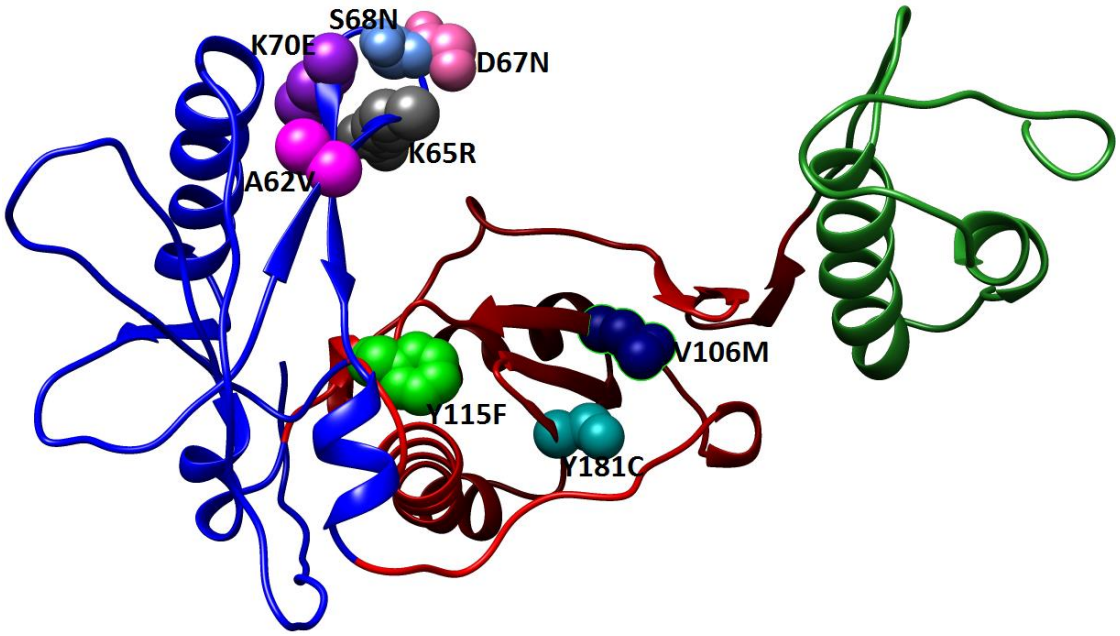

Figure 3.5: RT structure showing the location of mutations that are associated with TDF-resistance. Finger (blue), palm (red) and thumb (green**)**.

3.3.2.1 K65R mutation

The ball and stick representation shows differences between lysine (K65) and arginine (65R). The model shows a slight shift of 65R away from the wild type K65 residue (Figure 3.6A). This substitution may modify the positioning of this residue and affect the binding of the NRTI. S68G shifts away from the position of the wild type, and consequently moves away from codon 65. The distance between the wild type amino acids K65 and S68 (3Å) located within the β3 and β4 loop, is smaller than between the mutant residues 65R and 68N (4Å) (Figure 3.6B & C). Y115F (yellow) which is located in the α4 helix, loses the OH group present in the Tyrosine wild type residue (green) (Figure 3.6D) and therefore only contains a carbon ring. Similarly, same 181C loses an OH group and parts of the aromatic ring (Figure 3.6E).

