Multidisciplinary viral analyses in People Living with HIV-1C and receiving second-line combination antiretroviral therapy (cART) in South Africa

by

Adetayo Emmanuel Adegbenga Obasa

Thesis presented in fulfilment of the requirements for the degree Doctor of Science in Medical Virology at the University of Stellenbosch

Supervisor: Dr. Graeme Brendon Jacobs (University of Stellenbosch)

Co-supervisor: Associate Prof. Ujjwal Neogi (Karolinska Institute, Sweden)

Co-supervisor: Prof. Kamalendra Singh (University of Missouri, United States)

Co-supervisor: Dr. Ruben Cloete (University of the Western Cape, South Africa)

Division of Medical Virology

Department of Pathology

Faculty of Medicine and Health Sciences

December 2019

Declaration

This dissertation includes three original publications in peer reviewed journals and three unpublished papers under review. The development and writing of the papers (published and unpublished) were the principal responsibility of myself and my co-author. A declaration is included in the dissertation indicating the nature and extent of the contributions of the co-authors.

December 2019

Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. Graeme Brendon Jacobs and Professor Ujjwal Neogi for their guidance and advice throughout the course of this work. I would like to acknowledge the immerse contribution of my co-supervisors Dr. Ruben Cloete and Professor Kamalendra Singhi for all their contributions. Your contribution to this work is evident in many pages. I would like to express gratitude to Miss. Zimvo Maqeda for her unwavering support, motivation and encouragement.

Many thanks to all members of the Division of Medical Virology, past and present, for your encouragement, great attitude and help during the course of my study. Many thanks to Miss Shaheida Isaacs for taking time to proof read my thesis. I would also like to appreciate Dr. Shambhu Prasad Ganeshappa Aralaguppe, Mr. Mikasi Sello Given, Miss Olivette Varathan, Miss Michaela Lucienne Venter, Mrs Janca Ferreira and Mr. Njenda Duncan for their encouragement and support throughout my studies, you guys are simply the best. I will also like to appreciate The Redeemed Christian Church of God Christian community for their prayers and moral support.

My sincere appreciation goes to the following funding agencies: National Research Foundation (NRF), Poliomyelitis Research Foundation (PRF), Harry Crossley Foundation, Stellenbosch University and the National Health Laboratory Service (NHLS) Research Trust. Without your financial support, my project would not have been a success.

Finally, I would also like to appreciate my parents for believing in me and supporting me at all through my years of studying. I am forever grateful for God's wisdom, protection and guidance!

"*I have set the Lord always before me:*
Because he is at my right hand,
I shall not be shaken." Psalm 16:8

Dedication

“My help comes from the LORD, the maker of heaven and earth” (Ps 121: 02). I dedicate the thesis to my parents, siblings, Zimvo Maqeda and to all health care worker, people working and living with HIV

List of Abbreviations

3'EP	**3'-end processing**
3TC	**lamivudine**
ABC	**abacavir**
AIDS	**Acquired Immunodeficiency Syndrome**
ARV	**Antiretroviral**
ATP	**Adenosine triphosphate**
ATV	**atazanavir**
AZT	**zidovudine**
bPI	**boosted Protease Inhibitor**
BTG	**bictegravir**
CAB	**cabotegravir**
CA	**Capsid**
cART	**Combination antiretroviral treatment**
CCD	**Catalytic core domain**
CRF	**Circulating Recombinant Form**
CTD	**C terminal domain**
CCR5	**Chemokine receptor 5**
CXCR4	**Chemokine receptor 4**
CryoEM	**Cryo electron microscopy**
d4T	**stavudine**
ddI	**didanosine**
DNA	**Deoxyribonucleic acid**
DRMs	**Drug Resistance Mutations**
DRV	**darunavir**

DTG	**dolutegravir**
EDTA	**Ethylenediamine tetraacetic acid**
EFV	**efavirenz**
ETR	**etravirine**
EVG	**elvitegravir**
FTC	**emtricitabine**
GP	**Glycoprotein**
GRT	**Genotypic Resistance Testing**
HIV	**Human Immunodeficiency Virus**
HIV-1B	**HIV-1 Subtype B**
HIV-1C	**HIV-1 Subtype C**
HIVDR	**HIV Drug Resistance**
HTLV	**Human T-lymphotropic virus**
HTS	**High-throughput sequencing**
IN	**Integrase**
InSTIs	**Integrase Strand Transfer Inhibitors**
LA	**Long acting**
LAV	**Lymphadenopathy Associated Virus**
LMICs	**Low-middle income countries**
LPV/r	**lopinavir/ritonavir**
LTR	**Long Terminal Repeats**
MA	**Matrix**
MRC	**Medical Research Council**
mRNA	**messenger Ribonucleic acid**
Nef	***Negative regulatory factor***

NGS	**Next-Generation Sequencing**
NHLS	**National Health Laboratory Service**
NNRTIs	**non-nucleoside reverse transcriptase inhibitors**
NRTIs	**nucleoside reverse transcriptase inhibitors**
NTD	**N – terminal domain**
NVP	**nevirapine**
PCR	**Polymerase chain reaction**
PHI	**Primary HIV Infection**
PIs	**Protease inhibitors**
PLHIV	**People Living with HIV**
PMs	**Polymorphisms**
PR	**Protease**
PrEP	**Pre-exposure prophylaxis**
RAL	**raltegravir**
RAMs	**Resistance associated mutations**
Rev	***Regulator gene***
RNA	**ribonucleic acid**
RT	**Reverse Transcriptase**
RTI	**Reverse Transcriptase Inhibitors**
RLS	**Resource Limited Settings**
SIV	**Simian Immunodeficiency Virus**
ST	**Strand Transfer**
START	**strategic timing of antiretroviral therapy**
STC	**Strand transfer complex**
TAM	**Thymidine analogue resistance mutation**

Tat	***Trans-Activator of Transcription***
TDF	**tenofovir**
TDR	**Transmitted Drug Resistance**
TV	**Tygerberg Virology**
UN	**United Nations**
URF	**Unique Recombinant Forms**
VF	**Virological Failure**
Vif	***Viral infectivity factor***
Vpr	***Viral Protein R***
Vpu	***Viral Protein U***
WHO	**World Health Organisation**

Scientific contributions

List of accepted scientific papers included in this thesis

- Analyses of HIV-1 integrase sequences prior to South African national HIV treatment program and available of integrase inhibitors in Cape Town, South Africa. Brado D, **Obasa AE**, Ikomey GM, Cloete R, Singh K, Engelbrecht S, Neogi U, Jacobs GB. Sci Rep. 2018 Mar 16;8(1):4709. doi: 10.1038/s41598018-22914-5 Sci Rep. 2018 Apr 16;8(1):6262. PMID: 29549274.
- Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. Rogers L, **Obasa AE**, Jacobs GB, Sarafianos SG, Sönnerborg A, Neogi U, Singh K. Front Microbiol. 2018 Aug 2;9:1754. doi: 10.3389/fmicb.2018.01754. eCollection 2018. PMID: 30116231.
- Mutations in Long Terminal Repeats κB Transcription Factor Binding Sites in Plasma Virus Among South African People Living with HIV-1**. Obasa AE**, Ashokkumar M, Neogi U, Jacobs G. AIDS Res Hum Retroviruses. 2019 Feb 22. doi: 10.1089/AID.2018.0293. PMID: 30793917.

Under review

- Analyses of HIV-1 Protease, Reverse Transcriptase and Integrase sequences of patients failing antiretroviral therapy and first evidence of RAL resistance in South African patients. **Obasa AE**, Sello Given Mikasi Kamalendra Singh, Shambu Prasad, Ujjwal Neogi, Graeme Brendon Jacobs. Reference number: Nature – Scientific Report SREP-19-20242. .
- Increased protease inhibitor drug resistance mutations in minor viral quasispecies on HIV-1 infected patients suspected of failing on national second line therapy in South Africa. **Obasa AE**, Anoop T. Amibikan Gupta Soham, Ujjwal Neogi, Graeme Brendon Jacobs. Reference number: Journal of Antimicrobial Chemotherapy JAC-2019-1065.
- Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding. Rumbidzai Chitongo, **Adetayo Emmanuel Obasa**, Sello Given Mikasi, Graeme Brendon Jacobs, Ruben Cloete. Manuscript is under review PLOSOne. Pre-print available: http://biorxiv.org/cgi/content/short/781120v1. **doi:** https://doi.org/10.1101/781120

List of scientific papers not included in this thesis

This paper was part of my Masters (MSc) dissertation and falls outside the scope of the PhD thesis objectives.

- Near full-length HIV-1 subtype B sequences from the early South African epidemic, detecting a BD unique recombinant form (URF) from a sample in 1985. **Obasa AE**, Engelbrecht S, Jacobs GB.Sci Rep. 2019 Apr 17;9(1):6227. doi: 10.1038/s41598-019-42417-1. PMID: 30996293.

List of conference presentations

Poster presentations

- 22nd International AIDS Conference (AIDS 2018) Amsterdam, Netherlands entitled "Evidence of high-level Raltegravir (RAL) resistance in patients from South Africa failing second-line antiretroviral therapy (ART)." **Adetayo Emmanuel Obasa**, Ruben Cloete, Mathilda Classen, Kamal Singi, Ujjwal Neogi, Graeme Brendon Jacobs.
- 62nd annual academic day Stellenbosch University Faculty of Medicine and Health Sciences (Tygerberg campus) entitled "Genotypic analyses of HIV-1 Integrase from South African patients failing first and second line antiretroviral treatment." **Adetayo Emmanuel Obasa**, Kamalendra Singh, Mathilda Claassen, Ruben Cloete, Ujjwal Neogi, Graeme Brendon Jacobs
- 26th International HIV Dynamics and Evolution Cascais, Portugal entitled "Near full-length HIV-1 subtype B sequences from the early South African epidemic, detecting a BD unique recombinant form (URF) from a sample in 1985." **Adetayo Emmanuel Obasa,** Susan Engelbrecht, Graeme Brendon Jacobs

Oral presentations

- 63rd annual academic day Stellenbosch University Faculty of Medicine and Health Sciences (Tygerberg campus) entitled "Increased protease inhibitor drug resistance mutations in minor HIV-1 quasispecies from infected patients suspected of failing on national second line therapy in South Africa." **Adetayo Emmanuel Obasa**, Anoop T Ambikan, Soham Gupta, Graeme Brendon Jacobs, Ujjwal Neogi.

- 63rd annual academic day Stellenbosch University Faculty of Medicine and Health Sciences (Tygerberg campus) entitled “Structural impact of selected Raltegravir resistance variants on Dolutegravir binding to South African HIV-1 Integrase subtype C protein.” Rumbidzai Chitongo, **Adetayo Emmanuel Obasa**,, Graeme Brendon Jacobs, Ruben Cloete.

Research visit

- Karolinska Institutet, Division of Clinical Microbiology, Department of Laboratory Medicine, Stockholm, Sweden. 15 September to 15 December 2018. The aim of the visit was to perform high-throughput sequence analyses on PLHIV-1 subtype C from South Africa.

Summary

The use of combination Antiretroviral Therapy (cART) has grown since its first introduction into the South African public sector. cART has significantly reduced the mortality rate caused by human immunodeficiency virus (HIV) in both high- and low-to-middle-income countries. The development of drug resistance has challenged the outcome of cART. This has led to the introduction of Integrase (IN) strand transfer inhibitors (InSTIs) as part of the first-line cART regimen. Due to their superior efficacy and high genetic barrier, this class of drugs was previously reserved as salvage therapy. The World Health Organization (WHO) supports InSTIs as first-line regimen non-nucleoside reverse transcriptase inhibitors (NNRTIs) particularly in regions where pre-treatment drug resistance to NNRTIs reaches 10%. Therefore, this study aimed to (i) to investigate the prevalence of InSTI mutations in treatment-naïve and treatment-experienced PLHIV using genotypic assays, which included Sanger sequencing, next-generation sequencing (NGS) and molecular modelling; (ii) analysed Long Terminal Repeats (LTR) to identify transcription factor binding sites.

Chapter 2: Ninety-one (n = 91) treatment-naïve patients were obtained before the start of antiretroviral treatment in South Africa. Furthermore, we included 314 South African patient sequences obtained from the Los Alamos National Library database (www.lanl.gov). The IN gene ~ 900 base pairs [bps] was amplified and sequenced using conventional DNA Sanger sequencing. Homology structure was generated using the cryoEM structure of HIV-1B IN intasome (PDB file 5U1C) using 'Prime' of Schrodinger Suit. Chapter 3: Ninety-six (n = 96) treatment-experienced patients receiving boosted protease inhibitors (bPIs) as part of their cART treatment regimen were obtained for further analyses. We performed conventional DNA Sanger sequencing to analyse the complete pol gene (~3011bps) and sequences were analysed using the Stanford HIV drug resistance database to assess genotypic resistance associated mutations (RAMs). Chapter 4: Fifty-six (n = 56) treatment-experienced patients receiving boosted protease inhibitors (bPIs) as part of their cART treatment regimen were obtained. We performed a high-throughput (HT) sequence analyses on the complete pol gene using Illumina HiSeq2500, followed by bioinformatics analysis to quantify the RAMs according to the Stanford HIV drug resistance database. Chapter 5 and 6: We performed in-silico analyses on diverse HIV-1 subtypes based on 8114 sequences. These included treatment naïve and downloaded sequences from the HIV Los Alamos National Library Database (www.lanl.gov). Homology derived molecular models of HIV-1 IN tetramers from different subtypes were

generated using cryoEM structure of the HIV-1B IN intasome. Chapter 7: Fifty-six (n = 56) treatment-experienced patients receiving boosted protease inhibitors (bPIs) as part of their cART treatment regimen were obtained. We performed Sanger sequencing to analyse the LTR gene (~ 474 bps) followed by bioinformatics analyses to identify transcription factor binding sites.

The data indicates that in South Africa, the prevalence of RAMs against InSTIs is low and InSTIs can be used as a potential viable salvage therapy option and/or first-line regimen. Molecular modelling was done for IN structural analyses, which revealed how naturally occurring polymorphisms might affect structural stabilities, viral DNA binding and drug-binding propensity. This study represents a true baseline InSTI resistance rate as the treatment-naïve patients were obtained before the cART introduction. We propose GRT for people living with HIV (PLHIV) before treatment initiation and we recommend continued InSTIs drug resistance monitoring when introduced on a larger scale in South African.

Opsomming

Kombinasie- antiretrovirale terapie (cART) het heelwat toegeneem sedert dit die eerste keer in die Suid-Afrikaanse openbare sektor bekend gestel is. cART het ’n aansienlike verlaging in die sterftesyfer in sowel hoë- as lae-/middelinkomstelande teweeggebring. Die ontwikkeling van middelweerstandigheid en oorgeërfde middelweerstandige mutasies hou egter ’n uitdaging vir die uitkoms van cART in. Daarom is integrasestringoordraginhibitors (InSTI’s) by die eerstelinie-ART-regime ingesluit. Hierdie klas middels is voorheen uitsluitlik as reddingsterapie gebruik weens hulle uitmuntende doeltreffendheid en hoë genetiese skans. Die Wêreldgesondheidsorganisasie ondersteun InSTI’s as eerstelinieregime, veral in streke waar voorbehandelingsweerstand teen nie-nukleosied-omgekeerdetranskriptase-inhibitors 10% bereik.

Vir hierdie studie is ’n deursneeondersoek uitgevoer wat bestaan het uit die genotipiese en molekulêre modellering van InSTI’s vir die volgende:

Behandelingsnaïewe pasiënte voor bekendstelling van die Suid-Afrikaanse nasionale menslike-immuniteitsgebreksvirus-(MIV-)behandelingsprogram en die beskikbaarheid van InSTI’s in Suid-Afrika.

Behandelingservare pasiënte wat tweedelinie-ART in die Suid-Afrikaanse nasionale behandelingsprogram ontvang.

Hoëdeursetreeksvorming (HTS) by behandelingservare pasiënte wat tweedelinie-ART in die Suid-Afrikaanse nasionale behandelingsprogram ontvang.

Die strukturele implikasie van genotipiese variasies in MIV-1-integrase (IN) by sowel behandelingservare as behandelingsnaïewe pasiënte.

Vir behandelingsnaïewe pasiënte is genetiese ontledings van 91 plasmamonsters sowel as 314 pasiënte se databasisafkomstige reekse gedoen. Genotipiese weerstandigheidstoetsing (GRT) het op die Q148-roete afgekom, wat tweedegenerasie-dolutegravir uit die databasisafkomstige reekse kan beïnvloed. Nietemin is geen weerstandigheidsverwante mutasies (RAM’e) opgemerk in monsters wat voor die bekendstelling van die nasionale MIV-behandelingsprogram bekom is nie.

Vir behandelingservare pasiënte is die genotipiese Sanger-reeksvormingsessai gebruik om belangrike RAM’e soos T66I, Y143R en T97A te identifiseer. Hierdie RAM’e bied weerstand

teen raltegravir (RAL) en elvitegravir (EVG). HTS is gebruik om geringe (<20% van die populasie) en erge variante binne die MIV quasispecies (>20% van die populasie) by behandelingservare pasiënte te kwantifiseer. HTS het InSTI-RAM'e soos Y143R, S147G en E138R aan die lig gebring, wat weerstand bied teen RAL en EVG. Molekulêre modellering is vir IN- strukturele ontledings uitgevoer, wat getoon het hoe natuurlike polimorfismes strukturele stabiliteit, virale DNS-binding en middelbindingsgeneigdheid kan beïnvloed.

Die studiedata dui op 'n lae voorkoms van RAM'e teen InSTI's in Suid-Afrika. InSTI's kan as 'n potensieel lewensvatbare reddingsterapie en/of eerstelinieregime gebruik word. Die studie bied 'n ware basislynaanduiding van InSTI-weerstandigheid omdat die behandelingsnaïewe pasiënte voor die bekendstelling van cART gewerf is. Nogtans beveel die navorsers GRT aan vir MIV-geïnfekteerde persone voor die aanvang van behandeling, en word die gebruik van HTS-GRT as 'n standaard-GRT voorgestel.

Table of Contents

List of Figures

List of Tables

Chapter 1 Background and literature review

1.1. Background

Acquired immunodeficiency syndrome (AIDS) is one of the most common causes of morbidity and mortality in Africa. The human immunodeficiency virus (HIV), which is a retrovirus, is the causative agent of AIDS. There are two different types of HIV that infect humans: HIV-1 and HIV-2.[1,2] Phylogenetically, HIV-1 is classified into four main groups: groups M (major), O (outlier), N (non-M, non-O) and P (putative). Group M is responsible for the majority of HIV/AIDS infections, representing > 90% of all HIV infections and driving the global epidemic.[1,2] HIV-1 group M has been further classified into nine different subtypes (A-D, F, G, H, J and K), at least 98 circulating recombinant forms (CRFs) and numerous unique recombinant forms (URFs) (Los Alamos HIV database, www.lanl.gov accessed on 17 May 2019). HIV-2 is also phylogenetically categorised into groups A-H, but is restricted to Central and West Africa and is thought to be much less severe than HIV-1.[1,2]

More than 30 years since the first identification of HIV in 1981, due to the absence of a preventative vaccine or cure, the illness continues to affect millions of individuals. The United Nations AIDS (UNAIDS) organisation estimated that there were more than 43.9 million people living with HIV (PLHIV) worldwide at the end of 2018.[3] In 2017, 22 million PLHIV were residing in Eastern and Southern Africa.[3] Sub-Saharan Africa bears a significant global burden of PLHIV.[3]

The introduction of combination antiretroviral therapy (cART) has changed the AIDS epidemic to a chronic, but manageable disease.[4] Besides problems related to drug adherence, tuberculosis and HIV co-infections, cART has been constantly challenged by the development and spread of drug resistance. This poses a significant threat to the long-term management of PLHIV in public health settings where patients are often monitored using clinical or immunological parameters.[5]

South Africa has one of the largest populations of PLHIV, with 7.1 million infected, thus accounting for 19% of the global number of PLHIV. HIV prevalence is high amongst the general population at 18.9%, with approximately (61%) 4.5 millions of adults receiving cART.[3] In 2016, the South African government adopted the UNAIDS test and treat, and the 90-90-90 strategy. This strategy aims at having 90% of infected individuals diagnosed by 2020, providing cART to 90% of those diagnosed and achieving viral suppression for 90% of those on cART by 2020.[3] South Africa has made progress towards the first UNAIDS 90% target, but progress towards the second and the third 90% targets has been less impressive. Achieving the 90-90-90 target will require that at least 80% of PLHIV are on ART and virally supressed by the end of 2020.[6]

1.2. Research question and study aim

In South Africa, due to its high cost, genotypic resistance testing is not widely available for treatment-naïve patients and/or newly infected patients. The World Health Organization (WHO) supports transition to integrase strand transfer inhibitors (InSTIs)-based first-line treatment regimens in regions such as South Africa where pre-treatment drug resistance to non-nucleoside reverse transcriptase inhibitors (NNRTIs) reaches 10% [3]. With more people becoming eligible for cART, can pre-treatment or acquired mutations that confer resistance to nucleoside reverse transcriptase inhibitors (NRTIs) and NNRTIs affect the efficacy of an InSTIs-based regimen? To combat resistance challenges, a combination of two NRTIs and a third agent, either an InSTI, a boosted protease inhibitor (bPI) or an NNRTI, is needed. The cART approach, using three to four antiretrovirals, was designed to prevent the virus from developing escape mutations and to target HIV-1 at different replication stages. The main aim of my study was to investigate the prevalence of InSTI mutations in treatment-naïve and treatment-experienced PLHIV using genotypic assays, which included Sanger sequencing, next-generation sequencing (NGS) and molecular modelling. In addition, we also generated Long Terminal Repeat (LTR) HIV-1 sequences to identify transcription factor binding sites. NF-κB plays an important role in regulating the viral gene expression from the viral promoter and controlling viral latency.

1.3. Study objectives

Objective one: To identify the HIV-1 resistance mutation profiles of patients against available first- and second-generation InSTIs in treatment-naïve patients prior to the launching of the South African national HIV-treatment programme (Paper 1 – published).

Objective two: To identify the HIV-1 resistance mutation profiles of patients failing second-line cART regimens in South Africa (manuscript under review by Nature – Scientific Report. Reference number: SREP-19-20242).

Objective three: To perform NGS of the HIV-1 integrase (IN) gene fragment from South African patients failing second-line cART regimens, using the HiSeq 2500 sequencing system (Illumina) (manuscript under review by Journal of Antimicrobial Chemotherapy. Reference number: JAC-2019-1065).

Objective four: To explore the structural impact of IN polymorphisms from diverse HIV-1 subtypes (Paper 4 – published).

Objective five: To explore the consensus wild type subtype C IN sequence to build an accurate 3D model of HIV-1C IN. This is in order to understand the effect of three statistically enriched mutations on DTG drug binding (manuscript submitted to PLOS One). Pre-print available: http://biorxiv.org/cgi/content/short/781120v1. **doi:** https://doi.org/10.1101/781120

Objective six: To analyse LTR sequences from HIV participants in order to identify transcription factor binding sites. NF-κB in regulating the viral gene expression from the viral promoter and controlling viral latency (Paper 6 – published).

1.4. Origin and history of HIV infection

The lentivirus genus is a subset of the Retroviridae family of ribonucleic acid (RNA) viruses. AIDS was initially described as an immune deficiency state that was associated with multiple microbial infections and was commonly associated with homosexual patients in the early 1980s.[7,8] In 1983, the viral agent was isolated from a lymph node biopsy of a Kinshasa patient; as a result, the causative agent HIV was initially termed 'lymphadenopathy-associated virus' (LAV).[9] LAV was later termed 'human T-lymphotropic virus' (HTLV)-III as the viral structure was similar to that of HTLV-I.[10] In 1986, HTLV-III was determined to be the causative agent of AIDS, but due to immune depletion, it was later named as HIV type 1 (HIV-1).[10,11.] The first evidence of HIV infection was found in a serum sample stored from 1959 obtained from a male in Kinshasa. HIV was found to be closely related to a simian immunodeficiency virus (SIV) strain that originated from chimpanzees of the subspecies Pan Troglodytes and was classified as HIV-1 in 1998.[12] In 1960 a lymph node biopsy of a female stored in Leopoldville was recovered and analysed.[13] Comparison of HIV-1 sequences in the 1959 and 1960 samples showed a genetic variation of 12%. These findings suggested that HIV-1 had been present in humans before 1960[13] and possibly originated from multi-cross-species transmissions of SIV from primates to humans in Central and West Africa.[14]

1.5. Virion structure and genome organisation

The HIV-1 virion is surrounded by an external bilayer membrane and is cylindrically shaped, as shown in Figure 1.1. It is approximately 120 nm in diameter.[10,15] The Envelope protein is covered with the outer trimeric surface glycoprotein (gp120) and transmembrane glycoprotein (gp41), derived from the precursor protein (gp160). A Matrix (MA) protein (p17) appears on the inside of the lipid bilayer shell, which forms the inner surface. The Capsid (CA) protein (p24) layer forms the conical capsid.[16] The genome is a full-length diploid linear RNA.[17] HIV-1 enzymes encoded by the HIV genome are Protease (PR), Reverse transcriptase (RT) and IN. Figure 1.1 B displays the HIV-1 genome, which contains nine unique genes and encodes with various structural and non-structural viral proteins. These encoded viral proteins and enzymes play a crucial role in viral replication and establish infection in the host cell. HIV-1 also has regulatory genes called transactivator of transcription (*tat*) and regulatory of virion expression (*rev*). HIV-1 enzymes encoded by the HIV genome are Protease (PR), Reverse transcriptase (RT) and IN. Accessory genes include viral protein unique (*vpu*), viral protein R (*vpr*), viral infectivity factor (*vif*) and negative regulatory factor (*nef*).[16,17]

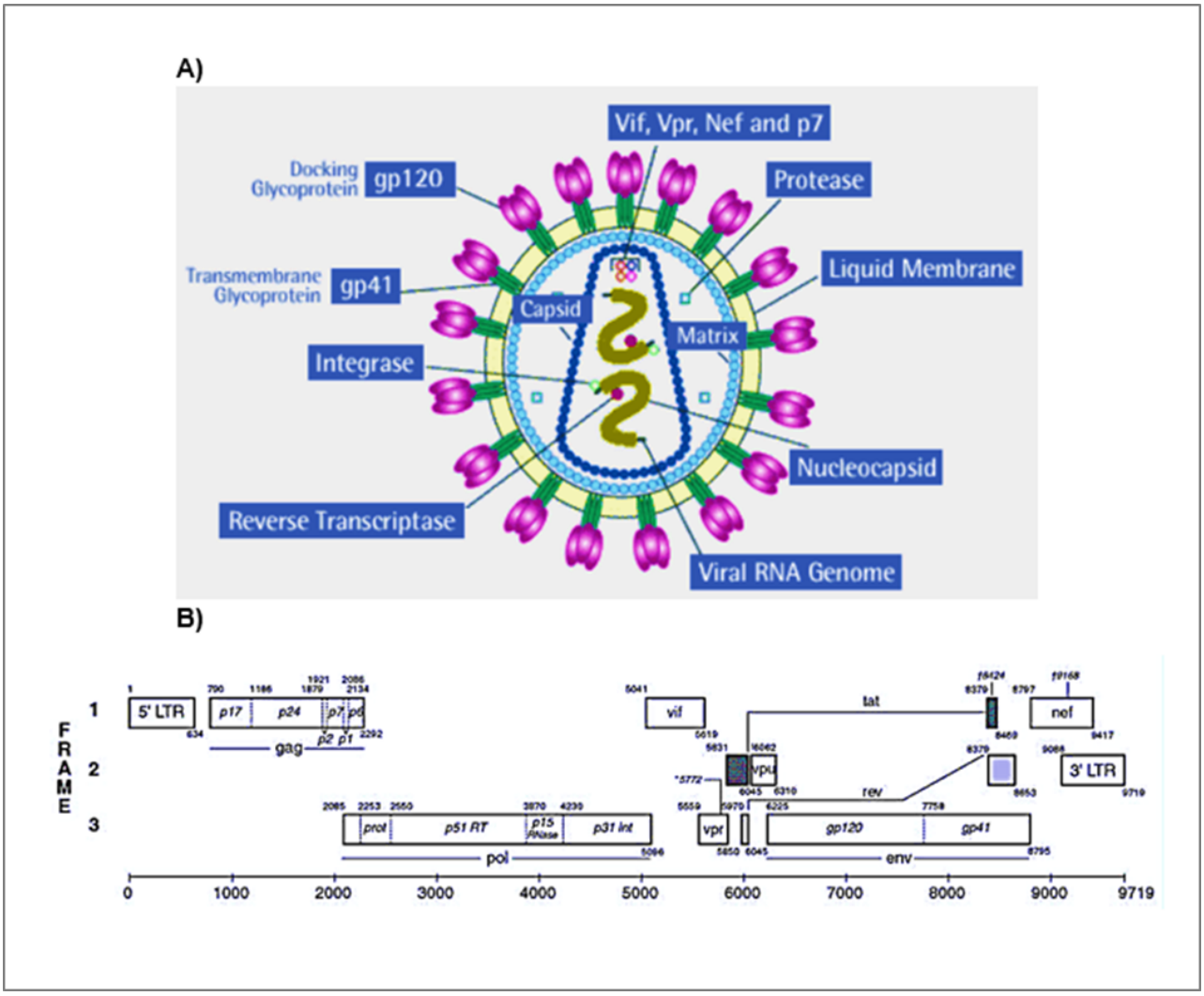
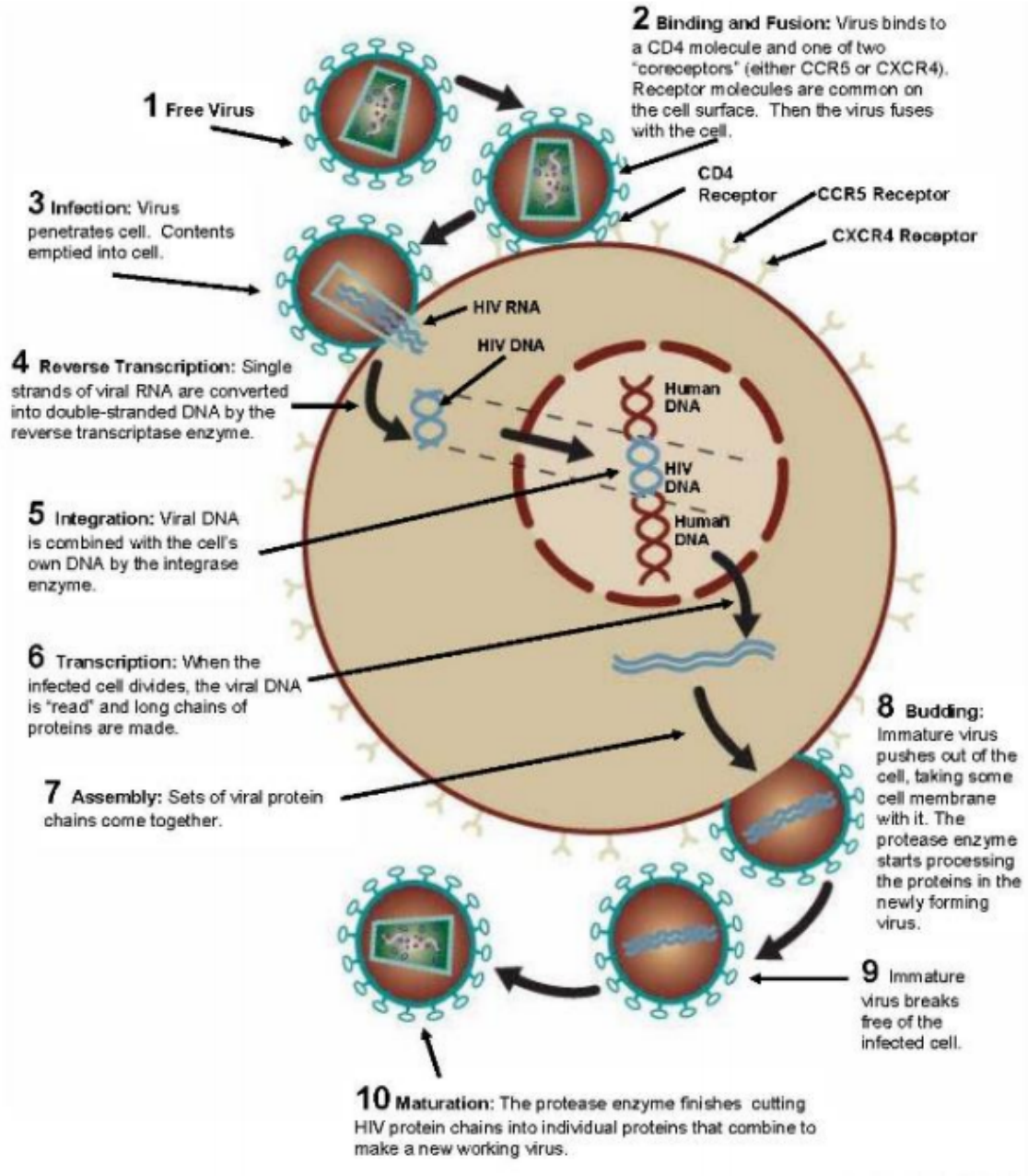


Figure 1.1 Schematic and Genetic Organisation of the HIV-1 (A and B). **A)** The diagram gives a transectional view of the HIV virion and displays the virus matrix encloses the capsid, which protects the two copies of genomic RNA, Protease, Reverse Transcriptase and Integrase. The Envelope consists of protruding glycoprotein gp120, which stems from the Fusion protein gp41**. B)** Genetic organisation of the HIV genome (approximately 9.7kb) showing the three main structural genes *gag, pol* and *env***.** Regulatory (*tat*, *rev*) and accessory genes (*vif*, *vpr*, *vpu*, *nef*) genes. All these genes are necessary for effective HIV replication. (Adopted from Los Alamos HIV database, www.lanl.gov accessed on 17 May 2019). Reprinted with permission.

1.6. The HIV life cycle and antiretroviral drugs

Viruses depend on living hosts to replicate. The *pol*-encoded viral enzymes that are essential for viral replication are PR, RT (this includes a polymerase and an RNAse H domain) and IN. Antiretroviral drugs target these viral enzymes and inhibit stages in the viral lifecycle to reduce the host's cellular function and replication steps. The Envelope protein of HIV-1, gp120, binds to CD4+ T-cell receptors on the host cells, followed by binding to co-receptors, the most important being chemokine receptor

5 (CCR5) and chemokine receptor 4 **(**CXCR4). After co-receptor binding, the gp41 Envelope protein undergoes a conformational change that enables the fusion of the viral membrane with the host cell membrane. At this stage, the viral capsid is released into the cytoplasm where uncoating begins. Thereafter, p24, which is the major Capsid protein, dissociates. The RT complex containing *pol*-encoded viral enzymes (RT, RNAse H and IN) remains intact to the two copies of viral RNA.[18] Reverse Transcription is the first step of the HIV replication cycle. RNA is transcribed to an RNA-DNA duplex, catalysed by the RT enzyme. The RNA is digested by RNAse H, followed by synthesis of double-stranded DNA through the DNA-dependent DNA polymerase activity of RT. The newly synthesised double-stranded DNA is transported through an active mechanism involving pre-integration complex (PIC) that contains IN and CA proteins. IN cleaves the host's chromosomal DNA, and the viral DNA is inserted into the host's chromosomal DNA, now referred to as 'proviral DNA'. At this stage, the virus can stay latent for years and establish a viral reservoir. RNA polymerase II catalyses the transcription of viral genes into viral messenger RNA (mRNA) or full genomic viral RNA, which is exported to the cytoplasm using host cell machinery. At this point, cellular ribosomes translate viral mRNAs into viral proteins. Viral proteins self-assemble during a process referred to as assembly. Upon maturation, the Gag protein is cleaved into the Matrix, Capsid and Nucleocapsid to form an infectious particle. In order for viral particles to mature to full infectivity, twelve (12) proteolytic reactions are needed.[19] In precursor proteins, Protease cleavage occurs at a particular site(s) and this process results in mature virions that are capable of infecting another cell, resulting in repeated cycles of viral replication. The main steps of HIV viral replication – binding and entry, reverse transcription, integration, viral assembly and budding form the basis for the targets of the six different cART drugs shown in Figure 1.2.



Source: http://www.aidsinfonet.org

Figure 1.2 The HIV lifecycle. The stages of viral lifecycle shown above. Thin, dark arrows show entry and integration. Curved and slightly bent arrows show early and late replication. 1- A piece of genetic material surrounded by protein. 2- Adsorption to the CD4 receptor (CXCR4 or CCR5). 3- Fusion and un-coating of a viral genomic dimer. 4- Reverse transcription. 5- Integration of pro-viral DNA into host genome using the integrase enzyme. 6- Transcription of viral DNA Nuclear import of PIC. 7- Assembly of viral chain 8- Viral budding, protease enzyme starts the formation of a new virus. 9- Breaking off the immature virus. 10- Viral maturation and formation of the new virus (Pomerantz and Horn, 2003; Freed, 2015).

1.7. Mechanisms of HIV-1 drug resistance

HIV drug resistance (HIVDR) emerges through different pathways, depending on the drug regimen. The goal of the treatment regimen is to preserve a functional immune system in PLHIV through suppression of HIV replication. In order to achieve this, cART is designed to target essential steps of the viral lifecycle. This section briefly explains the main mechanisms that can lead to resistance to each of the main drug classes.

