INVESTIGATING THE STRUCTURAL IMPACT OF HIV-1 INTEGRASE NATURAL OCCURRING POLYMORPHISMS AND NOVEL MUTATIONS IDENTIFIED AMONG GROUP M SUBTYPES CIRCULATING IN SUB-SAHARAN AFRICA

by

Sello Given Mikasi

Thesis presented in fulfilment of the requirements for the degree Doctor

Of Science in Medical Virology at the Stellenbosch University

Supervisor: Dr. Graeme Brendon Jacobs (Stellenbosch University, South Africa)

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

Co-supervisor: Prof. Gert Van Zyl (Stellenbosch University, South Africa)

Co-Supervisor: Prof. Susan Engelbrecht (Stellenbosch University, South Africa)

Co-Supervisor: Dr. George Mondinde Ikomey (University of Yaoundé I, Cameroon)

Co-Supervisor: Dr. Christa Kasang (German Leprosy and Tuberculosis Relief Association, Germany)

Division of Medical Virology

Department of Pathology

Faculty of Medicine and Health Sciences

December 2020

DECLARATION

By submitting this thesis, I declare that the work contained herein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it at any University for obtaining qualifications. I am the sole author of the abstract, introduction and summative comment sections and first author in five of the seven articles included in the thesis.

Sello Given Mikasi

01 August 2020

ACKNOWLEDGEMENTS

My Ph.D. project has been a long journey of leaning. I have gained valuable experience of teamwork, writing articles and also to be able to answer critical scientific questions. I will forever be indebted to many people, who made this piece of crafted book of knowledge to see the light. Only words will not reveal my inner most sincere gratitude to my one and only undisputed Ph.D. supervisor, Dr. Graeme Brendon Jacobs for taking me under his wing in his research group unapologetic. He guided me through the journey of my study up until I saw the light at the end of the tunnel. His tenacity, resilience, belief, caring and scientific knowledge, coupled with his financial funding support made my life easy throughout my study.

I would also like to extend my hands of appreciation to my co-supervisor: Dr. Ruben Cloete, Prof. Gert Van Zyl, Dr. George Mondinde Ikomey, Dr. Christa Kasang and Prof. Susan Engelbrecht for their excellence expertise and contribution towards my Ph.D. work.

I would like to acknowledge my research group, for making it a memorable time. Especially – Duncan Njenda, Emmanuel Obasa and Olivette Varathan, – who contributed in one way or the other to my research.

I am much obliged to my collaborators, Dr. Rubben Cloete, Rumbizai Chitongo, Darren Isaacs, Henerico Shimba, Dr. Christa Kasang and Dr. George Ikomey, who made this art of work possible.

Many thanks goes to my wife (Joy Mavhetha), son (Atendaho Jaden Mikasi), Mother (Sarah Ratapala) and family for their support and courageous words. Finally yet importantly, I would like to send my revolutionary salutations to the Economic Freedom Fighter's (EFF), particularly fighters from branch 107 (Parklands) and 113 (Table view) for their revolutionary support.

Many thanks goes to the staff and students in the Division of Medical Virology, Stellenbosch University and South African National Bioinformatics Institute (SANBI), University of the Western Cape. Special thanks goes to my humble fighters in academic; Duncan Njenda, Emmanuel Obasa, Darren Isaacs and Rumbidzai Chitongo, for their effort and positive contribution towards my Ph.D. study.

My appreciation goes to my funding's 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, the project would just be a dream.

Psalm 28:7:

"The Lord is my strength and my shield; in him my heart trusts, and I am helped; my heart exults, and with my song I give thanks to him."

DEDICATION

I dedicate my thesis to my wife *Aluwani Joy Mavhetha*, my Son *Atendaho Jaden Mikasi*, my mother *Sarah Ratapala*, siblings, friends, family and all people living with HIV/AIDS.

Che Guevara

"The first duty of a revolutionary is to be educated".

ABBREVIATIONS

3D	Three-dimensional
3TC	Lamivudine
AA	Amino acid
ABC	Abacavir
AIDS	Acquired Immuno Deficiency Syndrome
APOBEC3G	Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G
ART	Anti-retroviral therapy
ATV	Atazanavir
AZT	Zidovudine
bPI	boosted protease inhibitors
BIC	Bictegravir
cART	combination Antiretroviral therapy
CCR5	Chemokine core-receptors 5
cDNA	Complementary deoxynucleic acid
COMET	Context-based Modeling for Expeditious Typing
CPR	Calibrated Population Resistance
CryoEM	Cryo-electron microscopy
CTLs	Cytotoxic T-lymphocytes
CXCR4	Chemokine X core-receptors
d4T	Stavudine
ddNTPs	Dideoxyribo-nucleoside triphospahte
DRMs	Drug resistance mutations
DRC	Democratic Republic of Congo
DOR	Doravirine
DRV	Darunavir
DTG	Dolutegravir
EFV	Efavirenz
Env	Envelope
ESCRT	Endosomal sorting complexes required for transport machinery
ETV	Etravirine
EVG	Elvitegravir
FDA	Food and Drug Administration
FIs	Fusion Inhibitors
FTC	Emtricitabine
Gag	Group-specific antigen
gp	Glyco protein
HM	Homology modelling
HIV	Human Immunodeficiency virus
HREC	Human Research Ethics Committee
IN	Integrase
InSTIs	Integrase strand transfer inhibitors
LTR	Long terminal repeats
MA	Matrix
MD	Molecular modelling

MRC	Medical Research Council
mRNA	messenger RNA
NC	Nucleocapsid
Nef	Negative factor
NOPs	Natural occurring polymorphisms
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTIs	Nucleoside reverse transcriptase inhibitors
NTD	N-terminal domain
NVP	Nevarapine
PB	Phosphate buffer
PBMCs	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PFVs	prototype foamy viruses
PMCT	Prevent mother to child trans-mission
pol	Polymerase gene
PPT	3′ polypurine tract
PR	Protease
PrEP	Pre-exposure prophylaxis
RAL	Raltegravir
RAMs	Resistance-associated mutations
Rev	Regulator of expression of virion proteins
RMSD	Root-mean square deviation
RNA	Ribonucleic acid
RPV	Rilpivirine
RT	Reverse transcriptase
RTV	Ritonavir
SBVS	structure-based virtual screening
STC	strand transfer complex
SU	Surface
Tat	Transcriptional transactivator protein
tat	Transcriptional transactivator gene
TAF	Tenofovir alafenamide
TDF	Tenofovir
TM	Transmembrane
TRIM5α	Tripartite motif-containing 5α
UNAIDS	United nation AIDS
URFs	Unique recombinant forms
Vif	Virion infectivity factor
Vpr	Viral protein R
Vpu	Viral protein U gene
WT	Wild-Type

LIST OF FIGURES

Figure 1: The HIV-1 genome structure.. 23

Figure 2: This infographic illustrates the HIV replication cycle, which begins when HIV fuses with the surface of the host cell.. 25

Figure 3 : Global distribution of major HIV subtypes.. 28

Figure 4: DNA cutting and joining steps of retroviral DNA integration.. 31

Figure 5: FDA approved antiretroviral drugs for HIV treatment shown to act on different stages of the HIV-1 replication cycle 32

Figure 6: Outline of the homology modeling process and its applications in drug discovery.. 41

Figure 7: Schematic representation of a molecular dynamics cycle. 43

Figure 8. Molecular docking flow chart. 44

LIST OF TABLES

Table 1 FDA approved Fixed-dose antiretroviral drugs for HIV currently marketed 33

SCIENTIFIC CONTRIBUTIONS

All published papers were reproduced with permission from the publisher, according to publisher's copyright and "Open Access" conditions.

LIST OF SCIENTIFIC PAPERS INCLUDED IN THIS THESIS – PUBLISHED

I. ***Mikasi SG***, Gichana JO, Van der Walt C, Brado D, Obasa AE, Njenda D, Messembe M, Lyonga E , Okomo O, Cloete R, Ikomey GM, Jacobs GB. *HIV-1 Integrase Diversity and Resistance-Associated Mutations and Polymorphisms among Integrase Strand Transfer Inhibitor-Naive HIV-1 Patients from Cameroon*. AIDS Res Hum Retroviruses. 2020 May;36 (5):450-455. doi: 10.1089/AID.2019.0264.

II. ***Mikasi SG***, Isaacs D, Ikomey GM, Shimba H, Cloete R, Jacobs GB. *HIV-1 drug resistance analyses of Cameroon derived Integrase sequences.* AIDS Res Hum Retroviruses. July2020. https://doi.org/10.1089/AID.2020.0022.

III. Obasa AE, ***Mikasi SG***, Brado D, Cloete R, Singh K, Neogi U and Jacobs GB (2020) *Drug Resistance Mutations Against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living With HIV-1 Receiving Boosted Protease Inhibitors in South Africa.* Front. Microbiol. 11:438. doi: 10.3389/fmicb.2020.00438.

IV. Chitongo R, Obasa A.E, ***Mikasi S.G***, Jacobs G.B, Cloete R. *Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding*. PLoS ONE 15(5): e0223464. https://doi.org/10.1371/journal. Pone.0223464.

V. ***Mikasi SG***, Ikomey GM, Obasa AE, Cloete R, Jacobs GB. HIV-1 diversity and the implementation of Integrase strand-transfer inhibitors as part of combination antiretroviral therapy. S.Afr.Med.J. 2020;110(9):827. https://doi.org/10.7196/SAMJ.2020.v110i9.14848.

VI. Isaacs, D.; ***Mikasi, S.G***.; Obasa, A.E.; Ikomey, G.M.; Shityakov, S.; Cloete, R.; Jacobs, G.B. Structural Comparison of Diverse HIV-1 Subtypes using Molecular Modelling

and Docking Analyses of Integrase Inhibitors. Viruses 2020, 12, 936; doi:10.3390/v12090936

UNDER REVIEW

I. ***Mikasi SG***, Isaacs D, Chitongo R, Ikomey GM, Cloete R, Jacobs GM. *Investigating the structural effects of statistically enriched mutations identified in Cameroon recombinant subtype CRF02_AG that might lead to Dolutegravir drug resistance*. **Manuscript ID: INFD-D-20-02350**

LIST OF RELATED SCIENTIFIC PAPERS NOT INCLUDED IN THESIS

I. Ikomey GM, Assoumou O, Gichana JO, Njenda D, ***Mikasi SG***, Mesembe M, Lyonga E, Jacobs GB *Observed HIV drug resistanceassociated mutations amongst naïve immunocompetent children in Yaoundé, Cameroon*. GERMS 2017;7(4):178-185. doi: 10.18683/germs.2017.1124.

II. Ikomey GM, Chendi BH, van der Walt C, Obasa AE, ***Mikasi SG***, Mkong E, Mesembe M, Chegou NN, Sanderson M, Doh G, Estella T, Okomou M-C, Fokunang C, Jacobs GB. *Capacity strengthening of biomedical research within the framework of South-South collaboration between bilateral University partnerships of Cameroon and South Africa*. Frontiers: public Health .Manuscript ID: **516754**. Under review.

III. Benjamin L, Louw J, ***Mikasi SG***, Obasa AE, Jacobs GB. HIV/AIDS, education and the South African school curriculum (**Manuscript draft**).

LIST OF CONFERENCE PRESENTATIONS

International or national

I. ***Mikasi SG***, Jacobs GB, Engelbrecht S. *Analysis of the HIV-1 diversity in the remote areas of the Cape Winelands, Overberg and West Coast districts of the Western Cape Province of South Africa*. 22nd International AIDS Conference (AIDS, 2018), Amsterdam. Presented (23-27 July 2018).

II. ***Mikasi S.G***, Gichana JO, Van der Walt C, Brado D, Obasa AE, Njenda D, Messembe M, Lyonga E, Assoumou O, Cloete R, Ikomey GM and Jacobs GB. *Resistance-associated mutations and polymorphisms among integrase inhibitor-naïve HIV-1 patients in Cameroon* (Virology Africa 2020), Cape Town, South Africa. Presented (10-14 February 2020).

Division of Medical Virology presentations

I. ***Mikasi SG*** , Van Zyl G ,Engelbrecht S, Ikomey GM, Kasang C, Cloete R and Jacobs GB. *Investigating the structural impact of HIV-1 Integrase natural occurring polymorphisms and novel mutations identified among group M subtypes circulating in Sub-Saharan Africa*. **PhD project.**

- Dates of presentations
  - ✓ 22 July 2017
  - ✓ 11 November 2018
  - ✓ 25 February 2019
  - ✓ 19 February 2020
  - ✓ 15 April 2020

International research visit

I. **Catholic University in Mwanza (Tanzania).** The aim of the visit was to strengthen and support a long-lasting and close collaboration triangular partnership to fight HIV/AIDS in Sub-Saharan Africa between the University of Würzburg (Germany), the University of Stellenbosch (South Africa) and the Catholic University in Mwanza (Tanzania): 01 -15 July 2017 and 10-20 December 2017.

ABSTRACT

Introduction

HIV/AIDS remains a major health concern worldwide, with sub-Saharan Africa (SSA) carrying the largest burden. HIV is characterised by extremely high genetic diversity, with all the major groups and subtypes circulating in SSA. Combination antiretroviral therapy (cART) have substantially reduced HIV related deaths, but this is counteracted by the development of HIV drug resistance, caused by certain drug resistance-associated mutations (RAMS). Integrase (IN) strand transferase inhibitors (INSTIs), the newest class of antiretroviral drugs, has a high genetic barrier and can be used in individuals that previously exhibited resistance to other classes of drugs. The World Health Organisation (WHO) approved the use of Dolutegravir (DTG) as part of first-line cART.

Methods

This is a descriptive experimental design study, which aimed to identify IN natural occurring polymorphisms (NOP) among different HIV-1 group M subtypes and Drug resistance mutations within the HIV-1 pol gene fragment of INSTI naïve patients from South Africa (SA) and Cameroon (CR), using the Stanford University genotypic resistance interpretation algorithm. Structural computational methods that included; homology modelling, molecular docking, molecular dynamics simulations and interaction analysis was performed to understand the structural impact of mutations from diverse HIV-1 subtypes on DTG drug binding.

Results

We observed low-level RAMs against INSTIs in SA (2.2%) and CR sequences (5.4%). Through Fisher's exact test we noted that the two NOPs occurred: VI72I and R269K, with p-values ≤0. 05, were statistically enriched. The impact of having these mutations are yet to be fully understood. Through molecular modelling and stability predictions, we observed a destabilizing effect of the known G140S mutant on the HIV-1C IN protein structure and simulation analysis showed that it affected structural stability and flexibility of the protein structure. Interactions analysis of different drug binding conformations to different HIV-1 IN subtypes reported differences in the number of binding interactions to different HIV-1 IN subtypes, but we did not observe any significant differences in binding affinity for each INSTIs.

This implies no significant alteration to the binding site in the wild type IN, which may not prevent INSTIs drug binding. In addition, all accessory mutations that resulted in a change in the number of interactions encompassing residues were found within the stable alpha-helix secondary structure element and not in close proximity to the drug active site.

Conclusion

The study data indicate that RAMS against INSTIs remain low both in SA and in CR. Subtype C in SA and CRF02_AG in CR continues to be the driving force of the epidemic. We further reported on the impact of various NOPs on drug susceptibility. The analyses suggested that NOPs does not have an impact on IN protein structure and stability, and does not affect drug binding in the WT IN, but the known mutation G140S affect DTG binding. The study support recommendations made by the WHO to use DTG as part of salvage therapy in patients with RAM's and accessory mutations. Data obtained from this study can help to tailor effective treatment strategies in the African population, where diverse HIV subtypes circulate.

OPSOMMING

Inleiding

MIV/vigs bly wêreldwyd ’n ernstige gesondheidskwessie, en Afrika suid van die Sahara dra die swaarste las. MIV word deur uiters hoë genetiese diversiteit gekenmerk, waarvan al die vernaamste groepe en subtipes in Afrika suid van die Sahara in omloop is. Kombinasie-antiretrovirale terapie (kART) het ’n aansienlike vermindering in MIV-verwante sterftes tot gevolg, hoewel dít teengewerk word deur die ontwikkeling van MIV-middelweerstandigheid vanweë sekere middelweerstandigheidsverwante mutasies (oftewel RAM’s). Integrasestringtransferase-inhibitors (INSTI’s), die jongste klas antiretrovirale middels, het ’n hoë genetiese skans en kan gebruik word by individue wat voorheen weerstandigheid teen ander klasse middels getoon het. Die Wêreldgesondheidsorganisasie (WGO) het die gebruik van dolutegravir (DTG) as deel van eerstelinie-kART goedgekeur.

Metodes

Hierdie studie gebruik ’n beskrywende proefondervindelike ontwerp om natuurlike integrase-(IN-) polimorfismes (NOP’s) in verskillende MIV-1-groep-M-subtipes en middelweerstandige mutasies in die MIV-1-*pol*-geenfragment van INSTI-naïewe pasiënte van Suid-Afrika (SA) en Kameroen (KR) te identifiseer. Dit word met behulp van die Universiteit van Stanford se algoritme vir genotipiese weerstandigheidsvertolking gedoen. Strukturele berekeningsmetodes soos homologiemodellering, molekulêre koppeling, molekulêre dinamikasimulasies en interaksieontleding is uitgevoer om die strukturele impak van mutasies uit diverse MIV-1-subtipes op DTG-middelbinding te verstaan.

Resultate

Laevlak-RAM’s teen INSTI’s is in reekse van SA (2,2%) én KR (5,4%) opgemerk. Fisher se eksakte toets het twee NOP’s opgespoor – VI72I en R269K – met p-waardes van ≤0,05, wat statisties verryk was. Die impak van hierdie mutasies is nog nie ten volle duidelik nie. Deur molekulêre modellering en stabiliteitsvoorspellings het ons bepaal dat die bekende G140S-mutant ’n destabiliseringsuitwerking het op die MIV-1C-IN-proteïenstruktuur. Simulasieontleding het getoon dat dít die strukturele stabiliteit en buigbaarheid van die proteïenstruktuur beïnvloed. Interaksieontleding van middelbindingskonformasies met MIV-1-IN-subtipes het verskille in die getal bindingsinteraksies met verskillende subtipes opgelewer, maar geen beduidende verskille in bindingsaffiniteit is vir enige van die INSTI’s

opgemerk nie. Dít impliseer dat daar geen beduidende aanpassing is in die bindingsetel by die wilde-tipe IN wat INSTI-middelbinding kan verhoed nie. Daarbenewens is alle bykomstige mutasies wat ’n verandering in die getal interaksies in residu’s veroorsaak het in die stabiele alfaheliks- sekondêre struktuurelement aangetref, en nie naby die aktiewe setel van die middel nie.

Gevolgtrekking

Die studiedata toon dat RAM’s teen INSTI’s steeds laag is in sowel SA as KR. Subtipe C in SA en CRF02_AG in KR bly die dryfkrag agter die epidemie. Daarbenewens is daar oor die impak van verskillende NOP’s op middelvatbaarheid verslag gedoen. Die ontledings toon dat NOP’s nie ’n impak op IN-proteïenstruktuur en -stabiliteit het nie, en ook nie middelbinding in die WT-IN beïnvloed nie. Nogtans het die bekende mutasie G140S wél ’n invloed op DTG-binding. Die studie ondersteun die WGO se aanbeveling dat DTG as deel van reddingsbehandeling by pasiënte met RAM’s en bykomstige mutasies gebruik word. Die data uit hierdie studie kan doeltreffende behandelingstrategieë help ontwikkel vir die bevolking van Afrika, waar diverse MIV-suptipes in omloop is.

TABLE OF CONTENTS

DECLARATION i

ACKNOWLEDGEMENTS ii

DEDICATION iv

ABBREVIATIONS v

LIST OF FIGURES vii

LIST OF TABLES viii

SCIENTIFIC CONTRIBUTIONS ix

ABSTRACT xii

OPSOMMING xiv

TABLE OF CONTENTS xvi

CHAPTER 1: Introduction and literature review 21

1.1. INTRODUCTION 21

1.2. HIV-1 GENOME STRACTURE 22

1.3. HIV-1 LIFE CYCLE 23

1.4. HIV-1 IN South Africa and Cameroon 26

1.4.1. South Africa 26

1.4.2. Cameroon 26

1.5. HIV-1 DIVERSITY IN SUB-SAHARAN AFRICA 27

1.6. NATURALY OCCURING POLYMORPHISIMS (NOPs). 28

1.7. HIV-1 INTEGRASE (IN) 29

1.7.1. STRUCTURE 29

1.7.2. ACTIVITY 30

1.8. HIV DRUG TARGETS AND ARV CLASSES 31

1.9. MECHANISMs OF ACTION OF DIFFERENT cART drugs Error! Bookmark not defined.

1.9.1. *Reverse Transcriptase* 33

1.9.2. Integrase strand-Transfer Inhibitors (INSTIs) 34

1.9.3. *Protease Inhibitors (PIs)* 34

1.9.4. *Inhibitors of co-receptor usage (CCR5 antagonists)* 35

1.9.5. *Fusion Inhibitors (FIs)* 35

1.10. HIV drug resistance mechanisms, viral fitness and role of secondary mutations in virus evolution 35

1.10.1. *NRTIs resistance mechanisms* 36

1.10.2. *NNRTIs resistance mechanisms* 37

1.10.3. *INSTIs resistance mechanisms* 38

1.10.4. *PI resistance mechanisms* 38

1.11. MOLECULAR MODELLING 39

1.11.1. HOMOLOGY MODELLING 40

1.11.2. MOLECULAR DYNAMIC SIMULATION 41

1.11.3. MOLECULAR DOCKING 43

2. CHAPTER TWO: STUDY FINDINGS 45

2.1. AIMS AND OBJECTIVES 45

Chapter 3: HIV-1 Integrase Diversity and Resistance-Associated Mutations and Polymorphisms among Integrase Strand Transfer Inhibitor-Naive HIV-1 Patients from Cameroon. 47

3.1. **Journal article** 47

3.2. **Author's list** 47

3.3. **Author's contribution** 47

3.4. **Background** 47

3.5. **Main findings** 47

3.6. **Study significance** 48

3.7. **Conclusion** 48

3.8. **Published article** 49

Chapter 4: HIV-1 drug resistance analyses of Cameroon derived Integrase sequences 55

4.1. Journal article 55

4.3. Author's contribution 55

4.4. Background 55

4.5. Main findings 55

4.6. Study significance 55

4.7. Conclusion. 56

4.8. Published article 56

Chapter 5: Drug resistance mutations against protease, reverse transcriptase and integrase inhibitors in people living with HIV-1 receiving boosted protease inhibitors (bPIs) in South Africa. 60

5.1. **Journal article** 60

5.2. **Author's list** 60

5.3. **Authors contribution** 60

5.4. **Background** 60

5.5. **Main findings** 60

5.6. **Study significance** 61

5.7. **Conclusion** 61

5.8. Published article 61

Chapter 6: Investigating the structural effects of statistically enriched mutations identified in Cameroon recombinant subtype CRF02_AG that might lead to Dolutegravir drug resistance. 71

6.1. Journal article 71

6.2. Author's list 71

6.3. Authors contribution 71

6.4. Background 71

6.5. **Main findings** 71

6.6. **Study significance** 72

6.7. **Conclusion** 72

6.8. Preprint manuscript 72

Chapter 7: Structural comparison of diverse HIV-1 subtypes using Molecular modelling and Molecular Docking of integrase inhibitor. 89

7.1. **Journal article** 89

7.2. **Author's list** 89

7.3. **Authors contribution** 89

7.4. **Background** 89

7.5. **Main findings**.....90

7.6. **Study significance**.....90

7.7. **Conclusion** 90

7.8. **Published article**.....91

Chapter 8: Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding. 103

8.1. **Journal article** 103

8.2. **Author's list** 103

8.3. **Authors contribution** 103

8.4. **Background** 103

8.5. **Main findings**.....104

8.6. **Study significance**.....104

8.7. **Conclusion** 104

8.8. **Published article**.....104

Chapter 9: Summary of the study: South African Medical Journal (SAMJ). 120

Chapter 10: Overall discussion and future remarks 121

Chapter 11: References 127

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1. INTRODUCTION

There has been many challenges in the search for a cure for HIV/AIDS. The improvements of care through standard combination antiretroviral therapy (cART) regimens has reduced the severity of HIV/AIDS to a more manageable, chronic disease. cART includes the use of one non-nucleoside reverse transcriptase inhibitors (NNRTIs) and two nucleoside reverse transcriptase inhibitors (NRTIs) as part of the first-line regimen and two NRTIs, ritonavir (RTV)-boosted PI (bPI), as part of the second-line regimen. Due to the emergence of HIV-1 drug resistance, Integrase (IN) strand-transfer inhibitors (INSTIs) has now become a viable option to include in standardized cART. Available INSTIs include; Raltegravir (RAL) and Elvitegravir (EVG) as first-generation INSTIs and Dolutegravir (DTG) and Bictegravir (BIC) as second-generation inhibitors. First-generation INSTIs have been reported to have a relatively low genetic barrier to resistance, while second-generation INSTIs, including DTG, exhibit longer dissociation half-life in biochemical analyses of wild-type (WT) Integrase/DNA complexes, resulting in a high genetic barrier to resistance and able to achieve complete viral suppression that will likely reduce the rate of viral rebound. Furthermore high prevalence of HIV diversity worldwide, particularly in sub-Saharan Africa, poses a major challenge on a wide spectrum of fields, such as vaccine development, diagnostics and cART outcomes. It is hypothesized that diverse HIV-1 subtypes from sub-Saharan Africa have specific naturally occurring polymorphisms (NOPs) that might reduce the efficacy and binding strength of second-generation INSTIs, like DTG, since current available cART drugs are designed in relation to HIV1-B, predominate in the Western Countries, with less research done in Africa focusing on non-B subtypes prevalent in Africa. Our work uses a genotypic and structural modelling approach to help us try to understand the effect of NOPs from diverse subtypes on drug susceptibility and binding affinity to DTG. Molecular modelling provides an approach that can be applied to prioritize the effect of mutations/variants on IN drug binding before expensive experimental assays are performed. In this chapter, we will highlight the origin and history of HIV, including the impact it had on people's lives and how the world responded to the AIDS epidemic. I will also provide an overview of the steps that cART target to inhibit HIV-1 to replicate further. The study used cohort samples from Cameroon and South Africa. Central and West Africa, including Cameroon, which is seen, as the birthplace of HIV, whereas South Africa is a country that is heavily affected by the HIV-1 pandemic.

1.2. HIV-1 GENOME STRACTURE

A structure of the HIV-1 genome is presented in Figure 1. HIV is a retrovirus, composed of double-stranded genomic RNA that is approximately 9kb in size (1). The virus encodes nine open reading frames (ORFs) of which three of these (*gag*, *pol* and *env*) are found in all retrovirus and provide the instructions to make proteins that will form new virus particles (2). For example, env provides the code to make the proteins that form the envelope, or shell, of HIV. gag makes the structural proteins such as the matrix and the capsid, and pol makes the enzymes that are essential for making new viruses. The structural components of the virion compose of six proteins that form the building blocks of the virus. These are the four Gag proteins Capsid (CA), Matrix (MA), Nucleocapsid (NC) and p6, and the two Env proteins, Surface (SU or gp120) and Transmembrane (TM or gp41). The polymerase gene (*pol)* encodes three of the major enzymatic components, Protease (PR), Reverse Transcriptase (RT) and Integrase (IN) that plays unique roles in other retroviruses.

HIV encodes at least six additional proteins that play a role in the viral replication cycle. These proteins are called accessory and regulatory proteins. Three of the accessory proteins, [Virion infectivity factor (Vif), Viral protein R (Vpr), and Negative factor (Nef)] are packed in the viral particle core and they play a role in increasing production of the HIV proteins. The *vif* gene increases the production of the HIV particle in the peripheral blood lymphocytes (3). Vpr facilitates the infection of non-dividing cells by HIV , while Nef plays a role in down modulation of CD4 and MHC class I (4). Two other regulatory proteins, Transcriptional transactivator protein (Tat) and Regulator of expression of virion proteins (Rev) are essential for regulating the production of HIV *in vitro* (5) and the last Viral protein U gene (Vpu), helps in the assembly of the virion indirectly (6).

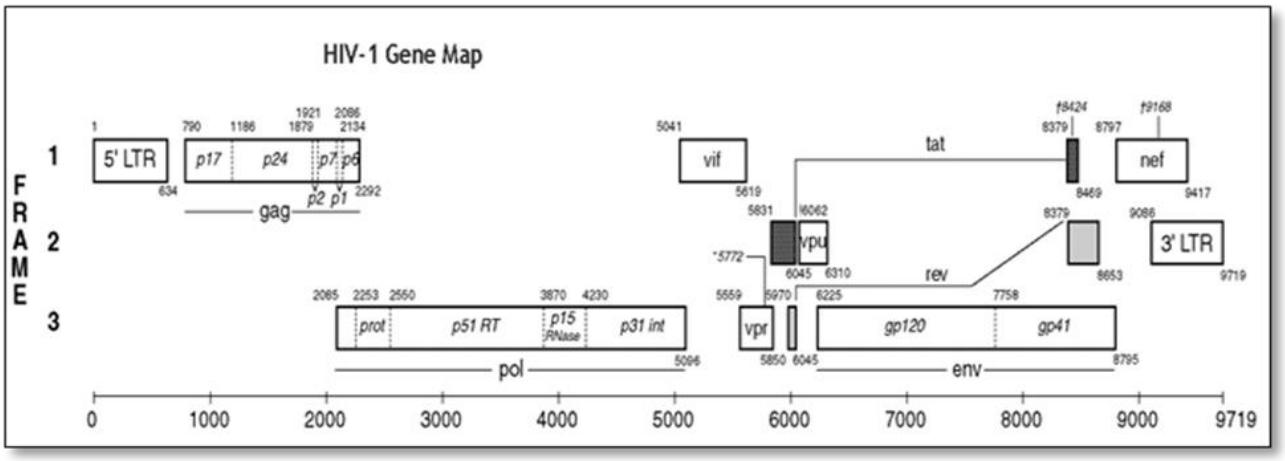

Figure 1: The HIV-1 genome structure.The Structural genes (*gag, pol* and *env*) regulatory genes (*tat* and *rev*) and accessory genes (*nef, vif, vpr* and *vpu*) are indicated in Figure 1, as well as the proteins encoded by each genomic region. Image reproduced with permission and adapted from (Source:www.hiv.lanl.gov).

1.3. HIV-1 LIFE CYCLE

Retroviruses uses cells in the host body to replicate (7) Figure 2. HIV infection starts when the virus attaches its own glycoprotein (gp120/gp41) to the host CD4+ T-cell receptor and the chemokine core-receptors , either Chemokine core-receptors 5 (CCR5) and / or Chemokine X core-receptors 4 (CXCR4), and then penetrate the human host cell, usually white blood cells (WBC) (8). The CCR5 and CXCR4 co-receptors are the main chemokine receptors that are used by HIV for entry *in vivo* (9). Following attachment, the viral Envelope (Env) fuses with the cellular membrane. This process of fusion allows the HIV capsid that hold the core of the virus, or nucleus to enter in the cytoplasm of the infected cell (10). The capsid contains two enzymes essential for HIV replication, the RT, PR and IN. The capsid houses the viral RNA genome from cytosolic DNA sensors. Soon after entry into the cytoplasm, the virus loses its outer shell through a process called “uncoating” which is disassembly of a protective, conical capsid around the HIV-1 genome, during which most of capsid (CA) sheds off, while nucleocapsid (NC), Vpr, RT, PR and IN are still associated (11). Thereafter, NC disintegrate and leaves two strand of viral RNA naked and exposed and Viral RNA is then reverse transcribed into full-length double-stranded complementary deoxynucleic acid (cDNA) with the help of the viral Reverse Transcriptase (RT) enzyme. Newly formed complementary DNA (cDNA) interacts with viral and cellular proteins to form the pre-integration complex (PIC).

The PIC consists of viral proteins (including Vpr, Matrix and Integrase). After the proteins, enzymes, and newly formed viral cDNA are transported to the host cell nucleus, double-stranded linear viral DNA is inserted into the host genome in a process catalyzed by the virus-encoded Integrase (IN) to form a proviral DNA (12). The mechanism involves a series of nucleophillic attacks, the first of which removes the terminal 2 bases from the 3′ ends of the long terminal repeats and of the second which inserts the viral DNA into the host genome (13). When viral DNA is successfully integrated into the human genomic DNA, the provirus RNA can be transcribed and with the help of Transcriptional Transactivator gene (Tat), which binds to the 5' end of the long terminal repeats (LTR) region of the incorporated viral DNA. Thereafter, host Polymerase recognizes the integrated viral DNA as part of the host genomic DNA (13). During the late phase of the HIV-1 replication cycle, viral genes are transcribed and viral RNAs are exported from the nucleus to the cytoplasm, where mRNA strands are used as blueprint to make long chains of HIV-1 precursor proteins that are not able to function within the viral cycle to give rise to virus products. PR cuts up these long strands of new HIV particles into small individual of subunits to make the virus infectious and continues even after virus assembly. (14). Full length unspliced RNA serves as the mRNA template for Gag and Gag–Pol synthesis, as well as the genome for packaging (15). The Gag precursor contains MA, CA, NC and p6 domains, as well as two spacer peptides, SP1 and SP2. As part of the uncleaved Gag precursor, the MA domain targets Gag to the plasma membrane and promotes incorporation of the viral Env glycoproteins into the forming virions (16). CA drives Gag multimerization during assembly; NC recruits the viral RNA genome into virions and facilitates the assembly process; and the p6 domain recruits the endosomal sorting complex required for transport apparatus, which catalyses the membrane fusion step to complete the budding process (17). HIV-1 virion assembly occurs at the plasma membrane, within specialized membrane micro domains. The HIV-1 Gag (and Gag-Pro-Pol) polyprotein itself mediates all of the essential events in virion assembly, including binding the plasma membrane, making the protein–protein interactions necessary to create spherical particles, concentrating the viral Env protein, and packaging the genomic RNA via direct interactions with the RNA packaging sequence. These events all appear to occur simultaneously at the plasma membrane, where conformational change(s) within Gag couples membrane binding, virion assembly, and RNA packaging (18). Through a complex combination of Gag– lipid, Gag–Gag, and Gag–RNA interactions, a multimeric budding structure forms at the inner leaflet of the plasma membrane. The budding virus particle is ultimately released from the cell surface in a process that is promoted by an interaction between the late domain in the p6 region of Gag and host proteins,

most mediated by the host ESCRT endosomal sorting complexes required for transport (ESCRT) machinery (19, 20). These processing events generate the mature Gag proteins MA, CA, NC and p6, and two small Gag spacer peptides (SP1 and SP2). Gag cleavage triggers a structural rearrangement termed maturation, during which the immature particle transits to a mature virion characterized by an electron-dense, conical core (19–21). Among the Gag processing cascade, cleavage of SP1 from the C terminus of CA is the final event required for final CA condensation and formation of the conical core of virus particles. Virion maturation is essential for the released virus particles to become infectious and initiate a new round of infection (22).