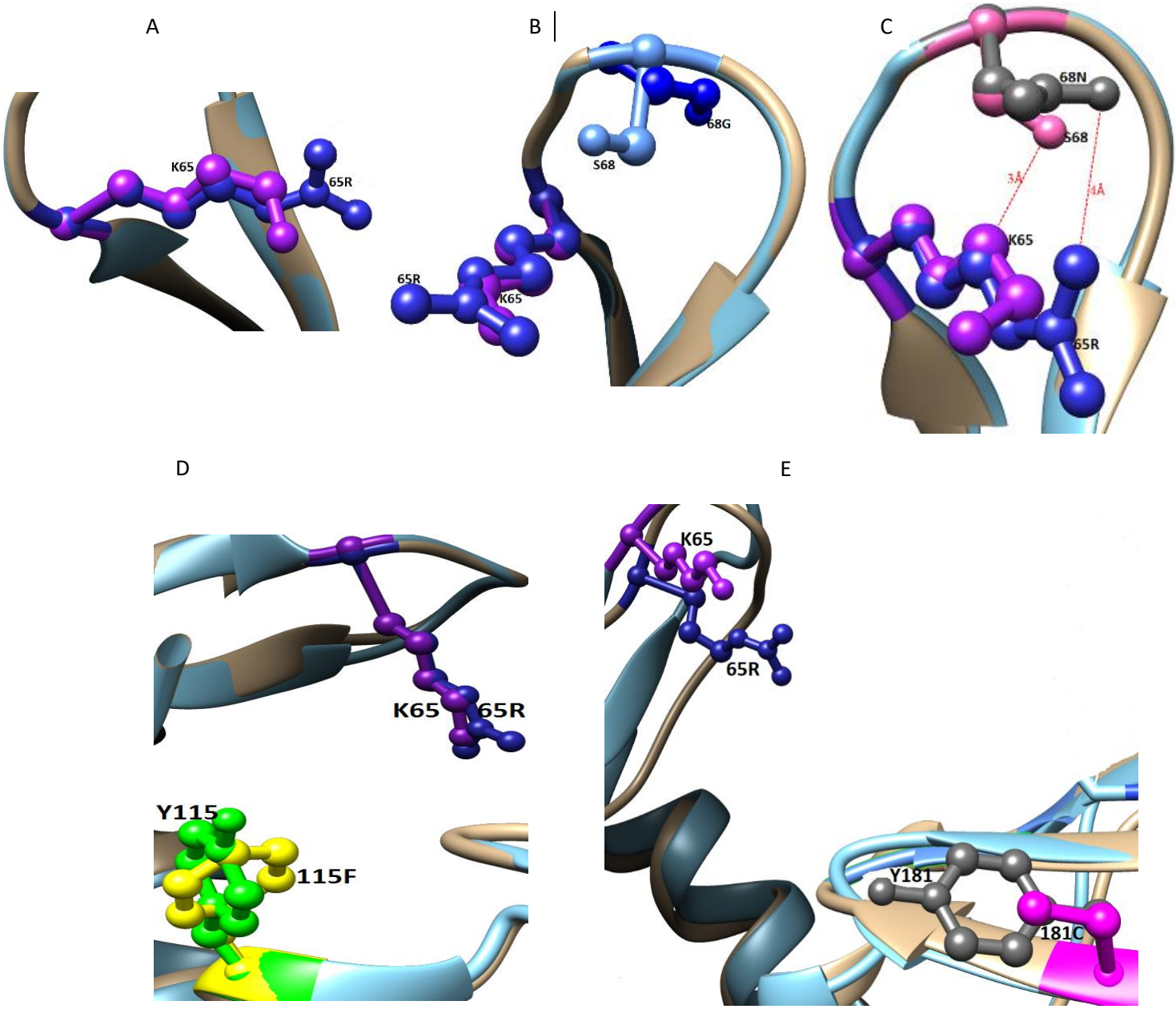

Figure 3.6: Conformational difference among mutations that interact with K65R. (A) Structural difference between wild type and mutant at position 65 (purple- K65 and blue-R65). (B) Interaction of K65R with S68G, (C) S68N, (D) Y115F and (E) Y181C.

We also compared the positioning of M184I and M184V, since M184I correlated with K65R while M184V correlated with K70E. M184 is located near the active site and usually mutates to 184V (purple) or 184I (yellow), which is more considered a transitional mutation. These mutations cause the residues to move distally away from wild type position, moving towards the active site (Figure 3.7 A & B), therefore blocking the binding of the incoming NRTIs.

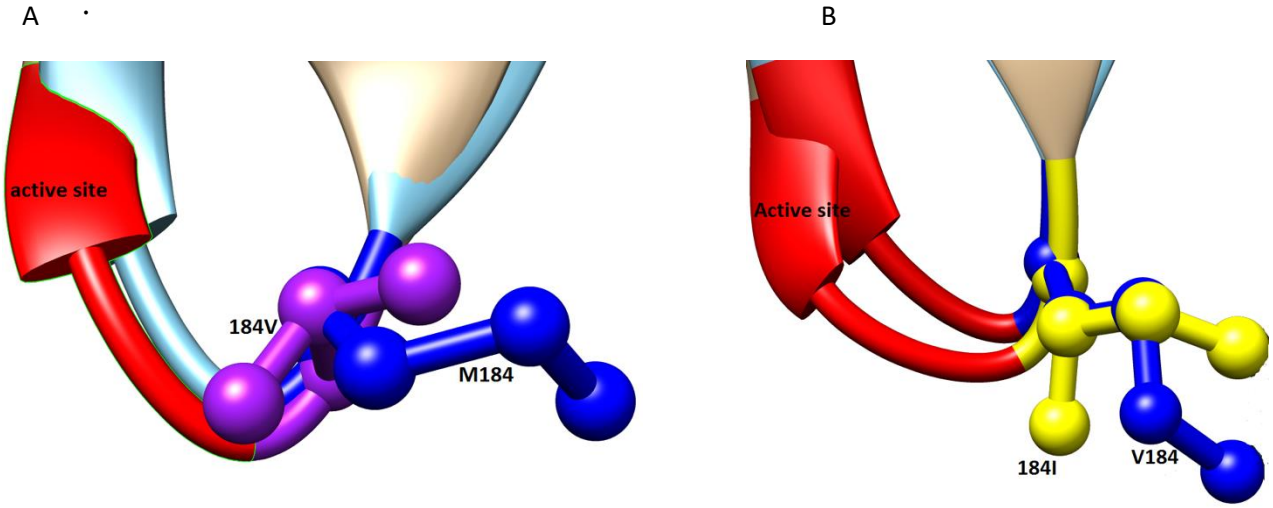

Figure 3.7: Residue 184 interaction with the active site. Structural difference between wild type and mutant in position 184 (blue: M184: yellow: 184I, purple: 184V). (A) Structural change from M to V in position 184. (B) Structural change from M to I in position 184.

3.3.2.2 K70E mutation

K70E is located on the same loop as K65R. Both 67N and 70E mutants move towards each other, while the wild type D67 and K70 are located distally away from each other (Figure 3.8B). The distance between the mutants (5Å) are less then between the wild types (9Å)

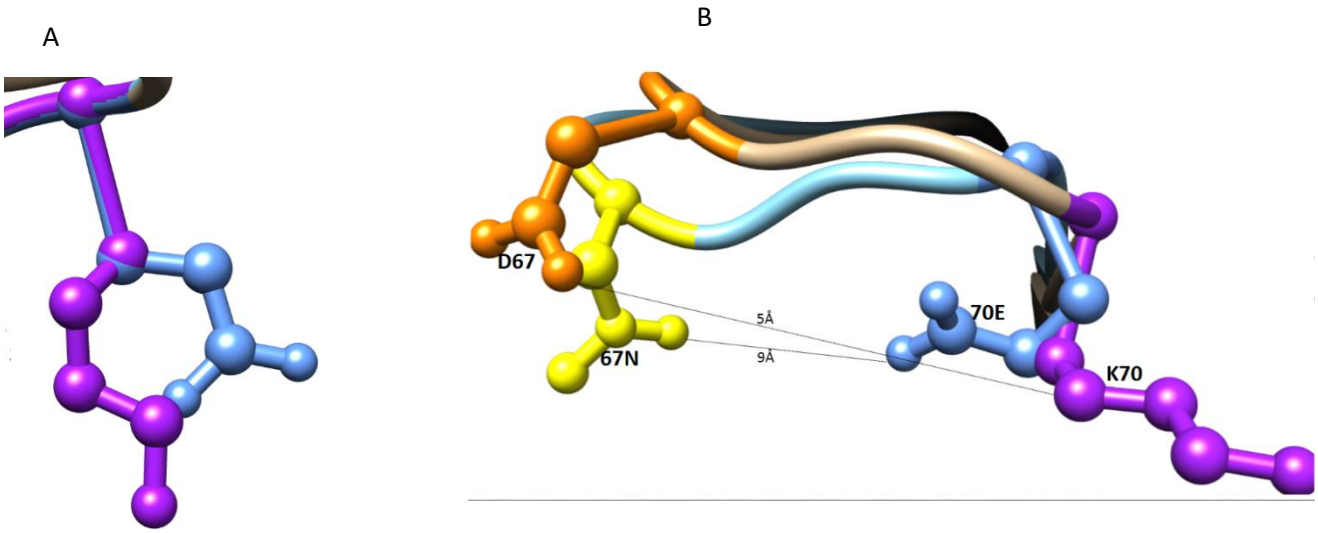

Figure 3.8: Conformational changes and interaction of double mutant K70E and D67N. Structural difference between wild type and mutant in position 70 and interaction of between mutation K70E and D67N.

