



**Nucleoside Reverse Transcriptase Inhibitors-associated Mutations in the
RNase H Region of HIV-1 isolates in South African Adults and Children
failing Highly Active Antiretroviral Therapy**

Sinaye Ngcapu

Submitted in fulfilment of requirements of the degree Master of Medical Science in the Department of Molecular Virology, Faculty of Health Science, Nelson R. Mandela School of Medicine at the University of KwaZulu-Natal.

2012

Supervisor: Dr. Michelle L. Gordon

PREFACE

The experimental work described in this dissertation was performed in the Hasso Plattner Research Laboratory, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa. This work was supervised by Dr Michelle Lucille Gordon. This study represents original work by the author and has not been submitted in any form for any degree or diploma to any other university. When the work of others has been used, it is clearly acknowledged in the text.



Sinaye Ngcapu (Student)



Michelle Lucille Gordon (Supervisor)

Plagiarism:

Declaration

I Sinaye Ngcapu declare that:

- (i) The research reported in this thesis, except where otherwise indicated, is my original work.
- (ii) This thesis has not been submitted for any degree or examination at any other university.
- (iii) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
- (iv) This thesis 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 thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed: _____


Sinaye Ngcapu

Date: September 2012

Ethics Approval

Full ethics approval, from the Biomedical Research Ethics Committee of the Nelson R. Mandela School of Medicine, University of KwaZulu-Natal (Ref: BE 106/09) was obtained for this study.

Presentations

1. Ngcapu S, Ndung'u T and Gordon ML. *Nucleoside Reverse Transcriptase Inhibitors (NRTI)-associated Mutations in the RNase H Region in South African Adults and Children failing Highly Active Antiretroviral Therapy (HAART)*. 10th Annual Symposium on Antiviral Drug Resistance: Targets and Mechanisms, Richmond, Virginia, US, 15 – 18 November 2009. (poster presentation).
2. Ngcapu S, Ndung'u T and Gordon ML. *Nucleoside Reverse Transcriptase Inhibitors-associated Mutations in the RNase H Region in South African Adults and Children failing Highly Active Antiretroviral Therapy*. 5th SA AIDS Conference, held at the Durban International Convention Centre, Durban, South Africa in June 7-10, 2011 (poster presentation).
3. Ngcapu S, Ndung'u T and Gordon ML. *Bayesian Network Analyses of Nucleoside Reverse Transcriptase Inhibitors-associated Mutations in the RNase H Region in HIV-1 subtype C infected individuals*. 12th Annual Symposium on Antiviral Drug Resistance: Targets and Mechanisms, Hershey, Pennsylvania, US, 6 – 9 November 2011. (poster presentation).

Dedication

I dedicate this thesis to **Biziwe “Nonyameko” Ngcapu**, my late grandmother and best friend, for her bravery, strength and faith, and to Jesus, I could not have done this without you.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Michelle Lucille Gordon for her guidance, support and encouragement over the last three years. It was always reassuring to know that she had faith in my abilities, and patience with my progress when frustration would get the better of me. I am particularly grateful for the freedom and opportunity to pursue a project that was different from the norm and allowed me to explore the many facets of my interests in molecular virology. Under her watch, the boundaries of my abilities were tested and expanded as I developed further confidence in myself and matured into a well-rounded independent scientist. Thank you for being the person you are: kind, sensitive and thoughtful. You are unselfish always, putting others before yourself, making us feel special and important. It is a privilege and a pleasure to know you, “mother”.

I also owe a great deal of thanks to my colleagues in the lab, both past and present who helped me in various ways. Mrs Taryn Green and Jennifer Giandhari provided a great deal of unparalleled scientific insight and valuable suggestions on my research. Both were instrumental in teaching or exposing me to skills and techniques that would help me along the way, and I greatly thank them for their friendship and support. Thanks to Mrs Keshni Hiramani for her much-appreciated assistance and motivation with this project.

Special thanks as well to Prof Thumbi Ndung'u for believing in me and for the endless opportunities that you have provided me during my term at HPP. I was so blessed to be part of the team and it was really a great opportunity to be working with you.

I would like to thank my family and friends, for their unwavering support and confidence in me. I am forever in debt to my mother (Noludumo E. Ngcapu), who supported me both

emotionally and financially. Lastly, I would like to thank my love (Ntombikhona F. Maphumulo) and my angel (Sinesihle Ambesa Maphumulo) for their sacrifices and unconditional loving support they made in helping me to achieve my goal. Just knowing that they supported me and were always available to listen were all that I needed to keep persevering and to see this through to fruition.

This work was supported by an IMPAACT grant.

Abstract

Background: The South African national treatment program includes NRTIs in both first and second line highly active antiretroviral therapy regimens. Recently, mutations in the RNase H domain have been associated with resistance to NRTIs. Here we investigated the prevalence and association of RNase H mutations with NRTI resistance in isolates of HIV-1 subtype C infected individuals.

Methods: RNase H sequences from 134 NRTI treated (104 adults and 30 children) and 134 drug-naïve sequences (30 KZN isolates and 104 downloaded from the Los Alamos Database) were analyzed. Spearman's rank correlation and a Bayesian network were used to explore the relationship between mutations occurring within the RNase H domain and NRTI treatment.

Results: 130 of 134 samples clustered phylogenetically with HIV-1 subtype C, with one subtype A, two subtype B and two subtype D. All 30 sequences from HAART-naïve patients were classified as subtype C. Five mutations in the RNase H region had significantly higher frequency when comparing ART-naïve and NRTI-experienced patients. These were: (E438GKR, L517ISV, K527GENQR, E529DK and Q547HKR) (Table 1). Three mutations (E432D, A446SVY and Q507HK) showed decreased proportions in treatment-experienced isolates when compared to ART-naïve isolates. E438GKR was seen in 6.72% of treated versus only 0% of naïve isolates ($p= 0.0034$), L517IV was found in 17.16% of treated isolates versus 7.46% of naïve isolates ($p= 0.0245$). Similarly, K527GENQRS was found in 41.04% of treated isolates versus 26.12% of naïve isolates ($p= 0.0138$), and E529DK was more prevalent in treated (17.91%) when compared to 2.99% of naïve subtype isolates ($p < 0.001$). Finally, Q547HKR was seen in 5.22% of treated versus 0% of naïve subtype C patients ($p= 0.0144$). Interestingly, samples of twenty treatment experienced individuals that did not show of the classical NRTI mutations in the RT domain harbored E438GKR, L517ISV, K527GENQR, E529DK and Q547HKR.

Conclusion: Results obtained from this study suggested that drug resistance could be caused by mutations in the RNase H domain either alone (T470S), or in combination with mutations in the *pol* region (D67N and L491P). Phenotypic studies are required to understand the prevalence and impact of RNase H mutations, particularly E438GKR, T470S, L517ISV,

K527GENQR, E529DK and Q547HKR on NRTI resistance in HIV-1 subtype C as suggested by our data. Further studies using site-directed mutagenesis may also reveal the impact of these mutations on viral fitness.

Table of Contents

<i>Plagiarism</i>	Error! Bookmark not defined.
Ethics Approval.....	iv
Presentations.....	v
Dedication	vi
Acknowledgements	vii
Abstract	ix
List of Figures	xiv
List of Tables.....	xvii
Abbreviations	xviii
Chapter 1: Introduction	1
1. Introduction	1
1.2 Hypothesis.....	3
1.3 Project aims.....	3
1.3.1 Specific objectives.....	3
Chapter 2: Review of Literature.....	4
2.1 Human Immunodeficiency Virus.....	4
2.1.1 HIV-1 Life Cycle	6
2.1.2 Classification of HIV-1 Strains.....	8
2.2 The Structure and Function of RT.....	10
2.2.1 The Polymerase Domain.....	12
2.2.2 RNase H Domain	13
2.2.2.1 RNase H Primer Grip Motif.....	15
2.2.2.2 C-helix and Loop.....	17
2.2.2.3 Connection domain	17
2.2.2.4 Enzyme activity.....	18
2.2.2.4.1 Substrate Interactions	18
2.2.2.4.2 DNA 3' end-directed cleavage	19
2.2.2.4.3 RNA 5' end-directed cleavage.....	20
2.2.2.4.4 Internal cleavage	20
2.2.2.5 Use of different binding modes in reverse transcription.....	21
2.2.2.6 Cleavage specificity of RNase H	21
2.2.2.6.1 Nucleotide preferences.....	21
2.2.2.6.2 Removal of tRNA primer.....	22
2.2.2.6.3 Generation of PPT primer	23
2.2.2.7 RNase H and Antiretroviral Therapy	24

2.2.2.7.1.1	Reverse Transcriptase Inhibitors	24
2.2.2.7.1.2	The Nucleoside Reverse Transcriptase Inhibitors (NRTIs)	25
2.2.2.7.1.3	The Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)	27
2.2.2.7.2	RNase H and NRTI Resistance	29
2.2.2.7.2.1	RNase H dependent mechanism for NRTI drug resistance	29
2.2.2.7.3	RNase H as an Anti-viral Target	31
2.2.3	Drug Resistance Testing.....	31
2.2.4	Bayesian Network	33
2.2.5	Polymerase Chain Reaction (PCR)	38
2.2.6	DNA Sequencing.....	40
Chapter 3: Research Design and Methodology		43
3.1	Study Design	43
3.1.1	Study Population	44
3.1.2	Inclusion Criterion:	44
3.1.3	Exclusion Criterion:	44
3.1.4	Sample Collection:	45
3.2	Methods.....	45
3.2.1	RNA extraction:	45
3.2.2	Amplification and Sequencing of the RNase H region	46
3.2.2.1	HIV-1 Reverse Transcription	46
3.2.2.2	PCR Amplification.....	47
3.2.2.3	PCR Purification	48
3.2.2.3.1	ExoSAP-IT® PCR Clean-up Protocol	48
3.2.2.3.2	Gel Purification Protocol.....	48
3.2.3	Sequencing of the RNase H region	51
3.2.3.1a	The Plate Clean Up	51
3.2.3.1b	The Tube Clean Up	51
3.2.3.1c	Preparing Sequencing Products.....	52
3.2.4	Sequence Data Analysis	52
3.2.5	Statistical and Bioinformatics Analysis	53
3.2.5.1	Graphpad Prism®.....	53
3.2.5.2	B-Right.....	54
Chapter 4: Results		54
4.1	Optimization of PCR of the RNase H region.....	54
4.2	Characterization of the RNase H Domain.....	55
4.2.1	Phylogenetic Analysis	55
4.2.2	Amino acid sequence variation in the RNase H in subtype C.	57

4.2.3	Prevalence of RNase H mutations in NRTI-treatment-experienced and ART-naïve subtype C isolates.....	60
4.2.4	Comparison of RNase H mutations between Subtype B and C isolates in treatment naive versus treatment experienced patients.	66
4.2.4.1	Comparison of the RNase H sequence variability between subtype B and C isolates.	68
4.3	Bayesian Network Analysis	69
4.3.1	Dataset for Bayesian Network Learning	69
4.3.2	Treatment Experienced Bayesian Network Learning	72
Chapter 5:	Discussion and Conclusion.....	73
Chapter 6:	References	78
Chapter 7:	Appendices	99
Appendix I:	Preparation of Reagents	99
Appendix III:	Sequence Identities.....	101

List of Figures

Figure 1.	Diagram illustrating the structure of HIV.....	4
Figure 2.	Diagram of the HIV-1 genome (HXB2 strain).....	6
Figure 3.	Diagrammatic representation of the HIV-1 life cycle.....	7
Figure 4.	The HIV-1 classification according to groups and subtypes.....	9
Figure 5.	The global spread of HIV-1 subtypes and CRFs.....	10
Figure 6.	The crystal structure of HIV-1 reverse transcriptase.....	11
Figure 7.	Diagram representation of HIV-1 reverse transcriptase.....	14
Figure 8.	(A) Crystal structure of HIV-1 RT with an RNA/DNA hybrid bound.....	16
Figure 9.	The three discrete modes of RNase H cleavages.....	19
Figure 10.	Schematic representation showing the positions and base preferences flanking internal and RNA 5' end-directed RNase H cleavage sites.....	22
Figure 11.	Schematic representation showing the positions and sequences surrounding cleavage sites for the extended tRNA primer and PPT region of HIV-1.....	23
Figure 12.	HIV-1 genome sequence at the PPT (underlined) and U3 region.....	24
Figure 13.	Diagram illustrating the mechanism of NRTIs.....	25
Figure 14.	Diagram indicating pyrophosphorolysis, one of the major mechanisms associated with NRTI resistance.....	27

Figure 15.	Schematic representation demonstrating binding mechanism of action of non nucleoside reverse transcriptase inhibitor to the RT enzyme.....	28
Figure 16.	A Bayesian Network representing conditional and unconditional dependencies	34
Figure 17.	Annotated nelfinavir experience Bayesian network expressing direct association between nelfinavir-associated mutations, polymorphisms and nelfinavir treatment (eNFV).....	37
Figure 18.	Schematic drawing of the PCR cycle. (1) Denaturing at 94–96 °C. (2) Annealing at ~65 °C (3) Elongation at 72 °C.....	39
Figure 19.	Schematic representation of dye-labeled terminator reactions.....	41
Figure 20.	2.0 % Agarose gel picture loaded with 4 µl of Low DNA Mass Ladder stained with ethidium bromide.....	49
Figure 21.	Gel picture of the primer titration. Primers were titrated at the following concentrations: 0.1µM, 0.2µM, 0.4µM and 0.8µM (lanes 2 to 5 respectively)	54
Figure 22.	Maximum Likelihood Tree (1000 replicates for Bootstrap) showing the phylogenetic relations between all the sequences (164) that had all the RNase H codons.....	56
Figure 23a.	Amino acid sequence variation in the RNase H region from ART naïve isolates. Amino acid residues are represented as the HXB2 consensus.....	58

Figure 23b.	Amino acid sequence variation in the RNase H region from NRTI experience isolates. Amino acid residues are represented as the HXB2 consensus.....	59
Figure 24.	Frequency of the NRTI-associated mutations in the subtype B naïve, subtype B-treated, subtype C naïve, subtype C-treated.....	67
Figure 25.	Alignment of the consensus sequences of the HIV-1 group M subtypes with HXB2 and the KZN consensus sequences.....	69
Figure 26.	Annotated NRTI experience Bayesian network (BN) expressing direct association between the NRTI-associated mutations in the RNase H domain and NRTI treatment.....	71

List of Tables

Table 1.	Enzyme Mix used for Reverse Transcription.....	45
Table 2.	1 st and 2 nd Round PCR Master Mix.....	46
Table 3.	The amount of each DNA fragment using 4µl of Low DNA Mass Ladder...49	
Table 4.	RNase H mutations in HIV-1 Subtype C-infected naive versus experienced participants.....	61
Table 4 (cont).	RNase H mutations in HIV-1 Subtype C-infected naive versus experienced participants.....	62
Table 5.	Frequency of resistance-associated mutations in the RNase H domain in isolates that did not harbor any of the classical NRTI mutations.....	64
Table 5 (cont).	Frequency of resistance-associated mutations in the RNase H domain in isolates that did not harbor any of the classical NRTI mutations	65

Abbreviations

3TC	-	lamivudine
A	-	adenine
ABC	-	abacavir
ABI	-	applied biosystems
AIDS	-	acquired immunodeficiency syndrome
Ala	-	alanine
ART	-	antiretroviral therapy
Asn	-	asparagine
Asp	-	aspartic acid
ATP	-	adenosine triphosphate
AZT	-	zidovudine
BN	-	bayesian network
bp	-	base pairs
C	-	cytosine
C-terminal	-	carboxyl-terminus
CD4	-	cluster of differentiation 4
cDNA	-	complementary deoxyribonucleic acid
CRF	-	circulating recombinant forms
d4T	-	stavudine
DAPY	-	diarylpyrimidine
ddNTPs	-	dideoxy nucleoside or nucleotide triphosphates
DEPC	-	diethylpyrocarbonate
dH₂O	-	distilled water
ddI	-	didanosine

DNA	-	deoxyribonucleic acid
dNTPs	-	deoxyribonucleoside triphosphates
dT:	-	DNA thymidine
EFV	-	efavirenz
eNFV	-	nelfinavir treatment
env	-	envelope protein
FTC	-	emtricitabine
G	-	guanine
gag	-	group-specific antigen
GDE	-	genetic data environment
Gln	-	glutamine
Glu	-	glutamic acid
gp120	-	glycoprotein 120
HAART	-	highly active antiretroviral therapy
His	-	histidine
HIV	-	human immunodeficiency virus
HIV-1	-	human immunodeficiency virus type-1
ID	-	identity
Ile	-	isoleucine
L	-	litre
Lys	-	lysine
Met	-	methionine
Mg²⁺	-	magnesium ion
mM	-	millimolar
MMLV	-	Moloney Murine Leukemia Virus

Mn²⁺	-	manganese ion
mRNA	-	messenger ribonucleic acid
NRTIs	-	nucleoside reverse transcriptase Inhibitors
NNRTIs	-	non-nucleoside reverse-transcriptase inhibitors
N-terminal	-	amine-terminus
nef	-	negative regulatory factor
ng	-	nanogram
nm	-	nanomolar
NVP	-	navirapine
p24	-	protein 24
PCR	-	polymerase chain reaction
PIs	-	protease inhibitors
Pol	-	polymerase
PPT	-	polypurine tract
R	-	the terminal repeated sequence at 3' end
rA:	-	ribosomal adinine
rev	-	regulator of virion
RNA	-	ribonucleic acid
RNase H	-	ribonuclease H
rpm	-	revolutions per minute
RT	-	reverse transcriptase
SARCS	-	South African Resistance Cohort Study
ssRNA	-	single-stranded ribonucleic acid
T	-	thymidine
T cell	-	thymus lymphocytes

TAMs	-	thymidine analog mutations
tat	-	trans-activator of transcription
Thr	-	threonine
tRNA^{lys3}	-	lysine- transfer ribonucleic acid
tRNA	-	transfer ribonucleic acid
Tyr	-	tyrosine
U	-	uracil
UNAIDS	-	joint united nations programme on HIV/AIDS
vif	-	viral infectivity factor
vpr	-	viral protein R
vpu	-	viral protein U
μ	-	micro
μM	-	micromolar
μL	-	microlitre
μg	-	microgram
α	-	alpha
β	-	beta
[]	-	concentration
°C	-	degrees celsius
-C	-	carbon
-OH	-	hydroxyl group
e.g.	-	for example

Chapter 1: Introduction

1. Introduction

South Africa is currently one of the most affected countries by the HIV-1 subtype C epidemic in the world which raises an urgent need for an efficient treatment program. The global report on AIDS epidemic conducted by UNAIDS in 2010, approximated that 5.6 million out of 49 million South Africans were living with the HIV infection. Of these, 3 million were women above 15 years and 220 000 children (UNAIDS and WHO, 2010). Subtype-C is the major circulating HIV-1 subtype in Southern African countries where HIV-1 prevalence is the highest in the world. It is also responsible for 50% of HIV infection worldwide and acquires multi-drug resistance more rapidly than any other HIV-1 subtype (UNAIDS and WHO, 2010).

South Africa's reaction to the HIV and AIDS epidemic has evolved significantly over the years since the 2004 antiretroviral rollout. According to The National Comprehensive HIV and AIDS Plan in 2008, there has been a significant growth in numbers of South African patients receiving highly active antiretroviral therapy (HAART), with over 800,000 people on treatment in the public sector alone (South African Department of Health, 2008).

The South African national treatment program includes two nucleoside reverse transcriptase inhibitors (NRTIs) and one non-nucleoside reverse transcriptase inhibitor (NNRTI) in the first line regimen and two NRTIs and one protease inhibitor (PI) in the second line regimen (South African Department of Health, 2010). However, the efficacies of these drugs are

limited by HIV-1 drug resistance, which is usually caused by mutations in the protease and reverse transcriptase (RT) regions (Mracna *et al.*, 2001). Recently, mutations in the DNA polymerase domain of RT as well as in the RNase H domain have been associated with resistance to NRTIs. Mutations in the RNase H domain could disturb the interactions of RT with its nucleic acid substrate causing both RNase H and polymerase activity to be affected. This has an impact on replication. For example, some mutations (T473A) can terminate HIV-1 replication whereas others (N474A, K476A and Q500A) can either reduce the virus concentration or indirectly affect the positioning of the template-primer substrate at the RNase H active site (Jacobo-Molina *et al.*, 1993). It has also been shown that by site-directed mutagenesis, amino acid substitutions in the RNase H domain affect the binding of enzyme/substrate, resulting in a decrease in the RNase H activity, conferring high-level resistance to AZT and d4T by a mechanism that allows more time for NRTI excision (Julias *et al.*, 2004). However, the RNase H domain has not been fully characterized in the context of drug resistance since previous studies mainly concentrated on the N-terminal polymerase domain (at sites where the drugs target) and therefore only analyzed up to 334 amino acids in the RT region (Santos *et al.*, 2008).

As d4T (in 2010, d4T was substituted with tenofovir) and AZT were in the first and second line regimens in South Africa, mutations related to resistance to these drugs will have important public health implications. This is in addition to existing concerns such as low genetic barrier to resistance of certain drugs; poor adherence; beliefs and knowledge about medication; social stability and support; side effects and poor self-efficacy. Other medical concerns also include travelling long distances to the drug collection centres; monitoring and support systems including psychological support; difficulty of transportation and lack of funds for transportation (especially in rural areas); stigma and discrimination; food security

and nutrition; poor knowledge on how well the regimen works with daily activities and poor adherence by patient receiving HAART.

Unless an effective regimen and good patient management is implemented, the likelihood of drug resistant variants developing will increase and limit the effectiveness of antiviral therapy. If we have more information about drug resistance in subtype C, this information could be used by policy-makers to improve treatment guidelines.

1.2 Hypothesis

We hypothesize that mutations in the RNase H domain are associated with resistance to the NRTIs.

1.3 Project aims

The aim of the project was to characterize the RNase H domain of the HIV-1 reverse transcriptase gene of HIV-1 subtype C in South African adults and children failing HAART and to also correlate mutations in the RNase H domain with NRTI resistance.

1.3.1 Specific objectives

- To sequence and analyse the RNase H domain from NRTI treated but virologically failing individuals and HAART-naïve individuals.
- To establish whether the presence of mutations in the RNase H domain correlate with the appearance of NRTI-associated mutations.

Chapter 2: Review of Literature

2.1 Human Immunodeficiency Virus

Human immunodeficiency virus type 1 (HIV-1), the causative agent of AIDS, is a lentivirus and a member of the Retroviridae family (Douek *et al.*, 2009, Weiss, 1993). 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) (McGovern *et al.*, 2002).

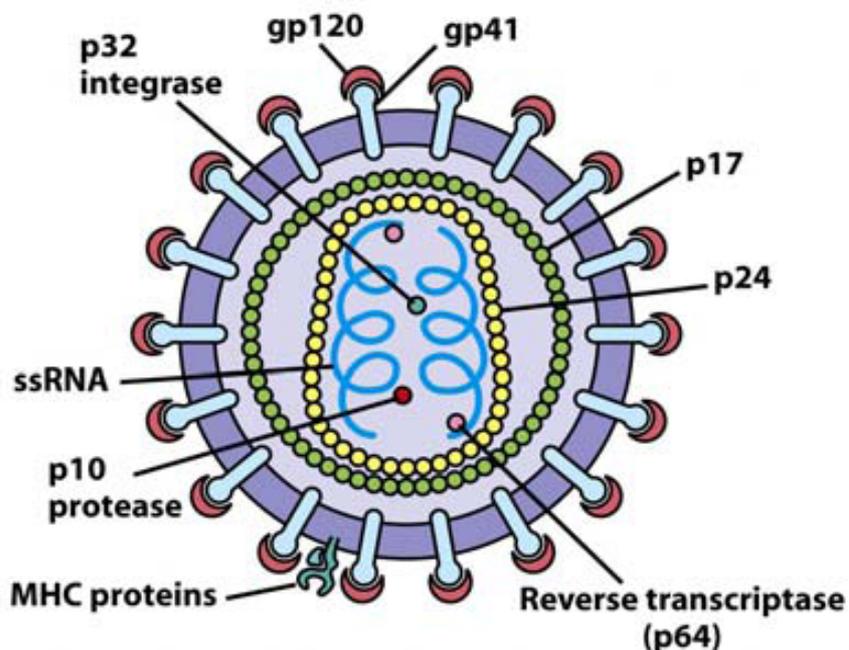


Figure 1. **The structure of HIV.** The proteins gp120 and gp41 together make up the spikes that project from HIV particles, while p17 forms the matrix and p24 forms the capsid (Kuby, 2007).

The ssRNA has nine genes, three of which are *gag*, *pol*, and *env* that comprise information required to form the structural proteins for new virus particles (Gibbs *et al.*, 1995). The *gag* gene encodes structural proteins of the conical capsid that can be cleaved by the viral protease into nucleocapsid (p6 and p7), matrix (p17) and the capsid (p24) (Gibbs *et al.*, 1995, Vingerhoets *et al.*, 2009). The nucleocapsids shield the RNA from degradation by nucleases, whereas the viral protein p17 provides a protective matrix for the virion particles. The *pol* region encodes a protein that contains four enzymes: protease, reverse transcriptase, RNase H and an integrase (Gibbs *et al.*, 1995, Vingerhoets *et al.*, 2009). The *env* gene codes for glycoprotein (gp160) that is broken down by the viral protease to form glycoproteins (gp120 and gp41) (Gibbs *et al.*, 1995, Vingerhoets *et al.*, 2009). In addition six regulatory genes *tat*; *rev*; *nef*; *vif*; *vpr* and *vpu* are also enclosed within the virion particles (Figure 2) (Gibbs *et al.*, 1995, Vingerhoets *et al.*, 2009). These genes are associated with deactivation of antiviral defence in host cells, enhancing viral replication and production of new virions (Vingerhoets *et al.*, 2009).