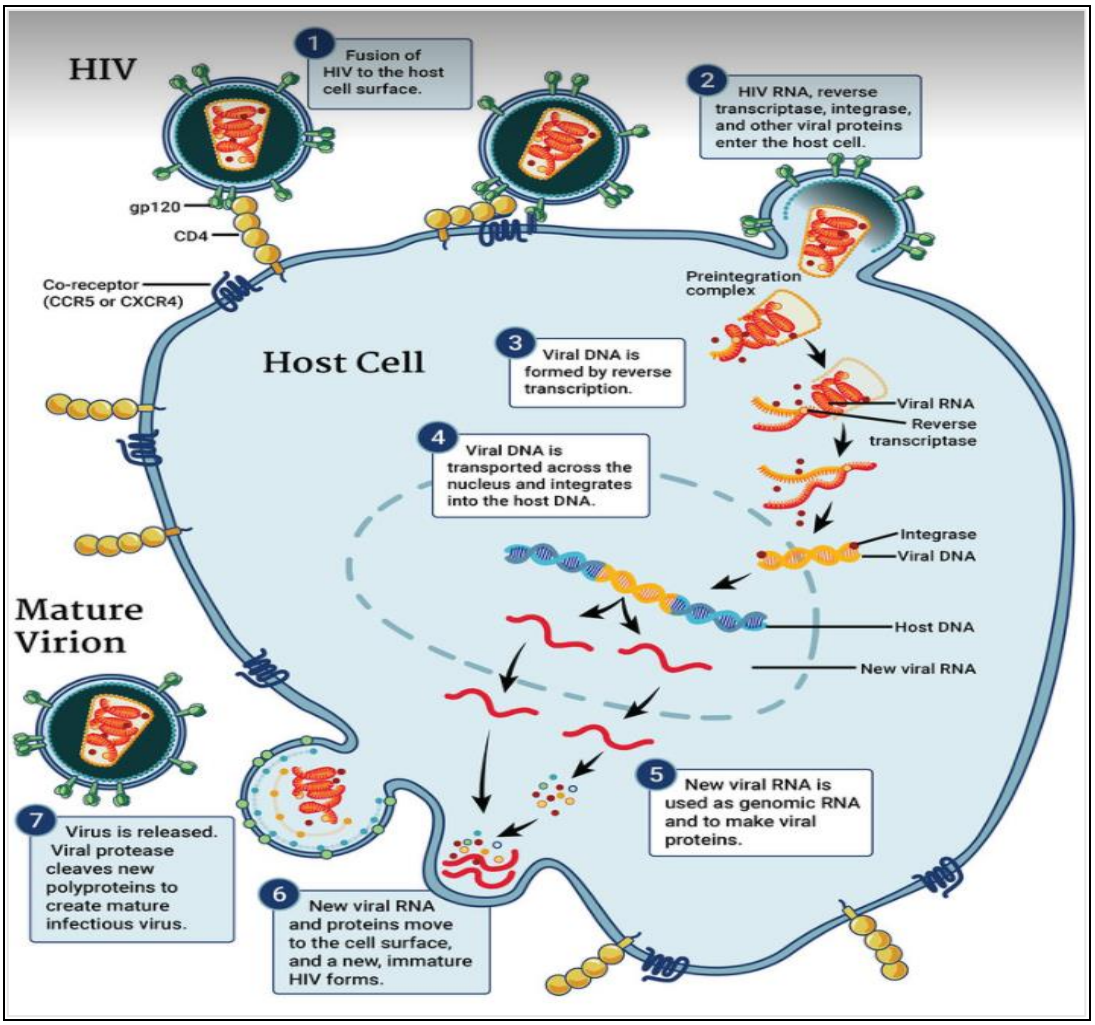

Figure 2: This infographic illustrates the HIV replication cycle, which begins when HIV fuses with the surface of the host cell. A capsid containing the virus's genome and proteins then enters the cell. The shell of the capsid

disintegrates and the HIV protein called reverse transcriptase transcribes the viral RNA into DNA. The viral DNA is transported across the nucleus, where the HIV protein integrase integrates the HIV DNA into the host's DNA. The host's normal transcription machinery transcribes HIV DNA into multiple copies of new HIV RNA. Some of this RNA becomes the genome of a new virus, while the cell uses other copies of the RNA to make new HIV proteins. The new viral RNA and HIV proteins move to the surface of the cell, where a new, immature HIV forms. Finally, the virus is released from the cell, and the HIV protein called protease cleaves newly synthesized polyproteins to create a mature infectious virus (23).

1.4. HIV-1 IN South Africa and Cameroon

1.4.1. South Africa

South Africa (SA) has a population of approximately 56 million people and remains the country that is the most heavily affected by HIV-1, with 7.7 million people living with the virus. SA has met the UNAIDS 90-90-90 target where by 90% of people living with HIV-1 were aware of their HIV-1 status by the year 2018 (24). The test and treat strategies in South Africa, regardless of the CD4+ T-cell count had made SA the country with the largest cART program in the world. According to UNAIDS in 2018, 240 000 new infections where recorded with an additional 71 000 AIDS related deaths. The Integrase (IN) strand-transfer inhibitor (InSTI), Dolutegravir (DTG) is now recommended by the World Health Organization (WHO) as part of salvage and / or first-line combination antiretroviral therapy (cART) (25). South Africa has roughly 4.8 million HIV-1 positive patients who are receiving cART (26). This equated to 62% of people living with HIV-1 in the country and 87% of all people living with HIV were virally suppressed (26).

1.4.2. Cameroon

In the West and Central Africa, Cameroon is the country with the highest prevalence rates of HIV/AIDS, with 560 000 people living with the virus out of a total population of 25 million people. In Cameroon, the lack of resources further limits the availability of treatment options in cases where patients require a change of their HIV-1 therapy regimen. Among 560 000 people living with HIV in Cameroon approximately 225 000 (39,3%) were accessing cART (27). In 2018, 23000 people were newly infected with HIV and about 18000 people died from AIDS related diseases. A huge effort has been made to meet UNAIDS 90-90-90 target were

about 74% of people know their HIV-1 status and 52% of people living with HIV are on treatment. In addition, 8% of pregnant women received cART to prevent Mother-to-child-Transmission (PMTCT) (26).

1.5. HIV-1 DIVERSITY IN SUB-SAHARAN AFRICA

HIV/AIDS is a major global health concern caused by two types of retrovirus, HIV-1 and HIV-2 (28). HIV is thought to have originated via cross-species infection (zoonotic transmission) from infected African primates to the human population. Cross-species infection gave birth to various types of HIV-1 groups; M (Major), O (Outlier), N (non- M, non-O), P and L (29). HIV-1 is characterized by an extensive genetic diversity, Figure 3. HIV-1 group M can be subdivided into various subtypes and many recombinant forms (30). Some of these subtypes, such as H and J, are not common and mostly found in Central Africa (31). Circulating Recombinant Forms (CRFs), Unique Recombinant Forms (URFs) and Subtype A are dominant in Eastern Africa and subtype D confined in Central Africa and Western Africa (32). HIV-1 group L was recently identified from the Democratic Republic of Congo (DRC) (33). HIV-1 group M subtype C is the driving force of epidemic and caused more than 75% of HIV-1 cases in sub-Saharan Africa, particularly in eastern and Southern Africa (34). South Africa, have extensively reported on HIV-1 subtype C as the predominant subtype that account for the majority of infections (35–37). The most diverse subtypes of HIV-1 is found in Cameroon. Approximately 1-6% of cases of HIV-1 infection in Cameroon and West-Central Africa is caused by HIV-1 group O. HIV-1 group N and P has only been isolated in Cameroonian individuals thus far and are rare (29,38,39). In Cameroon CRF02_AG subtype is the common cause of HIV-1 infection and account for approximately 58.2% of HIV infections, with at least 14.8% of infections caused by URFs. Other subtypes, such D, F2 and G, and CRF, 01, 11, 13, 22, 36, and 37 have been identified in Cameroon as well (40).

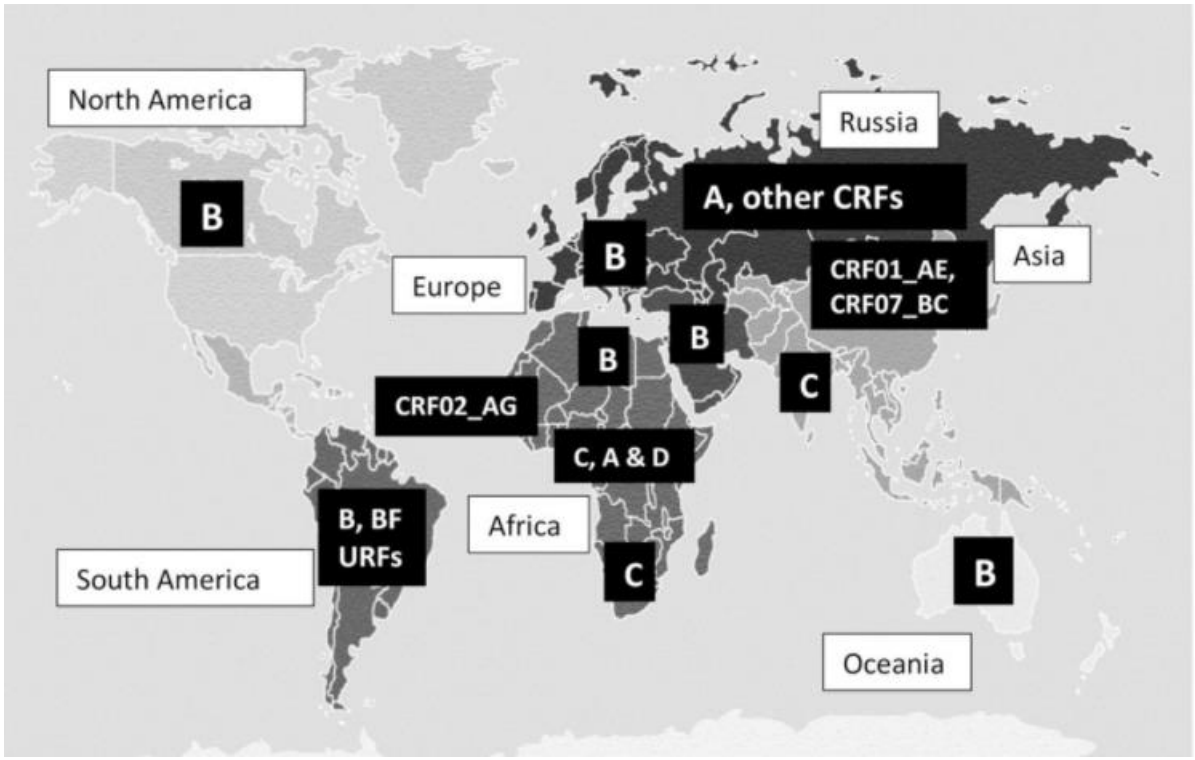

Figure 3 : Global distribution of major HIV subtypes. Map showing the global distribution of the major HIV subtypes and circulating recombinant forms from the review. Source: Map of the world modified to show HIV-1 subtype diversity worldwide. Map-menu.com.Source. HIV subtype diversity worldwide. Current Opinion in HIV and AIDS 14(3): 153-160, May 2019.

1.6. NATURALY OCCURING POLYMORPHISIMS (NOPs).

Naturally occurring polymorphisms (NOPs) are considered as secondary mutations that alone have no effect on resistance. The high error rate of reverse transcriptase (310–5 sites/ genome/replication cycle) provides tremendous scope for the generation of NOPs in HIV-1 open reading frames (ORFs) (41). To date, approximately 42 mutations within the HIV-1 IN gene have been associated with INSTI drug resistance. Naturally occurring IN gene polymorphisms may have important implications for INSTI resistance development. Of the 42 amino acids substitutions currently associated with INSTI resistance, 21 occurred as NOPs (42) However, their pre-existence might favour a more rapid evolution towards resistance, in combination with major mutations during therapy (43). For example, a novel next-generation INSTI termed MK-2078 with a higher genetic barrier for selection of resistance than either RAL or EVG was able to differentially select for a novel G118R substitution in IN in subtype C, compared to subtype B viruses (44). This mutation conferred only slight resistance to MK-2048, but gave rise to 25-fold resistance against RAL when it was present together with a natural polymorphic substitution at position L74M in CRF02-AG cloned patient isolates (45).

Other studies reported that NOPs can regulate INSTI susceptibility/resistance in non-B subtypes (46). A study by Franset *et al*., showed that the majority of viruses containing N155H also had at least one NOP mutation, L74M, E92Q, T97A, V151I, or G163R that confers large reduction in RAL susceptibility. The fact that patient viruses containing N155H without additional NOPs exhibited a broad range of RAL susceptibility indicates that other amino acid substitutions in the IN proteins of these viruses also influence RAL susceptibility (46). While, Ceccherini-Silberstein et al. reported several IN NOPs, for example, M154I, V165I, and M185L, that were positively associated with specific RT mutations (F227L and T215Y) in treated patients (47). *In silico* analysis of IN/DNA complexes predicted the impact of NOPs on the interaction between the DNA and INSTIs (48). Furthermore, two NOPs V82F and I84V present in HIV-1 A and C viruses were found to be associated with reduced binding affinity more than occurred in subtype B. (49).

1.7. HIV-1 INTEGRASE (IN)

1.7.1. STRUCTURE

The HIV-1 IN is a highly conserved protein. IN consists of 288-amino acids (32-kDa). It is synthesised from the portion of matured *pol* gene, of the Gag-Pol precursor from the C-terminal portion (46). IN is divided into three canonical domains. (i) N-terminal domain (NTD) (amino acids 1-49) that carries an HHCC motif analogous to a zinc finger, and effectively binds Zn^{2+} (50), possibly favouring protein multimerisation, a key process in integration (51) . (ii) The catalytic core domain (CCD) (50–212) which is indispensable for the catalytic activity and which is conserved between viral IN and transposases. (52) This CCD is also implicated in the binding of the viral DNA extremities mainly via the residuEs Q148, K156 and K159 (53) and the (iii) C-terminal domain (CTD) (213-288) binds non-specifically to DNA and therefore is mainly involved in the stability of the complex with DNA (54) . IN catalytic functions are maintained throughout the catalytic triad D64, D116, E152 consisting in two aspartates and one glutamate residues and the multimerisation of the protein (55). For example, Zn^{2+} facilitates the Mg^{2+}-dependent activity of IN by enhancing its multimerisation and cooperativity of DNA-binding (56). No complete structure has yet been determined for the integrase protomer (IN1–288), or for oligomers or complexes of these structures with DNA, due to poor solubility and interdomain flexibility problems. However, several structures of isolated domains or of two consecutive domains have been reported (57, 58).

1.7.2. ACTIVITY

HIV IN It is characteristic for all retroviruses, including HIV-1, to perform a catalytic integration of the proviral DNA (vDNA) copy of a RNA genome into the host target DNA (tDNA) (59). Viral IN is a key enzyme in the replication mechanism of retroviruses, mediating the covalent retroviral integration-insertion process of the vDNA into the tDNA (60). This integration process establishes productive permanent infection within the host cells, enabling replication and parallel transcription of the newly inserted provirus with other genes of the host organism (61). Once integrated, the provirus persist in the host cell and serves as a template for the transcription of viral genes and replication of the viral genome, leading to the production of new viruses (60). During this process the IN oligomerizes into a higher-order stable synaptic complex (SSC) containing two vDNA ends. This is a very important and crucial step in the replication cycle of HIV-1 and presents one of the major underlying difficulties in combating the HIV/AIDS pandemic to date (59)

In Figure 4 the DNA cutting and joining steps of retroviral DNA integration is shown. Initially, IN removes a GT dinucleotide from the 3′-terminus of each viral DNA end (3′-processing) and subsequently catalyzes concerted transesterification reactions (DNA strand transfer) to integrate the recessed viral DNA ends into the target DNA in a staggered fashion. Cellular chromatin-associated protein lens epithelium-derived growth factor (LEDGF)/p75 engages the IN tetramer in the pre-integration complex (PIC), which, in addition to the intasome, contains additional viral and cellular proteins, to target HIV-1 integration into active genes (62,63) . Recently in 2017, a full length three dimensional structure of the HIV-1B IN strand transfer complex (STC) intasome (PDB ID:5U1C) has been determined by Cryo-electron microscopy (CryoEM) methods (64). This structure (5U1C) has provided a first glimpse of nucleoprotein organization that could be used as a homolog model for the modelling of HIV-1C IN and other recombinant subtypes. Predicting a 3D structure for HIV-1C, and other subtypes, using 5U1C will assist in deducing the effect of known and/or novel HIV-1 cART resistant variants upon the IN structure. This is possible as 5U1C has a higher resolution (3.9 Å) and sequence identity (93.4%) to our target sequence, relative to other templates from prototype foamy viruses (PFVs). However, crystal structures of the PFV IN in complex with DNA and INSTIs are available for comparisons, as they contain both conserved DDE motifs and positions of the INSTI drugs making them useful for drug extractions (63, 65, 66).

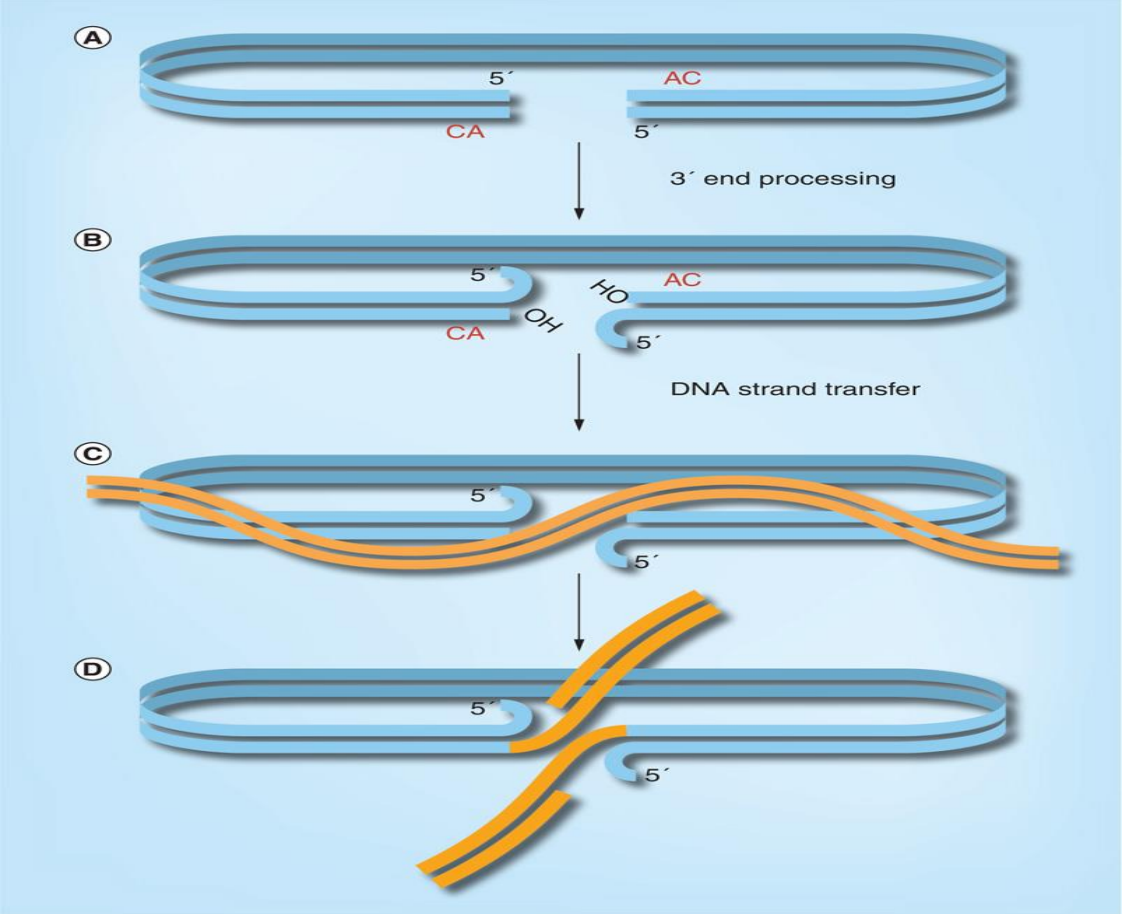

Figure 4: DNA cutting and joining steps of retroviral DNA integration. (A) The viral DNA synthesized by reverse transcription is initially blunt ended. (B) The 3′ end processing reaction removes two nucleotides from each 3′ end. (C) Next, in the DNA strand transfer reaction, the 3′ hydroxyls at the ends of the viral DNA attack a pair of phosphodiester bonds in the target DNA; in the case of HIV, the sites of attack are separated by five nucleotides on the two target DNA strands. (D) The result is the integration intermediate, in which the 3′ ends of the viral DNA are joined to the 5′ ends of the target DNA at the site of integration. Cellular enzymes to complete the integration process then repair the integration intermediate. Image reproduced with permission (67).

1.8. HIV DRUG TARGETS AND ARV CLASSES

The main steps of HIV viral replication include binding and entry, reverse transcription, integration, viral assembly, and budding. These steps form the basis for the targets of the 6 different ARV drug classes (Nucleos(t)ide reverse transcriptase inhibitors (NRTIs), Non-Nucleoside reverse transcriptase inhibitors (NNRTIs), Protease inhibitors (PIs), Integrase strand transfer inhibitors (INSTIs) and entry inhibitors (sub-divided as fusion inhibitors (FI) and inhibitors of co-receptor usage), Table 1. In total there are 31 United States of America (USA) Federal Drug Agency (FDA) approved HIV antiretroviral drugs currently being

marketed and the brand names of their co-formulated products are also given below in Figure 5.

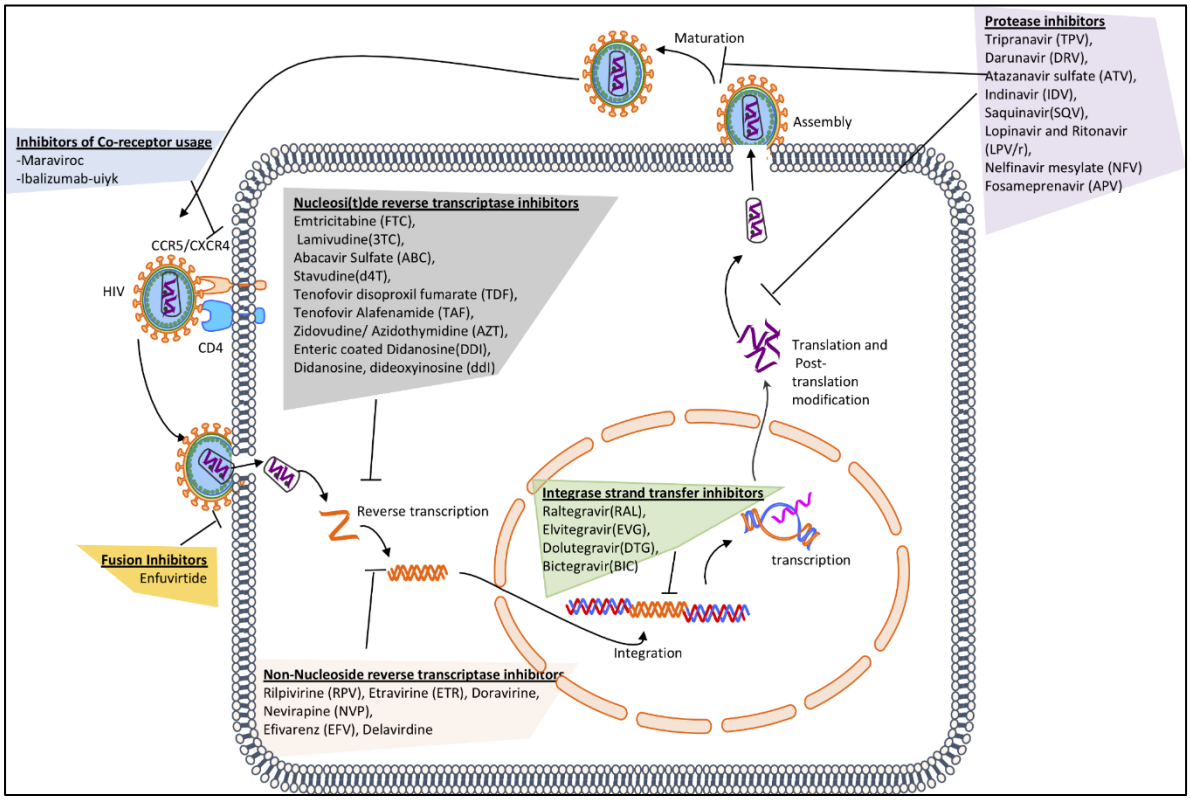

Figure 5: FDA approved antiretroviral drugs for HIV treatment shown to act on different stages of the HIV-1 replication cycle. Image reproduced with permission and adopted from Source: HIV-infected patients eje.bioscientifica.com.

Table 1: FDA approved Fixed-dose antiretroviral drugs for HIV currently marketed

Fixed-dose combinations	
Atripla (EFV+ FTC+TDF)	Prezcobix (DRV + Cobicistat)
Stribild (EVG + TDF + Cobicistat + FTC)	Trizivir (ABC +3TC +AZT)
Truvada (TDF + FTC)	Genvoya (EVG + cobicistat + FTC + TAF)
Triumeq (ABC + DTG + 3TC)	Biktarvy (BIC + FTC+TAF)
Complera (FTC + RPV + TDF)	Symtuza (DRV+ FTC +TAF + Cobicistat)
Combivir (3TC + AZT)	Dovato (DTG + 3TC)
Descovy (FTC + TAF)	Juluca (DTG + RPV)
Evotaz (ATV + Cobicistat)	Delstrigo (DOR + 3TC + TDF)
Epzicom (3TC + ABC)	Symfi (lo) (EFV +3TC +TDF)
Odefsey (FTC + TAF + RPV)	Cimduo (Temixys+ 3TC+TDF)

1.9. MECHANISMs OF ACTION OF DIFFERENT cART drugs

1.9.1. *Reverse Transcriptase*

Reverse Transcriptase (RT) is an enzyme that contains two enzymatic activities, a DNA polymerase that can copy either an RNA or DNA template, and an RNase H, which degrades RNA if the RNA is part of an RNA/DNA duplex. RT uses these two enzymatic activities to convert the single-stranded RNA genome of the virus into a double-stranded DNA that can be integrated into the genome of the host cell. The synthesis of a DNA copy of the viral genome is a crucial step in the life cycle of the virus, and RT has, for that reason, been the target of a number of different anti-HIV drugs (68).

1.9.1.1. *Nucleoside Reverse Transcriptase Inhibitors (NRTIS)*

NRTIs inhibit the replication of HIV via two channels, phosphorylating the 5′-triphosphate form by cellular kinase enzymes. The NRTIs becomes, incorporated into the enzyme-template-primer complex, where natural 5′-deoxynucleoside triphosphates attaches. The NRTIs do not

have the 3′-hydroxyl group on the deoxyribose moiety required for binding of nucleotide and this result in chain termination (69).

1.9.1.2 *Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIS)*

NNRTIs inhibit the replication process of HIV-1 by interacting with an HIV-RT pocket region (70,71). The binding site of the NNRTI is located very close to the substrate-binding site at a distance of 10 angstrong (Å) away from the polymerase active site. The relationship between these two sites helps in increasing the effectiveness of RT inhibitors. NNRTIs causes an effect on the three-stranded β-sheet in the p66 subunit by repositioning it. Once it is repositioned, the active catalytic site in the inactive p51 subunit becomes locked. All NNRTIs, when they bind to the HIV-1 RT pocket site, change the structural shape of the binding site into “butterfly-like “shape and then block the binding of the RT to the primer during reverse transcription (72).

1.9.2. Integrase strand-Transfer Inhibitors (INSTIs)

Two strategies used by the INSTIs to block integration of the virus into virus genome have been developed; 3’ processing and DNA strand transferase. INSTIs compete with the viral DNA for binding to the IN DNA complex. The viral DNA recognizes the binding site next to the catalytic triad, which opens after a change in structure caused by the binding and 3' processing of the viral DNA (73) . INSTIs chelate the Mg^{2+} cation required for the activity of IN. Secondly, INSTIs target and bind to the IN vDNA complex , near the 3' end of the host DNA and thereby blocking the binding of viral DNA, resulting in inhibition of the strand transfer reaction , ultimately inhibiting the Integration of viral DNA (74).

1.9.3. *Protease Inhibitors (PIs)*

Protease Inhibitors (PIs) interfere with the process of forming new infectious viral particles. The viral Protease is engaged in virion maturation (75). Protease targets the amino acid sequences in the Gag and Gag–Pol polyproteins, which must be cleaved before nascent viral particles (virions) can mature. Cleavage of the Gag polyprotein produces three large proteins (p24, p17 , and p7) that contribute to the structure of the virion and to RNA packaging, and three smaller proteins (p6, p2 and p1) of uncertain function (76). PIs are small molecules that bind to the active site of the Protease and therefore compete with its natural substrates. PIs contain a synthetic analogue of the amino acid sequence of the Gag–Pol polyprotein at position that is cleaved by the Protease. PIs prevent cleavage of Gag and Gag–Pol protein precursors in

acutely and chronically infected cells, arresting maturation and hence blocking the infectivity of nascent virion (77). This results in production of defective viral particles that are unable to continue with replication (78).

1.9.4. *Inhibitors of co-receptor usage (CCR5 antagonists)*

CCR5 and CXCR4 are the main HIV co-receptors involved in virus entry and cell-to-cell spread. R5-tropic viruses are nearly always involved in the initial infection, while HIV strains using the CXCR4 co-receptor are observed only seldomly in the early infection (79). The first step of the HIV-1 cell entry comprises the interaction of the Envelope glycoprotein gp120/gp41 with the host receptor CD4, and the binding to chemokine receptors CCR5 or CXCR4 (80). CCR5 antagonist bind in a side pocket region of the CCR5 molecule transmembrane cavity, thus preventing the interaction between the HIV gp120 and CCR5. The CCR5 imitates the functions of the chemokines, which are found to be natural ligands of the chemokine co-receptor, thus inhibiting their effect (81).

1.9.5. *Fusion Inhibitors (FIs)*

In order for the HIV virus to gain entry to the intracellular human machinery, which all viruses require for replication, the virus must fuse with the human cell membrane. This occurs in a complex sequence of events following attachment of the HIV-1 surface glycoprotein 120 (gp120) binding site to human cells expressing CD4 receptor molecules. After binding, gp120 changes shape to allow the viral glycoprotein 41 (gp41) to form a pore in the membrane through which the virus can enter (82). Fusion inhibitors act extracellularly to prevent the fusion of HIV to the CD4 or other target cell. Therefore, fusion inhibitors are drugs that blocks the second step in the fusion pathway by binding to the HR1 region of glycoprotein 41 (gp41). This mechanism does not allow HR1 and HR2 to fold properly, thereby preventing the conformational change of gp41 required to complete the final step in the fusion process (83).

1.10. HIV drug resistance mechanisms, viral fitness and role of secondary mutations in virus evolution

The high mutation rate of HIV-1 is an important biological factor contributing to HIV drug resistance. Approximately 1 - 3 mutations occur in each RNA genome target per round of replication (84). This is due to the error prone HIV-1 RT enzyme that lacks a 3’-> 5’ proofreading activity. Another contributing factor is that HIV-1 has a high recombination rate when co-infection of a cell occurs with more than one variant (85). The high recombination

rate is because of the mechanism (known as Copy choice) by which HIV-RT switches from one template to another during viral cDNA synthesis (86). Earlier studies have shown that the HIV replication efficiency is related at least in part to the processivity of RT. Additionally, *in vivo* peripheral blood mononuclear cell (PBMC)-based replication kinetic assays demonstrated that the L74V variant replicates two- to three fold slower than the M184V virus, and there was a 2.4% loss of fitness for M184V in comparison to WT virus and a 5.7% loss of fitness for L74V virus compared to M184V virus in growth competition assays (87).

In a particular study (88), the relationship between fitness and replication fidelity was investigated. The study suggested a mutation rate close to the HIV-1 WT strain represents the optimum for viral evolutionary adaptive forces. To add on, the study proposed a model that assumes that high replication fidelity implied evolution of HIV-1 to a lower mutation rate and factors such as changes in enzyme processivity and increased time for base recognition would require high energy demands that would inevitably result in a reduction of viral fitness. This could possibly explain why discriminatory mutations in the HIV-1 *pol* gene results in less fit viral mutants (85). The presence of secondary mutations in this gene, changes enzyme kinetics in a way that overcomes the energy constraints needed for correct base incorporation during viral synthesis and this compensates for fitness loss (88). Major NRTI resistance-associated mutations are known to decrease drug susceptibility of the virus at the expense of reduced viral fitness, whereas the general accepted dogma for secondary (accessory or compensatory) mutations is that these mutations maintain relative effect of decreased drug susceptibility imposed by major mutations but will enhance viral fitness (89). Secondary/compensatory mutations have been observed in some studies (90, 91), playing a role in restoring viral fitness. The precise molecular mechanisms for drug resistance of each ARV class is discussed below.

1.10.1. *NRTIs resistance mechanisms*

Drug resistance remains a central challenge in the success of HIV therapies, as resistant viruses can be transmitted. The prevalence of resistant viruses is increasing in untreated HIV-1 patients. For the virus to replicate and be transmitted, HIV-1 RT must be able to complete viral DNA synthesis, and NRTI-resistant RTs must retain the ability to incorporate normal dNTPs with reasonable efficiency. Two rare mutations associated with multiple NRTI drug resistance are insertions or deletions at codon 69 (known as T69ins/ T69del) and the Q151M (92). The T69ins is known to co-occur with Thymidine Analogue Mutations (TAMs) and causes high

level clinical resistance to all NRTIs (92). The Q151M mutation is known to usually co-occur with accessory mutations that tend to promote the efficiency of the phosphorolytic excision reaction of NRTIs (93). As part of the precursor of the multi-nucleoside resistance (MNR) complex that include accessory mutations F75I, F77L, F116Y and A62V, Q151M causes high-level clinical resistance to AZT, D4T, ddI and ABC and low – intermediate resistance to TDF, 3TC and FTC (94). Combinations of Q151M and T69ins mutations are rarely observed to occur, but when they appear severely limit therapy options as they incur resistance to all NRTIs (95). Major NRTI mutations are known to decrease drug susceptibility of the virus at the expense of reduced viral fitness whereas the general accepted dogma for secondary (accessory or compensatory) mutations is that these mutations maintain relative effect of decreased drug susceptibility imposed by major mutations but will enhance viral fitness (89). Secondary/compensatory mutations have been observed in some studies playing a role in restoring viral fitness (96).

1.10.2. *NNRTIs resistance mechanisms*

HIV-1 replication is continuous, occurs vigorously in infected individuals and can subsequently lead to HIV-1 drug resistance against RT inhibitors. Resistance of NNRTIs is caused by mutations that are located at the amino acid (aa) residues aligning the NNRTI-binding pocket site. The most common RT mutation pathways associated with resistance to NNRTIs include the K103N and Y181C mutations. These mutations can cause reduce susceptibility to nevirapine (NVP) (97). Other NNRTI mutations include L100I, E138K ,V106A, QA45M and P236L that can cause resistance to NNRTIs. NNRTI-resistance mutations can occur singly, or in combinations (98). K103R is a polymorphism that is not selected by NNRTIs and in isolation has minimal effect on NNRTI susceptibility. In combination with V179D, however, it is associated with about 10-fold resistance to each of the NNRTIs. Another study found that viruses carrying E138K or Y181C mutation each had reduced RT activity relative to that of the WT. The introduction of the secodary mutations into the RT-Y181C mutant virus significantly reduced RT activity and viral fitness (99). The two most common resistant mutants, K103N and Y181C, show a fitness that is very close to that of WT virus. Less-frequent mutations, such as V106A, Y188C and G190S, are associated with lower viral fitness.

1.10.3. *INSTIs resistance mechanisms*

Development of primary mutations against INSTIs usually arises under selective drug pressure and renders down the efficacy of the drugs against the virus at a cost of increasing replication capacity of the virus. This is due to change in the active site structure where the drug bind (100). Common pathways that lead to INSTIs resistance are located at IN residues T66I/A/K, E92Q/G, T97A, S147G, Q148R/H/K, and N155H (46,101,102). Another study in which IN mutations were purposely introduced in order to determine susceptibility and infectivity, reported that each of the G118R, Y143R, Q148R, R263K and G140S/Q148R mutations, when introduced into SIV, impaired infectiousness and replication fitness compared to WT virus (103).

In South Africa, a study conducted by Brado et al., 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 (104). some mutations such as S119R have been shown to increase the resistance to INSTIs when combined to the primary mutations Y143C, Q148H, and N155H (105). In addition, another study surprisingly, found no mutations in the IN gene but rather five mutations located in the nef region, one mutation six nucleotides upstream of the 3′ PPT and four other changes clustered in the 3′ end of the 3′ PPT, inside the G tract, resulting in GCAGT instead of GGGGGG. The location of these mutations in a highly conserved region is very surprising, and the mechanism involved in this change remains unknown. The disruption of the 3′ PPT could lead to modification of the reverse transcription (RT) process, resulting in linear DNA that is no longer fully compatible with integration, explaining the decrease in infectivity. This nonconventional linear DNA could impair DTG binding, explaining the resistance to DTG. (106). Secondary mutations that increase the fitness of the resistant viruses have been identified in both pathways. In particular, the secondary G140S mutation, observed in tandem with the Q148H mutation, rescues a replica- ive defect due to the presence of the primary mutation Q148H (107).