3.4 Discussion

Bayesian network learning identified 2 resistance mutations that are directly selected by TDF treatment in subtype C, suggesting two separate TDF resistance pathways in this subtype C: one that includes K65R and another that includes K70E. However, since the frequency of K70E in TDF treated patients is low, this pathway is rarely selected. Further analysis of the K65R and it's role in the pathway to TDF resistance identified positive associations with A62V, S68G/N, Y115F, Y181C and M230L, and a negative association with T215F. The network showed a strong asociation of K65R and S68N, and only a weak interaction with S68G, which is in contrast to a subtype B study that showed a strong interaction with S68G (Theys et al., 2009). A62V, S68G and Y115F have been previously reported to associate with K65R (Miller, 2003). A62V and S68G were reported to develop along with K65R, while Y115F was reported to be selected after the delelopment of K65R (Miller, 2003, Stone et al., 2004). The subtype C BN identified a strong association between K65R and Y115F. This was different from the subtype B BN of Theys et al which did not showed a direct interaction with Y115F (Theys et al., 2009). In addition, M230L was associated with M184I, which has also been previously reported; the effect of these two mutations together is still unknown (Deforche et al., 2008). The connection of K65R with the NNRTIs mutations in the BN suggests that there is cross resistance of K65R with the NNRTIs. It has been previously reported that reported the the presence of K65R and Y181C may contribute to the rapid failure of TDF and EFV/NVP treatment (Deforche et al., 2008).

The alternate K70E pathway includes D67N, A98G, K101A, K103S and K106M. This suggest that K70E is associated with the TAM II pathway and the NNRTI mutations. These results are suprising since Sluis-Cremer et al, reported an antagonistic effect of K70E with the TAMs in subtype B. The network also revealed the association between D67N and M184V, which interacts with A62V and G190A.

The Subtype B network showed an association beween K65R and A62V, S68G, L100I, Y181C and an antagonistic association with T215Y. This is similar to the network by They's network that also showed an association between K65R and S68G, L100I, Y181C as well as an antagonistic effect on T215Y (Theys et al., 2009). Taken together, these findings suggest that there are distinct differences in the pathways to TDF-resistance in subtype B and subtype C.

The modelling showed that K65R and S68G/N are located in the β3 and β4 finger loop which suggests that they make contact with dNTPs when the finger fold towards the active site in the palm. The modelling also showed that residue 65R shifted away from making the contact with γ-phosphate, disrupting the binding of NRTIs

Phenylalanine at position 115 only includes the carbon rings and lack the OH, this may interrupt the hydrogen bonding between template and 3'pocket. Y181C loses the OH and other parts of the aromatic ring that are found in Y181 making it difficult to for the interaction with the NNRTIs. Therefore the interaction of K65R, Y115F and Y181C suggest that there is a high instability of hydrogen bonding leading to a disruption of NRTI binding.

In addition, M184 is involved in the binding and positioning of the incoming dNTP and primer (Huang et al., 1998). Once mutated, β-branched of 184I/V block the active site and interfere with the binding of dNTP and primer.

In conclusion, results suggest two pathways associated with TDF-resistance in subtype C. Both K65R and K70E are associated with NNRTI mutations, suggesting cross resistance between TDF and NNRTI mutations. K65R, Y115F and Y181C all undergo structural changes that disrupt the binding of NRTIs and possibly also the NNRTIs.

Chapter 4 Discussion and conclusion

In South Africa, TDF is used in the current first line ARV regimen, as well as in the second-line regimen of the 2010 National guidelines (Organization, 2010). However its efficacy is limited by the development of K65R, a mutation associated with TDF resistance (Kagan et al., 2004). Prevalence reports of K65R in patients taking TDF in subtype C range between 23% and <65%, much higher than for subtype B (Hoffmann et al., 2013a, Sunpath et al., 2012b). This study investigated the mutational pathway and the role of K65R in the development of TDF resistance using a BN in both subtype B and C isolates.

We found that K65R was the dominant TDF-resistance associated mutation observed and was selected in 31% of TDF-treated isolates in subtype C. This was in agreement with other subtype C studies that reported a K65R prevalence of 46% and 23% (Hoffmann et al., 2013a, Van Zyl et al., 2013b). Interestingly, K65R was also found in d4T treated isolates, suggesting that it is also selected by d4T-treatment. This finding confirms a previous observation by Garcia-Lerma et al, who reported a K65R mediated-resistance pathway to d4T in nine recombinants using genotypic change analysis (García-Lerma et al., 2003). However, K65R still remains rarely selected in d4T-treated isolates. In addition, this study compared the prevalence of K65R in subtype B versus C and found a lower prevalence of 10% in subtype B.

We then identified other mutations in both subtype B and C that were associated with TDF-treatment failure. We found that S68G was associated with K65R in subtype B while S68N was more common in subtype C isolates. S68N was also selected in d4T-treated sequences (0.4%); this was in the same sequences that developed K65R, and this may suggest that S68N is selected for during or after the development of K65R. The A62V and Y115F mutations occurred at a higher frequency in subtype C than in B. The TAMs were infrequently selected in TDF-treated isolates in both subtype B and C, although in subtype C, mutations involved in the TAMs II pathway were found in TDF-treated isolates. However, most of these were not found in the same sequences containing K65R. Furthermore, we also observed a high prevalence of NNRTI mutations in TDF-treated isolates. K70E was also found in limited TDF-treated isolates. Previous studies have also reported a low level of K70E in TDF-treated isolates (Delaugerre et al., 2008, Kagan et al., 2007). However, K65R still remained the main mutation that was associated with TDF exposure (Wensing et al., 2015).