1.7.1. Nucleoside Reverse Transcriptase Inhibitors (NRTIs)

HIVDR to NRTIs mainly develops by two different mechanisms: discrimination and excision. In the discrimination mechanism, specific mutations enable RT to discriminate between the natural deoxyribose nucleotide triphosphate (dNTPs) and NRTI-TPs used during DNA replication, thus preventing the incorporation of the NRTI-monophosphate into the growing DNA strand.[20] Mutations responsible for such resistance are M184V/I, K65R, 70E/G, L74V, Y115F and the rare Q151M complex.[20–22] The most common discrimination mutation is M184V/I, which alone confer resistance to lamivudine (3TC) and emtricitabine (ETR).[22,23] Furthermore, NRTI resistance mechanism, involves the incorporation of adenosine triphosphate (ATP)-mediated phosphorolysis which facilitates the removal of NRTIs from the 3'-end of the DNA. Thymidine analogue resistance associated mutations (TAMs) confer resistance through this mechanism.[21] TAMs are primarily selected for by zidovudine (AZT) and stavudine (d4T); these are structural analogues of the cellular nucleotide thymidine.[24] TAMs develop along two characteristic patterns: TAM 1 (41L, 210W and 215Y) and TAM 2 (67N, 70R and 219E/Q).[25–27] Mutations can also arise as a result of induced diversification of apolipoprotein B mRNA-editing enzyme and catalytic polypeptide 3G/F (APOBEC3G/F).[28,29] The viral protein *vif* plays a vital role to inhibit APOBEC3G/F activity and in some cases can lead to certain levels of nonlethal genetic diversity.[28,29] Such diversity can result in the development of mutations that confer resistance to 3TC.[30]

1.7.2. Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

The NNRTI resistance associated mutations (RAMs) are mostly located with the hydrophobic pocket in the RT. Resistance to NNRTIs develops relatively easy.[20] These mutations can cause either alteration to the size of the binding pocket or physical changes to the structure of the binding pocket and thus prevent drug binding. A single NNRTI mutation alone is enough to confer resistance to several of the NNRTIs. The K103N mutation is the most common among patients receiving efavirenz (EFV) (50% of patients) and nevirapine (NVP) (30% of patients). The L1001 mutation confers

resistance to EFV, rilpivirine (RPV) and NVP.[31] Mutations E138K and M184I confer high-level resistance to both etravirine (ETR) and RPV[32]

1.7.3. Protease inhibitors (PIs)

Resistance to Protease inhibitors (PIs) is complex, and for ritonavir-boosted PIs, it requires the development of at least two mutations (either one or more major and one or more complementary) in the PR gene.[20] It also involves mutations in the Gag gene.[27,33] The binding of PR to its polyprotein target can be affected by major mutations in the PR gene.[22] Drug resistant mutations may reduce the viral fitness, whilst compensatory changes can repair lost fitness. Minor PI mutations are often referred to as compensatory mutations and are very common.[34] Minor PI mutations can improve viral fitness, and in the absence of major PI mutations, they have little effect on treatment outcomes.[34] bPIs have a higher genetic barrier as compared to NRTIs and NNRTIs. This could be because the development of resistance requires more than one mutation. This has led to suggestions that virological failure in patients on bPIs mostly occurs in patients who do not adhere to their treatment regimen, rather than as a result of drug resistance.[20]

1.7.4. Entry inhibitors

Resistance to maraviroc, a CCR5 receptor antagonist, mostly occurs through a shift in co-receptor usage by the virus or through a shift in the viral envelope that enables binding to the CCR5 receptor, despite the presence of the inhibitor.[35] At the start of a treatment regimen, changes would most likely occur if there is a mix of R5 and X4 tropic viruses.[20] Existing data indicate that changes in the viral Envelope are mostly likely a common cause of maraviroc resistance.[35] Resistance to enfuviritide (a fusion inhibitor) can develop through mutations in a domain of 10 amino acids.[22,35] However, entry inhibitors are not part of standard therapy and these regimens are rarely used in South Africa.[35]

1.7.5. Integrase strand transfer Inhibitors (InSTIs)

Due to the important role of IN in the HIV-1 lifecycle, it has been targeted as part of cART. All InSTIs target the strand transfer reaction, which is catalysed by the viral IN. They are thus referred to as either Integrase Inhibitors (INIs) or InSTIs.[36,37] InSTIs have two mechanisms of action: they bind only to the specific complex between IN and viral DNA, and they interact with the two essential divalent cations in the integrase active site in the presence of DNA. [38] InSTIs were first approved for clinical use in 2007 for the treatment of PLHIV. In terms of safety, tolerability, fast viral clearance and high potency, these classes of drugs have remarkable superiority over the other drug classes.[39] Currently, four InSTI drugs are approved: raltegravir (RAL), elvitegravir (EVG) dolutegravir (DTG) and bictegravir (BTG). Another InSTI cabotegravir (CAB) is in the advanced stages of clinical evaluation. The InSTI classes of drugs were previously reserved as salvage therapy; in high-income

countries, they have become the preferred first-line regimen and are recommended by ART guidelines.[40,41]

Due to high cost, access to InSTIs in low-middle income settings has been limited. Recent negotiations and agreements have led to the generic development of DTG and its adoption into WHO guidelines as an alternative first-line regimen.[42] To date, in sub-Saharan Africa, only Botswana and Kenya have administered DTG as part of their standard first-line cART. The recent WHO guidelines recommend DTG as the preferred first-line regimen in regions where the prevalence of pre-treatment NNRTI resistance is ≥ 10%. As countries adopt the guidelines and generic drugs become more readily available, wide-scale access across the region is anticipated.[43]

Even though these drugs are highly efficient, resistance to InSTIs has nevertheless been observed, which ultimately can lead to treatment failure.[44,45.] Some of these identified polymorphic mutations play a significant role in increasing the level of resistance and/or rescuing viral fitness. Primary mutation pathways identified involve substitutions at the T66, E92 (EVG), Y143 (EVG plus RAL), Q148, N155 and R263 (EVG, RAL and DTG) amino acids.[46] Several biochemical studies have shown subtype-dependent influence of natural polymorphisms on the occurrence and activity of InSTI resistance.[47–51,52] For example, subtype B viral strains associated with G118R have a higher fitness cost, recently proposed resistance pathway compared with subtypes C and CRF_02 AG.[48] Consequently, in subtype B viral strains, G118R has been rarely observed. It has been suggested that G118R could be an alternative pathway for DTG resistance in non-subtype B viruses, whereas R263K is the preferred pathway for subtype B viruses.[48] In vitro studies, have shown that mutations located outside the IN gene, can confer a high-level resistance to all InSTIs. Because of this, HIV-1 can use an alternative mechanism route to develop resistance to InSTIs by selecting mutations in the 3' PPT region.[51,53] A rare polymorphism at codon 118 (GGG or GGA) has a low genetic barrier through a single mutation to facilitate the transition of glycine to arginine (AGG or AGA), which in turn has an impact on the occurrence of G118R.[54] Furthermore, the prevalence of the G118R mutation may also differ by subtype but the impact of other polymorphic mutations in the resistance and viral fitness of G118R mutants was shown to be subtype dependent.[49] In South Africa, a study conducted by Brado et al., [52] identified the prevalence of Q148H in 1/314 (0.3%) if co-occurring with additional RAMs. This mutation can lead to resistance against all InSTIs.[52]

1.8. Combination antiretroviral treatment Mechanism of action

The classes of cARTs based on the different targets of the HIV replication cycle involves different mechanisms of action described below:

1.8.1. Nucleos(t)ide reverse transcriptase inhibitors (NRTIs)

NRTIs gains entry into the cell by passive diffusion. NRTIs can be transported by a variety of carrier and ATP-dependent mediated membrane transporters. [55] NRTI phosphorylation is catalyzed by the action of phosphotransferases and cellular kinases into triphosphate forms. [56] NRTI triphosphate forms actively compete with the cell's natural dNTP substrates at HIV-1 RT's polymerase active site for incorporation into the growing cDNA strand, synthesized by HIV-1 RT enzyme. Once incorporated the triphosphate forms cause premature chain termination as they lack a 3'- OH group that would have facilitates the incorporation of the next dNTP substrate.[57]

1.8.2. Non-Nucleoside reverse transcriptase inhibitors (NNRTIs)

NNRTIs gain entry to the cell by passive diffusion as highly lipophilic molecules.[58] Three different mechanism of action express antiviral activity. The first mechanism is the allosteric inhibition – NNRTIs bind in a HIV-RT pocket region. The polymerase active site is 10Å away from the pocket region and cause a conformation change in HIV-RT, a result this affect its ability to bind to the primer during reverse transcription.[59] The second mechanism (although controversial) involves the enhancement of the HIV-1 RT heterodimer formation. This inhibition downregulate the enzyme's activity to simultaneous executes its polymerase and RNA degradation functions. [58, 60] The third mechanism involves the acceleration of cytoplasmic Gag-Pol processing, which leads to the premature activation of the HIV-1 Protease. Premature activation of the Protease occurs when binding of NNRTIs to Gag-Pol proteins promotes oligomerization of the Pol molecules by binding to the p66 subunit of HIV-RT. Premature cleavage of these Gag-Pol constructs restricts proteins required for viral particle assembly [58]

1.8.3. Protease Inhibitors (PIs)

PI cellular entry can be mediated by Clathrin / Caveolae endocytic pathways or via passive diffusion [61] The tight binding interactions of the PI on the dimer surface of the homodimer HIV-1 Protease active site inhibits its function to execute aspartic acid- mediated cleavage of its substrates.[62] This reduces the rate of viral assembly or production of defective viral particles unable to establish another round of infection [63]

1.8.4. Integrase strand transfer inhibitors (INSTIs)

InSTI enter the cell by endocytic adsorption. Once in the cytoplasm, they are able to bind to the Integrase, but only access the enzyme's active site once there has been a conformational change induced, after the enzyme catalyzes the 3'-end processing reaction.[64] The 3'-end processing reaction exposes a dinucleotide (CA) at the end of the viral cDNA that will eventually allow for integration into the host genome via the strand transfer reaction.[64] To catalyze the strand transfer reaction in the

nucleus, Integrase requires divalent Magnesium ions to stabilize its active site, but these metal ions are sequestrated by INSTIs that have a β-diketo acid group (as part of their structure) that has high affinity for the divalent metal ions [46].

1.9. Human immunodeficiency virus treatment and the combination antiretroviral therapy roll-out programme in South Africa

1.9.1. In adults

South Africa launched a national cART rollout programme in 2004. Since its initiation, South Africa's national HIV treatment programme has grown to become the biggest in the world with an estimated 4.5 million people currently receiving treatment.[65] In concordance with the WHO guidelines, the recommended first-line cART in South Africa consists of an NNRTI-backboned regimen of EFV combined with two NRTIs, namely 3TC and tenofovir disoproxil fumarate (TDF) for adults. The recommended second-line cART consists of the NRTIs AZT and 3TC and a ritonavir-boosted PI, usually atazanavir (ATV).[66]

1.9.2. In Children

The South African guidelines recommend for children ≤ three years that first-line cART consist of a LPV/r and two NRTIs, namely abacavir (ABC), 3TC. For children ≥ three years and ≥ 10 kg, the first-line cART regimen includes ABC, 3TC and EFV. For adolescents ≤ 15 years, the regimen contains TDF, emtricitabine (FTC) or 3TC and EFV. If viral load is undetectable (< 50 RNA copies/mL), patients currently receiving a d4T-based regimen can be changed to ABC. The WHO guideline also recommends the PI lopinavir co-formulated with ritonavir (lopinavir–ritonavir) in a four-to-one ratio in first-line cART for children younger than three years, based on its superiority compared with NVP, regardless of previous NVP exposure to prevent mother-to-child HIV transmission.66 For children who have failed the first-line PI-based LPV/r regimen, experts must be consulted and resistance testing in patients > 12 months and adherent to treatment must be considered. Children who have failed the NNRTI-backboned regimen of EFV or NVP must be changed to an NRTI-backboned regimen of LPV/r.

1.9.3. When to initiate combination antiretroviral therapy

The question of when to initiate cART has long been the subject of intense debate.[67–69.] Initially, cART was recommended for only symptomatic PLHIV. The hit-hard-and-hit-early strategy covered a large percentage of asymptomatic PLHIV with HIV RNA > 10 000 RNA copies/mL or a CD4+ T-cell count < 500 cells/μL.[70] Due to long-term cART-related toxicities, in 2001 the WHO recommend

a shift in guidelines, postponing therapy until CD4+ T-cells declined to < 350 cells/μL. Clinical trials identified benefits for those initiated on cART with a CD4+ T-cell count around 350-500 cells/μL as opposed to < 350 cells/μL.[71,72] The outcome of these trials led the WHO to again reconsider its guidelines to cART initiation at a CD4+ T-cell count ≤ 500 cells/μL.[73]

International randomised clinical trials and the Strategic Timing of Antiretroviral Treatment (START) study provided compelling evidence of the importance and benefits of starting cART as soon as possible after HIV diagnosis.[74] This evidence gave birth to new guidelines that recommended that all PLHIV be initiated on cART regardless of their CD4+ T-cell count.[74] Early initiation of cART have clinical benefits, but do not necessarily reduce the size of the HIV reservoir. A study conducted by Cirion *et al* [75] identified 14 elite controllers (these patients were able to control their infection without using cART) after three years of cART initiated during primary HIV infection. In this study, the researchers showed that viral control with therapy, for several years, might be associated with early and prolonged cART. Their findings support the initiation of early cART. Early cART also limits the genetic diversity within HIV DNA reservoirs,[76,77] improves survival[78] and reduces the risk of HIV-1 transmission.[79]

1.9.4. Role of an extra κB transcription factor in HIV-1C LTR gene

Globally, HIV-1 subtype C (HIV-1C) is responsible for more than 75% of new infections and 50% of infections [1]. In South Africa, HIV-1C is responsible for the majority of infections. HIV-1 subtype C, unlike other HIV-1 subtypes, has a considerable variation in the number of NFκB sites 80 and most of the molecular variations are linked to the viral promoter, where host and viral transcription factors bind to regulate viral transcription [81.] The LTR is an important regulatory gene region for the viral life cycle and the HIV-1 promoter, located at the 5' LTR, contains several transcription factor-binding sites (TFBS). The HIV-1C LTR has the most genetically divergent sequences. One of the most significant genetic variations associated with the ancestral subtype C is the presence of three NFκB motifs in the LTR [82]. In most HIV viral subtypes, including subtype B, only two genetically identical NFκB motifs are present. [83] NFκB is a key nuclear transcription factor involved in HIV-1 transcription, triggering the viral replication through TAT/TAR mediation. Variation in LTR was observed in Circulating Recombinant Form (CRF-01), whereby NFκB was replaced by a GA binding protein (GABp) motif. The ancestral subtype C viruses contains three NFκB sites, but we have observed that at least three different promoter variants have been replacing the ancestral subtype C [84.] These viral strains contain additional NFκB, NFκB-like or RBEIII sites and the acquisition of an additional NFκB site remains a distinct property of the HIV-1C. As a result, this might strengthen the viral promoter activity at the levels of transcription initiation and elongation. This might also suggest

that the acquisition of a 4th NFκB site by emerging strains can lead to more infectious pathogenic strains [84,85,86,88.] In this thesis, we characterized the genetic variation in HIV-1 LTR from subtype C sequences derived from South African PLHIV-1.

1.10. In-silico integrase analyses

Computational approach was explored to the study of HIV-1 IN structure such as the usage of homology modelling, molecular docking and molecular dynamics have been very instrumental in studying the functional and structural details of HIV-1 IN.

1.10.1. Homology Modelling

Homology modelling also referred to as comparative modelling, is a computational method which is used to generate a structural model of a protein [89]. This method relies upon the observation that similar sequence implies similar structure. Provided with an experimentally solved structure, it is possible to generate a reliable homologous structure from an experimentally solved structure that has a sequence similarity of 30% or greater [89].

Homology modelling has played a major role in drug discovery and development [89]. Homology modelling has been used to gain structural understanding and drug development insights into HIV-1 IN. Early models exploited the homology shared with TN5 transposase to generate an homology model. One such Tn5 based homology model study found that residues 140-149 represents a mobile catalytic loop, the study also showed that residues responsible for DNA binding are highly conserved [90]. This mobile loop would be further studied in additional in-silico work as it plays key roles in drug resistance [48]. The experimental solving of the Protype Foamy Virus (PFV) IN protein provided a new structure for which reliable homology could be generated [91–93.]The IN active site in the catalytic core of each integrase is nearly identical. As a result of this high identity it is possible to bind InSTI's to PFV integrase, this enabled an accurate understanding of how InSTI's may bind to HIV-1 IN and the effect of resistance mutations upon this binding [94.]Models were also generated using domains of the IN proteins as the starting homologous structure from which to model. A study have shown that there is dimer interface between the integrase N-terminals was observed [95.] The Cryo-EM solving of a full length tetrameric HIV-1 integrase complex has made it possible to produce highly plausible full length homology models [96.]Studies have used this model as a template to generate an homology model, polymorphisms unique to certain subtypes have been investigated [52,97]. It was possible to assess their impact upon the structure of the IN protein [52,97.]

1.10.2. Human Immunodeficiency virus reservoir during combination antiretroviral therapy and reservoir establishment

The first documentation of a latent HIV reservoir was reported in 1997 by three groups (Chun *et al*., Finzi *et al*., and Wong *et al*.,[69–71]) This group of scientists stimulated viral production from cART individuals with undetectable viral load, using their peripheral resting CD4+ T-cells.[69–71.] Viral integration into the host DNA enables HIV to exist as a provirus through the suppression of viral expression. HIV reservoirs remain the major barrier to HIV cure as they persist and give rise to viral rebound in plasma once there is any interruption of cART.[72] Although HIV reservoirs are diverse and dynamic, HIV reservoirs in the memory CD4+ T-cells are considered as the major barrier to a cure. Two pathways through which HIV infects memory CD4+ T-cells have been hypothesised: through the infection of activated CD4+ T-cells that survive and revert to a memory phenotype, and via direct infection of and integration into the genome of the resting memory CD4+ T-cells.

1.11. Methods to test for drug resistance

The two main methods available to test for HIV-1 drug resistance are the genotypic and phenotypic methods. There are different genotypic approaches available to detect drug resistance mutations (DRMs), but in this study, I used direct population-based sequencing and Next Generation Sequencing (NGS). Genotypic resistance testing (GRT) involves the detection of known resistance mutation in the genes of HIV-1, which confers limited susceptibility to cART.

1.11.1. Population Based Sanger Sequencing

This approach is based on direct polymerase chain reaction (PCR) followed by Sanger (dideoxynucleotide) sequencing. In South Africa, this method is the most preferred and forms part of standard of care to guide cART regimen prescription and, most importantly, to monitor DRMs in PLHIV. Direct sequencing can be performed using either commercially available kits or in-house methods.[73,74] These assays are capable of producing a complete genome nucleotide[75]. The obtained virus sequence is a consensus sequence generated from a population of viral genomes, therefore the name ‘population sequencing’. In order to determine the presence of a possible DRM, derived viral sequences are aligned and compared to a reference sequence of a laboratory wild-type strain. Clinically relevant mutations can be determined by using the Stanford drug resistance database.[20,76] According to the international guidelines, population-based sequencing is recommended as the standard GRT approach to monitoring DRMs[77]; however, these approaches have their limitations. First, they are not sensitive enough to detect viral populations below 20% of the total viral population, which may allow clinically relevant minor variants to become undetected. In high-income countries, these assays are relatively inexpensive to monitor DRMs, but in resource-limited settings, they are

not cost-effective and not easy to implement. Second, interpretation of the viral sequence is complex and it is challenging and time-consuming to apply distinct viral mutations to cART management. Alternatively, more sensitive GRT has been developed, such as deep sequencing (NGS), which could be used as an option in resource-limited settings by targeting key DRMs.

1.11.2. Next-Generation Sequencing (NGS)

NGS platforms have revolutionised the field of sequencing and have made it possible to generate high throughput of sequence data from biological systems in a single reaction run. There are different platforms accessible for deep sequencing, such as Illumina, Ion Torrent and the Nanopore DNA sequencing technology. Due to cost and maintenance requirements, their availability has been limited. The introduction of benchtop sequencing platforms with the benefit of high-throughput data has been shown to be cost-effective and less time-consuming compared to the conventional population-based sequencing approach[106] and might be a more suitable option in clinical diagnostics. The sequencing approach in the Illumina MiSeq platform is characterised by synthesis technology that enables tracking of fluorescently labelled nucleotides as they are added in a parallel mode to the DNA strands. Illumina MiSeq, as compared to the other NGS platforms, has been shown to be more advantageous in terms of simpler laboratory workflows, reduced hands-on time and deeper sequencing coverage to detect single polymorphisms with minimum error rate.[107,108] Several studies have shown that the Illumina MiSeq platform is cost-efficient to generate high-throughput data and is a promising approach for the surveillance of primary DRMs in low-to-middle-income settings (LMICs).[109–111]

1.11.3. Phenotypic resistance testing

Phenotypic resistance can be referred to as resistance against cART regimen, which can be detected in viral culture as a reduction in susceptibility of a molecular clone or an isolate to any of the classes of cART treatment regimen in comparison to a standard viral isolate or molecular clone. Genotypic testing of HIV-1 viral strains can be used to infer phenotypic resistance. Phenotypic testing was developed when an HIV isolate was cultured in a CD4-expressing HeLa cell line and the reduction of syncytial foci in the presence of AZT was associated with the degree of AZT susceptibility [112]. This assay is only limited to be used for syncytium-inducing isolates that can only be found late in disease. This led to the development of an assay based on co-culture of infected cells with peripheral blood mononuclear cells (PMBCs) from HIV negative donor, which could detect anon-syncytium-inducing isolates [113]. In previous years, viral culture and co-culture requires biosafety level 3 facilities. The biosafety level 3 is required to protect laboratory technologist. Due to the development of reporter cell lines transfected with non-infectious plasmid vectors with inserts from the patient-derived RT and/ or protease genes and expressing a fluorescent or chemiluminescent reporter protein [114–116].

The results of this thesis are summarised in the following chapters:

1. **Chapter two:** Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and availability of integrase inhibitors in Cape Town, South Africa. *Manuscript published*, *Nature – Scientific Report*. PMID: 29549274.
2. **Chapter three:** Drug resistance mutations against protease, reverse transcriptase and integrase inhibitors in people living with HIV-1 (PLHIV-1) receiving second-line antiretroviral therapy on the South African National treatment program. *Manuscript under review, Nature – Scientific Report*. Manuscript reference number: SREP-19-20242.
3. **Chapter four:** Increased protease inhibitor drug resistance mutations in minor viral quasispecies on HIV-1 infected patients suspected of failing on national second line therapy in South Africa. *Manuscript under review,* Journal of Antimicrobial Chemotherapy (JAC). Manuscript reference number: JAC-2019-1065.
4. **Chapter five:** Structural Implications of Genotypic Variations in HIV-1 Integrase from Diverse Subtypes. *Manuscript published Frontiers Microbiology.* PMID: 30116231.
5. **Chapter six:** Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding.
6. **Chapter seven:** Mutations in Long Terminal Repeats κB Transcription Factor Binding Sites in Plasma Virus Among South African People Living with HIV-1**.** *Manuscript published AIDS Research and Human Retroviruses (ARHR)* PMID: 30793917.
7. **Chapter eight** is composed of a general discussion, conclusions and future remarks.

Chapter 2 Analyses of HIV-1 integrase sequences prior to South African national HIV treatment program and availability of integrase inhibitors in Cape Town, South Africa.

2.1. Article title

Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and availability of integrase inhibitors in Cape Town, South Africa. Permission to use as part of thesis was granted by Editor on 29.05.2019.

2.2. Authors list and citations

Brado D, **Obasa AE**, Ikomey GM, Cloete R, Singh K, Engelbrecht S, Neogi U, Jacobs GB.

Number of Citations: 6

Scientific Reports. 2018 Apr 16;8(1):6262. doi: 10.1038/s41598-018-24183-8. PMID: 29549274.

2.3. Author contribution

In the enclosed manuscript; Mr. Brado and I, are joint-first authors, with equal contributions to the work performed. Mr Brado was an undergraduate medical student for three months at the Division of Medical Virology as a DAAD intern. I retrieved the samples from the cold storage. I trained (assisted) Mr. Brado to perform the laboratory experiments, including Viral RNA extraction, polymerase chain reaction (PCR), PCR clean-up reactions and Sanger sequencing. Mr. Brado and I, performed the sequence and data analyses. I wrote the first draft and was responsible for subsequent manuscript updates received from the co-authors.

2.4. Background

Integrase strand transfer inhibitors (InSTIs) are highly efficacious and well tolerated antiretroviral, with fewer adverse side effects, relative to other classes of combination antiretroviral therapy (cART) drugs (or inhibitors). The use of InSTIs raltegravir (RAL), elvitegravir (EVG) and dolutegravir (DTG) has increased dramatically over recent years. There is limited information about the evolution and prevalence of InSTIs resistance mutations in people living with HIV from South Africa. The cohort contains treatment-naïve patient samples from multiple ethnic groups, as well as different sexual orientations. Patients were sampled between 2000 and 2001, before the initiation of South Africa's national HIV treatment program and the introduction of InSTIs.

2.5. Main findings

We identified 85 (92%) of our samples as HIV-1 subtype C (HIV-1C). The consensus sequences generated using the database-derived HIV-1CZA sequences (n = 314) and cohort sequences (n = 85) identified 17 naturally occurring polymorphisms. Among the 17 mutations, 11 were further increased our cohort. Genotypic analyses indicated the absence of major RAMs in the cohort collected before the broad availability of cART and InSTIs in South Africa. The database-derived sequences, InSTI occurred at a rate of 2.85% (9/314) in. RAMs were present at IN-positions 66, 92, 143, 147 and 148, all of which may confer resistance to RAL and EVG, but are unlikely to affect second-generation DTG, except mutations in the Q148 pathway.

2.6. Study significance

These are precious samples, as these patients were sampled nineteen (19) years ago, which means data generated from this study could be considered as baseline InSTIs resistance rate. As the analysed cohort was recruited before the initiation of the HIV treatment program in South Africa, the possibility of RAMs being transmitted by treatment-experienced individuals is highly unlikely. Therefore, we consider the described findings to be a true baseline InSTI resistance rate. There was no cART regimen when these patients were sampled, thus rules out the possibility of RAMs being transmitted amongst the treatment-experienced individuals. InSTIs have been introduced in South Africa, as a component of salvage therapy. Despite this, only a few studies have examined the IN gene in terms of viral genetic variability and resistance mutational patterns. Our study emphasizes on the need for proper drug resistance surveillance in future, in order to track the evolution of the virus in a subtype C predominated setting under the pressure of the new treatment.

2.7. Conclusions

The prevalence of InSTI resistance remains low in the treatment-naïve patients, but with the extension of InSTI as part of first-line cART regimen resistance toward this class of cART might increase. Furthermore, protein modelling showed, naturally occurring polymorphisms might influence the stability of the intasome-complex and therefore may contribute to an overall potency against InSTIs.

2.8. Open access

The article is open access and freely available online. Reprints can also be obtained from the author upon request: obasa@sun.ac.za. PubMed link: https://www.ncbi.nlm.nih.gov/pubmed/29549274.

2.9. Published manuscript

SCIENTIFIC REPORTS

Correction: Author Correction

OPEN

l: 7 December 2017
l: 1 March 2018
l online: 16 March 2018

Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and availability of integrase inhibitors in Cape Town, South Africa

Dominik Brado[1,2], Adetayo Emmanuel Obasa[2,3], George Mondinde Ikomey[4], Ruben Cloete[5], Kamalendra Singh[3,6,7], Susan Engelbrecht[2], Ujjwal Neogi[3] & Graeme Brendon Jacobs[2]

HIV-Integrase (IN) has proven to be a viable target for highly specific HIV-1 therapy. We aimed to characterize the HIV-1 IN gene in a South African context and identify resistance-associated mutations (RAMs) against available first and second generation Integrase strand-transfer inhibitors (InSTIs). We performed genetic analyses on 91 treatment-naïve HIV-1 infected patients, as well as 314 treatment-naive South African HIV-1 IN-sequences, downloaded from Los Alamos HIV Sequence Database. Genotypic analyses revealed the absence of major RAMs in the cohort collected before the broad availability of combination antiretroviral therapy (cART) and INSTI in South Africa, however, occurred at a rate of 2.85% (9/314) in database derived sequences. RAMs were present at IN-positions 66, 92, 143, 147 and 148, all of which may confer resistance to Raltegravir (RAL) and Elvitegravir (EVG), but are unlikely to affect second-generation Dolutegravir (DTG), except mutations in the Q148 pathway. Furthermore, protein modeling showed, naturally occurring polymorphisms impact the stability of the intasome-complex and therefore may contribute to an overall potency against InSTIs. Our data suggest the prevalence of InSTI RAMs, against InSTIs, is low in South Africa, but natural polymorphisms and subtype-specific differences may influence the effect of individual treatment regimens.

Combination antiretroviral therapy (cART) has dramatically reduced HIV infection to a chronic and manageable disease resulting in a near-normal life expectancy[1,2]. However, cART has also led to the development of resistance-associated mutations (RAMs) and transmitted drug resistances (TDRs), which are associated with a higher rate of virological failure[3–5]. With rising levels of drug resistance and first-line cART failure, more patients will require second-line and salvage cART, which shows treatment efficacy in terms of viral suppression in 60% of the cases, but may be up to three times more expensive than the first-line cART[6,7].

HIV-1 Integrase (IN) is responsible for the integration of the viral nucleic material into the host genomic DNA[8]. Integrase strand transfer inhibitors (INSTIs) are highly potent antiretroviral agents with durable efficacy, minimal toxicity and is internationally approved and used for both treatment-naïve and treatment-experienced patients[9–11].

[1]Division of Virology, Institute for Virology and Immunobiology, Faculty of Medicine, University of Wuerzburg, 97080, Wuerzburg, Germany. [2]Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, 7505, Cape Town, South Africa. [3]Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, University of Stockholm, Stockholm, Sweden. [4]CSCCD, Faculty of Medicine and Biomedical Sciences, University of Yaoundé, Yaoundé, Cameroon. [5]South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Western Cape, South Africa. [6]Department of Molecular Microbiology and Immunology, Columbia, MO, 65211, USA. [7]Christopher Bond Life Sciences Center, University of Missouri, Columbia, MO, 65211, USA. Dominik Brado and Adetayo Emmanuel Obasa contributed equally to the work. Correspondence and requests for materials should be addressed to A.E.O. (email: obasa@sun.ac.za)

There are currently three US-Food and Drug Administration (FDA)-approved InSTIs: Raltegravir (RAL), Elvitegravir (EVG) and Dolutegravir (DTG). Two newer INSTIs, bictegravir (BIC) and cabotegravir (CAB), are presently under consideration[12]. The use of higher genetic barrier drugs such as Dolutegravir (DTG) is crucial to the success of salvage therapy to mitigate the emergence of resistant variants[13]. In 2007, the first INSTI RAL was approved for the treatment of patients infected with HIV-1, followed by EVG in 2012. These first-generation InSTIs are highly effective in the treatment of HIV-1-infected patients, but have a low barrier to resistance, resulting in the rapid emergence of RAMs[14,15]. DTG is a second-generation InSTI that was approved by the FDA in 2014[16]. It has a higher resistance barrier than that of RAL and EVG[17]. In the case of DTG, resistance is selected slowly *in vitro*, but has not emerged in studies of therapy-naïve patients until today[18,19]. When compared to an EFV based first-line regimen, patients receiving DTG have shown to be superior regarding viral suppression rates and had stabilized $CD4^+$ T-Cell counts[20]. This is mainly attributed to better adherence and fewer discontinuation rates under treatment. The WHO, however, only names it as an alternative to the above-mentioned first-line regimen, as little research has been done on the use of DTG[21].

Since its initiation in 2004, South Africa's national HIV treatment program has grown to become the biggest in the world, currently treating approximately 3.4 million people[22]. Being in concordance with the World Health Organisations (WHO) guidelines, the recommended first-line combination antiretroviral therapy (cART) in South Africa consists of a non-nucleoside reverse transcriptase inhibitor (NNRTI) backboned regimen of Efavirenz (EFV), combined with two nucleoside reverse transcriptase inhibitors (NRTIs), namely Lamivudine (3TC) and either Tenofovir Disoproxil Fumarate (TDF) for adults or Abacavir (ABC) for children, respectively. The recommended second-line cART consists of the nucleoside reverse transcriptase inhibitors (NRTIs) Zidovudine (AZT) and Lamivudine (3TC) and a Ritonavir-boosted (/r) Protease Inhibitor (PI), usually Atazanavir[23].

In this study we aim to provide further information on the susceptibility and primary drug resistance mutations profile of InSTIs as well as to establish a protocol to screen for Integrase RAMs in an HIV-1 subtype C predominated setting in South Africa.

Results

Patient demographics. The patient demographics are summarized in Supplementary Table 1. We amplified and confirmed successful sequencing, containing all 288 amino acids of the IN gene, for 91 samples.

HIV-1 subtyping. Based on HIV-1 subtyping using online automated tools and phylogenetic analysis, 85 (92%) of the samples were identified as HIV-1 subtype C followed by five (5.4%) as HIV-1 subtype B (5,6%, TV122; TV431; TV404; TV420 and TV356) and one as HIV-1 subtype A1 strains (1.1%, TV412) (Fig. 1).

Resistance mutation analyses. Drug resistance analyses showed that no major InSTI RAMs were present in this study. One sample (TV367) carried the accessory drug mutation G140E, a non-polymorphic mutation, that has been selected *in vitro* before, but which alone does not seem to influence the susceptibility of the virus to InSTIs[24]. Minor, polymorphic, mutations were present in 6/91 samples (6.6%), of which four samples contained the mutation L74I (TV122, TV128, TV173, TV405), one other sample contained the mutation L74M (TV366) and another the polymorphism S230N (TV364).

Of note is, that 55/91 (60.4%) samples carried the M50I polymorphism, all of which were classified as subtype C. M50I does not confer resistance to any of the currently available InSTIs and therefore is not listed as RAM in the Stanford University HIV Drug Resistance Database (https://hivdb.stanford.edu/). However, it has been selected *in vitro*, following a bictegravir (BIC) resistance selection assay[25]. In this M50I succeeded an R263K mutation, and only conferred low-level resistance to BIC (2.8-fold) in this combination. R263K was not present in our cohort.

Database derived IN sequence resistance analyses. After excluding multiple sequences from a patient to avoid overestimation of the variant calling and problematic sequences, we used 314 sequences collected between 1999 and 2007. These, we subsequently screened for the presence of RAMs, and identified 6.4% (20/314) sequences to contain RAMs with only 2.86% (9/314) having major InSTI resistance mutations: Q148H, T66S, E92G, S147G, T66A, Y143YF as well as Y143H. Q148H, T66S, E92G, Y143YF, and S147G were present in one sequence each (0,3%) whereas T66A and Y143H could be detected in two sequences respectively (0.6%). 3,5% (11/314) of the sequences contained four different IN accessory mutations, namely E157Q, T97A, G163, and S230R.

Generation of consensus South African HIV-1 subtype C sequence (HIV-$1C_{ZA}$). The consensus sequences generated using the database-derived HIV-$1C_{ZA}$ sequences (n = 314) and cohort sequences (n = 85) identified 17 naturally occurring polymorphisms D25E, V31I, M50I, I72V, F100Y, L101I, T112V, T124A, T125A, K136Q, V201I, T218I, L234I, A265V, R269K, D278A, and S283G. (Fig. 2). Further profiling of sequences obtained through HIVseq Program, HIV-1B (n = 5278), HIV-1C (n = 1416), cohort sequences HIV-1C-ZA (n = 87) confirmed the findings. Among the 17 mutations 11 were further increased in our cohort.

Molecular modeling. The molecular models were created using the cryoEM structure of HIV-1B IN (Figure 3). The cryoEM structure has an active site mutation E152Q, in the modeled structure, this mutation was reverted to glutamate. Here, we focussed on our five naturally occurring polymorphisms: E25, I50, Y100, I101 and I201 in the modeled structure of HIV-$1C_{ZA}$ IN (Figure A). Our model showed that I50 (M50I mutation) is in the proximity of two strands of substrate DNA from two different monomers and therefore appears important in stabilizing/binding with DNA substrate (Figure B). Residue E25 (D25E mutation) from one monomer forms an ion-pair with K188 of a different monomer in a symmetric fashion (Figure C). The two monomers directly bind DNA substrate

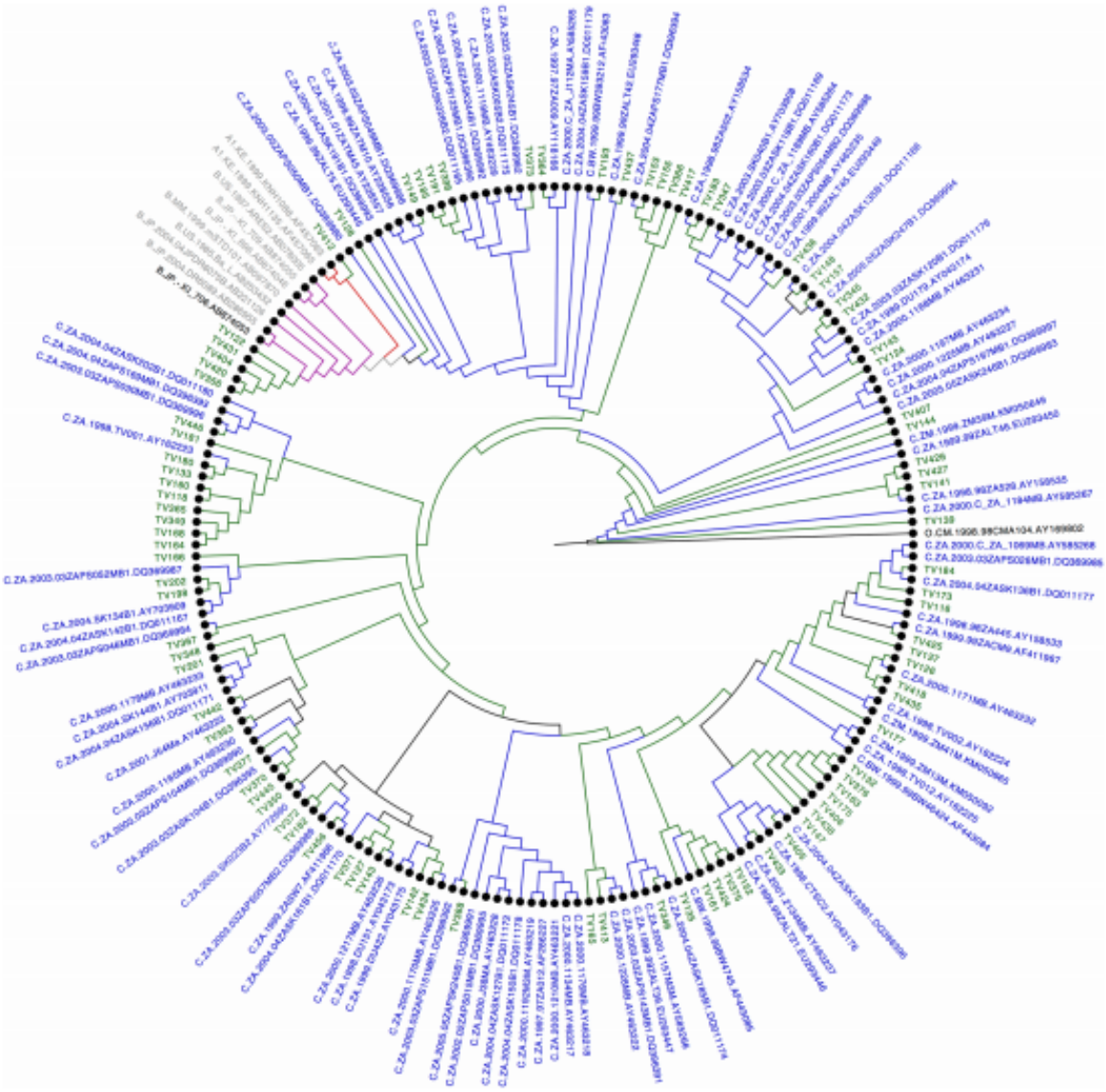


Figure 1. HIV-1 Integrase molecular phylogenetic analysis inferred by the maximum likelihood (ML) method. The ML phylogenetic tree inferred in RAxmL contains 91 patient sequences and HIV-1 reference sequences dataset from 2010 were acquired from HIV-1 LANL database. The evolutionary distances were computed using the general time reverse (GTR) model of nucleic acid substitution with an estimated Gamma shape parameter and invariant sites. The genetic distance is displayed in the scale bar at the bottom of the figure; while the majority of the sequences clusters with HIV-1 subtype C. 87 of the samples clustered with Subtype C reference strains (91.1%), five with Subtype B (5,6%) and one with Subtype A1 strains (1.1%).

and have their active site positioned close to DNA. This interaction is essential in maintaining the tetramer of IN. Mutations Y100 (F100Y) and I101 (L101I) are close to the active site of IN (D64, D116, and E152) in two monomers of which one subunit binds DNA through the active site, and the other one has an indirect interaction. Mutation I201 (V201I) is juxtaposed at the interface of two monomers of IN proteomer forming a hydrophobic interaction suggesting its importance in the maintenance of IN/DNA complex.

Discussion

InSTI containing regimens are considered a new and effective form of salvage therapy for cART-experienced patients failing first and/or second-line cART. South Africa, managing the most extensive HIV treatment program, has faced an increase in resistance rates against NNRTIs, NRTIs as well as PIs in the past years[26]. With second-generation InSTIs not being readily available until recently and drug resistance rates in cART naïve patients in some cases exceeding 10%, these drugs could play an essential role in maintaining treatment options against multi-drug-resistant virus variants and preventing resistant viruses from further spreading[5,27,28]

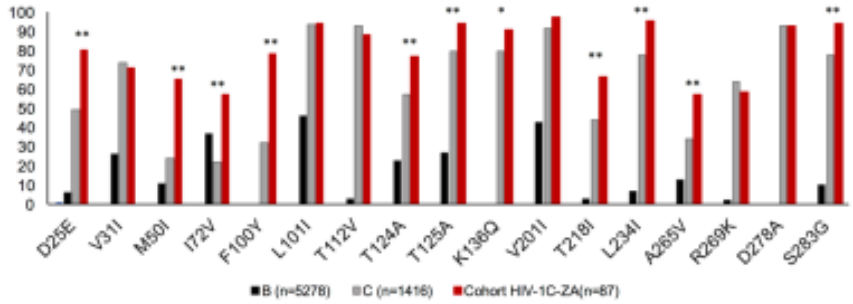


Figure 2. HIV-$1C_{ZA}$ Integrase mutation profiling. Integrase mutation profiling of consensus sequences generated using the database-derived HIV-$1C_{ZA}$ sequences HIV-1B (n = 5278), HIV-1C (n = 1416), cohort sequences HIV-1C-ZA (n = 87) identified 17 naturally occurring polymorphisms D25E, V31I, M50I, I72V, F100Y, L101I, T112V, T124A, T125A, K136Q, V201I, T218I, L234I, A265V, R269K, D278A and S283G. Among the 17 mutations 11 were further increased in our cohort

In this study we analyzed the IN region of HIV-1 infected, cART naïve patients, for the presence of InSTI treatment compromising polymorphisms and mutations. We showed that no primary major resistance mutations against InSTIs were circulating in our study cohort at the time of collection. One accessory mutation (G140E) was observed, while the highly polymorphic mutations L74I/M and S230N were present in 5,5% (5/91) and 1,1% (1/91), respectively. Neither of these mutations is associated with reduced susceptibility to InSTIs. L74I/M, a polymorphism, that has been described in both, cART naïve and RAL or EVG experienced patients before, does not diminish the effect of InSTIs by itself, but can contribute to a high-level resistance, only if co-occurring with major resistance mutations[29,30]. S230N has been reported as a natural variant with a polymorphism rate ranging between 0,5% to 2,0%[31]. It has also been selected by RAL and/or EVG before, *in-vivo* and *in-vitro*, but does not seem to confer resistance to any of the available drugs[31,32].