Tat (p16 and p14) are transcriptional transactivators comprising of 86 to 101 amino acids that induce the HIV virion production (Kuiken *et al.*, 2008, Allen and Altfeld, 2003). Rev (p19) is a regulatory protein that controls the export rate of mRNA from nucleus and cytoplasm. It consists of 116 amino acids including nuclear localization sequence (NLS) and nuclear export sequence (NES) (Pond *et al.*, 2009, Pollard and Malim, 1998, Zheng *et al.*, 2005). Vif (p23), another regulatory protein, prevents the function of APOBEC3G (a cell protein which deaminates DNA: RNA hybrids and/or interferes with Pol proteins (Zheng *et al.*, 2005, Gandhi *et al.*, 2008). Nef (p27) down-modulates cluster designation 4 (CD4) + T cells and allows infection by manipulating the host's cellular system. It is synthesized early in the HIV life cycle and is significant in the survival or replication of the virus (Lichterfeld *et*

al., 2004, Collins *et al.*, 1998). Nef also helps the virus to elude the host immune defense, resulting in AIDS disease progression (Stumptner-Cuvelette *et al.*, 2001, Cohen *et al.*, 1999, Adnan *et al.*, 2006). The Vpu protein (p16) promotes the release of new virus particles from infected cells and necessitates envelope maturation. It is an accessory protein with 81 amino acid residues and down-regulates CD4 + T cell (Pollard and Malim, 1998, Le Rouzic and Benichou, 2005, Kuiken *et al.*, 2008, Neil *et al.*, 2008).

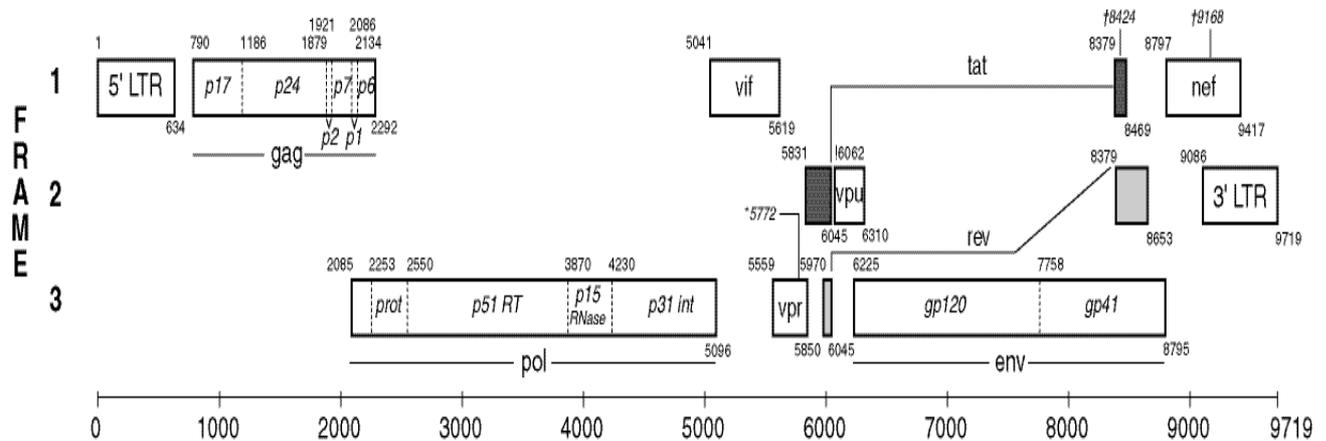


Figure 2. **The HIV-1 genome (HXB2 strain)** (Korber, 1998).

2.1.1 HIV-1 Life Cycle

HIV-1 infection occurs when the virus encounters a surface molecule called a CD4 + T cell. The CD4+T cells may be the major cell type that is infected by HIV viruses but other immune cells (monocytes and macrophages) without CD4 molecules on the surface are also infected by the virus (Chan *et al.*, 1997). The fusion of a HIV cell to the host cell surface leads to injection of the HIV RNA and various enzymes, including reverse transcriptase, integrase, ribonuclease and protease into the host cell (Figure 3) (Chan *et al.*, 1997, Wyatt *et al.*, 1998).

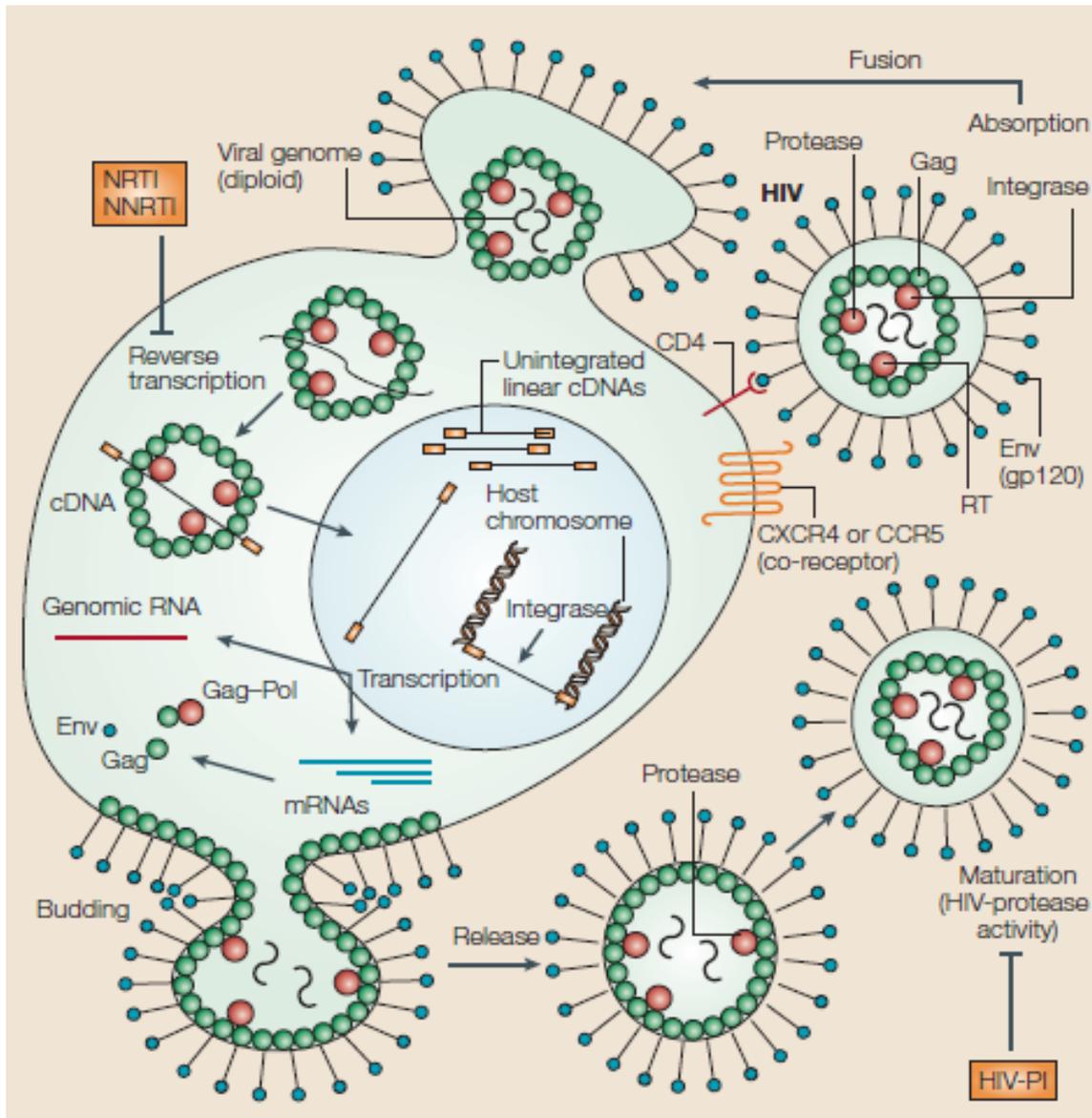


Figure 3. **Diagrammatic representation of the HIV-1 life cycle.** HIV infects the CD4 T cells using gp120 of the envelope, the viral RNA is then released inside the host cells. HIV RNA is reverse transcribed into cDNA which is then integrated into the host's DNA. Viral DNA is then synthesised into new viral RNA used to make viral proteins followed by protein translation inside the cytoplasm and a new, immature, HIV virus is released and infects more cells (Monini *et al.*, 2004).

Shortly after the viral capsid enters the cell, reverse transcription takes place, whereby viral RNA is reverse transcribed into complementary DNA (cDNA) molecule and the ribonuclease enzyme cleaves the ssRNA genome during the synthesis of cDNA. The DNA polymerase

replicates the cDNA into an *antisense* DNA to form viral DNA. The microtubule based transport moves the viral DNA across the cell's nucleus which is then integrated into a host chromosome. Inside the host chromosome, the viral DNA is integrated into the host DNA; this is catalyzed by integrase (Hiscott *et al.*, 2001, Pollard *et al.*, 1998).

The integrated viral DNA is then transcribed into messenger RNA (mRNA) using host cell's enzymes. After transcription of mRNA, it is spliced and exported from the nucleus into the cytoplasm. In the cytoplasm, pieces of mRNA are translated into elongated chains of viral protein (*tat* and *rev*) and enzymes (Hiscott *et al.*, 2001, Pollard *et al.*, 1998). The viral RNA and newly synthesized proteins gather to the surface of the plasma membrane and form immature viral particles and buds off from the host cell (Figure 3). As the immature virion buds off from the host cell, it acquires a phospholipid envelope that includes both cellular and HIV proteins from the host cell's membrane. The virus matures when the viral protease cleaves the precursor proteins into individual functional proteins and enzymes. Proteins gather at the surface to form infectious virions (Gelderblom *et al.*, 1997, Pollard *et al.*, 1998).

2.1.2 Classification of HIV-1 Strains

HIV is classified into two distinct types of HIV-1 and HIV-2. These are separated into four groups (Figure 4), the "main or major" group M, the "outlier" group O and recently discovered groups, N and P (Smith *et al.*, 1994, Osmanov *et al.*, 2002, Thomson 2002, Perrin *et al.*, 2003).

According to the nucleotide sequence analyses of the *env* and *gag* genes, at least nine genetically distinct HIV-1 group M subtypes have been discovered. These are: A, B, C, D, F, G, H, J and K. In addition to these, circulating recombinant forms (CRFs) that are produced

when two different subtypes infect a person at the same time and recombination of their genetic material recombination occurs (Osmanov *et al.*, 2002, Perrin *et al.*, 2003, Smith *et al.*, 1994, Thomson 2002).

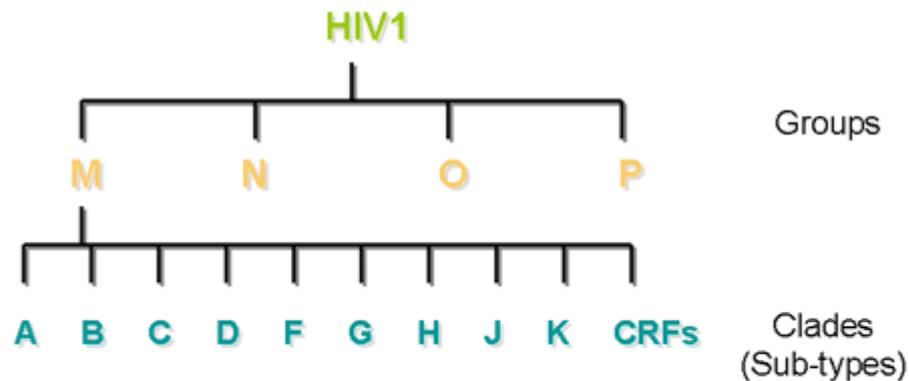


Figure 4. **The HIV-1 classification according to groups and subtypes** (Peeters, 2000).

The HIV-1 subtypes and CRFs are unequally spread all around the world, with the most predominant being subtypes B and C. Subtype A and CRF A/G are widely distributed in West and Central Africa, with A and D predominantly found in central and eastern Africa. Subtype B has been the major subtype in North America, Latin America and the Caribbean, Europe, Japan and Australia (Perrin *et al.*, 2003). Subtype C is largely predominant in southern and eastern Africa, India and Nepal. Subtype G and CRF A/G is found in western and eastern Africa and central Europe, Subtype H has only been observed in Central Africa; J only in Central America; and K only found in the Democratic Republic of Congo and Cameroon. A/E is found in south-east Asia, but developed in central Africa; F has been prevalent in central Africa, South America and eastern Europe (Figure 5) (Osmanov *et al.*, 2002, Perrin *et al.*, 2003, Smith *et al.*, 1994, Thomson 2002).

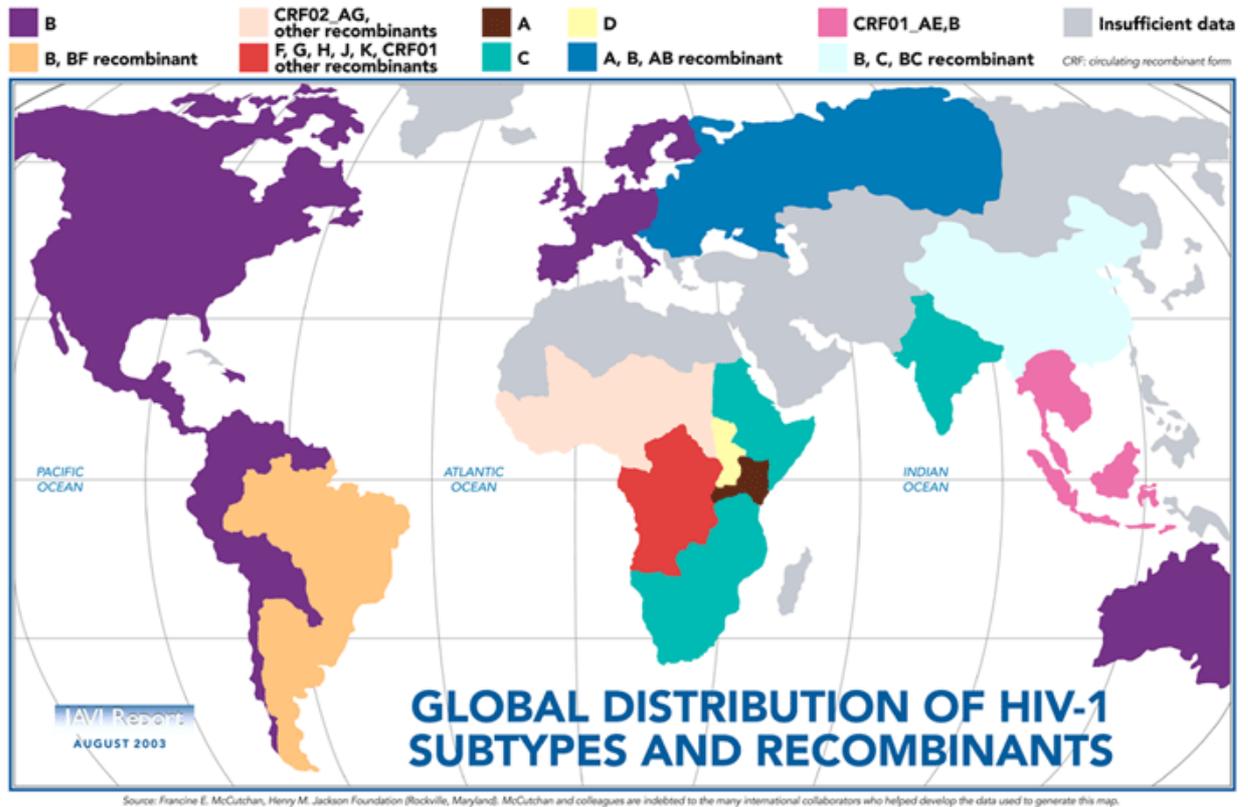


Figure 5. The global spread of HIV-1 subtypes and CRFs (Peeters, 2000).

2.2 The Structure and Function of RT

HIV-1 contains reverse transcriptase (RT) which is a RNA-dependent DNA polymerase enzyme responsible for synthesizing single-stranded RNA into double-stranded DNA and cleaving the RNA template strand of RNA/DNA replication intermediates using its RNase H activity (Gilboa *et al.*, 1979). The DNA polymerase and RNase H activities are situated within two large subunits found in the enzyme, the polymerase domain p51 (codons 1 – 440) and the ribonuclease H domain p66 (codons 1 - 560) linked together by the connection domain (codons 289–440) (Telesnitsky *et al.*, 1993). The p51 (51 kDa) subunit is formed

from the extensive p66 (66 kDa) subunit by protease-catalysed cleavages at the p51 RNase H cleavage site, resulting in the removal of a 15 kDa RNase H domain (Figure 6) (Telesnitsky *et al.*, 1993).

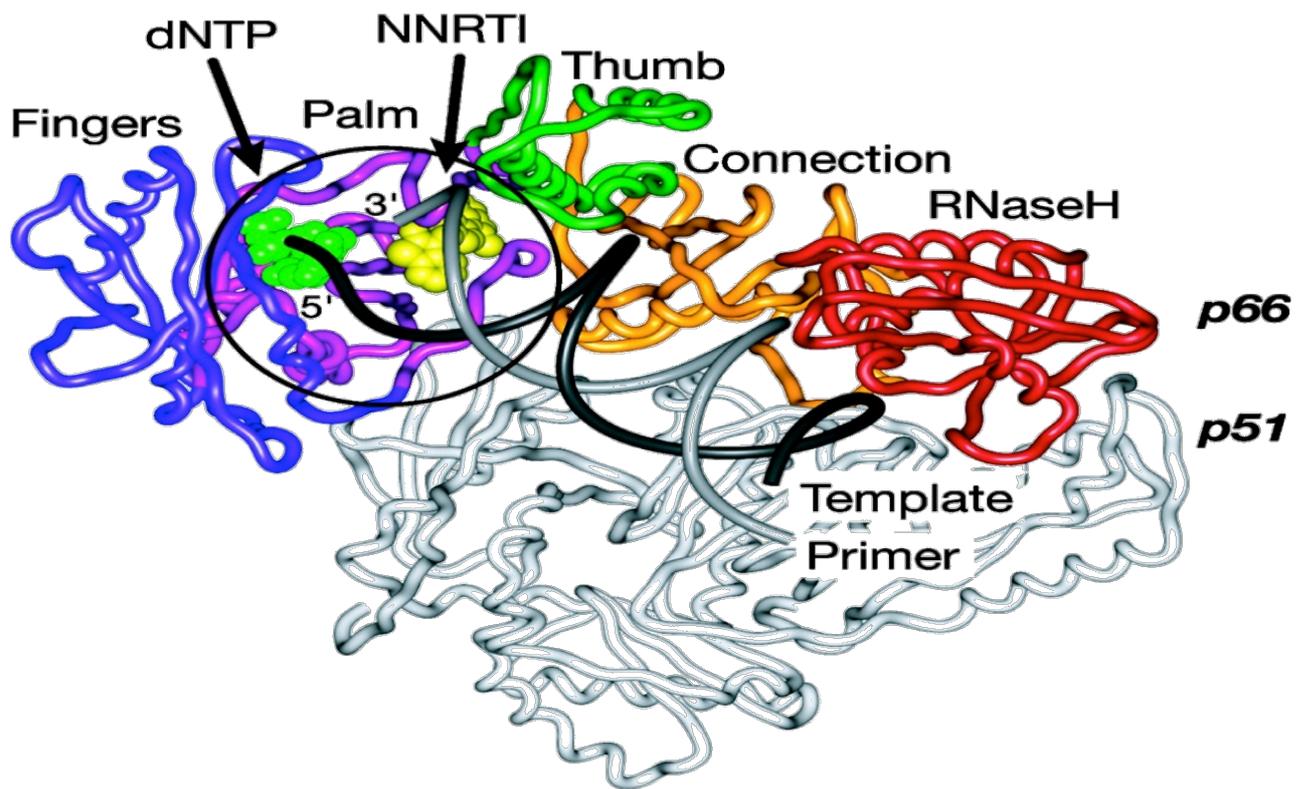


Figure 6. **The crystal structure of HIV-1 reverse transcriptase.** The polymerase, connection and RNase H domains of p66 are shown in blue, yellow and red, respectively; whilst p51 is shown in silver. The RNA template and DNA primer strands are shown in black and gray, respectively (Sarafinos *et al.*, 2001).

The polymerase domain of the p66 subunit consists of three sub-domains referred to as fingers (codons 118-155), palm (codons 156-237), and thumb (codons 238-289) that look like a human right hand when demonstrated using crystal structure (Huang *et al.*, 1998, Jacobo-Molina *et al.*, 1993, Kohlstaedt *et al.*, 1992, Rodgers *et al.*, 1995, Sarafianos *et al.*, 1999). The p66 subunit interact with its sub-domains using a comprehensive network of collision in

the sugar-phosphate backbone of the DNA strand (Huang *et al.*, 1998, Jacobo-Molina *et al.*, 1993, Kohlstaedt *et al.*, 1992, Rodgers *et al.*, 1995, Sarafianos *et al.*, 1999). The orientation and spatial arrangement of each subunit within the HIV-1 RT heterodimer differs, as a result, the surface on p51 involves different amino acid residues (Lys395 and Glu396) than on p66 (Gly359, Ala360, His361, Thr473, Asn474, Gln475, Lys476, Tyr501, and Ile505) for RT stability in the virion (Abram and Parniak, 2005, Sarafianos *et al.*, 1999).

2.2.1 The Polymerase Domain

The enzyme DNA polymerase found in the polymerase domain of the p51 subunit catalyzes the polymerization of both RNA dependent DNA and DNA dependent DNA. A stable HIV-RT interaction at the primer grip is crucial for the catalytic activity of polymerase. The stability between the enzyme and template is due to the van der Waals forces provided by a set of residues (Asp185, Asp186, Met230, Lys263, Glu258 and Asn255) that interacts with the DNA primer strand (Fleming *et al.*, 2002, Guitérrez-Rivas *et al.*, 1999, Jaeger *et al.*, 1998). These residues provide the van der Waals forces through donating their hydrogen bonds; the first set of amino acids (Asp185, Asp186, and Met230) donates the hydrogen bonds to the 3'OH of the growing DNA primer strand and renders six sets of van der Waals forces. The second set of amino acids (Lys263, Glu258, and Asn255) gives hydrogen bonds to the phosphate backbone and provides four sets of van der Waals forces to stabilize the polymerase domain (Fleming *et al.*, 2002, Guitérrez-Rivas *et al.*, 1999, Jaeger *et al.*, 1998).

2.2.2 RNase H Domain

RNase H (codons 441–560) is located in the p66 subunit and its activity is crucial in HIV-1 replication as it is responsible for cleaving the RNA template strand of RNA/DNA replication intermediates (Gilboa *et al.*, 1979). The hairpins in the template can delay the RT and such delays alleviate strand transfers and encourage RNase H catalytic activity (DeStefano *et al.*, 1994b, DeStefano *et al.*, 1992, Lanciault *et al.*, 2006, Peliska *et al.*, 1992). Its catalytic activity can occur in two modes, mainly the polymerization-dependent and –independent modes (Figure 7). Polymerase-dependent RNase H cleavage occurs during the synthesis of the minus-strand when the 3' end of the primer occupies the polymerase site between template amino acids 18 and 19 upstream (Figure 7a) and initiates the degradation of the RNA genome (DeStefano *et al.*, 1994b, DeStefano *et al.*, 1992, Lanciault *et al.*, 2006, Peliska *et al.*, 1992). The polymerase-independent RNase H cleavage occurs during the synthesis of the plus-strand and without the interaction between the 3' end of the primer and the polymerase active site (Figure 7b). Polymerase-independent RNase H cleavage can also produce the polypurine tract (PPT) primer (Figure 7c) and also cleaves the tRNA^{lys3} used in the synthesis of a DNA minus strand. The PPT RNA primer is used for the synthesis of the DNA plus strand (Furfine *et al.*, 1991a, Furfine *et al.*, 1991b, Gopalakrishnan *et al.*, 1992, Gotte *et al.*, 1995, Sarafianos *et al.*, 1999).

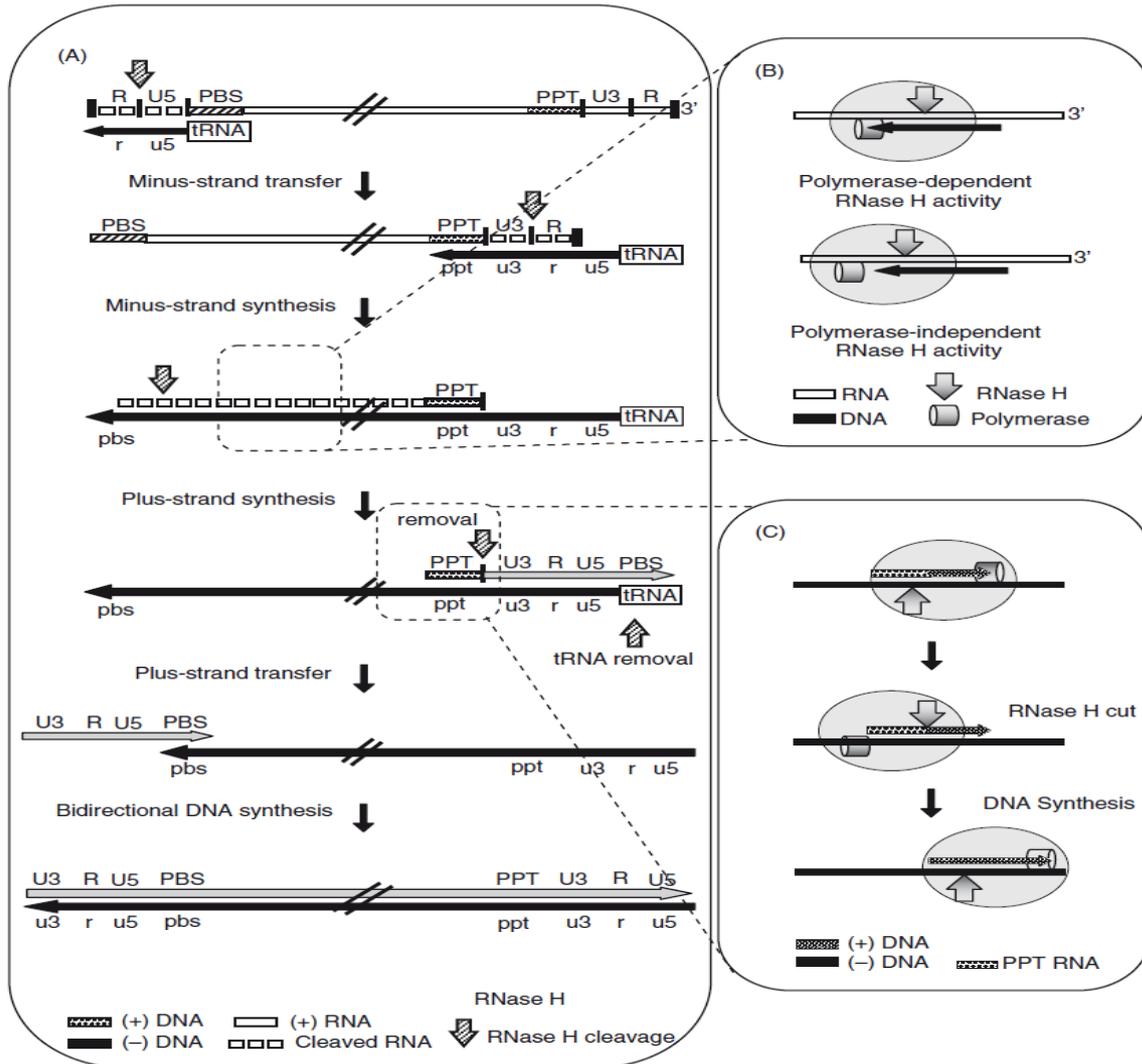


Figure 7. **Diagrammatic representation of HIV-1 reverse transcriptase.** (A) A stepwise process including DNA polymerization, RNA cleavages and DNA strand transfers that contribute to the synthesis of single stranded RNA into double-stranded DNA. (B) RNase H catalytic activity can occur in two modes, mainly the polymerization-dependent and –independent modes. Polymerase-dependent RNase H cleavage occurs when 3' end of the primer occupies the polymerase site and polymerase-independent RNase H cleavage occurs without the interaction between 3' end of the primer and the polymerase active site. (C) Degradation of the PTT primer used for plus-strand synthesis (Ehteshami and Götte, 2008).