The BN was used to identify mutations in addition to K65R involved in the pathway to TDF- resistance. In our subtype C sequences, the presence of K65R correlated with 5 NRTIs: A62V, S68G/N, Y115F and M184I, and negatively correlated with the TAMs. This was confirmed using the BN, showing a highly selected association of K65R with A62V, S68N, Y115F, Y181C, and weaker association with V106M and M230l. However, the BN did not show any direct association with M184I; this association

was through M230L. The strongly supported association of K65R with S68N has not been previously reported and suggests a novel TDF resistance-associated mutation. This mutation is currently not listed in International AIDS Society (IAS) or Stanford HIV-1 Database. Also the strongly supported association of K65R with Y181C and weakly supported association with V106M suggest that there is cross-resistance with the NNRTIs.

In subtype B, the TDF-resistance pathway, as seen in the BN, showed that only K65R was strongly supported to be associated with TDF exposure. While the K65R pathway showed a positive association with A62V, S68G, L100I and Y181C, it also showed a negative association with T215Y. This was expected as this has already been reported by Theys et al. However Theys' network did not show the association of K65R and A62V. The main difference between the subtype B and C BN was the association of K65R and S68N in subtype C and the association with S68G in subtype B; this needs further investigation. S68N was selected in subtype C only, while S68G was selected in both subtypes. More striking was the strongly selected association of K65R with Y115F, which was seen only in the subtype C network. This mutation has previously been reported to occur in subtype C isolates, although this is the first time that the association with K65R has been made (Miller, 2003, Skhosana et al., 2015, Stone et al., 2004).

Bayesian Network learning also identified a weakly supported direct association between TDF exposure and K70E. These results suggest that subtype C could have two different pathways to TDF resistance, one that involves K65R and the other that involves K70E. This is supported by the correlation results that showed a negative correlation between K65R and K70E, however this was not significant ($p=0.92$). As mentioned in chapter 2, K70E has been found in TDF treated isolates, albeit rarely. Furthermore, the K70E pathway included D67N and V106M, and weakly supported association with the NNRTI mutations A98G, K101A and K103S. D67N is one of the mutations in TAM II pathway, which are known to have an antagonistic effect with K65R (White et al., 2006a). Therefore, it is possible that the selection of TAM II pathway in TDF-treated sequences is driven by this association. K70E positively correlated with M184V which is in line with a previous study that reported that K70E was always accompanied by M184V (Kagan et al., 2007). In the BN, there is no direct association between K70E and M184V, however the association of these is mediated by mutation D67N.

We also investigated the structural changes that occur as a result of the K65R and associated mutations. A62V, K65R and S68G/N which are located in the finger region of RT, fold towards the active site in the palm to make contact with the incoming dNTPs (Sarafianos et al., 1999). More specifically, K65R and S68G/N are located in the β3 and β4 finger loop, while A62V is located outside of the β3 and β4 finger loop, suggesting that A62V does not make any direct contact with the dNTP. Furthermore, Y115F and Y181C are located in the palm region and they form part of the binding pocket. Both mutations decrease binding of the NRTIs, because they make contact with the template backbone. Y181C has

been reported to interact with NRTI resistance mutations and decrease binding of ATP as the excision substrate (Organization, 2010).

The modelling also showed that K65R does not form a salt bridge with the γ-phosphate, therefore affecting the binding of the inhibitors. Lastly, the interaction of K65R, Y115F and Y181C suggest that there is a high instability of hydrogen bonding leading to the disruption of NRTI binding.

In conclusion, our results suggest two distinct subtype C TDF-resistance pathways one involving K65R (which is the most common), and the other involving K70E. Both of these pathways are associated with NNRTI mutations (Y181C and V106M), suggesting some level of cross resistance. This is the first study to describe an association between K65R and S68N.The common pattern of K65R, Y115F and Y181C, which causes high level resistance to both TDF and EFV/NVP, raises concerns regarding the efficacy of the combination pill Truvada that is currently being used in the National ARV roll out. The effect of this combination needs further investigation using site-directed mutagenesis.

References

ACCOLA, M. A., HÖGLUND, S. & GÖTTLINGER, H. G. 1998. A putative α-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *Journal of Virology,* 72**,** 2072-2078.

AIDSINFO, A. 2013. Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents.

ANTINORI, A., TROTTA, M. P., NASTA, P., BINI, T., BONORA, S., CASTAGNA, A., ZACCARELLI, M., QUIRINO, T., LANDONIO, S. & MERLI, S. 2006. Antiviral efficacy and genotypic resistance patterns of combination therapy with stavudine/tenofovir in highly active antiretroviral therapy experienced patients. *Antiviral therapy,* 11**,** 233.

ARION, D., KAUSHIK, N., MCCORMICK, S., BORKOW, G. & PARNIAK, M. A. 1998. Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. *Biochemistry,* 37**,** 15908-17.

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

BOYER, P. L., SARAFIANOS, S. G., ARNOLD, E. & HUGHES, S. H. 2001. Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase. *J Virol,* 75**,** 4832-42.

BOYER, P. L., TANTILLO, C., JACOBO-MOLINA, A., NANNI, R. G., DING, J., ARNOLD, E. & HUGHES, S. H. 1994. Sensitivity of wild-type human immunodeficiency virus type 1 reverse transcriptase to dideoxynucleotides depends on template length; the sensitivity of drug-resistant mutants does not. *Proc Natl Acad Sci U S A,* 91**,** 4882-6.

BRENNER, B. G. & COUTSINOS, D. 2009a. The K65R mutation in HIV-1 reverse transcriptase: genetic barriers, resistance profile and clinical implications. *HIV Ther,* 3**,** 583-594.

BUONAGURO, L., TORNESELLO, M. & BUONAGURO, F. 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *Journal of virology,* 81**,** 10209-10219.