These findings are in line with previous studies on cART naïve patients confirming the variability of the genomic IN region as well as its lack of major resistance mutations. In 2013 Bessong and Nwobegahay reported the absence of major RAMs in a study conducted in the north-eastern part of South Africa, and a year before, in 2012, Oliveira *et al.* analyzed HIV-1 positive samples from Mozambique for genetic diversity of the IN gene[33,34]. While resistance-associated mutations were not present in this study, the L74M polymorphism was found in 3,4% of the cases. Similar results were observed in Brazil and Europe, before the widespread use of InSTIs[35,36].

Among the 9/314 (2.86%) major InSTI RAMs, present in the database-derived sequences, only Q148H, detected in 1/314 (0.3%), may profoundly affect second-generation InSTIs susceptibility. If co-occurring with additional RAMs, mutations in the Q148 pathway can lead to higher fold resistances against all InSTIs. Despite both first-generation InSTIs, RAL, as well as EVG, selecting for these mutations, they have not yet been described to emerge under initial second-generation InSTI treatment[37]. Y143H (2/314, 0.6%) is usually selected by RAL, and is considered to be a transitional mutation as part of the Y143R resistance pathway. Alone Y143H does not influence the effect of InSTIs, but by further mutating to Y143R it may confer moderate to high-level resistance to RAL, but minimal if any resistance to DTG[38,39].

T66S, T66A, E92G and S147G, found in 0.3%, 0.6%, 0.3%, and 0.3%, respectively, are non-polymorphic mutations, normally selected by EVG treatment. They are associated with moderate to high-level resistance against EVG, although T66 mutations also bear cross-resistance to and are selected by RAL[40,41].

The most frequent IN accessory RAM within the online, retrieved sequences was T97A, being present in 1.6% (5/314), followed by E157Q in 0.96% (3/314), G163R in 0.6% (2/314) and S230R in 0.3% (1/314) of the cases. All of these mutations are found to be within their natural prevalence rates, and although they can confer low-level resistance to both, RAL and EVG, none of these mutations are known to reduce DTG susceptibility, neither *in vitro* nor *in vivo*[41–44].

Interestingly, one case report found single E157Q to be associated with treatment failure of a DTG containing regimen[45]. Therefore, Anstett *et al.* investigated this association in 2016, but could not confirm the result[46]. On the other hand, however, a recent study has also shown that eight patients, who had E157Q mutation and initiated with DTG-based therapy, did not suppress the viremia below detection level after six months of therapy[47]. Hence, causality between E157Q and a reduced DTG susceptibility is debatable and needs further long-term follow-up studies.

Despite higher fold RAMs against InSTIs being absent in most treatment naïve settings, they can emerge under treatment, particularly with first generation InSTIs, as Rossouw *et al.* presented in their case report from May 2016. In this report, they describe the first South African patient to fail EFV based first-line consecutively, ritonavir-boosted Lopinavir backboned second-line, and RAL containing third-line therapy[48]. Poor adherence to the therapy was reported throughout the patient's history, and a final drug-resistance test, performed three years after the initiation of third-line treatment, for the first time included InSTI-resistance testing. This test ultimately confirmed a high-level resistance against RAL and high- or intermediate-level resistance against three of the other four drugs.

Furthermore, this test also revealed cross-resistance to EVG and a low-level resistance against DTG. This cross-resistance to DTG is seldom observed in the only RAL exposed patients and therefore is worrying, especially because InSTI resistances develop significantly less frequently if initially treated with DTG, instead of RAL[49,50].

Nevertheless, this case study raises the concern of emerging InSTI resistance patterns in the South African context. Hence, proper drug resistance surveillance within South Africa will be required, in particular also because, a recent study identified subtype-specific differences in DTG cross-resistance patterns in patients failing RAL[39]. Further, sequence and structure-based analyses showed that the subtype-specific effects were caused by polymorphic residues across subtypes, which significantly affected native protein activity, structure and function of importance for drug-mediated inhibition of enzyme activity[51]. Although DTG showed a high genetic barrier to resistance, subtype-specific differences have been observed in the selection of DTG resistance mutations.

Therefore, we analyzed the position of naturally polymorphic mutations in the context of their ability to impact the stability of intasome. The polymorphisms noted in our analyses appear to be essential for the stability of tetramer and/or binding of DNA substrate in catalytically competent mode. The topological positions of polymorphisms also suggest that the intasome complex stability may differ in different subtypes, which may alter the architecture of the complex and thereby affect InSTI-based therapy outcome.

As the analysed cohort was recruited before the initiation of the HIV treatment program in South Africa, the possibility of RAMs being transmitted by treatment-experienced individuals is highly unlikely. Therefore, we consider the described findings to be a true baseline InSTI resistance rate.

Our data suggest that the introduction of this class of ART drugs, especially second-generation InSTIs, into the national treatment program could help in managing the HIV epidemic in South Africa. However, the possible emergence of formerly described, as well as a subtype and setting specific resistance pathways, requires proper drug resistance surveillance in the future, in order to track the evolution of the virus in a subtype C predominated setting under the pressure of the new treatment.

Conclusion

In the absence of a cure for HIV, long-term cART outcomes need to be monitored efficiently for maximum efficiency. RAMs lead to therapy escape mutants, which can ultimately cause cART failure. We have shown that in the South African context InSTIs is potentially a viable option for salvage therapy. However, there is still a need to keep assessing the RAMs to ensure patients receive the best treatment and care possible.

Methods

Ethics statement. This study was approved by the Health Research Ethics Committee of Stellenbosch University, South Africa (N15/08/071). The study was conducted according to the ethical guidelines and principles of the international Declaration of Helsinki 2013, South African Guidelines for Good Clinical Practice and the Medical Research Council (MRC) Ethical Guidelines for Research. A waiver of consent was awarded to conduct analyses on the samples, as they were obtained between 2000 and 2001, stored since 2001.

Study design. The samples used for analyses (n = 91) were part of a previously described Tygerberg Virology (TV) cohort[52]. The cohort contains treatment naïve patient samples from multiple ethnic groups, as well as different sexual orientations. Patients were sampled between 2000 and 2001, before the initiation of South Africa's national HIV treatment program and the introduction of INSTIs.

Collection of patient samples. Whole blood was collected from patients with ethylenediaminetetraacetic acid (EDTA) tubes. Plasma was separated after centrifugation at 2000 rpm for 10 minutes at 4 °C. The samples were stored at −80 °C for long-term storage. We selected randomly from the stored samples for IN amplification.

Nucleic acid extraction. HIV-1 RNA extraction was performed using the QIAamp Viral RNA Mini Extraction Kit's Spin protocol according to the manufacturer's instructions (Qiagen, Germany). Briefly, 140 µl of plasma was used as a starting volume. Larger starting volumes of 280 µl of plasma was used for some samples with very low viral titers. Viral RNA was stored at −80°c until use.

cDNA synthesis and PCR amplification. The synthesis of complementary DNA (cDNA) and first-round PCR amplification was performed using the Invitrogen SuperScript® III Reverse Transcriptase (RT) reagents (Invitrogen, Germany). HIV-1 Integrase specific primers used for amplification were Poli5 (5′- CACACAAAGGRATTGGAGGAAATG-3′) and Poli8 (5′-TAGTGGGATGTGTACTTCTGAAC-3′), position 4162–4185 and 5195–5217 on the HXB2 reference strain, respectively, with an expected fragment size of 1056 base pairs[53]. The 25 µl reaction volume contained 8,5 µl nuclease-free water 12,5 µl of 2 × reaction mix buffer, 0.5 µl of both forward and reverse primers at a concentration of 5nmol, 0.5 µl of RT/Platinum Taq mix and 2,5 µl of extracted RNA. Reverse transcription was performed at 50 °C for 20 minutes, followed by an initial denaturation at 94 °C for 2 minutes. Forty cycles of amplification were carried out at 94 °C for 15 seconds for denaturation, 30 seconds for primer annealing at 94 °C and 90 seconds for elongation at 68 °C. The final elongation step was done at 68 °C for 10 minutes. For second round PCR amplification, primers Poli7 (5′-AACAAGTAGATAAATTAGTCAGT-3′) and Poli6 (5′-ATACATATGRTGTTTTACTAARCT-3′), with an expected fragment size of 945 base pairs relative to position 4186–4209 and 5107–5130, were used, in combination with the GoTaq® Flexi DNA Polymerase Kit (Promega, USA)[53]. The 50 µl reaction mix consisted of 28.5 µl nuclease-free water, 10 µl of 5 × reaction mix buffer, 0.5 µl of both primers at 20 nmol, 3.0 µl of $MgCl_2$ at 75 nmol and 3 µl of amplified DNA. Second round of amplification started with a denaturation at 95 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 20 seconds, annealing at 55 °C for 30 seconds and elongation at 72 °C for 90 seconds. The final elongation step was done at 72 °C for 10 minutes and amplicons were stored at 4 °C until further use. Positive PCR amplicons were purified from agarose gel according to the manufacturer instructions using the QIAquick PCR Purification Kit (Qiagen, Germany).

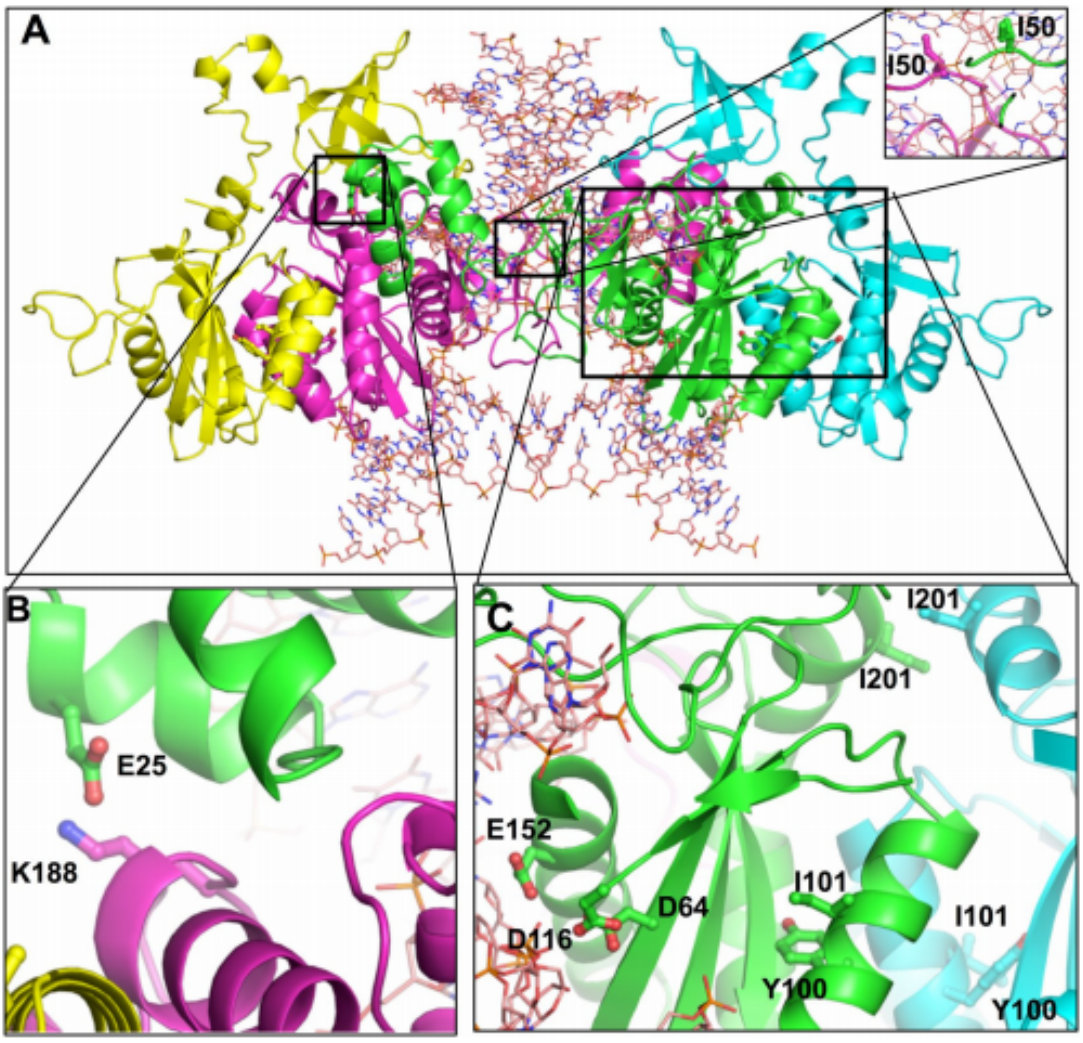


Figure 3. Homology derived molecular model of consensus HIV-1C_{ZA}. Homology derived a molecular model of Con_C_ZA. Figure A shows an intasome consisting of a tetramer of subtype C_ZA and substrate DNA. This structure was generated using the cryoEM structure of HIV-1B IN intasome (PDB file 5U1C) using 'Prime' of Schrodinger Suit using the protocol discussed in Neogi *et al.*, 2016. Inset in panel shows the proximity of I50 to DNA. Two I50 residues from two different subunit interact with DNA from two different sides. Figure B shows the position of E25 (in subunit colored green) that forms a ion-pair with K188 of subunit colored magenta. This is a symmetric interaction as E25 from magenta subunit interacts with K188 of green subunit. This interaction is important in maintaining the tetramer of IN. Figure C shows the active site residues D64, D116 and E152 of IN in one subunit (colored green) together with Y100 and I100 in the same and in the neighboring subunit. This figure also shows the position of I201 in two neighboring subunits. This interaction also appears critical for the maintenance tetramer organization of IN.

Sanger DNA Sequencing. All amplicons were sequenced on both strands with conventional Sanger DNA sequencing using the ABI Prism Big Dye® Terminator sequencing kit version 3.1 and run on the ABI 3130xl automated DNA sequencer (Applied Biosystems, USA). The sequencing reactions were performed according to the manufacturer's instructions. Briefly, sequencing PCR cycling conditions were as follows: Initial denaturation of 95 °C for 60 seconds, followed by 25 cycles of 95 °C for 60 seconds, 55 °C for 7 seconds and 60 °C for 4 minutes. Sequencing primers consist of the above Poli6 and Poli7. Additional sequencing primers were used namely; Poli2 (TAAARACARYAGTACWAATGGCA), relative to position 4745–4766 and KLVO83 (GAATACTGCCATTTGTACTGCTG), corresponding to position 4750–4772[54]. After that, sequences were assembled into contiguous fragments following (Phred quality score > 20) and edited manually using Sequencer version 5.0 (Gene Codes Corporation, USA). The bases were considered ambiguous is any nucleotide was present > 25% of the major peak.

HIV-1 Subtyping and phylogenetic analyses with online programs. HIV-1 subtyping based on integrase sequences was carried out using REGA v3 and COMET-HIV, followed by maximum likelihood phylogenetic analysis[55]. The best fitted general time reverse (GTR) model of nucleic acid substitution with an estimated Gamma shape parameter and invariant sites model using Randomized Axelerated Maximum Likelihood (RAxML) as described previously[56,57].

Additional sequences. To compare our sequences with the rest of the IN sequences from South Africa, we performed a search on the LANL HIV database (https://www.hiv.lanl.gov/components/sequence/HIV/search/search.comp). Our search inclusion criteria included all South African IN sequences and those identified from treatment naïve patients. We selected one sequence per patient and all problematic sequences were excluded from further analyses. Finally, 314 HIV-1 subtype C (HIV-1C) sequences were included in the analyses. Both cohort and database derived South African IN sequences were used to generate the consensus HIV-$1C_{ZA}$ sequence using the Consensus Maker tool available in HIV-1 Los Alamos database using majority value 0.5 (https://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html). HIVseq Program, a literature prevalence of mutations in submitted sequences were used to identify the prevalence of naturally occurring polymorphisms in the HIV-$1C_{ZA}$ sequences in HIV-1B[58].

Molecular Modelling. The homology model of HIV-$1C_{ZA}$ IN tetramer was generated using the cryoEM structure of HIV-1B IN intasome (PDB file 5U1C) in the presence of DNA substrate, using Prime version 4.2 of the Schrodinger Suite (Schrodinger, New York, NY, USA), integrated into Maestro of Schrodinger Suite, (Schrodinger Inc., NY) as described previously[59,60]. The homology model was subjected to energy minimization (5,000 steps) to reduce steric overlap between residues using the "Impact" utility of the Schrödinger Suite and the OPLS_2005 force field as described before[61]. The modeled structure was submitted to the Structure Analysis and Verification Server (SAVES) (https://services.mbi.ucla.edu/SAVES/) as well as Protein Structure Preparation tool of SYBYL-X (version 2.1). No bad contacts were noted in the structures. The backbone torsion angles were checked by Ramachandran plot for allowed conformations of ϕ and φ angles. All angles were in the allowed range. The mutant modeling was conducted with 'Prime' utility of Schrodinger Suite Fig. 3.

References

1. Johnson, L. F. *et al.* Life expectancies of South African adults starting antiretroviral treatment: collaborative analysis of cohort studies. *PLoS Med.* **10**, e1001418 (2013).
2. Trickey, A. *et al.* Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV* **4**, e349–e356 (2017).
3. Gupta, R. K. *et al.* Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. *Lancet (London, England)* **380**, 1250–8 (2012).
4. Kantor, R. *et al.* Pretreatment HIV Drug Resistance and HIV-1 Subtype C Are Independently Associated With Virologic Failure: Results From the Multinational PEARLS (ACTG A5175) Clinical Trial. *Clin. Infect. Dis.* **60**, 1541–1549 (2015).
5. Steegen, K. *et al.* Moderate Levels of Pre-Treatment HIV-1 Antiretroviral Drug Resistance Detected in the First South African National Survey. *PLoS One* **11**, e0166305 (2016).
6. Clinton Health Access Initiative, ARV MARKET REPORT Issue 8, 2017. Available at: https://clintonhealthaccess.org/content/uploads/2017/09/2017-ARV-Market-Report_Final-2.pdf (Accessed: 17th January 2018)
7. Paton, N. I. *et al.* Assessment of Second-Line Antiretroviral Regimens for HIV Therapy in Africa. *N. Engl. J. Med.* **371**, 234–247 (2014).
8. Hazuda, D. J. *et al.* Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287**, 646–50 (2000).
9. Rockstroh, J. K. *et al.* Durable efficacy and safety of raltegravir versus efavirenz when combined with tenofovir/emtricitabine in treatment-naive HIV-1-infected patients: final 5-year results from STARTMRK. *J. Acquir. Immune Defic. Syndr.* https://doi.org/10.1097/QAI.0b013e31828ace69 (2013).
10. Taha, H., Das, A. & Das, S. Clinical effectiveness of dolutegravir in the treatment of HIV/AIDS. *Infect. Drug Resist.* **8**, 339–52 (2015).
11. Heger, E. *et al.* Development of a phenotypic susceptibility assay for HIV-1 integrase inhibitors. *J. Virol. Methods* **238**, 29–37 (2016).
12. Yoshinaga, T. *et al.* Antiviral Characteristics of GSK1265744, an HIV Integrase Inhibitor Dosed Orally or by Long-Acting Injection. *Antimicrob. Agents Chemother.* **59**, 397–406 (2015).
13. Ndashimye, E. High Time to Start Human Immunodeficiency Virus Type 1–Infected Patients on Integrase Inhibitors in Sub-Saharan Africa. *J. Infect. Dis.* **216**, 283–284 (2017).
14. Quashie, P. K., Mesplède, T. & Wainberg, M. A. Evolution of HIV integrase resistance mutations. *Curr. Opin. Infect. Dis.* **26**, (2013).
15. Grobler, J. A. & Hazuda, D. J. Resistance to HIV integrase strand transfer inhibitors: *in vitro* findings and clinical consequences. *Curr. Opin. Virol.* **8**, 98–103 (2014).
16. US Food&Drug Administration, Approved Drug Products, New Drug Application 204790 Tivicay (Dolutegravir). Available at: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/204790Orig1s000Approv.pdf. (Accessed: 17th January 2018) (2013).
17. Llibre, J. M. *et al.* Genetic barrier to resistance for dolutegravir. *AIDS Reviews* **17** (2014).
18. Kobayashi, M. *et al. In Vitro* antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrob. Agents Chemother.* **55**, 813–21 (2011).
19. Raffi, F. *et al.* Once-daily dolutegravir versus twice-daily raltegravir in antiretroviral-naive adults with HIV-1 infection (SPRING-2 study): 96 week results from a randomised, double-blind, non-inferiority trial. *Lancet Infect. Dis.* **13**, 927–935 (2013).
20. Walmsley, S. L. *et al.* Dolutegravir plus Abacavir–Lamivudine for the Treatment of HIV-1 Infection. *N. Engl. J. Med.* **369**, 1807–1818 (2013).
21. World Health Organization (WHO). The use of antiretroviral drugs for treating and preventing hiv infection. 97 Available at: http://apps.who.int/iris/bitstream/10665/208825/1/9789241549684_eng.pdf?ua=1. (Accessed: 17th January 2018) (2016).
22. UNAIDS. Country Report South Africa. Available at: http://www.unaids.org/en/regionscountries/countries/southafrica. (Accessed: 17th January 2018) (2016).
23. Meintjes, G. *et al.* Adult antiretroviral therapy guidelines 2017. *South. Afr. J. HIV Med.* **18**, 24 (2017).
24. Goethals, O. *et al.* Primary mutations selected *in vitro* with raltegravir confer large fold changes in susceptibility to first-generation integrase inhibitors, but minor fold changes to inhibitors with second-generation resistance profiles. *Virology* **402**, 338–346 (2010).
25. Tsiang, M. *et al.* Antiviral Activity of Bictegravir (GS-9883), a Novel Potent HIV-1 Integrase Strand Transfer Inhibitor with an Improved Resistance Profile. *Antimicrob. Agents Chemother.* **60**, 7086–7097 (2016).
26. Van Zyl, G. U. *et al.* Trends in Genotypic HIV-1 Antiretroviral Resistance between 2006 and 2012 in South African Patients Receiving First-and Second-Line Antiretroviral Treatment Regimens. https://doi.org/10.1371/journal.pone.0067188 (2013).
27. Jacobs, G. B. *et al.* HIV-1 Subtypes B and C Unique Recombinant Forms (URFs) and Transmitted Drug Resistance Identified in the Western Cape Province, South Africa. *PLoS One* **9**, e90845 (2014).
28. NATIONAL INSTITUTE FOR COMMUNICABLE DISEASES. Annual Overview 2016/17. 23–24 (2017). Available at: http://www.nicd.ac.za/wp-content/uploads/2017/03/NICD_AR_2016_17.pdf (Accessed: 17th January 2018)

29. Garrido, C. *et al.* Broad phenotypic cross-resistance to elvitegravir in HIV-infected patients failing on raltegravir-containing regimens. *Antimicrob. Agents Chemother.* **56**, 2873–8 (2012).
30. Temesgen, Z. & Siraj, D. S. Raltegravir: first in class HIV integrase inhibitor. *Ther. Clin. Risk Manag.* **4**, 493–500 (2008).
31. Rhee, S.-Y. *et al.* Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology* **5**, 74 (2008).
32. Hombrouck, A. *et al.* Mutations in human immunodeficiency virus type 1 integrase confer resistance to the naphthyridine L-870,810 and cross-resistance to the clinical trial drug GS-9137. *Antimicrob. Agents Chemother.* **52**, 2069–78 (2008).
33. Bessong, P. O. & Nwobegahay, J. Genetic Analysis of HIV-1 Integrase Sequences from Treatment Naive Individuals in Northeastern South Africa. *Int. J. Mol. Sci.* **14**, 5013–24 (2013).
34. Oliveira, M. F. *et al.* Genetic diversity and naturally polymorphisms in HIV type 1 integrase isolates from Maputo, Mozambique: implications for integrase inhibitors. *AIDS Res. Hum. Retroviruses* **28**, 1788–92 (2012).
35. Casadellà, M. *et al.* Primary resistance to integrase strand-transfer inhibitors in Europe: Table 1. *J. Antimicrob. Chemother.* **70**, 2885–2888 (2015).
36. Passaes, C. B. *et al.* Lack of Primary Mutations Associated With Integrase Inhibitors Among HIV-1 Subtypes B, C, and F Circulating in Brazil. *JAIDS J. Acquir. Immune Defic. Syndr.* **51**, 7–12 (2009).
37. Anstett, K., Brenner, B., Mesplede, T. & Wainberg, M. A. HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology* **14**, 36 (2017).
38. Huang, W., Frantzell, A., Fransen, S. & Petropoulos, C. J. Multiple genetic pathways involving amino acid position 143 of HIV-1 integrase are preferentially associated with specific secondary amino acid substitutions and confer resistance to raltegravir and cross-resistance to elvitegravir. *Antimicrob. Agents Chemother.* **57**, 4105–13 (2013).
39. Doyle, T. *et al.* Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades. *J. Antimicrob. Chemother.* **70**, 3080–3086 (2015).
40. Hardy, I. *et al.* Evolution of a novel pathway leading to dolutegravir resistance in a patient harbouring N155H and multiclass drug resistance. *J. Antimicrob. Chemother.* **70**, (2015).
41. Shimura, K. *et al.* Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *J. Virol.* **82**, 764–74 (2008).
42. Abram, M. E. *et al.* Lack of impact of pre-existing T97A HIV-1 integrase mutation on integrase strand transfer inhibitor resistance and treatment outcome. *PLoS One* **12**, (2017).
43. Rhee, S.-Y. *et al.* Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology* **5**, 74 (2008).
44. Malet, I. *et al.* Mutations Associated with Failure of Raltegravir Treatment Affect Integrase Sensitivity to the Inhibitor *In Vitro*. *Antimicrob. Agents Chemother.* **52**, 1351–1358 (2008).
45. Danion, F. *et al.* Non-virological response to a dolutegravir-containing regimen in a patient harbouring a E157Q-mutated virus in the integrase region. *J. Antimicrob. Chemother.* **69**, 2118–22 (2015).
46. Anstett, K., Cutillas, V., Fusco, R., Mesplède, T. & Wainberg, M. A. Polymorphic substitution E157Q in HIV-1 integrase increases R263K-mediated dolutegravir resistance and decreases DNA binding activity. *J. Antimicrob. Chemother.* **71**, 2083–2088 (2016).
47. Neogi, U. *et al. Ex vivo* antiretroviral potency of newer integrase strand transfer inhibitors cabotegravir and bictegravir in HIV-1 non-B subtypes. *AIDS* 1 https://doi.org/10.1097/QAD.0000000000001726 (2017).
48. Rossouw, T. M., Hitchcock, S. & Botes, M. The end of the line? A case of drug resistance to third-line antiretroviral therapy. *South. Afr. J. HIV Med.* **17**, 3 (2016).
49. Fourati, S. *et al.* Cross-resistance to elvitegravir and dolutegravir in 502 patients failing on raltegravir: a French national study of raltegravir-experienced HIV-1-infected patients. *J. Antimicrob. Chemother.* **70**, 1507–1512 (2015).
50. Cahn, P. *et al.* Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naive adults with HIV: week 48 results from the randomised, double-blind, non-inferiority SAILING study. *Lancet* **382**, 700–708 (2013).
51. Quashie, P. K., Han, Y.-S., Hassounah, S., Mesplède, T. & Wainberg, M. A. Structural Studies of the HIV-1 Integrase Protein: Compound Screening and Characterization of a DNA-Binding Inhibitor. *PLoS One* **10**, e0128310 (2015).
52. Jacobs, G. B. *et al.* Emergence and diversity of different HIV-1 subtypes in South Africa, 2000–2001. *J. Med. Virol.* **81**, 1852–1859 (2009).
53. Swanson, P., Devare, S. G. & Hackett, J. Molecular Characterization of 39 HIV-1 Isolates Representing Group M (Subtypes A-G) and Group O: Sequence Analysis of *gag*p24, *pol* Integrase, and *env* gp41. *AIDS Res. Hum. Retroviruses* **19**, 625–629 (2003).
54. Laethem, K. V. *et al.* A genotypic assay for the amplification and sequencing of integrase from diverse HIV-1 group M subtypes. *J. Virol. Methods* **153**, 176–181 (2008).
55. Pineda-Peña, A. C. *et al.* V. A. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol.* **19**, 337–348 (2013).
56. Struck, D., Lawyer, G., Ternes, A.-M., Schmit, J.-C. & Bercoff, D. P. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Res.* **42**, e144–e144 (2014).
57. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
58. Rhee, S.-Y. *et al.* HIV-1 pol mutation frequency by subtype and treatment experience: extension of the HIVseq program to seven non-B subtypes.
59. Passos, D. O. *et al.* Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* **355**, 89–92 (2017).
60. Häggblom, A., Svedhem, V., Singh, K., Sönnerborg, A. & Neogi, U. Virological failure in patients with HIV-1 subtype C receiving antiretroviral therapy: an analysis of a prospective national cohort in Sweden. *lancet. HIV* **3**, e166–74 (2016).
61. Singh, K. *et al.* Biochemical mechanism of HIV-1 resistance to rilpivirine. *J. Biol. Chem.* **287**, 38110–23 (2012).

Acknowledgements

This study was funded by the National Research Foundation (NRF) of South Africa and the Poliomyelitis Research Foundation (PRF) of South Africa. DB was supported by a grant of the German *Excellence Initiative* to the Graduate School of Life Sciences, University of Würzburg. KS acknowledges partial support by National Institute of Health, Clinical and Translational Science CTSA grant UL1 TR002345. UN acknowledges support received from Swedish Research Council Establishment grant (2017–01330). Grant support for this study was received from the National Research Foundation (NRF) of South Africa and the Poliomyelitis Research Foundation (PRF) of South Africa. DB was supported by a grant of the German *Excellence Initiative* to the Graduate School of Life Sciences, University of Würzburg. KS acknowledges partial support by National Institute of Health, Clinical and Translational Science CTSA grant UL1 TR002345. UN acknowledges support received from Swedish Research Council Establishment grant (2017–01330). We also thank Stellenbosch University and the Harry Crossley Foundation for additional financial support.

Author Contributions

G.B.J. and S.E. conceptualized the study SE obtained the T.V. samples. D.B. performed the PCR amplification and sequence analyses with assistance from A.E. D.B. and A.E. both wrote the first draft. K.S. and U.N. performed data analysis and prepared Figures 1–3. G.M.I. and R.C. read, and edited the final manuscript. All authors performed quality control check, data analysis, edited, read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-22914-5.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter 3 Drug resistance mutations against Protease, Reverse Transcriptase and Integrase inhibitors in people living with HIV-1 (PLHIV) receiving second-line antiretroviral therapy on the South African National treatment program.

3.1. Article title

Drug resistance mutations against Protease, Reverse Transcriptase and Integrase inhibitors in people living with HIV-1 (PLHIV) receiving second-line antiretroviral therapy on the South African National treatment program.

3.2. Authors list and citations

Obasa AE, Mikasi SG, Brado D, Cloete R, Singh K, Neogi U, Jacobs GB. **Manuscript submitted and under review by Nature – Scientific Reports.** Manuscript Reference number: SREP-19-20242.

3.3. Author contribution

In the enclosed manuscript, I confirmed that I am the first author. I obtained the patient samples, performed the laboratory experiments, including Viral RNA extraction, polymerase chain reaction (PCR), PCR clean-up reactions and Sanger sequencing. I performed the sequence and data analyses. I wrote the first draft and was responsible for subsequent manuscript updates received from the co-authors.

3.4. Background

In South Africa, 7.1 million people are infected and living with HIV, representing 19% of the global infected population. The South African national cART rollout program started in 2004, with over 4.5 million accessing treatment by the end June 2018. HIV-1 drug resistance can hamper the success of cART. This chapter describes the patterns of RAMs against PIs, RTIs, and InSTIs from patients suspected of failing on the South African second-line national cART programme.

3.5. Main findings

Of the 96 sequences analysed (n= 52, 54%) had M184V/I as the most frequent NRTI RAM. The most frequent NNRTI RAM was K103N/S, which was identified in (n=40, 42%) patients E157Q was identified in (n= 2, 2%) patients as the most common InSTI RAM, alongside with T66I, Y143R and T97A, which are major InSTI RAMs.

3.6. Study significance

The majority of the patients had the M184V RAM. This could have been carried over from first-line cART. Data generated from this study can assist in the development of cART guidelines for patients who experience treatment failure in resource-limited settings where genotyping is not available. We analysed the integrase gene for the presence of treatment-compromising polymorphisms and DRMs against InSTIs. We observed the presence of Y143R in combination with T97A in one of our patients receiving ABC, 3TC, LPV/r. N155H, Q148H/R/K and Y143R/C/H are the three major recognised pathways of genotypic resistance against InSTIs.

3.7. Conclusion

Our study showed patterns of RAMs against RTIs and PIs from patients suspected of failing on the South African second-line National cART program. We observed very low or no primary InSTI RAMs. Non-identification of any RAMs in one third of the patients and presence of PI RAM in only one fifth of the patients indicating the failure may not be due to RAM, but might be due to non-adherence.

3.8. Manuscript under review by Nature – Scientific Reports

Drug resistance mutations against Protease, Reverse Transcriptase and Integrase inhibitors in people living with HIV-1 receiving second-line antiretroviral therapy on the South African national treatment programme.

Adetayo Emmanuel Obasa[1,2], Sello Given Mikasi[1], Dominik Brado[3], Ruben Cloete[4], Kamalendra Singh[2,5,6], Ujjwal Neogi[2], Graeme Brendon Jacobs[1]

[1] Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg 7505, Cape Town, South Africa

[2] Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, University of Stockholm, Sweden

[3] Division of Virology, Institute for Virology and Immunobiology, Faculty of Medicine, University of Würzburg, 97080, Würzburg, Germany

[4] South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Western Cape, South Africa

[5] Department of Molecular Microbiology and Immunology, Columbia, Missouri, 65211, USA

[6] Christopher Bond Life Sciences Center, University of Missouri, Columbia, Missouri, 65211, USA

Corresponding authors: Adetayo Emmanuel Obasa (obasa@sun.ac.za) and Graeme Brendon Jacobs (graeme@sun.ac.za)

Abstract

The South African national combination antiretroviral therapy (cART) rollout program started in 2006, with over 4.4 million people accessing treatment by the end of June 2018. HIV-1 drug resistance can hamper the success of cART. This study determined the pattern of HIV-1 drug-resistance associated mutations (RAMs) after the suspected failure of first- and/or second-line cART regimens in South Africa. During 2017 and 2018, 96 patient plasma samples were selected, including those of 17 children and infants. All patients were receiving a boosted protease inhibitor as part of their cART regimen. The viral sequences were analysed for RAMs through genotypic resistance testing. Of the 96 polymerase sequences analysed, 52 (54%) had M184V/I like the most frequent nucleoside reverse transcriptase inhibitor RAM. The most frequent non-nucleoside reverse transcriptase inhibitor RAM was K103N/S, which was identified in 40 sequences (42%). Protease inhibitor RAMs M46I and V82A were present in 12 (13%) of the sequences analysed. The RAM E157Q was identified in two (2%) of the integrase sequences as the most frequent integrase strand-transfer inhibitor (InSTI) RAM identified, along with one (1%) each of T66I, Y143R and T97A. We propose continued viral load monitoring for better management of infected patients.

Keywords

HIV-1; reverse transcriptase inhibitor (RTI); protease inhibitor (PI); integrase strand-transfer inhibitor (InSTI); resistance; South Africa; resistance-associated mutations (RAMs); combination antiretroviral therapy (cART)

Introduction

Exceptional improvements in combination antiretroviral therapy (cART) regimens has changed HIV/AIDS from a deadly pandemic to a chronic and manageable disease [1]. cART has made significant contributions to reducing the rates of morbidity and mortality in people living with HIV (PLHIV) and has led to better management of infection at an individual level, not only in high-income countries, but also in low- and middle-income countries [2,3]. South Africa's national HIV cART programme was introduced in 2006 with a public health approach. It is the largest in the world, giving treatment access to 4.4 million people by June 2018 [4]. Besides problems related to adherence, the development and spread of drug resistance has constantly challenged the long-term management of PLHIV in public health settings, where patients are often monitored using clinical or immunological parameters [5].

In accordance with the World Health Organization's (WHO) guidelines, the recommended first-line cART in South Africa consists of a non-nucleoside reverse transcriptase inhibitor (NNRTI) backboned regimen of efavirenz (EFV), combined with two nucleoside reverse transcriptase inhibitors (NRTIs), namely lamivudine (3TC) and either tenofovir disoproxil fumarate (TDF), for adults, or abacavir (ABC), for children. The recommended second-line cART consists of the NRTIs zidovudine (AZT) and 3TC and a ritonavir-boosted (/r) protease inhibitor (PI), usually atazanavir (ATV) [6]. The WHO guidelines also recommend the PI lopinavir co-formulated with ritonavir (lopinavir/ritonavir [LPV/r]) in a four-to-one ratio in first-line cART for children younger than three years, based on its superiority when compared with nevirapine (NVP), regardless of previous NVP exposure to prevent mother-to-child HIV transmission [6].

In vitro studies on PI-naïve PLHIV-1 subtype C, have indicated wide variations in their respective susceptibility to the PIs LPV/r and ATV [7]. Observational studies from sub-Saharan Africa have shown a 14–32% prevalence of virological failure to second-line boosted PI- (bPI) based cART [8,9]. In South Africa, reports of drug resistance patterns in patients receiving bPIs are scarce [10]. With this study, we aimed to identify the pattern of acquired drug resistance mutations (DRMs) among PLHIV in South Africa receiving bPI second-line cART. Furthermore, we characterised the presence of primary integrase strand-transfer inhibitor (InSTI) DRMs in this specific population.

Results

In this study we included patients suspected of failing second-line cART regimens. We confirmed the successful amplification and Sanger sequencing of the protease, reverse transcriptase and integrase gene fragments of the HIV-1 polymerase gene for 96 samples. Figure 3.1. shows the observed DRMs in patients failing second-line cART. Among the patients, 10% (10/96) of the samples did not have an indicated in their patient file as being either male or female. Hence, they were classified as unknown. The ages ranged from 2 to 66 years.

Seventeen (n = 17; 18%) of the patients were 16 years or younger. Of these patients, three (n = 3; 3%) were female, 12 (13%) were male, while for two (2%) the gender had not been disclosed by the physician. The NRTI cART regimen combinations administered were ABC plus 3TC (n = 6; 6% of patients), AZT plus 3TC (n = 6; 6%), stavudine [d4T] plus 3TC (n = 1; 1%), TDF plus emtricitabine [FTC] (n = 1; 1%), ritonavir and truvada (n = 1; 1%) and TDF plus AZT (n = 1; 1%). Fourteen (15%) had received LPV/r, two (2%) had received ATV and one (1%) had received darunavir (DRV) as their bPIs.

We had a total of 76 adults; 55 (57%) were women and 22 (23%) were men, while with two (2%) the gender was not disclosed by the physician. The most commonly used NRTI combination was AZT plus 3TC (n = 56; 58%), compared with those receiving ABC plus 3TC (n = 11; 11%) and TDF plus 3TC (n = 8; 8%). Other regimens given were d4T plus 3TC (n = 2; 2%), truvada (n = 2; 2%), TDF plus etravirine [ETR] (n = 1; 1%) and AZT plus TDF (n = 1; 1%). Fifty-eight (60%) had received LPV/r and six (6%) had received ATV as their bPIs and three (3%) are currently receiving DRV.

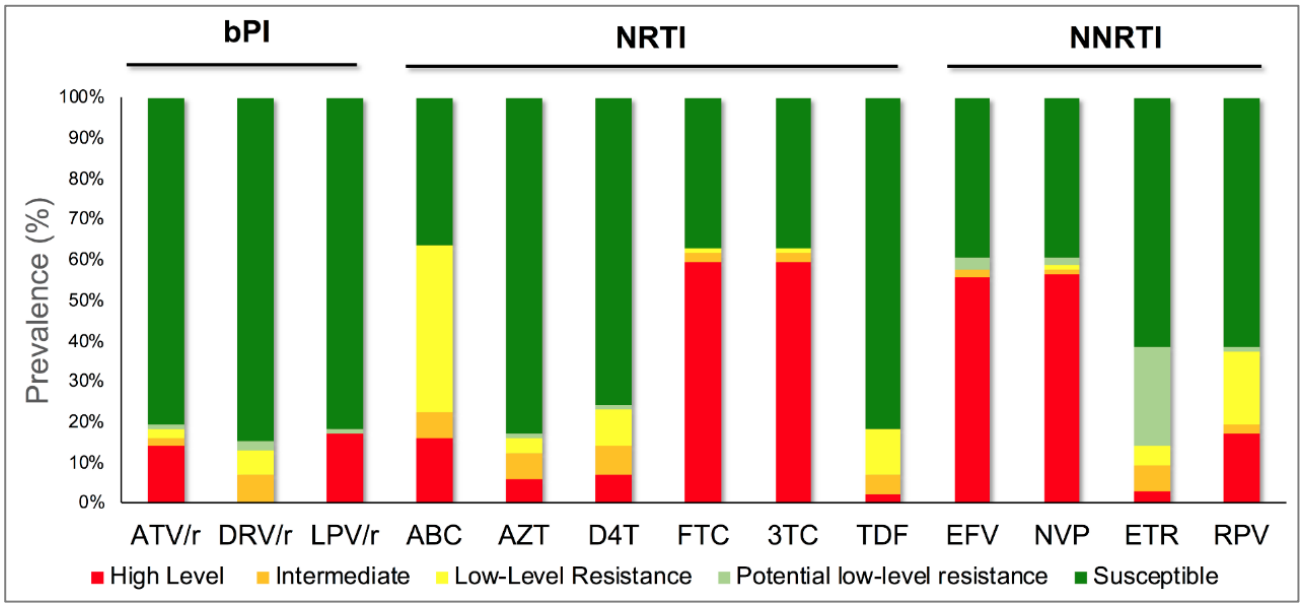


Figure 3.1 Observed drug resistance rates in patients receiving cART

The observed DRMs in patients failing second-line cART. High-level resistance was shown in 17 (17%) and 14 (14%) of PLHIV against LPV/r and (ATV), respectively, while seven (7%) showed intermediate cross-resistance to DRV. Despite not being NNRTIs based treatment more than half of the patients were shown to have high-level resistance to NVP (57%, n = 56) and EFV (56%, n = 55).