Many studies have demonstrated that RNase H cleavage involves the recognition of specific sequences. Uncontrolled degradation of the RNA genome can cause synthesis termination by disassociating the primer from the template strand. Also insufficient RNase H cleavage can

either allow the remaining RNA strand to decrease DNA plus-strand synthesis or disturb the primer removal (Finston *et al.*, 1984, Fuentes *et al.*, 1995; Huber *et al.*, 1990, Lou *et al.*, 1990, Powell *et al.*, 1996, Pullen *et al.*, 1990, 1992, 1993; Schultz *et al.*, 1995, Smith *et al.*, 1992).

2.2.2.1 RNase H Primer Grip Motif

The RNase H primer grip motif is situated close to the RNase H active site that contacts the DNA primer strand at positions proportional to the RNA strand scissile phosphate. Several amino acid residues (Thr473, Asn474, Gln475, Lys476, Tyr501 and Ile505) at the RNase H domain form the primer grip motif which is known to be conserved (Figure 8) (Lorain *et al.*, 2005, Sarafianos *et al.*, 2001). The primer grip stabilizes the attaching and positioning of the template at both the DNA polymerase and RNase H catalytic site, for optimal catalysis. The RNase H primer grip motif at Tyr501 connects with the phosphate backbone of the DNA nucleotides that are base paired with the RNA nucleotides and Gln475 connects with the deoxyribose rings of the DNA nucleotides that are base paired the RNA strand (McWilliams *et al.*, 2006, Sarafianos *et al.*, 2001).

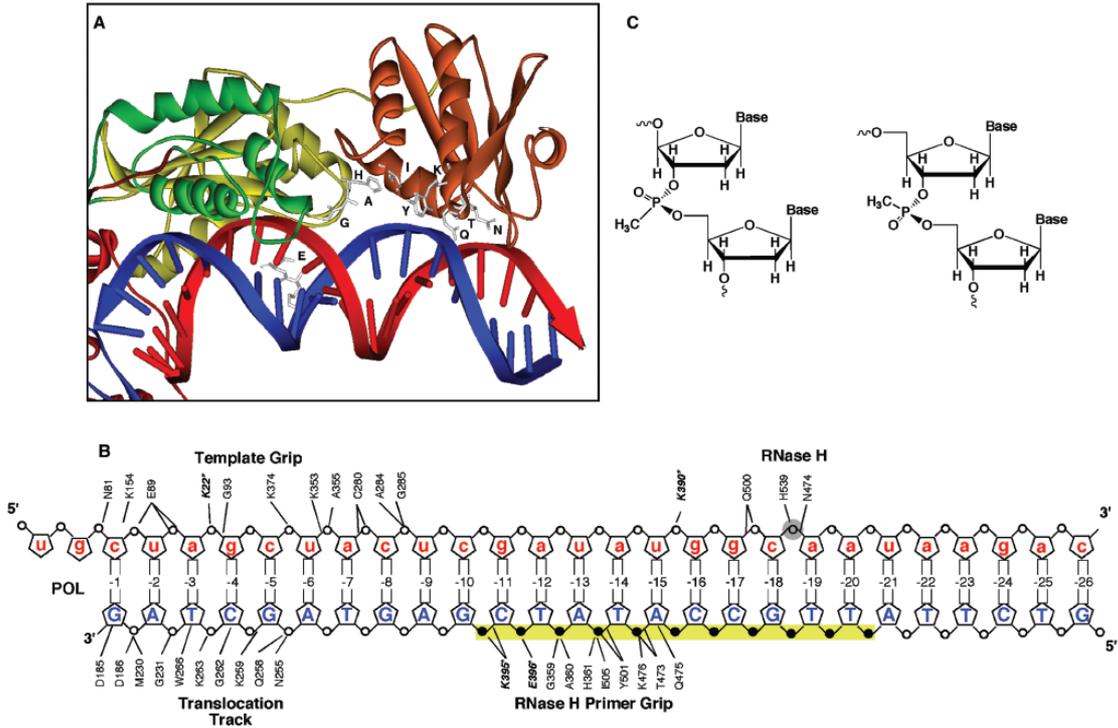


Figure 8. (A) Crystal structure of HIV-1 RT with an RNA/DNA hybrid bound. The letter formats represent amino acid residues and the DNA and RNA strands are in blue and red, respectively. (B) Representation of polymerase domain (template grip and translocation track), and RNase H active site. Shaded circles represent the DNA phosphates aimed to contact with the RNase H primer grip residues. (C) Chemical structure methylphosphonate Rp and Sp stereoisomers, where a methyl group replaces oxygen of the phosphate (Sarafianos *et al.*, 2001).

It has been shown that mutations in the connects with residues (Gln475 and Tyr501) can reduce RNase H activity, modify RNase H cleavage specificity and decrease the efficiency of strand transfer and initiation of DNA synthesis (Finston *et al.*, 1984, Fuentes *et al.*, 1995, 1991a, 1991b; Huber *et al.*, 1990, Lou *et al.*, 1990, Powell *et al.*, 1996, Pullen *et al.*, 1990, 1992, 1993; Schultz *et al.*, 1995, Smith *et al.*, 1992). In addition, mutation Gln475 and/or Tyr501 in association with mutations in the PPT sequence can reduce the rate RNase H degrades the PPT primer used for DNA plus strand synthesis (Figure 7) (Furfine *et al.*, 1991a, 1991b).

2.2.2.2 C-helix and Loop

Most of the information on C-helix and loop have been obtained from structures of the Moloney Murine Leukemia Virus (MMLV). The MMLV C-helix consists of 5 β -strands and 4 α -helices, one more α -helix than HIV-1 (Lim *et al.*, 2002). From this model, we see that the C-helix residues make contact with the RNA/DNA hybrid and play a role in the substrate binding. Mutations in the C-helix residues can cause replication defects in minus and plus strand priming, as well primer removal (Lim *et al.*, 2002, Telesnitsky *et al.*, 1993).

2.2.2.3 Connection domain

The connection domain (codons 289–440) links the p51 and p66 subunits. It is located between the polymerase and RNase H domains of HIV-1 and plays a crucial role in the RNase H activity. Amino acid residues (Lys390, Lys395 and Glu396 in p51; Ala360, His361 and Thr362 in p66) in the connection domain make precise contact with the RNA and DNA strands (Julias *et al.*, 2003). Substitutions at these positions usually interfere with the RNase H primer grip and can have a negative effect on enzyme positioning. For example, substitution of Alanine (Ala) for Histidine (His) in position 361 can reduce RNase H activity, since His361 interacts with the phosphate backbone of the DNA template strand close to Tyr501 of the RNase H primer grip (Julias *et al.*, 2003). Julias *et al.*, 2003, also demonstrated, using site-directed mutagenesis that mutations at other residues (A360V, A371V, K390R and A400T) in the connection domain seem to reduce, but not eliminate, the activity of RNase H.

2.2.2.4 Enzyme activity

The enzyme RNase H is a ribonuclease that hydrolyses the RNA from RNA-DNA hybrids, by cleaving the phosphodiester bond of RNA to generate 3'hydroxyl and 5' phosphate terminated products (Jacobo-Molina *et al.*, 1993). The RNase H endonuclease activities are non specific and its catalytic activities are through a hydrolytic mechanism, aided by enzyme-bound divalent cations (Mg^{2+} or Mn^{2+}) that bind to its active site (Katayanagi *et al.*, 1993). The RNase H domain is inactive but the endonuclease activity is activated by fusing the p51 thumb and connection sub-domains onto the RNase H domain (Evans *et al.*, 1991, Hostomsky *et al.*, 1991, Sluis-Cremer *et al.*, 2008, Smith *et al.*, 1992, 1993, 1994, 1997, 1998, 1999, 2009).

The binding and specificity of RNase H with the substrate can be influenced by the positioning of several amino acid residues in the polymerase domain of the reverse transcriptase (Evans *et al.*, 1991, Hostomsky *et al.*, 1991, Schultz *et al.*, 1996, 1994, Smith *et al.*, 1993). For example, studies by Gao *et al.*, 1998 and Huang *et al.*, 1998, showed that mutations in position 61 (fingers sub-domain) and 266 (thumb sub-domain) can reduce the RNase H activity from removing an elongated PPT primer. However, not all mutations in the polymerase domain that interact with the substrate influence the RNase H activity, e.g. mutation in position 64 (fingers sub-domain) has no influence on RNase H activity (Paulson *et al.*, 2007).

2.2.2.4.1 Substrate Interactions

The contact between the reverse transcriptase and its substrate influence the alignment of RNase H for RNA/DNA hybrid cleavage. In addition, the RNase H primer grip and a primer

grip in the polymerase domain also determine the positioning of the RNase H for cleavage (Jacobo-Molina *et al.*, 1993, Kohlstaedt *et al.*, 1992, Sarafianos *et al.*, 2001). The RNase H then makes both specific and nonspecific cleavages using the sugar phosphate backbone of the RNA. RNase H cleavage has 3 discrete modes known as DNA 3' end-directed, RNA 5' end-directed, and internal cleavages (Figure 9) (Schultz *et al.*, 2008).

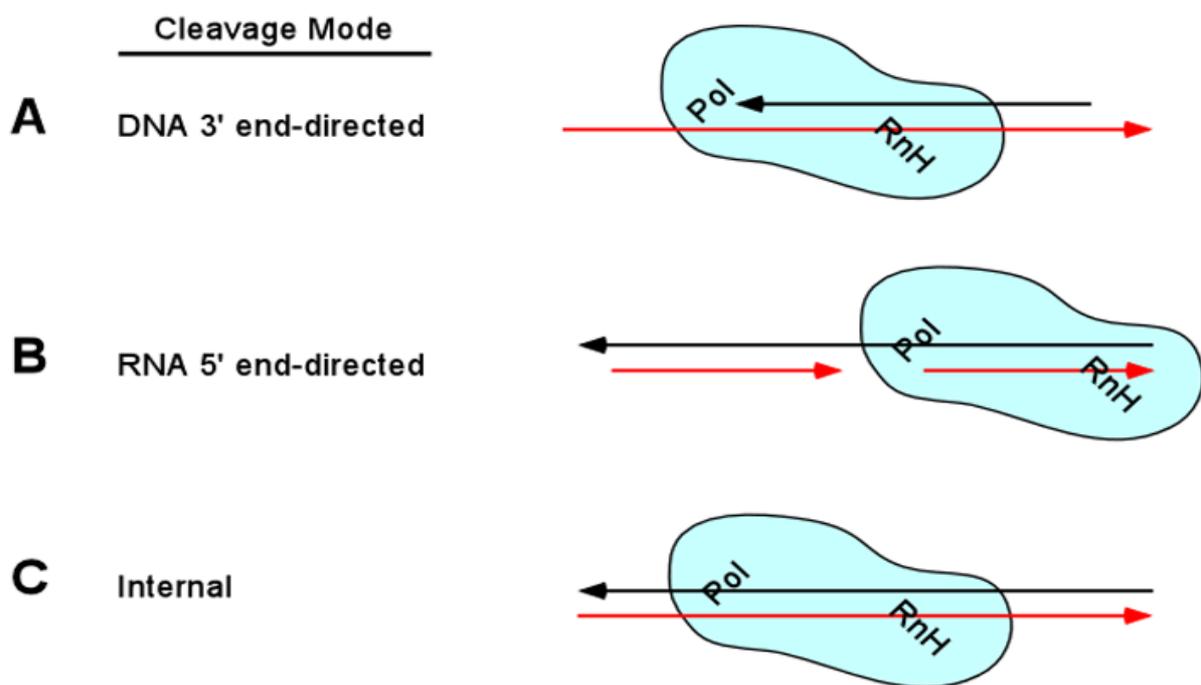


Figure 9. **The three discrete modes of RNase H cleavages:** (A) DNA 3' end-directed cleavage. (B) RNA 5' end-directed cleavage. (C) Internal cleavage. RT is depicted in blue, with RNase H active centre (RnH) and the active site of the polymerase (Pol). The hybrid substrate comprises RNA (red) and DNA strands (black), with the 3' ends indicated by an arrowhead (Schultz and Champoux, 2008).

2.2.2.4.2 DNA 3' end-directed cleavage

To accomplish the DNA 3' end-directed cleavages (Figure 9a), the polymerase domain binds with the 3' end of a recessed DNA primer annealed to an extended RNA strand, aligning the RNase H domain 15 to 20 nucleotides away from the recessed end. The alignment then

allows the RNase H to cleave the RNA template strand (DeStefano *et al.*, 1994b, Furfine *et al.*, 1991b, Schultz *et al.*, 2008). These cleavages can occur during synthesis of DNA and are called synthesis-dependent cleavage, or in the absence of DNA synthesis, they are referred to as synthesis independent cleavage (DeStefano *et al.*, 1994b, Furfine *et al.*, 1991b, Schultz *et al.*, 2008).

2.2.2.4.3 RNA 5' end-directed cleavage

To carry out the RNA 5' end-directed cleavage (Figure 9b), the polymerase domain of reverse transcriptase binds to the 5' end of a recessed RNA in a RNA/DNA hybrid to align the polymerase active site on the DNA strand near the RNA 5' end. This alignment places the RNase H domain on the RNA strand and the cleavage starts roughly 13 to 19 nucleotides away from the RNA 5' end. This cleavage is through a polymerization-independent form of RNase H activity (DeStefano *et al.*, 1994b, Gao *et al.*, 1997, Gotte *et al.*, 1995, Palaniappan *et al.*, 1996, Schatz *et al.*, 1990, Wohrl *et al.*, 1997).

2.2.2.4.4 Internal cleavage

In this mode, the RNase H cleavage can occur within the RNA/DNA hybrid without DNA synthesis and does not involve the positioning required for the other two forms of cleavage. Internal cleavage occurs when the reverse transcriptase binds at internal positions on the RNA template (Figure 9c) (Finston *et al.*, 1984, Fuentes *et al.*, 1995, Schultz *et al.*, 2003, Schultz *et al.*, 2004, Schultz *et al.*, 2000).

2.2.2.5 Use of different binding modes in reverse transcription

The degradation of the RNA template is through DNA 3' end -directed cleavage. However, this mode does not completely degrade the tRNA^{lys3} due to the delay caused by the hairpin structure on the RNA template strand during polymerization (DeStefano *et al.*, 1994b, Kati *et al.*, 1992). The internal and RNA 5' end-directed cleavages are responsible for further degradation of the RNA template to facilitate the plus-strand DNA synthesis and second strand transfer. Internal RNase H cleavages initiate RNA degradation and generates a 2 to 3 base pair gap that allows the kinetically preferred RNA 5' end-directed cleavages to start (Kelleher *et al.*, 2000).

2.2.2.6 Cleavage specificity of RNase H

2.2.2.6.1 Nucleotide preferences

Internal cleavage prefers certain nucleotides at specific positions of the RNA/DNA hybrid, these are: positions +1 (A/U), -2 (C/G), -4 (C/g/u), -7 (G/C), -12 (A/g/u), and -14 (A/g) (Schultz *et al.*, 2004). The strongly preferred nucleotides for each position are indicated in uppercase and weakly preferred are indicated in lowercase. The preferred nucleotide positions for the RNA 5' end-directed mode are: +1 (A/U/c), -2 (C/G), -4 (U/c/g) (Figure 10) (Schultz *et al.*, 2004). DNA 3' end-directed cleavage has no preferred position, but result from the preferential connection of the polymerase domain of reverse transcriptase to the recessed DNA 3' end in RNA/DNA hybrid (Palaniappan *et al.*, 1996, Schultz *et al.*, 2004).

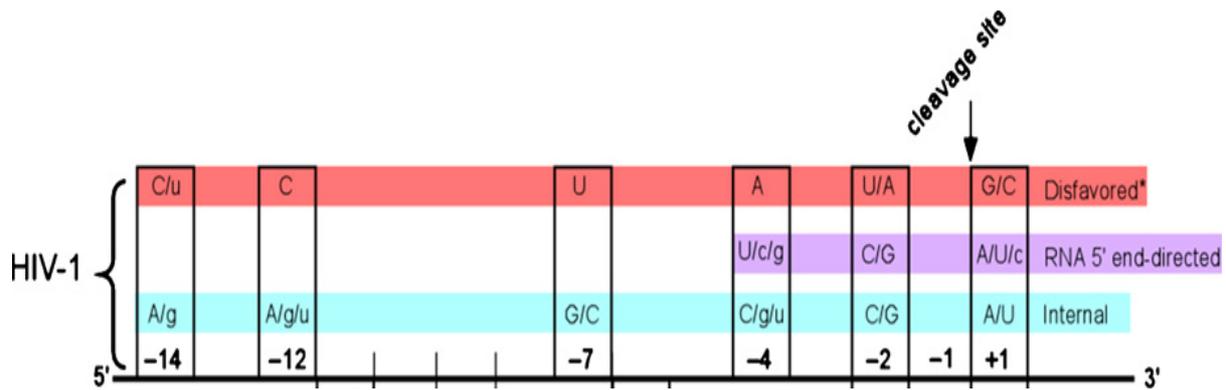


Figure 10. **Schematic representation showing the positions and base preferences flanking internal and RNA 5' end-directed RNase H cleavage sites.** The locations of preferred positions relative to a cleavage site at the scissile phosphate (arrow) located between nucleotides -1 and $+1$ are indicated on an RNA strand (thick black line). The preferences for or against nucleotides for HIV-1 RNase H are shown above the line. The preferred nucleotides for each position are indicated in uppercase (strongly preferred) or lowercase (preferred) for internal (blue) and RNA 5' end-directed (violet) cleavage sites (Schultz and Champoux, 2008).

2.2.2.6.2 Removal of tRNA primer

The degradation of the tRNA^{lys3} requires RNase H cleavage to occur at specific sequences shown in Figure 11 (Schultz *et al.*, 2009). The extended tRNA primer removal occurs in the DNA 3' end-directed mode by a stalled reverse transcriptase. The tRNA degradation leaves a ribo A (rA) at the 5' end of minus-strand DNA that pairs the preferential nucleotides for internal cleavage (Furfine *et al.*, 1991a, Pullen *et al.*, 1992, Schultz *et al.*, 1995).

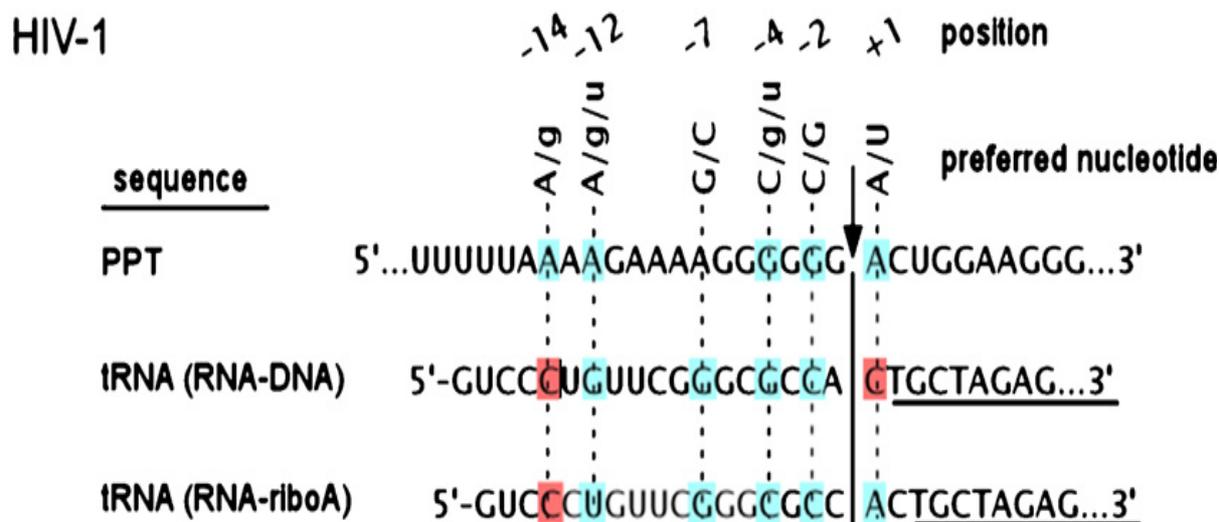


Figure 11. Schematic representation showing the positions and sequences surrounding cleavage sites for the extended tRNA primer and PPT region of HIV-1. The $-1/+1$ site for cleavage is indicated by an arrow and a vertical line. The match between the preferred positions and an internal cleavage site is indicated as preferred (blue box) or disfavored (red box). The DNA portion of the extended tRNA is underlined. Alignment of the tRNA sequence is shown for the cleavage at the RNA–DNA hybrid and for cleavage between two ribonucleotides that leaves a ribo A at the end of the minus strand (RNA-ribo A) (Schultz, 2009).

2.2.2.6.3 Generation of PPT primer

The minus strand DNA synthesis starts when the tRNA primer extends through U5 and R sequences. The tRNA is then transferred to a polypurine tract (PPT) at the 3' end of the viral RNA; it is then replicated and subsequently cleaved to generate the PPT primer (Sorge *et al.*, 1982). The PPT RNA primer is followed by 1 to 3 adenosines after a 5' rA:dT-rich area and a 3' rG:dC tract, with the cleavage occurring amongst the stretch of guanines and adjacent adenosines (Kvaratskhelia *et al.*, 2002, Rausch *et al.*, 2004, Sorge *et al.*, 1982). Both the 5' and 3' ends of the DNA at the polymerase active site along with preferred nucleotides of the PPT are significant for the HIV-1 RT replication capacity (Figure 12) (McWilliams *et al.*, 2003, Miles *et al.*, 2005). For example, Julias *et al.*, 2004 showed that mutations in the PPT reduce the viral replication and disrupt the PPT catalysis by RNase H cleavages.

<u>PPT</u>	*	U3
RNA 5' . . . CAGCCACUUUUU <u>AAAAGAAAAGGGGGG</u>		ACUGGAAGGG . . . 3'
(-) DNA 3' . . . GTCGGTGAAAAA <u>TTTTCTTTTCCCCC</u>		TGACCTTCCC . . . 5'
		+1

Figure 12. **HIV-1 genome sequence at the PPT (underlined) and U3 region.** The minus strand synthesis initiation site is marked with an asterisk; +1 is the first nucleotide of U3 (Sarafianos, 2001).

2.2.2.7 RNase H and Antiretroviral Therapy

The HIV-1 RT remains an important therapeutic target for HIV-1 infected patients. Most studies on reverse transcriptase inhibitor drug resistance have concentrated on the N-terminal half of RT, at sites where the drugs target, and consequently have only analyzed up to 334 amino acids in RT (Santos *et al.*, 2008). The C-terminal (RNase H) domains have not been fully characterized in the context of drug resistance.

2.2.2.7.1.1 Reverse Transcriptase Inhibitors

Reverse-transcriptase inhibitors (RTIs) are antiretroviral drugs used to inhibit reverse transcriptase's enzymatic role and stop synthesis of the double-stranded viral DNA (Shafer *et al.*, 2002). There are two distinct groups of RT inhibitors in clinical use; the nucleoside reverse-transcriptase inhibitors (NRTIs) and non-nucleoside reverse-transcriptase inhibitors (NNRTIs). The efficacies of these drugs are limited by HIV-1 drug resistance, which is usually caused by mutations in RT (Shafer *et al.*, 2002).

2.2.2.7.1.2 The Nucleoside Reverse Transcriptase Inhibitors (NRTIs)

The NRTIs are used in the first-line regimens of most countries worldwide. These RT inhibitors are dideoxy nucleoside or nucleotide triphosphates (ddNTPs) that lack a hydroxyl group (-OH) on the 3'-C as well as the 2'-primed carbon (2'-C) of the deoxyribose sugar (Figure 13) (Goldschmidt *et al.*, 2004). Most NRTIs need to be phosphorylated by host cellular kinase enzymes before incorporation into nascent viral DNA. The phosphorylated forms of NRTIs inhibit the viral replication by competing with natural deoxynucleoside triphosphate (dNTP) pools for incorporation into the nucleotide binding site of the newly synthesized DNA chains. Upon incorporation, the lack of a 3' hydroxyl group on the NRTI prevents further synthesis of viral DNA and act as a chain terminator (Goldschmidt *et al.*, 2004).