CLAVEL, F. & HANCE, A. J. 2004. HIV drug resistance. *New England Journal of Medicine,* 350**,** 1023-1035.

COFFIN, J. M. 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science,* 267**,** 483-9.

COUTSINOS, D., INVERNIZZI, C. F., MOISI, D., OLIVEIRA, M., MARTINEZ-CAJAS, J. L., BRENNER, B. G. & WAINBERG, M. A. 2011. A template-dependent dislocation mechanism potentiates K65R reverse transcriptase mutation development in subtype C variants of HIV-1. *PloS one,* 6**,** e20208.

COUTSINOS, D., INVERNIZZI, C. F., XU, H., MOISI, D., OLIVEIRA, M., BRENNER, B. G. & WAINBERG, M. A. 2009. Template usage is responsible for the preferential acquisition of the K65R reverse transcriptase mutation in subtype C variants of human immunodeficiency virus type 1. *J Virol,* 83**,** 2029-33.

COZZI-LEPRI, A., RUIZ, L., LOVEDAY, C., PHILLIPS, A. N., CLOTET, B., REISS, P., LEDERGERBER, B., HOLKMANN, C., STASZEWSKI, S., LUNDGREN, J. D. & EURO, S. S. G. 2005. Thymidine analogue mutation profiles: factors associated with acquiring specific profiles and their impact on the virological response to therapy. *Antivir Ther,* 10**,** 791-802.

DAS, K. & ARNOLD, E. 2013. HIV-1 reverse transcriptase and antiviral drug resistance. Part 2. *Current opinion in virology,* 3**,** 119-128.

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

DEFORCHE, K., CAMACHO, R. J., GROSSMAN, Z., SOARES, M. A., VAN LAETHEM, K., KATZENSTEIN, D. A., HARRIGAN, P. R., KANTOR, R., SHAFER, R., VANDAMME, A. M. & NON, B. W. 2008. Bayesian network analyses of resistance pathways against efavirenz and nevirapine. *AIDS,* 22**,** 2107-15.

DEFORCHE, K., SILANDER, T., CAMACHO, R., GROSSMAN, Z., SOARES, M. A., VAN LAETHEM, K., KANTOR, R., MOREAU, Y., VANDAMME, A. M. & NON, B. W. 2006. Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. *Bioinformatics,* 22**,** 2975-9.

DELANO, W. L. 2002. The PyMOL molecular graphics system.

DELAUGERRE, C., FLANDRE, P., MARCELIN, A., DESCAMPS, D., TAMALET, C., COTTALORDA, J., SCHNEIDER, V., YERLY, S., LEGOFF, J. & MORAND-JOUBERT, L. 2008. National survey of the prevalence and conditions of selection of HIV-1 reverse transcriptase K70E mutation. *Journal of medical virology,* 80**,** 762-765.

DEPARTMENT OF HEALTH 2014. National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults. Pretoria, South Africa.

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

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

DOUALLA-BELL, F., AVALOS, A., BRENNER, B., GAOLATHE, T., MINE, M., GASEITSIWE, S., OLIVEIRA, M., MOISI, D., NDWAPI, N., MOFFAT, H., ESSEX, M. & WAINBERG, M. A. 2006. High prevalence of the K65R mutation in human immunodeficiency virus type 1 subtype C isolates from infected patients in Botswana treated with didanosine-based regimens. *Antimicrob Agents Chemother,* 50**,** 4182-5.

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

ENGELMAN, A. & CHEREPANOV, P. 2012b. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat Rev Microbiol,* 10.

ETIEBET, M.-A. A., SHEPHERD, J., NOWAK, R. G., CHARURAT, M., CHANG, H., AJAYI, S., ELEGBA, O., NDEMBI, N., ABIMIKU, A. & CARR, J. K. 2013. Tenofovir-based regimens associated with less drug resistance in HIV-1-infected Nigerians failing first-line antiretroviral therapy. *Aids,* 27**,** 553-561.

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

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

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

GARCÍA-LERMA, J. G., MACINNES, H., BENNETT, D., REID, P., NIDTHA, S., WEINSTOCK, H., KAPLAN, J. E. & HENEINE, W. 2003. A novel genetic pathway of human immunodeficiency virus type 1 resistance to stavudine mediated by the K65R mutation. *Journal of virology,* 77**,** 5685-5693.

GARFORTH, S. J., LWATULA, C. & PRASAD, V. R. 2014a. The lysine 65 residue in HIV-1 reverse transcriptase function and in nucleoside analog drug resistance. *Viruses,* 6**,** 4080-4094.

GARFORTH, S. J., LWATULA, C. & PRASAD, V. R. 2014b. The lysine 65 residue in HIV-1 reverse transcriptase function and in nucleoside analog drug resistance. *Viruses,* 6**,** 4080-94.

GU, Z., ARTS, E. J., PARNIAK, M. A. & WAINBERG, M. A. 1995. Mutated K65R recombinant reverse transcriptase of human immunodeficiency virus type 1 shows diminished chain termination in the presence of 2',3'-dideoxycytidine 5'-triphosphate and other drugs. *Proc Natl Acad Sci U S A,* 92**,** 2760-4.

HARRICH, D. & HOOKER, B. 2002. Mechanistic aspects of HIV-1 reverse transcription initiation. *Reviews in medical virology,* 12**,** 31-45.

HECKERMAN, D., GEIGER, D. & CHICKERING, D. M. 1995. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning,* 20**,** 197-243.

HEMELAAR, J. 2012. The origin and diversity of the HIV-1 pandemic. *Trends in molecular medicine,* 18**,** 182-192.

HOFFMANN, C. J., LEDWABA, J., LI, J.-F., JOHNSTON, V., HUNT, G., FIELDING, K. L., CHAISSON, R. E., CHURCHYARD, G. J., GRANT, A. D. & JOHNSON, J. A. 2013a. Resistance to tenofovir-based regimens during treatment failure of subtype C HIV-1 in South Africa. *Antiviral therapy,* 18**,** 915.