NRTI resistance-associated mutations

Table 3.1 shows the number of resistance-associated mutations (RAMs) observed among the 96 sequences analysed. We observed M184V/I as the most prevalent NRTI mutation. M184V/I was detected in 52 (54%) patients suspected of failing cART. Of these, 36 (69%) patients were receiving a combination of AZT plus 3TC, compared with 13 (14%) patients receiving an ABC plus 3TC combination. M184V/I was also found in two (4%) patients receiving TDF plus 3TC, compared with those receiving FTC plus TDF (n = 1; 2%). L74V was detected in five (5%) patients – three (3%) patients receiving ABC plus 3TC, and two (2%) patients receiving AZT plus 3TC. The K65R/N mutation occurred in five (5%) patients; K65R occurred in two (2%) patients receiving AZT plus 3TC, and in one (1%) patient receiving TDF plus FTC. The Y115F mutation occurred in five (5%) patients. However, Y115F occurred more often in patients receiving AZT plus 3TC (n = 4; 4%); it occured in one (1%) patient receiving ABC plus 3TC. The A62V mutation occurred in one (1%) patient receiving AZT plus 3TC. V75I occurred in one (1%) patient receiving AZT plus 3TC.

Thymidine analogue mutations (TAMs) were grouped into TAMs1 and TAMs2. The most frequent TAMs observed were K70R/E in 10 (10%) patients receiving AZT plus 3TC, in three (3%) patients receiving ABC plus 3TC, and in one (1%) patient receiving FTC plus TDF. D67N occurred in five (5%) patients receiving AZT plus 3TC and in three (3%) patients receiving ABC plus 3TC. D67N

occurred in one (1%) patient receiving both TDF plus 3TC and an FTC plus TDF-based regimen. The K219E/Q mutation occurred in six (6%) patients receiving AZT plus 3TC, and in two (2%) patients receiving FTC plus TDF. The M41L mutation occurred in three (3%) patients receiving AZT plus 3TC, and in one (1%) patient receiving 3TC plus ATV. T215Y occurred in two (2%) patients receiving AZT plus 3TC.

Table 3:1 Characteristics and patterns of mutations in 96 patients receiving bPIs

Variable	N (%)
Gender	
Female	**58 (60%)**
Male	**34 (35%)**
Unknown	**4 (4%)**
Second line treatment regimen	
AZT, 3TC, LPV/r	**47 (49%)**
ABC, 3TC, LPV/r	**17 (18%)**
TDF, 3TC, LPV/r	**8 (8%)**
AZT, 3TC, ATV	**6 (6%)**
TDF,FTC,LPV/r	**5 (5%)**
Others	**13 (13%)**
Major PI Mutations	
Any PI Major Mutations	**18 (19%)**
>1 PI Major Mutations	**17 (18%)**
I47A/V	**3 (3%)**
I50L/V	**2 (2%)**
I54V	**10 (10%)**
I84V	**7 (7%)**
L76V	**7 (7%)**
M46I	**12 (13%)**

V32I	**2 (2%)**
V82A	**12 (13%)**
Major NRTI resistance mutations	
Any NRTI Mutations	**65 (68%)**
>1 NRTI Mutations	**30 (31%)**
M184V/I	**52 (54%)**
T69D	**2 (2%)**
L74V	**5 (5%)**
K65R/N	**5 (5%)**
Y115F	**5 (5%)**
TAM-1 pathway	
M41L	**4 (4%)**
T215Y	**2 (2%)**
TAM-2 pathway	
D67N	**11 (11%)**
K70R/E	**17 (18%)**
K219E/Q	**11 (11%)**
Major NNRTI resistance mutation	
Any NNRTI Mutations	**62 (65%)**
>1 NNRTI Mutations	**41 (43%)**
Y181C	**1 (1%)**
K103N/S	**40 (42%)**
G190A/S	**10 (10%)**

K101EP	**6 (6%)**
E138AGKQ	**11 (11%)**
H221Y	**2 (2%)**
M230L	**1 (1%)**
P225H	**14 (15%)**
V106M	**13 (14%)**
V108I	**3 (3%)**
Y188L	**8 (8%)**
L100I	**2 (2%)**
TAMS*	**45 (47%)**
D67N, M41L, T215Y/F, K219E/Q, K70R, and L210W	
M184V and TAMS	**15 (16%)**
Integrase (IN) mutations	
Major IN mutation	
T66I	**1 (1%)**
Y143R	**1 (1%)**
Y143R and T97A	**1 (1%)**
IN Accessory mutations	
E157Q	**2 (2%)**
T97A	**1 (1%)**

NNRTI resistance-associated mutations

Table 3.1 shows the number of NNRTI RAMs observed. We observed that the K103N/S mutation occurred in 41 (43%) of those patients failing cART. Of the patients with this mutation, 24 (59%) patients were receiving AZT plus 3TC, and five (5%) patients were receiving ABC plus 3TC. P225H occurred in 10 (10%) patients receiving AZT plus 3TC, in three (3%) patients receiving ABC plus 3TC, and in two (2%) patients receiving TDF plus 3TC. V106M occurred in 10 (10%) patients receiving AZT plus 3TC, and in two (2%) patients receiving ABC plus 3TC.

Y188L occurred in five (5%) patients receiving AZT plus 3TC, and in two (2%) patients receiving TDF plus ATV, and in one (1%) patient receiving ABC plus 3TC. G190G/A occurred in six (6%) patients receiving AZT plus 3TC, and in three (3%) patients receiving ABC plus 3TC. K101EP occurred in three (3%) patients receiving AZT plus 3TC, and in one (1%) patient receiving ABC plus 3TC. E138QGA occurred in four (4%) patients receiving AZT plus 3TC, in three (3%) patients receiving ABC plus 3TC, and in one (1%) patient receiving TDF plus 3TC. V108I occurred in one (1%) patient receiving AZT plus 3TC. H221Y and M230L occurred in one (1%) patient each; both these patients were receiving AZT plus 3TC.

PI resistance-associated mutations

Of the 96 patients, 75 (81%) were receiving the LPV/r-containing regimen, followed by the ATV- (n = 8; 8%) and DRV- (n = 4; 4%) containing regimens. We identified 18 (18%) patients with major PI RAMs. Of those, a substantial majority of 16 (89%) patients were receiving LPV/r as their bPI regimen, while two (11%) patients were receiving ATV (Table 2). The most common major PI RAMs observed were M46I and V82A (n = 12; 12%); I54V (n = 10; 10%); I84V and L76V (n = 7; 7%); I47A/V (n = 3; 3%); I50L/V (n = 2; 2%) and V32I (n = 2; 2%) (Table 1).

InSTIs resistance-associated mutations

In our cohort, we identified major InSTI RAMs in patients who had received the DRV containing regimen. One (1%) patient had a virus sequence with the Y143R major InSTI mutation in combination with the accessory T97A mutation, which confers high-level resistance to raltegravir (RAL), intermediate resistance to elvitegravir (EVG), and potential low-level resistance to bictegravir and dolutegravir (DTG). A patient receiving AZT, 3TC, EFV, d4T and LPV/r had a virus sequence with the T66I mutation (Table 3.1). The mutation identified and classified as an ‘accessory’ integrase, E157Q, occurred in two (2%) patients. One patient was receiving EFV, TDF and 3TC and the other AZT, 3TC and ATV.

Discussion

In this study we analysed 96 HIV-1 sequences from PI-experienced and InSTI-naïve patients for major RAMs. The patients were being treated with a bPI second-line cART regimen and were suspected of virological failure. Second-line cART consisted of two NRTIs, backboned by a PI, if previously treated with an NNRTI-based regimen, and vice versa.

As expected, major DRMs against NRTIs and NNRTIs were present at a rate of 65% (n = 65) and 62% (n = 62), respectively. Despite being on a bPI, only 18% (n = 18) of our study sequences harboured major PI RAMs. This is in line with a previous study conducted in Sweden, where it was predicted that HIV-1 subtype C would be more prone to failure in bPIs [11]. We identified 27 (27%) sequences not showing any DRM against the drug classes mentioned above and therefore could indicate a problem of poor adherence, rather than the selection of resistant variants.

M184V, the most common NRTI RAM, occurred more frequently in patients receiving AZT plus 3TC, in comparison with patients receiving the ABC plus 3TC regimen. Our findings correspond with previous studies conducted in South Africa with PLHIV, showing M184V/I as the most prevalent NRTI mutation [12–17] The K65R and Y115F RAMs occurred more frequently in patients receiving AZT plus 3TC, rather than in patients receiving ABC plus 3TC. TAMS 1 and 2 pathway mutations occurred more frequently in patients receiving the AZT plus 3TC cART regimen, but were low in patients receiving ABC plus 3TC, TDF plus 3TC, and FTC plus TDF. The most frequent TAM was K70R/E, which occurred mostly in patients receiving AZT plus 3TC, as opposed to ABC plus 3TC and FTC plus TDF. In our study, M184V, L74V, K65R and Y115F were the most common major NRTI RAMs in patients receiving LPV/r as their bPIs.

The K103N mutation occurred at a higher frequency in patients receiving AZT plus 3TC or ABC plus 3TC than in those receiving TDF plus 3TC, and 3TC plus d4T. The high rate of K103N RAM is also well documented and has been observed in several previous studies[15,19-19]. The group receiving AZT plus 3TC or ABC plus 3TC showed the highest rates of NNRTI mutations such as P225H, V106M, E138A/G/K/Q, G190A/S, and Y188L occurred most frequently in patients receiving AZT plus 3TC or ABC plus 3TC.

The majority of our patients were receiving LPV/r as their bPIs. The M46I and V82A RAMs were the most common mutations observed in patients receiving LPV/r compared with ATV. We identified a low frequency of M46I and V82A in patients receiving ATV as their bPIs. The group receiving AZT plus 3TC or ABC plus 3TC showed the highest rate of PIs such as I54V, I84V, L76V, I47A/V, I50L/V and V32I. Our findings are in agreement with Neogi et al., [13] study where major PI RAMs were observed in 5% of patients; among them, V82A 65% (28/43), I54V 63% (27/43), L76V 23%

(10/43) and L90M 16% (7/43) were the most frequent. Our findings also shown M46I and V82A RAMs as the most prevalent major PI RAMs. Chimbetete et al. in 2018 observed similar results, with all three drug classes showing their DRMs at similar rates [20]. The most common PI RAM reported by Chimbetete et al. was M46I 28 (33%), followed by I50V 18 (21%) and V82A 18 (21%). We observed more high-level resistance to patients receiving LPV/r compared with ATV and DRV. Van Zyl et al.'s [17] findings are consistent with ours, as they also identified more high-level resistance in patients receiving LPV/r compared with those receiving ARV and DRV.

We analysed the integrase gene for the presence of treatment-compromising polymorphisms and DRMs against InSTIs. We observed the presence of Y143R in combination with T97A in one of our patient receiving ABC, 3TC, LPV/r. N155H, Q148H/R/K and Y143R/C/H are the three major recognised pathways of genotypic resistance against InSTIs [21]. We confirmed Y143R in our study and this mutation in combination with T97A also impaired EVG susceptibility and showed possible low-level resistance. Furthermore, our data suggest that EVG activity is compromised in the presence of any RAL RAM, in this case Y143R. We also identified the presence of E157Q in 2 (2%) patients. The presence of this mutation is concerning, as these mutations are associated with potential low-level resistance to both RAL and EVG. In a previous study conducted by Brado et al [22]. We also found E157Q on HIV-1-infected South African sequences retrieved from the HIV database. Viruses having E157Q were found to be associated with treatment failure of a DTG-containing regimen [22] A study has shown that eight patients who had the E157Q mutation and were initiated with DTG-based therapy did not experience viremia suppression below detection level [23].

Furthermore, we identified the presence of T66I mutation in 1 (1%) patient. T66I confers resistance mutation to low-level resistance to RAL and high-level resistance to EVG. The low prevalence of DRMs to InSTIs in our cohort should not be underestimated. RAMs against InSTI raises the question about the positioning of DTG in the treatment guidelines for South Africa. Previous studies have shown that a considerable minority of patients develop cross-resistance to DTG after exposure to RAL and EVG; resistance to DTG has not yet been reported in patients from South Africa [24,25]. As DTG was proposed as the first-line drug, it is essential to conduct studies in real-life clinical settings to identify the efficacy of DTG with limited viral load monitoring and without drug resistance genotyping may compromise the roll-out of the InSTIs.

Our study had some limitations that merit comments. First, the sample size was small compared to the total number of patients who are receiving cART in South Africa. It is unlikely to be a biased sample, however, because to the best of our knowledge there has not been any statistical study that reports on the number of patients receiving second-line cART from South Africa. Second, the majority of our patients were receiving LPV/r as their bPIs compared to other bPI regimens. We

cannot tell for certain whether the patients having RAL resistance according to the sequence have had access to an RAL-based treatment regimen. Fourth, we did not have any adherence data for these patients. To the best of our knowledge, this is the first study from South Africa to show the resistance pattern against InSTIs-based regimen patients on cART failure. Finally, the DRM data were only based on population sequencing, therefore we could not detect minor mutations below 20% of the population.

In conclusion, we identified patterns of RAMs against reverse transcriptase inhibitors and PIs from patients suspected of failing on the South African second-line national cART programme. Very low or no primary InSTI RAMs were detected in second-line failure patients. The majority of them had M184V mutations that could have been carried over from the first-line cART. The non-identification of any RAMs in one-third of the patients and the presence of PI RAM in only one-fifth of the patients indicate that the failure may not be due to RAM, but might be due to adherence. Given the limited cART drug availability and high public health burden, we strongly propose for continued monitoring of RAMs in our setting. Data generated from this study can assist in the development of cART guidelines for patients who experience treatment failure in resource-limited setting where genotyping is not available. Studies that address operational issues, such as the optimal use of treatment monitoring tools, should be a research priority.

Patients and methods

Ethics statement

The Health Research Ethics Committee of Stellenbosch University, South Africa (N15/08/071) approved the study. The study was conducted according to the ethical guidelines and principles of the Declaration of Helsinki 2013, the South African Guidelines for Good Clinical Practice and the Medical Research Council Ethical Guidelines for Research. The Health Research Ethics Committee of Stellenbosch University, South Africa awarded a waiver of a written informed consent to conduct sequence analyses on these samples.

Study design

HIV-1-positive patient samples were obtained randomly, without any knowledge of drug-resistance pattern, from the diagnostic section at the Division of Medical Virology, Stellenbosch University, and the South African National Health Laboratory Services (NHLS). Samples were collected between March 2017 and February 2018. Patients had their samples submitted for HIV-1 genotypic resistance testing to the NHLS. The NHLS provides routine genotypic antiretroviral drug resistance testing for clinics from the Western Cape, Gauteng and Eastern Cape provinces.

We included samples from children (aged below 12 years) suspected of failing on bPI – which is used as first-line therapy in children – and adults suspected of experiencing virological failure on a bPI second-line cART regimen, for which treatment information, as provided by the physicians, was available. The treatment history was collected retrospectively. The selection consisted of plasma samples (n = 96) obtained from patients receiving second-line cART, according to the South African national cART guidelines [6]. These patients are eligible for InSTI treatment consideration.

Genotypic resistance testing

We performed genotypic resistance testing using viral RNA extracted from plasma. The HIV-1 protease and reverse transcriptase gene fragments were PCR-amplified using a slightly modified protocol as previously described by us [26]. Briefly, HIV-1 protease and reverse transcriptase first-round cDNA synthesis through reverse transcription was done using amplification primers HIV-PR outer 50prot1 (5'- TAA TTT TTT AGG GAA GAT CTG GCC TTC C -3') and HIVRT outer Mj4 (5'- CTG TTA GTG CTT TGG TTC CTC T –3'), position 2085-2109 and 3399-3420 of the HXB2 reference numbering, with an expected fragment size of approximately 1314 base pairs (bps).[27] For second-round PCR amplification, primers 50prot2 (5'- TCA GAG CAG ACC AGA GCC AAC AGC CCC A –3') and NE13 (5'- CCT ACT AAC TTC TGT ATG TCA TTG ACA GTC CAG CT –3'), position 2136-2163 and 3334-3300 of the HXB2 reference numbering, with an expected fragment size of approximately 1300bps, were used [27]. The integrase gene fragment amplification steps were performed as previously described by us [22]. Sequencing reactions were performed as previously described by us [22]. As part of quality control, each of the viral sequences were inferred on a phylogenetic tree in order to eliminate possible contamination. DRMs were interpreted using the Stanford University HIV Drug Resistance Database version 8.7 (https://hivdb.stanford.edu/).

Acknowledgements

The National Research Foundation of South Africa, the Poliomyelitis Research Foundation of South Africa and the NHLS Research Trust funded this study. We also thank Stellenbosch University and the Harry Crossley Foundation for additional financial support. UN is supported by a grant from the Swedish Research Council (2017-01330). We acknowledge Shambhu Prasad Ganeshappa Aralaguppe for performing the sequencing experiments. We also acknowledge Mathilda Claassen for sample collection.

Author contributions

GBJ and UN conceptualised and designed the study. AEO performed laboratory experiments, detailed sequence analyses and wrote the first draft of the manuscript. MSG helped with demographic data

and sample collection. UN performed sequencing experiments and sequence analyses. RC and KS helped with manuscript proofreading and editing. All authors read and approved the final manuscript.

Potential conflicts of interest. All authors: no conflicts.

References

1. Trickey, A. *et al.* Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV* **4**, e349–e356 (2017).

2. Hightower, K. E. *et al.* Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than raltegravir and elvitegravir from wild-type and integrase inhibitor-resistant HIV-1 integrase-DNA complexes. *Antimicrob. Agents Chemother.* **55**, 4552–4559 (2011).

3. Unaids. Data 2017. *Program. HIV/AIDS* 1–248 (2017). doi:978-92-9173-945-5

4. UNAIDS. *UNAIDS. Country Report South Africa. Available at: http://www.unaids.org/en/regionscountries/countries/southafrica. (2018).* (2018).

5. Rousseau, C. M. *et al.* Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *J. Virol. Methods* **136**, 118–125 (2006).

6. Meintjes, G. *et al.* Adult antiretroviral therapy guidelines 2017. *South. Afr. J. HIV Med.* **18**, 24 pages (2017).

7. Sutherland, K. A. *et al.* Wide variation in susceptibility of transmitted/founder HIV-1 subtype C Isolates to protease inhibitors and association with in vitro replication efficiency. *Scientific Reports* **6**, (2016).

8. Sigaloff, K. C. E. *et al.* Second-Line Antiretroviral Treatment Successfully Resuppresses Drug-Resistant HIV-1 After First-Line Failure: Prospective Cohort in Sub-Saharan Africa. *J. Infect. Dis.* **205**, 1739–1744 (2012).

9. Ajose, O., *et al.* Treatment outcomes of patients on second-line antiretroviral therapy in resource-limited settings: A systematic review and meta-analysis. *Aids* **26**, 929–938 (2012).

10. Collier, D. *et al.* Virological outcomes of second-line protease inhibitor-based treatment for human immunodeficiency virus type 1 in a high-prevalence Rural South African setting: A competing-risks prospective cohort analysis. *Clin. Infect. Dis.* **64**, 1006–1016 (2017).

11. Amanda, H. *et al*. "Virological failure in patients with HIV-1 subtype C receiving antiretroviral therapy: an analysisi of a prospective national cohort in Sweden." *Lancet HIV* **3**,(4), pp 1–18 doi: 10.1186/s40945-017-0033-9 (2016).

12. Rossouw, T. M. *et al.* HIV drug resistance levels in adults failing first-line antiretroviral therapy in an urban and a rural setting in South Africa. *HIV Med.* **18**, 104–114 (2017).

13. Neogi, U. *et al.* Mutational Heterogeneity in p6 Gag Late Assembly (L) Domains in HIV-1 Subtype C Viruses from South Africa. *AIDS Res. Hum. Retroviruses* **32**, 80–84 (2016).

14. Penrose, K. J. *et al.* Frequent cross-resistance to rilpivirine among subtype C HIV-1 from first-line antiretroviral therapy failures in South Africa. *Antivir. Chem. Chemother.* **26**, 204020661876298 (2018).

15. Wallis, C. L., *et al*. Protease Inhibitor Resistance Is Uncommon in HIV-1 Subtype C Infected Patients on Failing Second-Line Lopinavir/r-Containing Antiretroviral Therapy in South Africa. *AIDS Res. Treat.* **2011**, 769627 (2011).

16. Steegen, K. *et al.* Moderate Levels of Pre-Treatment HIV-1 Antiretroviral Drug Resistance Detected in the First South African National Survey. *PLoS One* **11**, e0166305 (2016).

17. Van Zyl, G. U. *et al.* Trends in Genotypic HIV-1 Antiretroviral Resistance between 2006 and 2012 in South African Patients Receiving First-and Second-Line Antiretroviral Treatment Regimens. doi:10.1371/journal.pone.0067188 (2013).

18. Wallis, C. L., *et al.* Varied patterns of HIV-1 drug resistance on failing first-line antiretroviral therapy in South Africa. *J. Acquir. Immune Defic. Syndr.* **53**, 480–484 (2010).

19. Steegen, K. *et al.* HIV-1 antiretroviral drug resistance patterns in patients failing NNRTI-based treatment: Results from a national survey in South Africa. *J. Antimicrob. Chemother.* **72**, 210–219 (2017).

20. Chimbetete, C. *et al.* HIV-1 drug resistance and third-line therapy outcomes in patients failing second-line therapy in Zimbabwe. *Open Forum Infect. Dis.* **5**, 1–8 (2018).

21. Doyle, T. *et al.* Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades. *J. Antimicrob. Chemother.* **70**, 3080–3086 (2015).

22. Brado, D. *et al.* Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and availability of integrase inhibitors in Cape Town, South Africa. *Sci. Rep.* **8**, 4709 (2018).

23. Neogi, U. *et al.* Ex-vivo antiretroviral potency of newer integrase strand transfer inhibitors cabotegravir and bictegravir in HIV type 1 non-B subtypes. *Aids* **32**, 469–476 (2018).

24. Mesplède, T. & Wainberg, M. A. Resistance against integrase strand transfer inhibitors and relevance to HIV persistence. *Viruses* **7**, 3703–3718 (2015).

25. Fourati, S. *et al.* Cross-resistance to elvitegravir and dolutegravir in 502 patients failing on raltegravir: a French national study of raltegravir-experienced HIV-1-infected patients. *J. Antimicrob. Chemother.* **70**, 1507–1512 (2015).

26. Jacobs, G. B. *et al.* Phylogenetic diversity and low level antiretroviral resistance mutations in HIV type 1 treatment-naive patients from Cape Town, South Africa. *AIDS Res. Hum. Retroviruses* **24**, 1009–12 (2008).

27. Jean-Christophe Plantier; *et al.* Generic screening test for HIV infection. *AIDS* **20**, 1345–7. (2006).

Chapter 4 Increased protease inhibitor drug resistance mutations in minor viral quasispecies on HIV-1 infected patients suspected of failing on national second line therapy in South Africa.

4.1. Article title

Increased protease inhibitor drug resistance mutations in minor HIV-1 quasispecies from infected patients suspected of failing on national second line therapy in South Africa.

4.2. Author's lists and citations

Obasa A.E, Anoop T Ambikan, Sohamm Gupta, Graeme Brendon Jacobs, Ujjwal Neogi. **Manuscript submitted and it is under review by Journal of Antimicrobial Chemotherapy (JAC) Manuscript Reference Number: JAC-2019-1065**

4.3. Author contribution

In the enclosed manuscript, I confirmed that I am the first author. I obtained the patient samples, performed the laboratory experiments, including Viral RNA extraction, polymerase chain reaction (PCR), PCR clean-up reactions and preparation for high throughput Next Generation Sequencing. I performed the sequence and data analyses. UN, GBJ and I, wrote the first draft and was responsible for subsequent manuscript updates received from the co-authors.

4.4. Background

High throughput sequencing has unique advantages and significantly improves the sensitivity in quantifying the minority HIV drug resistance (HIVDR) variants within the HIV quasispecies. Increased identification of pre-treatment minority drug resistance mutations (DRM) compared to Sanger based sequencing genotypic resistance testing (GRT) were reported from both resource rich as well as resource limited settings (RLS). This chapter evaluated the HIV-1 drug resistance associated mutations (RAMs) at the minor viral populations, by high throughput sequencing genotypic resistance testing (HTS-GRT), in patients suspected of failing on South African national second-line cART regimen with bPIs.

4.5. Main findings

HIV-1 subtyping identified 55 samples as HIV-1 subtype C and one as CRF02_AG. The major PI, NRTI, NNRTI and INI RAMs were observed 27% (15/56), 57% (32/56), 50% (28/56) and 7% (4/56) (Figure 4.1a). PI RAMs were observed in minor viral quasispecies (25%; 14/56), compared to NRTI (11%; 6/56), NNRTI (9%; 5/56) and INI RAM (4%; 2/56). 41% (23/56) did not have any PI RAMs.

4.6. Study significance

We have used high throughput sequencing to type the DRM in both minor (<20% of the population) and major (>20% of the population) viral quasispecies and identified increased PI RAM in minor viral populations. Our study also indicated low level of transmitted INI RAMs in patients failing on bPIs based regimen in South African PLHIV-1.

4.7. Conclusion

This study shows that use of high throughput resistance testing for GRT can greatly improve the identification of PI RAMs in bPI failing patients. Deep sequencing could be of greater value in detecting the acquired resistance mutations early.

4.8. Manuscript under review by Journal of Antimicrobial Chemotherapy (JAC)

Increased protease inhibitor drug resistance mutations in minor HIV-1 quasispecies from infected patients suspected of failing on national second line therapy in South Africa

Adetayo Emmanuel Obasa[1,2], Anoop T Ambikan[2], Soham Gupta[2], Graeme Brendon Jacobs[1#], Ujjwal Neogi[2#*],

[1]Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, 7505, Cape Town, South Africa

[2]Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, Stockholm, Sweden

[#]Equal contribution

***Correspondence:** Ujjwal Neogi, Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden, e-mail- ujjwal.neogi@ki.se, Phone: +46 852483680

Abstract:

Introduction: HIV-1C have been shown to have a greater risk of virological failure towards boosted-protease inhibitors (bPIs), a component of second-line combination antiretroviral therapy (cART) in South Africa. This study aimed at evaluating the HIV-1 drug resistance associated mutations (RAMs) at the minor viral populations, by high throughput sequencing genotypic resistance testing (HTS-GRT), in patients suspected of failing on South African national second-line cART regimen with bPIs.

Materials and methods: Plasma was obtained from 56 patients. The complete *pol* gene was amplified and sequenced with Illumina HiSeq2500, followed by bioinformatics analysis to quantify the RAMs according to the Stanford HIV drug resistance database.

Results: PI RAMs statistically significant ($p<0.001$) were observed in minor viral quasispecies (25%; 14/56), compared to nucleoside reverse transcriptase inhibitors (NRTIs) (11%; 6/56), non-nucleoside reverse transcriptase inhibitors NNRTIs (9%; 5/56) and integrase inhibitors (INIs) RAM (4%; 2/56). The majority of the DRMs in the minor viral quasispecies were observed in the V82A mutation (n=13) in protease, K65R (n=5), K103N (n=7) and M184V (n=5) in reverse transcriptase.

Conclusion: HTS-GRT improved identification of PI and RTI RAMs in second-line cART patients from South Africa, compared to the conventional GRT with ≥20% used in Sanger-based sequencing. Several RTI RAMs like K65R, M184V or K103N and PI RAM V82A were identified in <20% population. Deep sequencing could be of greater value in detecting the acquired resistance mutations early.

Introduction:

High throughput sequencing has unique advantages and significantly improves the sensitivity in quantifying the minority HIV drug resistance (HIVDR) variants within the HIV quasispecies.[1] Increased identification of pre-treatment minority drug resistance mutations (DRM) compared to Sanger based sequencing genotypic resistance testing (GRT) were reported from both resource rich as well as resource limited settings (RLS).[2, 3] However, the role of minority drug resistant variants and its clinical consequences in failure of combination antiretroviral therapy (cART) is debatable. Studies have shown that even in adherent patients, those with pre-existing Y181C mutants have a three times higher risk of virological failure on an efavirenz (EFV) based cART regimen.[4] Several studies have shown that minority pre-treatment drug resistance was associated with reduced treatment efficacy.[3, 5] In contradiction, other studies have indicated that in a population with a relatively low prevalence of DRM, the utility of deep sequencing to detect minority HIV-1 DRM have limited clinical benefit.[6] A more recent study reported that incorporating the minor DRM might improve the predictive value of GRT, but very low thresholds of minority mutations can compromise the test specificity.[7] Data on acquired minority mutations on treatment failure patients are limited.

HIV-1 subtype C (HIV-1C) is the major HIV-1 subtype in South Africa, responsible for more than 90% of infections. The recommended second-line cART consists of the nucleoside reverse transcriptase inhibitors (NRTIs) zidovudine (AZT) or tenofovir (TDF) and lamivudine (3TC) and a ritonavir-boosted (/r) protease inhibitor (PI), usually lopinavir (LPV/r).[8] Our earlier study from Sweden reported that despite good adherence there is an increased risk of virological failure in patients with HIV-1C, on bPI-based regimens.[9] The *ex vivo* and *in vitro* experiments also indicated large variation in susceptibility of HIV-1C viruses in absence of PI resistance associated mutations (RAMs).[10]

Studies have reported that the rates of virological failure on second-line cART are high in resource-limited settings, including South Africa and associated with duration of exposure to previous drug regimens and poor adherence,[11] mostly without any protease RAM.[12] In South Africa, with more than 4.5 million HIV-infected individuals accessing cART, with about 145,000 (~4%) accessing second-line cART.[13] However, the drug resistance pattern in patients failing on bPIs are limited and often described by GRT by Sanger Sequencing.[14] An earlier study with only seven patients indicated the presence of PI RAMs in bPI failure patients, which was missed by bulk Sanger sequencing.[15] Therefore, the primary aim of the present study is to find the level and pattern of HIVDR in minor (<20%) and major viral populations in patients suspected of failing bPIs.

Material and Methods:

Patient material: Plasma samples were obtained from bPI suspected of failing patients' (as referred by the clinician) from the diagnostic section at the Division of Medical Virology, Stellenbosch University, and the South African National Health Laboratory Services (NHLS) that were collected between March 2017 and February 2018. Viral RNA was extracted using the Viral RNA Extraction kit (Qiagen, GMBH) and stored in -80^{O}C. The study was approved by the Health Research Ethics Committee of Stellenbosch University, South Africa (N15/08/071).

PCR amplification and high throughput sequencing:

Reverse transcriptase PCR (RT-PCR), which consists of cDNA synthesis followed by first round PCR, were performed using the SuperScript™ III One-Step RT-PCR System with Platinum™ Taq DNA Polymerase (Invitrogen/Life Technology, Cat No. 12574026) using the primers 1810F (5'-GCTACACTAGAAGAAATGATGACAGCATG-3') and 5220R (5'-CCCTAGTGGGATGTGTACTTCTGA-3'). The second round nested PCR was performed with 2001F (5'-TGCAGGGCCCCTAGGAAAAAGGGCTGTT-3') and 5087R (5'-ATCCTGTCTACYTGCCACACAAYC-3') primers using KAPA HiFi HotStart ReadyMix PCR kit. The amplified products were purified using the QIAamp gel extraction kit (Qiagen, Germany). For HTS, the purified amplicons were fragmented, and the library was prepared using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolab, USA) with multiplexed NEB next adaptors. The samples were then pooled together with other unrelated non-viral indexed libraries. Paired end sequences of read length 250 bp were carried out on the Illumina HiSeq2500. The sequences are available in SRA (Submission ID : SUB5871663).

Bioinformatics analysis:

The raw reads were adapter trimmed using TrimGalore version 0.6.2, followed by the removal of the low-quality bases (phred value score <Q30) by Sickle version 1.33. Duplicate reads were removed using FastUniq. The *de novo* assembly was performed using the Iterative Virus Assembler (IVA). The processed reads were aligned against individual *pol* gene sequence in very sensitive local mode using Bowtie2, in order to select reads originated from *pol* gene and create a consensus gene. The subtyping was performed using REGA version 3. The selected reads were then aligned against Pol protein sequences using the blastx program from BLAST package. Best blastx hit was chosen for each read for the amino acid counting which was performed by in-house script. The resistance was interpreted as per the mutation lists provided in Stanford University HIV Drug Resistance Database, accessed on 6 Jan 2019 [16]. The complete script is available in github: https://github.com/neogilab/MiDRMPol_SouthAfrica.

Results:

Among the 67 samples sequenced, current treatment regimen data were not available for 11 samples, thus they were excluded from further analyses. Among the 56 patients 5.3% (n=3) were on boosted ATV, while only one patient was on DRV and the rest 93% (n=52) were receiving LPV/r. The median (range) viral load was 71814 (937-5500000) RNA copies/mL. HIV-1 subtyping identified 55 samples as HIV-1 subtype C and one as CRF02_AG. The major PI, NRTI, NNRTI and INI RAMs were observed 27% (15/56), 57% (32/56), 50% (28/56) and 7% (4/56) (Figure 1a). Statistically significant ($p<0.001$) higher PI RAMs were observed in minor viral quasispecies (25%; 14/56), compared to NRTI (11%; 6/56), NNRTI (9%; 5/56) and INI RAM (4%; 2/56). 41% (23/56) did not have any PI RAMs. The complete mutation profile is presented in supplementary table 1. There is no statistical difference in viral load (log_{10} copies/mL) in patients who only had PI DRM in minor population variants, compared to patients who had only DRM in the major viral population [Mean (SD): 4.92 (0.74) vs 4.66 (0.88); p=0.43]. All 56 patients harbored at least one DRM. Most of the DRM in the minor viral quasispecies were observed in V82A mutation (n=13) in protease, K65R (n=5), K103N (n=7) and M184V (n=5) in reverse transcriptase (Figure 1b). Despite no mention of use of any INIs by the clinical reports, three patient sequences had the Y143R mutation in major viral quasispecies and one in minor viral quasispecies, which confers resistance to raltegravir (RAL). However, resistance to INI inhibitors were low in the settings. The predicted resistance pattern (as per the Stanford HIVDB) was given in Figure 1c. Half (28/56) of the patients had doravirine cross-resistance. There were two patients (ZA94 and ZA97) who are resistant to all classes of drugs indicates the presence of extremely drug resistance HIV-1 strains in South Africa.

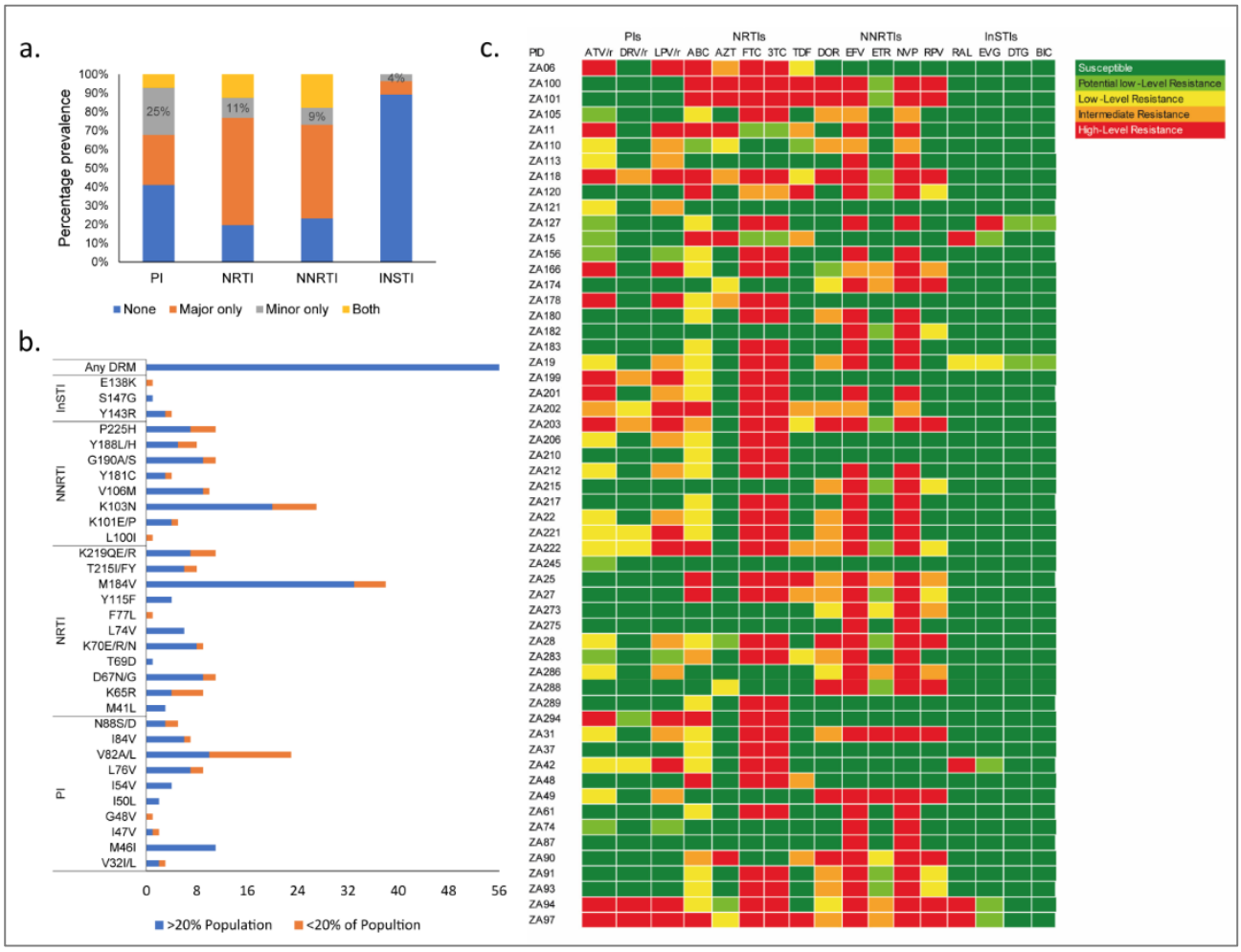


Figure 4.1. Percentage prevalence, number of different RAMs and predicted level of resistance

(a) Percentage prevalence of PI, NRTI, NNRTI and INI RAMs in minor (<20% of the populations) and major (≥20% of the populations) viral populations alone and together. The RAM in minor viral quasispecies are indicated. (b) Number of different RAMs in minor (<20% of the populations) and major (≥20% of the populations) viral populations (c) Predicted level of resistance to different antiretroviral drugs based on Stanford HIV database.

Discussion:

In this study we have used high throughput sequencing to type the DRM in both minor (<20% of the population) and major (>20% of the population) viral quasispecies and identified increased PI RAM in minor viral populations. Our study also indicated very lower level of transmitted INI RAMs in patients failing on bPIs based regimen.

Earlier studies have indicated that PI RAMs were uncommon in patients failing on second line cART, with only 7% patients on bPIs showing PI RAMs.[17] A recent report using Sanger Sequencing, indicates that 35% of the patients had PI DRMs failed on bPIs.[18] Our study also showed that 34% of the patients had major PI RAMs in major viral population. In addition, 25% of the patients had PI

DRM in the minor viral population. All the patients had at least one DRM. As most of the patients had high viral loads, there is a high chance of transmission of DRM if not treated timely. Furthermore, the PI RAMs in minor viral population can evolve to become majority under drug selection pressure.

Earlier studies indicated that HIV-1C has reduced susceptibility towards the PIs without emergence of major PI RAM [19] and there is higher risk of virological failure.[9] Gag mutations can also confer reduced susceptibility towards PIs.[20] Clinical studies also reported that the second line failure was frequent in South African settings,[14] this could further increase chance of transmission of primary DRMs.