1.10.4. *PI resistance mechanisms*

Protease inhibitors (PIs) which compete with natural cleavage sites, strongly impair viral infectivity and have proven to be highly valuable in the treatment of HIV-infected individuals. The emergence of specific mutations in Protease that directly decrease the inhibitor binding affinity for the enzyme active site is considered the primary mechanism of PI resistance (108). Both primary and secondary mutations modify the shape and size of the substrate-binding

cavity of the Protease (109), by enhancing the cleavage of Gag and Gag-Pol precursors (110). This enlargement seems to have more important consequences on the binding of the inhibitors, most of which are dependent upon a strong and tight interaction with the active site of the enzyme, than on the binding of the natural substrates of the protease in Gag and Gag-Pol. D30N is a nonpolymorphic NFV-selected substrate-cleft mutation that causes high-level resistance to NFV. The Protease mutations A71V/T, T74S, M89I/V and L90M (111) showed that mutation M89I alone did not confer primary resistance to PIs, but in viral isolates that displayed M89I/V and L90M together, the loss of susceptibility to nelfinavir (NFV) was significantly higher than in isolates with L90M alone (112). Besides the well documented effect of Gag cleavage site mutations on resistance-associated loss of viral fitness, evidence has been accumulating that these mutations could also directly affect HIV susceptibility to PIs in a manner that is independent of their effect on fitness. (113)

1.11. MOLECULAR MODELLING

Molecular modelling can help in determining the 3D structure of a protein sequence and predict its function from a sequence. It can also be used to understand protein-drug interactions and help us to assess the influence of mutations on the protein structure. Molecular modelling methods, such as molecular docking, molecular dynamics (MD) simulation and binding free energy calculation, have been proved to be very useful tools for protein-ligand and protein-protein interactions study(48,114). In a previously a molecular modelling study combining molecular docking, molecular dynamics simulation, and binding free energy calculation was performed to investigate the interaction details of HIV-1 IN catalytic core domain (CCD) with two recently discovered LEDGINs BI-1001 and CX14442, as well as the LEDGF/p75 protein (115). In addition Molecular modelling was useful in the discovery of known and widely used drugs, such as HIV PIs (116). In our study, we will use molecular modelling approach to understand if second-generation drug, DTG, will be able to retain efficacy against selected RAL and EVG known resistance mutations in an various subtypes on IN protein. This methods will provided us with a unique opportunity to model the structure of other IN HIV-1 to interrogate the effect of other known drug. Overall, computational methods play a significant part of the drug lead discovery and indispensable to study drug resistance mutations that can develop in future to prevent possible treatment failures (117).

1.11.1. HOMOLOGY MODELLING

Predicting the 3D structure of a protein derived from its own amino acid continues to be a scientific problem. The first and most accurate approach used to predict 3D structure is "comparative" or "homology" modelling (118). Homology modelling is a method used to predict a protein structure from its sequence based on the evidence that proteins with similar sequences have same structures (118). It relies on identifying known protein structures that they share similarity with the targeted structural sequence and relies on the production of sequence alignment that maps the residues on template sequence and target sequence. It has been reported that protein of a sequence is less conserved when compared to a structure of a protein amongst homologues (119). 3D protein structures gives valuable knowledge about the molecular and functionality of the protein structure. Although, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been used towards solving the structural solution experimentally, but these methods requires more time to be performed with no guarantee of success. About 84 645 of experimentally solved data are found on Protein Data Bank (PDB) (120). While, the number of known different structures is much less when compared to the number of known protein sequences in SWISS 64 PROT and TrEMBL databases (about 850,000 sequence entries) (121). There is a lack of knowledge and information on protein structures. Comparative homology modelling stands out to be the reliable methods that can be used to generate a 3D model of a target protein using its own amino acid sequence (122). Experimentally solved 3D structural template that have similar amino acid sequence with a target protein is required to build a successful model. At least, 30% or more of similarity between target sequence and template can be used to generate a model (118). Homology is made up of four processes (Figure 6). (1) identification of known 3D structure(s) of a related protein that can serve as template; (2) sequence alignment of target and template proteins; (3) model building for the target based on the 3D structure of the template and the alignment; (4) refining/validation/ evaluation of the models. All four steps can be repeated until a perfect model is built (123). Models by definition are abstraction and may have errors. MODELLER is a computer program for comparative protein structure modelling (124). Various parameters such as discrete optimized protein energy (DOPE) score, (125) template. modelling (TM) score, (126) and RMSD value (127) are used for comparisons and identifying best models. Depending on the similarity of the sequence and the alignment quality, the homology models accuracy, when compared with the actual experimental structure can be up to approximately ~1–2 Å C_α atom RMSD (root-mean-square deviation distance between

corresponding C_α atoms) (128,129). As a general rule, models with similarity sequence of ≥50% are accurate enough for applications in drug discovery, models built with similarity between 25-50% can be applied in assessing the druggability and experimental study design such as mutagenesis and models with similarity of between 10-25% are speculative at best (130). The presence of inserts and loops are thought to be the limitations in homology modelling. This makes modelling challenging without availability of template data (131).

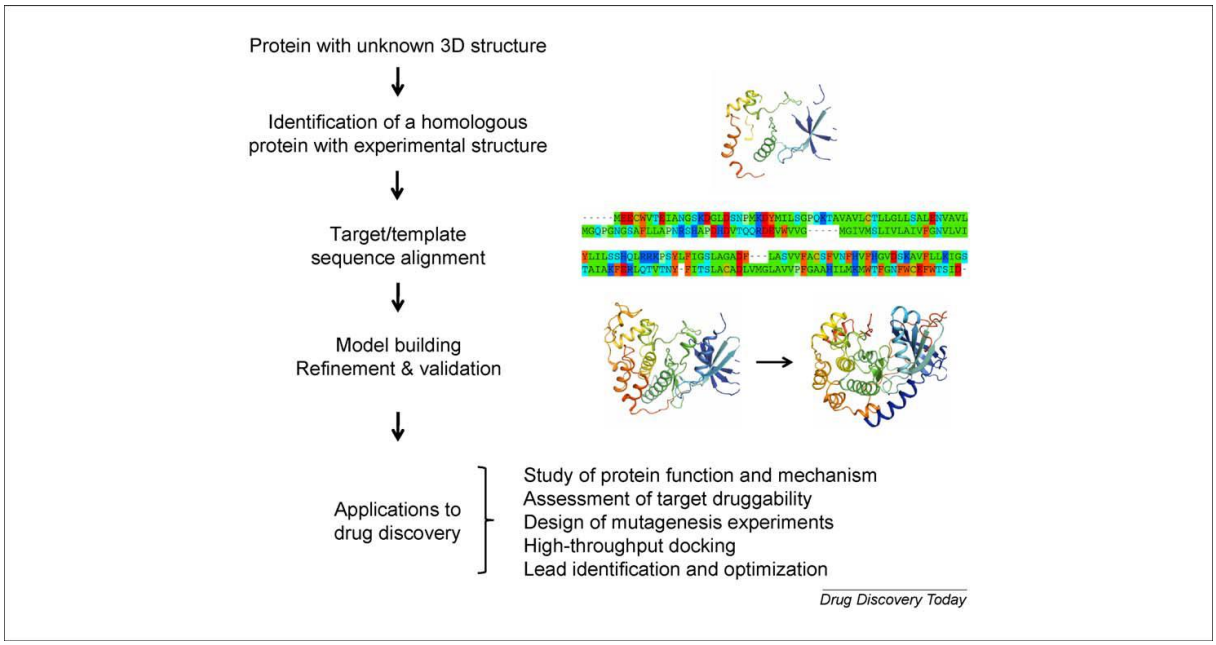

Figure 6: Outline of the homology modeling process and its applications in drug discovery. Image reproduced with permission from drug discovery today and adapted from informatics reviews (118). Given the sequence of a protein with unknown structure, the first step is the identification of a related protein with known 3D structure that serves as template. An alignment of the target and template sequences is necessary to assign the correspondence between target and template residues. A model is then built for the target based on the alignment and structure of the template, and further refined and validated. This figure was prepared using pymol (http://www.pymol.org).

1.11.2. MOLECULAR DYNAMIC SIMULATION

Molecular Dynamics (MD) is a simulation of atoms movement in a physical state and in a given system. Interactions of molecules and atoms is time-dependent, taking approximately

ten's-hundreds of nanoseconds (ns), reaching up to milliseconds generating a view of the atoms movement (132).

The dynamical system of these entities are calculated by solving the Newton's equations of motion (d2ri (t)/ dt2 =Fi (t) / mi) for a system of interacting particles (Figure 7). Forces between the potential energy of the system and atoms are described by the force fields (133). Commonly applied force fields for molecular dynamics simulation and analysis include several computer software versions from OPLSAA (134), CHARMM (135) AMBER (136) and GROMOS (137). Common MD softwares are GROMACS (138), AMBER (139) and NAMD (140). The molecular structures encode the conformational dynamics of a protein and this is a critical element for the functioning of a protein. The knowledge for understanding how proteins function, requires basic understanding of the link between 3D structure, generated with increased rapidity's by X-ray crystallography or nuclear NMR spectroscopy, and MD, that is not easy to obtain experimentally. MD simulations is a powerful computational tool that can be used for the explorations of changes on structural energy. Landscape accessible to these molecules and a sharp increase in power of computational analysis with better methods that makes the application of simulation on structural biology more interesting and exciting (141). MD simulations provide connections between dynamics and structure by allowing the manipulations of the structural energy landscape accessible to protein molecules (142,143). The first MD simulation of small proteins in vacuum (9.2-ps) was first reported in 1977. The availability of super power computing programs, make it easy to perform simulations of large proteins, in different systems. Molecular dynamic simulations can provide detail information of individuals at a level of atomic motions as a function of time. This can be used as a method to respond to questions, concerning the properties of a model system better than experiments (141). A successful MD simulation is dependent on the choice of a suitable energy function for describing the inter and intramolecular interactions (144). Besides the success of MD, principal challenges still makes MD difficult (145). The accuracy of classical MD simulations is not good enough to capture protein behaviour at high concentration, as the current force field parameters cannot describe the crowded microenvironment/ secondary structure of biomolecules (146).

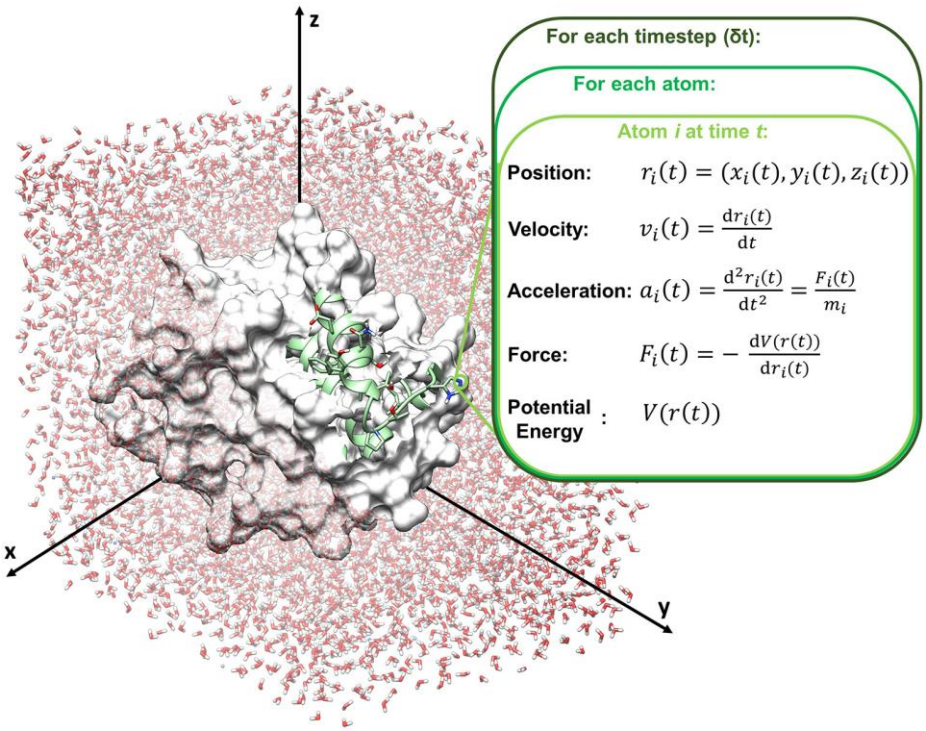

Figure 7: Schematic representation of a molecular dynamics cycle. Image reproduced with permission from Frontiers in pharmacology and adapted from experimental pharmacology and drug discovery informatics reviews (147)

1.11.3. MOLECULAR DOCKING

Molecular Docking is a structure-based virtual screening (SBVS) that is used to put computer-generated 3D structures of ligands into a target protein structure in different orientations, position and conformations (148). Docking is more applied in predicting the best binding orientation of a drug against a target protein structure, enabling the predictions of binding strength and drug activity (147). (Figure 8). Molecular docking generates possible structural orientations for protein-binding site, where drugs can possibly bind. (147). Molecular docking consists of two processes; pipelines to perform structural orientations sampling and a scoring function, which gives a score to each generated position that can be used when docking a ligand to a receptor (149–151). The favoured orientation is possibly used to predict binding strength between the protein-ligand using the scoring functions (152). Scoring functions play the role of poses selector, used to assume positions to differentiate a supposed right binding modes and binders from non-binders from a pools of possible assumed positions generated by the sampling system engines. Association of the native-bound to the global minimum of the energy hypersurface is performed by the scoring function. Three different types of scoring functions

are empirical scoring functions, knowledge-based scoring functions and force-field based scoring functions. The most popular used software programs for docking is AutoDock(153), DOCK (154), FlexX (155), Glide (89), GOLD (157), Molegro Virtual Docker (158), AutoDock Vina (159) and Surflex (160), to name the few. Docking process involves the following steps, preparations of active site prediction, ligand, the protein and docking. The docked ligand against protein and interactions are analysed (148). The best docked-ligand complex, get picked and scored by the scoring function. MM-PBSA and MM-GBSA are methods that employs two combination of molecular mechanisms (MM) energies, polar and nonpolar solvation terms, and an entropy term for calculations of free binding energy using the equation (ΔGbind; Massova and Kollman) from the change between the bound complex (ΔGcom) and unbound receptor (ΔGrec) and ligand (ΔGlig) in solution [Eq: ΔGbind= ΔGcom -ΔGrec - ΔGlig] (161). The most explored and successful breakthrough in drug discovery, it was the development of NNRTI RPV that was approved by FDA in 2011. RPV was developed by combining different synthetic chemicals structures with extensive antiviral screening; bioavailability and followed by clinical trials on the safety using animals models and use of molecular modelling approach, that include the analysis of 3D structures and molecular modelling (162). Lack of confidence and ability by the scoring function in giving accurate binding energies is a major limitation in molecular docking.

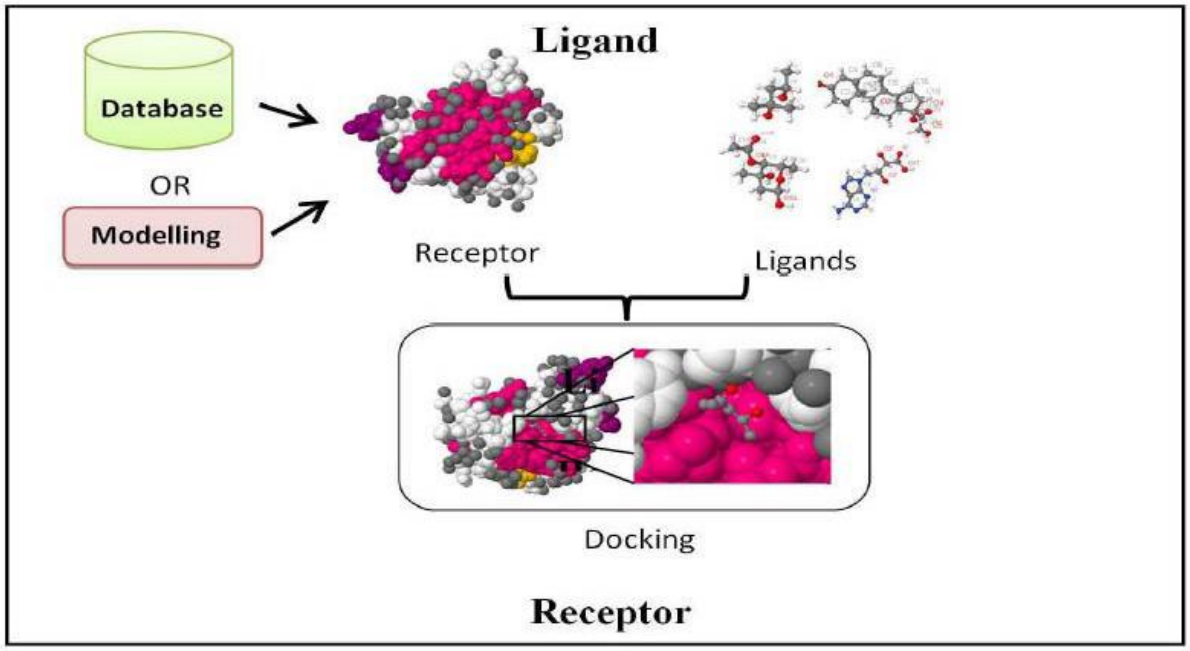

Figure 8. Molecular docking flow chart. A typical docking workflow. This flowchart shows the key steps common to all docking protocols. The 3D structures for the target macromolecule and the small molecule must first be chosen, and then each structure must be prepared in accordance with the requirements of the docking method

being used. Following the docking, the results must be analyzed, selecting the binding modes with the best scores. Source, and included with permission.Image reproduced with permission from JSM Chemistry and adapted from a Review on Molecular Docking (163).

2. CHAPTER TWO: STUDY FINDINGS

2.1. AIMS AND OBJECTIVES

The proposed study includes structural work to understand the impact of natural occurring polymorphisms (NOPs) and known INSTI mutations on the protein structure of HIV-1 Integrase. The aim of this project was to determine which polymorphisms and mutations screened may affect the efficacy of the INSTIs, and further understand the effect on binding of DNA and DTG to non-B subtypes from South Africa and Cameroonian sequences.

My thesis is driven by the hypothesis that HIV-1 IN gene fragment sequences obtained from sub-Saharan Africa, particularly Cameroon and South Africa, have specific naturally occurring polymorphisms that might reduce the efficacy and binding strength of second-generation INSTIs, like DTG.

This study is a descriptive experimental design study, that investigated the impact of Integrase NOPs among different HIV-1 group M subtypes. The specific major objectives were:

I. To determine the primary and secondary resistance-associated mutations (RAMs) to INSTI naïve patients from Cameroon (**Chapter III)**.
II. To determine the RAMS of HIV database derived sequences from Cameroon (**Chapter IV**).
III. To identify the HIV-1 resistance mutation profiles of patients failing second-line cART regimens in South Africa (**Chapter V**).
IV. To determine the structural effects of NOPs on DTG binding to IN (**Chapter VI**).
V. To compare the effects of NOPs on INSTI drug binding to various HIV-1 IN subtypes using molecular modelling, docking and simulation studies (**Chapter VII)**.
VI. To investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding (**Chapter VIII**).

The results of the study are summarized below in the form of manuscript or article published.

The goal of **Chapter III, IV** and **V** was to investigate prevalence of HIV-1 diversity and the HIV-1 drug resistance associated mutations (RAMs) from INSTI naïve patients from Cameroon and South African. Natural occurring polymorphisms of CRF02_AG subtype identified from **Chapter III, IV and V**, and those that are statistically enriched, were explored to further studied to try and understand their impact on the IN protein structure stability **Chapter VI** and see if these polymorphisms they exert any impact on INSTIs binding ability, using molecular modelling approach (**Chapter VII).** Known HIV-1 drug resistance mutations against INSTIs identified from South Africa and Cameroon cohorts (**Chapter III-V**), were further investigated to study their structural impact on the IN HIV-C that is the driving force of epidemic in South Africa (**Chapter VIII**). In (**Chapter IX**), I give a summary of the study. In **Chapter X**, I give general discussion, conclusions and future remarks.

Chapter 3: HIV-1 Integrase Diversity and Resistance-Associated Mutations and Polymorphisms among Integrase Strand Transfer Inhibitor-Naive HIV-1 Patients from Cameroon.

3.1. **Journal article**

The article is published in AIDS Res Hum Retroviruses. 2020 May; 36(5):450-455. doi: 10.1089/AID.2019.0264.

3.2. **Author's list**

Mikasi SG, Gichana JO, Van der Walt C, Brado D, Obasa AE, Njenda D, Messembe M, Lyonga E , Okomo O, Cloete R, IkomeGM, Jacobs GB.

3.3. **Author's contribution**

In the enclosed manuscript, I confirmed that I am the first author. I collected the samples, performed laboratory experiments and data analysis. I wrote the first draft and was responsible for subsequent manuscript updates received from the co-authors.

3.4. **Background**

In Cameroon, resistance testing remains challenging, due to the lack of laboratory infrastructure and cost involvement. The high genetic diversity of HIV-1 in Cameroon contribute negatively in optimizing laboratory protocols to amplify, sequence and characterize the IN gene for diverse subtypes. Though HIV-1 CRF02_AG is dominant in the country, all HIV strains can be found in the country, complicating HIV diagnostics.

3.5. **Main findings**

For this study we PCR amplified 56 patient samples targeting the HIV-1 IN gene fragment. Subsequently, for subtyping and genotypic drug resistance mutation analysis we only used 37 sequences with good quality and fragment size of ≥500 bp. The majority of the sequence were infected with subtype CRF02_AG (54%) and 45.9% were infected with other subtypes, such

as (A, CRF36_cpx, F,G and C). RAMs against INSTIs were identified in 18.9% of the sequences, of which major RAMs (Y143R/C/D/G and P145S) were detected in 5.4% of the sequences. Three (8.1%) of the sequences had accessory mutations, with one sequence having solely E157Q and the other sequence had Q95K. Another patient sequence had a combination of triple accessory mutations (G140E, E157Q, and G163R). Of the 288 IN positions, 53 (18.4%) had at least one amino acid NOP present in 0.5% or more sequences, including 49 (17.1%), of which two or more polymorphisms were present. Through our test, the two NOPs namely: VI72I and R269K occurred with p-values ≤0. 05.

3.6. **Study significance**

This was a novel study to report the possible appearance of IN mutations against INSTIs naïve patients. Mutations can interfere with the success of therapy. Therefore, proper monitoring of patients on therapy should be done, to reduce the risk of transmitted drug resistance (TDR).

3.7. **Conclusion**

Our study identified RAMs against INSTIs among treatment naïve patients from Cameroon. The study supports the use of DTG as part of first-line therapy in Cameroon for patients who are failing first-line and second line therapy regimens. However, resistance testing is recommended before initiation of treatment.

3.8. **Published article**

AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 36, Number 5, 2020

DOI: 10.1089/aid.2019.0264

HIV-1 Integrase Diversity and Resistance-Associated Mutations and Polymorphisms Among Integrase Strand Transfer Inhibitor-Naive HIV-1 Patients from Cameroon

Sello Given Mikasi,[1] Josiah Otwoma Gichana,[2] Cheri Van der Walt,[1] Dominik Brado,[3] Adetayo Emmanuel Obasa,[1] Duncan Njenda,[1,4] Martha Messembe,[5] Emilia Lyonga,[5] Okomo Assoumou,[5] Ruben Cloete,[6] George Mondinde Ikomey,[5] and Graeme Brendon Jacobs[1]

Abstract

The World Health Organization (WHO) has put forth recommendations for the use of integrase (IN) strand transfer inhibitors (INSTIs) to be part of the first-line combination antiretroviral therapy regimen to treat HIV infections. The knowledge of pretreatment drug resistance against INSTIs is still scarce in resource-limited settings (RLS). We characterized the *integrase* gene to identify resistance-associated mutations (RAMs) in 56 INSTI-naive patient viral sequences from Cameroon. Study analysis used 37 sequences with fragment size ≥500 bp or of good quality .The majority of the sequences were identified as CRF02_AG 54.% ($n=20/37$) and 45.9% ($n=17/37$), other subtype viral sequences include (A, CRF36_cpx, F,G, and C). A total of 18.9% ($n=7/37$) of the sequences had RAMs, with only 5.4% ($n=2/37$) having major RAMs (Y143R/C/D/G and P145S), against INSTIs. Accessory RAMs were present in 8.1% ($n=3/37$) of the sequences, of which one sequence contained solely E157Q, and another Q95K. One patient sequence had three accessory RAMs (G140E, E157Q, and G163R). We identified major RAMs to INSTIs, which might have a potential clinical impact to dolutegravir rollout in RLS, including Cameroon. This is the first study to describe RAMs among INSTI-naive people living with HIV-1 (PLHIV-1) infected with CRF02_AG and other subtypes in Cameroon.

Keywords: HIV-1, integrase, resistance, Cameroon, diversity, CRF02_AG

INTEGRASE (IN) STRAND TRANSFER INHIBITORS (INSTIs) are the newest class of drugs used as combination antiretroviral therapy (cART) for the treatment of HIV-1 infections. These drugs are clinically effective against HIV-1 strains that had previously exhibited resistance-associated mutations (RAMs) against other cART drugs.[1] IN consists of three functional domains: The N-terminal domain (amino acids 1–50), which coordinates zinc binding, the catalytic core domain (amino acids 51–212), and the C-terminal domain (amino acids 213–288), which facilitates the host binding and IN oligomerization.[2] Most studies that focus on HIV-1 IN resistance, protein structure, and drug binding involve HIV-1 subtype B (HIV-1B), which represents less than 10% of the total worldwide pandemic. In contrast, HIV-1 non-B-subtypes dominate the pandemic, with the majority of subtypes circulating in sub-Saharan Africa. Sub-Saharan Africa is home to more than 90% of people living with HIV-1 (PLHIV-1).[3]

Cameroon and the surrounding areas are seen as the "birthplace" of HIV, and CRF02-AG is the most common and highly prevalent variant in the region. HIV-1 CRF02_AG is also becoming frequent in developed countries, such as France and Brazil.[4] Studies have shown that INSTIs are clinically effective against HIV-1B strains.

[1]Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
[2]Division of Oral Surgery, Department of Pathology Laboratory, University of Nairobi, Nairobi, Kenya.
[3]Division of Virology, Faculty of Medicine, Institute for Virology and Immunobiology, University of Wuerzburg, Wuerzburg, Germany.
[4]Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, University of Stockholm, Stockholm, Sweden.
[5]Faculty of Medicine and Biomedical Sciences, Centre for the Study and Control of Communicable Diseases, University of Yaoundé I, Yaoundé, Cameroon.
[6]South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa.

FIG. 1. HIV-1 Integrase phylogenetic analysis inferred by ML. The ML phylogenetic tree inferred in MEGA contains 37 patient-derived sequences. HIV-1 reference sequences were acquired from the HIV-1 LANL database, using the 2010 data set. An ML tree was constructed using Mega version 7.0, with the Kimura 2 parameter. Bootstrapping was performed with 1,000 replicates, with significant values indicated on tree. Majority of the sequence's clusters with HIV-1 subtype CRF02_AG (50%), (31.2%) as HIV-1 subtype CPX, two as subtype G (6.2%), two as subtype F2 (6.2%), and two as HIV-1 subtype C strains (6.2%). LANL, Los Alamos National Library; MEGA, Molecular Evolutionary Genetics Analysis; ML, maximum likelihood.

2

There is a lack of data concerning the efficacy of these inhibitors on HIV-1 non-B-subtype strains. Naturally occurring polymorphisms (NOPs) can have an impact on INSTI susceptibility/resistance in non-B-subtypes.[5] We evaluated the genetic diversity and NOPs of HIV-1 IN sequences from INSTI-naive patients from Yaoundé, Cameroon. This is the first study to investigate the prevalence of dolutegravir (DTG) RAMs in PLHIV-1 from Yaoundé, Cameroon.

We obtained blood plasma samples ($n = 56$) from INSTI treatment-naive PLHIV-1. Our cohort consisted of patient samples from multiple ethnic groups, different age groups, as well as different sexual orientations. Patients were sampled between January 2016 and December 2017, and were stored until 2019. Ethical permission for this study was obtained from the Health Research Ethics Committee of Stellenbosch University (N14/10/130 and N15/08/071). The laboratory experiments were 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.

Viral RNA was extracted from the 56 patient plasma samples, using the QIAamp Viral RNA Mini Extraction Kit (Qiagen, Germany), according to the manufacturer's instructions and stored at −80°C until used. The synthesis of complementary DNA and first-round polymerase chain reaction (PCR) amplification was performed using the Invitrogen SuperScript® III Reverse Transcriptase reagent (Invitrogen, Germany), as per the manufacturer's instructions. In-house amplification of the IN region (867 bp, positions 4230–5096, HXB2 strain) by nested reverse transcription-polymerase chain reaction (RT-PCR) was carried out as previously described by our laboratory.[6,7] Purified 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), according to the manufacturer's instructions. Primers spanning the full-length IN (867 bp) were used to sequence the PCR products in both directions. These include sequencing primers consisting of the above primers, Poli6 and Poli7, and additional sequencing primers were used, namely, Poli2 (TAAARACA RYAGTACWAATGGCA), relative to positions 4745–4766, and KLVO83 (GAATACTGCCATTTGTACTGCTG), corresponding to positions 4750–4772.

In IN sequence, data were assembled and manually edited using sequencer version 6.0 (Gene Codes, Ann Arbor). Viral sequences were also subtyped using the Rega algorithm (http://hivdb.stanford.edu).

Multiple sequence alignment of the full IN region, with reference sequences from HIV-1 subtypes (A–K) and recombinants viruses prevalent in Cameroon, was obtained from the Los Alamos National Library database (www.lanl.gov) and performed by MAFFT version7.0. Aligned sequences were manually edited using the BioEdit software package[8] and subsequently used in phylogenetic analysis. A maximum likelihood (ML) tree was constructed using Mega version 7.0, with the Kimura 2 parameter. Bootstrapping was performed with 1,000 replicates, with significant values of >70% indicated on the tree. One reference sequence belonging to the phylogenetic group O was used as an outgroup. Nucleotide sequences were converted to amino acid and compared with the 2004 consensus C sequences available from the HIV Los Alamos National Library database (hivlanl).

The amino acid sequence alignment was extensively screened for the presence of primary and secondary RAMs and NOPs associated with resistance to known INSTIs. Furthermore, the Fisher exact test was used to compare amino acid frequencies across all IN codons, in INSTI-naive versus raltegravir (RAL)-experienced sequences available from the Stanford HIV drug resistance database (http://hivdb.stanford) and to check NOPs that are significantly enriched for INSTI RAMs in subtype CRF02_AG HIV-1.

Multiple sequences that were ≤500 bp or of low quality were excluded from further subtyping analysis. We used 37 sequences for phylogenetic inference and identified 54.1% (20/37) of the samples as HIV-1 subtype CRF02_AG, followed by 18.1% (9/37) as subtype A, 8.1% (3/37) as subtype CRF36_cpx, 8.1% (3/37) as subtype F, 5.4% (2/37) as subtype G, and 5.4% (2/37) as HIV-1 subtype C strains (Fig. 1). The nucleotide sequence of the IN gene was available for 37 plasma samples. These were subsequently screened for the presence of RAMs and identified 18.9% (7/37) sequences to contain RAMs, with 5.4% (2/37) having major INSTI RAMs: P145S (Baby44IN) and Y143R/C/D/G) were present in one sequence each 2.7% (1/37). Accessory RAMs were present in 3/37 samples (8.1%), of which one sequence contained solely E157Q (Baby33IN), another Q95K (Ad51), and a third had three accessory mutations, namely G140E, E157Q, and G163R (Ad55IN).

Distribution of detected mutations, including primary and secondary drug resistance mutations (DRMs), is shown in Table 1. The distribution of amino acid variation among CRF02_AG was identified using INSTI-exposed ($n = 51$)

TABLE 1. ANTIRETROVIRAL DRUG SUSCEPTIBILITY IN INTEGRASE STRAND TRANSFER INHIBITORS AMONG INTEGRASE STRAND TRANSFER INHIBITOR-NAIVE HIV-1-INFECTED STUDY PARTICIPANTS

	Mutations		*Drugs*				
Specimen	*Major*	*Accessory*	*BIC*	*DTG*	*RAL*	*EVG*	*Drugs' remaining effect*
BABY33IN	—	E157Q	S	S	PL-LR	PL-LR	BIC and DTG
BABY44IN	P145PS	—	S	S	S	H-LR	BIC,DTG and EVG
AD51IN	—	Q95K	S	S	PL-LR	PL-LR	BIC and DTG
AD55IN	—	G140E, E157Q, G163R	S	S	L-LR	L-LR	BIC and DTG
CM 47	Y143YCD	—	S	S	H-LR	PL-LR	BIC and DTG
CM 52	—	T97A	S	S	PL-LR	PL-LR	BIC and DTG

BIC, bictegravir; DTG, dolutegravir; EVG, elvitegravir; PL-LR, potential low level resistance; RAL, raltegravir; S, susceptible.

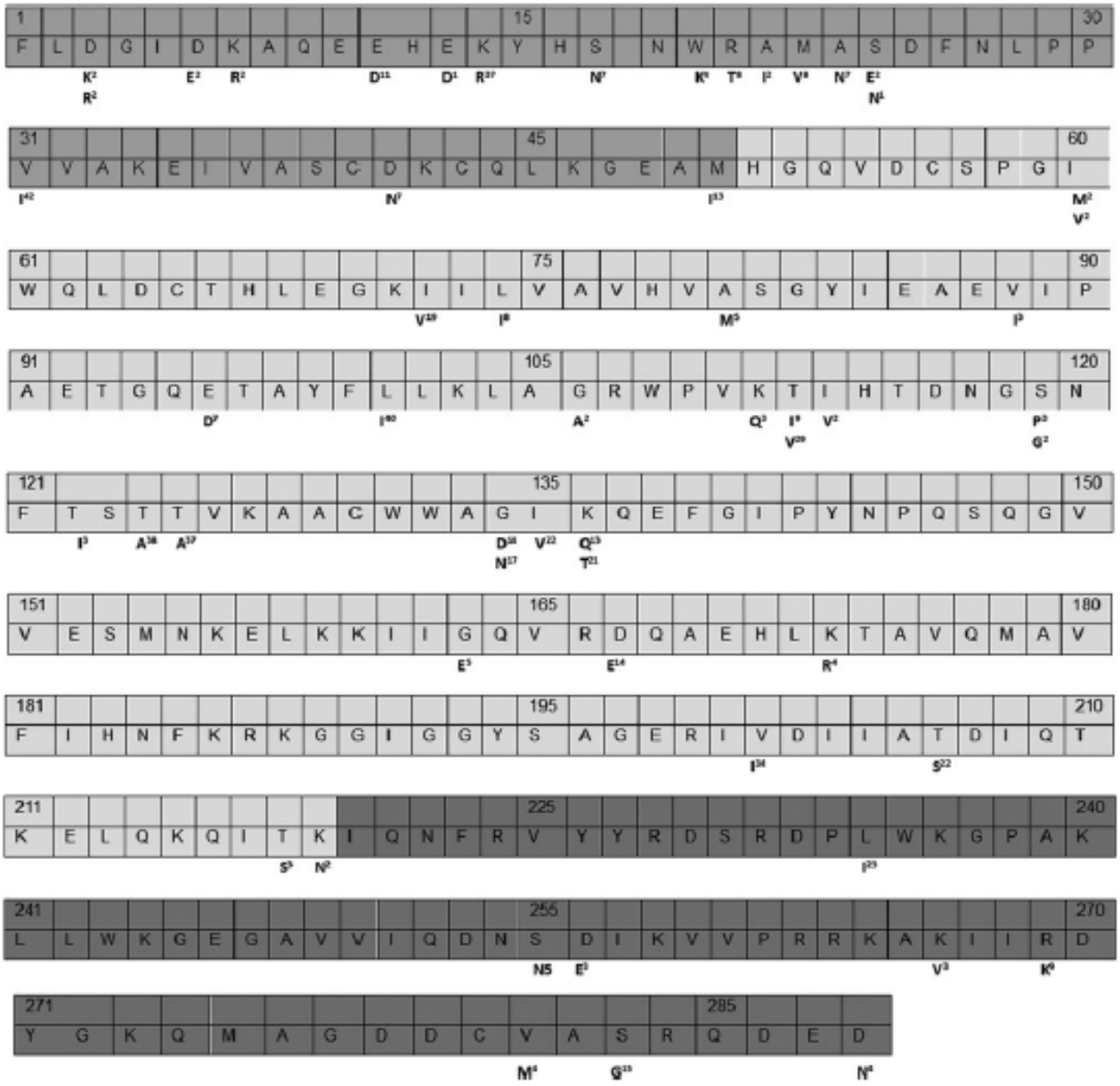

FIG. 2. Distribution of variants among CRF02_AG HIV-1 IN sequences among the newly characterized full-length IN sequences and sequences derived online. The consensus subtype CRF02_AG sequence is shown at the top of each 30 amino acid sequence section. Beneath are the variants with the superscript numbers indicating the number of times that they occurred. Domains and important residues are highlighted. The N-terminal domain is highlighted in *light gray*, the catalytic core domain in *white*, and the C-terminal domain in *dark gray*. Relevant IN inhibitor resistance mutations are in *green*. IN, integrase.

patient sequences downloaded from a specific database (https://hivdb.stanford.edu) and INSTI-unexposed patient study sequences ($n=56$), shown in (Fig. 2). Of the 288 IN positions, 53 (18.4%) had at least one amino acid NOP present in 0.5% or more sequences, including 49 (17.1%), of which two or more polymorphisms were present. All sequences that occur at a frequency of ≥2% were used to study the association of polymorphisms with INSTI resistance. HIV-1 subtype CRF02_AG patients, INSTI-exposed ($n=51$) versus INSTI-unexposed study sequence patients ($n=56$), were compared using two by two table (Fisher exact test). Through our test, the two NOPs occurred, namely: VI72I and R269K* with p-values ≤.05 (Fig. 3). This indicates strong evidence that the observed polymorphisms are significantly and statistically enriched.