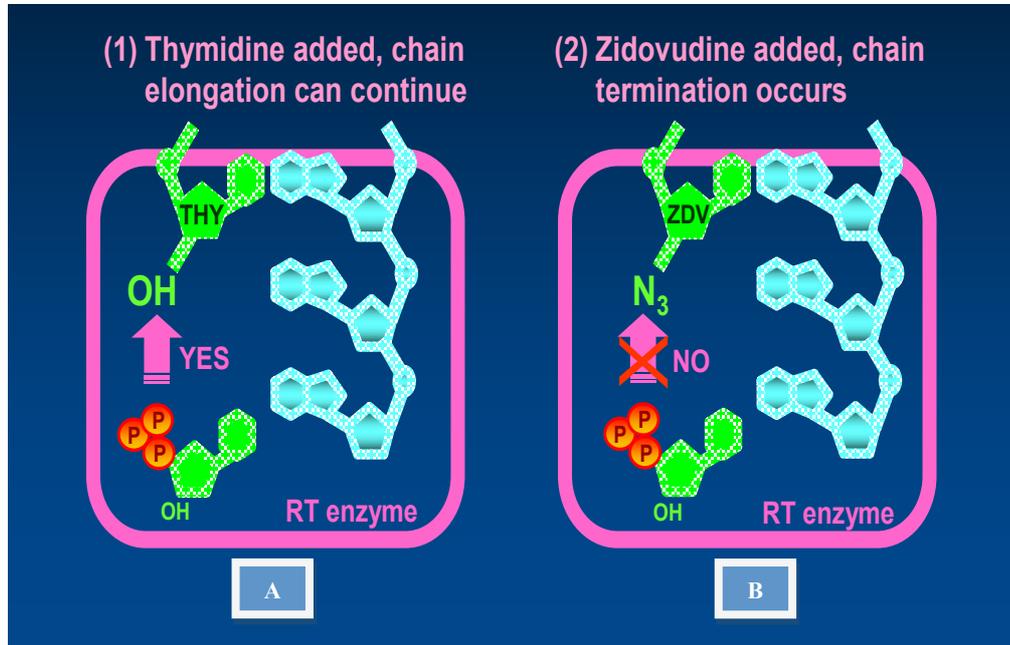


Figure 13. **The mechanism of NRTIs.** (A) Incorporation of dNTP to the free 3' OH group at the polymerase active centre. (B) Phosphorylated dNTP attached to the RT and chain termination occurs (Gotte *et al.*, 2000).

As the name suggests, the NRTI analogs mimic the four nucleotides listed below:

- ✓ thymidine analogs - zidovudine (AZT) and stavudine (d4T);

- ✓ cytidine analogs - lamivudine (3TC), and emtricitabine (FTC);
- ✓ guanosine analog - abacavir (ABC);
- ✓ adenosine analogs - didanosine (ddl) and tenofovir.

The discrimination against the NRTIs and NRTIs pyrophosphorolysis are two major mechanisms associated with NRTI drug resistance (Figure 14) (Menendez-Arias *et al.*, 2008). The first mechanism is caused by mutations (K65R, L74V, Q151M and M184V) that decrease the binding of the NRTIs to the RT enzyme, allowing the RT enzyme to discriminate against the NRTI incorporation to the extended DNA chain during polymerisation (Whitcomb *et al.*, 2003). These mutations are also associated with reduced catalytic activity of the RT enzyme while the incorporation of dNTPs remains undisturbed causing the virus to continue replicating (Whitcomb *et al.*, 2003). The second major mechanism involves the presence of pyrophosphate or a pyrophosphate donor such as adenosine triphosphate (ATP) or thymidine analogues mutations (TAMs) (D67N, K70R, L210W, T215Y and K219Q) 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).

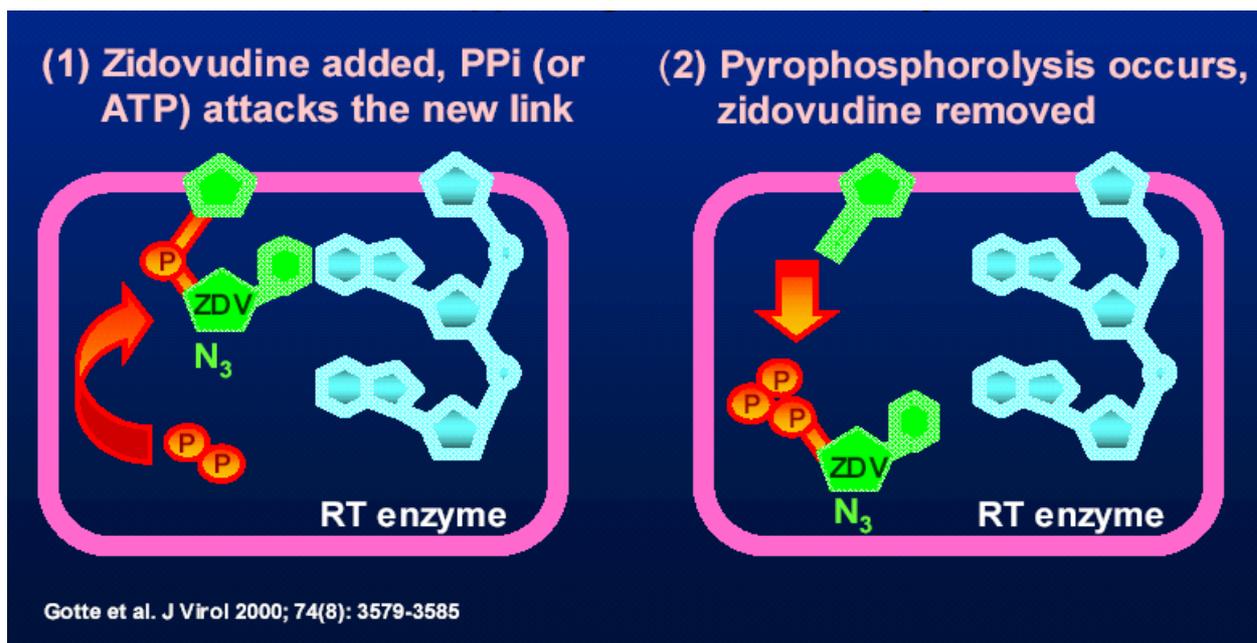


Figure 14. Diagram indicating pyrophosphorolysis, one of the major mechanisms associated with NRTI resistance (Gotte *et al.*, 2000).

2.2.2.7.1.3 The Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

NNRTIs are non-competitive inhibitors that block the reverse transcriptase enzyme allosterically by binding to a hydrophobic pocket roughly 10Å away from polymerase active site of the p66 subunit (Shafer *et al.*, 2002). The binding at these less well conserved active sites lead to conformational changes that interfere with the formation of the binding pocket leading to a reduced RT activity (Figure 15) (Shafer *et al.*, 2002).

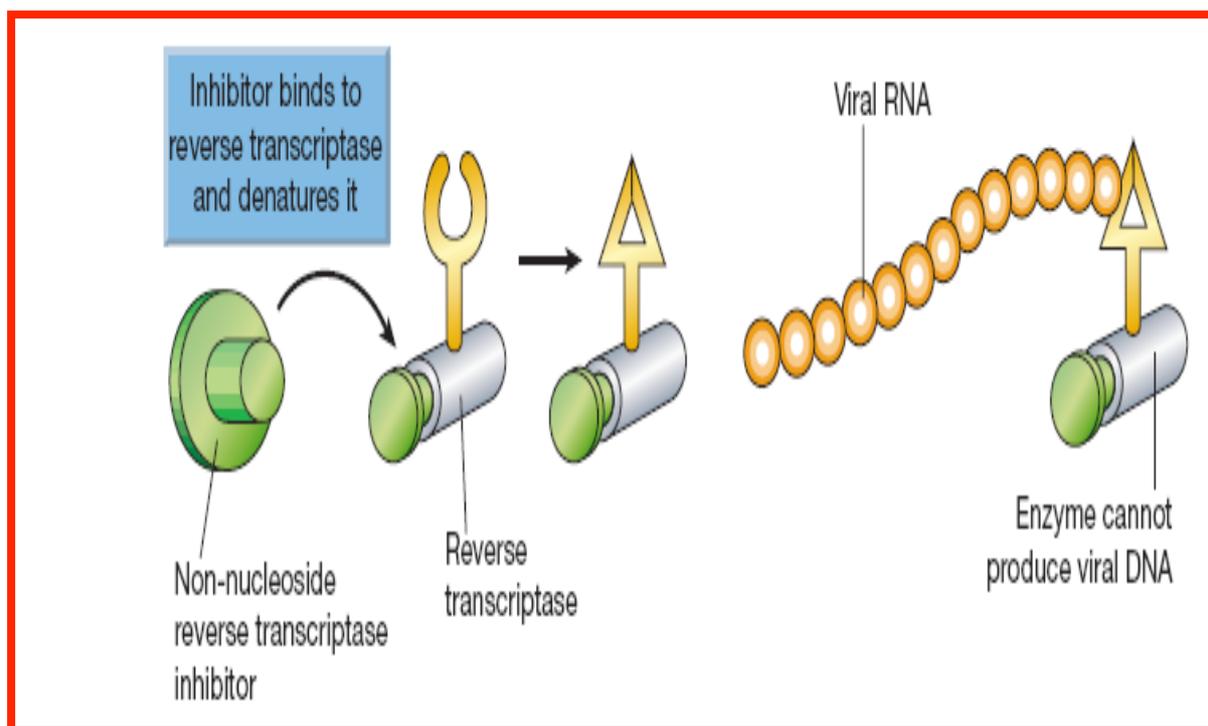


Figure 15. **Schematic representation demonstrating binding mechanism of action of non-nucleoside reverse transcriptase inhibitor to the RT enzyme** (Richman, 2001).

There are four NNRTIs in clinical use:

- ✓ Navirapine (NVP)
- ✓ Delavirdine (rarely used)
- ✓ Efavirenz (EFV)
- ✓ Etravirine

Most of the mutations (K103N, Y181C/I/V) associated with NNRTI drug resistance are found in the hydrophobic binding pocket (Llina *et al.*, 2008, Sluis-Cremer *et al.*, 2008). These mutations result in reduced susceptibility not only to the NNRTIs in the drug regimen but to all members of this class of drug. For example, K103N results in greater reductions in susceptibility to both efavirenz and navirapine (Scherrer *et al.*, 2009, Vingerhoets *et al.*,

2009). However, K103N does not affect the etravirine response (Scherrer *et al.*, 2009, Vingerhoets *et al.*, 2009). Etravirine, a new generation NNRTI, is a diarylpyrimidine (DAPY) that can bind to the reverse transcriptase enzyme at allosteric sites, even in the presence of K103N (Scherrer *et al.*, 2009, Vingerhoets *et al.*, 2009).

2.2.2.7.2 RNase H and NRTI Resistance

The data on NRTI resistance associated with mutations harbored in the RNase H domains is very sparse. However, several studies have identified mutations in the RNase H domain that are associated with resistance to NRTIs. Furthermore, these studies have provided essential understanding into the RNase H dependent mechanisms of NRTI drug resistance.

2.2.2.7.2.1 RNase H dependent mechanism for NRTI drug resistance

The mechanism for NRTI drug resistance was proposed by Julias *et al.* 2002 and Nikolenko, *et al.* 2005, in which mutations in the RNase H domain, specifically H539N and D549N, may contribute to NRTI resistance through reduction of the RNase H activity. This results in reverse transcriptase having more time for AZT-MP excision by decreasing RT template switching. In addition, it was also noted that these mutations can decrease the rate of RNA degradation during reverse transcription, inducing continuation of viral DNA synthesis (Nikolenko, *et al.*, 2007). This is suggested based on the observation that the integration of AZT-MP retards the RT complex while the AZT-MP excision resumes DNA synthesis (Nikolenko, *et al.*, 2005, 2007). These findings led to a model in which AZT-terminated reactions will reduce RNase H cleavage, subsequently allowing RT to form longer extensions of homology between the template:primer strands during reverse transcription (Nikolenko, *et*

al., 2005, Volkmann *et al.*, 1993). This process allows more time for NRTI excision, resulting in the continuation of DNA synthesis and high-level resistance to NRTIs.

Furthermore, mutations in the RNase H region that decrease the RNase H activity contribute to NRTI resistance by directly or indirectly affecting the positioning of the RNA template and DNA primer strand at the RNase H active region. Several amino acid residues of the RNase H region interact with the DNA primer strand and create the RNase H primer grip structure which aligns the RNA/DNA hybrid at the RNase H active region to assist efficient RNA cleavage (Sarafianos *et al.*, 2001). Site-directed mutagenesis studies on the RNase H primer grip have shown reduced RNase H activity, poor RNA template degradation, reduced RT template switching and deficient DNA synthesis (Julias *et al.*, 2002, Nikolenko *et al.*, 2005). These observations were confirmed by Delviks-Frankenberry *et al.*, 2008, who showed that alanine substitution mutations (T473M, N474A, Q475A, K476A, Q500A, Y501A and I505A) at the RNase H primer grip decrease RNase H activity, resulting in NRTI resistance.

It has been previously shown that the RNase H mutations near the RNase H primer grip (L469F/I, T470N) and the RNase H active site (A554S and K558E/R) can affect the interactions of RT with its nucleic acid substrate, which results in reduced RNase H activity and contributes to AZT resistance (Julias *et al.*, 2002, Nikolenko *et al.*, 2005). Additionally, L469F/I, T470N could partially reduce strand transfer efficiency and deficient DNA synthesis can contribute to the NRTI-resistant phenotype by providing more time for RT to carry out nucleotide excision and resume productive DNA synthesis (Julias *et al.*, 2002, Nikolenko *et al.*, 2005). Additionally, the RNase H domain mutation (Q509L) was found to increase NRTI resistance synergistically with TAMs, but had less effect in the absence of TAMs.

Studies performed by several groups reported differences in RNase H between naïve and NRTI treatment experienced patients (Santos *et al.*, 2008, Julias *et al.*, 2002, Nikolenko, *et al.*, 2005). However, few studies included subtype C sequences from NRTI treatment experienced patients.

2.2.2.7.3 RNase H as an Anti-viral Target

A review by Schultz and Champoux in (2008), proposed that a new class of drugs, called RNase H inhibitors, be developed. These drugs would either target the interactions between the RNase H domain and substrate or alter the positioning of substrate to the RNase H active site, in this way negatively affect polymerization. They also proposed a drug that is non-specifically associated with the minor groove of DNA/DNA and RNA/RNA substrates which would expand the breadth of the minor groove while reducing that of the major groove. Such alterations of the structure could lead to retarded RNase H cleavages, resulting in impaired polymerization (Schultz *et al.*, 2008).

2.2.3 Drug Resistance Testing

HIV-1 drug-resistant testing remains the most important practical instrument for executing an assessment of ARV treatment regimes. Genotyping and phenotyping assays are widely used to decide whether drug-resistant mutations are present in virions in the body of an infected patient (Potter *et al.*, 2004). Standard genotyping methods are more commonly used in clinical settings than phenotypic tests because of their lower cost, quicker turnaround, wider availability and ability to detect minor virions carrying resistance-associated mutations (Potter *et al.*, 2004). The ViroSeq™ (Applied Biosystems), Trugene® (Bayer, Pittsburgh,

PA) and GeneSeq™ (ViroLogic, Inc., South San Francisco, CA) assays are major sequence-based assays used for HIV-1 drug resistance analysis. The only disadvantage is that such methods require virions containing mutations to be present at levels greater than 20% of the total viral population in order to be detected and thus, it is easy to miss relevant mutations or combinations of mutations (Potter *et al.*, 2004).

The ViroSeq™ (Applied Biosystems) assay produces a 1.8-kb *pol* gene product that includes the entire protease gene and 335 codons of the reverse transcriptase gene. The size and the quality of the PCR product is verified using agarose gel electrophoresis. The resulting PCR products are sequenced directly using six overlapping sequencing primers and with premixed BigDye sequencing reagents. The ViroSeq™ kit (Applied Biosystems) recommends 0.5 ml of plasma to be used and viral loads greater than 1,000 copies/ml for analysis (Cunningham *et al.*, 2001).

The Trugene® assay amplifies a 1.3 kb product that covers protease codons (1 to 99) and RT codons (41-247) (Grant *et al.*, 2003). The PCR products are sequenced directly using seven overlapping sequencing primers. The OpenGene® DNA Sequencing System (Visible Genetics, Inc., Canada) is used to sequence data collected according to the manufacturer's recommendations (Grant *et al.*, 2003). The Trugene® assay can use plasma greater than 1 ml and viral loads greater than 500 copies/ml for analysis (Grant *et al.*, 2003).

The Antivirogram® (Virco BVBA; Mechelen, Belgium) and PhenoSense™ (ViroLogic, Inc., South San Francisco, CA) assays are rapid phenotypic assays for assessment of drug susceptibility of human immunodeficiency virus type 1 isolates (Lorain *et al.*, 2005). The PhenoSense™ assay is a single-cycle assay that is achieved by using recombinant viruses that

are limited to a single round of viral replication by a specific deletion in the HIV *env* gene within the vector, while the Antivirogram® assay is a recombinant virus assay that generates high virus stocks by cultivating the recombinant viruses for several replication cycles (1-2 weeks) then the stock is titrated before testing for drug susceptibility. The same format can be used to evaluate both protease inhibitors and reverse transcriptase inhibitors by adding serial dilutions of drug to the culture at the time of virus inoculation (Lorain *et al.*, 2005).

2.2.4 Bayesian Network

A Bayesian network (BN) is a graphical model that describes statistical correlation between multiple variables that hold in the domain (Pearl *et al.*, 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 (Deforche *et al.*, 2006). The directed acyclic structures of a BN are also capable of representing relationships between the variables through direct conditional or unconditional dependencies (Figure 16) (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). For example, if a dependency model has four variables A, B, C and D, it could be analyzed as follows (Myllymäki *et al.*, 2002):

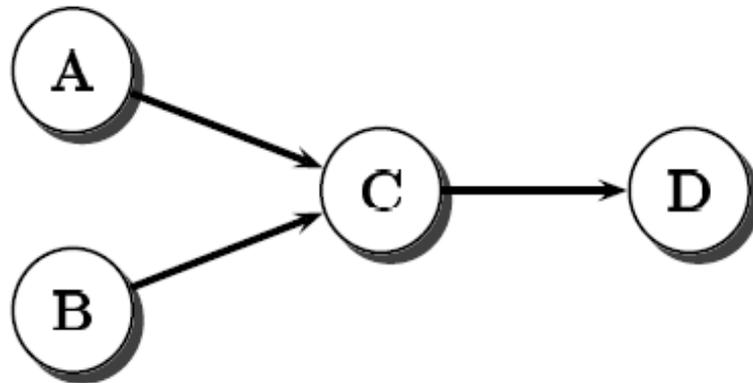


Figure 16. A Bayesian Network representing conditional and unconditional dependencies (Myllymäki, 2002).

- “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”.

Furthermore, the search strategy for a good BN model is also used to explore the possible structures by testing for conditional independence of the arc to the given event (Deforche, 2008, Myllymäki *et al.*, 2002, Pearl *et al.*, 1998). The search strategy uses joint probability distributions (p) to mathematically identify the behavior of each variable, given each combination of states of its parents. For example,

Scenario:

- a) For A and B to happen, B has happened and if B has happened then A has also happened (and vice versa). The joint probability distribution of A and B is then:
$$p(A,B) = p(A|B)p(B) = p(B|A)p(A).$$
- b) If A and B are independent then: $p(A|B) = p(A)$ and as a result, the joint distribution of A and B is written as: $p(A,B)=p(A|B)p(B) = p(A)p(B)$.
- c) If A is independent of B, conditionally on knowing (values of) C and D then: $p(A|B,C,D) = p(A|C,D)$ (*Bayes rule 1*)), thus the joint distribution of A,B,C and D is given as: $p(A,B,C,D) = p(A|B,C,D) \times p(B,C,D)$ and if *Bayes rule 1* holds, this is written as $p(A,B,C,D) = p(A|C,D) \times p(B,C,D)$

A Bayesian scoring metric is used to evaluate the quality of the most probable network by computing the expected frequencies under the assumption that the variables are dependent or independent (Myllymäki *et al.*, 2002, Pearl *et al.*, 1998). The scoring metric comparative measure of the posterior probability is through marginal likelihood of the model given the data (Deforche, 2008, Myllymäki *et al.*, 2002, Pearl *et al.*, 1998). The marginal likelihood model controls the network complexity to identify the best probable network using Bayes Factor (BF) (Deforche, 2008, Myllymäki *et al.*, 2002). The BF computes how many times a model is more likely than another as the ratio of their marginal likelihood. For example:

1. All models use the same data: $p(A|B) = p(B|A) p(A)$
2. Models have equal prior probability: $p(A|B) = p(B|A)$
3. $BF(A,C) = p(B|A) / p(B|C) = p(A|B) / p(C|B)$

After the most credible BN network structure has been found, assessments of the network robustness are searched using bootstrapping (Deforche, 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). The existence and thickness of arcs showing a direct influence amongst the corresponding variables are relative to bootstrap values and their importances are colored according to the arc weight. For example (Figure 17), black arc weighed $\geq 10^9$, purple weighed $\geq 10^6$, green weighed $\geq 10^3$, and blue weighed $\geq 10^3$ while grey weighed ≥ 1 (Myllmäki *et al.*, 2002, Pearl *et al.*, 1998). 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 favoritism in treatment combinations. Dotted arcs represent a bootstrap of 35% or more and robust arcs represent bootstraps over 65% (Deforche *et al.*, 2006).

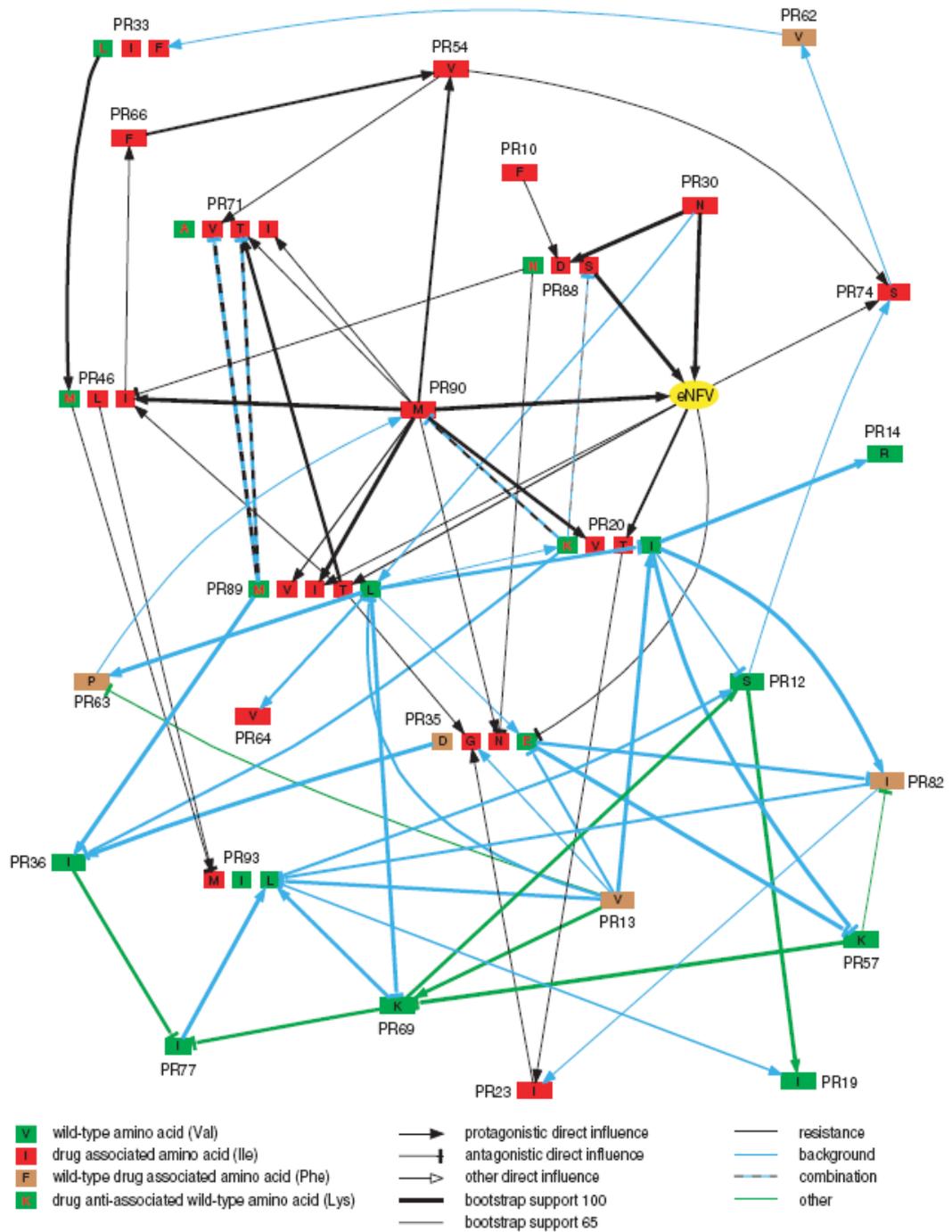


Figure 17. **Annotated nelfinavir experience Bayesian network expressing direct association between nelfinavir-associated mutations, polymorphisms and nelfinavir treatment (eNFV)** (Deforche *et al.*, 2006).

2.2.5 Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is an *in-vitro* technique widely used in immunology, diagnostics, molecular biology, genetics, clinical and forensic laboratories. It allows the amplification of a single or few copies of DNA to produce more copies of known DNA sequences (Bartlett & Stirling, 2003). The known DNA sequence is achieved by synthesis of new DNA strands complementary to the template strands. Prerequisites for successful PCR include the use of appropriate concentrations of primers, DNA polymerase enzyme, deoxynucleotide triphosphates (dNTPs), DNA template, sterile RNase free water, magnesium, buffer and suitable reaction conditions (Pavlov *et al.*, 2004).

The DNA template contains the target sequence that is used by the thermostable DNA polymerase enzyme to incorporate nucleotides for the synthesis of a new complementary DNA strand (cDNA) (Pavlov *et al.*, 2004). Primers used in the PCR reaction are short (between 20 – 30 nucleotides long) single-stranded DNA molecules complementary to the ends of a target sequence of DNA template (Sambrook and Russel 2001). The end of a primer is also used by the DNA polymerase to synthesize a new strands of DNA. Deoxynucleotide triphosphates (dNTPs) are single units of the bases A, T, G, and C, which are essential for the incorporation into the growing DNA strand in the PCR reaction (Pavlov *et al.*, 2004). The optimal magnesium (Mg^{2+}) concentration stimulates the activation of the polymerase enzyme and forms a soluble complex with dNTPs while the PCR buffer influences the pH of the reaction that affects the polymerase enzyme activity and fidelity (Sambrook and Russel 2001).