Even though there is no indication of the patients being administered RAL, three of the patients in our study harbored Y143R mutation in >20% of the population. Previous studies in HIV-1C have shown major INI mutations at baseline in less than 5% of the patients from Ethiopia (T66I, E138K, Q148R, and Q148H) and South Africa (Q148H, T66S, E92G, S147G, T66A, Y143YF and Y143H). [3,12] However, presence of INI RAMs in minor viral populations are deemed not have any clinical consequences.[21]

In conclusion, this study shows that use of high throughput resistance testing for GRT can greatly improve the identification of PI RAMs in bPI failing patients. Using HTS-GRT, PI RAMs (V82A) and RTI RAMs (K65R, M184V or K103N) were identified in <20% population that Sanger-based sequencing failed to identify strengthening their role in detecting the acquired mutations early.

Acknowledgements

This study was funded by the National Research Foundation (NRF) of South Africa, the Poliomyelitis Research Foundation (PRF) of South Africa and the National Health Laboratory Services (NHLS) Research Trust. We also thank Stellenbosch University and the Harry Crossley Foundation for additional financial support. UN is supported by a grant from the Swedish Research Council (2017-01330). GBJ is supported through the South African new Generation of Academics Programme (nGAP). We also acknowledge Mathilda Claassen for sample collection.

Potential conflicts of interest. All authors: no conflicts.

References:

1. Ji H, Enns E, Brumme CJ et al. Bioinformatic data processing pipelines in support of next-generation sequencing-based HIV drug resistance testing: the Winnipeg Consensus. *Journal of the International AIDS Society* 2018; **21**: e25193.

2. Derache A, Iwuji CC, Baisley K et al. Impact of next generation sequencing defined HIV pre-treatment drug resistance on virological outcomes in the ANRS 12249 treatment as prevention trial. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2018.

3. Telele NF, Kalu AW, Gebre-Selassie S et al. Pretreatment drug resistance in a large countrywide Ethiopian HIV-1C cohort: a comparison of Sanger and high-throughput sequencing. *Sci Rep* 2018; **8**: 7556.

4. Paredes R, Lalama CM, Ribaudo HJ et al. Pre-existing minority drug-resistant HIV-1 variants, adherence, and risk of antiretroviral treatment failure. *The Journal of infectious diseases* 2010; **201**: 662-71.

5. Li JZ, Paredes R, Ribaudo HJ et al. Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. *JAMA : the journal of the American Medical Association* 2011; **305**: 1327-35.

6. Metzner KJ, Scherrer AU, von Wyl V et al. Limited clinical benefit of minority K103N and Y181C-variant detection in addition to routine genotypic resistance testing in antiretroviral therapy-naive patients. *AIDS (London, England)* 2014; **28**: 2231-9.

7. Inzaule SC, Hamers RL, Noguera-Julian M et al. Clinically relevant thresholds for ultrasensitive HIV drug resistance testing: a multi-country nested case-control study. *Lancet HIV* 2018; **5**: e638-e46.

8. Meintjes G, Moorhouse MA, Carmona S et al. Adult antiretroviral therapy guidelines 2017. *South Afr J HIV Med* 2017; **18**: 776.

9. Haggblom A, Svedhem V, Singh K et al. Virological failure in patients with HIV-1 subtype C receiving antiretroviral therapy: an analysis of a prospective national cohort in Sweden. *Lancet HIV* 2016; **3**: e166-74.

10. Sutherland KA, Collier DA, Claiborne DT et al. Wide variation in susceptibility of transmitted/founder HIV-1 subtype C Isolates to protease inhibitors and association with in vitro replication efficiency. *Sci Rep* 2016; **6**: 38153.

11. Ajose O, Mookerjee S, Mills EJ et al. Treatment outcomes of patients on second-line antiretroviral therapy in resource-limited settings: a systematic review and meta-analysis. *AIDS (London, England)* 2012; **26**: 929-38.

12. Brado D, Obasa AE, Ikomey GM et al. Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and available of integrase inhibitors in Cape Town, South Africa. *Sci Rep* 2018; **8**: 4709.

13. Moorhouse M, Maartens G, Venter WDF et al. Third-Line Antiretroviral Therapy Program in the South African Public Sector: Cohort Description and Virological Outcomes. *Journal of acquired immune deficiency syndromes (1999)* 2019; **80**: 73-8.

14. Collier D, Iwuji C, Derache A et al. Virological Outcomes of Second-line Protease Inhibitor-Based Treatment for Human Immunodeficiency Virus Type 1 in a High-Prevalence Rural South African Setting: A Competing-Risks Prospective Cohort Analysis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2017; **64**: 1006-16.

15. Fisher R, van Zyl GU, Travers SA et al. Deep sequencing reveals minor protease resistance mutations in patients failing a protease inhibitor regimen. *Journal of virology* 2012; **86**: 6231-7.

16. Rhee SY, Gonzales MJ, Kantor R et al. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic acids research* 2003; **31**: 298-303.

17. Wallis CL, Mellors JW, Venter WD et al. Protease Inhibitor Resistance Is Uncommon in HIV-1 Subtype C Infected Patients on Failing Second-Line Lopinavir/r-Containing Antiretroviral Therapy in South Africa. *AIDS research and treatment* 2011; **2011**: 769627.

18. Pillay M, Khan A, Govender K et al. HIV drug resistance in adults failing protease inhibitor (PI)-based antiretroviral therapy (ART) in KwaZulu-Natal. *27th International Workshop on HIV Drug Resistance and Treatment Strategies*, 2018.

19. Sutherland KA, Parry CM, McCormick A et al. Evidence for Reduced Drug Susceptibility without Emergence of Major Protease Mutations following Protease Inhibitor Monotherapy Failure in the SARA Trial. *PloS one* 2015; **10**: e0137834.

20. van Domselaar R, Njenda DT, Rao R et al. HIV-1 Subtype C with PYxE Insertion Has Enhanced Binding of Gag-p6 to Host Cell Protein ALIX and Increased Replication Fitness. *Journal of virology* 2019; **93**.

21. Nguyen T, Fofana DB, Le MP et al. Prevalence and clinical impact of minority resistant variants in patients failing an integrase inhibitor-based regimen by ultra-deep sequencing. *The Journal of antimicrobial chemotherapy* 2018; **73**: 2485-92.

Chapter 5 Structural Implications of Genotypic Variations in HIV-1 Integrase from Diverse Subtypes

5.1. Article title

Structural Implications of Genotypic Variations in HIV-1 Integrase from Diverse Subtypes.

5.2. Authors and citations

Rogers L, **Obasa AE**, Jacobs GB, Sarafianos SG, Sönnerborg A, Neogi U, Singh K.

Number of citations: 3

Front Microbiol. 2018 Aug 2;9:1754. doi: 10.3389/fmicb.2018.01754. eCollection 2018. PMID: 30116231.

5.3. Author's contribution

In the enclosed manuscript; I confirmed that I am a co-author. I performed the laboratory based experiments (Viral RNA extractions, PCRs, Sanger sequencing and initial sequence analyses) for the Tygerberg Virology (TV) samples. I confirmed the sequences information (patient's demographics, raw data and aligned sequences) for further analyses to Prof. Kamalendra Singh. Prof Singh and I were both in frequent communication regarding the Integrase polymorphism structural analyses in the context of South African HIV-1C.

5.4. Background

Integrase (IN) integrates viral DNA into the host genome using its 3'-end processing and strand-transfer activities. As InSTIs are becoming increasingly accessible worldwide, it is important to understand the mechanism(s) of InSTI susceptibility. There is strong evidence suggesting differences in the patterns and mechanisms of drug resistance between HIV-1 subtype B, which dominates in United States, Western Europe and Australia, and non-B infections that are most prevalent in countries of Africa and Asia. Polymorphisms (PM) are defined as the change in genomic sequence variation that is common in more than 1% of the sequences. IN PM and other genetic differences among diverse subtypes are likely responsible for these different patterns, but lack of a full-length high-resolution structure of HIV-1 IN has been a roadblock in understanding the molecular mechanisms of InSTI resistance and the impact of polymorphisms on therapy outcome. Here we use molecular modelling to explore the structural impact of IN polymorphisms on the IN reaction mechanism and InSTI susceptibility.

5.5. Main findings

Among all non-B subtype sequences, 17 naturally occurring IN PM positions with 18 PMs were observed. Three of these (K14R, D25E, and V31I) belong to the N-terminal domain (NTD), whereas M50I belongs to the loop region connecting the NTD and C-terminal domain (CTD). Eight PMs (I72V, L74M/I, F100Y, L101I, T124A, K136Q, and D167E) are part of the catalytic core domain (CCD), and the remaining six (V201I, T218I, L234I, A265V, R269K, and S283G) belong to the CTD. Eight PMs (I72V, L74M/I, F100Y, L101I, T124A, K136Q, and D167E) are part of the CCD, and the remaining six (V201I, T218I, L234I, A265V, R269K, and S283G) belong to the CTD. The other two were in regions for which structural data are unavailable. IN PM T218I lies within the missing region (residues 205–222).

5.6. Study significance

We have presented extensive analyses of IN PMs in the structural context. This structure provides the first glimpse of nucleoprotein organisation, which could be used to deduce the effect of a PM. Our analyses suggest that there are several naturally occurring polymorphisms that may affect the structural stabilities of the IN, vDNA binding, and drug binding propensity. The cryoEM structure of the HIV-1 strand transfer complex (STC) is a milestone in the structural biology of integrase. To elucidate the IN structural aspects, this study model the IN nucleoprotein complexes from different subtypes in combination with sequence analyses.

5.7. Conclusion

In this study, we conducted deep analyses with the IN Polymorphisms (PMs) and our data shows how these natural occurring PMs may affect the structural stabilities of the IN and vDNA binding. Furthermore, data generated from this study can provide necessary guidance for investigating natural occurring polymorphisms.

5.8. Open access

The article is open access and freely available online. Reprints can also be obtained from the author upon request: obasa@sun.ac.za. PubMed link: https://www.ncbi.nlm.nih.gov/pubmed/30116231

5.9. Published manuscript

frontiers in Microbiology

ORIGINAL RESEARCH
published: 02 August 2018
doi: 10.3389/fmicb.2018.01754

Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes

Leonard Rogers[1], Adetayo E. Obasa[2,3], Graeme B. Jacobs[2], Stefan G. Sarafianos[4], Anders Sönnerborg[3,5], Ujjwal Neogi[3] and Kamalendra Singh[1,3,6]*

[1] Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO, United States, [2] Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa, [3] Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, Stockholm, Sweden, [4] Laboratory of Biochemical Pharmacology, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, United States, [5] Division of Infectious Diseases, Department of Medicine Huddinge, Karolinska Institute, Stockholm, Sweden, [6] Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, MO, United States

Human immunodeficiency virus type 1 (HIV-1) integrase (IN) integrates viral DNA into the host genome using its 3′-end processing and strand-transfer activities. Due to the importance of HIV-1 IN, it is targeted by the newest class of approved drugs known as integrase strand transfer inhibitors (INSTIs). INSTIs are efficient in maintaining low viral load; however, as with other approved antivirals, resistance mutations emerge in patients receiving INSTI-containing therapy. As INSTIs are becoming increasingly accessible worldwide, it is important to understand the mechanism(s) of INSTI susceptibility. There is strong evidence suggesting differences in the patterns and mechanisms of drug resistance between HIV-1 subtype B, which dominates in United States, Western Europe and Australia, and non-B infections that are most prevalent in countries of Africa and Asia. IN polymorphisms and other genetic differences among diverse subtypes are likely responsible for these different patterns, but lack of a full-length high-resolution structure of HIV-1 IN has been a roadblock in understanding the molecular mechanisms of INSTI resistance and the impact of polymorphisms on therapy outcome. A recently reported full-length medium-resolution cryoEM structure of HIV-1 IN provides insights into understanding the mechanism of integrase function and the impact of genetic variation on the effectiveness of INSTIs. Here we use molecular modeling to explore the structural impact of IN polymorphisms on the IN reaction mechanism and INSTI susceptibility.

Keywords: HIV-1 integrase, strand transfer inhibitor, polymorphism, drug-resistance, HIV-1 subtypes

OPEN ACCESS

Edited by:
Akio Adachi,
Kansai Medical University, Japan

Reviewed by:
Takao Masuda,
Tokyo Medical and Dental University, Japan
Tino Sanchez,
Loma Linda University, United States
Nanjie Deng,
Pace University, United States

****Correspondence:***
Ujjwal Neogi
ujjwal.neogi@ki.se
Kamalendra Singh
singhka@missouri.edu

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *28 May 2018*
Accepted: *13 July 2018*
Published: *02 August 2018*

Citation:
Rogers L, Obasa AE, Jacobs GB, Sarafianos SG, Sönnerborg A, Neogi U and Singh K (2018) Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. Front. Microbiol. 9:1754. doi: 10.3389/fmicb.2018.01754

INTRODUCTION

Combination antiretroviral therapy (cART) targets several steps of viral replication. In many cases, cART can suppress viral load below the detection level and make HIV infection a chronic yet manageable disease with a near-normal life expectancy (Antiretroviral Therapy Cohort Collaboration, 2017). However, use of cART has been constantly challenged by the emergence of both acquired and transmitted drug resistance mutations (DRMs). In addition, anti-HIV drugs have associated toxicity and bioavailability issues, although to different extents. These challenges

have spurred the development of new antiretrovirals that have a high genetic barrier to resistance and low toxicity and that are effective against resistant viruses. Integrase strand transfer inhibitors (INSTIs) are the newest class of approved anti-HIV drugs. As the name implies, INSTIs inhibit HIV-1 integrase (IN), which is one of the three enzymes encoded by the *pol* gene.

Retrovirus INs have two major catalytic activities: a 3′-end processing (3′EP) that excises a dinucleotide at the 3′-end, and a strand transfer (ST) activity that integrates HIV DNA into the host chromosome. Both activities are conducted by the same active site (Craigie, 2001, 2012; Chiu and Davies, 2004). In addition, two more IN associated activities have been reported. These are (i) specific endonucleolytic cleavage at the terminal sequences of each LTR, and (ii) disintegration, which can be considered as the reverse of the ST reaction (Thierry et al., 2016).

HIV-1 IN, a 32 kDa protein, functions as a tetramer and/or higher-order oligomer (Craigie, 2012; Craigie and Bushman, 2012). HIV-1 IN has three distinct domains: the N-terminal domain (NTD) (residues 1–46), the catalytic core domain (CCD) (residues 56–186), and the C-terminal domain (CTD) (195–288). The active site (D64, D116, and E152) resides in the CCD. The NTD contains a conserved Zn^{2+}-binding motif (His-His-Cys-Cys). The CTD is basic in nature and adopts an SH3-fold. The CTD has been implicated in DNA binding and oligomerization of HIV-1 IN (Kessl et al., 2009). The high resolution structure of full-length HIV-1 IN is not known (Craigie, 2012), although structures of IN domains (NTD/CCD and CCD/CTD) have been determined (Cai et al., 1997; Chen et al., 2000; Wang et al., 2001).

Currently used first-line cART includes an INSTI in the backbone of two nucleoside reverse transcriptase inhibitors (NRTIs). At present, four INSTIs have been approved. These are raltegravir (RAL), elvitegravir (EVG), dolutegravir (DTG), and bictegravir (BIC). Of these, RAL and EVG are referred to as first-generation, whereas DTG and BIC are second-generation INSTIs (Anstett et al., 2017). Another INSTI, cabotegravir (CAB), is in the late phases of clinical trials, and it has shown great potential, especially as a long-acting (LA) antiviral (Whitfield et al., 2016). Resistance pathways to first-generation INSTIs and DTG are well established. However, the same for CAB and BIC are not well understood. Molecular details of resistance to INSTIs are also not well known, mostly due to the lack of a high-resolution full-length structure of HIV-1 IN.

Recently, great strides have been made in the structural biology of INs. CryoEM structures of the HIV-1 intasome were reported in early 2017 (Passos et al., 2017). These structures, together with a 3.8 Å resolution X-ray crystal structure of the Rous sarcoma virus (RSV) intasome (Yin et al., 2016), the cryoEM structures of the prototype foamy virus (PFV) intasome (Ballandras-Colas et al., 2017), and that of the mouse mammary tumour virus (MMTV) intasome (Ballandras-Colas et al., 2016), provide significant insights into the assembly of retroviral IN nucleoprotein complexes at different stages of the ST reaction (Engelman and Cherepanov, 2017). These structures reveal many common features of the IN mechanism of action across retroviruses (Engelman and Cherepanov, 2017). In addition, the crystal structures of PFV INs in complex with DNA and DNA/INSTI have been solved (Maertens et al., 2010; Hare et al., 2010a, 2012). The crystal structures of PFV IN in complex with DNA and INSTIs have been serving as the model systems for understanding INSTI resistance mechanisms.

Limited solubility of HIV-1 IN in buffers containing isotonic salt concentrations has hindered structure determination of full-length HIV-1 IN. To overcome this limitation, Li et al. (2014) generated a chimera containing full-length HIV-1 IN fused with *Sulfolobus solfataricus* protein Sso7d that greatly enhanced the solubility and activity of the HIV-1 IN. This construct facilitated the solution of a cryoEM structure of the HIV-1 IN strand transfer complex (STC) intasome (Passos et al., 2017). Despite some limitations (discussed later), this structure provided the first glimpse of nucleoprotein organization that could be used to deduce the effect of a polymorphism (PM, defined as the change in genomic sequence variation that is common in more than 1% of the sequences) in the IN of different HIV-1 subtypes at the atomic level. Here we present the identification of IN PMs from diverse HIV-1 subtypes and assess the impact of these PMs in the structure-function mapping of HIV-1 IN.

MATERIALS AND METHODS

Identification of PMs From Diverse HIV-1 Subtypes

HIV-1 integrase sequences (n = 8114) of viruses isolated from individual patients were downloaded from the HIV-1 Stanford Database (Rhee et al., 2003). After excluding the low quality and shorter sequences, we included HIV-1 A1/A2 (n = 483), HIV-1B (n = 4379), HIV-1C (n = 1155), CRF01_AE (n = 1581), and CRF02_AG (n = 522) sequences. These five subtypes and Circulating Recombinant Forms represent >90% of global infections (Hemelaar, 2012). Multiple sequence alignment was performed using ClustalX (Larkin et al., 2007) against the HIV-1 HXB2 sequence (the reference sequence). Variant calling of each residue was performed using an in-house R script (Krzywinski et al., 2009). We also used the IN sequences (n = 91) from the Tygerberg Virology (TV) cohort (Jacobs et al., 2009). This cohort contains treatment-naïve patients from a variety of ethnic groups and sexual orientations. The patient samples of the TV cohort were collected between 2000 and 2001, before the initiation of South Africa's national HIV treatment program. Recently, we amplified and sequenced IN genes from the TV cohort (n = 91) for identification of INSTI DRMs (Brado et al., 2018). Subtype specific consensus sequences were generated using the Consensus Maker[1] tool. Naturally occurring polymorphisms (PMs) were defined as any mutations that were present in >50% of sequences.

Structures of IN Intasomes From Different Subtypes

Homology-derived molecular models of HIV-1 IN tetramers from different subtypes were generated using the cryoEM structure of the HIV-1B IN intasome (PDB file 5U1C) (Passos

[1]https://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html

et al., 2017) as a template, using Prime (version 4.2) through Maestro (Schrodinger, New York, NY, United States) as previously described (Haggblom et al., 2016). The homology models were minimized using Prime and MacroModel to eliminate steric overlaps and to optimize sidechain conformations, respectively (Singh et al., 2012). The cryoEM intasome structure (PDB file 5U1C) is missing residues A205 to N222 (Passos et al., 2017). In addition, residues 1–55 (comprising the entire NTD and the linker region between the NTD and CCD), residues 135–150, and residues 186–195 could not be resolved in the outer IN molecules (definition of outer and inner INs presented below). These loops were constructed with Prime's loop modeling utility (Schrödinger Suite). All modeled structures were submitted to the Structure Analysis and Verification Server (SAVES)[2] as well as the Protein Structure Preparation tool of SYBYL-X (version 2.1). No bad contacts were noted in the final models of INs. The backbone torsion angles were checked by Ramachandran plot for allowed conformations of φ and ψ angles.

RESULTS AND DISCUSSION

IN PMs

Among all non-B subtype sequences, 17 naturally occurring IN PM positions with 18 PMs were observed (**Figure 1A**). Distribution of these PMs in different domain contexts is depicted in **Figure 1B**. Fifteen of these PMs were successfully mapped structurally. Three of these (K14R, D25E, and V31I) belong to the NTD, whereas M50I belongs to the loop region connecting the NTD and CTD. Eight PMs (I72V, L74M/I, F100Y, L101I, T124A, K136Q, and D167E) are part of the CCD, and the remaining six (V201I, T218I, L234I, A265V, R269K, and S283G) belong to the CTD. The other two were in regions for which structural data are unavailable. IN PM T218I lies within the missing region (residues 205–222). While we have modeled this region as a loop, an ambiguity remains in the secondary structure assignment, as this region in the crystal structure of CCD + CTD assumes a helical conformation. A polymorphism at position 283 could not be structurally mapped, since the structure of the C-terminal end of IN (amino acids 270–288) is unavailable.

Structure of HIV-1 Intasome

The tetrameric cryoEM structure represents a post-catalysis synaptic STC of the HIV-1 intasome formed by four IN subunits arranged in two-fold symmetry (Passos et al., 2017). A pair of IN dimers encapsulates the viral/host DNA chimera, in which the two inner molecules directly interact with DNA, while the outer molecules have protein–protein interactions with the inner molecules (**Figure 2**). This arrangement of IN oligomers bound to the viral/host DNA chimera has been referred to as the intasome core structure (Engelman and Cherepanov, 2017). The structure of the NTD in the outer IN molecules is not available in PDB coordinate file 5U1C (Passos et al., 2017). The structure of residues 205–222 is missing in all IN structures, whereas the structures corresponding to residues 135–150 and 186–195 of the outer IN molecules are also unresolved. Residues 135–139 assume a β-strand conformation in the structure of NTD+CCD [PDB file 1K6Y, (Wang et al., 2001)], whereas residues 186–195 exist as a loop structure in the crystal structure of CCD+CTD [PDB file 1EX4, (Chen et al., 2000)]. In an earlier structure of IN CCD and CTD domains, residues 205–222 are part of the helix (PDB file 1EX4). However, due to the topological position of the CCD and CTD in the STC, a loop conformation of this region appears more feasible, which is the conformation in our modeled structures. We set the conformation of residues 135–139 as a β-strand, whereas residues 186–195 were kept in loop conformation in all IN molecules.

PMs in the NTD

Our analyses showed a total of three NTD PM positions in different subtypes. These are K14R, D25E, and V31I. PM M50I is located in the linker region connecting the NTD and CCD. Of these, D25E, V31I, and M50I were also noted in the cohort of South African HIV-1C patients (Brado et al., 2018). PM K14R was most prominent in CRFs 01_AE and 02_AG, followed by subtypes A1/A2. Approximately 33% of HIV-1C sequences had 14R. Amino acid K14 is located on αA-helix. The sidechain of K14 in HIV-1B forms a salt-bridge with the backbone C = O of W131 of the outer IN subunits, suggesting that K14 is involved in tetramer formation (**Figure 3A**). Substitution of K to R results in an additional salt-bridge formation (with C = O of W132). Typical salt-bridge energy ranges between 1 and 4 kCal/mol (Honig and Hubbell, 1984; Wimley et al., 1996), suggesting that R14 may induce a change in interaction energy due to a pair of salt-bridges and reduced compared to one in case of K14. In addition, both K14 and R14 form H-bonds with Y15. Previously, Y15 has been shown to be crucial for the assembly of IN and HIV-1 RT on viral RNA through the RT-IN precursor form (Takahata et al., 2017). In addition to the above mentioned interactions of K/R14, the K14 side chain forms a cation-π interaction with the indole ring of the W131 side chain, which may further stabilize the tetramer. Residue K14R is between zinc-binding residues H12 and H16. Due to the location of these residues on the α-helix, the sidechain of K14/R14 is extended away from the zinc-binding residues, and does not appear to affect the geometry of the zinc-binding motif.

PM D25E was noted exclusively in HIV-1C in the sequences from the HIV database as well as in our TV cohort (Brado et al., 2018). D25 is located at the beginning of αB-helix. D25 forms symmetrical interactions between the inner INs. D25 from one inner IN forms a salt-bridge with K188 of the other inner IN molecule (**Figure 3B**). K188 is one of the three (K186, R187, and K188) positively charged residues in the vicinity of D25. Of these, R187 directly interacts with the phosphate backbone of viral DNA (**Figure 3B**). In the modeled structure of HIV-1C IN, E25 interacts with both K186 and K188 (**Figure 3B**). Additional interactions of E25 may influence DNA binding by IN, which in turn can affect INSTI binding, since the INSTI binding pocket is formed by both DNA and protein molecules. The third PM of the NTD, V31I, exists in ~75–85% of sequences among non-B subtypes (**Figure 1**). In the STC, V31 is close to viral DNA

[2]https://services.mbi.ucla.edu/SAVES/

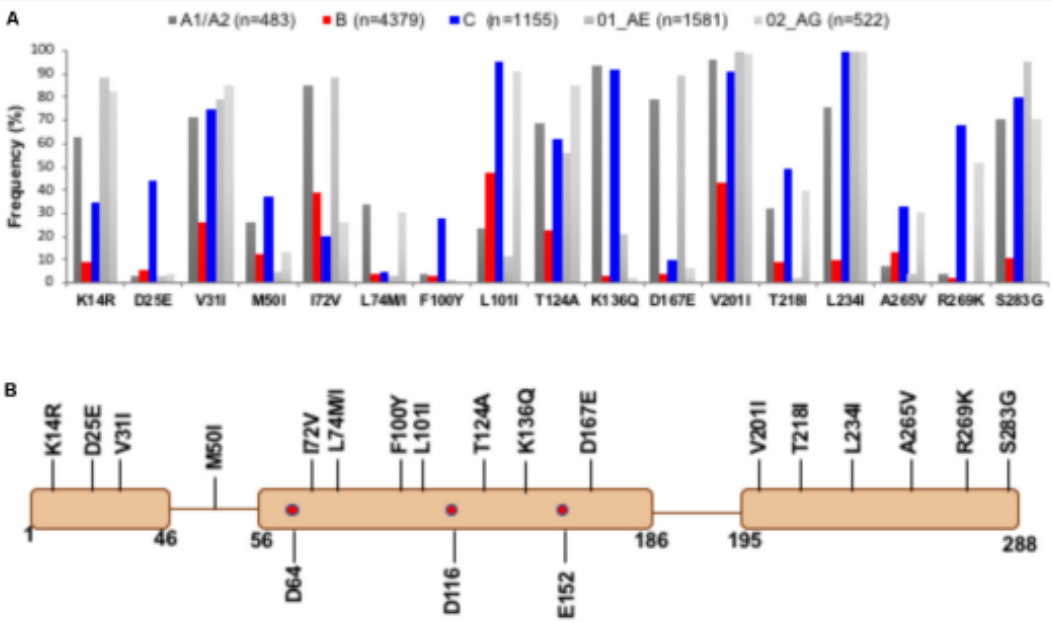


FIGURE 1 | Prevalence of PMs in IN genes from diverse subtypes – Panel **A** shows the quantitative distribution (frequency in percentage) of naturally occurring polymorphisms in HIV-1C (in blue) and A-like (A1/A2, 01_AE, and 02_AG, in shades of gray) which are statistically significantly different compared to HIV-1B (shown in red). Panel **B** shows the positions of PMs in the context of IN domains. The active site residues (D64, D116, and E152) are marked as red dots.

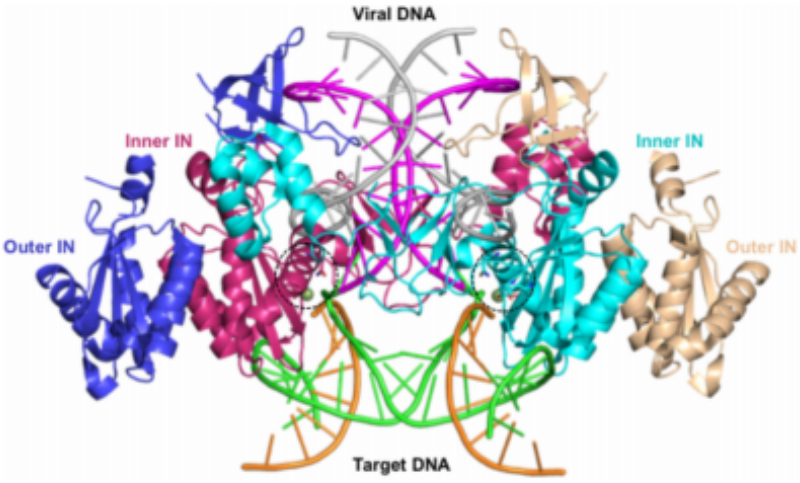


FIGURE 2 | Structure of HIV-1B intasome – This figure shows the structure of the HIV-1 synaptic stable complex (SSC) intasome as determined using cryoEM (PDB file 5U1C). This structure represents the complex after the strand transfer reaction. The integrated viral DNA is rendered as ladders in magenta and gray colors, whereas the host DNA is colored light green and orange. Two IN molecules that directly bind to DNA are shown as ribbons in cyan and maroon. These IN molecules are referred to as inner INs. The other two IN molecules that bind to inner INs are rendered as ribbons in medium blue and wheat colors. These IN subunits are called outer IN molecules.

(**Figure 3C**), although not within interacting distance. The longer sidechain of I31 brings DNA and protein within interacting distance. Hence, the V31I PM can influence binding of DNA with IN in non-B subtypes.

Amino acid M50 is located on the linker region between the NTD and CCD. In the STC, M50 from both inner IN molecules interact with viral DNA through hydrophobic interactions with base moieties (**Figure 3D**). Substitution I50 will enhance the

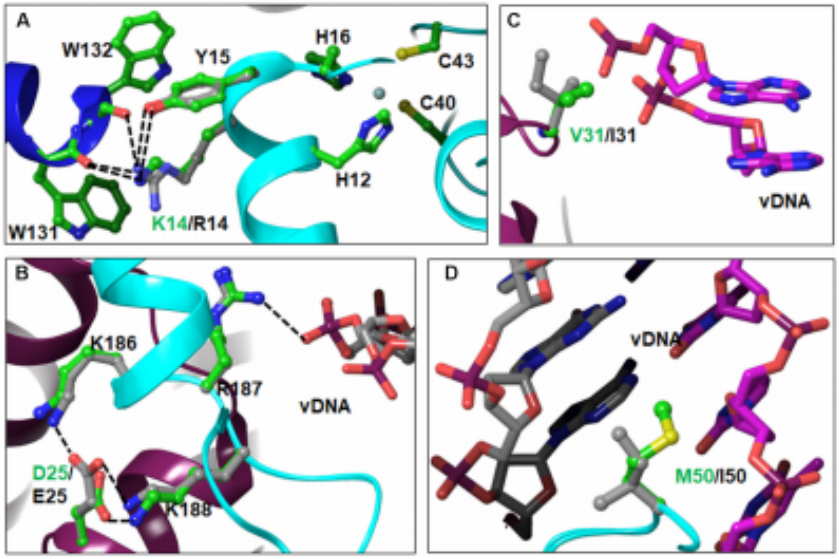


FIGURE 3 | Locations and interactions of NTD PMs – Panels **A–C** show PMs and their interactions in the NTD. The M50I PM belongs to the linker region and is shown in panel **D**. The interactions (hydrogen bonding and/or ion-pairs) are shown as dotted lines. In this and following figures, the amino acid residues are shown in ball-and-stick representation and colored as green carbons for HIV-1B and gray carbons in non-B HIV-1. The other atoms are colored as red (oxygen), blue (nitrogen) and yellow (sulfur). The DNA is shown as sticks. The viral DNA is colored with magenta and gray carbons, and target DNA is colored with green and orange carbons. For reference purposes, we have included the zinc-binding motif in panel **A**.

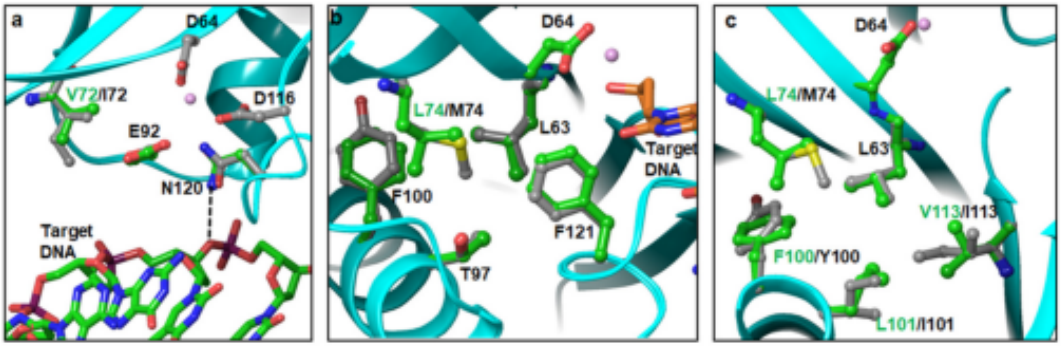


FIGURE 4 | Topological positions of IN PMs near the active site of CCD – Panels **a**, **b**, and **c** show the positions of V72I, L74M and F100Y/L101I, respectively. As a reference point, D64 and metal ion at the active site are also shown.

hydrophobic interaction due to greater hydrophobicity of I compared to M. In breakthrough selections, M50I mutations have been seen to emerge after R263K (Tsiang et al., 2016) and to provide a replication advantage to R263K-containing viruses. While M50I alone did not change the EC_{50} in seven of 24 recombinant viruses in our previous report (Neogi et al., 2018), it is possible that the M50I PM in non-B subtypes may be advantageous for R263K-containing viruses.

PMs in the CCD

A total of eight PMs (I72V, L74M/I, F100Y, L101I, T124A, K136Q, and D167E) in the CCD were identified in our sequence analyses. In the cryoEM intasome structure, position 72 is occupied by valine. It is not surprising, since I72 is highly polymorphic in HIV-1B (**Figure 1**). To evaluate the contribution of isoleucine, we modeled I72 in place of V72 and conducted energy minimization of the structure. Our results showed some rearrangement of neighboring sidechains including E92. Mutations E92Q/A are known DRMs against first-generation INSTIs RAL and EVG (Wensing et al., 2017). E92 is positioned between I/V72 and N120, and does not interact directly with DNA (**Figure 4a**). However, N120 has a direct interaction with the bridging phosphate oxygen of target DNA (shown as dotted line in **Figure 4a**). In the model of I72 containing IN, we noticed that the positions of the sidechain carbonyl and NH_2 groups of N120 are flipped relative to the conformation of N120 in the V72

structure. This flipped position of the C = O requires a water-mediated contact, indicating that the I72 PM may influence emergence of the DRM at E92. To assess the effect of the E92Q/A mutation, we superposed the crystal structure of PFV IN bound to RAL and DNA (PDB file 3OYA) (Hare et al., 2010b) onto the structure of the HIV-1 intasome (PDB file 5U1C) (Passos et al., 2017). This superposition showed that at least two nucleotides at the 3′ end of the 8-nt long target DNA must be displaced to accommodate the oxadiazole moiety of RAL. A second observation that we made was that there are several water molecules in the vicinity of RAL, and three of these water molecules appear to complete the coordination geometry of Mg^{2+} ions. In the superposed structure, these water molecules are within interacting distance of E92. The mutation E92Q/A is expected to disrupt the spatial arrangement of these waters which, in turn, may affect the coordination of Mg^{2+} ions and thereby result in reduced binding of RAL. L74 is part of a hydrophobic cluster including L63, T97, F100/Y100, L101, L113/I113, and F121 near the active site of IN (**Figures 4b,c**). F100Y and L101I are also highly polymorphic in our sequence and in the HIV-1 sequence database. Both T97 and F121 are known INSTI DRM positions (Wensing et al., 2017). In addition, a recent report showed that the L74F mutation increased resistance to second-generation INSTIs (Hachiya et al., 2017). Hence, the impact of PMs in this hydrophobic cluster appears rather complex. In our modeled HIV-1C IN structure, M74 is closer to T97 and F121 than L74 (**Figure 4b**). Previous *in vitro* selection studies have shown that Q148H/R and G140S in combination with mutations L74I/M, E92Q, T97A, E138A/K, G140A, or N155H are associated with 5- to 20-fold reduced DTG susceptibility (Kobayashi et al., 2011). It is possible that polymorphism L74M is related to T97A mutation evolutionarily. F100Y and L101I PMs can impact the core structure and thereby affect the local geometry of the active site. The sequence containing I101 also had I113 in our model of HIV-1C IN. It is possible that two mutations arise simultaneously for preferred replication of the virus.

T124 interacts with target DNA (**Figure 5a**). This is a highly polymorphic position, noted here as well as reported previously (Kobayashi et al., 2011). In a previous report, the T124A mutation alone or in combination with L101I (T124A/L101I) was identified more frequently in RAL failing patients than in INSTI-naïve HIV-1B patients (Malet et al., 2011). In another study, the T124A mutation was highly prevalent in INSTI-naïve and RAL-failed patients, and it was significantly associated with HIV-nonB (Saladini et al., 2012). While the T124A mutation was selected under DTG pressure, it does not significantly affect the efficacy of DTG (Kobayashi et al., 2011; Vavro et al., 2013). These results suggest that the T124A mutation affects the binding of RAL more than the binding of DTG. T124 from the inner subunit is within interacting distance with the phosphate group bridging two nucleotides that base pair with the forth and fifth nucleotides from the 3′-end of the 8-nt long target DNA (**Figure 5a**). This interaction will be lost with T124A. It is possible that this mutation does not affect DTG binding significantly compared to RAL, as RAL is a larger molecule due to the presence of an oxadiazole ring. In the outer subunits, T124 is exposed at the surface. Hence, the implication of T124A on other viral functions such as viral fitness cannot be deduced from available structures. From the modeled structures, the interactions of K136Q are difficult to determine for two reasons: (i) the position of K136 in the outer molecules of 5U1C is not known, and (ii) K136 is near the segment that is missing in the cryoEM structure (205–222). Nonetheless, one can speculate that K136 in the inner molecule interacts with residues in the region 205–222. It should also be pointed out here that residues in the 205–222 region in the crystal structure of CCD+CTD (PDB file 1EX4) (Chen et al., 2000) assume a helical conformation, whereas in the HIV-1 intasome structure (PDB file 5U1C) (Passos et al., 2017), they form an unordered structure that appears close to K126/Q136. Amino acid D167 from one inner IN molecule interacts with K42 of the other inner IN molecule (**Figure 5b**). This is a symmetric interaction similar to that seen for the D25E polymorphism. Due to the longer sidechain of E167 compared to that of D167, there are two interactions with E167, which contributes to additional stability of the intasome tetramer in non-B.

PMs in the CTD

There are six PMs (V201I, T218I, L234I, A265V, R269K, and S283G) in the CTD. Of these six positions, L234I and S283G are highly polymorphic in HIV-nonB. V201I is highly polymorphic in both HIV-1B and non-B (**Figure 1**). Amino acid positions 218 and 283 are missing in structures of IN. Hence, the structural impact of PMs at these positions cannot be deduced unambiguously. V201 is at the interface of inner and outer IN molecules (**Figure 5c**), and it participates in a symmetric interaction. Mutation V201I may increase the buried surface area between two IN molecules by a factor of 265 $Å^2$, which empirically can account for ~15 kCal/mol greater binding energy between the two molecules.

In the structures of IN tetramers bound to DNA, amino acid L234 from the outer IN molecules is adjacent to viral DNA. However, it is not within interacting distance. With mutation L234I, although the sidechain is closer to the DNA, it is still far from any sort of interaction with DNA. In the inner IN molecules, L234 is exposed to the solution and does not participate in either protein–protein interactions or protein–DNA interactions.

Amino acid position 265 resides in the SH3-like fold of CTD. We noticed PM A263V in our sequence analyses. In HIV-1B, A/V265 is part of a hydrophobic cluster constructed by V225, V249, V260, A265, and I267 (**Figure 5e**). A critical amino acid residue with respect to INSTI resistance (R263) is part of the loop connecting V260 and A265 (**Figure 5e**). The sidechain of R263 interacts with the phosphate backbone of viral DNA. Mutation A265V is expected to change the geometry of the hydrophobic cluster, which may affect the interaction of R263 with viral DNA. In fact, our modeled structure of HIV-1C IN shows a slight change in the conformation of R263K, resulting in altered distances of NH1 and NH2 atoms from the phosphate groups of viral DNA.