The majority (54.1%) of our sequences were assigned to CRF02_AG, and the remaining 45.9% were infected with other subtypes (A, CRF36_cpx, F, G, and C). The parental strains CRF02_AG is known to circulate in Central and Western Africa, particularly in Cameroon.[9] Major RAMs to INSTIs were detected in 12.5% of the study patient sequences. These mutations were found in positions 143 and 145. Mutations at position 143 are associated with high-level RAL resistance. These mutations alone have minimal effects

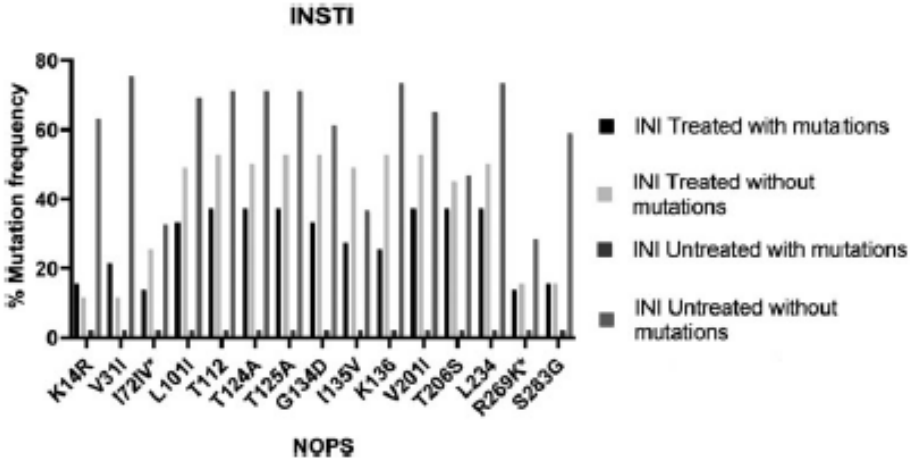

FIG. 3. Data derived from Fisher exact test indicating all naturally occurring polymorphisms of the IN gene that are significantly enriched. *Asterisk* (*) indicates the *p*-values of all naturally occurring polymorphisms that are significantly enriched: VI72I*: $p \leq .002$ and R269K $p \leq .000408249$. IN, integrase; INSTIs, IN strand transfer inhibitors.

on elvitegravir (EVG) susceptibility, but they are associated with intermediate reductions in EVG susceptibility when they occur in combination with T97A mutations.[10] Y143R/C/D/G mutations do not reduce DTG or bictegravir (BIC) susceptibility. Mutations at position 145 are associated with high-level resistance to EVG (Table 1).

Other studies have reported that major INSTI-associated mutations were very uncommon in INSTI-naive patients,[11,12] as opposed to our findings that reported the presence of RAMs against INSTIs from naive treated patients. Accessory RAMs have been found at positions 95,140,157, and 163. The presence of the E157Q mutation in our study cohort, two patients (Baby 33 IN and add 55IN) who were infected with subtype CRF02_AG, is in agreement with a previous study that found a CRF02_AG INSTI-naive infected patient from Cameroon who harbored the same mutation.[13] E157Q is a polymorphic accessory mutation selected *in vitro* by EVG and in RAL treatment-experienced patients. On its own, E157Q has no significant effect on INSTI susceptibility, but reduces the susceptibility of RAL and EVG in the presence of an N155H major RAM. We identified a viral sequence with T97A mutation; this mutation is mainly selected among treatment-experienced patients on RAL and EVG.[13] Similarly, T97A has a minimal effect on INSTI susceptibility on its own, but leads to a significant reduction in RAL and EVG susceptibility in combination with Y143C/D and N155H major RAMs. The Q95K mutation is a nonpolymorphic accessory mutation selected in patients on RAL and EVG. On its own, it does not affect INSTI susceptibility, but enhances the resistance of N155H mutants to RAL and EVG (https://hivdb.stanford.edu). We further identified polymorphic G140E and G163R RAMs.

All of these mutations were 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 is known to reduce DTG susceptibility, neither *in vitro* nor *in vivo*.[14] Three of the 37 patients (8.1%) had the M50I polymorphic mutation that is selected *in vitro* by DTG and BIC in combination with R263K and it appears to reduce DTG susceptibility in combination with R263K. Moreover, 9/37 samples (24.3%) harbored L74M/I polymorphic accessory mutations commonly selected by each of the INSTIs. In antiretroviral-naive patients, L74M occurs in 0.5% to 10% of patients and L74I occurs in 3% to 20% of patients depending on the subtype. Alone, L74M/I have minimal, if any, effect on INSTI susceptibility. However, they contribute to reduced susceptibility to each of the INSTIs when they occur with major INSTI-resistance mutations. Polymorphisms at positions that alter the binding of the drug to the IN may influence the genetic barrier to the development of INSTI resistance.[15] High frequency of amino acid change, occurring at positions 72 and 269, from our study viral sequences infected with CRF02_AG was also similar to a Mozambique study that observed a high prevalence of amino acid change in patients infected with subtype C.[16] Therefore, these polymorphisms (VI72IV and R269K) in our study indicated that they are significantly enriched in RAL-treated patients infected with the HIV-1 CRF02_AG subtype compared with INSTI-naive patients and may thus influence the evolution of DTG resistance (Fig. 3). Finally, our results indicate that the presence of primary resistance mutations in Cameroonian IN sequences, along with high amino acid variation across diverse HIV-1 subtypes, is of great concern as DTG is anticipated as a generic drug in our region.

Conclusion

None of the reported mutations (P145S and Y143R/C/D/G) in our study is associated with resistance to DTG, as it is contemplated to be used in our region. We suggest that IN sequencing of all HIV-1-infected patients before INSTI commencement to be performed to help in tailoring effective treatment strategies in the population infected with diverse HIV-1 subtypes.

Sequence Data

The newly reported integrase sequences are available in GenBank under the following accession numbers: MN816445-MN816488.

Author Contributions

S.G.M., G.M.I., A.E.O., and J.O.G. collected the samples. S.G.M., D.B., J.O.G., and C.V. performed laboratory experiments. S.G.M. performed detailed sequence analyses and subsequent analysis. S.G.M. wrote the first draft of the

article. G.B.J. conceptualized the study and helped with the article draft. All authors edited the article draft and read and approved the final article.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

The National Research Foundation (NRF) of South Africa, the Poliomyelitis Research Foundation (PRF) of South Africa, the Harry Crossley Foundation, and the National Health Laboratory Services (NHLS) Research Trust funded this study.

References

1. Messiaen P, Wensing AMJ, Fun A, Nijhuis M, Brusselaers N, Vandekerckhove L: Clinical use of HIV integrase inhibitors: A systematic review and meta-analysis. PLoS One 2013;8:e52562.
2. Van Maele B, Busschots K, Vandekerckhove L, Christ F, Debyser Z: Cellular co-factors of HIV-1 integration. Trends Biochem Sci 2006;31:98–105.
3. Martine P, Matthieu J, Ahidjo A: The origin and molecular epidemiology of HIV. Expert Rev Anti Infect Ther 2013; 11:885–896.
4. Derache A, Traore O, Koita V, *et al.*: Genetic diversity and drug resistance mutations in human immunodeficiency virus type 1 from untreated patients in Bamako, Mali. Antivir Ther 2007;12:123–129.
5. Fransen S, Gupta S, Danovich R, *et al.*: Loss of raltegravir susceptibility by human immunodeficiency virus type 1 is conferred via multiple nonoverlapping genetic pathways. J Virol 2009;83:11440–11446.
6. Brado D, Obasa A, Ikomey G, *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.
7. Jacobs G, Laten A, van Rensburg E, *et al.*: Phylogenetic diversity and low-level antiretroviral resistance mutations in HIV Type 1 treatment-naive patient from Cape Town, South Africa. AIDS Res Hum Retroviruses 2008;24:1009–1012.
8. Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 1999;41:95–98.
9. Tongo M, Martin DP, Zembe L, Eitel MN, Carolyn W, Wendy AB: Characterization of HIV-1 gag and nef in Cameroon: Further evidence of extreme diversity at the origin of the HIV-1 group M epidemic. Virol J 2013;10:29.
10. Reigadas S, Masquelier B, Calmels C, *et al.*: Structure-analysis of the HIV-1 integrase Y143C/R raltegravir resistance mutation in association with the secondary mutation T97A. Antimicrob Agents Chemother 2011;55:3187–3194.
11. Nguyen HL, Ruxrungtham K, Delaugerre C: Genetic barrier to the development of resistance to integrase inhibitors in HIV-1 subtypes CRF01_AE and B. Intervirology 2012; 55:287–295.
12. Piralla A, Paolucci S, Gulminetti R, Comolli G, Baldanti F: HIV integrase variability and genetic barrier in antiretroviral naive and experienced patients. Virol J 2011;8:149.
13. Rhee S-Y, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW: Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucl Acid Res 2003;31:298–303.
14. Abram ME, Ram RR, Margot NA, *et al.*: Lack of impact of pre-existing T97A HIV-1 integrase mutation on integrase strand transfer inhibitor resistance and treatment outcome. PLoS One 2017;12:e0172206.
15. Low A, Prada N, Topper M, Vaida F, Castor D, Mohri H: Natural polymorphisms of human immunodeficiency virus type 1 integrase and inherent susceptibilities to a panel of integrase inhibitors. Antimicrob Agents Chemother 2009;53: 4275–4282.
16. Michelli O, Montagna C, Falasca F, *et al.*: Short communication: analysis of the integrase gene from HIV type 1-positive patients living in a rural area of West Cameroon. AIDS Res Hum Retroviruses 2012;28:1729–1733.

Address correspondence to:
Sello Given Mikasi
Division of Medical Virology
Department of Pathology
Faculty of Medicine and Health Sciences
Stellenbosch University
Tygerberg
Cape Town 7505
South Africa

E-mail: mikasi@sun.ac.za

Graeme Brendon Jacobs
Division of Medical Virology
Department of Pathology
Faculty of Medicine and Health Sciences
Stellenbosch University
Tygerberg
Cape Town 7505
South Africa

E-mail: graeme@sun.ac.za

Chapter 4: HIV-1 drug resistance analyses of Cameroon derived Integrase sequences

4.1. Journal article

The article is published in AIDS Res Hum Retroviruses.2020 July. https://doi.org/10.1089/AID.2020.0022.

4.2. Author's list

Mikasi SG, Isaacs D, Ikomey GM, Shimba H, Cloete R and Jacobs GB

4.3. Author's contribution

In the enclosed manuscript, I confirmed that I am the first author. I collected all the sequences and performed data searches and performed data analyses. I wrote the first manuscript draft and was responsible for subsequent manuscript updates received from the co-authors.

4.4. Background

cART is widely used to fight against HIV infections. However, it has led to emerging HIV strains known to contain RAMs. This could consequently lead to treatment failure or unsatisfactory therapy outcomes. INSTIs are the most recent drug class being administered due to their higher genetic barrier to develop resistance.

4.5. Main findings

From our online derived sequences, we observed (1.4%) INSTIs mutations that confer resistance to raltegravir (RAL) and elvitegravir (EVG), as well as 10.1% of the sequences having accessory mutations.

4.6. Study significance

The study used sequences that were available as early as 2004, before the rollout of the HIV-1 treatment program and the introduction of INSTIs in Cameroon. Data generated from this study can be considered as a baseline INSTIs resistance rate. This study shows the need and importance for continuous surveillance of drug resistance mutations (DRM). This information could possibly be used to improve the effectiveness of future drugs.

4.7. Conclusion.

HIV-1 CRF02_AG was the predominant subtype (44.7.%) in this study analyses. The occurrence of INSTI RAMs among the sequences at baseline needs to be monitored carefully.

4.8. Published article

AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 00, Number 00, 2020

DOI: 10.1089/aid.2020.0022

HIV-1 Drug Resistance Mutation Analyses of Cameroon-Derived Integrase Sequences

Sello Given Mikasi,[1,*] Darren Isaacs,[2,*] George Mondinde Ikomey,[3] Henerico Shimba,[1,4] Ruben Cloete,[3] and Graeme Brendon Jacobs[1]

Abstract

HIV-1 integrase (IN) is a primary target for combination antiretroviral therapy. Only a limited number of studies report on the emergence of resistance-associated mutations (RAMs) in Cameroon. We observed that 1.4% of sequence from treatment-naive patients had IN strand transfer inhibitor (INSTI) RAMs. These mutations confer resistance to raltegravir and elvitegravir. We also observed that 10.1% of the sequences have INSTI accessory RAMs. HIV-1 CRF02_AG was the predominant subtype (44.7%) in this study analyses. The occurrence of INSTI RAMs among the sequences at baseline needs to be monitored carefully.

Keywords: HIV-1, integrase, resistance, Cameroon, diversity, CRF02_AG

HIV/AIDS remains a health concern that endangers the lives of millions of people worldwide. HIV-1 has three essential enzymes used for its replication, namely, protease, reverse transcriptase, and integrase (IN). HIV IN is a 32 kDa protein encoded at the 3′-end of the HIV pol gene.[1] IN consists of 288 amino acids, and can be divided into three structural and functional domains: a zinc-binding N-terminal domain, a catalytic core domain, and a DNA binding C-terminal domain.[2]

IN strand transfer inhibitors (INSTIs) are a class of antiretroviral drugs used to target the IN enzyme, to prevent viral complementary DNA integration into the host genome. IN resistance-associated mutations (RAMs) located at amino acid residues Y143C, Q148H, and N155H have shown to be the primary/major pathways that induce resistance to the INSTI raltegravir (RAL). Primary mutations that induce resistance against elvitegravir (EVG) include T66I, E92Q, Q148H/R/K, and N155H. RAL and EVG present broad cross-resistance.[3] Dolutegravir (DTG) retains its activity against viruses that exhibited RAMs against RAL and EVG.[4] However, combinations of mutations at position G140 and Q148 can result in significant drug resistance to DTG. Furthermore, DTG monotherapy in treated naive patients has been reported to be associated with more frequent development of RAMs such as R263K, G118R, and S230.[5]

Naturally occurring polymorphisms (NOPs) have been identified at sites associated with secondary RAMs: L74V, M154I, I72V, and T125A.[6] Studies have shown that difference across HIV-1 subtypes may play a role, by influencing NOPs in facilitating or compensating the development of major resistance pathway to antiretroviral drugs.[7] Owing to the development of drug resistance across currently available drugs, the World Health Organization (WHO) has put forth the use of the INSTI DTG as part of suggested first-line therapy. Therapeutic advantage of DTG is its ability to maintain high potencies against mutant strains of HIV-1 that previously exhibited resistance to other combination antiretroviral therapy (cART) drugs.[7] In Cameroon, there is a lack of data regarding the emergence of RAMs against INSTIs; particularity DTG as it is anticipated to be used in future treatment regimens.[8] Therefore, it is of paramount importance to identify RAMs from samples sequenced before the initiation of INSTI therapy in Cameroon.

In this study, we evaluated only IN sequences from INSTIs-naive HIV-infected individuals from Cameroon, dating between 1994 and 2009, obtained from the Los Alamos National Library (LANL) HIV-1 database (https://www.hiv.lanl.gov/components/sequence/HIV search.com), accessed between June 10, 2019, and June 12, 2019. We

[1]Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
[2]South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa.
[3]Centre for the Study and Control of Communicable Diseases, Faculty of Medicine and Biomedical Sciences, University of Yaoundé I, Yaounde, Cameroon.
[4]Bugando Medical Centre, Mwanza, United Republic of Tanzania.
*These authors contributed equally to this work.

screened for the presence of primary mutations, secondary mutations, and NOPs.

Ethical permission for this study was obtained from the Health Research Ethics Committee of Stellenbosch University (N15/08/071). Our search inclusion criteria included all Cameroonian IN sequences identified from treatment-naive patients. Online-derived sequences used were sampled between 1994 and 2009, before the start of INSTIs in Cameroon. We selected one sequence per patient and 655 HIV-1 sequences were included in the analyses.

For HIV-1 subtyping, the IN sequences were reanalyzed between June and July 2019, using the REGA version 3 online subtyping tool (www.bioafrica.net/rega-genotype/html/subtypinghiv.html), Stanford University HIV Drug Resistance Database (http://hivdb.stanford), and COMET (https://comet.lih.lu/).[9]

IN nucleotide, sequences, and predicted amino acid sequences were compared with the HXB2 HIV-1 subtype B consensus sequence (GenBank accession no. K03455). Mutations associated with resistance to INSTIs were identified using the Stanford University genotypic resistance interpretation algorithm, HIVdb version 8.3 (http://hivdb.stanford.edu/). Drug resistance mutation results were classified as either major or minor mutations, last accessed July 18, 2019.

Among the 577/655 (88.2%) group M viruses, sequences were classified as belonging to subtypes CRF02_AG (44.7%) $n=293/655$, A (12%) $n=79/655$, CPX (18.0%) $n=118/655$, CRF01_AE (0.6%) $n=4/655$, D (1.9%) $n=13/655$, F (5.3%) $n=35/655$, and G (5.3%) $n=35/655$. Sequences from (1.8%) $n=12/655$ individuals were classified as belonging to group N, group O (8.7%) $n=57/655$, and group U (1.4%) $n=9/655$.

In total, 1.4% (9/655) of the database-derived sequences had HIV-1 RAMS: this includes T66A detected in 4/655 (0.6%), R263K detected in 2/655 (0.3%), N155K/T detected in 2/655 (0.3%), Q148V detected in 1/655 (0.2%) and E92S detected in 2/655 (0.3%). Of note, 2/655 (0.3%) of sequences that contained E92S occurred in combination with another mutation, in this case, one sequence had E92S in combination with Q148V and the other sequence had E92S in combination with T66A.

In addition, 66/655 (10.1%) of the sequences contained seven different accessory RAMs, namely, A128T detected in 2/655 (3.0%), D232N detected in 3/655 (4.5%), E157Q detected in 19 (28.7%), G140A detected in 2/655 (3%), Q146R detected in 1/655 (1.5%), Q95K detected in 2/655 (3.0%), and T97A detected in 37/655 (5.6%).

Our findings are in agreement with other reported studies, where CRF02_AG is seen as the predominant subtype causing HIV-1 infection in Cameroon.[8,10] We further interrogated the IN region for the presence of INSTI RAMs and we identified major and minor RAMs in the sequences investigated. In total, 9/655 (1.4%) of the sequences exhibited major RAMs. Of note, these mutations reduce susceptibility to EVG and RAL. The R263K mutation was detected in our database-derived sequences. It occurs in a high proportion of patients developing virological failure (VF) on an incompletely suppressive DTG-containing regimen and rarely in patients receiving RAL. It reduces DTG and bictegravir susceptibility approximately twofold and reduces EVG susceptibility somewhat more.[11] In addition, Q148H may profoundly affect second-generation INSTI susceptibility. If co-occurring with additional RAMs, mutations in the Q148 pathway can lead to higher fold resistances against all INSTI. The results reported are in line with the previous findings by our laboratory, which reported the circulation of RAMs from cART-naive patients in Cameroon.[8]

We further detected large number of patients' cohort sequences harboring accessory mutations, with T97A (56.1%) being the most frequent accessory mutations detected, followed by E157Q detected in 28.7%. T97A substitution is a clinical worrisome mutation, since it reduces susceptibility to RAL and DTG, when combined with other major RAMs. E157Q is an accessory mutation usually selected in patients receiving RAL and *in vitro* by EVG. It has also been reported in a patient with VF on a DTG-containing regimen. It appears to reduce RAL and EVG susceptibility by two- to threefold and to increase DTG susceptibility.[12] NOPs that are mostly not detected at baseline testing are likely to compensate or delay the development of resistance, according to other reported findings.[13]

From this study, we observed the occurrence of INSTI RAMs from cART-naive patient's sequences. The occurrence of INSTI RAMs and NOPs among Cameroon-derived IN sequences warrants baseline resistance testing, before cART commencement, as these patients may be at higher risk of failing their INSTI-based therapy.

Authors' Contributions

G.B.J., G.M.I., and R.C. conceptualized the study. S.G.M. and D.I. wrote the first draft. S.G.M., D.I., H.S., and G.B.J. performed data analysis. All authors performed quality control check, edited, read, and approved the final article.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

The National Research Foundation (NRF) of South Africa and the Poliomyelitis Research Foundation (PRF) of South Africa funded this study. Grant support for this study was received from the National Research Foundation (NRF) of South Africa, the Poliomyelitis Research Foundation (PRF), Harry Crossley Foundation, South African Research Chairs Initiative of the Department of Science, Technology (DST) and South African Medical Research Council (SAMRC), Higher Education Department, next Generation of Programme (nGAP) for full time academic positions of Dr. Cloete and Dr. Jacobs.

References

1. Asante-Appiah E, Skalka AM: HIV-1 integrase: Structural organization, conformational changes, and catalysis. Adv Virus Res 1999;52:351–369.
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; 93:13659–13664.
3. Geretti AM, Armenia D, Ceccherini-Silberstein F: Emerging patterns and implications of HIV-1 integrase inhibitor resistance. Curr Opin Infect Dis 2012;25:677–686.

4. Blanco JL, Marcelin AG, Katlama C, Martinez E: Dolutegravir resistance mutations: Lessons from monotherapy studies. Curr Opin Infect Dis 2018;31:237–245.
5. Castagno A, Maggiolo F, Penco G, *et al.*: VIKING-3 Study Group Dolutegravir in antiretroviral-experienced patients with raltegravir- and/or elvitegravir-resistant HIV-1: 24-week results of the phase III VIKING-3 study. J Infect Dis 2014;210:354–362.
6. Loizidou EZ, Kousiappa I, Zeinalipour-Yazdi CD, Van de Vijver DA, Kostrikis LG: Implications of the HIV-1 M group polymorphisms on integrase inhibitor efficacy and resistance: Genetic and structural in silico analyses. Biochemistry 2009;13:48:4–6.
7. Quashie PK, Oliviera, M, Veres T, *et al.*: Substitutions in different subtypes of HIV integrase. J Virol 2015;89:3163–3175.
8. Mikasi SG, Gichana JO, Van der Walt C, *et al.*: HIV-1 integrase diversity and resistance-associated mutations and polymorphisms among integrase strand transfer inhibitor-naive HIV-1 patients from Cameroon. AIDS Res Hum Retroviruses 2020;36:450–455.
9. Pineda-Peña AC, Faria NR, Imbrechts S, *et al.*: 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 2013;19:337–348.
10. Tongo M, Martin DP, Zembe L, Eitel MN, Carolyn W, Wendy AB: Characterization of HIV-1 gag and nef in Cameroon: Further evidence of extreme diversity at the origin of the HIV-1 group M epidemic. Virol J 2013;10:29.
11. Cahn P, Pozniak AL, Mingrone H, *et al.*: Dolutegravir versus raltegravir in antiretroviralexperienced, integrase-inhibitor-naive adults with HIV: Week 48 results from the randomised, double-blind, non-inferiority SAILING study. Lancet 2013;382:700–708.
12. Anstett K, Cutillas V, Fusco R, Mesplede T, Wainberg MA: Polymorphic substitution E157Q in HIV-1 integrase increases R263Kmediated dolutegravir resistance and decreases DNA binding activity. J Antimicrob Chemother 2016;71:2083–2088.
13. Turriziani O, Montagna C, Falasca F, *et al.*: Short communication: Analysis of the integrase gene from HIV type 1- positive patients living in a rural area of West Cameroon. AIDS Res Hum Retroviruses 2012;28:1729–1733.

Address correspondence to:
Sello Given Mikasi
Division of Medical Virology
Department of Pathology
Faculty of Medicine and Health Sciences
Stellenbosch University
Tygerberg
Cape Town 7505
South Africa

E-mail: mikasi@sun.ac.za

Graeme Brendon Jacobs
Division of Medical Virology
Department of Pathology
Faculty of Medicine and Health Sciences
Stellenbosch University
Tygerberg
Cape Town 7505
South Africa

E-mail: graeme@sun.ac.za

Chapter 5: Drug resistance mutations against protease, reverse transcriptase and integrase inhibitors in people living with HIV-1 receiving boosted protease inhibitors (bPIs) in South Africa.

5.1. **Journal article**

The article is published in the journal: Frontiers in Microbiology. ORIGINAL RESEARCH published: 20 March 2020 doi: 10.3389/fmicb.2020.00438.

5.2. **Author's list**

Obasa AE, **Mikasi SG**, Brado D, Cloete R, Singh K, Neogi U, Jacobs GB.

5.3. **Authors contribution**

In the enclosed manuscript, I confirmed that I am the second, co- author. I helped collect samples for the study, collected demographic data of the patients, assisted with PCR, sequence and data analysis. I subsequently gave significant inputs in the manuscript.

5.4. **Background**

South Africa is home to more than 7.7 million people living with HIV-1 and has the largest cART rollout program in the world. A major drawback of the success of cART is the emergence of RAMS that can compromise treatment outcomes. In this study, we analyzed the RAMS against current available drugs (PIs, RTs and INSTIs) from patients suspected of failing first-line and or second-line cART in South Africa.

5.5. **Main findings**

Of the 96 samples from individuals suspected of failing first-line and or second-line cART (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, while E157Q mutation was identified in (n= 2, 2%) of patients as the most common INSTIs RAMs, alongside mutations;T66I, Y143R and T97A, which are major INSTIs RAMs.

5.6. **Study significance**

Amino acid sequence analysis revealed that the most frequent detected mutation was M184V. The mutation is associated with reduced susceptibility to the NRTIs. This mutation could have resulted from first-line therapy. Our findings are relevant given that DTG is recommended as a rescue intervention, as most patient who are on first-line therapy are already developing mutations that might compromise treatment success.

5.7. **Conclusion**

The study of 96 sequences covering three fragments; PI, RT and IN, from patients suspected to be failing second-line therapy disclosed a high level of polymorphisms against NRTIs and PIs of HIV-1-infected subjects from South Africa. No major mutations were detected against INSTIs. This study provide important information about the resistance profile in South Africa and supports the use of DTG in South Africa.

5.8. Published article

Drug Resistance Mutations Against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living With HIV-1 Receiving Boosted Protease Inhibitors in South Africa

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

[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, Stockholm University, Stockholm, Sweden, [3] Division of Virology, Institute for Virology and Immunobiology, Faculty of Medicine, University of Würzburg, Würzburg, Germany, [4] South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Bellville, South Africa, [5] Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, MO, United States, [6] Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO, United States

OPEN ACCESS

Edited by:
Francois Villinger, University of Louisiana at Lafayette, United States

Reviewed by:
David Katzenstein, Stanford University, United States
Philippe Colson, IHU Mediterranee Infection, France

***Correspondence:**
Adetayo Emmanuel Obasa
obasa@sun.ac.za;
emmanueladety@gmail.com
Graeme Brendon Jacobs
graeme@sun.ac.za

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

Received: *04 November 2019*
Accepted: *02 March 2020*
Published: *20 March 2020*

Citation:
Obasa AE, Mikasi SG, Brado D, Cloete R, Singh K, Neogi U and Jacobs GB (2020) Drug Resistance Mutations Against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living With HIV-1 Receiving Boosted Protease Inhibitors in South Africa. Front. Microbiol. 11:438. doi: 10.3389/fmicb.2020.00438

The South African national combination antiretroviral therapy (cART) roll-out program started in 2006, with over 4.4 million people accessing treatment since it was first introduced. HIV-1 drug resistance can hamper the success of cART. This study determined the patterns of HIV-1 drug-resistance associated mutations (RAMs) in People Living with HIV-1 (PLHIV-1). Receiving first (for children below 3 years of age) and second-line (for adults) cART regimens in South Africa. During 2017 and 2018, 110 patients plasma samples were selected, 96 samples including those of 17 children and infants were successfully analyzed. All patients were receiving a boosted protease inhibitor (bPI) as part of their cART regimen. The viral sequences were analyzed for RAMs through genotypic resistance testing. We performed genotypic resistance testing (GRT) for Protease inhibitors (PIs), Reverse transcriptase inhibitors (RTIs) and Integrase strand transfer inhibitors (InSTIs). Viral sequences were subtyped using REGAv3 and COMET. Based on the PR/RT sequences, HIV-1 subtypes were classified as 95 (99%) HIV-1 subtype C (HIV-1C) while one sample as 02_AG. Integrase sequencing was successful for 89 sequences, and all the sequences were classified as HIV-1C (99%, 88/89) except one sequence classified CRF02_AG, as observed in PR/RT. Of the 96 PR/RT sequences analyzed, M184V/I (52/96; 54%) had the most frequent RAM nucleoside reverse transcriptase inhibitor (NRTI). The most frequent non-nucleoside reverse transcriptase inhibitor (NNRTI) RAM was K103N/S (40/96, 42%). Protease inhibitor (PI) RAMs M46I and V82A were present in 12 (13%) of the sequences analyzed. Among the InSTI major RAM two (2.2%) sequences have Y143R and T97A mutations

while one sample had T66I. The accessory RAM E157Q was identified in two (2.2%). The data indicates that the majority of the patients failed on bPIs didn't have any mutation; therefore adherence could be major issue in these groups of individuals. We propose continued viral load monitoring for better management of infected PLHIV.

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 have changed HIV/AIDS from a deadly pandemic to a chronic and manageable disease (Trickey et al., 2017). 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 (Hightower et al., 2011; UNAIDS, 2017). South Africa's national HIV cART program was introduced in 2006, with a public health approach (UNAIDS, 2018). Besides problems related to adherence, the development and spread of drug resistance have constantly challenged the long-term management of PLHIV in public health settings, where patients are often monitored using clinical or immunological parameters (Rousseau et al., 2006).

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 lopinavir (LPV/r) which was revised to atazanavir (ATV/r) in 2017 (Meintjes et al., 2017). 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 3 years, based on its superiority when compared with nevirapine (NVP), regardless of previous NVP exposure to prevent mother-to-child HIV transmission (Meintjes et al., 2017).

In vitro studies on PI-naïve PLHIV-1 infected with HIV-1 subtype C (HIV-1C) viruses, have indicated wide variations in their respective susceptibility to the PIs LPV/r and ATV/r (Sutherland et al., 2016). Observational studies from sub-Saharan Africa have shown a 14–32% prevalence of virological failure to second-line boosted PI- (bPI) based cART (Ajose et al., 2012; Sigaloff et al., 2012). In South Africa, reports of drug resistance patterns in patients receiving bPIs are scarce (Collier et al., 2017). 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 characterized the presence of primary integrase strand-transfer inhibitor (InSTI) DRMs in this specific population.

MATERIALS AND METHODS

Ethics Statement

The 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 Declaration of Helsinki 2013, the South African Guidelines for Good Clinical Practice and the Medical Research Council Ethical Guidelines for Research. A waiver of written informed consent was awarded to conduct sequence analyses on these samples by the Health Research Ethics Committee of Stellenbosch University, South Africa.

Viral Load

HIV-1 Viral load was performed using the Abbott m2000sp and the Abbott m2000rt analyzers (Abbott laboratories, Abbott Park, IL, United States). RNA was isolated from patient samples according to the manufacturer's instructions using the Abbott RealTime HIV-1 amplification reagent Kit.

Study Design

HIV-1-positive patient samples were obtained randomly, without any knowledge of drug-resistance patterns, 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. We excluded patient samples with no previous cART regimen history and patients receiving first-line cART treatment regimen. Demographic and clinical information such as age, cART regimen, and viral load measurement (**Table 1**). 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 16 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 bPIs cART, according to the South African national cART guidelines (Meintjes et al., 2017). These patients are eligible for InSTI treatment consideration when PI mutations are present.

TABLE 1 | Characteristics and patterns of mutations in 96 patients at the time of treatment failure.

Variable	Value
Gender	
Female	58 (60%)
Male	34 (35%)
Unknown	4 (4%)
Viral Load (Log_{10} copies/mL), mean (SD)	Log 8.4, 4.64 (3.02 – 6.74)
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/r	6 (6%)
TDF, FTC, LPV/r	7 (7%)
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 mutations	
Major IN mutation	
T66I	1 (1%)
Y143R and T97A	2 (2%)
IN Accessory mutations	
E157Q	2 (2%)

The IN mutation is based on the 89 sequences that passed the quality control.

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 (Jacobs et al., 2008). 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) (Plantier et al., 2006). 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 1300 bps, were used (Plantier et al., 2006). The integrase gene fragment amplification steps were performed as previously described by us (Brado et al., 2018). Sequencing reactions were performed as previously described by us (Brado et al., 2018). As part of quality control, each of the viral sequences was 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[1]. Subtyping was carried out using REGAv3 and COMET (Pineda-Peña et al., 2013). Phylogenetic analysis was carried out using MEGAv6.

RESULTS

We included patients receiving bPI as part of their 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. In the Integrase region seven sequences did not pass the quality control. Those sequences were excluded from the final analyses. Among the patients, 4% (4/96) of the samples did not indicate

[1]https://hivdb.stanford.edu/

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 = 2; 2%), and TDF plus AZT (n = 1; 1%). Fourteen (15%) had received LPV/r, two (2%) had received ATV/r and one (1%) had received darunavir (DRV/r) 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%), TDF plus FTC (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/r as their bPIs and three (3%) are currently receiving DRV/r.

HIV-1 Subtyping

Subtyping was carried out using REGAv3 and COMET. While REGAv3 provide subtyping for all the RT/PR sequences as well as IN sequences, COMET failed to subtype the IN sequences as the majority of the IN sequences were typed as CPZ. Therefore we used HIV-1 BLAST to identify the nearest subtype for the IN sequences. Based on the PR/RT sequences (n = 96), 99% were identified as HIV-1C while one as 02_AG (PT405ZA). Based on the IN sequences (n = 89), patient samples PT405ZA identified as 02_AG while 88 (99%) samples as HIV-1C. The subtyping data is presented in **Supplementary Table S1**. We also performed phylogenetic analysis to identify any specific clusters. The neighbor-joining phylogenetic tree did not identify any specific cluster of a transmission network (**Figure 1**).

NRTI Resistance-Associated Mutations

Table 1 shows the number of resistance-associated mutations (RAMs) observed among the 96 sequences analyzed. 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 analog 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/r. T215Y occurred in two (2%) patients receiving AZT plus 3TC.

NNRTI Resistance-Associated Mutations

Table 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. The patients might have received the EFV or NVP based regiments as their first-line treatment though the past treatment history was not available.