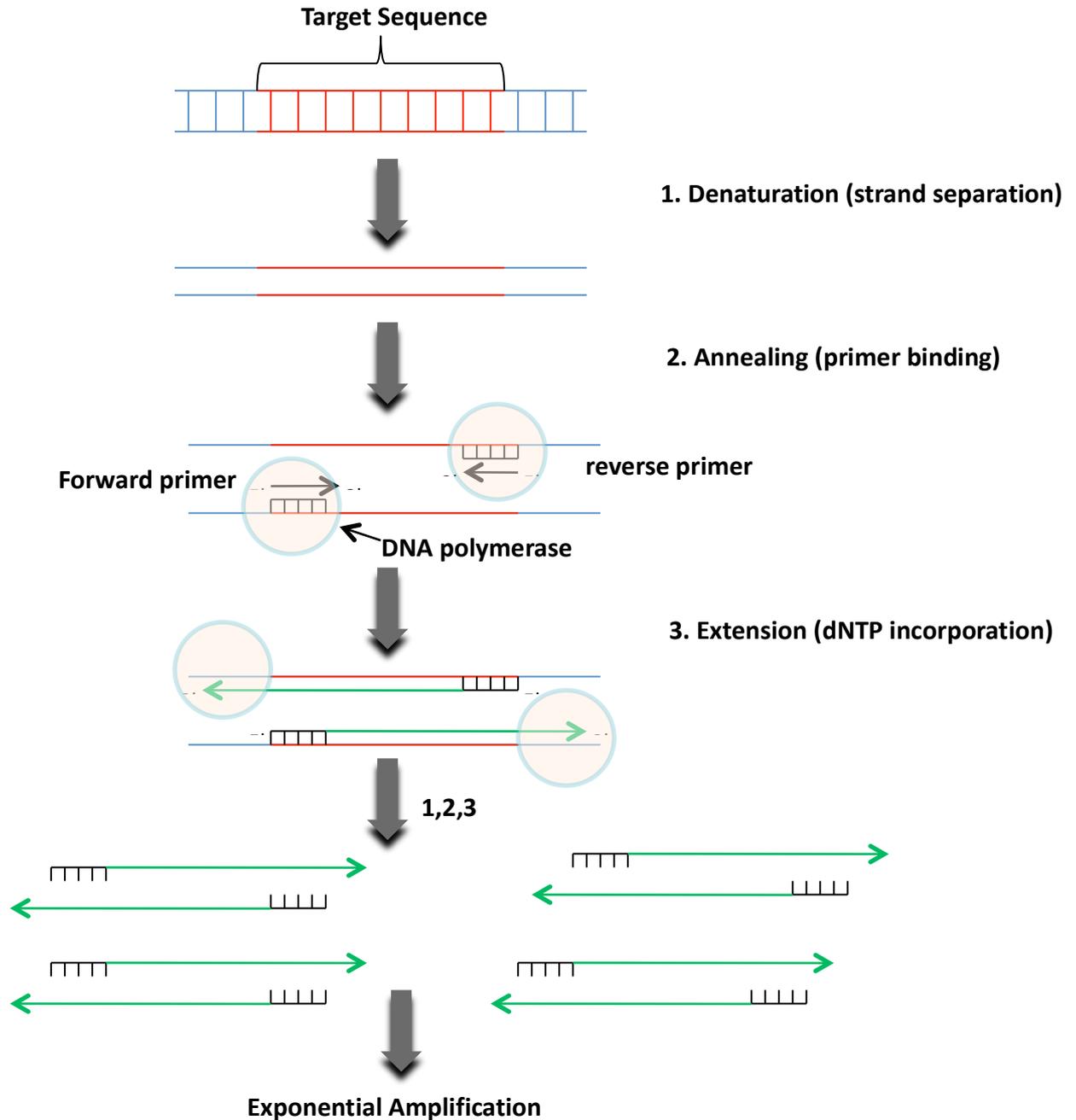


Figure 18. **Schematic drawing of the PCR cycle.** (1) Denaturing at 94–96 °C. (2) Annealing at ~65 °C (3) Elongation at 72 °C. Four cycles are shown here. The blue lines represent the DNA template with a target sequence (red lines) to which primers (black arrows) anneal that are extended by the DNA polymerase (light orange circles), to give shorter DNA products (green lines), which themselves are used as templates as PCR progresses.

Optimized cycling parameters (denaturation, annealing and extension temperatures) also play a crucial role in the PCR reaction (Figure 18) (Rychlik *et al.*, 1990). The denaturation temperature is required for denaturation of template DNA and complex genomic DNA during PCR. Annealing temperature is specific for primer binding more efficiently to the RNA template whereas extension temperature is used to ensure that all products are fully extended (Rychlik *et al.*, 1990).

2.2.6 DNA Sequencing

DNA sequencing is the essential molecular research tool used to determine sequences of nucleotides in a sample of DNA. In 1977, Frederic Sanger and colleagues invented the dideoxy DNA sequencing procedure. The DNA sequencing method is done using both natural dNTPs (dATP, dCTP, dGTP, dTTP) and synthetic dideoxynucleotides (ddNTPs) which lack the 3'-hydroxyl group (-OH) necessary for chain extension (Smith *et al.*, 1985, Smith *et al.*, 1986). Each of the 4 ddNTPs (ddATP, ddCTP, ddGTP, ddTTP) is tagged with a different fluorescent dye which can be detected by a special laser (Figure 19) (Smith *et al.*, 1985, Smith *et al.*, 1986). During PCR, the double-stranded DNA is denatured into single strands and an oligonucleotide primer is then annealed to the denatured template strands to initiate the DNA extension (Smith *et al.*, 1985, Smith *et al.*, 1986). The DNA polymerase synthesizes a new DNA chain from natural dNTPs added to the DNA template. However, whenever DNA polymerase incorporates the synthetic ddNTPs to the growing chain, they act as chain terminators because another nucleotide cannot be attached to it (Smith *et al.*, 1985, Smith *et al.*, 1986).

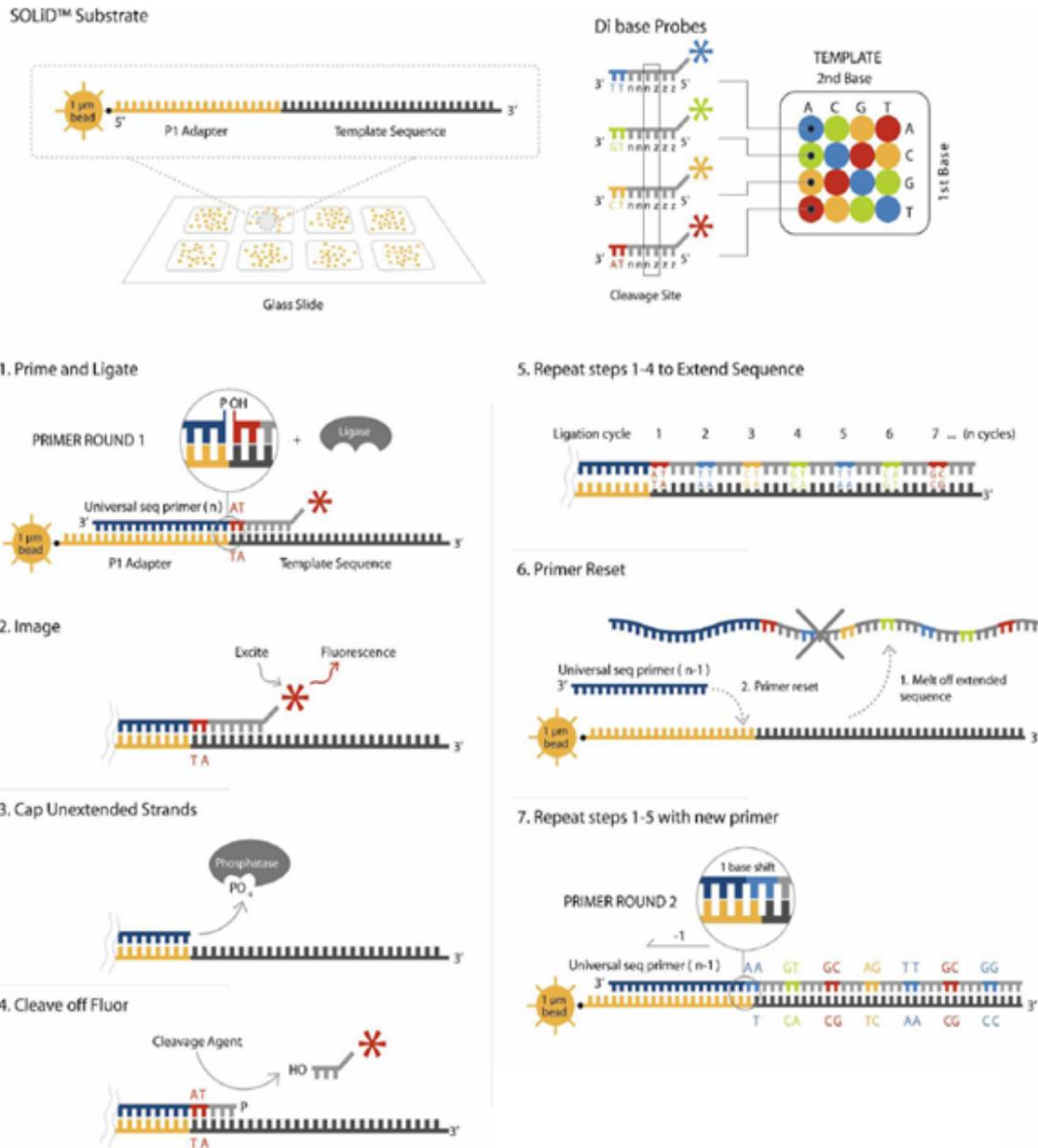


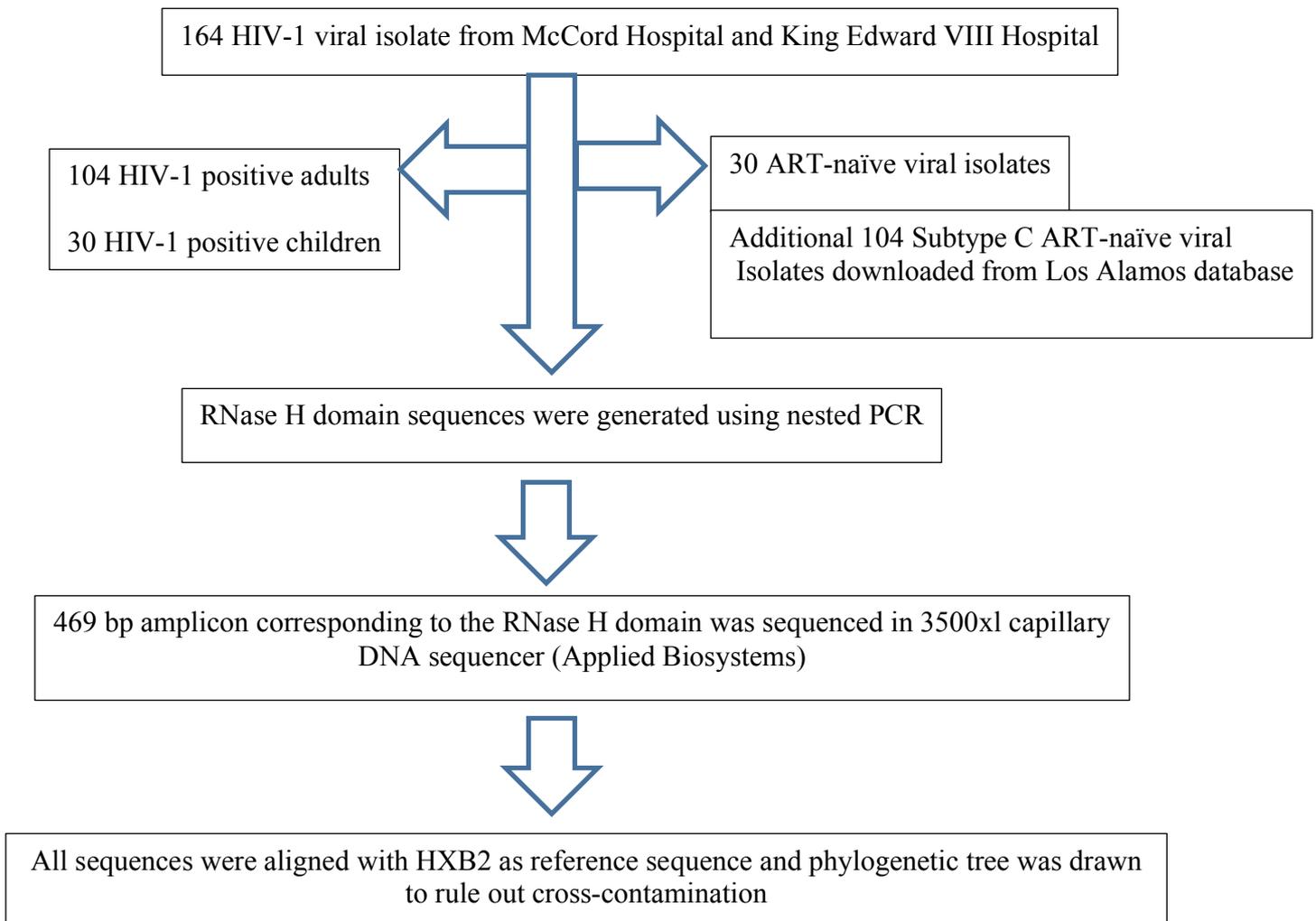
Figure 19. **Schematic representation of dye-labeled terminator reactions.** Each of the four dideoxy terminators (ddNTPs) is tagged with a different fluorescent dye and tagged products are generated in four separate base-specific reactions. When incorporated, they simultaneously terminated the growing chain and tagged with the dye that matches to that base. The products from these four reactions are then combined and loaded into a single capillary injection and translated by a photocell and recorded on a computer. (Valouev *et al.*, 2008).

At the end of the sequencing reaction, the newly synthesized DNA strands with fluorescent tags are sorted and distinguished by length from shortest to longest using capillary electrophoresis (Figure 19) (Valouev *et al.*, 2008). As the strands emerge out of the narrow capillaries tube filled with the polyacrylamide gel solution, they are scanned with a laser detection device. The laser excites the fluorescent dye attached to the ddNTPs, and the color of fluorescence is then translated by a photocell, which feeds the information to the computer. A computer program measures the information received from the photocell as an electropherogram and names the nucleotides in the rate they passed through the viewer (Valouev, 2008).

Chapter 3: Research Design and Methodology

3.1 Study Design

This was a retrospective study to investigate the presence of NRTI-associated mutations in the hiv-1 reverse transcriptase gene of hiv-1 in isolates from South African adults and children failing HAART. Patients were recruited from ARV clinics at McCord Hospital and King Edward VIII Hospital, Durban.



3.1.1 Study Population

Plasma samples were studied retrospectively from HIV-1 infected adults (104) and children (30) who had already been genotyped for drug resistance. All patients had experienced virologic failure with a viral load of greater than 1000 copies/ml. Samples were identified by their study number with no link to original patient ID numbers. Samples were obtained from the South African Resistance Cohort Study (SARCS) at McCord Hospital and the Tropism Paediatric study at King Edward VIII Hospital. Sample size was determined in consultation with a statistician, Ms Anneke Grobler [Centre for the Aids Programme of Research in South Africa (CAPRISA)].

3.1.2 Inclusion Criterion:

- Previously genotyped patients failing their first or second-line regimen on the South African National roll-out.
- The regimen must have included AZT or d4T.
- The patient must have been on treatment at the time of sample collection.

3.1.3 Exclusion Criterion:

- Patients who did not receive an NRTI.
- Patients not currently on treatment.

3.1.4 Sample Collection:

This study was conducted retrospectively. All plasma samples were stored at -80°C.

3.2 Methods

3.2.1 RNA extraction:

Stored RNA from a previous study by Marconi et al., 2008, was used when it was available. For those plasma samples for which RNA was not available, viral RNA was isolated using the QIAamp RNA kit (QIAGEN Inc.) according to the manufacturer's instruction. Briefly, 140 µl of sample plasma was transferred into a new 1.5 ml screw-capped tube (RNase/DNase free) containing Buffer AVL with carrier RNA and mixed by pulse-vortexing. This homogeneous solution was then incubated at room temperature (15 - 25°C) for 10 minutes to ensure efficient lysis of the viral particles. Five hundred and sixty microlitres of absolute ethanol (96 – 100%) was then added to ensure efficient binding of viral particles and the tube was briefly spun to collect drops from the tube lid. The solution was transferred in a 2ml collection tube (QIAamp Mini column) and centrifuged at 6000 x g [8000 revolution per minute (rpm)] for 1 minute. Five hundred microlitres of Buffer AW1 was added into a QIAamp Mini column and centrifuged at 6000 x g (8000 rpm) for 1 minute. A new sterile 2 ml collection tube was used after discarding the filtrate and 500µl AW2 was added and centrifuged at 20,000 x g (14,000 rpm) for 3 minutes. The column was incubated at room temperature for 1 minute and viral RNA from the column was eluted by adding 30µl of sterile RNase free water, and then spun at maximum speed for 1 minute. Extracted RNA was stored at -80°C until ready for use.

3.2.2 Amplification and Sequencing of the RNase H region

3.2.2.1 HIV-1 Reverse Transcription

Amplification of the RNase H region began with reverse transcription of the viral RNA to complementary DNA (cDNA) using the commercial ImProm-II Reverse transcriptase enzyme (Promega© Corp.). Two microliters of random hexamers (0.1 µg) were added to each reaction tube followed by the addition of 3 µl of the RNA templates. The mixture was then incubated at 65°C for 5 minutes, followed by 4°C for 1 minute. The enzyme mix used for reverse transcription was prepared as shown in table 1 below. Following incubation, 15 µl of the enzyme mix was added immediately, followed by incubation at 50°C for 60 minutes and 70°C for 17 minutes and held at 4°C. The generated cDNA was used either immediately in PCR reactions or stored at -70°C.

Table 1. **Enzyme Mix used for Reverse Transcription.**

Component	1X Reaction	Final Concentration
DEPC treated water	5 µl	-
5 x First Strand Buffer	4 µl	1 x
MgCl ₂ (25 mM)	3 µl	3.75 mM
dNTP (10 mM)	1 µl	0.5 mM
RNaseOUT (40 U/µl)	1 µl	2 U
ImProm-II™ Reverse Transcriptase (200 U/µl)	1 µl	10 U
TOTAL volume	15 µl	

3.2.2.2 PCR Amplification

A 469 bp product, corresponding to codons 431–560 in RT was amplified by nested PCR. The first round of PCR was amplified using AmpliTaq Gold polymerase (Invitrogen, CA, EUA) with the upstream primer: MS1 [5'-tatacgaagccacctggattc-3' (HXB2 positions 3762-3783)] and the downstream primer: MS2 [5'-cagtctactgtgccatgcatggcttc-3' (HXB2 positions 4371-4396)] (Santos *et al.*, 2008). The second round of PCR was performed using primers (MS3) 5'-ggtaccagttagagaaagaacca-3' (HXB2 positions 3826-3849) and MS4 5'-cattgcctctccaattactgtgatatttctcatg-3' (HXB2 positions 4263-4295). The reaction conditions are shown in Table 2.

Table 2. **1st and 2nd Round PCR Master Mix.**

<u>Components</u>	<u>1X</u>	<u>Final concentration</u>
10X Buffer	2.5 µl	1X
25 mM MgCl ₂	1.5 µl	1.5mM
10 mM dNTPs	0.5 µl	0.2mM
10 pmol <i>MSF</i> primer	0.5 µl	0.2µM
10 pmol <i>MSR</i> primer	0.5 µl	0.2µM
AmpliTaq Gold DNA polymerase	0.25 µl	1.25U
Sterile distilled water	17.25 µl	-
cDNA	2.0 µl	-
Total volume	25 µl	

Amplification was carried out in a GeneAmp9700 Thermal Cycler (Applied Biosystem™, International, Inc., Switzerland). Cycling conditions were as follows: Initial denaturation at

94°C for 5 minutes; followed by 40 cycles of 94°C for 40 seconds, 58°C for 1 minute, followed by final extension of 72°C for 5 minute. Amplicons were electrophoresed in a 1% agarose gel to confirm the presence of the 469 bp amplicon corresponding to the RNase H domain. The size of the product was determined using the GeneRuler™ 1kb DNA Ladder (Fermentas International Inc).

3.2.2.3 PCR Purification

PCR products were purified using the ExoSAP-IT® PCR clean-up protocol (©USB Corporation, USA) and QiAquick gel purification kit (QIAGEN Inc.) according to the manufacturer's instructions.

3.2.2.3.1 ExoSAP-IT® PCR Clean-up Protocol

The ExoSAP-IT® PCR clean-up was performed as follows, 5µl of the PCR product was mixed with 2µl of ExoSAP-IT® to remove free excess primers and nucleotides. The mix was then incubated in a thermocycler at 37°C for 15 minutes, followed by 80°C for 1 minute and held at 4°C. The purified product was then stored at -20°C until ready for use.

3.2.2.3.2 Gel Purification Protocol

Where the PCR products contained non-specific bands, the PCR products were purified using the QiAquick gel extraction kit (QIAGEN Inc.). A sterile scalpel was used to cut out an agarose band containing the PCR product of interest. The amount of agarose gel was weighed and three volumes of Buffer QG (from the kit) were added to one volume of gel (100

mg \equiv 100 μ l). For example, 300 μ l Buffer QG was added to a 100 mg gel slice. The buffer together with the gel slice were mixed by inversion and incubated at 50°C for 15 minutes until the agarose slice was completely dissolved. During the incubation, the gel was mixed by inversion and vortexed every 3 minutes. Once the agarose gel was completely dissolved, the color of the sample mix was checked to visually indicate whether the buffer QG plus sample mix were at the optimal pH for DNA to bind to the silica membrane. For efficient binding of DNA to the silica membrane the buffer QG plus sample mix requires a pH \leq 7.5. One gel volume of isopropanol was added into the tube containing the sample and mixed. For example, if the agarose gel slice was 100 mg, 100 μ l of isopropanol was added.

A QIAquick spin column was placed into a collection tube and the sample mix was transferred onto the assembled QIAquick spin column and incubated at room temperature for 1 minute to bind the DNA onto the silica membrane. The column was then centrifuged at 16,000 x g for 1 minute. The flow through was discarded by emptying the collection tube and the column was placed back inside the collection tube. This sample binding step was repeated as necessary until the entire sample was loaded. Traces of Agarose were removed by adding 500 μ l of Buffer QG into the column and centrifuged at 16,000 x g for 1 min. The flow through was discarded by emptying the collection tube. Seven hundred and fifty microlitres of Buffer PE was transferred into the column and incubated at room temperature for 3 minutes and centrifuged at 16,000 x g at 22°C for 1 minute. The flow through was discarded and the column was placed back inside the collection tube and continued with centrifugation for another 1 min. The collection tube with flow through was discarded and the QIAquick spin column was transferred to a fresh DNase free 1.5ml microcentrifuge tube. The DNA was eluted by adding 50 μ l of DNase free water to the center of the column and incubated for 1 minute before centrifuging at 16,000 x g for 1 minute.

The purified PCR product was subsequently quantified by electrophoresis in a 1% agarose gel with a DNA Low Mass ladder (Invitrogen®). The concentration of the PCR product was determined by comparing the intensity of the amplicon band to a band with known concentration from the DNA Low Mass ladder (Invitrogen®) (Figure 20). The concentrations of the bands of the low mass ladder are as follows:

Table 3. **The amount of each DNA fragment using 4µl of Low DNA Mass Ladder.**

Fragment Size	Amount of DNA (ng)
2000	200
1200	120
800	80
400	40
200	20
100	10

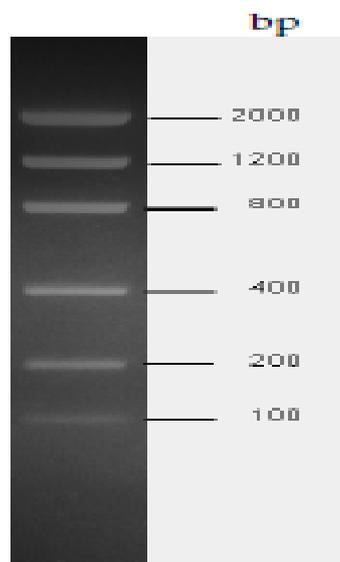


Figure 20. **2.0 % Agarose gel loaded with 4 µl of Low DNA Mass Ladder stained with ethidium bromide (Invitrogen insert, 2006).**

3.2.3 Sequencing of the RNase H region

The sequencing reaction was performed using 0.4 μ l pre-mixed BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), 2 μ l of 5 X Sequencing Buffer, 3.1 μ l sterile DNase free water, 2 μ l of DNA (4 ng/ μ l), and 1.6 pmol second round primers (MS3 and MS4). Cycling conditions were as follows: 96°C for 1 minute; 35 cycles of 96°C for 10 seconds, 50°C for 5 seconds, 60°C for 4 minutes and finally held at 4°C. The sequencing reaction product was then purified using either the plate cleanup or tube clean up method.

3.2.3.1a The Plate Clean Up

Briefly, 1 μ l of 125mM EDTA (pH 8.0) was added into each well and mixed by pipetting up and down 10 times to ensure that the EDTA was completely dissolved. Twenty six microlitres of 3M sodium acetate (NaOAc) (pH 5.2) mixed with 100% ice cold ethanol was also added into each well. The plate was sealed, mixed by lightly vortexing and centrifuged at 3000 x g for 20 minutes. The plate was then dried by carefully inverting it onto paper towel and centrifuging at 150 x g for 5 minutes. Immediately, 35 μ l of 70% cold ethanol was added to each well and centrifuged at 3000 x g for 5 minutes. A paper towel was placed onto the plate and carefully inverted and centrifuged at 150 x g for 1min to dry. The plate was placed in the dark at room temperature for 5 minutes to dry. It was then sealed with an adhesive cover and wrapped with foil to protect it from light and stored at -20°C.

3.2.3.1b The Tube Clean Up

Briefly, an orientation mark was placed on each PCR tube, followed by the addition of 80 μ l of freshly prepared 75% isopropanol. The tubes were then vortexed lightly and incubated for

15 minutes at room temperature in the dark. The tubes were then placed such that the orientation mark was facing the outside rim of the rotor and centrifuged at maximum speed for 20 minutes. Immediately, the supernatant was carefully aspirated in order not to disturb the pellet and 200µl of freshly prepared 70% ethanol was added to each tube. The tube was vortexed, centrifuged at 16,000 x g for 5 minutes at room temperature and aspirated. Again 200µl of freshly prepared 70% ethanol was added to each tube, vortexed and spun at 16,000 x g for 5 minutes at room temperature. Immediately, the supernatant was quick spun to pellet the contaminants and carefully aspirated to remove residual ethanol. The tube was placed in the dark at room temperature for 5 minutes to dry and stored at -20°C in the dark.