In the cryoEM structure, K269 is the last residue that could be resolved. Although our sequence analyses show very low R269K PM in HIV-1B, the cryoEM structure contains K269, and it does not interact with the viral DNA. Hence, we modeled

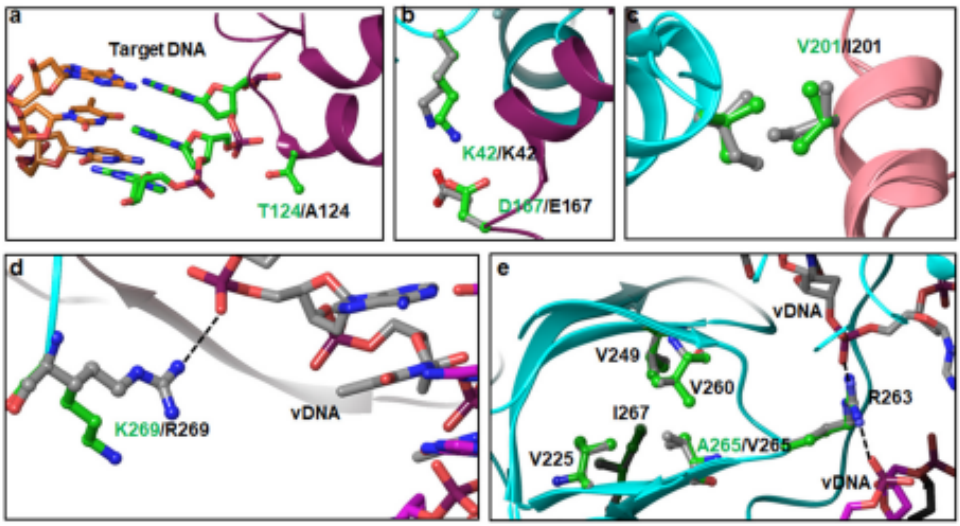


FIGURE 5 | Interactions of PMs in CCD and CTD domains – Panels **a** and **b** show the position and interactions of T124A and D167E PMs of the CCD. Panels **c** and **d** show the interactions of V201I and R269K PMs of the CTD. Panels **d** and **e** show the positions of K269R and A265V PMs, respectively.

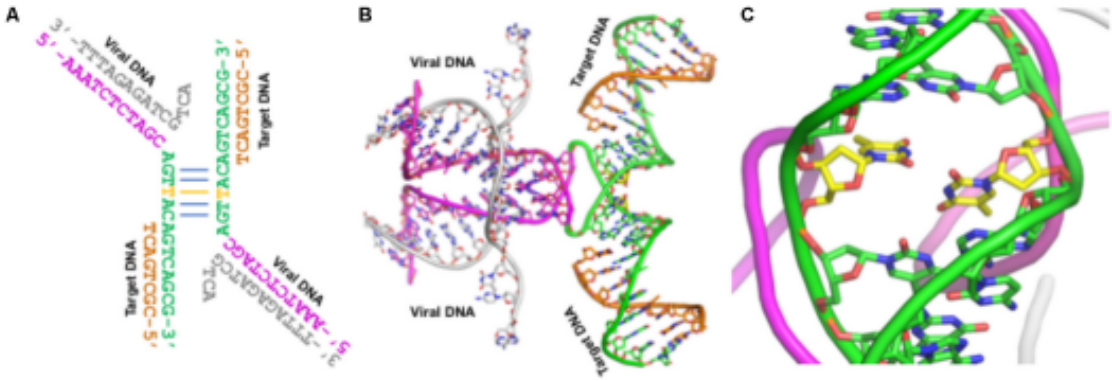


FIGURE 6 | The conformation and sequence of DNA used in the cryoEM structure of IN – Panel **A** shows the viral (gray and magenta) DNA and target (green and orange) DNA. The base-paired region is shown in the middle of this panel. The T–T mismatch in this and in panels **B** and **C** is shown with yellow carbons. The overall DNA conformation is shown in panel **B**. This DNA conformation represents the structure after the strand transfer reaction. Panel **C** shows an enlarged segment from panel **B**.

R269 in place of K269. The results shown in **Figure 5d** indicate that R269 interacts with the phosphate backbone of the viral DNA molecule (shown as dotted line). While this interaction is not seen with K269, there are two confounding factors associated with this phenomenon. First, since it is the last residue in the solved structure, the conformation of the sidechain cannot be unambiguously deduced. The second factor is that the lysine sidechain is quite flexible, leaving the possibility open that K269 will adopt conformations which can interact with DNA.

The cryoEM structure of the HIV-1 STC is a milestone in the structural biology of integrase. By modeling the IN nucleoprotein complexes from different subtypes in combination with sequence analyses, we have elucidated structural aspects and potential functional impacts of IN PMs. However, our analyses and the conclusions drawn from these analyses should be considered with caution, since there are some unusual features associated with the cryoEM structure of the IN intasome. First, the structure was solved as a fusion protein with Sso7. Whether Sso7 fusion has affected any domain rearrangement of IN remains unknown. Second, the IN construct that was used in structure determination contains an active site mutation (E152Q), although it is highly unlikely that this mutation may have affected the overall structure of the IN/DNA complex. The third and most important feature is that the DNA used in this structure determination contains a T–T mismatch in the double-stranded region (**Figure 6**). This mismatch (shown in yellow carbons

in **Figure 6**) may have affected DNA bending, which in turn would have affected spatial oligomerization of IN molecules in the intasome.

In summary, we have presented extensive analyses of IN PMs in the structural context. Our analyses suggest that there are several naturally occurring polymorphisms that may affect the structural stabilities of the IN and vDNA binding, and drug binding propensity. Future biochemical and virological experiments will provide deeper insights into the functional impacts of sequence variations among IN genes from different subtypes. Importantly, these studies will also provide guidance for investigating how naturally occurring polymorphisms can affect treatment response in large real-life cohorts.

AUTHOR CONTRIBUTIONS

KS and UN conceived and designed the study. UN conducted sequence analyses. LR and KS conducted structural analyses. GJ and AO provided sequence information on TV cohort. KS, and UN wrote the first draft of the manuscript reviewed by GJ, AO, SS, AS, and LR. All the authors approved the final version of the manuscript.

FUNDING

The study is partially funded by the Swedish Research Council Establishment grant (2017-01330) and Jeanssons Stiftelser (JS2016–0185) to UN, the Swedish Research Council grant 2016-01675 to AS and ALF-grant from Stockholm County Council (ALF 20160074) to UN and AS. Part of this research was supported by National Institute of Health grant GM118012 (SS). KS acknowledges support from NIH CTSA grant UL1 TR002345.

ACKNOWLEDGMENTS

The authors acknowledge the support received from Anoop T. Ambikan to perform statistical analysis and SNP calling. They also acknowledge comments on manuscript by Dr. Donald Burke.

REFERENCES

Anstett, K., Brenner, B., Mesplede, T., and Wainberg, M. A. (2017). HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology* 14:36. doi: 10.1186/s12977-017-0360-7

Antiretroviral Therapy Cohort Collaboration (2017). Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV* 4, e349–e356. doi: 10.1016/S2352-3018(17)30066-8

Ballandras-Colas, A., Brown, M., Cook, N. J., Dewdney, T. G., Demeler, B., Cherepanov, P., et al. (2016). Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* 530, 358–361. doi: 10.1038/nature16955

Ballandras-Colas, A., Maskell, D. P., Serrao, E., Locke, J., Swuec, P., Jonsson, S. R., et al. (2017). A supramolecular assembly mediates lentiviral DNA integration. *Science* 355, 93–95. doi: 10.1126/science.aah7002

Brado, D., Obasa, A. E., Ikomey, G. M., Cloete, R., Singh, K., Engelbrecht, S., et al. (2018). Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and available of integrase inhibitors in Cape Town, South Africa. *Sci. Rep.* 8, 4709–4718. doi: 10.1038/s41598-018-22914-5

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1997). Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* 4, 567–577. doi: 10.1038/nsb0797-567

Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., et al. (2000). Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8233–8238. doi: 10.1073/pnas.150220297

Chiu, T. K., and Davies, D. R. (2004). Structure and function of HIV-1 integrase. *Curr. Top. Med. Chem.* 4, 965–977. doi: 10.2174/1568026043388547

Craigie, R. (2001). HIV integrase, a brief overview from chemistry to therapeutics. *J. Biol. Chem.* 276, 23213–23216. doi: 10.1074/jbc.R100027200

Craigie, R. (2012). The molecular biology of HIV integrase. *Future Virol.* 7, 679–686. doi: 10.2217/fvl.12.56

Craigie, R., and Bushman, F. D. (2012). HIV DNA integration. *Cold Spring Harb. Perspect. Med.* 2:a006890. doi: 10.1101/cshperspect.a006890

Engelman, A. N., and Cherepanov, P. (2017). Retroviral intasomes arising. *Curr. Opin. Struct. Biol.* 47, 23–29. doi: 10.1016/j.sbi.2017.04.005

Hachiya, A., Kirby, K. A., Ido, Y., Shigemi, U., Matsuda, M., Okazaki, R., et al. (2017). Impact of HIV-1 integrase L74F and V75I mutations in a clinical isolate on resistance to second-generation integrase strand transfer inhibitors. *Antimicrob. Agents Chemother.* 61, e315–e317. doi: 10.1128/AAC.00315-17

Haggblom, A., Swedhem, V., Singh, K., Sonnerborg, A., and Neogi, U. (2016). Virological failure in patients with HIV-1 subtype C receiving antiretroviral therapy: an analysis of a prospective national cohort in Sweden. *Lancet HIV* 3, e166–e174. doi: 10.1016/S2352-3018(16)00023-0

Hare, S., Gupta, S. S., Valkov, E., Engelman, A., and Cherepanov, P. (2010a). Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464, 232–236. doi: 10.1038/nature08784

Hare, S., Vos, A. M., Clayton, R. F., Thuring, J. W., Cummings, M. D., and Cherepanov, P. (2010b). Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20057–20062. doi: 10.1073/pnas.1010246107

Hare, S., Maertens, G. N., and Cherepanov, P. (2012). 3′-processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J.* 31, 3020–3028. doi: 10.1038/emboj.2012.118

Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18, 182–192. doi: 10.1016/j.molmed.2011.12.001

Honig, B. H., and Hubbell, W. L. (1984). Stability of "salt bridges" in membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* 81, 5412–5416. doi: 10.1073/pnas.81.17.5412

Jacobs, G. B., Loxton, A. G., Laten, A., Robson, B., Van Rensburg, E. J., and Engelbrecht, S. (2009). Emergence and diversity of different HIV-1 subtypes in South Africa, 2000-2001. *J. Med. Virol.* 81, 1852–1859. doi: 10.1002/jmv.21609

Kessl, J. J., Mckee, C. J., Eidahl, J. O., Shkriabai, N., Katz, A., and Kvaratskhelia, M. (2009). HIV-1 Integrase-DNA Recognition Mechanisms. *Viruses* 1, 713–736. doi: 10.3390/v1030713

Kobayashi, M., Yoshinaga, T., Seki, T., Wakasa-Morimoto, C., Brown, K. W., Ferris, R., et al. (2011). In vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrob. Agents Chemother.* 55, 813–821. doi: 10.1128/AAC.01209-10

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., Mcwilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Li, M., Jurado, K. A., Lin, S., Engelman, A., and Craigie, R. (2014). Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLoS One* 9:e105078. doi: 10.1371/journal.pone.0105078

Maertens, G. N., Hare, S., and Cherepanov, P. (2010). The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* 468, 326–329. doi: 10.1038/nature09517

Malet, I., Wirden, M., Fourati, S., Armenia, D., Masquelier, B., Fabeni, L., et al. (2011). Prevalence of resistance mutations related to integrase inhibitor S/GSK1349572 in HIV-1 subtype B raltegravir-naive and -treated patients. *J. Antimicrob. Chemother.* 66, 1481–1483. doi: 10.1093/jac/dkr152

Neogi, U., Singh, K., Aralaguppe, S. G., Rogers, L. C., Njenda, D. T., Sarafianos, S. G., et al. (2018). Ex-vivo antiretroviral potency of newer integrase strand transfer inhibitors cabotegravir and bictegravir in HIV type 1 non-B subtypes. *AIDS* 32, 469–476. doi: 10.1097/QAD.0000000000001726

Passos, D. O., Li, M., Yang, R., Rebensburg, S. V., Ghirlando, R., Jeon, Y., et al. (2017). Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* 355, 89–92. doi: 10.1126/science.aah5163

Rhee, S. Y., Gonzales, M. J., Kantor, R., Betts, B. J., Ravela, J., and Shafer, R. W. (2003). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* 31, 298–303. doi: 10.1093/nar/gkg100

Saladini, F., Meini, G., Bianco, C., Monno, L., Punzi, G., Pecorari, M., et al. (2012). Prevalence of HIV-1 integrase mutations related to resistance to dolutegravir in raltegravir naive and pretreated patients. *Clin. Microbiol. Infect.* 18, E428–E430. doi: 10.1111/j.1469-0691.2012.03917.x

Singh, K., Marchand, B., Rai, D. K., Sharma, B., Michailidis, E., Ryan, E. M., et al. (2012). Biochemical mechanism of HIV-1 resistance to rilpivirine. *J. Biol. Chem.* 287, 38110–38123. doi: 10.1074/jbc.M112.398180 PMCID: PMC3488081 doi: 10.1074/jbc.M112.398180

Takahata, T., Takeda, E., Tobiume, M., Tokunaga, K., Yokoyama, M., Huang, Y. L., et al. (2017). Critical contribution of Tyr15 in the HIV-1 integrase (IN) in facilitating IN assembly and nonenzymatic function through the IN precursor Form with reverse transcriptase. *J. Virol.* 91, e2003–e2016. doi: 10.1128/JVI.02003-16

Thierry, E., Deprez, E., and Delelis, O. (2016). Different pathways leading to integrase inhibitors resistance. *Front. Microbiol.* 7:2165. doi: 10.3389/fmicb.2016.02165

Tsiang, M., Jones, G. S., Goldsmith, J., Mulato, A., Hansen, D., Kan, E., et al. (2016). Antiviral activity of bictegravir (GS-9883), a novel potent HIV-1 integrase strand transfer inhibitor with an improved resistance profile. *Antimicrob. Agents Chemother.* 60, 7086–7097. doi: 10.1128/AAC.01474-16

Vavro, C., Hasan, S., Madsen, H., Horton, J., Deanda, F., Martin-Carpenter, L., et al. (2013). Prevalent polymorphisms in wild-type HIV-1 integrase are unlikely to engender drug resistance to dolutegravir (S/GSK1349572). *Antimicrob. Agents Chemother.* 57, 1379–1384. doi: 10.1128/AAC.01791-12

Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001). Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* 20, 7333–7343. doi: 10.1093/emboj/20.24.7333

Wensing, A. M., Calvez, V., Gunthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., et al. (2017). 2017 Update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* 24, 132–133.

Whitfield, T., Torkington, A., and Van Halsema, C. (2016). Profile of cabotegravir and its potential in the treatment and prevention of HIV-1 infection: evidence to date. *HIV AIDS* 8, 157–164. doi: 10.2147/HIV.S97920

Wimley, W. C., Gawrisch, K., Creamer, T. P., and White, S. H. (1996). Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proc. Natl. Acad. Sci. U.S.A.* 93, 2985–2990. doi: 10.1073/pnas.93.7.2985

Yin, Z., Shi, K., Banerjee, S., Pandey, K. K., Bera, S., Grandgenett, D. P., et al. (2016). Crystal structure of the *Rous sarcoma virus* intasome. *Nature* 530, 362–366. doi: 10.1038/nature16950

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Chapter 6 Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding

6.1. Article title

Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding.

6.2. Authors and citations

Rumbidzai Chitongo[1*¶], Adetayo Emmanuel Obasa[2*¶], Sello Given Mikasi[2], Graeme Brendon Jacobs[2], Ruben Cloete[1¶]

Citation number: none

6.3. Author's contribution

In the enclosed manuscript; I confirmed that I am a joint-first author. I obtained the patient samples, performed the laboratory experiments, including Viral RNA extraction, PCR, PCR clean-up reactions and conventional Sanger sequencing. I obtained additional sequences from theHIV Los Alamos National Library Database (www.lanl.gov). I performed the sequence analyses, sent the aligned and consensus sequences for In-silico simulation to Dr. Ruben Cloete and Miss Rumbidzai Chitongo. I had in person weekly meetings with Dr. Cloete, discussing and analysing the data.

6.4. Background

Resistance associated mutations (RAMs) threaten the long-term success of combination antiretroviral therapy (cART) outcomes for HIV-1 treatment. HIV-1 Integrase (IN) strand transfer inhibitors (INSTIs) have proven to be a viable option for highly specific HIV-1 therapy. The INSTI, Dolutegravir is recommended by the World Health Organization for use as first-line cART. This study aims to understand how RAMs affect the stability of IN, as well as the binding of the drug onto the catalytic pocket of the protein. Computational methods have been used to assess the impact of known RAMs on the binding of Dolutegravir to the IN protein. Molecular modelling of HIV-1C IN was performed using the SWISS-MODEL webserver, with quality assessment performed using internal methods and external software tools. The site directed mutator webserver was applied for stability predictions on the effect of known RAMs while FoldX confirmed any changes in energy.

6.5. Main findings

All the MD trajectory analysis considered the single chain A (monomer) of the IN protein in contact with the drug DTG. Trajectory analysis of the root mean square deviation (RMSD) of the backbone indicated that the WT system reached equilibrium after 100 ns as well as the Q92, R143 and S140 mutant systems. We performed interaction analysis for five snapshots (every 50 ns) of each of the simulation systems to determine which residues played a role in the binding of DTG to the protein in the WT and mutant protein structures. For the WT system, interactions were observed between known active site residues D64, D116 and N148, MG ion and to DNA nucleotides

6.6. Study significance

Our findings suggest the G140S mutant has a strong effect on the HIV-1C IN protein structure and Dolutegravir binding and should be validated using laboratory-based experiments. This approach can be applied to determine the effect of other mutations/variants on HIV-1C integrase drug binding.

6.7. Conclusion

The findings from our study suggest that patients should be screened for mutations and or novel variants to determine if the drug DTG will be efficacious or not. The model generated in this study can be used to tease out the effects of novel variants.

6.8. Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding

Rumbidzai Chitongo[1*¶], Adetayo Emmanuel Obasa[2*¶], Sello Given Mikasi[2], Graeme Brendon Jacobs[2], Ruben Cloete[1¶]

[1] South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa.

[2] Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Cape Town, South Africa.

Corresponding author: Dr. Ruben Cloete; email address:ruben@sanbi.ac.za

*Both authors contributed equally

Abstract

Resistance associated mutations (RAMs) threaten the long-term success of combination antiretroviral therapy (cART) outcomes for HIV-1 treatment. HIV-1 Integrase (IN) strand transfer inhibitors (INSTIs) have proven to be a viable option for highly specific HIV-1 therapy. The INSTI, Dolutegravir is recommended by the World Health Organization for use as first-line cART. This study aims to understand how RAMs affect the stability of IN, as well as the binding of the drug onto the catalytic pocket of the protein. Computational methods have been used to assess the impact of known RAMs on the binding of Dolutegravir to the IN protein. Molecular modelling of HIV-1C IN was performed using the SWISS-MODEL webserver, with quality assessment performed using internal methods and external software tools. The site directed mutator webserver was applied for stability predictions on the effect of known RAMs while FoldX confirmed any changes in energy. Interaction analysis between neighbouring residues was done using PyMOL. Three randomly selected mutations were chosen for molecular dynamic simulation studies using Gromacs. Trajectory analysis included Root mean square deviation and fluctuation, Radius of gyration, Principal component analysis and Interaction analysis between Dolutegravir and protein residues. The structural quality assessment indicated high reliability of the HIV-1C IN tetrameric structure, with more than 90% confidence in modelled regions. Change in free energy for the G140S mutant indicated a stabilizing effect and simulation analysis showed it to affect structural stability and flexibility of the protein structure. This was further supported by the drug being expelled from the G140S mutant active site, as indicated by interaction analysis. Our findings suggest the G140S mutant has a strong effect on the HIV-1C IN protein structure and Dolutegravir binding and should be validated using laboratory-based experiments. This approach can be applied to determine the effect of other mutations/variants on HIV-1C integrase drug binding.

Introduction

The Integrase (IN) enzyme plays an important role in the Human Immunodeficiency Virus type 1 (HIV-1) replication cycle by catalysing two distinct reactions termed: 3'-end processing and strand transfer. During the 3' processing, IN removes two nucleotides from the 3' ends of both viral DNA strands and exposes the C-alpha hydroxyl group on the 3'ends. The subsequent step involves strand transfer whereby, IN attacks the phosphodiester backbone of the host DNA and links the exposed 3'-end to the 5' hydroxyl end of the host DNA [1]. This makes HIV-1 IN an important target for combination antiretroviral therapy (cART). HIV-1 IN is a 32 kilo Dalton (kDa) protein, and consist of three structural and functional domains; the N-terminal domain (NTD, residues 1-49), the catalytic core domain (CCD, residues 50-212), and C-terminal domain (CTD, residues 213-288). It also contains a conserved DDE motif consisting of residues Asp54, Asp116 and Glu152 in the CCD, important for drug binding and enzyme activity [2]. Several IN strand transfer inhibitors (INSTIs) have been developed [3–5]. These inhibitors include; Raltegravir (RAL) and Elvitegravir (EVG) as first-line INSTIs and the most recently approved second-line inhibitors Dolutegravir (DTG) and Bictegravir (BIC) [6].

DTG is a coplanar and linear molecule with a high barrier against resistance, it is safe and tolerable and shows little to none drug interactions [7]. Furthermore, a recent study provided evidence for the replacement of RAL with DTG based on the low prevalence of DTG resistance and the low risk for INSTI mutations when patients are on DTG treatment [8]. Although BIC has been recently approved, not much information is known about it and as a result, DTG still remains the preferred option. The functional mechanism of INSTIs is to bind to the catalytically essential magnesium ions, thereby displacing the reactive 3'-hydroxyl group of the terminal A17 away from the active site which disrupts the strand transfer process. Several mutations have emerged in patients receiving first-line INSTIs, RAL and EVG. Brado *et al*. reported that despite higher fold RAMs against INSTIs being absent in most treatment naïve patients, they can emerge under treatment, particularly with first generation IN STIs [9].

Several studies have used the prototype foamy virus intasome structure (medium sequence identity) to model the structure of HIV-1 IN in order to investigate the effect of mutations on HIV-1 IN using molecular dynamic simulations [10]. These studies revealed the binding mode of EVG and RAL to HIV-1 IN and the structural mechanism of drug resistant mutants that affect the 140's loop region spanning residues 140-149. The precise role of this loop is unknown, but molecular dynamics studies have demonstrated the importance of this loop's flexibility for catalysis [11]. This loop has been reported to regulate the active site of HIV-1 IN by decreasing or increasing flexibility under the influence of mutations G140A and G149A. Conformational flexibility of this loop is thought to be

important for the catalytic steps following DNA binding, as a decrease in flexibility induced, for example, by the G140A mutation, results in lower levels of activity, despite minimal effects on DNA binding [11]. Mouscadet *et al*. [12] reported that the G140 residue is not directly involved in the cooperative flexibility of the catalytic loop, but plays a critical role in controlling the overall motion of the loop and its precise position relative to the phosphodiester bond to be cleaved. The G140 residue also participates in the catalytic loop hinge formation and its mutation may restore specific contacts required for catalysis, between the loop of the double mutant and the end of the viral DNA [11,13,14]. The findings from these studies were inconclusive due to the poor quality of the protein models delineating the active site and viral DNA binding site for simulation studies.

In 2017, Cryogenic electron microscopy was used to solve the structure of HIV-1 strand transfer complex intasome for HIV-1 subtype B [15]. This provided us with a unique opportunity to model the structure of HIV-1 subtype C IN to interrogate the effect of known drug resistance associated mutations (RAMs) on the protein structure using molecular dynamic simulation studies. This is the first study that uses the consensus wild type subtype C IN sequence to build an accurate 3D model of HIV-1C IN to understand the effect of three statistically enriched mutations on DTG drug binding.

Materials and Methods

Generation of consensus HIV-1C Integrase sequence

To compare our sequences with the rest of the IN sequences from South Africa, we performed a search on the HIV Los Alamos National Library (LANL) database (https://www.hiv.lanl.gov/components/sequence/HIVsearch.com). Our search inclusion criteria included all South African HIV-1 subtype C IN sequences and those identified from treatment naïve patients. We selected one sequence per patient and all problematic sequences were excluded from further analyses. Finally, the consensus sequence was generated using the database-derived HIV-1C_{ZA} sequences (**n = 314**) and cohort sequences (**n = 91**) [9] . Nucleotide sequences were verified for stop codons, insertion and deletions using an online quality control program on the HIVLANL database (https://www.hiv.lanl.gov/content/sequence/QC/index.htm). Multiple sequence alignments were done with MAFFT version 7, from which the consensus sequence was derived [16]. As part of quality control, each of the viral sequences were inferred on a phylogenetic tree in order to eliminate possible contamination.

Molecular modelling and quality assessment

The crystal structure of the HIV-1B intasome (PDBID: 5U1C) was used to generate a three-dimensional tetrameric structure of HIV-1C IN using the consensus HIV-1C sequence that we generated. The SWISSMODEL webserver was used for model generation by first constructing a pairwise sequence-structure alignment between HIV-1C wild-type (WT) amino acid sequence and template 5U1C [17]. The quality of the resulting model was assessed using SWISSMODEL quality assessment scores, Root mean square deviation analysis compared to homologous template (PDBID: 5U1C) and with publicly available algorithms located at the SAVES webserver (https://servicesn.mbi.ucla.edu/SAVES/) namely; ERRAT, VERIFY3D and Ramachandran plot [18,19].

Structure preparation

The predicted 3D structure of HIV-1C IN was superimposed to 5U1C to extract proviral DNA, while the Magnesium (MG) ions and drug DTG were obtained from homologous template 3s3m (Prototype foamy virus). The wild-type (WT) structure of HIV-1C was energy minimized in complex with DNA, MG and DTG using Gromacs version 5.1 [20]. Thereafter, we predicted the stabilizing and/or destabilizing effect of mutations on the protein structure. For this purpose, the site directed mutator (SDM) webserver and the software FoldX was used to predict the change in Gibbs free energy after the introduction of the mutation. We also calculated the loss or gain of polar interactions between neighbouring residues located adjacent to the mutation using PyMOL [21].

Molecular dynamic simulation

For simulation studies we only considered the two inner dimers of the protein structure, as the other two monomers were similar in sequence and structure. Three different mutant systems were prepared by introducing a specific mutation into the WT structure through the mutagenesis wizard in PyMOL and energy minimizing the structures using Gromacs [20]. The WT and three mutant systems (E92Q, S140 and R143) were prepared by uploading the atomic coordinates of the Protein-DNA-MG-DTG complexes to CHARMM-GUI interface [22]. The three mutant systems selected for simulation studies represent three resistance pathways associated with RAL, EVG and possibly DTG resistance. The option solution builder was used as an input generator. Each system was solvated in a rectangular TIP3 water-box with 10Å distance between the edges of the box. The topology and coordinates for each system was generated using CHARMM36 force field and CHARMM general force field for DTG. Each system was neutralized by adding counter ions to each of the systems. For the WT system, 157 potassium ions (K) and 81 chloride ions (Cl) were added, for the mutant R143 system we added 156 K and 81 Cl ions, while for the mutant system S140 we added 157 K and 81 Cl ions and for the

mutant system Q92 we added 335 K and 276 Cl ions. Each system was at a final concentration of 0.15M for simulation dynamics.

Gromacs version 5.1 was used for running all the simulations [20]. Each system underwent 50000 steps of steepest descents energy minimization to remove steric overlap. Afterwards, all the systems were subjected to a short position restraint NPT (constant Number of particles, Pressure and Temperature) for 500 picoseconds (ps) to stabilize the pressure of the system by relaxing the system and keeping the protein restrained. For NPT, the Nose-Hoover pressure coupling was turned on with constant coupling of 1ps at 303.15K under conditions of position restraints (h-bonds) selecting a random seed. Electrostatic forces were calculated using Particle Mesh Ewald method [20]. All systems were subjected to a full 300 ns simulation under conditions of no restraints.

The analyses of the trajectory files were done using GROMACS utilities. The root mean square deviation (RMSD) was calculated using gmx rmsd and root mean square fluctuation (RMSF) analysis using gmx rms. The radius of gyration was calculated using gmx gyrate to determine if the system reached convergence over the 300 nanoseconds (ns) simulation. Pairwise distance analysis between the drug and MG was done using gmx pairdist tool. Afterwards we extracted structures every 50 ns over the last 200 ns of the equilibrated system to determine any structural changes and differences in the number of interactions between the protein and drug at different time intervals.

Principal component analysis

Principal component analysis (PCA) is a statistical technique that reduces the complexity of a data set in order to extract biologically relevant movements of protein domains from irrelevant localized motions of atoms. For PCA analysis the translational and rotational movements were removed from the system using gmx covar from GROMACS to construct a covariance matrix. Next the eigenvectors and eigenvalues were calculated by diagonalizing the matrix. The eigenvectors that correspond to the largest eigenvalues are called "principal components", as they represent the largest-amplitude collective motions. We filtered the original trajectory and project out the part along the most important eigenvectors namely: vector 1 and 2 using gmx anaeig from GROMACS utilities. Furthermore, we visualized the sampled conformations in the subspace along the first two eigenvectors using gmx anaeig in a two-dimensional projection.

Results

Sequence and structural analysis

The amino acid sequence of HIV-1C IN shared 93.4% sequence identity with the template 5u1c amino acid sequence. The protein structure built for HIV-1C IN had a Global mean quality estimate score of 0.59 and 60% sequence similarity to 5u1c. The tetrameric structure of HIV-1C is shown in Fig 1, with each monomer labelled and colour coded. The VERIFY 3D score was 80.1%, the ERRAT overall quality score was 90% and higher for all four chains (A, B, C and D) and the Ramachandran plot indicated more than 90% of residues fell within the most favoured regions of the plot suggesting the predicted structure is a reliable model. Stability predictions indicated 15 RAMs to be destabilizing and five to be stabilizing for the protein structure based on SDM delta-delta G free energy scores (Table 1). The FoldX change in unfolded energy values indicated that the S140 was stabilizing, Q92 destabilizing and R143 neutral based on comparison with the WT structure each having values of 162.89, 131.94, 146.47 and 151.83 Kcal/Mol, respectively. Interaction analysis showed ten mutations resulted in a loss of polar contacts; three resulted in an increase in polar contacts, while seven showed no change in the number of polar contacts with neighbouring residues (Table 1).

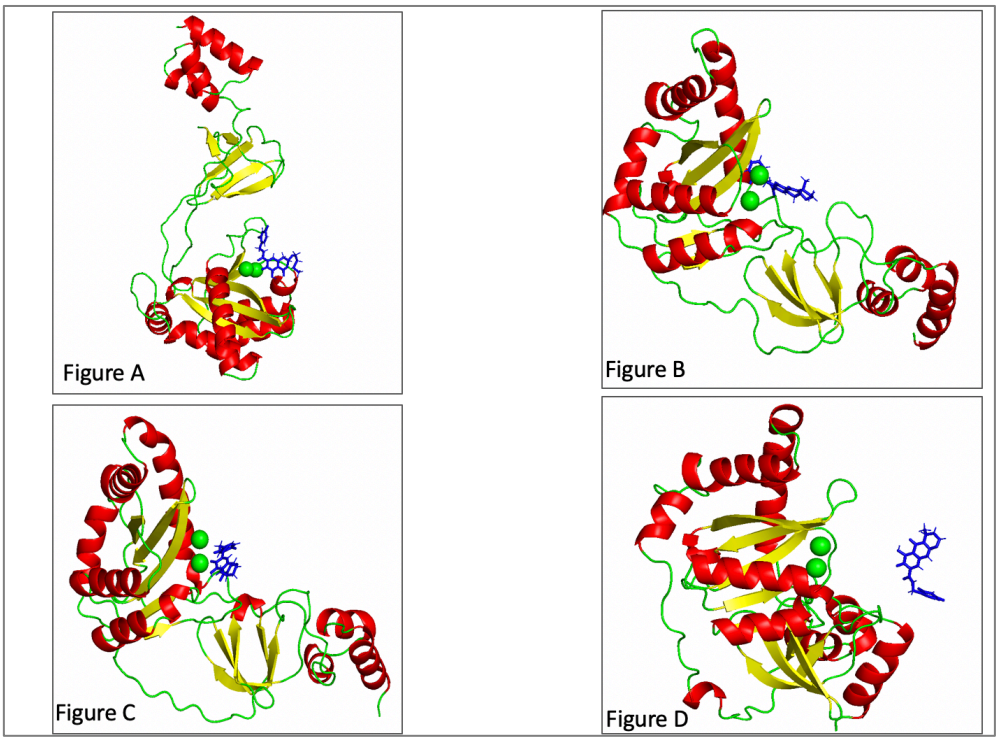


Figure 6.1 Tetrameric 3D structure of HIV-1C Integrase in complex with DNA, MG and drug Dolutegravir. Magnesium^{2+} ions (dirty violet spheres) are shown sitting in close proximity with Dolutegravir (brown) within the binding pocket (DDE motif) of the protein. A pair of dimers is seen encapsulating the viral/host DNA chimera, in which the two inner molecules (chain A and C) are shown in direct contact with DNA.

Table 6:1 Summary of stability predictions and polar interactions.

	SDM[1]	FoldX[2]		Polar Interactions	
Mutation	Predicted ΔΔG (Kcal/Mol)	Total Energy ΔG (Kcal/Mol)	Energy Difference	Wild Type	Mutant
WT	N.A	151.83	N.A	N.A	N.A
T66A	-1.2	152.69	0.86	3 (H67, I73, ADE21)	1 (I73)
T66I	0.08	153.63	1.80	3 (H67, I73, ADE21)	1 (I73)
T66K	-0.61	160.15	8.32	3 (H67, I73, ADE21)	1 (I73)
E92Q	-0.16	131.94	-19.89	3 (Q136, I113, T115)	0
E138K	-0.12	151.32	-0.51	3 (Q136, I113, T115)	2 (T115, I113)
E138A	-0.4	152.01	0.18	3 (Q136, I113, T115)	2 (T115, I113)
E138T	0.48	151.82	-0.01	3 (Q136, I113, T115)	2 (T115, I113)
G140S	-0.58	162.89	11.06	2 (T115, N117)	3 (Q148, T115, N117)
G140A	-0.68	152.43	0.60	2 (T115, N117)	2(T115, N117)
G140C	0.39	154.81	2.98	2 (T115, N117)	2(T115, N117)

Y143C	0.14	152.49	0.66	None	None
Y143R	-0.08	146.47	-5.36	None	1 (S230)
Y143H	-0.07	152.28	0.45	None	None
S147G	-0.18	151.16	-0.67	2 (Q148, N144)	1 (Q144)
Q148H	0.63	157.78	5.95	3 (V151, P145, S147)	2 (V151, P145)
Q148K	-0.78	151.33	-0.50	3 (V151, P145, S147)	3 (V151, P145, H114)
Q148R	-0.71	152.53	0.70	3 (V151, P145, S147)	4 (D116,P145, V150, V151)
Q148N	-0.82	151.58	-0.25	3 (V151, P145, S147)	3 (V151, P145, H114)
N155H	-0.23	152.01	0.18	3 (V151, P145, S147)	3 (E152, V151, K159)
R263K	-0.29	151.15	-0.68	4 (Q146, N144, GUA18, ADE18)	0

[1]negative values for ΔΔG indicate a stabilizing effect and positive values destabilizing.

[2]positive energy difference ΔG values >1 indicate a destabilizing effect, whereas values $1 \leq \Delta G \leq 0$ imply a neutral effect and ΔG values > -1 indicate a stabilizing effect. Abbreviations used: N.A- not applicable. The number in front of brackets is the total amount of interactions. Abbreviations of amino acids: A -Alanine; D-Aspartic acid; E-Glutamic acid; G-Glycine; H-Histidine; I-Isoleucine; K-Lysine; N-Asparagine; Q-Glutamine; R-Arginine; S-Serine; T-Threonine; Y-Tyrosine.

Molecular dynamic simulations

All the MD trajectory analysis considered the single chain A (monomer) of the IN protein in contact with the drug DTG. Trajectory analysis of the RMSD of the backbone indicated that the WT system reached equilibrium after 100 ns as well as the Q92, R143 and S140 mutant systems (Fig 2A). Only S140 showed higher RMSD fluctuation values compared to the WT, R143 and Q92 systems (Fig 2A). RMSF analysis clearly showed higher flexibility for the S140 mutant system, with four highly flexible regions (residues 68 - 70, 142 - 146, 166 - 170 and 253 - 256) compared to the WT, Q92 and R143 systems (Fig 2B). These flexible regions affect the 140's loop region that regulates drug binding. The Radius of gyration values indicated decreasing values for R143 and E92 compared to the WT and S140 mutant system (Fig 2C). Plotting the first two principal components provided insight into the collective movement of each protein atom. The 2D projections of the first and second principal components for the WT, Q92, R143 and S140 systems are shown in Figure 2D. Calculation of the covariance matrix values after diagonalization showed a significant increase for the S140 system (18.33 nm) compared to the other three systems WT, Q92 and R143 each having 9.58 nm, 8.98 nm and 10.41 nm lower values, respectively. Distance analysis indicated a smaller average distance of 0.21 □ 0.01 nm and 0.22 □ 0.01 nm between the WT, R143 MG ion and drug DTG systems compared to Q92 and S140 each having a distance of 0.41 □ 0.04 nm, 0.99 □ 0.20 nm, respectively (Fig 2D).

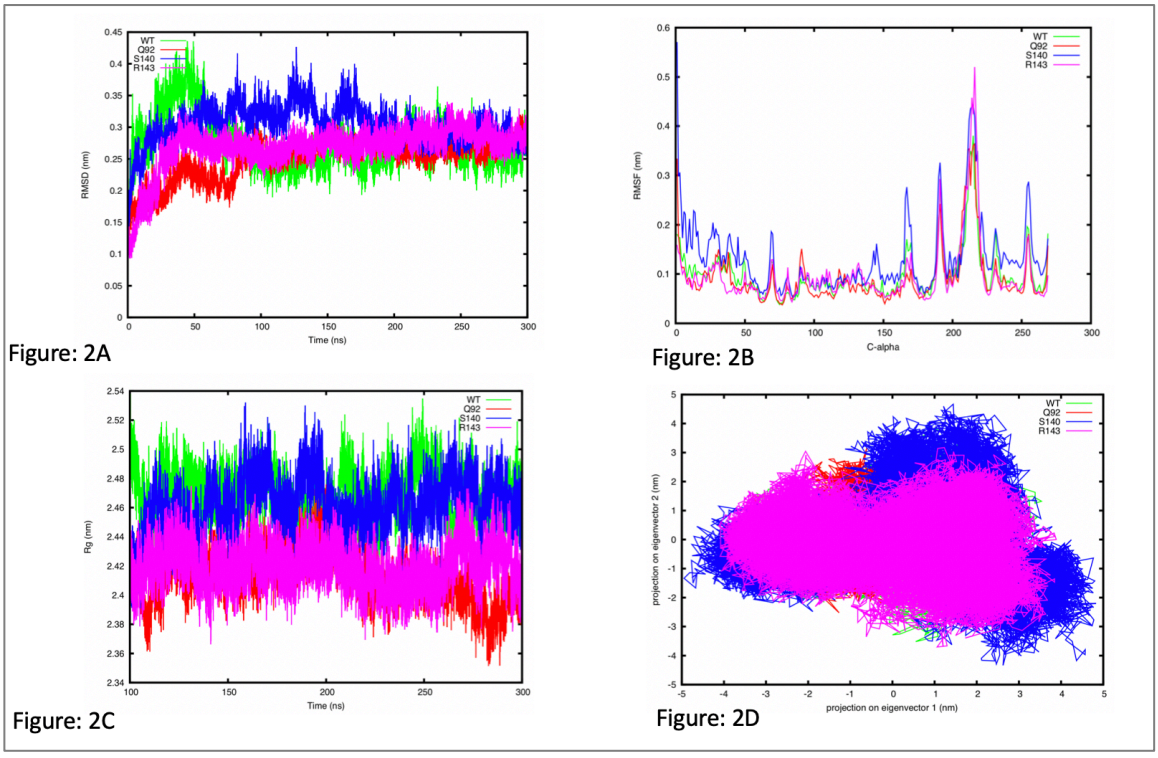


Figure: 2A

Figure: 2B

Figure: 2C

Figure: 2D

Figure 6.2 Trajectory analysis of the four simulation systems.

(A) Change in backbone RMSD for the WT, Q92, S140 and R143 systems plotted over 300 ns. (B) Change in RMSF for the C-alpha residues for the WT, Q92, S140 and R143 systems plotted over the last 200ns. (C) Measure of compactness for the WT, Q92, S140 and R143 systems plotted over the last 200 ns. (D) First two principal components plotted for the WT, Q92, S140 and R143 systems plotted over the last 200 ns.