Y188L occurred in five (5%) patients receiving AZT plus 3TC, and in two (2%) patients receiving TDF plus ATV/r, 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/r- (n = 8; 8%) and DRV/r- (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/r (**Table 1**). 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

We successfully sequence 89 INI samples. In our cohort, we identified major InSTI RAMs in patients who had received

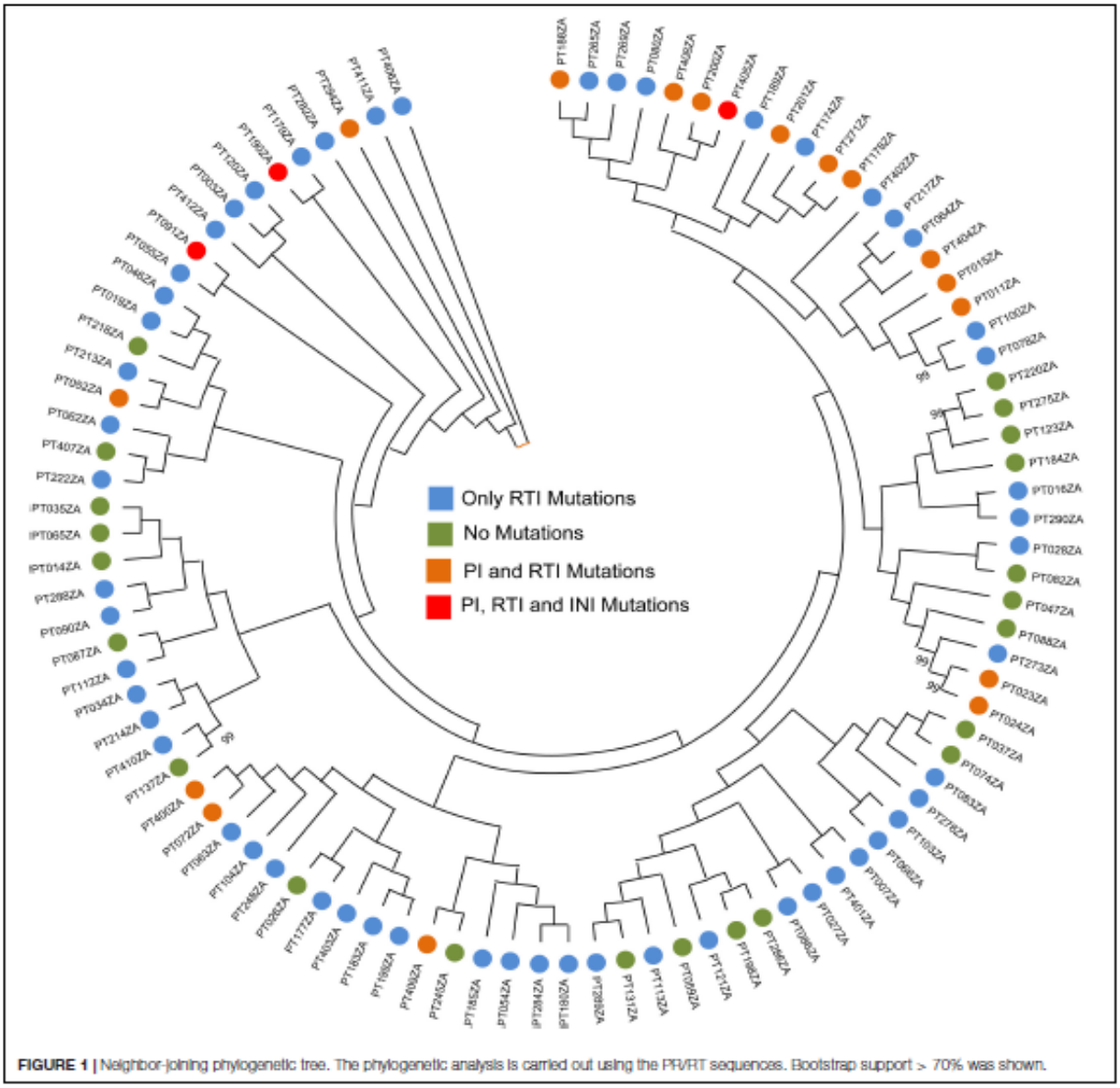

FIGURE 1 | Neighbor-joining phylogenetic tree. The phylogenetic analysis is carried out using the PR/RT sequences. Bootstrap support > 70% was shown.

the DRV/r-containing regimen. Two (2%) patient had a viral sequences with 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 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/r. The level of drug resistance against all the cARTs is presented in **Figure 2**. The observed RAMs in patient receiving bPI based treatment regimen shows high-level resistance was demonstrated in 17 (17%) and 14 (14%) of PLHIV against LPV/r and (ATV/r), respectively, while seven (7%) showed intermediate cross-resistance to DRV/r. Despite off-treatment with NNRTIs, more than half of the patients were shown to have high-level resistance to NVP (57%, $n = 56$) and EFV (56%, $n = 55$). High-level resistance were also observed with patients receiving NRTI-based treatment to FTC (60%, $n = 59$) and 3TC (60%, $n = 59$).

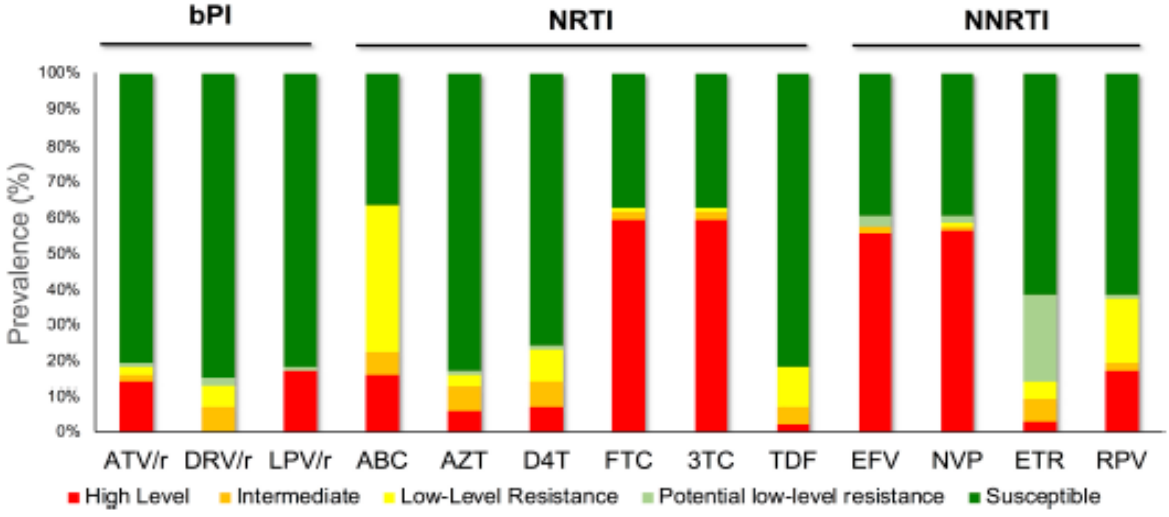

FIGURE 2 | The observed resistance in patients receiving bPIs as part of their cART. High-level resistance was shown in 17 (17%) and 14 (14%) of PLHIV against LPV/r and (ATV/r), respectively, while seven (7%) showed intermediate cross-resistance to DRV/r. Despite off-treatment with NNRTIs, more than half of the patients were shown to have high-level resistance to NVP (57%, $n = 56$) and EFV (56%, $n = 55$).

DISCUSSION

In this study, we analyzed 96 HIV-1 RT/PR and 89 IN 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 harbored 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 (Amanda et al., 2016). 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 (Marconi et al., 2008; Van Zyl et al., 2011, 2013; Wallis et al., 2011; Lombaard et al., 2016; Neogi et al., 2016; Steegen et al., 2016a,b; Rossouw et al., 2017; Penrose et al., 2018). 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 (Bronze et al., 2012; Steegen et al., 2017). 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 presence of NNRTI RAMs when they were off the NNRTI indicating either the carryover NNRTI RAMs from the failed first-line therapy or could be the transmitted DRMs in children from the mother due to vertical transmission prophylaxis.

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/r. We identified a low frequency of M46I and V82A in patients receiving ATV/r 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.'s (2016) 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 (Neogi et al., 2016). Our findings also showed M46I and V82A RAMs as the most prevalent major PI RAMs. Furthermore, these findings are in agreement with Rossouw et al. (2015) that observed both V82A and M46I has the most common mutation in infected

children receiving PI-based cART. Chimbetete et al. (2018) observed similar results, with all three drug classes showing their DRMs at similar rates. The most common PI RAM reported by Chimbetete et al. (2018) 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/r and DRV/r. Van Zyl et al. (2013) findings are consistent with ours, as they also identified more high-level resistance in patients receiving LPV/r compared with those receiving ATV/r and DRV/r.

We analyzed 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 recognized pathways of genotypic resistance against InSTIs (Doyle et al., 2015). 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. (2018), 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 (Brado et al., 2018). 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 (Neogi et al., 2018).

Furthermore, we identified the presence of T66I mutation in 1 (1%) patient. T66I confers 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 (Fourati et al., 2015; Mesplède and Wainberg, 2015). 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 as limited viral load monitoring and without drug resistance genotyping may compromise the next-generation InSTIs to be used.

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. However, 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. Finally, we did not have any adherence data for these patients and the DRM data were only based on population sequencing, therefore we could not detect minor mutations below 20% of the population.

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 program. 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. Given the negative effect of M184V mutations on viral fitness, it is more plausible to the recycling of 3TC/FTC in second-line cART maintains the presence of M184V. 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 propose for genotypic resistance testing should be performed before switching to InSTIs-based regimen in our setting. This will not only detect treatment failure earlier but will also identify poor treatment adherence. 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. Studies that address operational issues, such as the optimal use of treatment monitoring tools, should be a research priority.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The 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 Declaration of Helsinki 2013, the South African Guidelines for Good Clinical Practice and the Medical Research Council Ethical Guidelines for Research. A waiver of written informed consent was awarded to conduct sequence analyses on these samples by the Health Research Ethics Committee of Stellenbosch University, South Africa.

AUTHOR CONTRIBUTIONS

GJ and UN conceptualized and designed the study. AO performed the laboratory experiments, detailed the sequence

analyses, and wrote the first draft of the manuscript. SM helped with demographic data and sample collection. UN performed the sequencing experiments and the sequence analyses. DB, RC, and KS helped with manuscript proofreading and editing. All authors read and approved the final manuscript.

FUNDING

This study was funded by the National Research Foundation of South Africa, the Poliomyelitis Research Foundation of South Africa, and the NHLS Research Trust. We 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).

ACKNOWLEDGMENTS

We acknowledge Shambhu Prasad Ganeshappa Aralaguppe for performing the sequencing experiments. We also acknowledge Mathilda Claassen for sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00438/full#supplementary-material

TABLE S1 | The subtyping data is presented in **Supplementary Table S1**. Subtyping was carried out using REGAv3 and COMET. Based on the PR/RT sequences ($n = 96$), 99% were identified as HIV-1C while one as 02_AG (PT405ZA). Based on the IN sequences ($n = 89$), patient samples PT405ZA identified as 02_AG while 88 (99%) samples as HIV-1C.

REFERENCES

Ajose, O., Mookerjee, S., Mills, E. J., Boulle, A., and Ford, N. (2012). Treatment outcomes of patients on second-line antiretroviral therapy in resource-limited settings: a systematic review and meta-analysis. *AIDS* 26, 929–938. doi: 10.1097/QAD.0b013e328351f5b2

Amanda, H., Svedhem, V., Singh, K., Sonnerborg, A., and Neogi, U. (2016). Virological failure in patients with HIV1 subtype C receiving antiretroviral therapy: an analysis of a prospective national cohort in Sweden. *Lancet HIV* 3, 1–18. doi: 10.1186/s40945-017-0033-39

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 availability of integrase inhibitors in Cape Town, South Africa. *Sci. Rep.* 8:4709. doi: 10.1038/s41598-018-22914-22915

Bronze, M., Steegen, K., Wallis, C. L., De Wolf, H., Papathanasopoulos, M. A., Van Houtte, M., et al. (2012). ART-A Consortium. HIV-1 phenotypic reverse transcriptase inhibitor drug resistance test interpretation is not dependent on the subtype of the virus backbone. *PLoS One* 7:e34708. doi: 10.1371/journal.pone.0034708

Chimbetete, C., Katzenstein, D., Shamu, T., Spoerri, A., Estill, J., Egger, M., et al. (2018). HIV-1 drug resistance and third-line therapy outcomes in patients failing second-line therapy in Zimbabwe. *Open Forum Infect. Dis.* 5, 1–8. doi: 10.1093/ofid/ofy005

Collier, D., Iwuji, C., Derache, A., de Oliveira, T., Okesola, N., Calmy, A., et al. (2017). Virological outcomes of second-line protease inhibitor-based treatment for human immunodeficiency virus type 1 in a high-prevalence Rural South African setting: a competing-risks prospective cohort analysis. *Clin. Infect. Dis.* 64, 1006–1016. doi: 10.1093/cid/cix015

Doyle, T., Dunn, D. T., Ceccherini-Silberstein, F., De Mendoza, C., Garcia, F., Smit, E., et al. (2015). Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades. *J. Antimicrob. Chemother.* 70, 3080–3086. doi: 10.1093/jac/dkv243

Fourati, S., Charpentier, C., Amiel, C., Morand-Joubert, L., Reigadas, S., Trabaud, M. A., et al. (2015). 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. doi: 10.1093/jac/dku535

Hightower, K. E., Wang, R., Deanda, F., Johns, B. A., Weaver, K., Shen, Y., et al. (2011). 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. doi: 10.1128/AAC.00157-111

Jacobs, G. B., Laten, A., van Rensburg, E. J., Bodem, J., Weissbrich, B., Rethwilm, A., et al. (2008). Phylogenetic diversity and low level antiretroviral resistance mutations in HIV type 1 treatment-naive patients from Cape Town, South Africa. *AIDS Res. Hum. Retrovirus.* 24, 1009–1012. doi: 10.1089/aid.2008.0028

Lombaard, J., Bunupuradah, T., Flynn, P. M., Ramapuram, J., Ssali, F., Crauwels, H., et al. (2016). Rilpivirine as a treatment for HIV-infected antiretroviral-naïve adolescents: week 48 safety, efficacy, virology and pharmacokinetics. *Pediatr. Infect. Dis. J.* 35, 1215–1221. doi: 10.1097/inf.0000000000001275

Marconi, V. C., Sunpath, H., Lu, Z., Gordon, M., Koranteng-Apeagyei, K., Hampton, J., et al. (2008). South africa resistance cohort study team. Prevalence of HIV-1 drug resistance after failure of a first highly active antiretroviral therapy regimen in KwaZulu Natal, South Africa. *Clin. Infect. Dis.* 46, 1589–1597. doi: 10.1086/587109

Meintjes, G., Moorhouse, M. A., Carmona, S., Davies, N., Dlamini, S., van Vuuren, C., et al. (2017). Adult antiretroviral therapy guidelines 2017. *Southern Afr. J. HIV Med.* 18:24. doi: 10.4102/sajhivmed.v18i1.776

Mesplède, T., and Wainberg, M. A. (2015). Resistance against integrase strand transfer inhibitors and relevance to HIV persistence. *Viruses* 7, 3703–3718. doi: 10.3390/v7072790

Neogi, U., Engelbrecht, S., Claassen, M., Jacobs, G. B., van Zyl, G., Preiser, W., et al. (2016). Mutational heterogeneity in p6 gag late assembly (L) domains in HIV-1 subtype C viruses from South Africa. *AIDS Res. Hum. Retrovirus.* 32, 80–84. doi: 10.1089/aid.2015.0266

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

Penrose, K. J., Brumme, C. J., Scoulos-Hanson, M., Hamanishi, K., Gordon, K., Viana, R. V., et al. (2018). Frequent cross-resistance to rilpivirine among subtype C HIV-1 from first-line antiretroviral therapy failures in South Africa. *Antivir. Chem. Chemother.* 26:2040206618762298. doi: 10.1177/2040206618762985

Pineda-Peña, A., Faria, N. R., Imbrechts, S., Libin, P., Abecasis, A. B., Deforche, K., et al. (2013). 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. doi: 10.1016/j.meegid.2013.04.032

Plantier, J. C., Lemée, V., Nabias, R., Sire, J. M., and Simon, F. (2006). Generic screening test for HIV infection. *AIDS* 20, 1345–1347. doi: 10.1097/01.aids.0000232251.66533.ca

Rossouw, T. M., Feucht, U. D., Melikian, G., van Dyk, G., Thomas, W., du Plessis, N. M., et al. (2015). Factors associated with the development of drug resistance mutations in HIV-1 infected children failing protease inhibitor-based antiretroviral therapy in South Africa. *PLoS One* 10:e0133452. doi: 10.1371/journal.pone.0133452

Rossouw, T. M., Nieuwoudt, M., Manasa, J., Malherbe, G., Lessells, R. J., Pillay, S., et al. (2017). 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. doi: 10.1111/hiv.12400

Rousseau, C. M., Birditt, B. A., McKay, A. R., Stoddard, J. N., Lee, T. C., McLaughlin, S., et al. (2006). Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *J. Virol. Methods* 136, 118–125. doi: 10.1016/j.jviromet.2006.04.009

Sigaloff, K. C., Hamers, R. L., Wallis, C. L., Kityo, C., Siwale, M., Ive, P., et al. (2012). 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. doi: 10.1093/infdis/jis261

Steegen, K., Bronze, M., Papathanasopoulos, M. A., van Zyl, G., Goedhals, D., Van Vuuren, C., et al. (2016a). Prevalence of antiretroviral drug resistance in patients who are not responding to protease inhibitor – based treatment: results from the first national survey in South Africa. *J. Infect. Dis.* 214, 1826–1830. doi: 10.1093/infdis/jiw491

Steegen, K., Bronze, M., Papathanasopoulos, M. A., van Zyl, G., Goedhals, D., Variava, E., et al. (2017). 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. doi: 10.1093/jac/dkw358

Steegen, K., Carmona, S., Bronze, M., Papathanasopoulos, M. A., van Zyl, G., Goedhals, D., et al. (2016b). Moderate levels of pre-treatment HIV-1 antiretroviral drug resistance detected in the first south african national survey. *PLoS One* 11:e0166305. doi: 10.1371/journal.pone.0166305

Sutherland, K., Collier, D. A., Claiborne, D. T., Prince, J. L., Deymier, M. J., Goldstein, R. A., et al. (2016). Wide variation in susceptibility of transmitted/founder HIV-1 subtype C Isolates to protease inhibitors and association with in vitro replication efficiency. *Sci. Rep.* 6:38153. doi: 10.1038/srep38153

Trickey, A., May, M. T., Vehreschild, J. J., Obel, N., Gill, M. J., Crane, H. M., et al. (2017). Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV* 4:e00349-6. doi: 10.1016/S2352-3018(17)30066-30068

UNAIDS (2017). *Data 2017. Programme on HIV/AIDS.* Geneva: UNAIDS.

UNAIDS (2018). *Country Report South Africa.* Geneva: UNAIDS.

Van Zyl, G. U., Liu, T. F., Claassen, M., Engelbrecht, S., de Oliveira, T., Preiser, W., et al. (2013). Trends in genotypic HIV-1 antiretroviral resistance between 2006 and 2012 in South African patients receiving first-and second-line antiretroviral treatment regimens. *PLoS One* 8:e67188. doi: 10.1371/journal.pone.0067188

Van Zyl, G. U., van der Merwe, L., Claassen, M., Zeier, M., and Preiser, W. (2011). Antiretroviral resistance patterns and factors associated with resistance in adult patients failing NNRTI-based regimens in the Western Cape, South Africa. *J. Med. Virol.* 83, 1764–1769. doi: 10.1002/jmv.22189

Wallis, C. L., Mellors, J. W., Venter, W. D. F., Sanne, I., and Stevens, W. (2011). Protease Inhibitor resistance is uncommon in HIV-1 subtype C infected patients on failing second-line lopinavir/r-containing antiretroviral therapy in South Africa. *AIDS Res. Treat. Hindawi* 2011:769627. doi: 10.1155/2011/769627

Conflict of Interest: 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: Investigating the structural effects of statistically enriched mutations identified in Cameroon recombinant subtype CRF02_AG that might lead to Dolutegravir drug resistance.

6.1. Journal article

The manuscript was submitted to BMC infectious diseases' journal on the 04 July 2020. It was subsequently sent back by the reviewers to address the comments on 15 March 2020. Currently the manuscript is under review ID: INFD-D-20-02350. This is the preprint, submitted to the BMC infectious diseases doi: 10.21203/rs.3.rs-40608/v1.

6.2. Author's list

Mikas SG, Isaacs D, Chitongo R, Ikomey GM, Cloete R, Jacobs GB

6.3. Authors contribution

In the enclosed manuscript, I confirmed that I am first author. I performed data analyses and structural analyses. SGM wrote the first draft of the manuscript reviewed by GBJ, DI and RC. All the authors approved the final version of the manuscript.

6.4. Background

The majority of HIV-1 infections are caused by subtype C, with an increase in number of circulating recombinant forms (CRFs), such as CRF02_AG, in the Western regions of Africa. Genomic differences among various subtypes lead to sequence variations in encoded genes. This has consequences for gene targets used for cART, such as IN. Clinical aspects of HIV-1 diversity in resource-limited settings is rarely studied and constitute a great interest in the world. Structural modelling approaches can be a cost effective way to investigate the impact of mutations at a level of protein structure.

6.5. Main findings

The possible impact on protein structure caused by CRF02_AG subtype variations was addressed within the context of a 3D model of the HIV-1 IN complex. We observed 12.8 % (37/287) sequences to contain RAMs with only 1.04% (3/287) of the sequences having major

INSTI resistance mutations: T66A, Q148H, R263K and N155H. 11.8% (34/287) of the sequences contained five different IN accessory mutations, namely Q95K, T97A, G149A, E157Q and D232N. NOPs rates equal or above 50% were found for 66% of central core domain (CCD) positions, 44% C-terminal domain (CTD) positions and 35% of the N-terminal domain (NTD) positions. Our analysis showed that all accessory mutations that resulted in a change in the number of interactions were found within the stable alpha-helix secondary structure element, but any change may result in a change of protein conformation and could interfere with the functioning of the protein.

6.6. **Study significance**

Here we used computational analysis to explore the impact of IN natural occurring polymorphisms on the IN protein. This study shows that NOPs can have an impact on the stability of the protein, by either enhancing the functionality of the protein or rendering down the functionality of the protein.

6.7. **Conclusion**

Molecular modelling can be used to determine the location and effect of mutations on IN protein structure to ascertain their effect on drug binding. This is important as some INSTIs naïve patients sequences are already exhibiting resistance mutations before the roll-out of INSTIs in Cameroon. Integrase strand transfer inhibitors (INSTIs) can be used as the preferred option to be on the forefront of treatment options.

6.8. Preprint manuscript

In Review | BMC Infectious Diseases BMC Series

RESEARCH ARTICLE *Infectious Diseases*

Interaction Analysis of Statistically Enriched Mutations Identified in Cameroon Recombinant Subtype CRF02_AG that can Influence the Development of Dolutegravir Drug Resistance Mutations

Sello Given Mikasi, Darren Isaacs, Rumbidzai Chotongo, George Mondinde Ikomey, Ruben Cloete, Graeme Brendon Jacobs

DOI: 10.21203/rs.3.rs-40608/v1

Abstract

Background The Integrase (IN) strand transfer inhibitor (INSTI), Dolutegravir (DTG), has been given the green light to form part of first-line combination antiretroviral therapy (cART) by the World Health Organization (WHO). DTG is clinically effective against all HIV-1 isolates previously showing resistance to INSTIs.

Methods We evaluated the HIV-1 CRF02_AG IN gene sequences from Cameroon for the presence of resistance-associated mutations (RAMs) against INSTIs and naturally occurring polymorphisms (NOPs), using study sequences (n=20) and (n=278) sequences data derived from HIV Los Alamos National Library (LANL) database. The possible impact of NOPs on protein structure caused by HIV-1 CRF02_AG variations was addressed within the context of a 3D model of the HIV-1 IN complex.

Results We observed 12.8% (37/287) sequences to contain RAMs, with only 1.0% (3/287) of the sequences having major INSTI RAMs: T66A, Q148H, R263K and N155H. Of these,11.8% (34/287) of the sequences contained five different IN accessory mutations; namely Q95K, T97A, G149A, E157Q and D232N. NOP rates equal or above 50% were found for 66% of central core domain (CCD) positions, 44% C-terminal domain (CTD) positions and 35% of the N-terminal domain (NTD) positions.

Conclusions Our analysis indicated that all mutations that resulted in a change in the number of interactions encompassing residues were found within the stable alpha-helix secondary structure element and not in close proximity to the drug active site. Our findings highlight the structural basis for HIV-1 IN interactions and that INSTIs will remain effective against CRF02_AG.

1. Background

Sub-Saharan Africa (SSA) remains the region worst burdened by HIV infection at 70% of the global epidemic. SSA has an extremely high HIV-1 genetic diversity. It is well known that diverse subtypes may affect the clinical treatment outcome in patient management (1). In Cameroon, HIV-1 CRF02_AG continues to be the predominant subtype causing infection, with most other strains, including groups N, O and P, circulating in the country and accounting for a smaller proportion of infections (2, 3). Previous studies have shown that different HIV subtypes may support different mutational pathways, and this could lead to subtype-based differences in acquiring drug resistance (4–6). There is also evidence that natural occurring polymorphisms (NOP) associated with Integrase (IN) activity and occurrence of resistance to IN strand-transfer inhibitors (INSTIs) are subtype-dependent. Subtype-specific polymorphic mutations in the IN gene fragment affect IN DNA binding affinity, when additional mutations are present, and this can influence INSTIs efficacy (4, 5, 7, 8). Computational modelling of resistance-associated mutations (RAMs) against INSTIs, across different HIV-1 subtypes compared to subtype B, showed that the presence of M50I in subtypes A and C, L74I in subtypes A and CRF02_AG, G163R in CRF01_AE, and V165I in subtypes F and CRF01_AE would be associated with lower genetic barrier to resistance in non-B clades (9). Cameroon has seen a considerable reduction of HIV infection, since the introduction of

combination antiretroviral therapy (cART), especially with the rolling-out of programmes like prevention of mother-to-child transmission (PMTCT) and the 90-90-90 strategies to end the AIDS pandemic by 2030 (10). The ability of the HIV-1 to mutate during therapy, can lead to the emergence of HIV-1 drug resistance and this necessitates the need for more effective cART regimens with higher genetic barriers (1). In Cameroon, the HIV-1 drug resistance rates stand at approximately at 10% (11, 12).

The United States of America (USA) Food and Drug Administration (FDA) for the treatment of HIV/AIDS has approved four HIV-1 INSTIs, including raltegravir (RAL), elvitegravir (EVG), dolutegravir (DTG) and bictegravir (BIC) (13). Access to INSTIs in resource-limited countries has been restricted due to its high costs. However, the World Health Organization (WHO) gave the green light to the inclusion of DTG in an alternative 1st-line regimen (14). INSTIs bind to a catalytic site in the catalytic core domain (CCD) of IN and block the strand transfer reaction catalyzed by HIV-1-expressed IN (15, 16).

It is important to note that INSTIs do not lead to the disappearance of the viral genome. Mutations that confer resistance to INSTIs (for example G140S, Q148H and N155H) have been structurally mapped in proximity to the IN catalytic centre (17, 18). Primary resistance to INSTIs, together with residues associated with catalytic activity among different subtypes are highly conserved. HIV-1 sequence and structure-based analyses have shown that polymorphic residues can cause subtype-specific effects, which significantly affect native protein activity, structure and functions important for drug-mediated inhibition of enzyme activity (8).

There is limited information available on the IN structure of CRF02_AG (19). There is henceforth a need to continue monitoring patients to identify additional RAMs and polymorphic mutations that might affect the genetic barrier to the development of RAMs against INSTI (8). The goals of this study was to analyse the Cameroonian CRF02_AG IN gene sequences obtained from the Los Alamos National Library (LANL; www.lank.gov) HIV-1 database to assess the occurrence of mutations and natural occurring polymorphism (NOPs). NOPs are considered secondary mutations that by themselves play a limited, if any, role in resistance. However, their pre-existence might favor a more rapid evolution towards resistance when additional mutations are selected under therapy(20). Furthermore, the possible impact on protein structure caused by statistically enriched NOPs found in CRF02_AG subtype was addressed within the context of a three dimension (3D) model of the HIV-1 IN complex. Molecular modelling of HIV-1 integrase sequences showed that viral, subtypes and codon substitutions would affect the genetic barrier to the emergency of INSTIs resistance differently (8).

Methods

Generation of consensus HIV-1CRF02_AG Integrase sequence

To compare our study CRF02_AG sequences (n = 20) available in GenBank under the following accession numbers: MN816445- MN816488 (21), with the the CRF02_AG IN sequences from Cameroon between

1994–2010 before INSTIs treatment. We performed a search on the HIV Los Alamos National Library (LANL) database for additional (n = 278) sequences (https://www.hiv.lanl.gov/components/sequence/HIVsearch.com), completed on 01 February 2020. Our search criteria included all Cameroonian HIV-1 subtype CRF02_AG IN sequences identified from treatment naïve patients. We selected one sequence per patient and all problematic sequences were excluded from further analyses. The consensus sequence was generated using the database-derived HIV-1 CRF02_AG sequences (n = 278) and CRF02_AG cohort sequences from our previous study (n = 20) (21). Nucleotide sequences were checked for quality and 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 (22). As part of quality control, each of the viral sequences were inferred on a phylogenetic tree in order to eliminate possible contamination.

HIV-1 subtyping using online programs.

HIV-1 subtyping based on IN sequences was performed using REGA version 3 (http://www.bioafrica.net/rega-genotype/html/subtypinghiv. html) and COMET-HIV (https://comet.lih.lu/).

Drug resistance analysis

Mutations associated with resistance to INSTIs were identified using the Stanford University genotypic resistance interpretation algorithm, HIVdb version 8.3 (http://hivdb.stanford.edu/). Drug resistance mutations results were classified as either major or minor mutations, last accessed 01 April 2020.

Homology 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-1CRF02_AG IN using the consensus sequence of recombinant subtype CRF02_AG 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 5U1 (22). 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 (22).

Mutation analysis and stability predictions

The energy minimized structure of HIV-1 CRF02_AG IN was used to introduce variants identified from the sequence alignment using PyMol mutagenesis wizard. Once the WT and variant structures were generated, the change in energy after introduction of a mutation was calculated using mutation cut-off scanning matrix (mCSM). mCSM uses graph-based distance patterns and calculates a Delta-delta G-score for the impact of the mutation on the neighbouring residues and provides a phenotypic assessment namely destabilizing or stabilizing. The loss or gain of interactions between the variant and WT neighbouring residues was determined using Pymol

Results

HIV-1 subtyping.

HIV-1 subtyping was done using HIV-1 subtyping online-automated tools and all sequences were verified as HIV-1 subtype CRF02_AG.

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 287 sequences collected between 1994 and 2010. These sequences were subsequently screened for the presence of RAMs. We identified 12.8% (37/287) sequences to contain RAMs, with only 1.0% (3/287) having major INSTI RAMs: T66A, Q148H, R263K and N155H. Two mutations, Q148H and R263K, occurred together in one sequence (0.3%), whereas T66A and N155H were present individually in one sequence each. 11.8% (34/287) of the sequences contained five different IN accessory mutations, namely Q95K, T97A, G149A, E157Q and D232N. Mutations G149A and D232H occurred together in one sequence (0.3%). Notably, one sequence dating back from 2010 had two major mutations; Q148H and R263K in combination with two other minor mutations G149 and D232H.

Generation of consensus Cameroonian's HIV-1 CRF02_AG sequence

The consensus sequences generated using the database-derived HIV-1 CRF02_AG sequences (n = 287) and cohort sequences (n = 20), identified 20 naturally occurring polymorphisms (NOPS): E11D ,K14R, V31I, M50I, I72V, L74MVI, L101I, T112V, T124A, T124A, G134N, I135V, K136K/Q, V201I, T206S, T218I, L234I, A265V, R269K, S283G (Fig. 1). Three of these (E11D, K14R and V31I) belong to the NTD, whereas M50I belongs to the loop region connecting the NTD and CTD. Eleven NOPs (I72V, L74MVI, L101I, T112V, T124A, T124A, G134N, I135V, K136K/Q, V201I and T206S) are part of the CCD, and the remaining five (T218I, L234I, A265V, R269K and S283G) belong to the CTD.

Molecular modelling and structural analysis

Figure 2, shows the 3D tetrameric structure for HIV-1 CRF02_AG IN that consist of 288 amino acids, 10 alpha helices, 9 beta sheets and 19 coil regions. The homology model passed all the external 3D quality validation tools subjected to it. The Verify3D score for the model was predicted to be 71.1%, while ERRAT score for all the chains was 86.0% and higher, the Ramachandran plot indicated that 98.0% of residues occur in most favoured and allowed regions, and the Prosa Z-score was – 6.18 which is in range with proteins of similar size. Superimposing the template 5u1c onto the target energy minimized structure indicated an RMSD value of 0.212 Å, suggesting very little backbone deviation in main chain atoms (Fig. 2b). Figure 2c, shows the locations of the nine mutations relative to the active site. Furthermore, mCSM predictions indicated that eight of the nine variants, i.e. the M50I, L74I, L74M, T97A, G118S, S119R, P145S, E157Q substitutions resulted in destabilizing effects of -0.582, -1.069, -0.93, -1.051, -0.492, -0.091, -0.485 and – 1.111 Kcal/Mol each, respectively. Only substitution Q95K resulted in a slightly stabilizing effect of 0.146Kcal/Mol. Interaction analysis of the single amino acid changes indicated differences in the number and type of interaction between neighbouring residues and the DNA. The T97A showed four polar contacts for T97 compared to the three of A97 (Table 1). This suggests a loss of stable contacts in this region that could destabilize the protein structure. Moreover, the S119R substitution indicated interactions with the known active site residue E92 that could alter the IN active site reducing INI binding (Table 1). Inspection of the E157 residue showed four contacts with neighbouring residues while Q157 revealed five polar contacts of which two were with DNA (Table 1). In addition, the remaining other six substitutions; M50I, L74I, L74M, Q95K, G118S and P145S showed no changes in the number or type of interactions, implying no strong effect on the protein structure and function (Table 1).

Table 1
Summary of all interactions observed between the INSTIs and CRF02_AG IN subtype. 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, Bold-Change in amino acid

Number	Mutation	# Polar contacts	
		WT	Mutant
1	M50I	None	None
2	L74M	1 (Glu 87)	1 (Glu 87)
3	L74I	1 (Glu 87)	1 (Glu 87)
4	Q95K	2 (1(Ala98), 1(Tyr99))	2 (1(Ala98), 1(Tyr99))
5	T97A	4 (2(T93), 1(G94), 1(I101))	3 (1(T93), 1(G94), 1(I101))
6	G118S	None	None
7	S119R	3 (1(Thy29), 1(Asn120), 1(Thr122)	3(1(Thy29), 1(**Glu92**), 1(Thr122))
8	P145S	1 (Gln148)	1 (Gln148)
9	E157Q	4 (1(Ser153), 1(Met154), 1(Lys156), 1 (Ile161))	5 (1(**Thy20**), 1(**Ade21**), 1(Ser153), 1 (Met154), 1(Ile161))

Discussion

Despite INSTIs possessing a higher genetic barrier against resistance, studies performed from high-income countries shows that the development of RAMS against INSTIs can occur, via acquired drug resistance mutations (DRM) and/or transmitted DRM, leading to reduced susceptibility to INSTIS and possible treatment failure (23, 24). The IN mutations usually associated with reduced INSTIs susceptibility include both polymorphic mutations and non-polymorphic mutations (25, 26). Other studies have reported that several NOPs can affect structural stability and flexibility of the IN protein structure (27,28). Low rates of IN mutations against INSTIs were reported in Cameroon (21). The WHO has recommended the use of DTG as part of first-line regimens (14). With the approval of using INSTIs worldwide, it is projected that approximately 57% of people living with HIV will be receiving DTG based regimens, including new-borns and children (29). It is imperative to monitor the presence of mutations against INSTIs that can affect treatment outcomes. Furthermore, there is limited knowledge and data available for INSTI RAMs from studies that focuses on the SSA region, where over two-thirds of the current infected individuals reside (30, 31). Our previous studies found low level of RAMSs against INSTIs (21). Our recent studies on CRF02_AG, reported that NOPs can have an effect on DTG binding with or without combination of primary mutation (19, 24, 28). In the current, study we analysed the CRF02_AG IN gene sequences for the presence of polymorphic and non-polymorphic mutations. Four major INSTIs mutations were present in the database sequences: T66A, Q148H, N155H and R263K. R263K displayed

moderate level resistance against EVG (12-fold) (32) and seems to confer low-level resistance against DTG. Structural analyses have suggested that DTG shares a similar interfacial mechanism of inhibition with EVG and RAL, but is able to make more intimate contacts with the viral DNA (33). In addition, DTG is purported to be able to readjust its position and conformation in response to structural changes in the active sites of EVG- or RAL resistant IN enzymes and avoid some cross-resistance due to slower dissociation (24, 34). Two principal mutation pathways identified from our study that reduces susceptibility to RAL are Q148HKR and N155H. These mutations are located in close proximity to the Integrase's active site and each mutation significantly reduces viral fitness by 92-fold for Q148R, 30-fold for N155H (35). Q148H and N155H mutations are thought to trigger conformational changes within the catalytic pocket that result in an increase in the binding energy of INSTIs to IN (36). Variant T66A found in 0.3% of our sequence cohort, is a primary mutation normally selected by EVG treatment. It is associated with reduced susceptibility to EVG by 5-fold, although T66A mutations also bear cross-resistance to DTG and is selected by RAL (35). Abraham et al., 2013, showed that T66A lies within the 2 sheet distal from the DDE triad, its proximity to the viral DNA 3' end and N155H, suggests that T66I/A/K mutations may sterically affect viral DNA binding and/or metal ion coordination through N155H (35). The fact that only 1.0% of sequences analysed contained INSTI primary RAMs suggest that mutations against INSTIs will need to be monitored carefully against Cameroonians living with HIV. This result is in agreement with other studies done in Africa (2, 21, 37–39), Asia (40, 41), Europe (42, 43), and South America (44) where studies showed a low frequency of INSTI primary RAMs. IN accessory RAMs observed in our study include Q95K, T97A, G149A, E157Q and D232N. T97A mutation can reduce EVG susceptibility by 3-fold (35) and combination of T97A mutation with other INSTI RAMs lead to reduced susceptibility to RAL (45, 46) and DTG (47, 48). Similarly, E157Q individually has a negligble effect on the susceptibility to INSTIs; however, a combination of E157Q with other INSTI RAMs may lead to reduced susceptibility to INSTIs. Individuals containing E157Q (49) mutation in combination with other IN RAMs showed reduced susceptibility to DTG. Eleven of the INSTI NOPs, were in the IN CCD. This IN region contains the endonuclease and polynucleotide transferase site and is involved in DNA substrate recognition, binding and chromosomal integration of the newly synthesized double-strand viral DNA into the host genomic DNA (50–52). The other five NOPs mutation were in the CTD, a region that helps stabilize the integrase–viral DNA complex (53). The mutations on the CTD regions involved in recognition, binding, integration and stabilization of HIV into the host DNA shows the potential for these mutations to affect the IN function and response to INSTIs. We further analyzed the effect of NOPs on the stability of the structures and neighbouring residues. Most of the variants noted in our study were shown to destabilise the protein structure, except one residue Q95K, that showed to exert a slightly stabilising effect on the protein structure. We assume that the destabilising effects will render down the functionality of the protein or enzyme. However at this stage, we cannot confidently state this, as additional analyses like molecular dynamic simulations would be required for confirmation. Molecular dynamic simulations will allow us to investigate whether these variants or mutations exert an effect upon protein folding or unfolding. This would allow us to conclude what effect these variants has upon the functionality of IN. While interactions between Q157 variant residue and DNA could promote viral DNA integration and prevent IN drug binding. Interestingly, all the mutations (T97A, S119R and E157Q). In

addition, all accessory mutations that resulted in a change in the number of interactions encompassing residues were found within the stable alpha-helix secondary structure element and not in close proximity to the drug active site.