HOFFMANN, C. J., LEDWABA, J., LI, J. F., JOHNSTON, V., HUNT, G., FIELDING, K. L., CHAISSON, R. E., CHURCHYARD, G. J., GRANT, A. D., JOHNSON, J. A., CHARALAMBOUS, S. & MORRIS, L. 2013b. Resistance to tenofovir-based regimens during treatment failure of subtype C HIV-1 in South Africa. *Antivir Ther,* 18**,** 915-20.

HU, W.-S. & HUGHES, S. H. 2012. HIV-1 reverse transcription. *Cold Spring Harbor perspectives in medicine,* 2**,** a006882.

HU, Z. & KURITZKES, D. R. 2011. Interaction of reverse transcriptase (RT) mutations conferring resistance to lamivudine and etravirine: effects on fitness and RT activity of human immunodeficiency virus type 1. *Journal of virology,* 85**,** 11309-11314.

HUANG, H., CHOPRA, R., VERDINE, G. L. & HARRISON, S. C. 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science,* 282**,** 1669-1675.

INHIBITORS, A. C. M. U. P. 2002. from Virtual and High-Throughput Screening McGovern, Susan L.; Caselli, Emilia; Grigorieff, Nikolaus; Shoichet, Brian K. *Journal of Medicinal Chemistry,* 45**,** 1712-1722.

INVERNIZZI, C. F., COUTSINOS, D., OLIVEIRA, M., MOISI, D., BRENNER, B. G. & WAINBERG, M. A. 2009. Signature nucleotide polymorphisms at positions 64 and 65 in reverse transcriptase favor the selection of the K65R resistance mutation in HIV-1 subtype C. *J Infect Dis,* 200**,** 1202-6.

IVES, N. J., GAZZARD, B. G. & EASTERBROOK, P. J. 2001. The changing pattern of AIDS-defining illnesses with the introduction of highly active antiretroviral therapy (HAART)in a London clinic. *J Infect,* 42**,** 134-9.

JOHNSON, V. A., BRUN-VÉZINET, F., CLOTET, B., CONWAY, B., D'AQUILA, R., DEMETER, L., KURITZKES, D., PILLAY, D., SCHAPIRO, J. & TELENTI, A. 2003. Drug resistance mutations in HIV-1. *Top HIV Med,* 11**,** 215-221.

JOHNSON, V. A., CALVEZ, V., GUNTHARD, H. F., PAREDES, R., PILLAY, D., SHAFER, R., WENSING, A. M. & RICHMAN, D. D. 2011. 2011 update of the drug resistance mutations in HIV-1. *Top Antivir Med,* 19**,** 156-64.

KAGAN, R. M., LEE, T.-S., ROSS, L., LLOYD, R., LEWINSKI, M. & POTTS, S. 2007. Molecular basis of antagonism between K70E and K65R tenofovir-associated mutations in HIV-1 reverse transcriptase. *Antiviral research,* 75**,** 210-218.

KAGAN, R. M., MERIGAN, T. C., WINTERS, M. A. & HESELTINE, P. N. 2004. Increasing prevalence of HIV-1 reverse transcriptase mutation K65R correlates with tenofovir utilization. *Antiviral therapy,* 9**,** 827-828.

KARIM, Q. A., KARIM, S. S. A., FROHLICH, J. A., GROBLER, A. C., BAXTER, C., MANSOOR, L. E., KHARSANY, A. B., SIBEKO, S., MLISANA, K. P. & OMAR, Z. 2010. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *science,* 329**,** 1168-1174.

KIRCHMAIR, J., DISTINTO, S., ROMAN LIEDL, K., MARKT, P., MARIA ROLLINGER, J., SCHUSTER, D., MARIA SPITZER, G. & WOLBER, G. 2011. Development of anti-viral agents using molecular

modeling and virtual screening techniques. *Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders),* 11**,** 64-93.
KONDO, E. & GÖTTLINGER, H. 1996. A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr. *Journal of virology,* 70**,** 159-164.
KOSTRIKIS, L. G., BAGDADES, E., CAO, Y., ZHANG, L., DIMITRIOU, D. & HO, D. D. 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *Journal of Virology,* 69**,** 6122-6130.
KUIKEN, C., HRABER, P., THURMOND, J. & YUSIM, K. 2008. The hepatitis C sequence database in Los Alamos. *Nucleic acids research,* 36**,** D512-D516.
KURITZKES, D. R., BASSETT, R. L., HAZELWOOD, J. D., BARRETT, H., RHODES, R. A., YOUNG, R. K. & JOHNSON, V. A. 2004. Rate of thymidine analogue resistance mutation accumulation with zidovudine-or stavudine-based regimens. *JAIDS Journal of Acquired Immune Deficiency Syndromes,* 36**,** 600-603.
LAURA, W., NELSO, M., MANDALI, S., BOWE, M., POWLE, T., GAZZARD, B. & STEBBING, J. 2008. The risks and incidence of K65R and L74V mutations and subsequent virologic responses. *Clinical infectious diseases,* 46**,** 96-100.
LOLE, K. S., BOLLINGER, R. C., PARANJAPE, R. S., GADKARI, D., KULKARNI, S. S., NOVAK, N. G., INGERSOLL, R., SHEPPARD, H. W. & RAY, S. C. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of virology,* 73**,** 152-160.
MARGOT, N. A., ISAACSON, E., MCGOWAN, I., CHENG, A. & MILLER, M. D. 2003. Extended treatment with tenofovir disoproxil fumarate in treatment-experienced HIV-1-infected patients: genotypic, phenotypic, and rebound analyses. *Journal of acquired immune deficiency syndromes (1999),* 33**,** 15-21.
MARSDEN, M. D. & ZACK, J. A. 2013. HIV/AIDS eradication. *Bioorganic & medicinal chemistry letters,* 23**,** 4003-4010.
MAUCK, C. K. & STRATEN, A. 2008. Using objective markers to assess participant behavior in HIV prevention trials of vaginal microbicides. *J Acquir Immune Defic Syndr,* 49**,** 64-9.
MENÉNDEZ-ARIAS, L. 2010. Molecular basis of human immunodeficiency virus drug resistance: an update. *Antiviral research,* 85**,** 210-231.
MEYER, P. R., MATSUURA, S. E., SO, A. G. & SCOTT, W. A. 1998. Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. *Proc Natl Acad Sci U S A,* 95**,** 13471-6.
MILLER, M. D. 2003. K65R, TAMs and tenofovir. *AIDS reviews,* 6**,** 22-33.
MILLER, M. D. 2004. K65R, TAMs and tenofovir. *AIDS Rev,* 6**,** 22-33.
MILLER, V., AIT-KHALED, M., STONE, C., GRIFFIN, P., MESOGITI, D., CUTRELL, A., HARRIGAN, R., STASZEWSKI, S., KATLAMA, C. & PEARCE, G. 2000. HIV-1 reverse transcriptase (RT) genotype and susceptibility to RT inhibitors during abacavir monotherapy and combination therapy. *Aids,* 14**,** 163-171.
MONINI, P., SGADARI, C., TOSCHI, E., BARILLARI, G. & ENSOLI, B. 2004. Antitumour effects of antiretroviral therapy. *Nature Reviews Cancer,* 4**,** 861-875.
MYLLYMÄKI, P., SILANDER, T., TIRRI, H. & URONEN, P. 2002. B-course: A web-based tool for Bayesian and causal data analysis. *International Journal on Artificial Intelligence Tools,* 11**,** 369-387.
NELSON, M. R., KATLAMA, C., MONTANER, J. S., COOPER, D. A., GAZZARD, B., CLOTET, B., LAZZARIN, A., SCHEWE, K., LANGE, J. & WYATT, C. 2007. The safety of tenofovir disoproxil fumarate for the treatment of HIV infection in adults: the first 4 years. *Aids,* 21**,** 1273-1281.
NYAMWEYA, S., HEGEDUS, A., JAYE, A., ROWLAND-JONES, S., FLANAGAN, K. L. & MACALLAN, D. C. 2013. Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Reviews in medical virology,* 23**,** 221-240.