3.2.3.1c Preparing Sequencing Products

The sequencing products were reconstituted by adding 10µl of Hi-Di formamide to the dried pellets in each well and pipetting up and down 10 times; the plate was then sealed with an adhesive cover and quick spun at 1000 x g to collect the sequencing products at the bottom of the wells. The DNA product was denatured at 94°C for 2 minutes and held at 4°C. The plate was spun briefly at 1000 x g to collect the sequencing products at the bottom of the wells. Samples were run on an Applied Biosystem (ABI) 3130xl automated sequencer in the Pfizer lab, Doris Duke Medical Research Institute, University of KwaZulu-Natal.

3.2.4 Sequence Data Analysis

Sequence analysis was performed on a Linux Computer. Sequences were assembled to form a contig using the Phred and Phrap programs and chromatograms were viewed and manually edited using the consed tool (<http://www.phrap.org/phredphrapconsed.html>) (Ewing *et al.*,

1998). Multiple sequence alignments were generated using the ClustalW program (<http://www.clustal.org/>) (Larkin *et al.*, 2007) and manually edited using the Genetic Data Environment v2.2 (GDE) (<http://www.bioafrica.net/GDElinux/GDEmicrobial.html>) (de Oliveira *et al.*, 2003).

Subtypes were determined by drawing phylogenetic trees using Paup*4b10 package using reference strains downloaded from the Los Alamos HIV database (<http://www.HIV.lanl.gov/content/sequence/align.html>). To rule out cross-contamination between specimens, each newly generated sequence was aligned (using ClustalX) with all previously sequenced samples and a phylogenetic tree was drawn using the Paup*4b10 (<http://paup.csit.fsu.edu/>) package and visualized in FigTree v1.1.2 (<http://tree.bio.ed.ac.uk/>).

3.2.5 Statistical and Bioinformatics Analysis

3.2.5.1 Graphpad Prism®

All statistical analysis was performed using GraphPad Prism® 5 software (Graphpad© Software, Inc). Chi-square (and Fisher's exact) test was performed to compare the frequency of mutations in the RNase H in isolates of treatment naïve and NRTI experience patients. A *p*-value less than 0.05 was considered significant. Spearman's rank correlation (nonparametric correlation) was used to explore the relationship between mutations occurring within the RNase H domain and NRTI mutations.

3.2.5.2 B-Right

A Bayesian-based graphical modelling software program, B-Right (<http://www.mybiodata.eu/B-Right-0.1.zip>) was used to explore data for multivariate probabilistic independencies. The programme was used to model the association between the RNase H domain mutations and selected key NRTI resistance mutations during antiretroviral therapy. Bayesian network learning of HIV-1 sequence data was also used to analyze the direct and indirect correlation between the RNase H domain mutations and exposure to NRTI-treatment. Non-parametric bootstrap was calculated by running 100 replicates to derive network robustness and only interactions with bootstrap support over 65% were included for analysis. The Bayesian Network for this study was done in collaboration with Kristof Theys (kristof.theys@rega.kuleuven.be) from the Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium.

Chapter 4: Results

4.1 Optimization of PCR of the RNase H region

PCR for the amplification of the RNase H region was optimized by titrating the primer concentrations (Figure 21). Primers were titrated at the following concentrations: 0.1 μ M, 0.2 μ M, 0.4 μ M and 0.8 μ M (lanes 2 to 5 respectively). A 100bp Ladder electrophoresed in lane 1 and a negative control in lane 6. The optimum concentration taken as 0.2 μ M (lane 3).



Figure 21. **Agarose gel picture of the primer titration at 4 concentrations.**

4.2 Characterization of the RNase H Domain

The RNase H domain (RT codons 439 to 560) sequences of isolates from 134 NRTI treated but virologically failing [South African HIV positive adults (104) and children (30)] and 30 ART-naïve viral isolates were generated and analyzed. In addition, sequences from 104 subtype C drug-naïve isolates were downloaded from the Los Alamos HIV-1 sequence database (<http://www.HIV.lanl.gov/content/sequence/align.html>). This gave a total of 134 sequences from ART-naïve individuals for the analysis.

4.2.1 Phylogenetic Analysis

The subtypes were confirmed by phylogenetic analysis and 130/134 samples were classified as subtype C, with one subtype A, two subtype B and two subtype D. No specific clustering of the study sequences were observed (Figure 22). Sequences from the 30 ART-naïve patients were classified as subtype C (Figure 22).

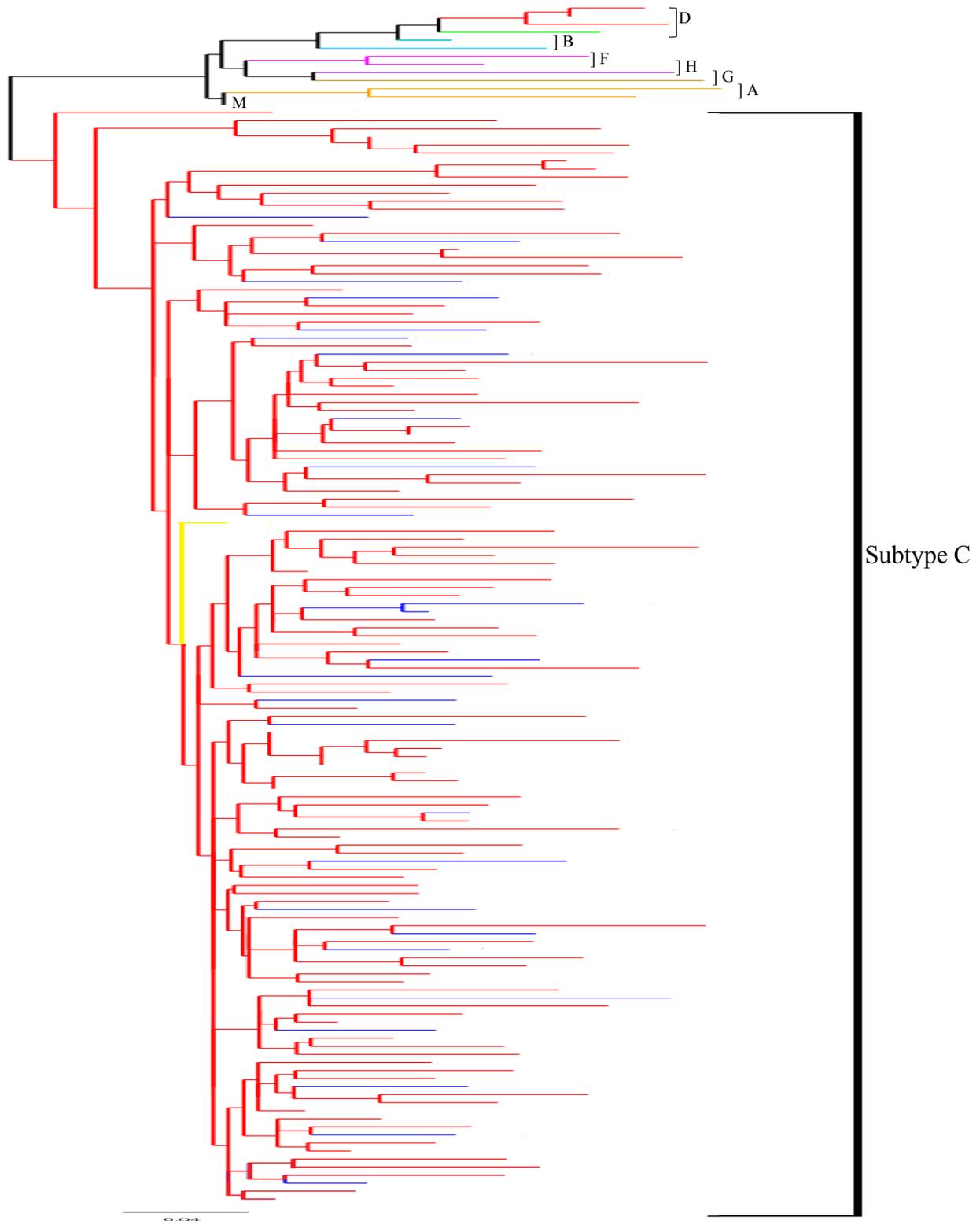


Figure 22. **Maximum Likelihood Tree (1000 replicates for Bootstrap) indicating the phylogenetic relations between all the sequences that had all the RNase H codons.** Red- NRTI-experienced, royal blue- ART-naïve sequences, Orange-subtype A references, Blue- subtype B reference, Yellow- subtype C references, Green- subtype D reference, Pink- subtype F references and Purple - subtype H references and Black- M group.

4.2.2 Amino acid sequence variation in the RNase H in subtype C.

Amino acids (from amino acid 440 to 560) for each subtype C sequence were evaluated by their degree of codon polymorphisms and compared with the HXB2 reference strain to identify naturally occurring mutations in the RNase H region. Chi-square (and Fisher's exact) test was used to calculate the frequencies of the RNase H mutations in both the ART-naïve and NRTI treated isolates. A similar number of highly conserved sites ($\leq 1\%$ variation) was found between NRTI treated 85/130 (65%) and ART-naïve isolates 84/130 (65%). Twenty (1.5%) of the 130 amino acids found in the RNase H region were highly polymorphic ($> 50\%$ variability) at positions 435, 452, 466, 468, 471, 483, 491, 519, 530 and 554 in both groups. Positions 446, 448, 469, 476, 488, 502, 513 and 559 had 2 – 25% variation in ART-naïve isolates, but had a high degree of conservation ($\leq 1\%$ variation) in NRTI treated isolates (Figures 23a and 23b).

Analysis of the residues at the primer grip motif (T473, Q475, Q500, Y501, and I505) of RNase H showed that they were highly conserved ($\leq 1\%$ variation) in to both groups with an exception of positions R448, N474 and K476 that showed $\geq 2\%$ variation found normally in ART-naïve (Figures 23a and 23b). Positions D443, E478, D498, S499, H539, N545 and D549 of the catalytic site were also highly conserved ($\leq 1\%$ variation). The RNase H motif (E438, T439, F440, Y441, V442, D443, G444 and A445) which overlaps the protease cleavages site (A437, T439, F440, Y441, V442 and D443) had a high degree of conservation ($\leq 1\%$ variation) with the exception of position E438 that presented a low degree of variability (6% variation) in ART-naïve isolates.

K431 T	E432 D	P433	I434 M	V435 AIMPT	G436 EQ	A437 DEIQRTV	E438	T439 P	F440	Y441
V442	D443	G444 E	A445	A446 VY	N447 S	R448 GK	E449 GKN	T450 N	K451 R	L452 AEILKMQ RSTV
G453	K454 R	A455	G456	Y457	V458 I	T459	D460 N	R461 IKT	G462 K	R463
Q464 H	K465 R	V466 AFGILT	V467 IM	S468 AHLNPQT	L469 FI	T470 ADGNPS	D471 EG	T472 A	T473 ASP	N474 D
Q475 HL	K476 EQR	T477 AS	E478	L479	Q480 EHK	A481 P	I482 FV	H483 LMNQY	L484 I	A485
L486 F	Q487 KP	D488 N	S489	G490	L491 AMPQSTV	E492 AKV	V493 AI	N494 HY	I495 Y	V496 F
T497 S	D498 HK	S499 A	Q500	Y501 N	A502 GV	L503	G504	I505	I506	Q507
A508 GS	Q509 HL	P510	D511	K512 NRS	S513 G	E514 D	S515 L	E516	L517 SV	V518 A
S519 NT	Q520 KPN	I521	I522	E523	Q524 ER	L525	I526	K527 ENQR	K528	E529 DK
K530 ER	V531 IT	Y532 F	L533	A534 S	W535	V536 G	P537	A538	H539	K540
G541	I542 K	G543	G544	N545	E546 V	Q547	V548 AI	D549	K550	L551
V552 N	S553 N	A554 DEGKN RST	G555 D	I556	R557 IG	K558 ER	V559 I	L560 F		

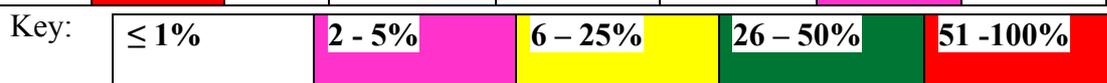


Figure 23a. **Amino acid sequence variation in the RNase H region from ART naïve isolates.** Amino acid residues are represented as the HXB2 consensus.

E431	E432 D	P433	I434 MV	V435 AEILMPST	G436 ER	A437 DEIQRTV	E438 GKR	T439	F440	Y441
E442	D443 G	G444	A445	A446 SY	N447 DHIST	R448 G	E449 GKN	T450 AN	K451 ERT	L452 AEILKM RSTV
E453 EK	K454	A455	G456	Y457 A	V458 AILT	T459 A	D460 A	R461 IKT	G462 Q	R463 GK
E464 H	K465 ENR	V466 AFGILT	V467 AEIM	S468 AHLNPQT	L469 FIMVT	T470 ADGLNSPT	D471 AEGN	T472	T473 S	N474 D
E475 HL	K476 R	T477 AS	E478	L479	Q480 EHIKLR	A481 TV	I482 V	H483 LNQRSY	L484 I	A485
E486 F	Q487 H	D488	S489 P	G490 ER	L491 AMPQSTV	E492 Q	V493 AIL	N494	I495 V	V496 I
E497 AS	D498 H	S499	Q500	Y501	A502	L503	G504	I505	I506	Q507 R
E508 GST	Q509 HKLR	P510	D511	K512 DEHNQ RST	S513 G	E514 D	S515 I	E516 G	L517 V	V518 A
E519 GNRT	Q520	I521	I522	E523 G	Q524 EKLR	L525	I526 K	K527 EGNQRS	K528	E529 D
E530 ER	V531 AIT	Y532 N	L533	A534 ST	W535	V536 G	P537	A538	H539	K540
E541	I542 K	G543	G544	N545	E546 V	Q547 LHK	V548 AI	D549	K550	L551 I
E552 N	S553 N	A554 DEGKN RST	G555 D	I556 L	R557 IG	K558 E	V559 C	L560 F		

Key:

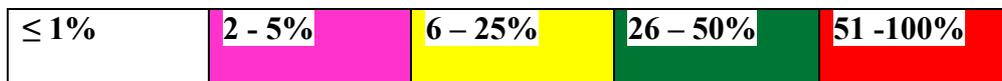


Figure 23b. Amino acid sequence variation in the RNase H region from NRTI experience isolates. Amino acid residues are represented as the HXB2 consensus.

4.2.3 Prevalence of RNase H mutations in NRTI-treatment-experienced and ART-naïve subtype C isolates

Five mutations in the RNase H region had significantly higher frequency when comparing ART-naïve and NRTI-experienced patients. These were: (E438GKR, L517ISV, K527GENQR, E529DK and Q547HKR) (Table 4). Three mutations (E432D, A446SVY and Q507HK) showed decreased proportions in treatment-experienced isolates when compared to ART-naïve isolates. E438GKR was seen in 6.72% of treated versus only 0% of naïve isolates ($p= 0.0034$), L517IV was found in 17.16% of treated isolates versus 7.46% of naïve isolates ($p= 0.0245$). Similarly, K527GENQRS was found in 41.04% of treated isolates versus 26.12% of naïve isolates ($p= 0.0138$), and E529DK was more prevalent in treated (17.91%) when compared to 2.99% of naïve subtype isolates ($p <0.001$). Finally, Q547HKR was seen in 5.22% of treated versus 0% of naïve subtype C patients.

Differences at three positions were statistically (with a p -value <0.05) more prevalent in naïve patients than in pre-treated patients: 432, 446 and 507 (Table 4). Mutation E432D had a statistically significant lower frequency in the treatment-experienced (0.75%) than in naïve patients (8.96%) ($p= 0.0027$). Likewise, the frequency of the Q507HK mutation was significantly lower in the treatment-experienced isolates (0%) when compared to drug naïve (4.48%) subtype C isolates ($p= 0.0295$), and A446SVY had decreased proportions in treatment-experienced isolates (0%) compared to naïve isolates (5.22%) ($p= 0.0144$).

Table 4. RNase H mutations in HIV-1 Subtype C-infected naive versus experienced participants.

Mutations in the RNase H domain	Frequency Subtype C naives (n=130)	Frequency Subtype C NRTI-treated (n=130)	<i>p</i> - value
E432D	8.96%	0.75%	0.0027
V435AEILMPST	90.30%	88.06%	0.6947
A437DEIQRTV	26.87%	36.57%	0.115
E438GKR	0%	6.72%	0.0034
T439P	0%	0.75%	1
F440Y	0%	0%	0
Y441C	0%	0%	0
V442I	0%	0%	0
D443GN	0%	0%	0
G444E	1.49%	0.75%	1
A445V	0%	0%	0
A446SVY	5.22%	0%	0.0144
R448GK	2.24%	0%	0.2472
L452AEIKMQRTV	77.61%	69.40%	0.166
D460AN	1.49%	0.75%	1
V466AFILT	66.42%	58.21%	0.2074
S468AHNPQT	59.7%	56.72%	0.7103
L469FIV	10.45%	17.16%	0.1559
T470ADGNPS	47.76%	52.99%	0.4636
D471EG	98.51%	96.27%	0.4469
T473ASP	0.75%	0%	1
N474DP	1.49%	2.99%	0.6839
Q475HLR	0.75%	1.49%	1
K476EQR	3.73%	0.75%	0.2135
H483KLMNQRY	97.01%	94.03%	0.3766
D488EN	2.24%	0.75%	0.4978
L491APQS	88.06%	88.81%	1

Note: numbers in bold represent significant *p*-values ≤ 0.05 level. Amino acids in bold are only present in naive patients and never in pre-treated patients.

Table 4 (continued). **RNase H mutations in HIV-1 Subtype C-infected naive versus experienced participants.**

Mutations in the RNase H domain	Frequency Subtype C naives (n=130)	Frequency Subtype C NRTI-treated (n=130)	p - value
Q500R	0.75%	0%	1
Y501N	0%	1.49%	0.4981
A502GV	2.99%	0.75%	0.3703
I505T	0.75%	0%	1
I506L	1.49%	0%	0.4981
Q507IK	2.99%	0.75%	0.3703
S513G	2.24%	0%	0.2472
L517ISV	7.46%	17.16%	0.0245
S519NT	96.27%	92.54%	0.2877
Q524EKLR	20.15%	31.34%	0.05
K527GENQRS	26.12%	41.04%	0.0138
E529DK	2.99%	17.91%	< 0.0001
K530R	89.55%	87.31%	0.703
V531AIT	20.90%	17.91%	0.6434
A534S	98.51%	94.03%	0.1028
Q547HKR	0%	5.22%	0.0144
A554HGKNQRST	100%	97.01%	0.1222
K558ER	1.49%	2.99%	0.6839
V559CI	2.99%	0.75%	0.3703

Note: numbers in bold represent significant p-values ≤ 0.05 level. Amino acids in bold are only present in naive patients and never in pre-treated patients.

The RNase H mutations (H539N and D549N) that were found to confer NRTI- resistance by a mechanism that allows more time for NRTI excision in subtype B were never found among the NRTI-treated subtype C sequences studied. Mutations L469, L491 and Q509 also associated with TAMs were extensively polymorphic in both treatment naïve and treatment-

experienced patients. All other position (N460, P468, L469, T470, H483, I506, Q512, S519, Q524, K530, A554, and K558) described in HIV-1 subtype B studies to affect the functionality of enzymatic processes, were not statistically significant in this study. In subtype C, twenty treatment-experienced samples had mutations in the RNase H domain but did not have any of the classical NRTI mutations (Table 5).

Table 5. Frequency of resistance-associated mutations in the RNase H domain in isolates that did not harbor any of the classical NRTI mutations.

Sequence ID	NRTI Mutations	NNRTI Mutations	PI Mutations	RNase H Mutations	Subtype
sw67	none	none	none	V435E, A437V, N460D , V466I, T468S, T470S , D471E, Y483Q, L491S, Q512K , K527R , K530R , A534S, A554S ,	C
sw84	none	none	none	V435A, G436E, N460D , V466I, L469I, T470S , D471E, T477AT, Y483L, L491S, Q512K , K527N , K530R , V531I, A534S, A554S	C
sw91	none	none	none	V435A, L452I, V458L, N460D , V466I, T468S, D471E, Q480KQ, Y483Q, L491S, Q512K , Q524EQ, K530R , A534S, V548I, A554S	C
sw40	none	none	none	V435A, G436E, L452I, N460D , V466I, V467I, T470S , D471E, Y483Q, L491S, Q512K , K530R , A534S, A554S	C
sw45	none	none	none	L452K, N460D , V466I, V467I, T468S, T470DGNS , D471E, Y483Q, L491P, Q512K , K530R , A534S, A554S	C
sw59	none	none	none	V435A, E449D, N460D , V466I, T468S, D471E, Y483Q, L491P, Q512K , K530R , A534S, A554S , K558KR	C
sw6	none	none	none	V435A, L452I, N460D , V466I, T468S, T470S , D471E, Y483Q, L491S, Q512R , K530R , A534S, V548I, A554S	C
iwn1	none	none	none	V435A, L452I, N460D , R461K, V466I, T468S, T470N , D471E, Y483Q, L491S, Q512K , Q524L, E529D, K530R , A554S	C
iwn2	none	none	none	V435A, E449D, L452K, K454R, N460D , V467I, T470N , D471E, Y483Q, L491S, Q512N , L517V, K527N , K530R , A534S, A554S	C

NB: Mutations in bold were associated with NRTI resistance in subtype B

Table 5 (continued). **Frequency of resistance-associated mutations in the RNase H domain in isolates that did not harbor any of the classical NRTI mutations.**

Sequence ID	NRTI Mutations	NNRTI Mutations	PI Mutations	RNase H Mutations	Subtype
ich12	none	none	none	I434M, V435A, A437V, N460D , T470AT , D471E, Y483Q, L491P, Q512K, Q524E, K530R, A534S, A554N	C
ich16	none	none	none	V435A, A437T, E449G, L452I, N460D , R461K, T470N , D471E, Y483Q, L491S, Q512K , K530R , A534S, A554K	C
ich22	none	none	none	V435A, L452I, N460D , V466I, T468S, T470S , D471E, T472A, Y483Q, L491S, Q512N , L517I, K527N , K530R , A534S, A554S	C
ich8	none	none	none	V435L, L452I, N460D , T468A, D471E, Y483Q, L484I, Q512K , Q524E, K527N , K530R , A534S, A554N	C
ich9	none	none	none	E432D, V435A, E449D, L452R, N460D , V466I, T468P, T470S , D471E, Q480H, Y483H, L491S, Q512K , K530R , A534S, A554R , R557I	C
pke10f	none	none	none	V435A, A437D, L452M, N460D , V466I, T468S, T470S , D471E, Y483Q, L491S, Q512K , K527R , K530R , A534S, A554S	C
pke12f	none	none	none	V435A, A437Q, E438K, K454R, N460D , R461K, V466I, T468S, T470S , D471E, T477A, Y483L, G490R, Q512K , K527N , K530R , A534S, A554K	C
pke18f	none	none	none	V435A, L452I, N460D , V466I, T470S , D471E, Y483Q, L491S, Q512K , Q524E, E529D, K530R , V531IV, A534S, A554S	C
pke19f	none	none	none	V435I, L452I, N460D , T468P, Y483H, L491LS, Q512K , N519S	C
pke31f	none	none	none	I434M, V435T, A437V, L452I, N460D , L469I, D471E, Q480H, Y483Q, L491S, Q509L, Q512K , N519S, Q524E, E529D, K530R , A534S, A554S	C
pke33f	none	none	none	V435A, L452I, N460D , V466I, T468S, L469V, T470S , D471E, Y483Q, L491S, Q509H, Q512K , Q524E, E529D, K530R , V53II, A534S, A554S	C

NB: Mutations in bold were associated with NRTI resistance in subtype B

4.2.4 Comparison of RNase H mutations between Subtype B and C isolates in treatment naïve versus treatment experienced patients.

Subtype B (the most studied) was used as a reference for this study, and was subsequently also used to compare frequencies with subtype C (most prevalent) sequences. Two hundred and sixty subtype B sequences were downloaded from the Los Alamos HIV-1 sequence database (<http://www.HIV.lanl.gov/content/sequence/align.html>) of which, 130 were from treatment-naïve and 130 were from NRTI-treated patients.

Eleven mutations had significantly high frequencies in subtype C drug-naïve patients compared to subtype B drug-naïve patients. These were E432D ($p = <0.0001$), A437DEIQRTV ($p = <0.0001$), S468AHNPQT ($p = 0.0002$), T470ADGNPS ($p = 0.0004$), H483KLMNQRY ($p = <0.0001$), L517ISV ($p = <0.0001$), S519NT ($p = <0.0001$), Q524EKLR ($p = 0.0007$), K527GENQR ($p = 0.0134$), K530R ($p = <0.0001$) and A554HGKNQRST ($p = <0.0001$) (Figure 24).

Frequencies of NRTI-associated RNase H mutations were used to compare the difference between sequences subtypes B naïves and subtype C NRTI treated patients. Fourteen mutations had significantly different frequencies when treatment-experienced subtype C patients and subtypes B naïves patients (Figure 24). Of those, 10 sequences had increased prevalence in treated patients (A437DEIQRTV, S468AHNPQT, T470ADGNPS, H483KLMNQRY, Q524EKLR, K527GENQR, E529DK, K530R, S519NT and A554HGKNQRST) with the $p = <0.0001$ respectively, except for E438GKR with $p = 0.0034$. Three mutations showed decreased proportions in NRTI-treatment experienced subtype C patients when compared to subtypes B naïves patients (D460AN, K512DGHNQRST and L517ISV) with a p -value <0.0001 each.