Conclusion

With the use of DTG and the high HIV-1 diversity noted in SSA, it is probable that some individuals on this regimen may well develop resistance mutations at some point during treatment. Proper surveillance for management of such cases should include modelling and screening for the presence of INSTIs major RAM. This would help in identifying gaps and actions needed to improve effectiveness of DTG once adopted.

Abbreviations

IN: Integrase ; INSTI: Integrase strand transfer inhibitors; NOP: Natural occurring polymorphism; SSA: Sub-Saharan Africa; RAMS: Resistance associated mutations; WHO: World Health Organisation; LANL: Los Alamos National Library; CCD: Central core domain (CCD) ; CTD: C-terminal domain; NTD: N-terminal domain; cART: combination antiretroviral therapy; PMCT: prevention of mother-to-child transmission; FDA: Food and Drug Administration; RAL: raltegravir; EVG: elvitegravir; DTG: dolutegravir; BIC: bictegravir; 3D: three dimension; MRC: Medical Research Council; Mcsm: mutation cut-off scanning matrix.

Declarations

Ethics approval and consent to participate

This study form part of a broader study, for which the ethic was obtained from the Health Research Ethics Committee of Stellenbosch University (N14/10/130 and N15/08/071). The database sequences were 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.

Consent for publication

Not applicable.

Availability of data and materials

The dataset used and/or analysed during the current study are found at the Stellenbosch University database and available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding's

The study was supported by the National Research Foundation (NRF) of South Africa, Poliomyelitis Research Foundation (PRF) of South Africa, Harry Crossley Foundation, South African Research Chairs Initiative of the Department of Science, Technology (DST), South African Medical Research Council (SAMRC) and Higher Education Department, next Generation of Programme (nGAP)

Author's contributions

GBJ, GMI and RC conceived and designed the study. SGM conducted sequence analyses and structural analyses. SGM wrote the first draft of the manuscript and reviewed by RC, SS, GBJ, DI and RC. All the authors approved the final version of the manuscript.

Acknowledgments

Appreciations goes to the collaborators from South African National Bioinformatics Institute (SANBI), University of the Western Cape, South Africa and University of Yaoundé I, Cameroon.

References

1. Han Y-S, Mesplède T, Wainberg MA. Differences among HIV-1 subtypes in drug resistance against integrase inhibitors. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2016 Dec;46:286–91.
2. Obasa AE, Mikasi SG, Brado D, Cloete R, Singh K, Neogi U, et al. Drug Resistance Mutations Against Protease , Reverse Transcriptase and Integrase Inhibitors in People Living With HIV-1 Receiving Boosted Protease Inhibitors in South Africa. Front Microbiol. 2020;11(March):1–9.
3. Plantier J, Djemai M, Leme V, Reggiani A, Leoz M, Burc L, et al. Census and Analysis of Persistent False-Negative Results in Serological Diagnosis of Human Immunodeficiency Virus Type 1 Group O Infections . 2009;47(9):2906–11.
4. Bar-magen T, Sloan RD, Faltenbacher VH, Donahue DA, Kuhl BD, Oliveira M, et al. Comparative biochemical analysis of HIV-1 subtype B and C integrase enzymes. 2009;2048.
5. Bar-magen T, Donahue DA, Mcdonough EI, Kuhl D, Faltenbacher VH, Xu H, et al. HIV-1 subtype B and C integrase enzymes exhibit differential patterns of resistance to integrase inhibitors in biochemical assays. 2010;(June):2171–9.

6. Oliveira MF, Ramalho DB, Abreu CM. Genetic Diversity and Naturally Polymorphisms in HIV Type 1 Integrase Isolates from Maputo , Mozambique : 2012;28(12):1788–92.
7. Quashie PK, Han Y, Hassounah S, De TM. Structural Studies of the HIV-1 Integrase Protein : Compound Screening and Characterization of a DNA-Binding Inhibitor. 2015;1–18.
8. Theys K, Libin P, Laethem K Van, Abecasis AB. An evolutionary-based approach to quantify the genetic barrier to drug resistance in fast-evolving viruses : an application to HIV-1 subtypes and integrase inhibitors. 2019;
9. Kristof Theys, Pieter JK Libin, Kristel Van Laethem ABA. An evolutionary-based approach to quantify the genetic barrier to drug resistance in fast- evolving viruses: an application to HIV-1 subtypes and integrase inhibitors. Antimicrob Agents Chemother. 2019;
10. Bain LE, Nkoke C, Noubiap JJN. UNAIDS 90 – 90 – 90 targets to end the AIDS epidemic by 2020 are not realistic : comment on " Can the UNAIDS 90 – 90 – 90 target be achieved ? A systematic analysis of national HIV treatment cascades ." 2020;2016–8.
11. Tchouwa GF, Eymard-Duvernay S, Cournil A, Lamare N, Serrano L, Butel C, et al. Prevalence of pretreatment HIV drug resistance in Cameroon following a nationally representative WHO survey. J Antimicrob Chemother [Internet]. 2018 Jun 19;73(9):2468–74. Available from: https://doi.org/10.1093/jac/dky221
12. Ikomey GM, Claire M, Assoumou O, Gichana JO, Njenda D. Original article Observed HIV drug resistance associated mutations amongst naïve immunocompetent children in Yaoundé , Cameroon. 2017;7(December):178–85.
13. Jacobson JM, Flexner CW. Universal antiretroviral regimens: thinking beyond one-pill-once-a-day. Curr Opin HIV AIDS. 2017 Jul;12(4):343–50.
14. WHO. Consolidated Guidelines On THE USE OF ANTIRETROVIRAL DRUGS FOR TREATING AND PREVENTING HIV INFECTION. Recommendations for a Public Health Approach. Second Edition. World Health Organization, [Internet]. 2019. Available from: https://apps.who.int/iris/bitstream/handle/10665/208825/%0A9789241549684%0A
15. Yoshinaga T, Kobayashi M, Seki T, Miki S, Wakasa-morimoto C, Suyama-kagitani A. Antiviral Characteristics of GSK1265744 , an HIV Integrase Inhibitor Dosed Orally or by Long-Acting Injection. 2015;59(1):397–406.
16. Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA, et al. Inhibitors of Strand Transfer That Prevent Integration and Inhibit HIV-1 Replication in Cells. 2000;287(January):646–51.
17. Menéndez-arias L. Molecular basis of human immunodeficiency virus type 1 drug resistance : Overview and recent developments. 2013;98:93–120.
18. Hare S, Smith SJ, Me M, Jaxa-chamiec A, Pommier Y, Hughes SH, et al. Structural and Functional Analyses of the Second-Generation Integrase Strand Transfer Inhibitor Dolutegravir (S / GSK1349572) □. 2011;80(4):565–72.
19. Isaacs D, Mikasi SG, Obasa AE, Ikomey GM. Structural comparison of diverse HIV-1 subtypes using Molecular Modelling and Docking analyses of Integrase inhibitors. 2020;(February).

20. Paraschiv S, Otelea D, Dinu M, Maxim D, Tinischi M. Polymorphisms and resistance mutations in the protease and reverse transcriptase genes of HIV-1 F subtype Romanian strains . 2007;(May 2004):123–8.
21. Mikasi SG, Gichana JO, Walt C Van Der, Brado D, Obasa AE, Njenda D, et al. Transfer Inhibitor-Naive HIV-1 Patients from Cameroon. 2020;36(5):450–5.
22. 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. PLoS One. 2013;4(5):1–11.
23. Lepik KJ, Harrigan PR, Yip B, Wang L, Robbins MA, Zhang WW, et al. Emergent drug resistance with integrase strand transfer inhibitor-based regimens. 2017;(February).
24. Hurt CB, Sebastian J, Hicks CB, Eron JJ. Resistance to HIV Integrase Strand Transfer Inhibitors Among Clinical Specimens in the United States , 2009 – 2012. 2014;58:2009–12.
25. Rhee S, Liu TF, Kiuchi M, Zioni R, Gifford RJ, Holmes SP, et al. Natural variation of HIV-1 group M integrase: Implications for a new class of antiretroviral inhibitors. 2008;11:1–11.
26. Anstett K, Brenner B, Mesplede T, Wainberg MA. HIV drug resistance against strand transfer integrase inhibitors. Retrovirology. 2017;1–16.
27. Rogers L, Obasa AE, Jacobs GB, Sarafianos SG, Sönnerborg A, Neogi U, et al. Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. 2018;9(August):1–9.
28. Chitongo R, Obasa AE, Mikasi SG, Id BJ, Id RC. Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase- Dolutegravir binding. PLoS One [Internet]. 2020;1–15. Available from: http://dx.doi.org/10.1371/journal.pone.0223464
29. Gupta A, Juneja S, Vitoria M, Habiyambere V, Nguimfack D, Doherty M, et al. Projected Uptake of New Antiretroviral (ARV) Medicines in Adults in Low- and Middle- Income Countries : A Forecast Analysis 2015- 2025. PLoS One. 2016;1–18.
30. Mabeya S, Nyamache A, Ngugi C, Nyerere A, Lihana R. Characterization of HIV-1 Integrase Gene and Resistance Associated Mutations Prior to Roll out of Integrase Inhibitors by Kenyan National HIV-Treatment Program in Kenya. Ethiop J Health Sci. 2020 Jan;30(1):37–44.
31. Masoud S, Kamori D, Barabona G, Mahiti M, Sunguya B, Lyamuya E, et al. Circulating HIV-1 Integrase Genotypes in Tanzania: Implication on the Introduction of Integrase Inhibitors-Based Antiretroviral Therapy Regimen. AIDS Res Hum Retroviruses. 2020 Jun;36(6):539–43.
32. Hassounah SA, Mesplède T, Quashie PK, Oliveira M, Sandstrom PA, Wainberg A. Effect of HIV-1 Integrase Resistance Mutations When Introduced into SIVmac239 on Susceptibility to Integrase Strand Transfer Inhibitors. 2014;88(17):9683–92.
33. Kobayashi M, Yoshinaga T, Seki T, Wakasa-Morimoto C, Brown KW, Ferris R, Foster SA, Hazen RJ, Miki S, Suyama-Kagitani A, Kawauchi- Miki S, Taishi T, Kawasuji T, Johns BA, Underwood MR, Garvey EP, Sato A FT. In vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. Antimicrob Agents Chemother. 2011;55:813– 821.

34. Hightower KE, Wang R, Deanda F, Johns BA, Weaver K, Shen Y, 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. 2011;55(10):4552–9.

35. Abram ME, Hluhanich RM, Goodman DD, Andreatta KN, Margot NA, Ye L, et al. Impact of Primary Elvitegravir Resistance-Associated Mutations in HIV-1 Integrase on Drug Susceptibility and Viral Replication Fitness. 2013;57(6):2654–63.

36. Hare S, Gupta SS, Valkov E, Engelman A, Place N. HHS Public Access. 2010;464(7286):232–6.

37. Brado D, Obasa AE, I 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. 2018;(December 2017):1–9.

38. Seth C. Inzaule, Raph L. Hamers, Marc Noguera-Julian MC, Parera M, Paredes TFR de W and R. Primary resistance to integrase strand transfer inhibitors in patients infected with diverse HIV-1 subtypes in sub-Saharan Africa. 2018;92(February):1167–72.

39. Villabona-arenas CJ, Domyeum J, Mouacha F, Delaporte E, Peeters M, Mpoudi-ngole E, et al. HHS Public Access. 2015;210–6.

40. Kim J-Y, Kim E-J, Choi J-Y, Kwon O-K, Kim GJ, Choi SY, et al. Genetic variation of the HIV-1 integrase region in newly diagnosed anti-retroviral drug-naïve patients with HIV/AIDS in Korea. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2011 Aug;17(8):1155–9.

41. Kotaki T, Khairunisa SQ, Sukartiningrum SD, Witaningrum AM, Rusli M, Diansyah MN, et al. Detection of drug resistance-associated mutations in human immunodeficiency virus type 1 integrase derived from drug-naive individuals in Surabaya, Indonesia. AIDS Res Hum Retroviruses. 2014 May;30(5):489–92.

42. Meixenberger K, Yousef KP, Smith MR, Somogyi S, Fiedler S, Bartmeyer B, et al. Molecular evolution of HIV-1 integrase during the 20 years prior to the first approval of integrase inhibitors. 2017;1–13.

43. Tostevin A, White E, Dunn D, Croxford S, Delpech V, Williams I, et al. Recent trends and patterns in HIV-1 transmitted drug resistance in the United Kingdom. HIV Med [Internet]. 2017 Mar; Available from: http://www.scopus.com/inward/record.url?scp=84994415935&partnerID=8YFLogxK

44. Passaes, C.B.; Guimaraes, M.L.; Fernandez, S.L.; Lorete Rdos, S.; Teixeira, S.L.; Fernandez, J.C.; Morgado MG. Lack of Primary Mutations Associated With Integrase Inhibitors Among HIV-1 Subtypes B , C , and F Circulating in Brazil. 2009;51(413):7–12.

45. Eron JJ, Clotet B, Durant J, Katlama C, Kumar P, Lazzarin A, et al. Safety and Ef fi cacy of Dolutegravir in Treatment-Experienced Subjects With Raltegravir-Resistant HIV Type 1 Infection : 24-Week Results of the VIKING Study. 2013;207.

46. Fransen S, Gupta S, Danovich R, Hazuda D, Miller M, Witmer M, et al. Loss of Raltegravir Susceptibility by Human Immunodeficiency Virus Type 1 Is Conferred via Multiple Nonoverlapping Genetic Pathways . 2009;83(22):11440–6.

47. Melikian GL, Rhee S, Varghese V, Porter D, White K, Taylor J, et al. Non-nucleoside reverse transcriptase inhibitor (NNRTI) cross-resistance: implications for preclinical evaluation of novel

NNRTIs and clinical genotypic resistance testing. 2014;(August 2013):12–20.

48. Naeger LK, Harrington P, Komatsu T, Deming D. Effect of dolutegravir functional monotherapy on HIV-1 virological response in integrase strand transfer inhibitor resistant patients. Antivir Ther. 2016;21(6):481–8.
49. Charpentier C, Malet I, Andre-garnier E, Storto A, Bocket L, Amiel C, et al. Phenotypic analysis of HIV-1 E157Q integrase polymorphism and impact on virological outcome in patients initiating an integrase inhibitor-based regimen. 2018;(January):1039–44.
50. Mulu A, Maier M, Liebert UG. Lack of integrase inhibitors associated resistance mutations among HIV - 1C isolates. J Transl Med. 2015;1–8.
51. Li Y, Xuan S, Feng Y, Yan A. Targeting HIV-1 integrase with strand transfer inhibitors. Drug Discov Today. 2015 Apr;20(4):435–49.
52. Mouscadet, J.F.; Delelis, O.; Marcelin, A.G.; Tchertanov L. Resistance to HIV-1 integrase inhibitors : A structural perspective. 2010;13:139–50.
53. Cai MI, Zheng RI, Caffrey M, Craigie R, Clore GM, Gronenborn AM. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. 1997;4(1):4–5.

Figures

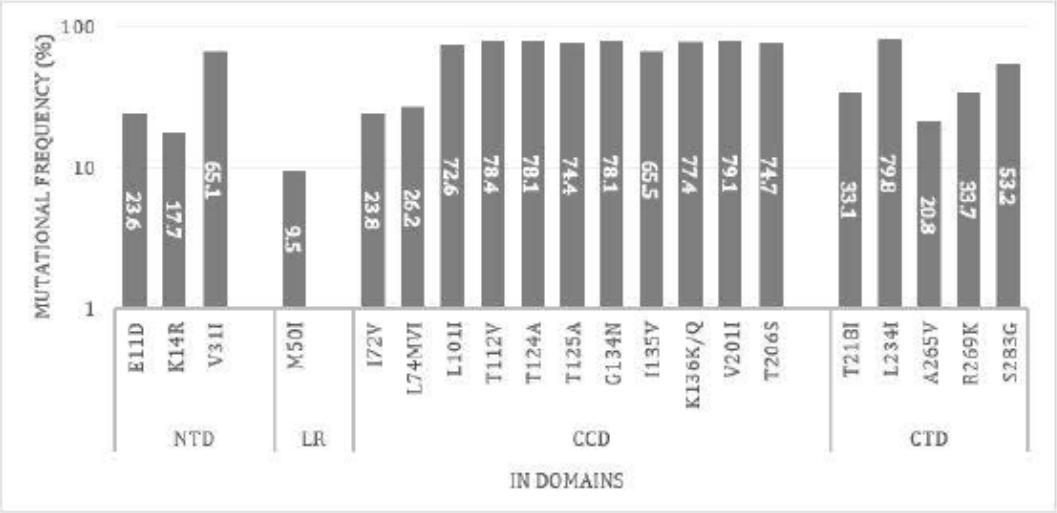

Figure 1

Prevalence of NOPs in IN genes from CRF02_AG subtypes. The figure shows the distribution of variants among the 293 and 20 CRF02_AG full-length integrase sequences. The proteins varied from 1 -288 amino acids in length. Divided into: N-terminal domain (NTD) (residues 1–50), catalytic core domain (CCD) (residues 50–212) and C-terminal domain (CTD) composes of amino acids 213–288.

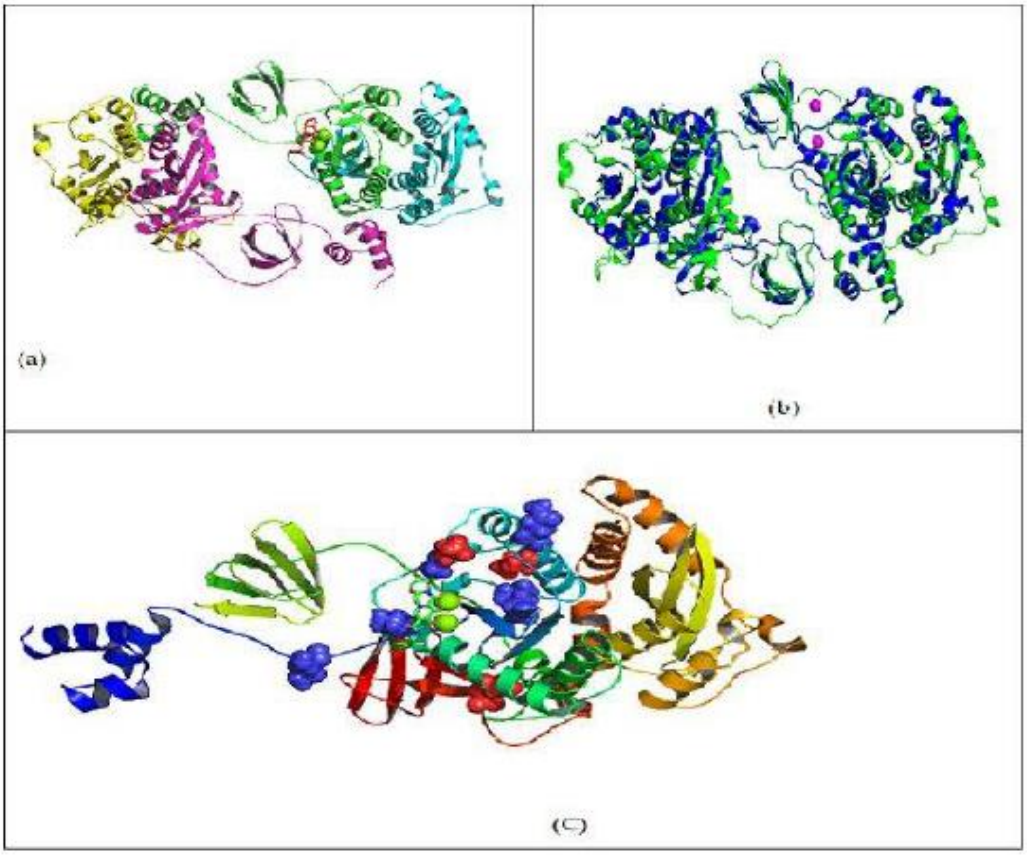

Figure 2

3D tetrameric structure for HIV-1 CRF02_AG IN. (A): Three-dimensional tetrameric structure predicted for HIV-1 CRF_02 AG IN in complex with magnesium ions. Chain A: green, Chain B: cyan, Chain C: Magenta, Chain D: yellow, Magnesium ions shown as spheres coloured in green and Dolutegravir shown as sticks coloured in red. No DNA shown. (B): Structural superimposition of HIV-1 AG IN onto HIV-1 B (5u1c) in complex with MG ions. CRF02_AG IN: green, 5u1c: blue, DNA= magenta, MG shown as spheres. (C): Locations of destabilizing mutations on HIV-1 integrase CRF02_AG structure. Mutations that affect the protein structure are shown as red spheres and mutations with no effects are shown as blue spheres.

Chapter 7: Structural comparison of diverse HIV-1 subtypes using Molecular modelling and Molecular Docking of integrase inhibitor.

7.1. **Journal article**

The manuscript was submitted to Viruses' journal on the 31 January 2020. It was subsequently sent back by the reviewers to address the comments on 15 March 2020. Currently the manuscript was re-submitted to the journal on 10 June 2020 after addressing the reviewers comments ID: Viruses-721103. This is the preprint ,submitted to the preprints.org > life sciences > virology > doi: 10.20944/preprints202002.0062.v1.

7.2. **Author's list**

Isaacs D, **Mikasi SG**, Obasa AE, Ikomey GM, Jacobs GB and Cloete R.

7.3. **Authors contribution**

In the enclosed manuscript; Mr. Isaacs and I, are joint-first authors, with equal contributions to the work performed. I provided raw sequence data. Mr. Isaacs and I, we performed sequence analysis .We both started the first manuscript draft and subsequently addressed the reviewers comment to the manuscript.

7.4. **Background**

HIV-1 integrates its viral DNA into the host genome using two processes: 3′-end processing and strand-transfer activities. INSTIs target the HIV-1 gene. These drugs have high genetic barrier against resistance mutations. As these drugs become preferred option to be used as part of first-line therapy world-wide, there is a need to understand the mechanism of resistance among subtype C that dominates in South Africa and subtype CRF02_AG that dominates in West Africa, particularly Cameroon. Studies have showed that mutations can be subtype-specific and for this reason, IN polymorphisms among various subtypes might be responsible for different mutational patterns. HIV-1 IN structure of high resolution has been a major drawback to the study and understanding the mutational pathways and the impact of NOPs on treatment outcome.

7.5. **Main findings**

In this study, we investigated the impact of NOPS on IN structure of HIV-1, using molecular modelling and docking approach. NOPs were identified among Cameroonian sequences (n=16) and South African sequences (n=17). NOPs did not show to exert any impact on INSTIs binding. All INSTIs showed to have same binding affinity to each IN structure. INSTIs shows to remain the effective drugs against INSTI treatment naïve individuals living with HIV. The study support wide scale access to DTG in Sub-Saharan Africa, where there is a high HIV pandemic and to combat the spread and control of HIV.

7.6. **Study significance**

The study provided important data that can provide guidance for investigating how NOPs affect IN structure and INSTIs binding affinity. The analysis of these molecular modelling approaches provides crucial clues to the possible future discovery of novel IN mutations

7.7. **Conclusion**

NOPs that were present in the study did not possess any form of effect against binding affinity of INSTIs. DTG will remain effective once it becomes available in this region.

7.8. **Published article**

viruses

MDPI

Article

Structural Comparison of Diverse HIV-1 Subtypes using Molecular Modelling and Docking Analyses of Integrase Inhibitors

Darren Isaacs [1,†], Sello Given Mikasi [2,†], Adetayo Emmanuel Obasa [2], George Mondinde Ikomey [3], Sergey Shityakov [4,5], Ruben Cloete [1,*] and Graeme Brendon Jacobs [2,*]

1 South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Cape Town 8000, South Africa; 3433660@myuwc.ac.za
2 Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Francie van Zijl Avenue, P.O. Box 241, Cape Town 8000, South Africa; mikasi@sun.ac.za (S.G.M.); obasa@sun.ac.za (A.E.O.)
3 Centre for the Study and Control of Communicable Diseases (CSCCD), University of Yaoundé 1, Yaoundé P.O. Box 8445, Cameroon; mondinde@yahoo.com
4 Department of Psychiatry & Mind-Body Interface Laboratory (MBI-Lab), China Medical University Hospital, Taichung 404, Taiwan; shityakoff@hotmail.com
5 Department of Bioinformatics, University of Würzburg, Würzburg 97074, Germany
* Correspondence: ruben@sanbi.ac.za (R.C.); graeme@sun.ac.za (G.B.J.); Tel.: +27-21-938-9744 (G.B.J.)
† These authors contributed equally to this work.

Received: 10 June 2020; Accepted: 5 August 2020; Published: 26 August 2020

check for updates

Abstract: The process of viral integration into the host genome is an essential step of the HIV-1 life cycle. The viral integrase (IN) enzyme catalyzes integration. IN is an ideal therapeutic enzyme targeted by several drugs; raltegravir (RAL), elvitegravir (EVG), dolutegravir (DTG), and bictegravir (BIC) having been approved by the USA Food and Drug Administration (FDA). Due to high HIV-1 diversity, it is not well understood how specific naturally occurring polymorphisms (NOPs) in IN may affect the structure/function and binding affinity of integrase strand transfer inhibitors (INSTIs). We applied computational methods of molecular modelling and docking to analyze the effect of NOPs on the full-length IN structure and INSTI binding. We identified 13 NOPs within the Cameroonian-derived CRF02_AG IN sequences and further identified 17 NOPs within HIV-1C South African sequences. The NOPs in the IN structures did not show any differences in INSTI binding affinity. However, linear regression analysis revealed a positive correlation between the Ki and EC50 values for DTG and BIC as strong inhibitors of HIV-1 IN subtypes. All INSTIs are clinically effective against diverse HIV-1 strains from INSTI treatment-naïve populations. This study supports the use of second-generation INSTIs such as DTG and BIC as part of first-line combination antiretroviral therapy (cART) regimens, due to a stronger genetic barrier to the emergence of drug resistance.

Keywords: integrase; naturally occurring polymorphisms; HIV-1; molecular modelling; molecular docking; diversity

---

1. Introduction

The HIV/AIDS pandemic continues to be a significant problem worldwide [1]. The viral integration process, which is the insertion of viral DNA into host genomic DNA, is an indispensable step of the retroviral life cycle and is catalyzed by the viral integrase (IN) enzyme [2]. Integration is achieved via two distinct sequential catalytic activities, 3′ processing and strand transfer. IN first processes

viral DNA by excising a dinucleotide at the 3′ end, exposing hydroxyl ends. IN then catalyzes the introduction of the prepared DNA into genomic DNA by facilitating a nucleophilic attack upon genomic DNA [3,4]. The same active site in IN, which contains a retroviral highly conserved DDE motif and magnesium ions, performs both activities [5,6]. HIV-1 IN is a 32 kDa protein that functions as a tetramer or multimer [3,4]. A monomer consists of three distinct domains; the N-terminal domain (NTD) comprising residues 1–46, the catalytic core domain (CCD) comprising residues 56–186 within which the active site DDE motif (aspartate (D64), aspartate (D116), and glutamate (E152) is present, and the C-terminal domain (CTD) comprising residues 195–288 [7,8]. Several integrase strand transfer inhibitors (INSTIs) have been developed to target HIV-1 IN to prevent viral integration into the host genome. The four INSTIs available thus far include raltegravir (RAL) and elvitegravir (EVG) that are considered as first-generation inhibitors, while dolutegravir (DTG) and bictegravir (BIC), along with the late-phase clinically trialed cabotegravir (CBT), are classified as second-generation INSTIs [9]. At present, first-line combination antiretroviral therapy (cART) regimens for HIV-1 are expected to include the INSTI DTG according to World Health Organization (WHO) recommendations [10], as it has been shown to possess a higher genetic barrier to drug resistance development as compared with RAL and EVG [11]. HIV-1 is a genetically highly diverse virus, forming different subtypes, recombinant and region-specific variants [12]. Development of INSTIs, as with most pharmaceutical agents, was primarily conducted by companies in First World nations, where subtype B is the most predominant variant or subtype [13,14]. It remains unclear what effects naturally occurring polymorphisms (NOPs) may have upon the IN structure and INSTI susceptibility. This lack of data poses a challenge in concluding the effects of NOPs on the binding of INSTIs to HIV-1 IN subtypes [15–19].

In this study, computational methods, which include molecular modelling and docking, were used to determine if NOPs affect INSTI binding to HIV-1 IN subtype C (HIV-1C) and a circulating recombinant form of HIV-1 IN CRF_02_AG. The recently resolved cryogenic electron microscopy full-length HIV-1 subtype B IN structure allowed us to build accurate and complete tetrameric three-dimensional structures of HIV-1C IN and of HIV-1 IN CRF_02_AG. The value of having accurate protein models allows us to infer the exact mode of interactions formed between active site residues of HIV-1 subtypes and drug atoms. HIV-1 subtype C derived from a South African cohort was chosen as one of our IN models, as it represents the most prevalent subtype both globally and for sub-Saharan Africa in particular [14,19,20]. Our focus on a Cameroonian cohort was spurred on by the previously reported HIV-1 diversity present in Cameroon with all known subtypes/variants found within Cameroon [21,22]. Furthermore, the full-length Cryo-EM HIV-1 IN structure, which was used as the template in our molecular modelling, served additionally as a subtype B IN model in our study, the predominant strain in developed nations.

2. Materials and Methods

2.1. Ethics Statement

The study used sequences from two African settings: South Africa and Cameroon.

Ethical permission for this study was obtained from the Health Research Ethics Committee of Stellenbosch University (N14/10/130—approved on 13 August 2019 and N15/08/071—approved on 26 March 2019). Ethics protocols are revised and renewed each year. 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 sequence analyses.

2.2. Study Design

HIV-1-positive plasma samples were obtained from the Centre for the Study and Control of Communicable Diseases (CSCCD), University of Yaoundé I, Cameroon (n = 37), and South Africa samples (n = 91) were requested, with permission, through the National Health Laboratory Services

(NHLS) within the Division of Medical Virology, Stellenbosch University, South Africa. Samples were collected between March 2017 and February 2018 [23]. We excluded patient samples with no previous cART history and patients receiving first-line cART. Patients had their samples sent to the NHLS for HIV-1 genotypic resistance testing. Treatment failure is defined according to the South Africa adult antiretroviral guidelines by a confirmed viral load of >1000 copies/mL on two measurements taken two to three months apart.

2.3. 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, Hilden, Germany). Briefly, 140 µL of plasma was used as a starting volume. A larger starting volume of 280 µL of plasma was used for samples with very low viral titers. Viral RNA was stored at −80°C until use.

2.4. PCR Amplification and Sequencing

The synthesis of complementary DNA (cDNA) and first-round PCR amplification were performed using the Invitrogen SuperScript® III Reverse Transcriptase (RT) reagents (Invitrogen, Karlsruhe, Germany), as per the manufacturer's instructions. In-house amplification of the IN region (867 bp, positions 4230–5096, HXB2 strain) by nested RT-PCR was performed as previously described by our laboratory [24,25]. Purified 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, Foster City, California, USA), according to manufactures instructions. Primers spanning the full-length integrase (867 bp) were used to sequence the PCR products in both directions. These included sequencing primers Poli6 and Poli7, and additional sequencing primers were used, namely Poli2 (TAAARACARYAGTACWAATGGCA), relative to position 4745–4766, and KLVO83 (GAATACTGCCATTTGTACTGCTG), corresponding to position 4750–4772.

2.5. Consensus Sequence Alignment and Mutation Detection

We performed a search on the HIV Los Alamos National Library (LANL) database (https://www.hiv.lanl.gov/components/sequence/HIV search.com). Our search inclusion criteria included all Cameroonian HIV-1 subtype CRF02_AG IN sequences identified from treatment-naïve patients. We selected one sequence per patient, and all problematic sequences were excluded from further analyses. The consensus sequence representing CRF02_AG was generated using the CRF02_AG study sequences (n = 37) as previously reported [25], accession number: MN816445-MN816488, while the consensus sequence for subtype C was derived from cohort sequences (n = 91,) as previously reported [26]. 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 [27]. As part of quality control, each of the viral sequences were inferred on a phylogenetic tree in order to eliminate possible contamination. The amino acid sequence alignment was extensively screened for the presence of primary and secondary resistance-associated mutations (RAMs) and NOPs associated with resistance to known INSTIs.

2.6. Protein Modelling

A three-dimensional model was constructed for HIV-1C IN and recombinant form CRF02_AG using Schrodinger Prime modelling software [28,29]. A suitable homologous template was identified by performing a Blastp search using the consensus amino acid sequences of HIV1C IN and recombinant form CRF02_AG. Prior to modelling, missing residues were fixed by re-modelling the structure of template 5U1C using Schrodinger PRIME modelling software [30]. The Cryo-EM-solved IN subtype B intasome structure (ID: 5U1C) was used as the homologous template for comparative modelling,

as it shared a high sequence identity and coverage with HIV-1C IN from a South African cohort and with the recombinant form CRF02_AG from a Cameroonian cohort. Additionally, as 5U1C contains a mutation at residue 152, this was mutated back into glutamic acid during re-modelling.

2.7. Protein Preparation

Processing of protein models was performed using Schrodinger Protein Preparation Wizard, which added hydrogen atoms, created disulphide bonds, assigned bond orders, filled in any missing side chains, and optimized the H-Bonds [31]. Magnesium ions were extracted from the prototype foamy virus and simian immunodeficiency virus IN experimental structures.

2.8. Model Validation

To assess the quality of the constructed IN protein models, a variety of structural parameters were tested within each model. The Structural Analysis and Verification Server (SAVES) was used for this purpose and includes the tools Procheck, Whatcheck, Prove, Verify3D, and ERRAT [32–36]. The cut-off values used by the tools were as follows: >80% for Verify3D, <1% for Prove, >50 for ERRAT. Whatcheck and Procheck are further subdivided into tests, but both make use of a Ramachrandran plot analysis, which is deemed passed if the majority of residues are within the allowable region. Furthermore, the root mean square deviation (RMSD) analysis was conducted using PYMOL/Maestro molecular visualizing software to compare backbone structural similarity to the experimentally solved 5U1C template structure [37].

2.9. INSTI Extractions and Molecular Docking

The INSTI's RAL, EVG, DTG, and BIC were extracted from known solved structures deposited inside the Protein Data Bank (PDB) with the respective identifiers 3oya, 3l2u, 3s3m, and 6rwm [38–41]. The structures were superimposed upon our generated IN models using PYMOL and subsequently saved as protein–ligand complexes.

The INSTI CBT has not been published in complex with an IN protein, therefore molecular docking was done to predict the binding mode, affinity, and chemical interactions. Docking was performed using SMINA, a fork of AUTODOCK VINA [42,43]. The three-dimensional (3D) structures for CBT were acquired from the ZINC database [44]. Conversion of receptor and ligand structures from the respective pdb and sdf formats to pdbqt was done using OBABEL [45]. The docking grid was centered on the active site with a box size of 20Å in all planes. CBT was docked to each subtype structural intasome HIV-1C, B and AG, respectively. Ranking of generated binding poses was done based on the binding affinity values calculated using the Vinardo Scoring function with the top binding pose complex selected for the next step [46].

2.10. Refinement and Energy Minimisation

The refinement process involved repeating the aforementioned protein preparation steps. The complexes were subsequently energy-minimized using the CHARMM-GUI webserver [47] for structural preparation and energy minimization using default CHARMM-GUI-generated parameter files. The molecular dynamics engine software utilized was GROMACS 2018.1 [48] and charmm36M forcefield [49].

2.11. Binding Affinity Calculation and Interaction Analysis

The energy-minimized structures were desolvated and the ions removed using PYMOL. The complexes were separated into two separate files, one for IN as the receptor(s) and one for the INSTIs as the ligand(s). To calculate the binding affinity, the score only function was used within SMINA in combination with the Vinardo scoring function. Pymol was used to calculate polar interactions formed between the IN protein, DNA, MG, and the drugs.