ORGANIZATION, W. H. 2010. *Antiretroviral drugs for treating pregnant women and preventing HIV infection in infants: recommendations for a public health approach-2010 version*, World Health Organization.

PARIKH, U. M., BACHELER, L., KOONTZ, D. & MELLORS, J. W. 2006. The K65R mutation in human immunodeficiency virus type 1 reverse transcriptase exhibits bidirectional phenotypic antagonism with thymidine analog mutations. *Journal of virology,* 80**,** 4971-4977.

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

POND, S. J., RIDGEWAY, W. K., ROBERTSON, R., WANG, J. & MILLAR, D. P. 2009. HIV-1 Rev protein assembles on viral RNA one molecule at a time. *Proceedings of the National Academy of Sciences,* 106**,** 1404-1408.

PRESTON, B. D., POIESZ, B. J. & LOEB, L. A. 1988. Fidelity of HIV-1 reverse transcriptase. *Science,* 242**,** 1168-71.

REN, J. & STAMMERS, D. K. 2008. Structural basis for drug resistance mechanisms for non-nucleoside inhibitors of HIV reverse transcriptase. *Virus research,* 134**,** 157-170.

ROBERTS, J. D., BEBENEK, K. & KUNKEL, T. A. 1988. The accuracy of reverse transcriptase from HIV-1. *Science,* 242**,** 1171-3.

SARAFIANOS, S. G., DASI, K., DINGI, J., BOYER, P. L., HUGHES, S. H. & ARNOLD, E. 1999. Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at the polymerase active site. *Chemistry & biology,* 6**,** R137-R146.

SCHWEDE, T., KOPP, J., GUEX, N. & PEITSCH, M. C. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research,* 31**,** 3381-3385.

SHAFER, R. W. 2002. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clinical microbiology reviews,* 15**,** 247-277.

SHI, J., BLUNDELL, T. L. & MIZUGUCHI, K. 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol,* 310**,** 243-57.

SKHOSANA, L., STEEGEN, K., BRONZE, M., LUKHWARENI, A., LETSOALO, E., PAPATHANASOPOULOS, M. A., CARMONA, S. C. & STEVENS, W. S. 2015. High prevalence of the K65R mutation in HIV-1 subtype C infected patients failing tenofovir-based first-line regimens in South Africa. *PloS one,* 10**,** e0118145.

SLUIS-CREMER, N., SHEEN, C.-W., ZELINA, S., TORRES, P. S. A., PARIKH, U. M. & MELLORS, J. W. 2007. Molecular mechanism by which the K70E mutation in human immunodeficiency virus type 1 reverse transcriptase confers resistance to nucleoside reverse transcriptase inhibitors. *Antimicrobial agents and chemotherapy,* 51**,** 48-53.

STONE, C., AIT-KHALED, M., CRAIG, C., GRIFFIN, P. & TISDALE, M. 2004. Human immunodeficiency virus type 1 reverse transcriptase mutation selection during in vitro exposure to tenofovir alone or combined with abacavir or lamivudine. *Antimicrobial agents and chemotherapy,* 48**,** 1413-1415.

SUNPATH, H., WU, B., GORDON, M., HAMPTON, J., JOHNSON, B., MOOSA, M. Y., ORDONEZ, C., KURITZKES, D. R. & MARCONI, V. C. 2012a. High rate of K65R for antiretroviral therapy-naive patients with subtype C HIV infection failing a tenofovir-containing first-line regimen. *AIDS,* 26**,** 1679-84.

SUNPATH, H., WU, B., GORDON, M., HAMPTON, J., JOHNSON, B., MOOSA, Y., ORDONEZ, C., KURITZKES, D. R. & MARCONI, V. C. 2012b. High rate of K65R for ART naïve patients with subtype C HIV infection failing a TDF-containing first-line regimen in South Africa. *AIDS (London, England),* 26**,** 1679.

SVAROVSKAIA, E. S., FENG, J. Y., MARGOT, N. A., MYRICK, F., GOODMAN, D., LY, J. K., WHITE, K. L., KUTTY, N., WANG, R. & BORROTO-ESODA, K. 2008. The A62V and S68G mutations in HIV-1 reverse transcriptase partially restore the replication defect associated with the K65R mutation. *JAIDS Journal of Acquired Immune Deficiency Syndromes,* 48**,** 428-436.