Interaction analysis

We performed interaction analysis for five snapshots (every 50 ns) of each of the simulation systems to determine which residues played a role in the binding of DTG to the protein in the WT and mutant protein structures. For the WT system, interactions were observed between known active site residues D64, D116 and N148, MG ion and also to DNA nucleotides (Table 2). Similarly, interactions were observed between known active site residues D64, R143, N148, MG ion and DNA nucleotides for the R143 system (Table 2). Interestingly, no active site residue interactions between DTG and S140 structure was observed and no MG ionic interactions for S140. Only the Q92 system showed interactions with active site residues but no MG ionic interactions (Table 2). Figs A-D in S1 Supporting Information File shows the relative location of the drug to the MG ion and active site for snapshot 1 taken at 100 ns. Compared to the WT, R143, Q92 the drug DTG is located close to the MG ion and active site while for S140 the drug is located outside the catalytic site suggesting the mutation is able to expel the drug from the binding pocket over time (Figs A-D in Supporting Information File).

Table 6:2 Summary of interaction analysis.

Structure	Cluster	Interactions	
		Hydrogen bonds	Ionic
WT	1 (100 ns)	2 (GUA22 D116)	MG
	2 (150 ns)	3 (THY11, D64, D116)	MG
	3 (200 ns)	2 (GUA22, D116)	MG
	4 (250 ns)	4 (THY11, GUA22, D64, D116)	MG
	5 (300 ns)	2 (THY11, N148)	MG
R143	1 (100 ns)	4 (THY11, GUA22, D64, N148)	MG
	2 (150 ns)	4 (GUA22, D64, R143, N148)	MG
	3 (200 ns)	4 (GUA22, D64, R143, N148)	MG
	4 (250 ns)	4 (GUA22, D64, R143, N148)	MG
	5 (300 ns)	5 (THY11; GUA22, GUA22, D64, N148)	MG
Q92	1 (100 ns)	3 (CYT20, D116, P145)	None
	2 (150 ns)	3 (D116, P145, E152)	None
	3 (200 ns)	3 (CYT20, H21, D116)	None
	4 (250 ns)	4 (CYT20, P142, P145, E152)	None
	5 (300 ns)	3 (CYT20, P145, N148)	None
S140	1 (100 ns)	3 (GUA22, ADE25, ADE27)	None
	2 (150 ns)	3 (GUA22, ADE25, ADE27)	None
	3 (200 ns)	3 (GUA22, ADE25, ADE27)	None
	4 (250 ns)	3 (THY11, GUA22, ADE27)	None
	5 (300 ns)	3 (THY11, GUA22, ADE27)	None

Discussion

Sequence variation between HIV-1B and HIV-1C can have a meaningful contribution to the protein structure, affecting protein drug interaction. We investigated known drug RAMs associated with RAL, EVG and DTG resistance to determine their effect on the protein structure and drug binding. The structural modelling of HIV-1C IN considered a homologous template of high sequence identity, and good overall target sequence coverage, compared to previous homology models that considered templates of low sequence identity. We could therefore accurately reconstruct HIV-1C using the close homolog HIV-1B crystal structure as template to infer accurate drug interactions. Further inspection of the overall structure confirmed accurate prediction of more than 90% of domains within the protein structure, compared to the template HIV-1B structure. The quality analysis provided support for the predicted model based on side chain conformations. Stability predictions showed contrasting results to interaction analysis, whereby amino acid substitutions that resulted in a gain of interactions was predicted to be destabilising. The FoldX changes in energy values were similar to interaction analysis for the three mutant structures under investigation. To fully comprehend the effects of individual mutations we opted to use molecular dynamic (MD) simulations to understand the effect of selected mutations on protein movement and drug interactions. MD analyses have shown to be successful in quantifying small changes in protein structures that can affect overall drug binding [23]. We randomly selected three mutations Q92, S140 and R143 to investigate the role of amino acid substitutions on protein structure and drug binding in comparison to the WT protein structure. Analysis of the change in trajectory of the mutant systems compared to the wild type suggested less stability and higher fluctuation of the G140S mutant system compared to the WT system. We also confirmed the destabilizing effect of the G140S mutant using principal component analysis which suggested larger randomized concerted movement for the G140S mutant compared to the WT, Q92 and R143 systems.

These findings are similar to Chen et al. [10] which showed that the G140S mutations can either stabilize or destabilize the 140's loop region. In our case, the 140's loop region is stabilized by the G140S mutation that could prevent drug binding. This is supported by pairwise distance analysis confirming a larger distance between the MG ion and drug DTG for the G140S mutant system compared to the WT and R143. Further interaction analysis was performed to confirm the hypothesis that the S140 mutation could prevent drug binding by extracting structures at different snapshots of the simulation. Here, we found that the G140S mutation resulted in the drug being expelled from the binding pocket. We also observed weaker interactions for the Q92 mutation but stronger interactions for R143 mutant. The findings from our study suggest that

patients should first be screened for mutations and or novel variants to determine if the drug DTG will be efficacious or not. The model generated in this study can be used to tease out the effects of novel variants. A few limitations of this study are the use of RAL, EVG mutants and not considering novel RAL or DTG mutations. We also only simulated single mutations instead of double however we have yet to identify double mutants within the South African cohort of HIV-1C infected patients. Future work will include viral fitness assays to determine the effect of the mutant S140 on the HIV-1C virus and the effect of the drug DTG on the virus ability to persist in the presence of the drug.

Acknowledgements

We would like to acknowledge the Poliomyelitis Research Foundation (PRF) of South Africa for funding this project. The Centre for High Performance Computing (CHPC), Rondebosch, South Africa for access to their high-performance computing cluster and the Higher Education Department, next Generation of Academic Programme (nGAP) for full time academic positions of Dr Cloete and Dr Jacobs. Finally, the National Research Foundation (NRF) for partial funding to kick-start this project.

References

1. Pommier Y, Johnson AA, Marchand C. Integrase inhibitors to treat HIV/AIDS. Nat Rev Drug Discov. 2005;4(3):236–48.
2. Zheng R, Jenkins TM, Craigie R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. Proc Natl Acad Sci U S A. 1996;93(24):13659–64.
3. Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, Ferrara M, et al. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. J Med Chem. 2008;51(18):5843–55.
4. Okeke NL, Hicks C. Role of raltegravir in the management of HIV-1 infection. HIV/AIDS - Res Palliat Care. 2011;3:81–92.
5. Marchand C, Johnson AA, Karki RG, Pais GCG, Zhang X, Cowansage K, et al. Metal-dependent inhibition of HIV-1 integrase by β-diketo acids and resistance of the soluble double-mutant (F185K/C280S). Mol Pharmacol. 2003;64(3):600–9.
6. Lenz JCC, Rockstroh JK. S/GSK1349572, a new integrase inhibitor for the treatment of HIV: Promises and challenges. Expert Opin Investig Drugs. 2011;20(4):537–48.
7. Di Santo R. Inhibiting the HIV integration process: Past, present, and the future. J Med Chem. 2014;57(3):539–66.
8. Jere KC, Hungerford D, Bar-zeev N, Kanjerwa O, Houpt ER, Operario DJ, et al. Ce Pt Cr Ip T Ce Pt Us Cr T. 2000;
9. Brado D, Obasa AE, Ikomey GM, Cloete R, Singh K, Engelbrecht S, et al. Analyses of HIV-1 integrase sequences prior to South African national HIV-Treatment program and available of integrase inhibitors in Cape Town, South Africa. Sci Rep. 2018;8(1):1–9.
10. Chen J, Shao J, Cai R, Shen Y, Zhang R, Liu L, et al. Anti-retroviral therapy decreases but does not normalize indoleamine 2,3-dioxygenase activity in HIV-infected patients. PLoS One. 2014;9(7):3–10.
11. Mouscadet JF, Delelis O, Marcelin AG, Tchertanov L. Resistance to HIV-1 integrase inhibitors: A structural perspective. Drug Resist Updat. 2010;13(4–5):139–50.
12. Mouscadet JF, Arora R, André J, Lambry JC, Delelis O, Malet I, et al. HIV-1 IN alternative molecular recognition of DNA induced by raltegravir resistance mutations. J Mol Recognit. 2009;22(6):480–94.

13. Canducci F, Marinozzi MC, Sampaolo M, Boeri E, Spagnuolo V, Gianotti N, et al. Genotypic/phenotypic patterns of HIV-1 integrase resistance to raltegravir. J Antimicrob Chemother. 2010;65(3):425–33.

14. Blanco JL, Varghese V, Rhee SY, Gatell JM, Shafer RW. HIV-1 integrase inhibitor resistance and its clinical implications. J Infect Dis. 2011;203(9):1204–14.

15. Passos DO, Li M, Yang R, Rebensburg S V., Ghirlando R, Jeon Y, et al. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. Science (80-). 2017;355(6320):89–92.

16. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, et al. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. MBio. 2013;4(5):1–11.

17. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947–8.

18. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci. 1993;2(9):1511–9.

19. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C., Beck JS. [2]»ÉøÍ¸ÐÔµÄ © 19 9 2 Nature Publishing Group. Nature. 1992;359:710–3.

20. Spoel D Van Der. Gromacs Reference Manual v5.1. 2011;

21. Lunin VY, Urzhumtsev A, Bockmayr A, Fokin A, Urzhumtsev A, Afonine P, et al. Theory and Techniques 12. Binary Integer Programming and its Use for Envelope Determination Bulk Solvent Correction for Yet Unsolved Structures Search of the Optimal Strategy for Refinement of Atomic Models Metal Coordination Groups in Proteins: Some Comm. 2002;(4). Available from: http://www.iucr.org

22. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput. 2016;12(1):405–13.

23. Nair PC, Miners JO. Molecular dynamics simulations: from structure function relationships to drug discovery. Silico Pharmacol. 2014;2(1):2–5.

Chapter 7 Mutations in Long Terminal Repeats Kb Transcription Factor Binding Sites in Plasma Virus Among South African People Living with HIV-1

7.1. Article title

Mutations in Long Terminal Repeats Kb Transcription Factor Binding Sites in Plasma Virus Among South African People Living with HIV-1.

7.2. Authors and citations

Obasa AE, Ashokkumar M, Neogi U, and Jacobs GB,

Citation number: none

AIDS Res Hum Retroviruses. 2019 Jun;35(6):572-576. doi: 10.1089/AID.2018.0293. Epub 2019 April

7.3. Author's contribution

In the enclosed manuscript; I confirmed that I am a the first author. I obtained the patient samples, performed the laboratory experiments, including Viral RNA extraction, PCR, PCR clean-up reactions and conventional Sanger sequencing. I performed the sequence analyses, sent the sequences for further analyses. Mr. Ashokkumar M and I wrote the first draft. I confirmed that I am the corresponding author for this manuscript.

7.4. Background

HIV-1 subtype C (HIV-1C) is responsible for the majority of infections in sub-Saharan Africa. We selected 63 plasma-derived samples and generated Long Terminal Repeats (LTR) amplicons from people living with HIV (PLHIV) in South Africa to identify transcription factor binding sites. NF-κB plays an important role in regulating the viral gene expression from the viral promoter and controlling viral latency.

7.5. Main findings

LTR amplicons were sequenced and phylogenetically analysed. In our dataset, we identified F-κB sites (n= 4; 6%) at position II and (n = 1; 1%) at position I among 63 sequences analysed. The majority of the sequences identified with H-κB at position II (n = 50; 79%) and position I (n = 55; 87%). Forty-nine (n = 49; 78%) sequences were found to exhibit C-κB site. ZA_LTR052 was identified with a single point mutation, but it resembles C-κB like. We

identified all three NF-κB-binding sites in (n=44; 70%) the viral promoter–enhancer regions in South African patients.

7.6. Study significance

This study investigate the diversity in the regulatory motifs within the LTR which may influence viral fitness by acting at the level of transcription factor. This study also contribute to the knowledge diversity of difference in the regulatory motifs within HIV-1C LTR.

7.7. Conclusion

In summary, we report that South African HIV-1C contain C-κB- binding site in their LTR regions, but no evolution of fourth NF-κB as reported previously. Viral fitness may be influenced by differences in regulatory motifs by acting at the level of transcription level. Majority of our sequences have additional NF-κB sites signature for HIV-1C viruses; however, it remains to be established the potential effect of a third NF-κB site has on HIV-1C replicative capacity and rate of progression in people living with HIV-1C in South Africa.

7.8. Open Access

The article is not freely available online. Reprints can also be obtained from the author upon request: obasa@sun.ac.za. PubMed link: https://www.ncbi.nlm.nih.gov/pubmed/30793917

AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 00, Number 00, 2019

DOI: 10.1089/aid.2018.0293

Mutations in Long Terminal Repeats κB Transcription Factor Binding Sites in Plasma Virus Among South African People Living with HIV-1

Adetayo Emmanuel Obasa,[1,2] Manickam Ashokkumar,[2,3] Ujjwal Neogi,[2] and Graeme Brendon Jacobs[1]

Abstract

HIV-1 subtype C (HIV-1C) is responsible for the majority of infections in sub-Saharan Africa. We selected 63 plasma-derived samples and generated long terminal repeats (LTRs) amplicons from people living with HIV in South Africa to identify transcription factor binding sites. NF-κB plays an important role in regulating the viral gene expression from the viral promoter and controlling viral latency. LTR amplicons were sequenced and phylogenetically analyzed. In our data set, we identified F-κB sites ($n=4$; 6%) at position II and ($n=1$; 1%) at position I among 63 sequences analyzed. The majority of the sequences identified with H-κB at position II ($n=50$; 79%) and position I ($n=55$; 87%). Forty-nine ($n=49$; 78%) sequences were found to exhibit C-κB site. ZA_LTR052 was identified with a single point mutation. We identified all three NF-κB-binding sites in ($n=44$; 70%) the viral promoter–enhancer regions in South African patients.

Keywords: HIV-1, subtype-C, transcription factors NF-κB, phylogenetic analyses, South Africa

THE LENTIVIRUS GENUS FAMILY (such as HIV-1, HIV-2, SIV, BIV, FIV, and visna virus) share a common replicative cycle, which is regulated by several key events for the successful establishment of infection and disease progression.[1] These events begin with viral-host binding and entry, followed by reverse transcription of the RNA, integration of the resulting pro-viral DNA into the host genome, gene expression, virus assembly, budding, and maturation of new virus. Several of these events are monitored and controlled by the activity of the long terminal repeat (LTR). However, the configuration changes in the region of LTR during the evolution of quasi-species results in (1) altering inducible rates of viral gene expression[2] and (2) the transition from a latency-like level of viral expression to a highly productive infection or vice versa.[3]

A set of transcription factors binding sites (such as Sp1 sites, a TATA box, the TAR region, and a strong enhancer composed of κB sites) within the LTR of viruses play a critical role in promoter activity and is important for optimal HIV-1 replication.[4] Distinct differences in the composition of the NF-κB motifs are important for gene expression regulation from the viral promoter. NF-κB selectively recognizes a 10-bp stretch of consensus DNA sequence 5′-GGGRNYY YCC-3′.[5] Alterations in the NF-κB sites results in a loss of LTR promoter activity in transient transfection with LTR-CAT constructs.[6] The two adjacent NF-κB binding sites were reported to be common and are highly conserved in the HIV-1 LTR among different viral subtypes. Unlike other HIV-1 subtypes, there is a considerable variation in the number of NF-κB sites in HIV-1 subtype C (HIV-1C) sequences. Subtype C LTRs with different numbers of κB sites have been reported to show a difference in transactivation activity.[7]

South Africa experiences the largest HIV-1 epidemic with 19% (7.2 million) of the global number of people living with the virus (UNAIDS 2018). With this study we aim to contribute to the knowledge diversity of differences in regulatory motifs within the LTR that may influence viral fitness by acting at the level of transcription.

Ethical permission for this study was obtained from the Health Research Ethics Committee of Stellenbosch University (N15/08/071). The laboratory experiments were conducted according to the ethical guidelines and principles

[1]Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
[2]Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, University of Stockholm, Stockholm, Sweden.
[3]HIV/AIDS Division, Department of Clinical Research, National Institute for Research in Tuberculosis, Indian Council of Medical Research, Chennai, India.

of the international Declaration of Helsinki 2013, South African Guidelines for Good Clinical Practice, and the Medical Research Council (MRC) Ethical Guidelines for Research. A waiver of consent was awarded to conduct sequence analyses on these samples.

A total of 63 samples were included in this study. A single vial of 6–8 mL of plasma was collected from each of the participants who acquired infection through horizontal transmission. Samples were collected between 2017 and 2018. The patient information was anonymized and delinked before the experiment.

Viral RNA was extracted from patient's plasma using the QIAamp viral RNA mini kit (Qiagen GmbH, Germany), following the manufacturer's instructions. cDNA was synthesized as previously described.[8] In brief, a ±474 base pairs (bps) region spanning ~150 bps of the *nef* gene and most of the U3 region of the LTR was amplified using a heminested PCR. The first round of PCR was performed with LTRF1 (5′CCAGTCAGACCTCAGGTGCCTTTAAGACCAATG AC3′) and 9555R/LTRR (5′-TCTACCTAGAGAGAC CCAGTACA3′) primers (HXB2 positions 9001–9035 and 9055–9533) followed by a second round with primers

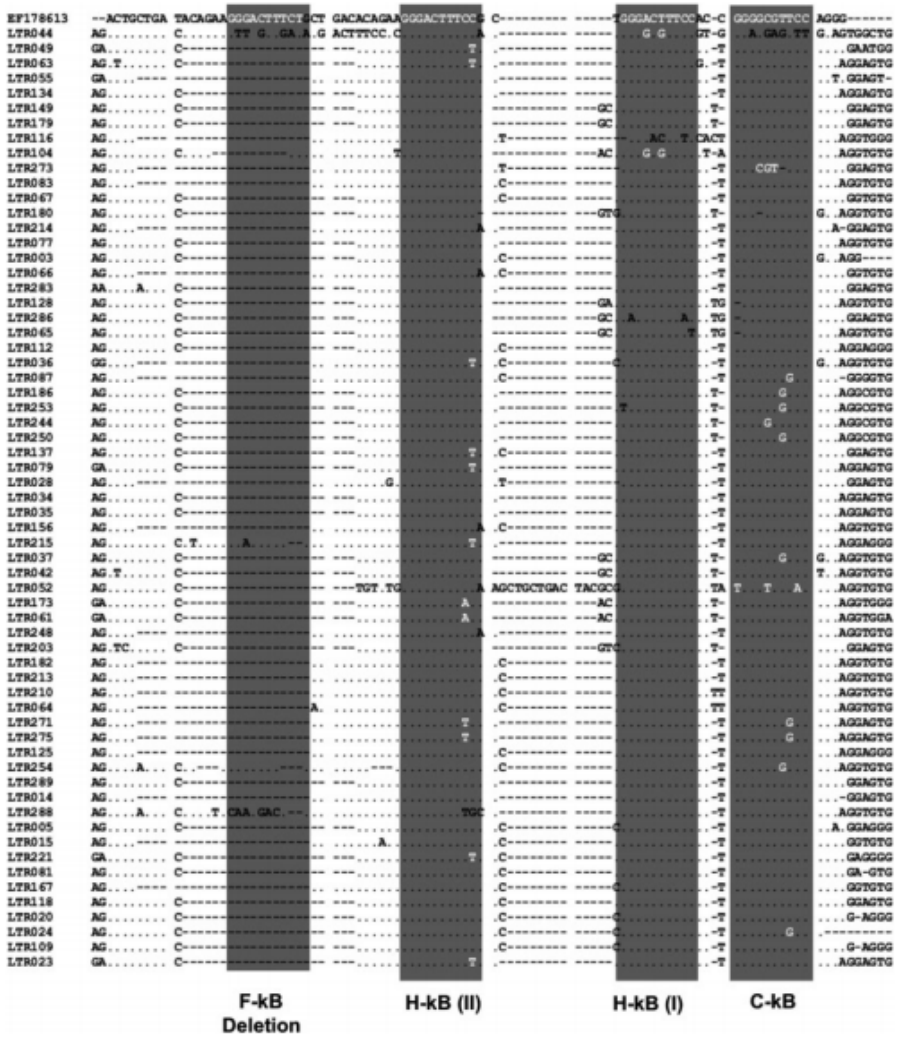


FIG. 1. The number of κB sites present in each sequence and also variations in NF-κB sites. Multiple sequence alignment of the LTR and schematic representations of transcription binding sites including the deletion (F-κB), position I and II (H-κB) of HIV-1 subtype C LTR. LTR, long terminal repeat.

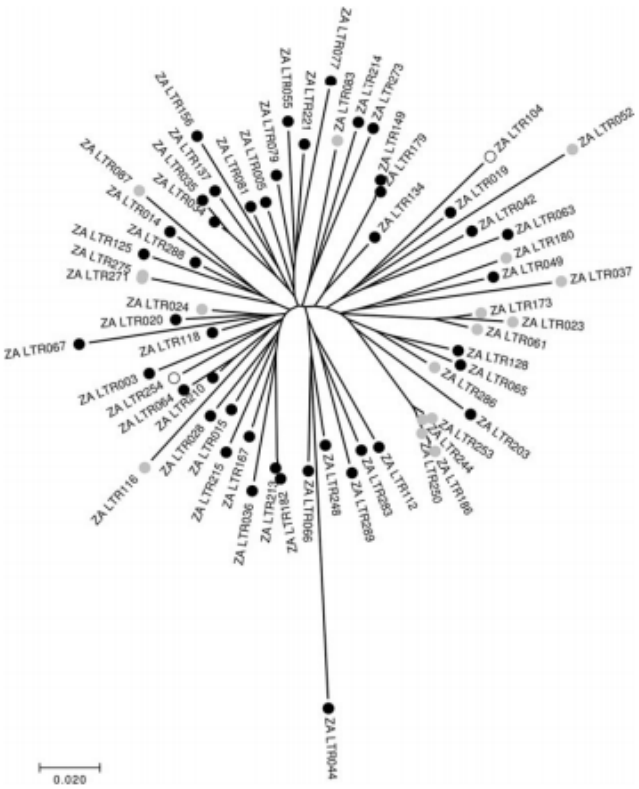


FIG. 2. Molecular phylogeny for κB sites of LTR sequences. Molecular phylogenetic analysis of LTR sequences of the 63 horizontally transmitted viruses. The phylogenetic tree was constructed from amino acid alignments using the neighbor-joining method. The sequences with one, two, and three κB site are denoted as *empty*, *gray-filled*, and *black-filled circles*, respectively. The tree was drawn to scale, with branch lengths measured as number of substitutions per site. The number of bootstrap replications was set at 100.

LTRF2 (5′GCTTCTACGCGTAAGAAAAGGGGGGACT GGA) (HXB2 position 9059–9089) and 9555R/LTRR primers. Cycling conditions were at 94°C for 2 min, 50°C for 2 min, and 72°C for 4 min, followed by 25 cycles of 94°C for 45 s, 50°C for 1 min, 72°C for 90 s, and a final extension at 72°C for 7 min.

The LTR sequences were edited based on the purity of the chromatogram peaks using the seqscape assembly program, version 2.5 (Thermofischer), with prescribed default parameters and were aligned against a South African subtype C sequence (AF388970.1) reported to have four κB sites (Fig. 2). The κB site that recognizes GGGACTTTCC, GGGACTTTCT, and GGGGCGTTCC were defined as H, F, and C-κB.[7] A multiple sequence alignment was done with MEGA version 7. Sequences were quality checked with the Los Alamos National Laboratory (LANL) database (www.hiv.lanl.gov/content/sequence/QC/index.html, accessed on the November 20, 2018). HIV-1 subtyping was performed using the online HIV-1 viral subtyping and recombinant detection tool, RIP (www.hiv.lanl.gov/content/sequence/RIP/RIP.html, accessed on November 20, 2018).

We identified three types of κB sites in the viral promoter region (Fig. 1), namely F-κB, H-κB (I and II), and C-κB in our study population sequences. We did not identify any F-κb sites in the F-κB site region; however, F-κB has been reported in most of the Indian HIV-1C sequence, leading to higher titer of infectivity. We identified 4/63 (6%) F-κB sites at position II and 1/63 (1%) at position I among the 63 sequences analyzed. The majority of the sequences were identified with H-κB at position II ($n=50$; 79%) and at position I ($n=55$; 87%). Forty-nine (49; 78%) sequences were found to exhibit C-κB site. ZA_LTR052 was identified with a single point mutation, but it resembles C-κB like (Fig. 1).

We observed variations in the NF-κB sites in 35/63 (56%) viral LTR sequences (Fig. 1). We used HXB2 as our reference strain. F-κb site (5′ GGGACTTTCT-3′), position 1, 2 represent H-κb site (5′-GGGACTTTCC-3′), and position 4 represent F-κb site (5′-GGGGCGTTCC-3′). At position 2, we observed that 17/63 (27%) viral sequences had variations. Among the 17 viral sequences, 2 viral sequences each at 9479 "T to C" and "A to C" variation, 5 viral sequences at 9481 had "A to G" variation, 7 viral sequences at 9480 had "T to C" variation. (Fig. 1). At position 3, we observed 5/63 (8%) viral sequences had variations: two viral sequences each had "G to C" and "G to T" variations at 9502 and 9504, respectively. One viral sequence had two variations at positions 9500 and 9509. One viral sequence had a "T to G" change at position 9499 HXB2 coordinate (Fig. 1). At position 4, we observed that 13/63 (21%) viral sequences had variations: One and 3 variations were observed in 10 and 2 viral

sequences, respectively. Viral sequences with three or more variations were not considered as transcription factors binding sites.

Sixty-three (63) sequences from HIV-1-infected South African patients were determined to be HIV-1C using the LTR region. The spread of HIV-1C in the heterosexual population in South Africa has been on the increase since it was first identified in the late 1980s.[9] Since the epidemic switch, HIV-1C has been in the most predominant subtype in sub-Saharan Africa and responsible for over 50% of the global subtype.[10] Studies have described that HIV-1C shown to have additional NF-κB or "NF-κB-like" as compared with other HIV-1 subtypes.[11] NF-κB plays a key regulator role in host immune and antiviral responses; HIV-1 uses NF-κB for effective transcription of viral genes.[12]

The majority of HIV-1 subtypes only have two NF-κB sites, except for members of subtype A and CRF01_AE viruses, in which only one NF-κB site is generally observed.[13] However, HIV-1C has been shown to contain three or more NF-κB sites, which has been hypothesized as to why HIV-1C spreads more rapidly than its other counterpart.[13] We identified a single variant change in our viral sequences κB site our finding are in agreement with previous studies that reported single mutation in κB site might be intact in binding toward transcription factor.[7,14]

Of the 63 LTR sequences described, 42 (67%) contained an intact third NF-κB site. We did not observe any of the viral sequence to have a fourth NF-κB site. Our findings are consistent with Hunt *et al.* that reported the presence of a third NF-κB site in South African HIV-1C sequences in the LTR region.[15] Bachu *et al.* observed four functional NF-κB binding sites in the South African HIV-1C-infected patients as opposed to our findings. It is evident that the presence of the additional fourth-κB LTR manifested significantly higher transcription as compared with the three-κB LTR.[7]

We also observed two NF-κB sites from 18/63 (29%) LTR sequences analyzed. Studies have shown that deletions of either one or two NF-κB sites from the enhancer region can lead to substantially decreased replication levels and that the virus becomes less infective.[16] Our study had the following limitations that merit mention. The analyzed LTR sequences were partial and were obtained from RNA. We did not perform any infectivity assay to confirm the true effect of the variations in the transcription factor binding site.

In summary, we report that South African HIV-1C contain C-κB-binding site in their LTR regions, but no evolution of fourth NF-κB as reported previously. Viral fitness may be influenced by differences in regulatory motifs by acting at the level of transcription level. Majority of our sequences have additional NF-κB sites signature for HIV-1C viruses; however, the potential effect of a third NF-κB site has on HIV-1C replicative capacity and rate of progression in people living with HIV-1C in South Africa remain to be established.

Acknowledgments

This study was funded by the National Research Foundation (NRF) of South Africa, the Poliomyelitis Research Foundation (PRF) of South Africa and the National Heath Laboratory Service (NHLS) Research Trust. We also thank Stellenbosch University and the Harry Crossley Foundation for additional financial support. UN is supported by a grant from the Swedish Research Council (2017-01330). We acknowledge Mathilda Claassen for sample collection. We acknowledge Shambhu Prasad Ganeshappa Aralaguppe for performing the sequencing experiments. GenBank number: MK585135-MK585197.

Authors' Contributions

A.E.O. performed laboratory experiments, detailed sequence analyses, and wrote the first draft of the article. M.A. performed sequence analyses and helped with article draft. U.N. and G.B.J. conceptualized the study. All authors read and approved the final article.

Author Disclosure Statement

No competing financial interests exist.

References

1. Bose D, Gagnon J, Chebloune Y: Comparative analysis of Tat-dependent and Tat-deficient natural lentiviruses. Vet Sci 2015;2:293–348.
2. Sgarbanti M, Remoli AL, Marsili G, *et al.*: IRF-1 is required for full NF-kappaB transcriptional activity at the human immunodeficiency virus type 1 long terminal repeat enhancer. J Virol 2008;82:3632–3641.
3. Wang XF, Liu Q, Wang YH, *et al.*: Characterization of equine infectious anemia virus long terminal repeat quasispecies in vitro and in vivo. J Virol 2018;92: e02150–17.
4. Quivy V, Adam E, Collette Y, *et al.*: Synergistic activation of human immunodeficiency virus type 1 promoter activity by NF-kappaB and inhibitors of deacetylases: Potential perspectives for the development of therapeutic strategies. J Virol 2002;76:11091–11103.
5. Wan F, Lenardo MJ: Specification of DNA binding activity of NF-kappaB proteins. Cold Spring Harb Perspect Biol 2009;1:a000067.
6. Verhoef K, Sanders RW, Fontaine V, Kitajima S, Berkhout B: Evolution of the human immunodeficiency virus type 1 long terminal repeat promoter by conversion of an NF-kappaB enhancer element into a GABP binding site. J Virol 1999;73:1331–1340.
7. Bachu M, Yalla S, Asokan M, *et al.*: Multiple NF-kappaB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. J Biol Chem 2012;287: 44714–44735.
8. Neogi U, Engelbrecht S, Claassen M, *et al.*: Mutational heterogeneity in p6 Gag late assembly (L) domains in HIV-1 subtype C viruses from South Africa. AIDS Res Hum Retroviruses 2016;32:80–84.
9. van Harmelen J, Wood R, Lambrick M, Rybicki EP, Williamson AL, Williamson C: An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. AIDS 1997;11:81–87.
10. Hemelaar J, Gouws E, Ghys PD, Osmanov S: Global trends in molecular epidemiology of HIV-1 during 2000–2007. AIDS 2011;25:679–689.
11. Rodenburg CM, Li Y, Trask SA, *et al.*: Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. AIDS Res Hum Retroviruses 2001;17:161–168.

12. Chan JK, Greene WC: Dynamic roles for NF-kappaB in HTLV-I and HIV-1 retroviral pathogenesis. Immunol Rev 2012;246:286–310.
13. Wallace A, West K, Rothman AL, Ennis FA, Lu S, Wang S: Post-translational intracellular trafficking determines the type of immune response elicited by DNA vaccines expressing Gag antigen of Human Immunodeficiency Virus Type 1 (HIV-1). Hum Vaccin Immunother 2013;9: 2095–2102.
14. Bachu M, Mukthey AB, Murali RV, *et al.*: Sequence insertions in the HIV type 1 subtype C viral promoter predominantly generate an additional NF-kappaB binding site. AIDS Res Hum Retroviruses 2012;28:1362–1368.
15. Hunt GM, Johnson D, Tiemesse CT: Characterisation of the long terminal repeat regions of South African human immunodeficiency virus type 1 isolates. Virus Genes 2001;23: 27–34.
16. Ilyinskii PO, Desrosiers RC: Efficient transcription and replication of simian immunodeficiency virus in the absence of NF-kappaB and Sp1 binding elements. J Virol 1996;70:3118–3126.

Address correspondence to:
Adetayo Emmanuel Obasa, MS
Division of Medical Virology
Department of Pathology
Faculty of Medicine and Health Sciences
Stellenbosch University
Tygerberg
Cape Town 7505
South Africa

E-mail: obasa@sun.ac.za

Chapter 8 General discussion and future remarks

8.1. Introduction

8.2. Introduction

The result chapters included in this thesis (two published manuscripts and two manuscripts under review) focused on the use of molecular techniques to analyse InSTIs in two cohorts (treatment-naïve and treatment-experienced patients). In the treatment-naïve cohort, we explored the structural impact of IN polymorphisms on the IN-reaction mechanism and InSTI susceptibility. In treatment-experienced patients, I focussed on the use of population-based Sanger sequencing and NGS to identify pre-existing mutations that could compromise InSTI rollout in PLHIV in South Africa.

8.3. Resistance testing and HIV drug resistance

RAMs against InSTIs and the efficacy of the use of DTG as a monotherapy have been established,[117] which suggest the use of a NRTI backbone must be considered. Currently, in LMICs, due to high cost, GRT drug resistance testing is not widely accessible and incomplete viral load testing could possibly lead to functional DTG monotherapy resistance in different situations. In Southern Africa, where long delays in the management of viraemia have been documented, this could allow NRTI resistance to emerge in patients receiving DTG.[118,119] Furthermore, adults with pre-treatment RAMs (especially treatment-experienced patients) might have undetected NRTI mutations associated with reduced susceptibility to TDF and 3TC.[120] Several studies have identified a high prevalence of M184V in adults with virological failure receiving first-line NNRTI-regimens.[23,27,121] This suggests that patients who switch from TDF, ETR and EFV to a combination of TDF, 3TC and DTG without prior confirmation of viral load suppression might be at increased risk of acquired HIVDR to DTG.[122] Knowledge of the risks and benefits of this strategy should be considered as a possible high-priority research question as several cART programmes might not use viral load testing in transition to DTG-based first-line cART. Although this transition could be justified by the advantages of DTG, programmes to accelerate viral load testing rollout should be supported, policy makers should be guided and HIVDR surveillance systems should be implemented across all public sectors.[122]

8.4. Management of viraemia in patients

In LMICs, the transition from EFV to a more efficacious InSTIs-based first-line cART is widely expected to increase levels of viral suppression. Challenges, such as inadequate retention in care, funding constraints, overworked healthcare personnel and inadequate healthcare

structures might reduce the effectiveness of this treatment regimen. Therefore, to measure the effectiveness of InSTIs in expanding viral load, coverage remains essential to identify patients who need urgent intervention because of viraemia. Furthermore, viraemia detection strategies for viral load monitoring should be reassessed for better management of patients who are failing treatment regimens.[123] The development of RAMs in patients receiving DTG-based regimens has been rare in both low- and high-income countries.[44,54] In some LMICs, resistance testing is not cost-effective and as a result is not widely used. If viraemia persists for more than three months regardless of adequate adherence, the WHO guidelines recommend switching to second-line cART. However, in LMICs these guidelines are not adequately implemented, leading to drug resistance, ongoing viraemia and morbidity.[118,120] The introduction of first-line DTG could allow these guidelines to be simplified as poor adherence rather than HIVDR is likely to cause viraemia, which should greatly reduce the number of patients switching from first-line DTG to a second-line regimen. Management of viraemia should focus more on improving adherence than on treating HIVDR.

8.5. Adherence interventions

In LMICs, numerous adherence interventions have been explored, but evidence of an effect on virological suppression is weak. Accurate measurement of adherence is challenging.[124] Furthermore, adherence relies on a complex combination of behavioural, social medical and structural factors.[125] For instance, HIVDR and not poor adherence, could cause a patient receiving NNRTIs-based first-line cART to have virological failure; however, a suppressed viral load could suggest good adherence. DTG as compared to EFV has a short half-life.[126] A patient receiving DTG, who does not adhere to her/his treatment, could experience quicker viral rebound. Improved adherence should reflect in a rapid return to viral suppression. As the use of InSTIs is set to increase, viral load testing coverage should also be expanded in LMICs, which would make monitoring adherence and response to adherence interventions easier. This strategy might help to develop and implement evidence-based adherence interventions that would help to achieve early viral suppression.[127] In South Africa, point-of-care viral load testing might be instrumental in helping to solve adherence challenges, monitor the effect of counselling and to guide patients' regimen diagnosis in differentiated care services.[128]

8.6. Concerns and questions around DTG roll-out in low-to-middle-income countries

The roll out of DTG as part of the first-line cART is arguably a significant step in HIV treatment evolution in LMICs. DTG has already been introduced in countries such as Tanzania, Nigeria, Malawi, Botswana and South Africa. This could be a major improvement in HIV care

in LMICs, but specific questions remain unanswered. First, there are still safety concerns for pregnant women receiving DTG and/or women in childbearing age with adverse birth outcomes [128.] According to the WHO, in the recently concluded International AIDS conference showed additional data from two clinical trials in Africa that compared the safety of DTG and EFV. The risk of neural tube defects are significantly lower than what was previously reported (IAS, 2019). Second, how will adherence support interventions be integrated with viral load monitoring to maximise virological re-suppression and effectively monitor patients with viraemia on DTG. Third, how will pre-treatment or acquired mutations that confer resistance to NRTIs affect the efficacy of DTG-based regimen either in first and/or second-line regimen? Fourth, will the replacement of first-line or second-line regimens or both with DTG be cost-effective? Will the roll out of DTG affect the cost-effectiveness of HIV drug resistance testing and viral load monitoring? Finally, without proper adherence, viral load monitoring and/or drug resistance testing, there will be high chance of evolution of DTG RAMs as observed in high-income countries. There is clear benefit of DTG when there is proper clinical management with optimal adherence, viral load monitoring and resistance testing. Data is lacking about the clinical evolution of DTG resistance in public health setting for this massive roll-out of DTG. Improper management of DTG could jeopardize the future use of newer InSTIs like CAB or BTG in the settings. These gaps should be addressed as part of DTG roll-out particularly regarding safety surrounding pregnant women and women of childbearing age and HIV drug resistance surveillance would assist health systems and clinicians to maximise the promising benefits of InSTIs regimens [128.]

8.7. Conclusion

The study addressed the existing knowledge gaps in the prevalence of InSTI drug resistance in South African treatment-naïve patients. The study also shows treatment-experienced patients suspected of failing a first- and/or second-line treatment regimen. The data generated from this study represent baseline findings regarding InSTI RAMs for both treatment-naïve and treatment-experienced patients in a South African context. These findings are important to verify whether the new InSTI regimens can ensure a supressed viral load and good treatment outcome for PLHIV in South Africa.

Below are the main conclusions:

- The introduction of the InSTIs class of cART regimen into the national treatment programme could potential help in the management of the HIV epidemic in South Africa. In the absence of a cure for HIV, long-term cART outcomes need to be

monitored closely for maximum efficiency to ensure that patients receive the best treatment. The possible emergence of RAMs as well as subtype-specific resistance to this class of cART regimen cannot be ignored.

- The pattern of RAMs identified was against PIs, NRTIs and NNRTIs from patients on the South African second-line national cART programme. Observed InSTI mutation levels were low. Major mutations, such as T66I and Y143R were identified at 1% each, and accessory mutations, such as E157Q and T97A were identified at 2% and 1% respectively. Due to the high public health burden, I strongly recommend continued viral load monitoring in South Africa. This will not only detect treatment failure earlier but will also detect poor treatment adherence.
- We presented extensive analyses of IN polymorphisms in the IN-reaction mechanism and InSTI susceptibility. These analyses suggest that naturally occurring polymorphisms may affect the structural stabilities of the IN gene, viral DNA binding and drug propensity. The data generated from this study provide guidance for investigating how polymorphism can affect treatment response.
- In this study, I used HTS to quantify the DRM in both minor (< 20% of the population) and major (> 20% of the population) viral quasi-species and identified increased PI RAM in minor viral populations. This study shows that the use of high-throughput resistance testing for GRT can greatly improve the identification of PI RAMs in bPI-failing patients. Using HTS-GRT, PI RAMs (V82A) and RTI RAMs (K65R, M184V or K103N) were identified in < 20% of the population that Sanger sequencing failed to identify, strengthening their role in detecting the acquired mutations early.

In this study, we observed two NF-kB sites from 18/63 (29%) LTR sequences analyzed. Studies have shown that deletions of either one or two NF-kB sites from the enhancer region can lead to substantially decreased replication levels and that the virus becomes less infective. We report that South African HIV-1C contain C-kB-binding site in their LTR regions, but no evolution of fourth NF-kB. Viral fitness may be influenced by differences in regulatory motifs by acting at the transcription level. Majority of our sequences have additional NF-kB sites signature for HIV-1C viruses; however, the potential effect of a third NF-kB site has on HIV-1C replicative capacity and rate of progression in people living with HIV-1C in South Africa remain to be established.

Baseline resistance testing and quantification of drug-resistant viral populations before the initiation of cART reduce the probability of developing virological treatment failure. In countries such as South Africa, baseline resistance testing is expensive as compared to high-income countries. In addition to the, South Africa guidelines recommend baseline resistance testing to patients for the following reasons: pre-exposure prophylaxis in the last six months, sexual exposure to a person with known drug-resistant HIV and patients who are known to have failed cART regimens. In South Africa, population-based Sanger sequencing has been used as genotype testing tool when a patient has met any of the required recommendations. We also propose the use of HTS.