2.12. Binding Site Analysis

The spatial and chemical features of the active sites containing the DDE motif of each IN subtype were compared to one another using PYMOL/MAESTRO. Briefly, the residues encompassing the binding site were determined by aligning each IN-INSTI complex to one another and extracting all residues within a 5Å radius of the inhibitor binding site. The binding sites were superimposed to determine the difference by measuring the Root Mean Square Deviation (RMSD) values for the backbone atoms of the protein chains.

3. Results

3.1. Sequence Alignments and Protein Structure Assessment

Two IN consensus protein sequences, corresponding to the South African and Cameroonian cohorts respectively, were aligned to the sequence of the subtype B IN structure (ID:5u1c) (Figure 1). The sequence identity was calculated to be 98% for both sequences when compared with subtype B. The alignment revealed 13 NOPs within the Cameroonian cohort-derived CRF02_AG IN sequences. The identified polymorphisms being K14R, V31I, V72I, L101I, T112V, T124A, T125A, K136T, I151V, V201I, T206S, V234I, S283G. The alignment also showed 17 NOPs within a South African cohort-derived Subtype C consensus sequence, namely D25E, V31I, M50I, V72I, F100Y, L101I, T112V, T124A, T125A, K136Q, I151V, V201I, T218I, V234I, R269K, D278A, S283G (Supplementary Figure S1). The constructed IN models were validated with the following scores obtained with the SAVES server tests. CRF02_AG IN: Verify3D, 71%; ERRAT, 93/100; Prove, 7.8%; and for the Ramachandran plot analysis, 86.1% of residues are within the most favored region. Subtype C IN: Verify3D, 74%; ERRAT, 92/100; Prove, 7.8%; and for the Ramachandran plot analysis, 86.1% of residues are within the most favored region. Subtype B IN: Verify3D, 71%; ERRAT, 92/100; Prove, 10.3%; and for the Ramachandran plot analysis, 82.6% of residues are within most favored region.

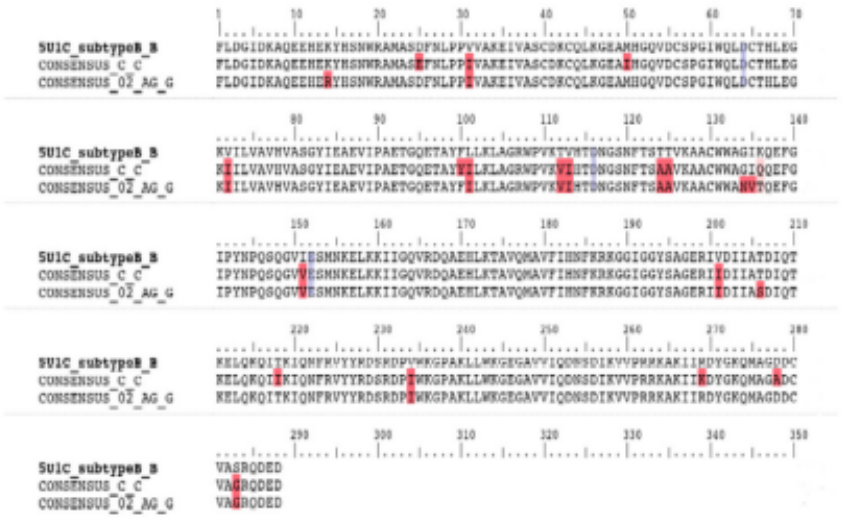

Figure 1. Amino acid sequence alignment of integrase variants subtype B, subtype C.a, and CRF_02_AG, respectively. Red highlighted residues indicate polymorphism locations in comparison with the Subtype B template. Blue highlighted residues indicate the aspartate (D64), aspartate (D116) and glutamate (E152) DDE motif.

The RMSD analysis score was found to be approximately 0.4Å when all three structures were compared to one another. In addition, there was minimal difference in secondary structural makeup (Figure 2). However, reference subtype B structure forms an extra helical turn absent in Subtype C and CRF02_AG.

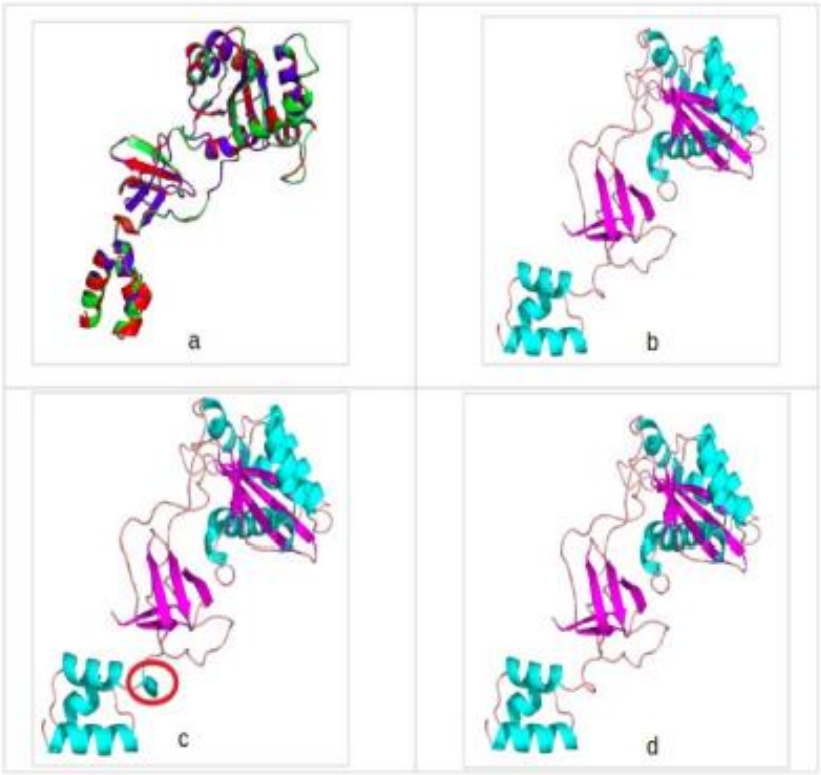

Figure 2. (**a**) A superimposition of all three integrase models showing high backbone identity and secondary structure conservation. (**b**) CRF02_AG integrase (IN) model colored according to secondary structure and in cartoon depiction. (**c**) Subtype B IN model colored according to secondary structure and in cartoon depiction. (**d**) Subtype C IN model colored according to secondary structure and in cartoon depiction. Encircled in red is the secondary structure difference observed between the template structure and the generated IN models. Light blue indicates the helices, purple indicates beta-sheets, and tint color indicates loops.

3.2. *Molecular Docking and Interaction Analysis*

The molecular docking methodology was validated using the genetics algorithm (AutoDock), providing a correlation of the experimental data (half-maximal effective concentration, EC50) with the binding affinities showing reliable statistics (Supplementary Tables S1 and S2, Figure 3) [50]. Subsequently, rescoring/docking assays was done using VINA of the five FDA approved and late-phase clinical trial INSTI's showed minimal differences in binding affinity between one another, with less than 2 kcal/mol difference observed. The obtained binding affinities furthermore were comparable when SMINA was used to rescore Prototype Foamy Virus (PFV)-INSTI or Simian Immunodeficiency (SIV)-INSTI complexes. (Table 1). In Table 2, most of the drugs made two ionic interactions with Magnesium ions, except for EVG and BIC in HIV-1B IN due to the different orientation of the active site as a result of remodelling missing residues.

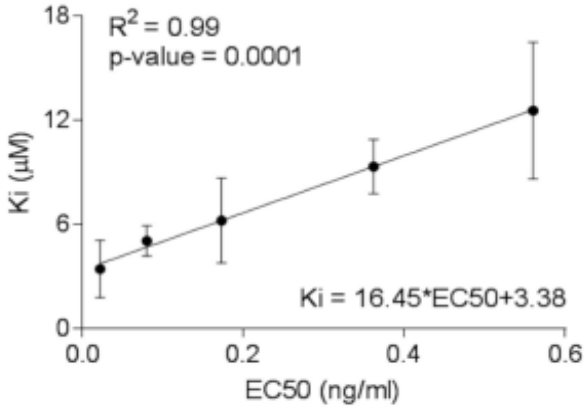

Figure 3. Linear relationship between the predicted binding constants calculated from the docking studies of integrase strand transfer inhibitors (INSTI) with HIV Ins and the experimental half-maximal effective concentration (EC50) values determined from the TZM-bl cells-based assay.

Table 1. Vinardo binding affinity scores predicted for each INSTI bound to the three IN subtypes CRF02_AG, B, and C and the PFV/SIV reported INs.

Drug	CRF02_AG (Kcal/Mol)	Subtype B (Kcal/Mol)	Subtype C (Kcal/Mol)	PFV/SIV (Kcal/Mol)
Raltegravir	−7.2	−4.4	−5.1	−6.1
Elvitegravir	−3.7	−3.4	−3.4	−4.0
Dolutegravir	−3.0	−3.4	−3.0	−5.4
Bictegravir	−4.0	−3.5	−3.5	−3.9
Cabotagravir	−6.5	−5.7	−7.0	N/A

Table 2. Summary of all interactions observed between the five INSTIs and three IN subtypes. Listed also are interactions which occur between INSTI and PFV/SIV nucleic acid.

INSTI	HIV-1B (ID:5u1c)		HIV-1C IN		CRF_02_AG IN		PFV/SIV IN	
	Hydrogen bonds	Ionic contact	Hydrogen bonds	Ionic contact	Hydrogen bonds	Ionic contact	Hydrogen bonds	Ionic contact
RAL	2 (CYS10, GLU152)	2MG	6 (THY11, GUA22, ASP64, ASP116, TYR143, GLU152)	2MG	5 (THY11, GUA22, ASP64, ASP116, GLU152)	2MG	4 (ASP128, ASP185, TYR212, GLU221)	2MG
DTG	5 (THY11, GUA22, ASP64, ASP116, GLU152)	2MG	4 (THY11, GUA22, ASP64, GLU152)	2MG	4 (ADE21, GUA22, ASP64, ASP116)	2MG	3 (ASP128, ASP185, GLU221)	2MG
EVG	5 (THY11, ADE21, GUA22, ASP116, ARG231)	1MG	3 (ADE21, GLU52, ASP64)	2MG	6 (THY11, GUA22, ASP64, ASP116, GLU152, ARG231)	2MG	1 (GLU221)	2MG

Table 2. *Cont.*

INSTI	HIV-1B (ID:5u1c)		HIV-1C IN		CRF_02_AG IN		PFV/SIV IN	
	Hydrogen bonds	Ionic contact	Hydrogen bonds	Ionic contact	Hydrogen bonds	Ionic contact	Hydrogen bonds	Ionic contact
BIC	4 (THY11, GUA22, ASP64, CYS65)	1MG	4 (THY11, GUA22, ASP64, CYS65, ASP116)	2MG	5 (THY11, ASP64, CYS65, GLU92, ASP116)	2MG	3 (ASP64, ASP116, GLU152)	2MG
CBT	7 (THY11, ADE21, GUA22, ASP64, THR66, HIS67, ASP116)	2MG	8 (THY11, ADE21, GUA22, ASP64, THR66, HIS67, ASP116, GLU152)	2MG	6 (THY11, ADE21, GUA22, ASP64, HIS67, ASP116)	2MG	N/A	N/A

Number outside bracket indicates total number of interactions. Supplementary Figures S2–S6. RAL, raltegravir; DTG, dolutegravir; EVG, elvitegravir; BIC, bictegravir; CBT, cabotegravir.

3.3. *Binding Site Analysis*

We calculated differences in total surface area, with 860Å^2 for subtype B IN, 969Å^2 for subtype C IN, and 1041Å^2 for CRF_02_AG IN. RMSD analysis show that subtypes C IN and CRF_02_AG deviate by 0.309Å and 0.44Å, respectively. The NOP I151V was found to occur within the binding site, but does not directly interact with INSTI's. NOP I151V is present in both subtype C and CRF_02_AG IN's.

4. Discussion

The NOPs that have been identified in subtype C and CRF_02_AG IN have not been previously associated with HIV-1 IN drug resistance, except for the polymorphism M50I. M50I was identified in our subtype C IN sequences and this polymorphism has been reported to reduce DTG susceptibility when found in combination with the mutation R263K in HIV-1 subtype B IN [51,52]. However, in our study, the mutation R263K was not present. Furthermore, M50I is not able to cause drug resistance on its own but increases the effect of resistance exhibited by R263K [51]. In our study, M50I had no effect on INSTI's binding to IN. NOPs were found to be occurring within their natural prevalence rates and these NOPs have minimal effect on INSTI susceptibility when occurring alone, such as in the case of M50I [51]. This is in contrast to the study by Brado et al. that reported an impact of NOPs on the stability of the protein complex, suggesting they may contribute to an overall potency against INSTIs. Our study is, however, in agreement with a study by Chitongo et al., 2020 that showed that one known major resistance mutation, G140S, had an effect on DTG drug binding in HIV-1C IN in combination with NOP's. Therefore, NOPs alone have a negligible effect on drug binding [53].

A comparison of the backbone structures of the modelled IN proteins showed high structural similarity with one another with RMSD values less than 0.4Å, suggesting similar fold between the protein structures. Our IN homology modelling showed one slight secondary structural feature alteration within the N-terminal domain. Subtype B displayed a helical turn absent in Subtype C and CRF02_AG IN. Secondary structural features, such as helices, may influence the accessibility of drugs to a protein's active site, either by directly changing the binding pocket properties or through affecting stability of the whole protein [54]. In our study, the structural alteration had no effect on drug binding as it is located further away from the active site.

Extractions showed that all INSTIs are able to bind to either of the tested IN subtype structures with plausible binding poses, which is in agreement to the previously reported PFV or SIV binding poses. No significantly reduced binding affinity was observed for each of the INSTIs, implying no negative alteration to the binding site which may prevent INSTI drug binding. Differences in binding affinity are present. The Binding affinity is an indication of how strong the ligand is binding to the active site. NOPs

associated with a reduction to EVG susceptibility have been previously reported to occur with a relatively high frequency within CRF02_AG IN, however those NOPs were not identified in our study [55–57].

Magnesium ions are responsible and important for the binding of DNA. INSTIs competitively inhibit this process by binding to the magnesium ions [58–60]. It was therefore expected that interaction analysis would reveal that an interaction(s) takes place between the INSTIs and magnesium ions present within IN active sites. The interaction(s) with magnesium ions are considered to be essential for inhibition to take place, whereas other interactions are considered nonessential for the activity of INSTIs to take place. Therefore, we considered an INSTI to be successfully bound if interaction analysis predicted interactions occurring with magnesium ions. It may, however, be plausible that NOPs favor or reduce the likelihood of additional interactions occurring and thereby enhancing binding affinity. In our study, all drugs make contact with MG ions but only 1MG contact is found for HIV-1B IN bound to EVG and BIC. The remodeling of missing residues could be a reason for HIV-1B not making two MG ion contacts with EVG and BIC, but this is a surprising finding that warrants further investigation.

Binding site analysis further supports the results from the rescoring/molecular docking assay. The analyses conducted show that binding sites between each IN structure are identical. One NOP I151V is present within the binding sites of our study, but does not induce any effect beyond a slight spatial change due to its different side chain orientation.

The free energy of binding values calculated between the different proteins and drugs using the Molecular Mechanics Poisson-Boltzmann Surface (MM-PBSA) package confirmed no significant difference in the strength of binding between the drugs and different HIV-1 IN protein subtypes. However, the inhibition constant (Ki) values and inhibitor potency EC50 values provided evidence for DTG and BIC as strong binders with high inhibition values, suggesting DTG and BIC as good candidates for treatment of the three caused by the three different HIV-1 subtypes.

Future work should include viral integration assays to determine if DTG, BIC, and CBT can prevent viral integration within plasmids containing different HIV IN subtype sequences.

5. Conclusions

Our study showed that unique polymorphisms within geographically distinct HIV-1-infected populations with different variants do not prevent INSTI binding. However, linear correlation between the inhibition constant (Ki) values derived from docking experiments and experimental EC50 values for INSTI's suggest DTG and BIC as strong inhibitors. We therefore put forward that second-generation DTG and BIC should be added to antiviral regimens as part of first-line regimens for HIV-1C IN subtype infections, this would account for cross-resistance which may occur between EVG and RAL as DTG and BIC may have a higher genetic barrier to resistance.

Supplementary Materials: The following are available online at http://www.mdpi.com/1999-4915/12/9/936/s1, Table S1: Binding affinities (ΔG in kcal/mol), Table S2: Inhibition constants (Ki in μM) [61], Figure S1: HIV-1C and CRF02_AG integrase mutation profiling, Figure S2: Interaction diagram showing different polar contacts being formed between RAL and different IN subtypes, Figure S3: Interaction diagram showing different polar contacts being formed between EVG and different IN subtypes, Figure S4: Interaction diagram showing different polar contacts being formed between DTG and different IN subtypes, Figure S5: Interaction diagram showing different polar contacts being formed between BIC and different IN subtypes, Figure S6: Interaction diagram showing different polar contacts being formed between CBT and different IN subtypes.

Author Contributions: Conceptualization, R.C.; methodology, D.I., S.G.M., A.E.O. and S.S.; formal analysis, D.I., S.G.M., S.S. and R.C.; investigation, D.I. and S.G.A.; data curation, D.I.; writing—original draft preparation, D.I. and S.G.M.; writing—review and editing, D.I., S.G.M., S.S., G.B.J., R.C., G.M.I., A.E.O. and R.C.; supervision, R.C. and G.B.J.; project administration, R.C.; funding acquisition, R.C. All authors have read and agreed to the published version of the manuscript.

Funding: The National Research Foundation (NRF) of South Africa and the Poliomyelitis Research Foundation (PRF) of South Africa funded this study. Grant support for this study was received from the National Research Foundation (NRF) of South Africa, the Poliomyelitis Research Foundation (PRF) of South Africa, Harry Crossley Foundation and National Health Laboratory Services (NHLS) research trust. Darren Isaacs was funded by the South African Research Chairs Initiative of the Department of Science and Innovation (DSI) and the National Research Foundation (NRF) of South Africa (award number UID 64591).

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

References

1. Piot, P.; Quinn, T.C. Response to the AIDS Pandemic—A Global Health Model. *N. Engl. J. Med.* **2013**, *368*, 2210–2218. [CrossRef] [PubMed]
2. Lesbats, P.; Engelman, A.N.; Cherepanov, P. Retroviral DNA Integration. *Chem. Rev.* **2016**, *116*, 12730–12757. [CrossRef] [PubMed]
3. Craigie, R. The Molecular Biology of HIV Integrase. *Future Virol.* **2012**, *7*, 679–686. [CrossRef] [PubMed]
4. Craigie, R.; Bushman, F.D. HIV DNA Integration. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, a006890. [CrossRef] [PubMed]
5. Grobler, J.A.; Stillmock, K.; Hu, B.; Witmer, M.; Felock, P.; Espeseth, A.S.; Wolfe, A.; Egbertson, M.; Bourgeois, M.; Melamed, J.; et al. Diketo Acid Inhibitor Mechanism and HIV-1 Integrase: Implications for Metal Binding in the Active Site of Phosphotransferase Enzymes. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6661–6666. [CrossRef]
6. Neamati, N.; Lin, Z.; Karki, R.G.; Orr, A.; Cowansage, K.; Strumberg, D.; Pais, G.C.G.; Voigt, J.H.; Nicklaus, M.C.; Winslow, H.E.; et al. Metal-Dependent Inhibition of HIV-1 Integrase. *J. Med. Chem.* **2002**, *45*, 5661–5670. [CrossRef]
7. Ceccherini-Silberstein, F.; Malet, I.; D'Arrigo, R.; Antinori, A.; Marcelin, A.-G.; Perno, C.-F. Characterization and Structural Analysis of HIV-1 Integrase Conservation. *AIDS Rev.* **2009**, *11*, 17–29. [PubMed]
8. Malet, I.; Soulie, C.; Tchertanov, L.; Derache, A.; Amellal, B.; Traore, O.; Simon, A.; Katlama, C.; Mouscadet, J.-F.; Calvez, V.; et al. Structural Effects of Amino Acid Variations between B and CRF02-AG HIV-1 Integrases. *J. Med. Virol.* **2008**, *80*, 754–761. [CrossRef]
9. Anstett, K.; Brenner, B.; Mesplede, T.; Wainberg, M.A. HIV Drug Resistance against Strand Transfer Integrase Inhibitors. *Retrovirology* **2017**, *14*, 36. [CrossRef]
10. World Health Organisation. *Update of Recommendations of First-and Second-Line Antiretroviral Regimens*; World Health Organisation: Geneva, Switzerland, 2019.
11. Rhee, S.-Y.; Grant, P.M.; Tzou, P.L.; Barrow, G.; Harrigan, P.R.; Ioannidis, J.P.A.; Shafer, R.W. A Systematic Review of the Genetic Mechanisms of Dolutegravir Resistance. *J. Antimicrob. Chemother.* **2019**, *74*, 3135–3149. [CrossRef]
12. Santoro, M.M.; Perno, C.F. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiol.* **2013**, *2013*, 1–20. [CrossRef]
13. Keyhani, S.; Wang, S.; Hebert, P.; Carpenter, D.; Anderson, G. US Pharmaceutical Innovation in an International Context. *Am. J. Public Health* **2010**, *100*, 1075–1080. [CrossRef] [PubMed]
14. Bbosa, N.; Kaleebu, P.; Ssemwanga, D. HIV Subtype Diversity Worldwide. *Curr. Opin. HIV AIDS* **2019**, *14*, 153–160. [CrossRef] [PubMed]
15. Bar-Magen, T.; Sloan, R.D.; Faltenbacher, V.H.; Donahue, D.A.; Kuhl, B.D.; Oliveira, M.; Xu, H.; Wainberg, M.A. Comparative Biochemical Analysis of HIV-1 Subtype B and C Integrase Enzymes. *Retrovirology* **2009**, *6*, 103. [CrossRef]
16. Lessells, R.; Katzenstein, D.; de Oliveira, T. Are Subtype Differences Important in HIV Drug Resistance? *Curr. Opin. Virol.* **2012**, *2*, 636–643. [CrossRef]
17. Depatureaux, A.; Quashie, P.K.; Mesplède, T.; Han, Y.; Koubi, H.; Plantier, J.-C.; Oliveira, M.; Moisi, D.; Brenner, B.; Wainberg, M.A. HIV-1 Group O Integrase Displays Lower Enzymatic Efficiency and Higher Susceptibility to Raltegravir than HIV-1 Group M Subtype B Integrase. *Antimicrob. Agents Chemother.* **2014**, *58*, 7141–7150. [CrossRef]
18. Han, Y.-S.; Mesplède, T.; Wainberg, M.A. Differences among HIV-1 Subtypes in Drug Resistance against Integrase Inhibitors. *Infect. Genet. Evol.* **2016**, *46*, 286–291. [CrossRef]
19. Llácer Delicado, T.; Torrecilla, E.; Holguín, Á. Deep Analysis of HIV-1 Natural Variability across HIV-1 Variants at Residues Associated with Integrase Inhibitor (INI) Resistance in INI-Naive Individuals. *J. Antimicrob. Chemother.* **2016**, *71*, 362–366. [CrossRef]
20. 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.* **2019**, *9*, 6227. [CrossRef]

21. Courtney, C.R.; Agyingi, L.; Fokou, A.; Christie, S.; Asaah, B.; Meli, J.; Ngai, J.; Hewlett, I.; Nyambi, P.N. Monitoring HIV-1 Group M Subtypes in Yaoundé, Cameroon Reveals Broad Genetic Diversity and a Novel CRF02_AG/F2 Infection. *AIDS Res. Hum. Retrovir.* **2016**, *32*, 381–385. [CrossRef]
22. Abongwa, L.E.; Nyamache, A.K.; Torimiro, J.N.; Okemo, P.; Charles, F. Human Immunodeficiency Virus Type 1 ((HIV-1) Subtypes in the Northwest Region, Cameroon. *Virol. J.* **2019**, *16*, 103. [CrossRef]
23. Obasa, A.E.; Mikasi, S.G.; Brado, D.; Cloete, R.; Singh, K.; Neogi, U.; Jacobs, G.B. Drug Resistance Mutations Against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living With HIV-1 Receiving Boosted Protease Inhibitors in South Africa. Front. *Microbiology* **2020**, *11*, 438. [CrossRef] [PubMed]
24. Jacobs, G.B.; Laten, A.; van Rensburg, E.J.; Bodem, J.; Weissbrich, B.; Rethwilm, A.; Preiser, W.; Engelbrecht, S. Phylogenetic Diversity and Low Level Antiretroviral Resistance Mutations in HIV Type 1 Treatment-Naive Patients from Cape Town, South Africa. *AIDS Res. Hum. Retrovir.* **2008**, *24*, 1009–1012. [CrossRef] [PubMed]
25. Brado, D.; Obasa, A.E.; Ikomey, G.M.; Cloete, R.; Singh, K.; Engelbrecht, S.; Neogi, U.; Jacobs, G.B. 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.* **2018**, *8*, 4709. [CrossRef] [PubMed]
26. Mikasi, S.G.; Gichana, J.O.; Van der Walt, C.; Brado, D.; Obasa, A.E.; Njenda, D.; Messembe, M.; Lyonga, E.; Assoumou, O.; Cloete, R.; et al. HIV-1 Integrase Diversity and Resistance-Associated Mutations and Polymorphisms Among Integrase Strand Transfer Inhibitor-Naive HIV-1 Patients from Cameroon. *AIDS Res. Hum. Retrovir.* **2020**, *36*, 450–455. [CrossRef]
27. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
28. Jacobson, M.P.; Friesner, R.A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *J. Mol. Biol.* **2002**, *320*, 597–608. [CrossRef]
29. Jacobson, M.P.; Pincus, D.L.; Rapp, C.S.; Day, T.J.F.; Honig, B.; Shaw, D.E.; Friesner, R.A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins* **2004**, *55*, 351–367. [CrossRef] [PubMed]
30. Passos, D.O.; Li, M.; Yang, R.; Rebensburg, S.V.; Ghirlando, R.; Jeon, Y.; Shkriabai, N.; Kvaratskhelia, M.; Craigie, R.; Lyumkis, D. Cryo-EM Structures and Atomic Model of the HIV-1 Strand Transfer Complex Intasome. *Science* **2017**, *355*, 89–92. [CrossRef] [PubMed]
31. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221–234. [CrossRef]
32. Colovos, C.; Yeates, T.O. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci.* **1993**, *2*, 1511–1519. [CrossRef] [PubMed]
33. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst.* **1993**, *26*, 283–291. [CrossRef]
34. Hooft, R.W.W.; Vriend, G.; Sander, C.; Abola, E.E. Errors in Protein Structures. *Nature* **1996**, *381*, 272. [CrossRef] [PubMed]
35. Eisenberg, D.; Lüthy, R.; Bowie, J.U. [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. In *Methods in Enzymology*; Elsevier: Amsterdam, The Netherlands, 1997; Volume 277, pp. 396–404.
36. Pontius, J.; Richelle, J.; Wodak, S.J. Deviations from Standard Atomic Volumes as a Quality Measure for Protein Crystal Structures. *J. Mol. Biol.* **1996**, *264*, 121–136. [CrossRef] [PubMed]
37. *The PyMOL Molecular Graphics System*, version 1.8; Schrodinger LLC.: New York, NY, USA, 2015.
38. Hare, S.; Vos, A.M.; Clayton, R.F.; Thuring, J.W.; Cummings, M.D.; Cherepanov, P. Molecular Mechanisms of Retroviral Integrase Inhibition and the Evolution of Viral Resistance. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20057–20062. [CrossRef]
39. Hare, S.; Gupta, S.S.; Valkov, E.; Engelman, A.; Cherepanov, P. Retroviral Intasome Assembly and Inhibition of DNA Strand Transfer. *Nature* **2010**, *464*, 232–236. [CrossRef]
40. Hare, S.; Smith, S.J.; Métifiot, M.; Jaxa-Chamiec, A.; Pommier, Y.; Hughes, S.H.; Cherepanov, P. Structural and Functional Analyses of the Second-Generation Integrase Strand Transfer Inhibitor Dolutegravir (S/GSK1349572). *Mol. Pharm.* **2011**, *80*, 565–572. [CrossRef]
41. Cook, N.J.; Li, W.; Berta, D.; Badaoui, M.; Ballandras-Colas, A.; Nans, A.; Kotecha, A.; Rosta, E.; Engelman, A.N.; Cherepanov, P. Structural Basis of Second-Generation HIV Integrase Inhibitor Action and Viral Resistance. *Science* **2020**, *367*, 806–810. [CrossRef]

42. Trott, O.; Olson, A.J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef]
43. Koes, D.R.; Baumgartner, M.P.; Camacho, C.J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904. [CrossRef]
44. Irwin, J.J.; Shoichet, B.K. ZINC—A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182. [CrossRef]
45. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**, *3*, 33. [CrossRef] [PubMed]
46. Quiroga, R.; Villarreal, M.A. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLoS ONE* **2016**, *11*, e0155183. [CrossRef] [PubMed]
47. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865. [CrossRef] [PubMed]
48. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef] [PubMed]
49. Huang, J.; MacKerell, A.D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34*, 2135–2145. [CrossRef]
50. Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639–1662. [CrossRef]
51. Wares, M.; Mesplède, T.; Quashie, P.K.; Osman, N.; Han, Y.; Wainberg, M.A. The M50I Polymorphic Substitution in Association with the R263K Mutation in HIV-1 Subtype B Integrase Increases Drug Resistance but Does Not Restore Viral Replicative Fitness. *Retrovirology* **2014**, *11*, 7. [CrossRef]
52. Rogers, L.; Obasa, A.E.; Jacobs, G.B.; Sarafianos, S.G.; Sönnerborg, A.; Neogi, U.; Singh, K. Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. *Front. Microbiol.* **2018**, *9*, 1754. [CrossRef]
53. Chitongo, R.; Obasa, A.E.; Mikasi, S.G.; Jacobs, G.B.; Cloete, R. Molecular Dynamic Simulations to Investigate the Structural Impact of Known Drug Resistance Mutations on HIV-1C Integrase-Dolutegravir Binding. *PLoS ONE* **2020**, *15*, e0223464. [CrossRef]
54. Lahti, J.L.; Tang, G.W.; Capriotti, E.; Liu, T.; Altman, R.B. Bioinformatics and Variability in Drug Response: A Protein Structural Perspective. *J. R. Soc. Interface* **2012**, *9*, 1409–1437. [CrossRef] [PubMed]
55. Kobayashi, M.; Nakahara, K.; Seki, T.; Miki, S.; Kawauchi, S.; Suyama, A.; Wakasamorimoto, C.; Kodama, M.; Endoh, T.; Oosugi, E. Selection of Diverse and Clinically Relevant Integrase Inhibitor-Resistant Human Immunodeficiency Virus Type 1 Mutants. *Antivir. Res.* **2008**, *80*, 213–222. [CrossRef] [PubMed]
56. Hatano, H.; Lampiris, H.; Fransen, S.; Gupta, S.; Huang, W.; Hoh, R.; Martin, J.N.; Lalezari, J.; Bangsberg, D.; Petropoulos, C.; et al. Evolution of Integrase Resistance During Failure of Integrase Inhibitor-Based Antiretroviral Therapy. *J. Acquir. Immune Defic. Syndr.* **2010**, *54*, 389–393. [CrossRef] [PubMed]
57. Abram, M.E.; Ram, R.R.; Margot, N.A.; Barnes, T.L.; White, K.L.; Callebaut, C.; Miller, M.D. Lack of Impact of Pre-Existing T97A HIV-1 Integrase Mutation on Integrase Strand Transfer Inhibitor Resistance and Treatment Outcome. *PLoS ONE* **2017**, *12*, e0172206. [CrossRef]
58. Ribeiro, A.J.M.; Ramos, M.J.; Fernandes, P.A. The Catalytic Mechanism of HIV-1 Integrase for DNA 3'-End Processing Established by QM/MM Calculations. *J. Am. Chem. Soc.* **2012**, *134*, 13436–13447. [CrossRef]
59. Rogolino, D.; Carcelli, M.; Compari, C.; De Luca, L.; Ferro, S.; Fisicaro, E.; Rispoli, G.; Neamati, N.; Debyser, Z.; Christ, F.; et al. Diketoacid Chelating Ligands as Dual Inhibitors of HIV-1 Integration Process. *Eur. J. Med. Chem.* **2014**, *78*, 425–430. [CrossRef]
60. Musyoka, T.; Tastan Bishop, Ö.; Lobb, K.; Moses, V. The Determination of CHARMM Force Field Parameters for the Mg2+ Containing HIV-1 Integrase. *Chem. Phys. Lett.* **2018**, *711*, 1–7. [CrossRef]
61. Hassounah, S.A.; Alikhani, A.; Oliveira, M.; Bharaj, S.; Ibanescu, R.-I.; Osman, N.; Xu, H.-T.; Brenner, B.G.; Mesplède, T.; Wainberg, M.A. Antiviral Activity of Bictegravir and Cabotegravir against Integrase Inhibitor-Resistant SIVmac239 and HIV-1. *Antimicrob. Agents Chemother.* **2017**, *61*, e01695-17. [CrossRef]

Chapter 8: Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding.

8.1. **Journal article**

The article is published in the journal: PLoS ONE 15(5): e0223464. https://doi.org/10.1371/journal. pone.0223464.

8.2. **Author's list**

Chitongo R, Obasa A.E, **Mikasi SG**, Jacobs GB and Cloete R.

8.3. **Authors contribution**

In the enclosed manuscript, I am the co-author. I assisted with sample collection and with laboratory work (RNA extraction, PCR, sequencing) and performed sequence analysis, prior to molecular modelling, also learned major principles of molecular modelling. I also had various interactions (discussions) with the first author (Ms. Chitongo) and the corresponding author (Dr. Cloete) on the analysis and findings.

8.4. **Background**

The emergence of drug RAMs as a result of cART poses a major concern to the control of the HIV-1 pandemic. The INSTI DTG has shown high efficiency in controlling HIV replication and viremia, IN structures have only been partially studied to understand the resistance mechanisms against INSTIs, like DTG. The development of resistance involves the accumulation of concoctions of mutations along the periphery and active site of the enzyme. Mutations that are found in close proximity to the active site decreases the drug binding. Conformational changes of the active site of an enzyme is typically associated with destabilizing of proteins. Mutations that successfully destabilizes the protein can therefore interfere with protein function.

8.5. **Main findings**

The HIV-1 subtype C structure was assessed and found to be reliable with support of 90% confidence in modelled regions. Through molecular modelling and stability predictions, we observed a destabilizing effect of the known G140S mutant on the HIV-1C IN protein structure. Simulation analysis showed that it affected structural stability and flexibility of the protein structure. Our results indicate that the G140S mutant has an effect on INSTI drug binding to HIV-1 subtype C IN structure. These results can also be validated using laboratory - experiments. This study method can also be applied to investigate the effect of other mutations on HIV-1 subtype C INSTIs binding.

8.6. **Study significance**

Our findings are in agreement with the hypothesis that conformational change on the active site of the enzyme caused by the development of resistance mutations; it has a negative effect on the stability of the protein and drug binding

8.7. **Conclusion**

Our findings explain the effects of mutations located close to the active site and underscore the importance of drug binding and protein stability during the presence of mutations.

8.8. **Published article**

PLOS ONE

RESEARCH ARTICLE

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

☯ These authors contributed equally to this work.
* ruben@sanbi.ac.za

Citation: Chitongo R, Obasa AE, Mikasi SG, Jacobs GB, Cloete R (2020) Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding. PLoS ONE 15(5): e0223464. https://doi.org/10.1371/journal.pone.0223464

Editor: Peter J. Bond, Bioinformatics Institute, SINGAPORE

Received: September 18, 2019

Accepted: April 21, 2020

Published: May 7, 2020

Data Availability Statement: Data cannot be shared publicly because of ethical issues. Data are available from the host Institutional Data Access / Ethics Committee (contact via Dr Graeme Jacobs, Senior Lecturer and Research Scientist, Division of Medical Virology, Stellenbosch University, +27 21 938 9744, graeme@sun.ac.za) for researchers who meet the criteria for access to confidential data.