TANG, M. W. & SHAFER, R. W. 2012. HIV-1 antiretroviral resistance. *Drugs,* 72**,** e1-e25.
TASTAN BISHOP, A., DE BEER, T. A. & JOUBERT, F. 2008. Protein homology modelling and its use in South Africa. *South African Journal of Science,* 104**,** 2-6.
TEMIZ, N. A. & BAHAR, I. 2002. Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. *Proteins: Structure, Function, and Bioinformatics,* 49**,** 61-70.
THEYS, K., VERCAUTEREN, J., ABECASIS, A. B., LIBIN, P., DEFORCHE, K., VANDAMME, A.-M. & CAMACHO, R. 2009. The rise and fall of K65R in a Portuguese HIV-1 Drug Resistance database, despite continuously increasing use of tenofovir. *Infection, Genetics and Evolution,* 9**,** 683-688.
TROTTA, M. P., BONFIGLI, S., CECCHERINI-SILBERSTEIN, F., BELLAGAMBA, R., D'ARRIGO, R., SOLDANI, F., ZACCARELLI, M., CONCETTA BELLOCCHI, M., LORENZINI, P. & MARCONI, P. 2006. Clinical and genotypic correlates of mutation K65R in HIV-infected patients failing regimens not including tenofovir. *Journal of medical virology,* 78**,** 535-541.
UNAIDS 2013. Report on the Global AIDS Epidemic.
UNAIDS 2016. Global AIDS Update.
VAN ZYL, G. U., LIU, T. F., CLAASSEN, M., ENGELBRECHT, S., DE OLIVEIRA, T., PREISER, W., WOOD, N. T., TRAVERS, S. & SHAFER, R. W. 2013a. Trends in Genotypic HIV-1 Antiretroviral Resistance between 2006 and 2012 in South African Patients Receiving First- and Second-Line Antiretroviral Treatment Regimens. *PLoS One,* 8**,** e67188.
VAN ZYL, G. U., LIU, T. F., CLAASSEN, M., ENGELBRECHT, S., DE OLIVEIRA, T., PREISER, W., WOOD, N. T., TRAVERS, S. & SHAFER, R. W. 2013b. Trends in genotypic HIV-1 antiretroviral resistance between 2006 and 2012 in South African patients receiving first-and second-line antiretroviral treatment regimens. *PloS one,* 8**,** e67188.
VON WYL, V., YERLY, S., BÖNI, J., BÜRGISSER, P., KLIMKAIT, T., BATTEGAY, M., BERNASCONI, E., CAVASSINI, M., FURRER, H. & HIRSCHEL, B. 2008. Factors associated with the emergence of K65R in patients with HIV-1 infection treated with combination antiretroviral therapy containing tenofovir. *Clinical infectious diseases,* 46**,** 1299-1309.
WENSING, A. M., CALVEZ, V., GÜNTHARD, H. F., JOHNSON, V. A., PAREDES, R., PILLAY, D., SHAFER, R. W. & RICHMAN, D. D. 2015. 2015 Update of the Drug Resistance Mutations in HIV-1. *Topics in Antiviral Medicine™***,** 132.
WHITCOMB, J. M., PARKIN, N. T., CHAPPEY, C., HELLMANN, N. S. & PETROPOULOS, C. J. 2003. Broad nucleoside reverse-transcriptase inhibitor cross-resistance in human immunodeficiency virus type 1 clinical isolates. *Journal of Infectious Diseases,* 188**,** 992-1000.
WHITE, K. L., CHEN, J. M., FENG, J. Y., MARGOT, N. A., LY, J. K., RAY, A. S., MACARTHUR, H. L., MCDERMOTT, M. J., SWAMINATHAN, S. & MILLER, M. D. 2006a. The K65R reverse transcriptase mutation in HIV-1 reverses the excision pheno type of zidovudine resistance mutations. *Antiviral therapy,* 11**,** 155.
WHITE, K. L., CHEN, J. M., FENG, J. Y., MARGOT, N. A., LY, J. K., RAY, A. S., MACARTHUR, H. L., MCDERMOTT, M. J., SWAMINATHAN, S. & MILLER, M. D. 2006b. The K65R reverse transcriptase mutation in HIV-1 reverses the excision phenotype of zidovudine resistance mutations. *Antivir Ther,* 11**,** 155-63.
WHITE, K. L., MARGOT, N. A., WRIN, T., PETROPOULOS, C. J., MILLER, M. D. & NAEGER, L. K. 2002. Molecular mechanisms of resistance to human immunodeficiency virus type 1 with reverse transcriptase mutations K65R and K65R+ M184V and their effects on enzyme function and viral replication capacity. *Antimicrobial agents and chemotherapy,* 46**,** 3437-3446.

Appendix A: K70E correlation with NRTIs and NNRTIs mutations

Table 1: K70E significant correlation table: NRTIs and NNRTIs Mutations that significantly correlated with K70E

Mutations	P-values	correlation coefficient
D67N	0.000**	0.315
V75I	0.009**	0.12
V75S	0.014*	0.112
V75T	0.000**	0.214
90I	0.031*	0.099
A98G	0.019*	0.108
K101A	0.000**	0.241
K101D	0.002**	0.144
K101R	0.044*	0.092
K103S	0.015*	0.111
V106M	0.008**	0.122
E122P	0.002**	0.144
139K	0015*	0.111
M184V	0.001**	0.152
G190A	0.006**	0.125
G190E	0.037*	0.097
196R	0.015*	0.111
203K	0014*	0.113
228I	0.002**	0.144
228R	0.000**	0.196
238T	0.047*	0.092

Appendix C: Bayesian Network causal arcs

Table B: Different types of arcs that can be found in causal model

Solid arc from A to B	A has direct causal influence to B (direct meaning that causal influence is not mediated by any other variable that is included in the study)
Dashed arc from A to B.	There are two possibilities, but we do not know which holds. Either A is cause of B or there is a latent cause for both A and B.
Dashed line without any arrow heads between A and B.	There is a dependency but we do not know whether A causes B or if B causes A or if there is a latent cause of them both the dependency (confounding).