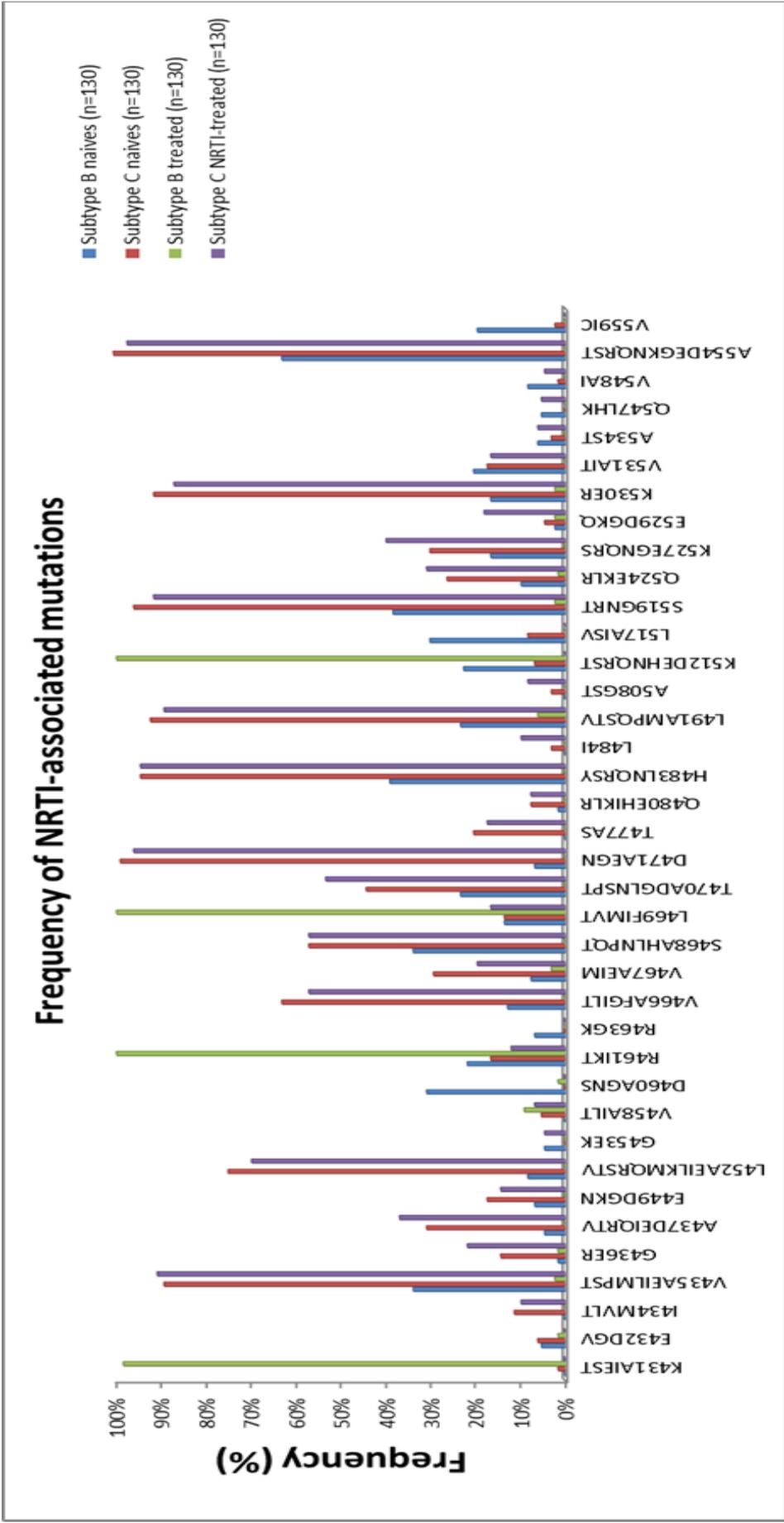


Figure 24. Frequency of the NRTI-associated mutations in the subtype B naïve, subtype B-treated, subtype C naïve, subtype C-treated.

4.2.4.1 Comparison of the RNase H sequence variability between subtype B and C isolates.

Twelve amino acid substitutions were identified in the consensus sequences from both KZN naïve and KZN treated subtype C patients when compare to HXB2 reference strain (Figure 25). The RNase H domain consensus sequences of KZN isolates differed from the HXB2 or subtype B consensus by 9% and 7% respectively. And mutations at positions 435, 452, 460, 466, 468, 471, 483, 491, 512, 530, 534 and 554 (shown in yellow in Figure 25) were found in the consensus sequences of both treatment naïve and NRTI treated subtype C isolates, as well as in other subtypes other than HXB2. However, the selected amino acids substitutions were not found in the sites which are critical for the reverse transcriptase enzyme activity. For example, the catalytic (red), primer binding (grey) and sites that overlap the protease cleavages sites (green) were highly conserved in subtype B and C (Figure 25).

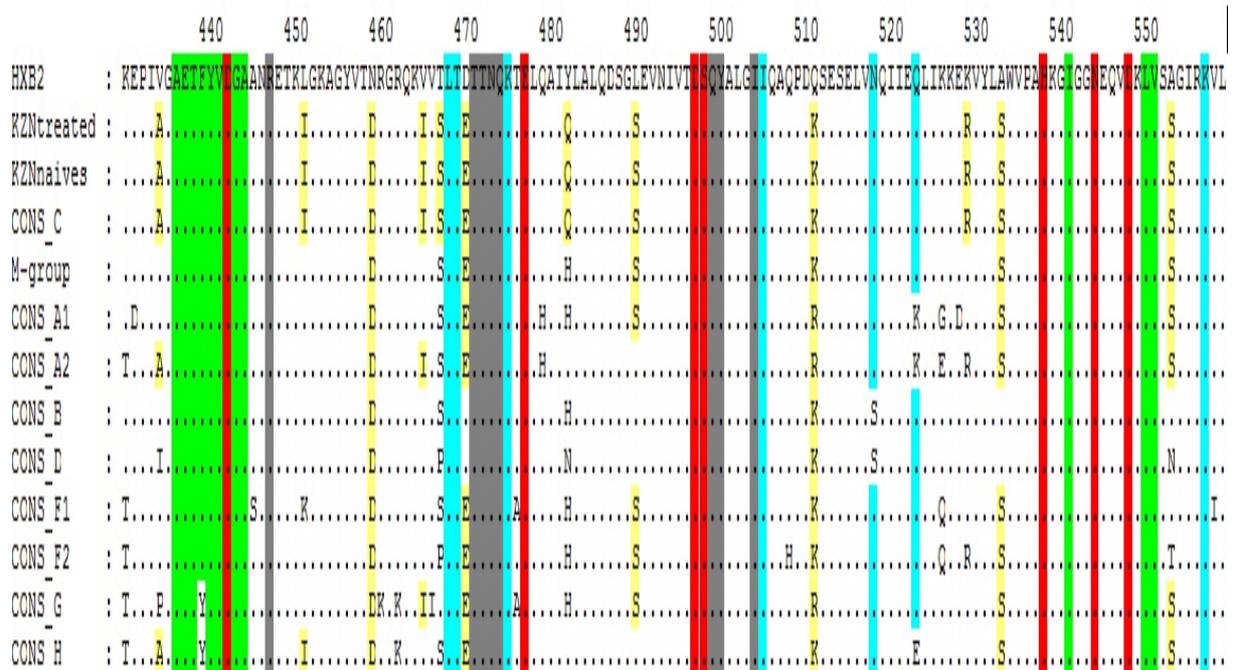


Figure 25. **Alignment of the consensus sequences of the HIV-1 group M subtypes with HXB2 and the KZN consensus sequences.** Putative functional sites were determined using PROSITE in GENEDOC. Functionally important RNase H sites are showed by the colored boxes: red – catalytic site; grey – primer grip; blue – other sites conferring resistance to AZT monotherapy; green – sites that overlaps the protease cleavages and subtype difference (amino acid) are shown in yellow color.

4.3 Bayesian Network Analysis

4.3.1 Dataset for Bayesian Network Learning

The initial dataset for BN learning included 304 (120 wild-type amino acid and 184 mutations) RNase H variables from each sequence. However, analysis of all the variables would have required very high computational power, and therefore only a select number of variables (those with frequencies greater than 14%) were analyzed. Subsequently, the number of variables used in this study was reduced to 70. Chi-square and Fisher’s exact were used to compare frequency of the C-terminal RT domain in treatment naïve and NRTI experience isolates. After stratifying the dataset using the Fisher’s test, the BN learning was performed

by searching the most credible network structure. The Bayesian network learning showed 15 variables with frequencies above 15% in treatment-experienced isolates: 13 were polymorphic sites and 2 were at sites associated with NRTI resistance. The polymorphic positions were V435, A437V, L452IM, R461K, V466, V467I, S468T, T477A, L491P, Q524E, K527N, K530, V531I and NRTI-associated mutations were T470S and E529D (Figure 26). Some of these positions (470, 527 and 530) were previously reported in subtype B and were associated with AZT-drug resistance (Santos *et al.*, 2008, Nikolenko *et al.*, 2007).

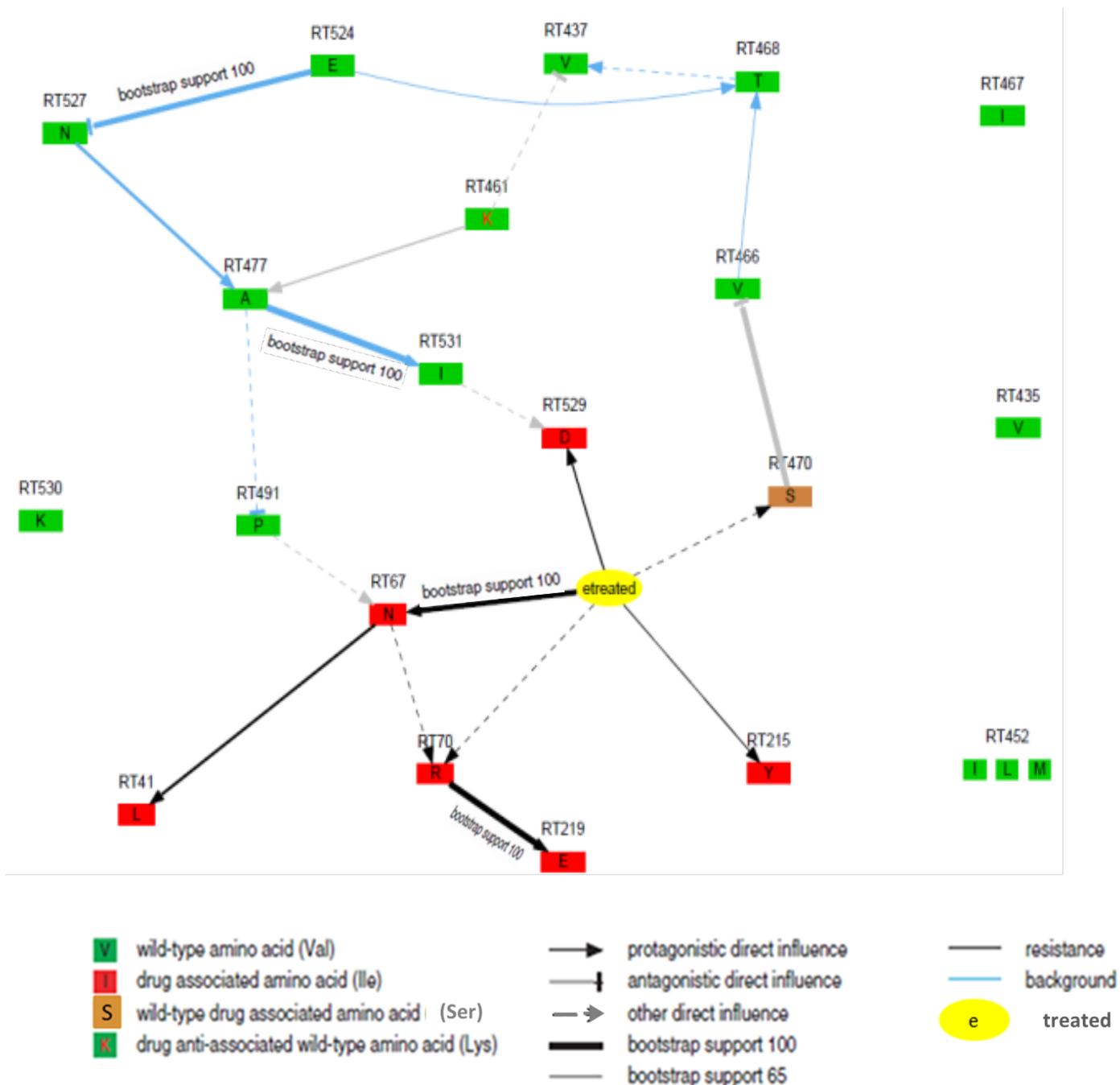


Figure 26. **Annotated NRTI experience Bayesian network (BN) expressing direct association between the NRTI-associated mutations in the RNase H domain and NRTI treatment.** An arc constitutes a direct dependency between the similar variables and robustness is relative to bootstrap value. Arc color shows antagonistic direct influence between NRTI-associated mutations (black), grey shows an influence from background polymorphisms on NRTI-associated mutations (blue) and green shows an association between background polymorphisms.

4.3.2 Treatment Experienced Bayesian Network Learning

The most probable network is shown in Figure 26 with 15 arcs and 8 bootstrap (over 65%). The BN was used to explore dependencies of the selected mutations to the treatment. Bayesian network showed that RNase H mutations T470S and E529D had a direct association with NRTI treatment. Mutation T470S ($p= 0.007$) was not dependent on any other mutations in the network while E529D appeared to be associated with drug resistance when in combination with V531I (Figure 26). In addition, mutation T470S showed a strong antagonistic effect on V466 with strong bootstrap support. Mutation L491P was associated with NRTI resistance in combination with the thymidine analogue mutation (RT67, RT70 and RT219) 2 pathway (TAM 2 pathway), but this had low bootstrap support. However, none of these statistically significant mutations (E438GKR, L517ISV, K527GENQR and Q547HKR) found in this study were associated with treatment in the BN.

Chapter 5: Discussion and Conclusion

South Africa is the epicenter of the HIV epidemic, with more than ten percent of the population being infected with HIV-1 (Department of Health, 2010), all of whom will eventually require ARV treatment. Currently, the South African national treatment program includes NRTIs in both first and second line highly active antiretroviral therapy regimens. The NRTIs target reverse transcriptase, a unique enzyme required for development of the virion. It consists of two enzymatic activities that include the RNase H cleavage activity and both the RNA dependent and DNA dependent DNA polymerization. The RNase H domain and DNA polymerase domain play significant roles throughout replication and are dependent on each other. However, the emergence of HIV-1 drug resistance against this treatment is a major concern for ARV therapy.

Previous studies showed that specific mutations in the *pol* region are responsible for NRTI drug resistance (Santos *et al.*, 2008 and Hirsch *et al.*, 2008). Most genotypic RTI resistance studies have analyzed approximately 300 amino acid of RT (Santos *et al.*, 2008). However, recent studies have shown that mutations in the RNase H domain can also increase resistance to NRTIs. Mutations in the RNase H domain can reduce the RNase H activity and confer high-level resistance to AZT and d4T by affecting the overall positioning of template:primer strand, template switching, polymerization and NRTI excision (Nikolenko *et al.*, 2005 and 2007, Delviks-Frankenberry *et al.*, 2008).

In this study, NRTI-associated mutations in the RNase H region of HIV-1 subtype C isolates were determined. Most of the amino acids present in the active site and primer grip were conserved, which is supported by literature (Santos *et al.*, 2008). However, there was variability at positions 474, 475, 476 and 501, but this was not significantly different to the

naïve samples. This suggests that any amino acid changes in these highly conserved positions may be detrimental in viral competency if selected *in-vivo* (Arion *et al.*, 2002, Volkmann *et al.*, 1993).

The RNase H sites overlapping the protease cleavage site were highly conserved with the exception of E438 that had a prevalence of 6.7% in treatment failures ($p= 0.0034$). Other RNase H mutations indicated in the literature to be associated with NRTI drug resistance were also included in our analysis. Some of these mutations were found in our samples, but at low frequencies, e.g. D460AN (0.75%), K476EQR (0.75%) and K558 (2.99%), and showed no significant difference to the untreated isolates. Mutations at position 435 (90.3%), 471 (98.51%), 483 (97.01%), 491 (88.06%), 519 (96.12%), 530 (89.55%), 534 (98.51%) and 554 (100%) were in the majority of naïve samples, indicating that these were not related to treatment failure.

When frequencies in subtype C ART naïve and NRTI treatment-experienced sequences were compared, five mutations (E438GKR, L517ISV, K527GENQR, E529DK and Q547HKR) had higher frequencies in treatment-experienced isolates (Table 4), the latter three being found close to the RNase H catalytic site. Mutation E438GKR was only prevalent in treatment experienced in contrast to naïve isolates, while mutations L517ISV was 10 times higher on NRTI-treated patients. Position 438 is in the proximity of the RNase H domain site that overlaps the protease cleavage site. In a study by Abram and Parniak, 2005, site-directed mutagenesis at this site resulted in defective viruses and impaired intergrase processing. Closer examination of these mutations is needed to assess their role in NRTI-resistance.

Positions 432, 446 and 507 showed significant decreased proportions in treatment-experienced isolates when compared to the ART naïves. This can be explained by the close

proximity of positions 446 and 507 to the RNase H primer grip, which plays a role in positioning the template:primer at the RNase H active site (Arts *et al.*, 1998). The other mutated amino acid E432D is found five amino acids away from the RNase H site that overlaps the protease cleavage site. Abram and Parniak, 2005, have shown by site-directed mutagenesis that mutations in or near the primer grip and protease cleavage site can indirectly produce defective virions.

Very little data exists on genetic diversity of the RNase H domain between subtype C and subtype B NRTI drug resistance expression. This study did not show any subtype-specific amino acid differences between the different subtypes at their functionally important sites. However, subtype-specific amino acid substitutions in functionally non-important sites were observed. These included: V435A, L452I, N460D, V466I, I468S, D471E, Y483Q, L491S, Q512K, K530R, A534S and A554S, some of which (Q512K, K530R and L491S) have previously been associated with NRTI exposure in subtype B (Santos *et al.*, 2008).

Using the Bayesian network to analyse direct associations between mutations and exposure to NRTI treatment sequences in the RNase H domain sequences, two resistance mutations were identified. The E529D mutation, which has not previously been associated with resistance, appeared to be dependent on two additional mutations in the pathway to resistance, namely the V531I mutation. This suggests that E529D can be classified as a minor or accessory mutation which increases resistance mostly in the presence of other mutations.

The second mutation T470S, was not dependent on any other mutations in the network, had good bootstrap support, suggesting that it could be classified as a major mutation (i.e. able to cause resistance on its own). Position 470 was adjacent to the RNase H primer grip which is

responsible for stabilizing the binding and alignment of the template at both the DNA polymerase and RNase H catalytic site (Sarafianos *et al.*, 2001). Therefore, it is reasonable to assume that the mechanism of resistance development by T470S could be through the disruption of the binding of RNase H enzyme to the template (Sarafianos *et al.*, 2001). In addition, mutation T470S showed a strong antagonistic effect on V466 with strong bootstrap support, suggesting that they are unlikely to occur together in the same patient.

The analysis of the BN data suggests that NRTI drug resistance could be caused by mutations (T470S and E529D) in the RNase H domain outside of the catalytic and primer binding motifs. As these mutations were not strongly associated with thymidine analogue mutations, it is possible that an alternate pathway to NRTI resistance exists in the RNase H domain. Interestingly, the BN did not directly associate some of the RNase H mutations (E438GKR, L517ISV, K527GENQR, E529DK and Q547HKR) that had high prevalence in NRTI-treated patients using the Fisher's exact test, with exposure to NRTIs.

In conclusion, results obtained from this study suggested that drug resistance could be caused by mutations in the RNase H domain either alone, or in combination with mutations in the *pol* region. Additionally, protein sequences of the RNase H domain in subtype B and C remained conserved in their functionally important primer grip and catalytic sites. However, there was distinct sequence diversity at sites not related to function, although these may still directly or indirectly affect viral fitness. Therefore, subtype specific amino acid differences should be considered when assessing RNase H mutations on NRTI resistance in HIV-1 subtype C.

Despite the limitations including small sample size, limited computer power for BN analysis, and the small numbers of NRTI-associated RNase H mutations that could not be detected in the BN, our genotypic data showed that RNase H region mutations of HIV-1 may contribute to drug resistance in HIV-1 subtype C infections. Therefore, further investigations using site-directed mutagenesis need to be conducted to determine whether the RNase H mutations (E438GKR, T470S, L517ISV, K527GENQR, E529DK and Q547HKR) affect viral fitness.

Chapter 6: References

Abram ME and Parniak MA. 2005. Virion instability of human immunodeficiency virus type 1 reverse transcriptase (RT) mutated in the protease cleavage site between RT p51 and the RT RNase H domain. *J. Virol* **79**: 11952–11961.

Adnan S, Balamurugan A, Trocha A, Bennett MSNGHL, Ali A, Brander C, Yang OO. 2006. Nef interference with HIV-1-specific CTL antiviral activity is epitope specific. *Blood*. **108**: 3414-9.

Allen TM and Altfeld M. 2003. HIV-1 superinfection. *J. Allergy Clin Immunol.* **112**: 829-35.

Arion D, Sluis-Cremer N, Min KL, Abram ME, Fletcher RS, Parniak MA. 2002. Mutational analysis of Tyr-501 of HIV-1 reverse transcriptase. Effects on ribonuclease H activity and inhibition of this activity by Nacylhydrazones. *J. Biol Chem.* **277**:1370–1374.

Arts EJ, Le Grice SF. 1998. Interaction of retroviral reverse transcriptase with template-primer duplexes during replication. *Mol. Biol.* **58**:339-393.

Bartlett JM and Stirling D. 2003. A Short History of the Polymerase Chain Reaction. In: *Methods Mol Biol.* **226**:3-6

Chan DC, Fass D, Berger JM, Kim PS. 1997. Core Structure of gp41 from the HIV Envelope Glycoprotein (PDF). *Cell* **89**:263–73.

Cohen GB, Gandhi RT, Davis DM, Mandelboim O, Chen BK, Strominger JL, Baltimore D. 1999. The selective down regulation of class I major histocompatibility complex proteins by HIV-1 protects HIVinfected cells from NK cells. *Immunity*. **10**: 661-71.

Collins KL, Chen BK, Kalams SA, Walker BD, Baltimore D. 1998. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature*. **391**: 397-401.

Cunningham S, Ank B, Lewis D, Lu W, Wantman M, Dileanis JA, Jackson JB, Palumbo P, Krogstad P, Eshleman SH. 2001. Performance of the Applied Biosystems ViroSeq Human Immunodeficiency Virus type 1 (HIV-1) genotyping system for sequence-based analysis of HIV-1 in pediatric plasma samples. *J. Clin Micro*. **39**:1254-7.

Deforche K. 2008. Slides on bioinformatics workshop on virus evolution using Bayesian Network learning: Cape Town, South Africa.

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

Delviks-Frankenberry KA, Nikolenko GN, Boyer PL, Hughes SH, Coffin JM, Jere A, Pathak VK. 2008. HIV-1 reverse transcriptase connection subdomain mutations reduce template RNA degradation and enhance AZT excision. *Proc Natl Acad Sci USA*. **105**: 10943-10948.

de Oliveira T, Miller R, Tarin M, Cassol S. 2003. An Integrated Genetic Data Environment (GDE)-Based LINUX Interface for Analysis of HIV-1 and Other Microbial Sequences. *Bioinfo.* **19**: 153-4.

Department of Health, South Africa. 2008. National Comprehensive HIV and AIDS plan statistics, South Africa. Pretoria.

Department of Health, South Africa. 2010. The National Antenatal Sentinel HIV and Syphilis Prevalence Survey, South Africa. Pretoria.

DeStefano JJ, Mallaber LM, Fay PJ, Bambara RA. 1994b. Quantitative analysis of RNA cleavage during RNA directed DNA synthesis by human immunodeficiency and avian myeloblastosis virus reverse transcriptases. *Nucleic Acids Res.* **22**:3793–3800.

DeStefano JJ, Mallaber LM, Rodriguez-Rodriguez L, Fay PJ, Bambara RA. 1992. Requirements for strand transfer between internal regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase. *J. Virol.* **66**:6370–6378.

Douek DC, Roederer M, Koup RA. 2009. Emerging concepts in the immunopathogenesis of AIDS. *Annu. Rev Med.* **60**:471–84.

Ehteshamil M and Götte M. 2008. Effects of Mutations in the Connection and RNase H Domains of HIV-1 Reverse Transcriptase on Drug Susceptibility. *AIDS Reviews.* **10**: 224-35

Evans DB, Brawn K, Deibel MR Jr, Tarpley WG, Sharma SK. 1991. A recombinant ribonuclease H domain of HIV-1 reverse transcriptase that is enzymatically active. *J. Biol Chem.* **266**:20583–20585.

Ewing B, Green P. 1998. Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8**:186-194.

Finston WI and Champoux JJ.1984. RNA-primed initiation of Moloney murine leukemia virus plus strands by reverse transcriptase in vitro. *J. Virol.* **51**:26–33.

Fleming, P. Byers RH, Sweeney PA, Daniels D, Karon JM, Janssen S. 2002. HIV prevalence in the United States, 2000 [Abstract 11]. Presented at the Ninth Conference on Retroviruses and Opportunistic Infections. Seattle, WA. February 24-28.

Fuentes GM, Rodriguez-Rodriguez L, Fay PJ, Bambara RA. 1995. Use of an oligoribonucleotide containing the polypurine tract sequence as a primer by HIV reverse transcriptase. *J. Biol Chem.* **270**:28169–28176.

Furfine ES and Reardon JE. 1991a. Human immunodeficiency virus reverse transcriptase ribonuclease H: specificity of tRNA(Lys³)-primer excision. *Biochemistry* **30**:7041–7046.

Furfine ES and Reardon JE. 1991b. Reverse transcriptase. RNase H from the human immunodeficiency virus Relationship of the DNA polymerase and RNA hydrolysis activities. *J. Biol Chem.* **266**:406–412.

Gandhi SK, Siliciano JD, Bailey JR, Siliciano RF, Blankson JN. 2008. Role of APOBEC3G/F-mediated hypermutation in the control of human immunodeficiency virus type 1 in elite suppressors. *J. Virol.* **82**: 3125-30.

Gao G, Orlova M, Georgiadis MM, Hendrickson WA, Goff SP. 1997. Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc Natl Acad Sci USA.* **94**:407–411.