8.8. Overall strength and limitations of the work

There are several strengths and limitations to this work. Samples were stored since 2001, as they were obtained between 2000 and 2001 (**Paper I**). These samples are unique because they were obtained before cART roll-out in South Africa. Despite this advantage, our study was limited to the number of samples that were analysed. No viral load data is available and $CD4^+$ T cell analysis was not performed on these samples. This is the first study to look at the prevalence of InSTIs in patients receiving LPV/r and ATV as their bPIs from South Africa, this manuscript therefore contributed to expanding current HIV-1 knowledge regarding LPV/r and ATV resistance. The majority of our patients were receiving LPV/r as their bPIs compared to ATV. We cannot tell whether the patients having RAL resistance, according to the sequences, have had access to a RAL-based treatment regimens. No adherence data was available for these patients. This manuscript provided the first glimpse of nucleoprotein organisation in viral sequences obtained from South African treatment-naïve patients and different subtypes. Data generated from this manuscript could be used to deduce the effect of a polymorphism in the IN of different HIV-1 subtypes. The limitation for this manuscript is the DNA used in this structure determination contains a T–T mismatch in the double-stranded region (Paper IV). This mismatch may have affected DNA bending, which in turn would have affected spatial oligomerization of IN molecules in the intasome.

This study also shows the use of HTS resistance testing for GRT can greatly improve the identification of PI RAMs in bPI failing patients. Using HTS-GRT, PI RAMs (V82A) and RTI RAMs (K65R, M184V or K103N) were identified in <20% population that sanger-based sequencing failed to identify strengthening their role in detecting the acquired mutations early (**Paper II and IV**). In Chapter 2, there were no major mutations against InSTIs detected (prior to 2004); in chapter 3, 3.1% (3 of 96) had a major InSTI mutation by Sanger sequencing (2017

– 2018), and 5/56 (8.9%) patients harboured 3 virus with high level resistance to InSTIs as detected by NGS (Chapter 4; 2017 – 2018). This is an evolving epidemic, it would be critical to keep monitoring patients for therapy outcomes, when switched to InSTIs treatment regimen.

References

1. Hemelaar, J., Gouws, E., Ghys, P. D., Osmanov, S. & WHO-UNAIDS Network for HIV Isolation and Characterisation. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* **25**, 679–89 (2011).

2. Vidal, N. *et al.* Unprecedented Degree of Human Immunodeficiency Virus Type 1 (HIV-1) Group M Genetic Diversity in the Democratic Republic of Congo Suggests that the HIV-1 Pandemic Originated in Central Africa. *J. Virol.* **74**, 10498–10507 (2000).

3. UNAIDS. 2017 Global HIV Statistics. 5 (2018).

4. Trickey, A. *et al.* Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV* **4**, e349–e356 (2017).

5. Rousseau, C. M. *et al.* Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *J. Virol. Methods* **136**, 118–125 (2006).

6. UNAIDS Country Report 2018. South Africa | UNAIDS.

7. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, S. A. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med.* **305(24):**,.

8. Gottlieb, M. S., Schanker, H. M., Fan, P. T., Saxon, A. & Weisman, J. D. *Pneumocystis Pneumonia*. (1981).

9. F Barre-Sinoussi, JC Chermann, F Rey, MT Nugeyre, S Chamaret, J Gruest, C Dauguet, C Axler-Blin, F Vezinet-Brun, C Rouzioux, W Rozenbaum, L. M. REPORTS Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (80-.).* **220**, 868–871 (1983).

10. Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V. S., Mann, D., Sidhu, G. D., Stahl, R. E., Zolla-Pazner, S., Leibowitch, J., and Popovic, M. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science (80-.).* **220**, 865–7 (1983).

11. Popovic, M., Sarngadharan, M. G., Read, E., and Gallo, R. C. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science (80-.).* **224**, 497–500 (1984).

12. Zhu T *et al.* An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**, 594–597 (1998).

13. Worobey, M. *et al.* Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**, 661–664 (2008).

14. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* **1**, 1–22 (2011).

15. Novitsky, V. A. *et al.* Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana. *J. Virol.* **73**, 4427–32 (1999).

16. Girard, M. P., Osmanov, S., Assossou, O. M. & Kieny, M. P. Human immunodeficiency virus (HIV) immunopathogenesis and vaccine development: A review. *Vaccine* **29**, 6191–6218 (2011).

17. Freed, E. O. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* **13**, 484–496 (2015).

18. Gleenberg, I. O., Herschhorn, A. & Hizi, A. Inhibition of the Activities of Reverse Transcriptase and Integrase of Human Immunodeficiency Virus Type-1 by Peptides Derived from the Homologous Viral Protein R (Vpr). *J. Mol. Biol.* **369**, 1230–1243 (2007).

19. de Oliveira, T. *et al.* Variability at Human Immunodeficiency Virus Type 1 Subtype C Protease Cleavage Sites: an Indication of Viral Fitness? *J. Virol.* **77**, 9422–9430 (2003).

20. Tang, M. W. & Shafer, R. W. HIV-1 Antiretroviral Resistance. *Drugs* **72**, e1–e25 (2012).

21. Menéndez-Arias, L. Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. *Antiviral Res.* **98**, 93–120 (2013).

22. Johnson, V. A. *et al.* Update of the drug resistance mutations in HIV-1: March 2013. *Top. Antivir. Med.* **21**, 6–14 (2013).

23. Van Zyl, G. U. *et al.* Trends in Genotypic HIV-1 Antiretroviral Resistance between 2006 and 2012 in South African Patients Receiving First-and Second-Line Antiretroviral Treatment Regimens. (2013) doi:10.1371/journal.pone.0067188.

24. De Luca, A. *et al.* Declining prevalence of HIV-1 drug resistance in antiretroviral treatment-exposed individuals in Western Europe. *J. Infect. Dis.* **207**, 1216–1220 (2013).

25. Wensing, A. M. *et al.* 2017 Update of the Drug Resistance Mutations in HIV-1. *Top. Antivir. Med.* **24**, 132–133 (2014).

26. Cozzi-Lepri, A. *et al.* Thymidine analogue mutation profiles: Factors associated with acquiring specific profiles and their impact on the virological response to therapy. *Antivir. Ther.* **10**, 791–802 (2005).

27. Neogi, U. *et al.* Mutational Heterogeneity in p6 Gag Late Assembly (L) Domains in HIV-1 Subtype C Viruses from South Africa. *AIDS Res. Hum. Retroviruses* **32**, 80–84 (2016).

28. Sadler, H. A., Stenglein, M. D., Harris, R. S. & Mansky, L. M. APOBEC3G Contributes to HIV-1 Variation through Sublethal Mutagenesis. *J. Virol.* **84**, 7396–7404 (2010).

29. Neogi, U. *et al.* Human APOBEC3G-mediated hypermutation is associated with antiretroviral therapy failure in HIV-1 subtype C-infected individuals. *J. Int. AIDS Soc.* **16**, 1–8 (2013).

30. Mulder, L. C. F., Harari, A. & Simon, V. Cytidine deamination induced HIV-1 drug resistance. *Proc. Natl. Acad. Sci.* **105**, 5501–5506 (2008).

31. N., M. *Resistance to non-nucleoside reverse transcriptase inhibitors. In: Geretti AM, editor. Antiretroviral Resistance in Clinical Practice. London: Mediscript; 2006. Chapter 2. Available from: https://www.ncbi.nlm.nih.gov/books/NBK2249/*. (2006).

32. Asahchop, E. L. *et al.* Distinct resistance patterns to etravirine and rilpivirine in viruses containing nonnucleoside reverse transcriptase inhibitor mutations at baseline. *Aids* **27**, 879–887 (2013).

33. Neogi, U. *et al.* Research letter: Novel tetra-peptide insertion in Gag-p6 ALIXbinding motif in HIV-1 subtype C associated with protease inhibitor failure in Indian patients. *Aids* **28**, 2319–2322 (2014).

34. Scherrer, A. U. *et al.* Minor protease inhibitor mutations at baseline do not increase the risk for a virological failure in HIV-1 subtype B infected patients. *PLoS One* **7**, (2012).

35. Briz, V., Poveda, E. & Soriano, V. HIV entry inhibitors: Mechanisms of action and resistance pathways. *J. Antimicrob. Chemother.* **57**, 619–627 (2006).

36. Espeseth, A. S. *et al.* HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc. Natl. Acad. Sci.* **97**, 11244–11249 (2002).

37. Hazuda, D. J. *et al.* Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science (80-.).* **287**, 646–650 (2000).

38. Hazuda, D. J. HIV integrase as a target for antiretroviral therapy. *Curr. Opin. HIV AIDS* **7**, 383–389 (2012).

39. Wong, E., Trustman, N. & Yalong, A. HIV pharmacotherapy. *J. Am. Acad. Physician Assist.* **29**, 36–40 (2016).

40. AIDSinfo. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV. *Dep. Heal. Hum. Serv.* 298 (2018).

41. Battegy, M., Lundgren, J. D. & Ryom, L. Eacs Guidelines Version 8.2. 97 (2017).

42. WHO. *Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection*. *World Health Organization. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection Recommendations for a Public Health Approach—Second edition. 2016. http://www.who.int/hiv/pub/arv/arv-2016/en/.* (2016). doi:10.1097/00022744-199706000-00003.

43. WHO. *The public health response to pretreatment HIV drug resistance*. *WHO Guidelines* (2017).

44. Lepik, K. J. *et al.* Emergent drug resistance with integrase strand transfer inhibitor-based regimens. *Aids* **31**, 1425–1434 (2017).

45. Hurt, C. B., Sebastian, J., Hicks, C. B. & Eron, J. J. Resistance to HIV integrase strand transfer inhibitors among clinical specimens in the united states, 2009-2012. *Clin. Infect. Dis.* **58**, 423–431 (2014).

46. Rhee, S.-Y. *et al.* Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology* **5**, 74 (2008).

47. Bar-Magen, T. *et al.* HIV-1 subtype B and C integrase enzymes exhibit differential patterns of resistance to integrase inhibitors in biochemical assays. *Aids* **24**, 2171–2179 (2010).

48. Quashie, P. K., Han, Y.-S., Hassounah, S., Mesplède, T. & Wainberg, M. A. Structural Studies of the HIV-1 Integrase Protein: Compound Screening and Characterization of a DNA-Binding Inhibitor. *PLoS One* **10**, e0128310 (2015).

49. Quashie, P. K. *et al.* Differential Effects of the G118R, H51Y, and E138K Resistance Substitutions in Different Subtypes of HIV Integrase. *J. Virol.* **89**, 3163–3175 (2015).

50. Mesplède, T. & Wainberg, M. A. Resistance against integrase strand transfer inhibitors and relevance to HIV persistence. *Viruses* **7**, 3703–3718 (2015).

51. Malet, I. *et al.* Short communication Genetic barriers for integrase inhibitor drug resistance in HIV type-1 B and CRF02 _ AG subtypes. *Antivir. Ther.* 123–129 (2009).

52. Brado, D. *et al.* Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and availability of integrase inhibitors in Cape Town, South Africa. *Sci. Rep.* **8**, 4709 (2018).

53. Das, A. T. crossm How Polypurine Tract Changes in the HIV-1 RNA Genome Can Cause Resistance against the Integrase Inhibitor Dolutegravir. 10–13 (2018).

54. Anstett, K., Brenner, B., Mesplede, T. & Wainberg, M. A. HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology* **14**, 1–16 (2017).

55. Russel, F. G. M., Koenderink, J. B. & Masereeuw, R. Multidrug resistance protein 4 (MRP4 / ABCC4): a versatile efflux transporter for drugs and signalling molecules. **4**, (2008).

56. Hurwitz, S. J. & Schinazi, R. F. Practical considerations for developing nucleoside reverse transcriptase inhibitors. *Drug Discov. Today Technol.* **9**, e183–e193 (2012).

57. Clercq, E. D. E. Antiviral Therapy for Human Immunodeficiency Virus Infections. **8**, 200–239 (1995).

58. Sluis-cremer, N. & Tachedjian, G. Mechanisms of inhibition of HIV replication by non-nucleoside reverse transcriptase inhibitors. **134**, 147–156 (2008).

59. Usach, I. & Melis, V. Review article Non-nucleoside reverse transcriptase inhibitors : a review on pharmacokinetics , pharmacodynamics , safety and tolerability. 1–14 (2013) doi:10.7448/IAS.16.1.18567.

60. Figueiredo, A. *et al.* Potent Nonnucleoside Reverse Transcriptase Inhibitors Target HIV-1 Gag-Pol. **2**, (2006).

61. He, Y. *et al.* Enhancement of cellular uptake , transport and oral absorption of protease inhibitor saquinavir by nanocrystal formulation. *Nat. Publ. Gr.* 1151–1160 (2015) doi:10.1038/aps.2015.53.

62. Lefebvre, E. & Schiffer, C. A. NIH Public Access. **10**, 131–142 (2009).

63. Adamson, C. S. Protease-Mediated Maturation of HIV : Inhibitors of Protease and the Maturation Process. **2012**, (2012).

64. Casadellà, M. *et al.* Primary resistance to integrase strand-transfer inhibitors in Europe: Table 1. *J. Antimicrob. Chemother.* **70**, 2885–2888 (2015).

65. UNAIDS. *UNAIDS. Country Report South Africa. Available at: http://www.unaids.org/en/regionscountries/countries/southafrica. (2018).* (2018).

66. Meintjes, G. *et al.* Adult antiretroviral therapy guidelines 2017. *South. Afr. J. HIV Med.* **18**, 24 pages (2017).

67. Brady, M., Fidler, S., Oxenius, A., Phillips, R. & Weber, J. Short Course Antiretroviral Therapy in Primary Hiv-1 Infection. *Clin. Sci.* **103**, 25P.3-26P (2015).

68. Kai Deng & Robert F. Siliciano. Early treatment may Tooth structure re - engineered. *Nature* **512**, 35–36 (2014).

69. Sandler, N. G. & Sereti, I. Can ART treatment reduce long-term complications by reducing inflammation? *Curr. Opin. HIV AIDS* **9**, 72–9 (2014).

70. Stanley, S. . Centers for Disease Control & Prevention (CDC) Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents Author (s): Sharilyn K . Stanley Source : Morbidity and Mortality Weekly Report : Recommendations and Reports , Vol . **47**, 43–82 (1998).

71. Grinsztejn, B. *et al.* Effects of early versus delayed initiation of antiretroviral treatment on clinical outcomes of HIV-1 infection: Results from the phase 3 HPTN 052 randomised controlled trial. *Lancet Infect. Dis.* **14**, 281–290 (2014).

72. To, W. & Consortium, S. Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: a collaborative analysis of 18 HIV cohort studies. *Lancet* **373**, 1352–1363 (2009).

73. WHO, 2013. WHO, Consolidated guidelines on general HIV care and the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. 2013. www.who.int. www.who.int.

74. UNAIDS 2014. *UNAIDS. Fast-Track: ending the AIDS epidemic by 2030. Geneva: JointUnited Nations Programme on HIV/AIDS; 2014. AIDS epidemic update* vol. 58 http://www.unaids.org/sites/default/files/media_asset/JC2686_WAD2014report_en.pdf

(2014).

75. Sáez-Cirión, A. *et al.* Post-Treatment HIV-1 Controllers with a Long-Term Virological Remission after the Interruption of Early Initiated Antiretroviral Therapy ANRS VISCONTI Study. *PLoS Pathog.* **9**, (2013).

76. Josefsson, L. *et al.* The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proc. Natl. Acad. Sci.* **110**, E4987–E4996 (2013).

77. Von Stockenstrom, S. *et al.* Longitudinal Genetic Characterization Reveals That Cell Proliferation Maintains a Persistent HIV Type 1 DNA Pool during Effective HIV Therapy. *J. Infect. Dis.* **212**, 596–607 (2015).

78. Cain LE, Logan R, Robins JM, Sterne JA, Sabin C, E. Al. When to Initiate Combined Antiretroviral Therapy to Reduce Mortality and AIDS-Defining Illness. *Ann. Intern. Med.* **154**, 509–515 (2011).

79. Cohen, M. et al. Prevention of HIV-1 Infection with Early Antiretroviral Therapy Myron. *N Engl J Med.* **365**, 493–505 (2012).

80. Hunt, G. M., Johnson, D. & Tiemessen C, T. Characterisation of the long terminal repeat regions of South African human immunodeficiency virus type 1 isolates. *Virus Genes* **23**, 27–34 (2001).

81. Montano, M. A. *et al.* Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. *J. Virol.* **71**, 8657–65 (1997).

82. Hunt, G. & Tiemessen, C. T. Occurrence of Additional NF- k B-Binding Motifs in the Long Terminal Repeat Region of South African HIV Type 1 Subtype C Isolates. *AIDS Res. Hum. Retroviruses* **16**, 305–6 (2000).

83. Verhoef, K., Sanders, R. W., Fontaine, V., Kitajima, S. & Berkhout, B. Evolution of the Human Immunodeficiency Virus Type 1 Long Terminal Repeat Promoter by Conversion of an NF-□B Enhancer Element into a GABP Binding Site. *J. Virol.* **73**, 1331–1340 (1999).

84. Bachu, M. *et al.* Multiple NF-κB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. *J. Biol. Chem.* **287**, 44714–44735 (2012).

85. O, H.-G. *et al.* The Replicative Fitness of Primary Human Immunodeficiency Virus Type

1 (HIV-1) Group The Replicative Fitness of Primary Human Immunodeficiency Virus Type 1 (HIV-1) Group M , HIV-1 Group O , and HIV-2 Isolates. *J. Virol.* **1**, 8979–8990 (2005).

86. Njai, H. F. *et al.* The predominance of Human Immunodeficiency Virus type I (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its replicative fitness. *Retrovirology* **3**, 1–11 (2006).

87. el Kharroubi, A. & Martin, M. A. Cis-Acting Sequences Located Downstream of the Human Immunodeficiency Virus Type 1 Promoter Affect Its Chromatin Structure and Transcriptional Activity. *Mol. Cell. Biol.* **16**, 2958–2966 (2015).

88. Cavasotto, C. N. & Phatak, S. S. Homology modeling in drug discovery : current trends and applications. **14**, (2009).

89. Wielens, J., Crosby, I. T. & Chalmers, D. K. A three-dimensional model of the human immunodeficiency virus type 1 integration complex. 301–317 (2005) doi:10.1007/s10822-005-5256-2.

90. Hare, S. *et al.* Structural and Functional Analyses of the Second-Generation Integrase Strand Transfer Inhibitor Dolutegravir (S / GSK1349572) □. **80**, 565–572 (2011).

91. Hare, S., Maertens, G. N. & Cherepanov, P. by retroviral integrase in crystallo. *EMBO J.* **31**, 3020–3028 (2012).

92. Hare, S. *et al.* Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc. Natl. Acad. Sci.* **107**, 20057–20062 (2010).

93. Dewdney, T. G., Wang, Y., Kovari, I. A., Reiter, S. J. & Kovari, L. C. Reduced HIV-1 integrase flexibility as a mechanism for raltegravir resistance. *J. Struct. Biol.* **184**, 245–250 (2013).

94. Hightower, K. E. *et al.* Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than raltegravir and elvitegravir from wild-type and integrase inhibitor-resistant HIV-1 integrase-DNA complexes. *Antimicrob. Agents Chemother.* **55**, 4552–4559 (2011).

95. Passos, D. O. *et al.* Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. **92**, 89–92 (2017).

96. Rogers, L. *et al.* Structural implications of genotypic variations in HIV-1 integrase from diverse subtypes. *Front. Microbiol.* **9**, 1–9 (2018).

97. Walker, C. M. *et al.* Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy. *Science (80-.).* **278**, (1997).

98. TAE-WOOK CHUN, LIEVEN STUYVER, STEPHANIE B. MIZELL, LINDA A. EHLER, JO ANN M. MICAN, MICHAEL BASELER, ALUN L. LLOYD, MARTIN A. NOWAK, A. A. S. F. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Med. Sci.* **94**, 13193–13197 (1997).

99. Wong, J. K. *et al.* Recovery of Replication-Competent HIV Despite Prolonged Suppression of Plasma Viremia Recovery of Replication-Competent HIV Despite Prolonged Suppression of Plasma Viremia. *Science (80-.).* **278**, 1291–1294 (1997).

100. Davey, R. T. *et al.* HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 15109–14 (1999).

101. Kuritzkes, D. R. *et al.* Performance Characteristics of the TRUGENE HIV-1 Genotyping Kit and the Opengene DNA Sequencing System. *J. Clin. Microbiol.* **41**, 1594–1599 (2003).

102. Eshleman, S. H. *et al.* Performance of the Celera Diagnostics ViroSeq HIV-1 genotyping system for sequence-based analysis of diverse human immunodeficiency virus type 1 strains. *J. Clin. Microbiol.* **42**, 2711–2717 (2004).

103. Obasa, A. E., Engelbrecht, S. & Jacobs, G. B. Near full-length HIV-1 subtype B sequences from the early South African epidemic, detecting a BD unique recombinant form (URF) from a sample in 1985. *Sci. Rep.* **9**, 1–7 (2019).

104. Rhee, S. *et al.* Predictive Value of HIV-1 Genotypic Resistance Test Interpretation Algorithms. *J. Infect. Dis.* **200**, 453–463 (2009).

105. Vandamme, A. M. *et al.* European recommendations for the clinical use of HIV drug resistance testing: 2011 update. *AIDS Rev.* **13**, 77–108 (2011).

106. Li, X. *et al.* Towards Clinical Molecular Diagnosis of Inherited Cardiac Conditions: A Comparison of Bench-Top Genome DNA Sequencers. *PLoS One* **8**, (2013).

107. Pallen, M. J. Reply to Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* **31**, 296–296 (2013).

108. Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439 (2012).

109. Ekici, H. *et al.* Cost-efficient HIV-1 drug resistance surveillance using multiplexed high-throughput amplicon sequencing: Implications for use in low- and middle-income countries. *J. Antimicrob. Chemother.* **69**, 3349–3355 (2014).

110. Aralaguppe, S. G. *et al.* Multiplexed next-generation sequencing and de novo assembly to obtain near full-length HIV-1 genome from plasma virus. *J. Virol. Methods* **236**, 98–104 (2016).

111. Telele, N. F. *et al.* Pretreatment drug resistance in a large countrywide Ethiopian HIV-1C cohort: A comparison of Sanger and high-throughput sequencing /631/326/2521 /631/337/151/1431 /38/23 /38/77 /38/90 /14/63 /38/43 /38/47 /42/40 article. *Sci. Rep.* **8**, 1–10 (2018).

112. Larder, B. A., Darby, G. & Richman, D. D. HIV with Reduced Sensitivity to Zidovudine (AZT) Isolated during Prolonged Therapy Author (s): Brendan A . Larder , Graham Darby and Douglas D . Richman Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.o. **243**, 1731–1734 (2017).

113. Mayers, D. L. *et al.* Dideoxynucleoside resistance emerges with prolonged zidovudine monotherapy. *Antimicrob. Agents Chemother.* **38**, 307–314 (1994).

114. Gervaix, A. *et al.* A new reporter cell line to monitor HIV infection and drug susceptibility in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4653–4658 (1997).

115. Chiba-Mizutani, T. *et al.* Use of new T-cell-based cell lines expressing two luciferase reporters for accurately evaluating susceptibility to anti-human immunodeficiency virus type 1 drugs. *J. Clin. Microbiol.* **45**, 477–487 (2007).

116. Baxter, J. D. *et al.* A randomized study of antiretroviral management based on plasma genotypic antiretroviral resistance testing in patients failing therapy. *Aids* **14**, 83–93 (2000).

117. Wijting, I. *et al.* Dolutegravir as maintenance monotherapy for HIV (DOMONO): a phase 2, randomised non-inferiority trial. *Lancet HIV* **4**, e547–e554 (2017).

118. Murphy, R. A., Court, R., Maartens, G. & Sunpath, H. Second-Line Antiretroviral Therapy in Sub-Saharan Africa: It Is Time to Mind the Gaps. *AIDS Res. Hum. Retroviruses* **33**, 1181–1184 (2017).

119. Manasa, J. *et al.* High-Levels of Acquired Drug Resistance in Adult Patients Failing First-Line Antiretroviral Therapy in a Rural HIV Treatment Programme in KwaZulu-

Natal, South Africa. *PLoS One* **8**, (2013).

120. Gupta, R. K. *et al.* HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. *Lancet Infect. Dis.* **18**, 346–355 (2018).

121. Gregson, J. *et al.* Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: A multicentre retrospective cohort study. *Lancet Infect. Dis.* **16**, 565–575 (2016).

122. Dorward, J. *et al.* Dolutegravir for first-line antiretroviral therapy in low-income and middle-income countries: uncertainties and opportunities for implementation and research. *Lancet HIV* **5**, e400–e404 (2018).

123. Barnabas, R. V., Revill, P., Tan, N. & Phillips, A. Cost-effectiveness of routine viral load monitoring in low- and middle-income countries: A systematic review. *J. Int. AIDS Soc.* **20**, 50–61 (2017).

124. Kanters, S. *et al.* Interventions to improve adherence to antiretroviral therapy: a systematic review and network meta-analysis. *Lancet HIV* **4**, e31–e40 (2017).

125. Heestermans, T., Browne, J. L., Aitken, S. C., Vervoort, S. C. & Klipstein-Grobusch, K. Determinants of adherence to antiretroviral therapy among HIV-positive adults in sub-Saharan Africa: A systematic review. *BMJ Glob. Heal.* **1**, 1–13 (2016).

126. Cottrell, M. L., Hadzic, T. & Kashuba, A. D. M. Clinical pharmacokinetic, pharmacodynamic and drug-interaction profile of the integrase inhibitor dolutegravir. *Clin. Pharmacokinet.* **52**, 981–994 (2013).

127. K., B., A., M., T., R., N., F. & Cohn J. Viral load monitoring as a tool to reinforce adherence: A systematic review. *J. Acquir. Immune Defic. Syndr.* **64**, 74–78 (2013).

128. Dorward, J., Drain, P. K. & Garrett, N. Point-of-care viral load testing and differentiated HIV care. *Lancet HIV* **5**, e8–e9 (2018).

Chapter 9 Appendices

Appendix 1 – Ethics letter

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Approval Letter
Progress Report

27/03/2019

Project ID #: 2215

Ethics Reference #: N15/08/071

Title: Tracking the molecular epidemiology and resistance patterns of HIV-1 in South Africa.

Dear Dr. Graeme Jacobs,

Your request for extension/annual renewal of ethics approval dated 05/03/2019 14:28 refers.

The Health Research Ethics Committee reviewed and approved the annual progress report you submitted through an expedited review process.

The approval of this project is extended for a further year.

Approval date: 26 March 2019

Expiry date: 25 March 2020

Kindly be reminded to submit progress reports two (2) months before expiry date.

Where to submit any documentation

Kindly note that the HREC uses an electronic ethics review management system, *Infonetica*, to manage ethics applications and ethics review process. To submit any documentation to HREC, please click on the following link: https://applyethics.sun.ac.za.

Please remember to use your **Project ID** 2215 and Ethics Reference Number N15/08/071 on any documents or correspondence with the HREC concerning your research protocol.

Yours sincerely,

Mrs. Ashleen Fortuin

Health Research Ethics Committee 2 (HREC2)

National Health Research Ethics Council (NHREC) Registration Number:
REC-130408-012 (HREC1)•REC-230208-010 (HREC2)

Federal Wide Assurance Number: 00001372
Office of Human Research Protections (OHRP) Institutional Review Board (IRB) Number:
IRB0005240 (HREC1)•IRB0005239 (HREC2)

The Health Research Ethics Committee (HREC) complies with the SA National Health Act No. 61 of 2003 as it pertains to health research. The HREC abides by the ethical norms and principles for research, established by the World Medical Association (2013). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects; the South African Department of Health (2006).Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa (2nd edition); as well as the Department of Health (2015). Ethics in Health Research: Principles, Processes and Structures (2nd edition).

The Health Research Ethics Committee reviews research involving human subjects conducted or supported by the Department of Health and Human Services, or other federal departments or agencies that apply the Federal Policy for the Protection of Human Subjects to such research (United States Code of Federal Regulations Title 45 Part 46); and/or clinical investigations regulated by the Food and Drug Administration (FDA) of the Department of Health and Human Services.

Appendix 2 – Supplementary table for Chapter 1

Title: Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and availability of integrase inhibitors in Cape Town, South Africa

Short title: **INSTI DRM from South Africa**

Dominik Brado[1#], Adetayo Emmanuel Obasa[2,3#*], George Mondinde Ikomey[4], Ruben Cloete[5], Kamalendra Singh[3,6,7], Susan Engelbrecht[2], Ujjwal Neogi[3], Graeme Brendon Jacobs[2]

[1]Division of Virology, Institute for Virology and Immunobiology, Faculty of Medicine, University of Wuerzburg, 97080, Wuerzburg Germany

[2]Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, 7505, Cape Town, South Africa

[3]Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, University of Stockholm, Sweden

[4]CSCCD, Faculty of Medicain and Biomedical Sciences, University of Yaoundé I, Cameroon

[5]South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape.

[6]Department of Molecular Microbiology and Immunology and [7]Christopher Bond Life Sciences Center, University of Missouri, Columbia MO 65211, USA

[#]Both authors contributed equally to the work

*Corresponding author: Mr. Adetayo Emmanuel Obasa (obasa@sun.ac.za)

Supplementary Table 1: Demographical and clinical data of patients

Isolate	Race / Gender	Orientation	Symptoms	Subtype (IN gene)
TV0116	CF	Hetero	Asymptomatic	C
TV0118	CF	Hetero	Asymptomatic	C
TV0122	CM	MTCT	Asymptomatic	B
TV0124	BF	Hetero	Shingles, Miositis	C
TV0126	BM	Hetero	Asymstomatic	C
TV0127	BM	Hetero	symp to stage II	C
TV0128	BF	Hetero	TB	C
TV0132	BF	Hetero	Weight loss, Oral candidiasis, Pneumonia	C
TV0133	BM	Hetero	TB	C
TV0135	CF	Hetero	Asymstomatic	C
TV0137	BF	Hetero	DVT, Ca CervixIIIB, prev Radiotherapy	C
TV0139	BF	Hetero	Asymstomatic	C
TV0141	BM	Hetero	Asymstomatic	C
TV0142	BF	Hetero	Asymstomatic	
TV0143	BM	Hetero	TB treatment completed	C
TV0144	BF	Hetero	Asymstomatic	C
TV0145	BF	Hetero	STD	C
TV0147	BF	Hetero	Asymstomatic	C
TV0148	BM	Hetero	Current pleural effusion	C
TV0149	CF	MTCT	Pneumonia x2	C
TV0152	BM	Hetero	Asymptomatic	C
TV0153	BM	Hetero	Asymptomatic	B
TV0155	BM	Hetero	Asymptomatic	C
TV0157	BM	Hetero	TB	C
TV0160	CM	Hetero	Asymptomatic	C
TV0161	BF			C
TV0163	CF	Hetero	Sin I with MPV	C
TV0164	CF		Asymptomatic	C
TV0165	CF	Hetero	PTB	C
TV0166	BF			C

Isolate	Race / Gender	Orientation	Symptoms	Subtype (IN gene)
TV0168	CF	Hetero	Candida	C
TV0173	BM	Hetero	Symptomatic of TB	C
TV0175	CF	Hetero	Seborrhea	C
TV0177	BF	Hetero	Anal Warts	C
TV0180	BF	MTCT	Septicemia, PID, Lymphoma,	C
TV0181	BF	Hetero	Asymptomatic	C
TV0182	BF	Hetero	Asymptomatic	C
TV0183	BF	Hetero	Asymptomatic	C
TV0184	BF	Hetero	Asymptomatic	C
TV0185	IF			C
TV0191	BF			C
TV0193	BF			C
TV0198	BF			C
TV0201	BF			C
TV0202	BF			C
TV0345	BF	Hetero	Asymptomatic	C
TV0346	BF	Hetero	Asymptomatic	C
TV0347	CF	Hetero	Asymptomatic	C
TV0348	BM	Hetero	Oral thrush, Prurigo, Weight Loss, Diarrhoea	C
TV0349	BM	Hetero	Asymptomatic	C
TV0350	BM	Hetero	Weight Loss, Oral thrush	C
TV0353	CF	Hetero		C
TV0356	CF	Hetero	Asymptomatic	B
TV0364	BF	Hetero	Focal neurologic deficits	C
TV0365	CM	Hetero	Opthamic Zoster, Candidiasis, Severe Psoriases	C
TV0366	CF	Hetero	Thrombocytopenia	C
TV0367	CF	Hetero	Asymptomatic	C
TV0370	BM	Hetero	Asymptomatic	C
TV0371	BF	Hetero	Asymptomatic	C
TV0372	BF	Hetero	TB	C
TV0373	BM	Hetero	Prurigo, Weight loss, Lymphadenopathy	C

Isolate	Race / Gender	Orientation	Symptoms	Subtype (IN gene)
TV0375	BF	Hetero	TB<1year, Genital Warts, Candidiasis	C
TV0376	BF	Hetero	Asymptomatic	C
TV0377	BM	Hetero	Asymptomatic	C
TV0388	CF			C
TV0398	BM	Hetero		C
TV0404	CF	Hetero	Asymptomatic	B
TV0405	BF	Hetero	Asymptomatic	C
TV0406	BF	Hetero		C
TV0407	BF	Hetero	Prurigo, Lymphadenopathy	C
TV0412	BM	Hetero	Chronic staph. Aur. skin sepsis	A
TV0413	CF	Hetero	Asymptomatic	C
TV0417	CF	Hetero		C
TV0418	BF	Hetero		C
TV0420	CF			B
TV0424	BM	MTCT	PTB	C
TV0425	BF	Hetero	Asymptomatic	C
TV0426	BF	Hetero	Prurigo	C
TV0427	BF	Hetero	Asymptomatic	C
TV0431	WM	Hetero	Asymptomatic	B
TV0432	BF		Asymptomatic	C
TV0433	CF	Hetero	Asymptomatic	C
TV0434	BF	Hetero	TB	C
TV0435	BF	Hetero	Asymptomatic	C
TV0436	BM	Hetero	Asymptomatic	C
TV0437	BM	Hetero	Asymptomatic	C
TV0438	BM	Hetero	TB	C
TV0442	WF	Hetero		C
TV0445	Mix	Hetero	Asymptomatic	C
TV0446	CF	Hetero	Asymptomatic	C
TV0456	CF	Hetero	Asymptomatic	C

BF: Black Female, **BM**: Black Male, **CF**: Coloured Female, **CM**: Coloured Male, **WF**: White Female, **WM**: White Male, **PTB**: Pulmonary Tuberculosis, **DVT**: Deep Vein Thrombosis: **STD**: Sexually Transmitted Diseases

Appendix 3 – Supplementary table Chapter 4

PID	PI_Major	NRTI	NNRTI	INI
ZA22	V82A (1.13)	M184V (2.60)	K103N (1.63), P225H (1,66)	
ZA174		D67N (75.75),	L100I (1,025)	
ZA212	V82A (2.17)	M184V (99,88)	K103N (2.79),	
ZA28	V82A (1.15)	M184V (84.04), T215I (13.12)	K103N (2.53), K103S (1.27), V106M (1.08), Y188L (95.77), P225H (15.05)	
ZA87			K103N (1.31)	
ZA 288		D67N (8.98),	Y188L (99.75)	
ZA 275			K103N (99.38)	
ZA 166	M46I (93.84), V82A (99.18), I54V (99.6%)	L74 (99.28), Y115 (99.12), M184V (99.46)	Y181C (99.41)	
ZA 180		M184V (99.64),	K103N (99.52), P225H (99,51)	
ZA 19	V82A (17.68)	M184V (99.85),	K103N (98.09), P225 (99,34)	E138K (2,59)
ZA 201	I50L (99.41), I54V (99.3) V82A (99.10)	M184V (99,67)	K103N (99,45)	
ZA105	N88D (1.38)	M184V (99.69)	P225 (99,07)	
ZA11	M46I (90.44), I54V (99.3), V82A (99.11)	M41L (81.34), D67G (81.14), K70R (81.87), L74I (81.40), T215 (99.51), K219E (99.59)	K103N (38.46))	
ZA118	M46I (95.23), L76V (97.59), I84V (97.48)	D67N (99.60), K70R (99.41), M184V (99.77), K219Q (99.29)	Y188L (99.54)	
ZA127	N88D (2.77)	M184V (98.87)	K103N (98.58)	S147G (29.14)
ZA156	V82L (96.78)	M184V (97,48)	K103N (97.31)	
ZA202	M46I (97.52), L76V (91.28), V82A (99.66)	K65R (99.51), M184V (99,89)	P225H (99.41),	
ZA217		M184V (99.77)	K103N (99.63)	
ZA221	L76V (34.33), V82A (12.27), N88D (33.25)	M184V (99.84)	K103N (99.54), P225H (99.63)	
ZA61		M184V (99.70)	K103N (99.48)	
ZA91		M184V (99.71)	G190A (99.25), V106M (41.11)	
ZA93		M184V (99.60)	V106M (41.11), G190A (99.57)	
ZA94	V32I (98.25), M46I (92.29), I47V (2.92), L76V (96.76), V82A (2.12), I84V (96.50)	M184V (99.65), T215F (99.10),	G190A (98.96), K101E (98.50)	Y143R (99.07)
ZA25		M184V(99.63), K65R (98.78), Y115F (98.76),	K103N (99.47), V106M (98.99), Y181C (99.02)	
ZA90		M41L (98.60), T69D (98.39), T215Y (98.41), K219R (98.42),	G190A (99.25), V106M (99.36), Y188L (98.71)	
ZA110	V82A (10.34)	F77L (2.26), K219Q (37.05),	P225H (36.77)	
ZA113	V82A (1.82)		K103N (21.51)	
ZA120		K65R (9.26), L74V (61.11), Y115F (98.68)	K103N (98.58), G190A (99.42)	
ZA182			K103N (99.38), G190A (99,62)	
ZA203	M46I (98.86), I47V (93.55), L76V (97,69) I84V (97.20)	M184V (99.83), D67N (40.81), K70N (99.40), K219 (7.52)	Y188L (99.31),	
ZA273			K101E(99.73), K103N (99.69),	
ZA49	V82A(6.92)		K101E (76.26), K103N (22.46), Y181C (73.65), G190S (73.34),	
ZA74	M46I (88.2)		K103N (95.04)	
ZA101		K65R (3.81), D67N (8.27%), K70R (2.41%), M184V (5.69), T215I (97.68), K219E (2.63)	V106M (99.56), Y188L (1.61)	
ZA222	L76V (1.25), V82A (1.15)	M184V (1.12), K65R (1.26)	G190A (98.62), P225H (1,24)	
ZA27		M184V (2.11), K65R (1.09)	K103N (1.87), K103S (95.78), V106M (1,24), G190A (97.46)	
ZA37		M184V (1.10)		
ZA 199	M46I (96.51), M46L (1.16), L76V (58.73), I84V (62.88)	M184V (99.96), L74V (99.83), Y115F (99.79),		
ZA06	M46I (95.36), I54V(99.4), V82A (99.48)	D67N (99.38), K70R (99.51), M184V (98.60), K219Q (99.34)		
ZA15	I50L (99.82)	M41L (81.34), D67G (81.14), K70R (81.87), L74I (81.40), T215Y (99.51), K219E (99.59)		Y143R (99.47)
ZA178	M46I (92.37), G48V (1,19), V82A (99.12)	D67N (99.45), K70R (99,36), M184V (99,06),		
ZA206	V82A (99.01)	M184V (99.43)		
ZA210		M184V (76.70)		
ZA283	V82L (2.16)	D67N (99.05), K70E (99.38), M184V (99.70)	Y188H (4.95), V106M (99.28)	
ZA289		M184V (99.70)		
ZA294	M46I (90.51), V82A (99.20), I84V (61.54), N88S (37.85)	M184V (99.80), L74V (99.64)		
ZA31	V82A (1.10)	M184V (99.61)	K101P (97.05), K103N (1.70), K103S (97.85), P225H (1.49)	
ZA42	V32L (42.74), L76V (93.86), V82A (99.52), I884V (99.52)	M184V (99.78)		Y143R (99.07)
ZA48		K65R (99.14), M184V (99.70)		
ZA 286	V82A (31.26)		K103N (29.17), Y181C (10.43)	
ZA100		K65R (1.85), D67N (97.8%), K70R (97.92%), M184V (99.73), T215I (97.68), K219E (97.81)	V106M (99.56), Y188L (1.61),	
ZA121	V82A (5.32)			
ZA183		M184V (99.52), K219Q (2,04)	K103N (15,39)	
ZA215			K103N (30.88), V106M (8.39), G190A (11,69)	
ZA245	N88D (39.19)			
ZA97	V32I (1.57), L76V (1.39), I84V (1.47)	K65R (98.29), M184V (99.84), T215F (1.33),	K101E(1.12), K103N (98.19), G190A (1.34), P225H (97.88),	Y143R (1.23)