Gao HQ, Boyer PL, Arnold E, Hughes SH. 1998. Effects of mutations in the polymerase domain on the polymerase, RNase H and strand transfer activities of human immunodeficiency virus type 1 reverse transcriptase. *J. Mol Biol* **277**:559–572.

Gelderblom, H. R 1997. Fine structure of HIV and SIV. in Los Alamos National Laboratory. HIV Sequence Compendium. Los Alamos National Laboratory. 31–44. <http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/1997/partIII/Gelderblom>.

Gibbs JS, Lackner AA, Lang SM, Simon MA, Sehgal PK, Daniel MD, Desrosiers RC. 1995. Progression to AIDS in the absence of a gene for vpr or vpx. *J Virol.* **69**: 2378-2383.

Gilboa E, Mitra SW, Goff S, Baltimore D. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell.* **18**:93–100.

Goldschmidt V and Marquet R. 2004. Primer unblocking by HIV-1 reverse transcriptase and resistance to nucleoside RT inhibitors (NRTIs). *Int J. Biochem Cell Biol.* **36**:1687-705.

Gopalakrishnan V, Peliska JA, Benkovic SJ. 1992. Human immunodeficiency virus type 1 reverse transcriptase:spatial and temporal relationship between the polymerase and RNase H activities. *Proc Natl Acad Sci USA.* **89**:10763–767.

Götte M, Arion D, Michael A, Parniak MA, Wainberg MA. 2000 .The M184V Mutation in the Reverse Transcriptase of Human Immunodeficiency Virus Type 1 Impairs Rescue of Chain-Terminated DNA Synthesis. *J. Virol.* **74**:3579-3585.

Gotte M, Fackler S, Hermann T, Perola E, Cellai L, Gross HJ, Le Grice SF, Heumann H. 1995. HIV-1 reverse transcriptase-associated RNase H cleaves RNA/RNA in arrested complexes: implications for the mechanism by which RNase H discriminates between RNA/RNA and RNA/DNA. *EMBO J.* **14**:833–841.

Grant RM, Kuritzkes DR, Johnson VA, Mellors JW, Sullivan JL, Swanstrom R, D'Aquila RT, Van Gorder M, Holodniy M, Lloyd RM Jr, Reid C Jr, Morgan GF, Winslow DL. 2003. Accuracy of the TRUGENE HIV-1 genotyping kit. *J. Clin Micro.* **41**:1586-593.

Guitérrez-Rivas, M. Ibáñez A, Martinez MA, Domingo E, Menéndez-Arias L. 1999. Mutational Analysis of Phe160 within the "Palm" Subdomain of Human Immunodeficiency Virus Type 1 Reverse Transcriptase. *J. Mol Bio.* **290.** 615-625.

Hirsch M, Gunthard H, Schapiro JM, Brun-Vezinet F, Clotet B, Hammer SM, Johnson VA, Kuritzkes DR, Mellors JW, Pillay D, Yeni PG, Jacobsen DM, Richman DD. 2008.

Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. *Clin. Infect Dis.* **47**:266-285.

Hiscott J, Kwon H, Genin P. 2001. Hostile takeovers: viral appropriation of the NF-kappaB pathway. *J. Clin Invest.* **107**:143–151.

Hostomsky Z, Hostomska Z, Hudson GO, Moomaw EW, Nides BR. 1991. Reconstitution in vitro of RNase H activity by using purified N-terminal and C-terminal domains of human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci USA.* **88**:1148–1152.

Huang H, Chopra R, Verdine GL, Harrison SC. 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**:1669–675.

Huber HE and Richardson CC. 1990. Processing of the primer for plus strand DNA synthesis by human immunodeficiency virus 1 reverse transcriptase. *J Biol Chem.* **265**:10565–573.

Jacobo-Molina A, Ding J, Nanni RG, Clark AD Jr, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P, Hizi A, Hughes SH, Arnold E. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc Natl Acad Sci USA* **90**:6320–324.

Jaeger, J, Restle T, Steitz TA. 1998. The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *The European Molecular Biology Organization Journal.* **17**:4535-542.

Joint United Nations Programme on HIV/AIDS and World Health Organization. 2010. Report on the global AIDS epidemic.

Julias JG, McWilliams MJ, Sarafianos SG, Alvord WG, Arnold E, Hughes SH. 2003. Mutation of amino acids in the connection domain of human immunodeficiency virus type 1 reverse transcriptase that contact the template-primer affects RNase H activity. *J. Virol.* **77**:8548–554.

Julias JG, McWilliams MJ, Sarafianos SG, Arnold E, Hughes SH. 2002. Mutations in the RNase H domain of HIV-1 reverse transcriptase affect the initiation of DNA synthesis and the specificity of RNase H cleavage in vivo. *Proc Natl Acad Sci USA* **99**:9515–520.

Katayanagi K, Okumura M, Morikawa K. 1993. Crystal structure of *Escherichia coli* RNase HI in complex with Mg^{2+} at 2.8 Å resolution: proof for a single Mg^{2+} -binding site. *Proteins* **17**:337–346.

Kati WM, Johnson KA, Jerva LF, Anderson KS. 1992. Mechanism and fidelity of HIV reverse transcriptase. *J. Biol Chem.* **267**:25988–997.

Kelleher CD and Champoux JJ. 2000. RNA degradation and primer selection by Moloney murine leukemia virus reverse transcriptase contribute to the accuracy of plus strand initiation. *J. Biol Chem.* **275**:13061–070.

Korber, B. 1998. Numbering positions in HIV relative to HXB2CG, p. 102–111. In B. Korber, C. L. Kuiken, B. Foley, B. Hahn, F. McCutchan, J. W. Mellors, and J. Sodroski (ed.), Human retroviruses and AIDS. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA. 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**:1783–790.

Kuby TK. 2007. Immunology. New York: W.H. Freeman and Company. Fig. 20.9a.

Kuiken C, Hraber P, Thurmond J, Yusim K. 2008. The hepatitis C sequence database in Los Alamos. *Nucleic Acids Res.* **36**: 512-516.

Kvaratskhelia M, Budihas SR, Le Grice SF. 2002. Pre-existing distortions in nucleic acid structure aid polypurine tract selection by HIV-1 reverse transcriptase. *J. Biol Chem.* **277**:16689–696.

Lanciault C and Champoux JJ. 2006. Pausing during reverse transcription increases the rate of retroviral recombination. *J. Virol.* **80**:2483–494.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinfo.* **23**:2947-948.

Le Rouzic E and Benichou S. 2005. The Vpr protein from HIV-1: distinct roles along the viral life cycle. *Retrovirology*. **2**:11.

Lichterfeld M, Yu XG, Cohen D, Addo MM, Malenfant J, Perkins B, Pae E, Johnston MN, Strick D, Allen TM, Rosenberg ES, Korber B, Walker BD, Altfeld M. 2004. HIV-1 Nef is preferentially recognized by CD8 T cells in primary HIV-1 infection despite a relatively high degree of genetic diversity. *AIDS*. **18**: 1383-392.

Lim D, Gregorio GG, Bingman C, Martinez-Hackert E, Hendrickson WA, Goff SP. 2006. Crystal structure of the moloney murine leukemia virus RNase H domain. *J. Virol*. **80**:8379–389.

Lim D, Orlova M, Goff SP. 2002. Mutations of the RNase H C helix of the Moloney murine leukemia virus reverse transcriptase reveal defects in polypurine tract recognition. *J. Virol*. **76**:8360–373.

Llina T and Parniak M. 2008. Inhibitors of HIV-1 reverse transcriptase. *Adv Pharmacol*. **56**:121-167.

Lorain V. 2005. *Antibiotics in Laboratory Medicine: Making a Difference*. Edition: **5**:566 – 568.

Marconi VC, Sunpath H, Lu Z, Gordon M, Koranteng-Apegyei K, Hampton J, Carpenter S, Giddy J, Ross D, Holst H, Losina E, Walker BD, Kuritzkes DR. 2008. Prevalence of HIV-1 Drug Resistance after Failure of a First Highly Active Antiretroviral Therapy Regimen in KwaZulu Natal, South Africa. *HIV/AIDS* **46**: 1589-597.

McGovern SL, Caselli E, Grigorieff N, Shoichet BK. 2002. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med Chem.* **45**:1712–722.

McWilliams MJ, Julias JG, Sarafianos SG, Alvord WG, Arnold E, Hughes SH. 2003. Mutations in the 5' end of the human immunodeficiency virus type 1 polypurine tract affect RNase H cleavage specificity and virus titer. *J Virol.* **77**:11150–157.

McWilliams MJ, Julias JG, Sarafianos SG, Alvord WG, Arnold E, Hughes SH. 2006. Combining mutations in HIV-1 reverse transcriptase with mutations in the HIV-1 polypurine tract affects RNase H cleavages involved in PPT utilization. *Virology* **348**:378–388.

Menendez-Arias L. 2008. Mechanisms of resistance to nucleoside analogue inhibitors of HIV-1 reverse transcriptase. *Virus Res.* **134**:124-146.

Miles LR, Agresta BE, Khan MB, Tang S, Levin JG, Powell MD. 2005. Effect of polypurine tract (PPT) mutations on human immunodeficiency virus type 1 replication: a virus with a completely randomized PPT retains low infectivity. *J. Virol.* **79**:6859–867.

Monini P, Sgadari C, Toschi E, Barillari G, Ensoli B. 2004. Antitumor effects of antiretroviral therapy. *Nat Rev Cancer.* **4**:861-75.

Mracna M, Becker-Pergola G, Dileanis J, Guay LA, Cunningham S, Jackson JB, Eshleman SH. 2001. Performance of Applied Biosystems ViroSeq HIV-1 Genotyping System

for Sequenced-Based Analysis of Non-subtype B Human Immunodeficiency Virus Type 1 from Uganda. *J. Clin. Microbiol.* **30**:4323-4327.

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

Neil SJ, Zang T, Bieniasz PD. 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature.* **451**: 425-430.

Nikolenko GN, Delviks-Frankenberry KA, Palmer S, Maldarelli F, Fivash MJ Jr, Coffin JM, Pathak VK. 2007. Mutations in the connection domain of HIV-1 reverse transcriptase increase 3'-azido-3'- deoxythymidine resistance. *Proc Natl Acad Sci USA* **104**:317–322.

Nikolenko GN, Palmer S, Maldarelli F, Mellors JW, Coffin JM, Pathak VK. 2005. Mechanism for nucleoside analog-mediated abrogation of HIV-1 replication: balance between RNase H activity and nucleotide excision. *Proc Natl Acad Sci USA* **102**:2093–098.

Osmanov S, Pattou C, Walker N, Schwardlander B, Esparza J, WHO-UNAIDS Network for HIV Isolation and Characterization. 2002. Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *Acquir. Immune. Defic. Syndr.* **29**:184–190.

Palaniappan C, Fuentes GM, Rodriguez-Rodriguez L, Fay PJ, Bambara RA. 1996. Helix structure and ends of RNA/DNA hybrids direct the cleavage specificity of HIV-1 reverse transcriptase RNase H. *J. Biol Chem.* **271**:2063–070.

Pavlov AR, Pavlova NV, Kozyavkin SA, Slesarev AI. 2004. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends Biotechnol.* **22**: 253–260.

Paulson BA, Zhang M, Schultz SJ, Champoux JJ. 2007. Substitution of alanine for tyrosine-64 in the fingers subdomain of M-MuLV reverse transcriptase impairs strand displacement synthesis and blocks viral replication in vivo. *Virology* **366**: 361- 376.

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

Peeters M. 2000.Recombinant HIV sequences: Their role in the global epidemic. pp. I-39-54 in *HIV Sequence Compendium 2000.* Edited by: Kuiken CL, Foley B, Hahn B, Korber B, McCutchan F, Marx PA, Mellors JW, Mullins JI, Sodroski J, and Wolinsky S. Published by: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Perrin L, Kaiser L, Yerly S. 2003. Travel and the spread of HIV-1 genetic variants. *Lancet Infect Dis.* **3**:22–27.

Peliska JA and Benkovic SJ. 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science*. **258**:1112–1118.

Pollard, VW and Malim, M. H. 1998. The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **52**:491–532.

Pond SJ, Ridgeway WK, Robertson R, Wang J, Millar DP. 2009. HIV-1 Rev protein assembles on viral RNA one molecule at a time. *Proc Natl Acad Sci USA.* **106**: 1404-408.

Potter SJ, Chew CB, Steain M, Dwyer DE, Saksena NK. 2004. Obstacles to successful antiretroviral treatment of HIV-1 infection: problems & perspectives. *Indian J. Med Res.* **119**:217-237.

Powell MD and Levin JG. 1996. Sequence and structural determinants required for priming of plus-strand DNA synthesis by the human immunodeficiency virus type 1 polypurine tract. *J. Virol.* **70**:5288–296.

Pullen KA and Champoux JJ. 1990. Plus-strand origin for human immunodeficiency virus type 1: implications for integration. *J. Virol.* **64**:6274–277.

Pullen KA, Ishimoto LK, Champoux JJ. 1992. Incomplete removal of the RNA primer for minus-strand DNA synthesis by human immunodeficiency virus type 1 reverse transcriptase. *J. Virol.* **66**:367–373.

Pullen KA, Rattray AJ, Champoux JJ. 1993. The sequence features important for plus strand priming by human immunodeficiency virus type 1 reverse transcriptase. *J. Biol Chem.* **268**:6221–227.

Rausch JW and Le Grice SF. 2004. 'Binding, bending and bonding': polypurine tract-primed initiation of plusstrand DNA synthesis in human immunodeficiency virus. *Int J. Biochem Cell Biol.* **36**:1752–766.

Richman DD. 2001. HIV chemotherapy. *Nature.* **410**:995-1001.

Rodgers DW, Gamblin SJ, Harris BA, Ray S, Culp JS, Hellmig B, Woolf DJ, Debouck C, Harrison SC. 1995. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA* **92**:1222–226.

Rychlik W, Spencer WJ, Rhoads RE. 1990. Optimization of the annealing temperature for DNA amplification in vitro. *Nucl Acids Res* **18**: 6409–412.

Sambrook J and Russel DW. 2001. *Molecular Cloning: A Laboratory Manual* (3rd ed.). Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. Chapter 8: In vitro Amplification of DNA by the Polymerase Chain Reaction.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* **74**:5463-7.

Santos AF, Lengrubler RB, Soares EA, Jere A, Sprinz E, Martinez AM, Silveira J, Sion FS, Pathak VK, Soares MA. 2008. Conservation patterns of HIV-1 RT connection and

RNase H domains: identification of new mutations in NRTI-treated patients. PLoS ONE **3**:1781.

Sarafianos SG, Das K, Clark AD Jr, Ding J, Boyer PL, Hughes SH, Arnold E. 1999. Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. Proc Natl Acad Sci USA **96**:10027-032.

Sarafianos SG, Das K, Tantillo C, Clark AD Jr, Ding J, Whitcomb JM, Boyer PL, Hughes SH, Arnold E. 2001. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. Embo J. **20**:1449–461.

Schatz O, Mous J, Le Grice SF. 1990. HIV-1 RT-associated ribonuclease H displays both endonuclease and 3'- 5' exonuclease activity. EMBO J. **9**:1171–176.

Scherrer AU, Hasse B, von Wyl V, Yerly S, Böni J, Bürgisser P, Klimkait T, Bucher HC, Ledegerber B, Günthard HF. 2009. Prevalence of etravirine mutations and impact on response to treatment in routine clinical care: Swiss HIV Cohort Study (SHCS). HIV Med. **10**:647-656.

Schultz SJ and Champoux JJ. 1996. RNase H domain of Moloney murine leukemia virus reverse transcriptase retains activity but requires the polymerase domain for specificity. J. Virol. **70**:8630–638.

Schultz SJ, Whiting SH, Champoux JJ. 1995. Cleavage specificities of Moloney murine leukemia virus RNase H implicated in the second strand transfer during reverse transcription. *J. Biol. Chem.* **270**:24135–145.

Schultz SJ, Zhang M, Champoux JJ. 2003. Specific cleavages by RNase H facilitate initiation of plus-strand RNA synthesis by Moloney murine leukemia virus. *J. Virol.* **77**:5275–285.

Schultz SJ and Champoux JJ. 2008. RNase H Activity: Structure, Specificity, and Function in Reverse Transcription. *Virus Res.* **134**:86–103.

Schultz SJ, Zhang M, Champoux JJ. 2004. Recognition of internal cleavage sites by retroviral RNases H. *J. Mol Biol.* **344**:635–652.

Schultz SJ, Zhang M, Champoux JJ. 2006. Sequence, distance, and accessibility are determinants of 5'-end directed cleavages by retroviral RNases H. *J. Biol Chem.* **281**:1943–955.

Schultz SJ, Zhang M, Kelleher CD, Champoux JJ. 2000. Analysis of plus-strand primer selection, removal, and reutilization by retroviral reverse transcriptases. *J. Biol Chem.* **275**:32299–309.

Shafer RW. 2002. Genotypic testing for human immunodeficiency virus type-1 drug resistance. *Clin. Microbiol. Rev.* **15**:247–277.

Sluis-Cremer N and Tachedjian G. 2008. Mechanisms of inhibition of HIV replication by non-nucleoside reverse transcriptase inhibitors. *Virus Res.* **134**:147-156.

Smith CM, Leon O, Smith JS, Roth MJ. 1998. Sequence requirements for removal of tRNA by an isolated human immunodeficiency virus type 1 RNase H domain. *J. Virol.* **72**:6805–812.

Smith CM, Potts WB 3rd, Smith JS, Roth MJ. 1997. RNase H cleavage of tRNA^{Pro} mediated by M-MuLV and HIV-1 reverse transcriptases. *Virology* **229**:437–446.

Smith CM, Smith JS, Roth MJ. 1999. RNase H requirements for the second strand transfer reaction of human immunodeficiency virus type 1 reverse transcription. *J. Virol.* **73**:6573–581.

Smith JS, Gritsman K, Roth MJ. 1994. Contributions of DNA polymerase subdomains to the RNase H activity of human immunodeficiency virus type 1 reverse transcriptase. *J. Virol.* **68**:5721–729.

Smith JS and Roth MJ. 1992. Specificity of human immunodeficiency virus-1 reverse transcriptase-associated ribonuclease H in removal of the minus-strand primer, tRNA(Lys₃). *J. Biol Chem.* **267**:15071–079.

Smith JS and Roth MJ. 1993. Purification and characterization of an active human immunodeficiency virus type 1 RNase H domain. *J Virol.* **67**:4037–049.

Smith L. 2009. Woman found carrying new strain of HIV from gorillas. London: The Independent. <http://www.independent.co.uk/lifestyle/health-and-families/health-news/woman-found-carrying-new-strain-of-hiv-from-gorillas-1766627.html>.

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321**: 674–679.

Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood LE. 1985. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Res.* **13**:2399–412.

Sorge J and Hughes SH. 1982. Polypurine tract adjacent to the U3 region of the Rous sarcoma virus genome provides a cis-acting function. *J. Virol.* **43**:482–488.

Stumptner-Cuvelette P, Morchoisne S, Dugast M, Le Gall S, Raposo G, Schwartz O, Benaroch P. 2001. HIV-1 Nef impairs MHC class II antigen presentation and surface expression. *Proc Natl Acad Sci USA.* **98**:12144-149.

Telesnitsky A and Goff SP. 1993. Strong-stop Strand Transfer during Reverse Transcription. In: Skalka, AM.; Goff, SP., editors. *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press; Plainview, New York: p. 49-83.

Thomson MM, Perez-Alvarez L, Najera R. 2002. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect. Dis.* **2**:461–471.

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18**:1051–063.

Vingerhoets J, Tambuyzer L, Paredes R, Nijs S, Clotet B, Schapiro JM, Picchio G. 2009. Effect of mutations in the RT connection domain on phenotypic susceptibility and virologic response to etravirine. *Antivir Ther.* **14**:A35.

Volkman S, Wohrl BM, Tisdale M, Moelling K. 1993. Enzymatic analysis of two HIV-1 reverse transcriptase mutants with mutations in carboxyl-terminal amino acids residues conserved among retroviral ribonucleases H. *J Biol Chem* **75**:71–78.

Weiss RA. 1993. How does HIV cause AIDS? *Scie.* **260**:1273–279.

Whitcomb JM, Parkin NT, Chappey C, Hellmann NS, Petropoulos CJ. 2003. Broad nucleoside reverse-transcriptase inhibitor cross-resistance in human immunodeficiency virus type 1 clinical isolates. *J. Infect. Dis.* **188**:992–1000.

Wohrl BM, Krebs R, Thrall SH, Le Grice SF, Scheidig AJ, Goody RS. 1997. Kinetic analysis of four HIV-1 reverse transcriptase enzymes mutated in the primer grip region of p66. Implications for DNA synthesis and dimerization. *J. Biol Chem.* **272**:17581–587.

Wyatt R and Sodroski J. 1998. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **280**:1884–888.

Zheng YH, Lovsin N, Peterlin BM. 2005. Newly identified host factors modulate HIV replication. *Immunol Lett.* **97**:225-234.

Chapter 7. Appendices

Appendix I: Preparation of Reagents

1. 1% Agarose Electrophoresis gel

0.4g of agarose was added to 40ml of 1X TBE to which 2 μ l of ethidium bromide was added. Five microliters of each PCR product was loaded in the gel along with 1 μ l of loading dye.

2. 1X TBE Buffer

One liter was prepared by adding 10ml of 10X TBE buffer to 90ml of distilled water. This was stored at room temperature.

3. 10mM dNTPs calculations

This was prepared by adding 20 μ l of each deoxyribonucleotide triphosphate (A, T, C, and G) from dNTPs stock (100mM) to 120 μ l of distilled water. This was stored at -20°C.

4. 10pmol Primer calculations

This was prepared by adding 10 μ l of a primer (100pmol) to 90 μ l of distilled water. This was stored at -20°C.

5. 1.6pmol Primer calculations

This was prepared by adding 1 μ l of 10pmol primer plus 5.25 μ l of distilled water. This was stored at -20°C.

6. 50ng Random Hexamers

This was prepared by adding 5 μ l of random Hexamers (3 μ g/ μ l) to 295 μ l of DEPC water. This was stored at -20°C.

Appendix II: Materials and equipment

1. Materials to be supplied by user

- ✓ Disposable gloves
- ✓ Sterile pipette tips with aerosol barrier
- ✓ 1.5 ml RNase/DNase free microcentrifuge tubes
- ✓ Serological pipettes
- ✓ Isopropanol
- ✓ Ethanol
- ✓ 0.2ml thin-walled PCR tubes
- ✓ MicroAmp[®] 96 Well Reaction Plate
- ✓ 10% Bleach
- ✓ Ice

2. Equipment to be supplied by user

- ✓ Microcentrifuge
- ✓ Vortex Mixer
- ✓ micropipettes
- ✓ Pipette Aid
- ✓ Ice-bucket
- ✓ Pipette Aid
- ✓ -20°C Freezer
- ✓ GeneAmp9700 Thermal Cycler
- ✓ Timer
- ✓ Genetic Analyzer 3130xl (Applied Biosystems Inc, Foster City, CA)

Appendix III: Sequence Identities

Sequence ID subtype C naïves			
704mc10n	AY901976	AY585267	DQ011170
704mc1n	AY901977	AY585268	DQ011171
704mc2n	AY901978	AY703908	DQ011172
704mc3n	AY901979	AY703909	DQ011173
704mc4n	AY901980	AY703910	AY463217
704mc6n	AY901981	AY703911	AY463218
704mc8n	DQ011165	AY772690	AY463219
704mc9n	DQ011166	AY772700	AY463220
pke10n	DQ011167	DQ396364	AY463230
pke11n	DQ011169	DQ396365	AY463231
pke13n	DQ011170	DQ396366	AY463232
pke15n	DQ011171	DQ396367	
pke16n	DQ011172	DQ396368	
pke18n	DQ011173	DQ396369	
pke19n	DQ011175	DQ396370	
pke1n	AY463217	DQ396371	
pke21n	AY463218	DQ396372	
pke22n	AY463219	DQ396373	
pke23n	AY463220	DQ396374	
pke24n	AY463221	DQ396375	
pke2n	AY463222	DQ396376	
pke3n	AY463223	DQ396377	
pke4n	AY463224	DQ396378	
pke5n	AY463225	DQ396379	
pke6n	AY463226	DQ396380	
pke8n	AY463227	DQ396381	
pke9n	AY463228	DQ396382	
pke12n	AY463229	DQ396383	
pke17n	AY463230	DQ396384	
AY901967	AY463231	DQ396385	
AY901968	AY463232	DQ396386	
AY901969	AY463233	DQ396387	
AY901970	AY463234	DQ396388	
AY901971	AY463236	DQ396389	
AY901972	AY463237	DQ396390	
AY901973	AY585264	DQ011165	
AY901974	AY585265	DQ011166	
AY901975	AY585266	DQ011167	

Sequence ID subtype C NRTI-treated			
704mc14f	pke4f	sw51	pcsm10
704mc16f	pke5f	sw52	sw50
704mc1f	pke6f	sw53	pke31f
704mc20f	pke7f	sw54	pke33f
704mc21f	pke9f	sw55	sw48
704mc23f	sc1	sw57	sw49
704mc25f	sc3	sw58	pcsm8
ich1	sc8	sw59	pcsm9
ich10	sp1	sw6	
ich11	sp11	sw61	
ich12	sp13	sw63	
ich13	sp14	sw66	
ich15	sp15	sw67	
ich16	sp24	sw68	
ich17	sp25	sw69	
ich18	sp34	sw74	
ich19	sp37	sw81	
ich2	sp5	sw82	
ich20	sp7	sw83	
ich21	sw10	sw84	
ich22	sw100	sw88	
ich23	sw102	sw89	
ich25	sw14	sw91	
ich5	sw16	sw92	
ich8	sw18	sw94	
ich9	sw21	sw97	
iwn1	sw22	sw99	
iwn2	sw25	pcsk9	
pke10f	sw26	pcsk10	
pke11f	sw27	pcsk11	
pke12f	sw29	pcsk12	
pke13f	sw30	pcsk13	
pke17f	sw31	pcsk14	
pke18f	sw32	pcsk2	
pke19f	sw34	pcsk3	
pke21f	sw35	pcsk6	
pke22f	sw36	pcsk7	
pke23f	sw37	pcsk13	
pke24f	sw38	pcsm4	
pke28f	sw40	pcsm5	
pke29f	sw45	pcsm5	
pke2f	sw47	pcsm7	