Funding: This study was funded by the Poliomyelitis Research Foundation (PRF) of South

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 Dolutegravir to the catalytic pocket of the protein. A homology model of HIV-1 subtype C IN was successfully constructed and validated. The site directed mutator webserver was used to predict destabilizing and/or stabilizing effects of known RAMs while FoldX confirmed any changes in protein energy upon introduction of mutation. Also, interaction analysis was performed between neighbouring residues. Three mutations known to be associated with Raltegravir, Elvitegravir and Dolutegravir resistance were selected; E92Q, G140S and Y143R, for molecular dynamics simulations. 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 three mutants indicated different effects, while simulation analysis showed G140S to have the largest affect on protein stability and flexibility. This was further supported by weaker non-bonded pairwise interaction energy and binding free energy values between the drug DTG and E92Q, Y143R and G140S mutants suggesting reduced binding affinity, as indicated by interaction analysis in comparison to the WT. Our findings suggest the G140S mutant has the strongest effect on the HIV-1C IN protein structure and Dolutegravir binding. To the best of our knowledge, this is the first study that uses the consensus wild type HIV-1C IN sequence to build an accurate 3D model to understand the effect of three known mutations on DTG drug binding in a South Africa context.

Africa grant no. 19/06 to R. Cloete and in-part by the National Research Foundation (NRF). The Higher Education Department, next Generation of Academic Programme (nGAP) provided support for this study in the form of full time academic positions and salaries to R. Cloete and GBJ. Additionally, the Centre for High Performance Computing (CHPC), Rondebosch, South Africa provided high performance computing cluster.

Competing interests: The authors have declared that no competing interests exist.

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 Asp64, 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-generation INSTIs and Dolutegravir (DTG) and Bictegravir (BIC) are second-generation inhibitors [6]. All first-generation INSTIs have been reported to have relatively low genetic barrier to resistance while second-generation INSTIs including DTG (a coplanar and linear molecule) are associated with a higher genetic barrier against resistance associated mutations (RAMs), and are rendered safe and tolerable, showing little to none drug-drug interactions resistance [7].

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 INSTIs [8].

Genetic resistance pathways including primary mutations at codons Y143C/H/R, Q148H/K/R or N155H together with one or more additional associated secondary mutations at L74M, E92Q, T97A, E138E/A/K or G140S/A, has been reported to result in higher levels of resistance with RAL treatment [9–11]. On its own, the Y143R non-polymorphic mutation reduces RAL susceptibility by ~20-fold, but has no effect on DTG susceptibility [12,13]. On the other hand, EVG specific resistance pathways involve IN mutations T66I/A/K, E92Q/G and S147G pathways [14,15]. The E92Q mutation alone reduces EVG susceptibility to >20-fold and results in limited (<5-fold) cross-resistance to RAL [14–16]. The E92Q mutation is also selected *in vitro* by DTG and reduces DTG susceptibility by ~1.5-fold [11, 17]. With more people living with HIV-1 in resource limited countries still receiving RAL and EVG treatment as first-line antiretroviral ARV therapy, these treatments have suffered an extensive cross-resistance of mutations, highlighting the need for a switch to INSTIs possessing a more robust resistance profile such as DTG. 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 [18]. 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 single and double mutations on HIV-1 IN and drug binding using molecular dynamic simulations [19–23]. Here, molecular dynamics studies have demonstrated the importance of this 140's loop's flexibility as a mechanism of drug resistance [19–23]. However, the findings from some studies were inconclusive due to the poor quality of the protein models delineating the active site and viral DNA binding site for simulation studies.

HIV-1 subtype C (HIV-1C) accounts for nearly 50% of all global HIV-1 infections, while HIV-1 subtype B (HIV-1B) accounts for only approximately 12% [24]. However, a vast majority of research on HIV-1 infections focussed on the development of combination anti-retroviral therapy (cART) drugs for HIV-1B and studying the mechanisms of drug resistance in HIV-1B, with less information known about HIV-1C. As a result, all antiretroviral drugs have been developed in relation to HIV-1B. They have also been reported to be effective against a wide range of HIV-1 subtypes [25]. Other clinical studies have however revealed very poor cART treatment outcome when associated with HIV-1C infections [26–28]. Although HIV-1C has not been considered an effective predictor for therapy failure earlier, a recent trial indicated that HIV-1C has independent predictors for viral failure [28]. Recent studies also have identified subtype specific differences in DTG cross-resistance pattern in patients failing the first-generation RAL treatment [8,14].

Our work carries on from such previously reported molecular dynamics simulation findings to try and assess the molecular mechanisms of resistance in a subtype C IN protein and see how the reported known resistance mutations will affect binding affinity. Based on this insight, this study has been dedicated to employ *in silico* methods to understand whether the second-generation drug, DTG, will be able to retain efficacy against selected RAL and EVG known resistance mutations in an HIV-1C IN protein. In 2017, Cryogenic electron microscopy was used to solve the structure of HIV-1 strand transfer complex intasome for HIV-1 subtype B [29]. 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 known RAL, EVG and DTG 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**) [8]. Nucleotide sequences were verified for stop codons, insertion and deletions using an online quality control program on the HIV-LANL 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 [30]. 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 [31]. 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 [32–34].

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 extracted and obtained from homologous template 3S3M (Prototype foamy virus) onto their specific positions within the predicted HIV-1C IN using PyMOL. The atomic coordinates of the wild-type (WT) structure of HIV-1C in complex with DNA, MG and DTG, were uploaded onto the CHARMM-GUI solution builder webserver to generate a series of input files for energy minimization of the molecule in an aqueous solvent environment [35, 36]. 50,000 steps of energy minimization using the steepest descent minimization integrator was used to equilibrate the system of the solvated complex structure using CHARMM36 force field [37], and applying constraints to hydrogen bonds using the LINCS constraint algorithm. All this was performed with Gromacs software version 5.1 [38]. 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 [39, 40]. We also calculated the loss or gain of polar interactions between neighbouring residues located adjacent to the mutation using PyMOL [36].

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 applying the same parameters used to energy minimize the WT structure [38, 41]. The WT and three mutant systems (E92Q, G140S and Y143R) were prepared by uploading the atomic coordinates of the Protein-DNA-MG-DTG complexes to the CHARMM-GUI interface [36]. These three mutant systems were selected for simulation studies because they represent three resistance pathways associated with RAL, EVG and possibly DTG resistance. Both E92Q and G140S mutations have been reported to reduce susceptibility of DTG ~1.5-fold and up to 10-fold respectively [13,17]. The individual systems were built using the solution builder option in the input generator [35,36]. Each system was solvated in a rectangular TIP3 waterbox with 10Å distance between the edges of the box. The topology and coordinates for each system was generated using the CHARMM36 all-atom force field [37] and CHARMM general force field [37] 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 Y143R system we added 156 K and 81 Cl ions, while for the mutant system G140S we added 157 K and 81 Cl ions and for the mutant system E92Q we added 156 K and 81 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 [38]. Each system underwent 50000 steps of steepest descents energy minimization to remove steric overlap. Afterwards, all the systems were subjected to a two-step equilibration phase namely NVT (constant number of particles, Volume and Temperature) and NPT (constant Number of particles, Pressure and Temperature). The NVT equilibration was run for 500 picoseconds (ps) to stabilize the temperature of the system and a short position restraint NPT was run for 500 ps to stabilize the pressure of the system by relaxing the system and keeping the protein restrained. The V-rescale

temperature-coupling [38,39] method was used for the NVT ensemble, with constant coupling of 1 ps at 303.15K. For NPT, the Nose-Hoover pressure coupling [42–44] was turned on with constant coupling of 1 ps at 303.15K under conditions of position restraints (h-bonds) selecting a random seed. Electrostatic forces were calculated for both NVT and NPT using Particle Mesh Ewald method [45]. All systems were subjected to a full 300 nanoseconds (ns) simulation under conditions of no restraints, an integration time step of 0.002 ps and an xtc collection interval of 5000 steps for 10 ps. Each simulation was repeated (duplicating each simulation separately) to validate reproducibility of results.

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 rmsf**. The radius of gyration was calculated using **gmx gyrate** to determine if the system reached convergence over the 300 ns simulation. Pairwise distance analysis between the drug and MG was done using **gmx pairdist** tool. The total number of hydrogen bonds between the protein and drug was calculated using **gmx hbond**. The total pairwise non-bonded interaction energy (which is not a free energy or binding energy) between the protein and the drug DTG was calculated using gmx energy over 100 ns. The energy terms included were average short range Coulombic interactions and short range Lennard Jones interactions. The free energy of binding was calculated using Molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) protocols implemented in the g_mmpbsa package [46]. The ΔG binding energy was calculated for 800 frames between the protein and the drug over 8 ns from 100–108 ns of the simulation trajectory with a sampling interval of 10 ps. 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. The technique is known for its ability to transform a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables, called principal components (PCs), while retaining those characteristics of the data set that contribute most to its variance [19]. 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 (S1 Fig). The tetrameric protein structure built for HIV-1C IN had a Global mean quality estimate score of 0.59 and 60% sequence similarity to 5U1C (S2 Fig). The active site residues, MG and DNA nucleotides involved in DTG binding to HIV-1C IN are shown in Fig 1. The VERIFY 3D score was 80.1%, the ERRAT overall quality score was

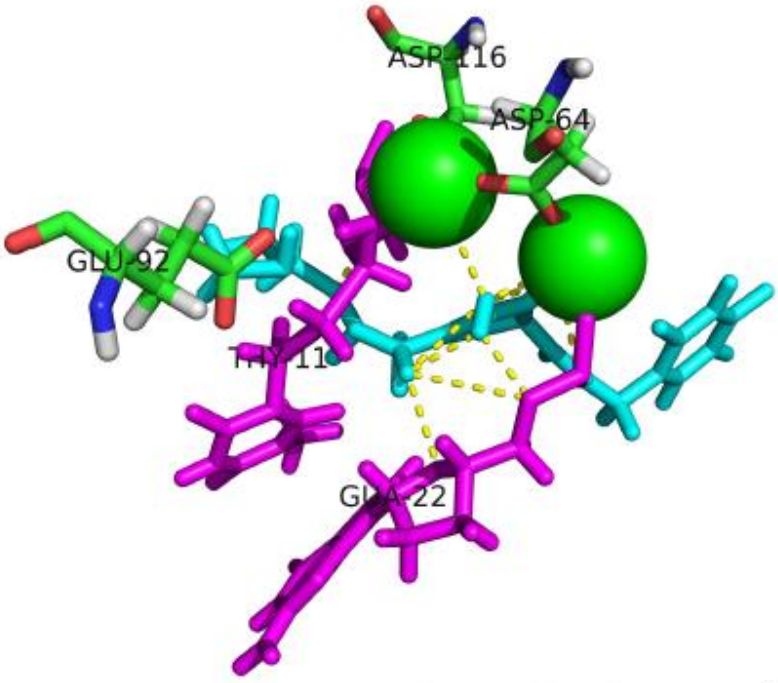

Fig 1. HIV-1C Integrase active site showing interactions with DNA, MG and drug Dolutegravir. Magnesium $^{2+}$ ions (green spheres) are shown sitting in close proximity with Dolutegravir (cyan) within the binding pocket (DDE motif) residues are labelled and shown as sticks. Two DNA residues: THY11 and GUA22 (magenta sticks) are shown expressing polar interactions with Dolutegravir and the drug also interacts with both MG ions as shown. Dashed yellow lines show polar contacts.

https://doi.org/10.1371/journal.pone.0223464.g001

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 G140S was stabilizing, E92Q destabilizing and Y143R 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).

Molecular dynamic simulations

All the MD trajectory analysis considered the single chain A (monomer) of the IN protein in contact with the drug DTG and the DNA. Trajectory analysis of the RMSD of the backbone indicated that the WT system reached equilibrium after 100 ns as well as the E92Q, Y143R and G140S mutant systems (Fig 2A). Only G140S showed higher RMSD variation values compared

Table 1. Summary of stability predictions and polar interactions.

	SDM[1]	FoldX[2]		Polar Interactions	
Mutation	Predicted ΔG (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.

https://doi.org/10.1371/journal.pone.0223464.t001

to the WT, Y143R and E92Q systems (Fig 2A). RMSF analysis clearly showed higher flexibility for the G140S mutant system, with four highly flexible regions (residues 68–70, 142–146, 166–170 and 253–256) compared to the WT, E92Q and Y143R 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 Y143R and E92Q compared to the WT and G140S 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 vs E92Q, WT vs Y143R and WT vs G140S systems are shown in S3A, S3B and S3C Fig. Calculation of the covariance matrix values after diagonalization showed a significant increase for the G140S system (18.33 nm) compared to the other three systems WT, E92Q and Y143R 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, Y143R MG ion and drug DTG systems compared to E92Q and G140S each having a distance of 0.41 ± 0.04 nm, 0.99 ± 0.20 nm, respectively. The average number of hydrogen bonds formed between the protein-DNA-MG and drug were calculated to be 1.58, 0.34, 0.20 and 0.54 for the WT, E92Q, G140S and Y143R, respectively (S4A–S4D Fig). The repeats of the simulation runs showed similar results for each of the first simulation runs (S5A and S5B Fig). The RMSD showed equilibrium after 100 ns for the WT system while the G140S mutant system showed increasing RMSD values comparable to the first run (S5A Fig).

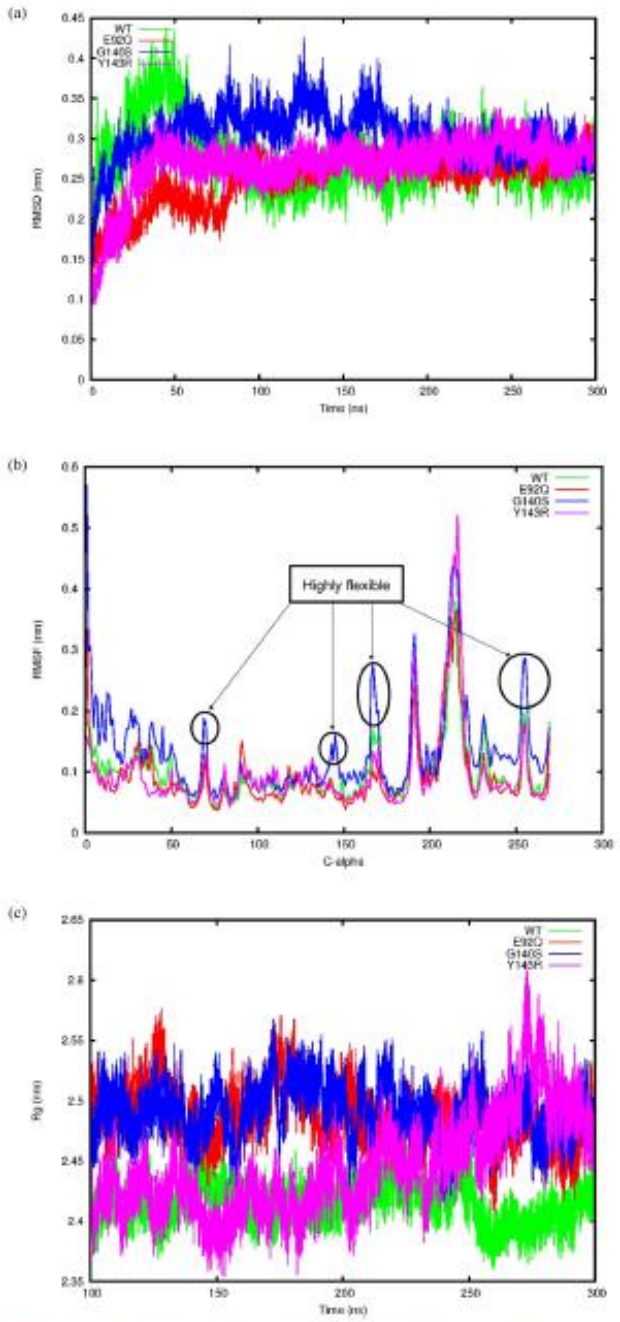

Fig 2. Trajectory analysis of the four simulation systems. (A) Change in backbone RMSD for the WT, E92Q, G140S and Y143R systems plotted over 300 ns. (B) Change in RMSF for the C-alpha residues for the WT, E92Q, G140S and Y143R systems plotted over the last 200 ns. (C) Measure of compactness for the WT, E92Q, G140S and Y143R systems plotted over the last 200 ns.

https://doi.org/10.1371/journal.pone.0223464.g002

Table 2. Binding free energies of DTG to various Protein complexes using MMPBSA method.

Energy (Kcal/Mol)	WT	E92Q	Y143R	G140S
ΔE_{vdW}	-43.88 ± 22.54	-23.28 ± 9.08	-21.07 ± 10.13	-14.88 ± 12.22
ΔE_{elec}	-1.12 ± 11.27	-2.10 ± 5.07	10.23 ± 6.64	-5.00 ± 9.11
ΔE_{pcal}	20.90 ± 15.84	8.20 ± 9.01	-8.52 ± 11.33	-0.31 ± 19.84
ΔE_{SASA}	-5.55 ± 2.43	-3.47 ± 1.33	-3.84 ± 1.95	-1.73 ± 2.04
ΔG_{bind}	-29.65 ± 18.54	-20.65 ± 9.36	-23.20 ± 10.52	-21.93 ± 23.11

ΔE_{vdW}: van der Waals energy, ΔE_{elec}: Electrostatic interaction energy, ΔE_{pcal}: polar solvation energy, ΔE_{SASA}: Solvent accessible surface area energy, ΔG_{bind}: Total binding energy.

https://doi.org/10.1371/journal.pone.0223464.t002

Additionally, the radius of gyration value was lower for the WT compared to the mutants suggesting the WT structure is more compact compared to the mutant structures, again similar to the first run (S5B Fig). The total non-bonded pairwise interaction energies between the HIV-1C IN protein and DTG were found to be higher for the WT (-94.54 ± 13.20) compared to the three mutant structures (E92Q, Y143R and G140S) each having, -38.74 ± 6.70, -27.97 ± 2.37 and -16.49 ± 1.02 KJ/Mol, respectively. To understand the binding free energy of DTG to the WT and the mutant HIV-1C IN protein structures we performed MMPBSA binding energy calculations. Table 2 summarises the different contributors to the binding free energy. The highest total binding free energy was observed for the WT (-29.65 ± 18.54 Kcal/Mol) followed by the Y143R mutant system (-23.20 ± 10.52 Kcal/Mol) and G140S mutant system (-21.93 ± 23.11 Kcal/Mol), while the E92Q mutant system showed the weakest binding free energy of (-20.65 ± 9.36 Kcal/Mol) (Table 2). The major contributors to the total binding free energy in the WT was the van der Waals energy, electrostatic interaction energy and SASA energy, while the polar solvation had no contribution (because of its positive value) to the binding of the drug and similarly for the E92Q mutant (Table 2). For the Y143R system the van der Waals, polar solvattion and SASA energy were the major contributors to the binding free energy, while the electrostatic interaction energy had no contribution. Both the van der Waals and the electrostatic interaction energies contributed significantly to the total binding energy observed in the G140S mutant, while the polar solvation and SASA energy each had smaller contributions to the binding of the drug. This suggests that all the mutants considered may trigger conformational changes in the active site resulting in significantly weak binding of DTG to HIV-1C IN.

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 the drug DTG and known active site residues D64, D116 and N148, MG ion and also to DNA nucleotides (Table 3). Similarly, interactions were observed between the drug DTG and known active site residue D64, Y143R, N148, MG ion and DNA nucleotides for the Y143R system (Table 3). Interestingly, no active site residue and MG ionic interactions were formed between DTG and the G140S mutant system, resulting in the dissociation of the drug from the binding pocket over time (S6D Fig). On the other hand, the E92Q system showed interactions with one of the active site residues (D116) but no MG ionic interactions (Table 3). S6A–S6D Fig in Supporting Information File shows the different interactions formed between the drug, MG ion and active site residues for snapshots taken at 100 ns for each simulation system.

Table 3. Summary of interaction analysis.

Structure	Cluster	Interactions	
		Hydrogen bonds	Ionic
WT	1 (100 ns)	2 (GUA22[a], D116)	MG
	2 (150 ns)	3 (THY11[a], D64, D116)	MG
	3 (200 ns)	2 (GUA22[a], D116)	MG
	4 (250 ns)	4 (THY11[a], GUA22[a], D64, D116)	MG
	5 (300 ns)	2 (THY11[a], N148)	MG
Y143R	1 (100 ns)	4 (THY11[a], GUA22[a], D64, N148)	MG
	2 (150 ns)	4 (GUA22[a], D64, R143, N148)	MG
	3 (200 ns)	4 (GUA22[a], D64, R143, N148)	MG
	4 (250 ns)	4 (GUA22[a], D64, R143, N148)	MG
	5 (300 ns)	5 (THY11[a]; GUA22[a], GUA22[a], D64, N148)	MG
E92Q	1 (100 ns)	3 (CYT20[a], D116, P145)	None
	2 (150 ns)	3 (D116, P145, E152)	None
	3 (200 ns)	3 (CYT20[a], H21, D116)	None
	4 (250 ns)	4 (CYT20[a], P142, P145, E152)	None
	5 (300 ns)	3 (CYT20[a], P145, N148)	None
G140S	1 (100 ns)	3 (GUA22[a], ADE25[a], ADE27[a])	None
	2 (150 ns)	3 (GUA22[a], ADE25[a], ADE27[a])	None
	3 (200 ns)	3 (GUA22[a], ADE25[a], ADE27[a])	None
	4 (250 ns)	3 (THY11[a], GUA22[a], ADE27[a])	None
	5 (300 ns)	3 (THY11[a], GUA22[a], ADE27[a])	None

[a]Interactions with DNA nucleotide residues.

Abbreviations of DNA nucleotides: ADE-Adenine; CYT-Cytosine; GUA-Guanine; THY-Thymine. Abbreviations of amino acids: D-Aspartic Acid; E-Glutamic Acid; H-Histidine N-Asparagine; P-Proline; R-Arginine.

https://doi.org/10.1371/journal.pone.0223464.t003

Discussion

Previous studies by Chen *et al.* and Dewdney *et al.* [19, 21] showed the structural impact of mutations Q148H/R and G140S/A on the flexibility of the HIV-1 IN as a mechanism for RAL resistance. Furthermore, Xue and team [46] found that the cross-resistance mutation E138K/Q148K resulted in a reduction in the chelation ability of oxygen atoms in INSTIs to Mg^{2+} in the active site of the mutated intasomes resulting in a reduced binding affinity of RAL and EVG to the protein. Another simulation study also revealed the binding mode of EVG and RAL to HIV-1 IN and the structural mechanism of drug resistant mutants (G140A and G149A) that affect the 140's loop region spanning residues 140–149 [23]. However, all of these studies only considered HIV-1B IN and protein models of low sequence identity. In this study, we selected three known mutations E92Q, G140S and Y143R associated with RAL, EVG and DTG resistance to investigate their effect on the protein structure of HIV-1C IN and DTG 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 three known 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 [47]. 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, E92Q and Y143R systems. These findings are contradictory to Chen et al. [19] who performed 150 ns simulation studies of the G140S HIV-1B IN mutant system with NAMD and discovered that the 140's loop of the single G140S mutant system displayed reduced movements using principal component analysis. Their results showed that the single G140S mutation did not adversely affect drug binding. In our case, the 140's loop region is stabilized by the G140S mutation and we assume that could reduce 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 Y143R. Furthermore, the total pairwise non-bonded interaction energy was significantly lower for the G140S mutant compared to the WT, suggesting weaker affinity of the drug DTG for HIV-1C IN in the presence of the mutant. Similarly, the binding free energy calculations also showed higher binding energy between the WT HIV-1C IN and DTG and reduced binding for the E92Q, Y143R and G140S mutant systems. These results are in stark contrast to the study of Chen et al. [19] that showed no difference in binding affinity of RAL to the WT and G140S single mutant. Interestingly, the binding free energy in our study for the WT and DTG (-29.65 ± 18.54) was comparable to that found in the Xue *et al.* [22] study (-30.95 ± 0.10), although having ~1.3 Kcal/Mol energy difference. Further interaction analysis was performed to confirm the hypothesis that the G140S mutation could reduce drug binding by extracting structures at different snapshots of the simulation. Here, we found that the G140S mutation resulted in the drug moving further away from the binding pocket. We also observed weaker interactions for the E92Q mutation but stronger interactions for Y143R mutant based on the average number of hydrogen bonds and the total number of polar contacts between the protein and the drug. 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 and EVG mutants and not considering novel RAL or DTG mutations and also simulating single instead of double mutations. However, we have yet to identify double mutants within the South African cohort of HIV-1C infected patients. Another limitation is the exclusion of entropy effects due to the lack of computational resources this might have led to under or overestimation of the binding free energy. However, our total pairwise interaction energies also correlate well with RAL binding energies observed in the Chen et al. [19] study with the WT showing higher pair interaction energy compared to the G140S/Q148H double mutant. Future work will include viral fitness assays to determine the effect of mutants E92Q, Y143R and G140S on the HIV-1C virus replication in the presence of DTG.

Supporting information

S1 Fig. Pairwise amino acid sequence alignment between HIV-1C consensus and HIV-1B (PDBID: 5U1C). The conserved DDE motif residues (D64, D116 and E152) are shown in black boxes.
(TIFF)

S2 Fig. Tetrameric 3D structure of HIV-1C Integrase in complex with DNA, MG and drug Dolutegravir. Magnesium^{2+} ions (dirty violet spheres), Dolutegravir (brown), DDE motif residues of the protein represented as navy blue sticks and the DNA as a ladder. Each chain/monomer of the protein is labelled and coloured differently.
(TIFF)

S3 Fig. PCA analysis for the first two principal components. (A) Graphical representation of PCA of WT vs E92Q systems plotted over the last 200 ns, (B) Graphical representation of PCA of WT vs G140S systems plotted over the last 200 ns and (C) Graphical representation of PCA of WT vs Y143R systems plotted over the last 200 ns.
(TIFF)

S4 Fig. The average number of hydrogen bonds formed between the HIV-1C IN protein-DNA-MG and DTG. A) WT, B) E92Q, C) G140S and D) Y143R.
(TIFF)

S5 Fig. Trajectory analysis of the repeat of the four simulation systems. A) RMSD backbone deviation of the four HIV1C IN protein simulations and B) The change in Raduis of gyration values for the backbone atoms of the four HIV1C IN protein simulations.
(TIFF)

S6 Fig. Interaction analysis for the four simulation systems. (A) Interactions formed between WT HIV-1C integrase structure and DTG taken at 100 ns. (B) Interactions formed between Y143R HIV-1C integrase structure and DTG taken at 100 ns. (C) Interactions formed between E92Q HIV-1C integrase structure and DTG taken at 100 ns. (D) Interactions formed between G140S HIV-1C integrase structure and DTG taken at 100 ns.
(TIFF)

Author Contributions

Conceptualization: Graeme Brendon Jacobs, Ruben Cloete.

Data curation: Rumbidzai Chitongo.

Formal analysis: Ruben Cloete.

Methodology: Ruben Cloete.

Supervision: Graeme Brendon Jacobs, Ruben Cloete.

Visualization: Rumbidzai Chitongo.

Writing – original draft: Rumbidzai Chitongo, Adetayo Emmanuel Obasa.

Writing – review & editing: Adetayo Emmanuel Obasa, Sello Given Mikasi, Graeme Brendon Jacobs, Ruben Cloete.

References

1. Pommier Y, Johnson AA, Marchand C. Integrase inhibitors to treat HIV/AIDS. Nat Rev Drug Discov. 2005; 4(3):236–48. https://doi.org/10.1038/nrd1660 PMID: 15729361
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. https://doi.org/10.1073/pnas.93.24.13659 PMID: 8942990

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. https://doi.org/10.1021/jm800245z PMID: 18763751
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. https://doi.org/10.1124/mol.64.3.600 PMID: 12920196
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. https://doi.org/10.1517/13543784.2011.562189 PMID: 21381981
7. Di Santo R. Inhibiting the HIV integration process: Past, present, and the future. J Med Chem. 2014; 57(3):539–66. https://doi.org/10.1021/jm400674a PMID: 24025027
8. 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.
9. Hu Z, Kuritzkes DR. Effect of raltegravir resistance mutations in HIV-1 integrase on viral fitness. J Acquir Immune Defic Syndr. 2010 Oct 1; 55(2):148–55. https://doi.org/10.1097/QAI.0b013e3181e9a87a PMID: 20634701
10. 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. https://doi.org/10.1093/infdis/jir025 PMID: 21459813
11. Garrido C, Villacian J, Zahonero N, Pattery T, Garcia F, Gutierrez F, et al. Broad phenotypic cross-resistance to elvitegravir in HIV-infected patients failing on raltegravir-containing regimens. Antimicrob Agents Chemother. 2012 Jun; 56(6):2873–8. https://doi.org/10.1128/AAC.06170-11 PMID: 22450969
12. Kobayashi M, Yoshinaga T, Seki T, Wakasa-Morimoto C, Brown KW, Ferris R, et al. In vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. Antimicrob Agents Chemother. 2011 Feb; 55(2):813–21. https://doi.org/10.1128/AAC.01209-10 PMID: 21115794
13. Cahn P, Pozniak AL, Mingrone H, Shuldyakov A, Brites C, Andrade-Villanueva JF, 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. 2013; 382(9893):700–8. https://doi.org/10.1016/S0140-6736(13)61221-0 PMID: 23830355
14. Doyle T, Dunn DT, Ceccherini-Silberstein F, De Mendoza C, Garcia F, Smit E, et al. Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades. J Antimicrob Chemother [Internet]. 2015 Nov [cited 2019 Nov 22]; 70(11):3080–6. Available from: https://academic.oup.com/jac/article-lookup/doi/10.1093/jac/dkv243 PMID: 26311843
15. Seki T, Suyama-Kagitani A, Kawauchi-Miki S, Miki S, Wakasa-Morimoto C, Akihisa E, et al. Effects of raltegravir or elvitegravir resistance signature mutations on the barrier to dolutegravir resistance in vitro. Antimicrob Agents Chemother. 2015 May 1; 59(5):2596–606. https://doi.org/10.1128/AAC.04844-14 PMID: 25691633
16. Brenner BG, Wainberg MA. Clinical benefit of dolutegravir in HIV-1 management related to the high genetic barrier to drug resistance. Vol. 239, Virus Research. Elsevier B.V.; 2017. p. 1–9.
17. Abram ME, Hluhanich RM, Goodman DD, Andreatta KN, Margot NA, Ye L, et al. Impact of primary elvitegravir resistance-associated mutations in HIV-1 integrase on drug susceptibility and viral replication fitness. Antimicrob Agents Chemother. 2013 Jun; 57(6):2654–63. https://doi.org/10.1128/AAC.02568-12 PMID: 23529738
18. Vermeulen M, Lelie N, Coleman C, Sykes W, Jacobs G, Swanevelder R, et al. Assessment of HIV transfusion transmission risk in South Africa: a 10-year analysis following implementation of individual donation nucleic acid amplification technology testing and donor demographics eligibility changes. Transfusion. 2019; 59(1):267–76. https://doi.org/10.1111/trf.14959 PMID: 30265757
19. Chen Q, Buolamwini JK, Smith JC, Li A, Xu Q, Cheng X, et al. Impact of Resistance Mutations on Inhibitor Binding to HIV-1 Integrase. J Chem Inf Model [Internet]. 2013 Dec 23 [cited 2019 Nov 22]; 53(12):3297–307. Available from: https://pubs.acs.org/doi/10.1021/ci400537n PMID: 24205814
20. Chen Q, Cheng X, Wei D, Xu Q. Molecular dynamics simulation studies of the wild type and E92Q/N155H mutant of Elvitegravir-resistance HIV-1 integrase. Interdiscip Sci Comput Life Sci. 2015; 7(1):36–42.
21. 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. https://doi.org/10.1016/j.drup.2010.05.001 PMID: 20570551

22. Wilkinson AL, El-Hayek C, Spelman T, Fairley C, Leslie D, McBryde E, et al. "Seek, Test, Treat" Lessons From Australia. J AIDS J Acquir Immune Defic Syndr [Internet]. 2015 Aug [cited 2019 Nov 22]; 69 (4):460–5. Available from: https://insights.ovid.com/crossref?an=00126334-201508010-00012

23. Geretti AM, Harrison L, Green H, Sabin C, Hill T, Fearnhill E, et al. Effect of HIV-1 Subtype on Virologic and Immunologic Response to Starting Highly Active Antiretroviral Therapy. Clin Infect Dis [Internet]. 2009 May [cited 2019 Nov 22]; 48(9):1296–305. Available from: https://academic.oup.com/cid/article-lookup/doi/10.1086/598502 PMID: 19331585

24. Brenner BG, Thomas R, Blanco JL, Ibanescu R-I, Oliveira M, Mesplède T, et al. Development of a G118R mutation in HIV-1 integrase following a switch to dolutegravir monotherapy leading to cross-resistance to integrase inhibitors. J Antimicrob Chemother [Internet]. 2016 Jul [cited 2019 Nov 22]; 71 (7):1948–53. Available from: https://academic.oup.com/jac/article-lookup/doi/10.1093/jac/dkw071 PMID: 27029845

25. 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. 2016 Apr 1; 3(4):e166–74. https://doi.org/10.1016/S2352-3018(16)00023-0 PMID: 27036992

26. Sutherland KA, Collier DA, Claiborne DT, Prince JL, Deymier MJ, Goldstein RA, 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 Nov 30; 6.

27. Passos DO, Li M, Yang R, Rebensburg SV., 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.

28. 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.

29. 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. https://doi.org/10.1093/bioinformatics/btm404 PMID: 17846036

30. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci. 1993; 2(9):1511–9. https://doi.org/10.1002/pro.5560020916 PMID: 8401235

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

32. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem [Internet]. 2008 Aug [cited 2019 Nov 22]; 29(11):1859–65. Available from: http://doi.wiley.com/10.1002/jcc.20945 PMID: 18351591

33. 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 [Internet]. 2016 Jan 12 [cited 2019 Nov 22]; 12(1):405–13. Available from: https://pubs.acs.org/doi/10.1021/acs.jctc.5b00935 PMID: 26631602

34. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. J Comput Chem [Internet]. 2013 Sep 30 [cited 2019 Nov 22]; 34(25):2135–45. Available from: http://doi.wiley.com/10.1002/jcc.23354 PMID: 23832629

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

36. 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

37. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem [Internet]. 2009 [cited 2019 Nov 22];NA–NA. Available from: http://doi.wiley.com/10.1002/jcc.21367

38. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys [Internet]. 2007 Jan 7 [cited 2020 Feb 18]; 126(1):014101. Available from: http://aip.scitation.org/doi/10.1063/1.2408420 PMID: 17212484

39. Wong-ekkabut J, Karttunen M. Assessment of Common Simulation Protocols for Simulations of Nanopores, Membrane Proteins, and Channels. J Chem Theory Comput [Internet]. 2012 Aug 14 [cited 2020 Feb 18]; 8(8):2905–11. Available from: https://pubs.acs.org/doi/10.1021/ct3001359 PMID: 26592129

40. Nosé S. A molecular dynamics method for simulations in the canonical ensemble. Mol Phys. 1984; 52 (2):255–68.

41. Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A. 1985; 31 (3):1695–7.

42. Evans DJ, Holian BL. The Nose-Hoover thermostat. J Chem Phys. 1985; 83(8):4069–74.

43. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Acc Chem Res [Internet]. 2000 Dec [cited 2019 Nov 22]; 33(12):889–97. Available from: https://pubs.acs.org/doi/10.1021/ar000033j PMID: 11123888

44. Wang J, Wang W, Kollman PA, Case DA. Antechamber, An Accessory Software Package For Molecular Mechanical Calculations Correspondence to. Journal of Chemical Information and Computer Sciences.

45. Dewdney TG, Wang Y, Kovari IA, Reiter SJ, Kovari LC. Reduced HIV-1 integrase flexibility as a mechanism for raltegravir resistance. J Struct Biol. 2013 Nov; 184(2):245–50. https://doi.org/10.1016/j.jsb.2013.07.008 PMID: 23891838

46. Xue W, Jin X, Ning L, Wang M, Liu H, Yao X. Exploring the Molecular Mechanism of Cross-Resistance to HIV-1 Integrase Strand Transfer Inhibitors by Molecular Dynamics Simulation and Residue Interaction Network Analysis. J Chem Inf Model [Internet]. 2013 Jan 28 [cited 2019 Nov 22]; 53(1):210–22. Available from: https://pubs.acs.org/doi/10.1021/ci300541c PMID: 23231029

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

Chapter 9: Summary of the study: South African Medical Journal (SAMJ).

HIV-1 diversity and the implementation of integrase strand-transfer inhibitors as part of combination antiretroviral therapy

To the Editor: The integrase (IN) strand-transfer inhibitor (InSTI) dolutegravir (DTG) is now recommended by the World Health Organization as part of salvage and/or first-line combination antiretroviral therapy (cART).[1] DTG has a high genetic barrier against developing resistance and is effective against all strains that previously exhibited resistance-associated mutations (RAMs) against other cART regimens.[2] Recommendations to use DTG were delayed owing to preliminary findings from Botswana that indicated potential safety concerns in pregnancy, with a small increased risk of neural tube defects.[3] Studies that investigated the safety and efficacy of DTG now support its use in all populations, including pregnant women and those of childbearing potential.[4,5]

HIV-1 genetic diversity continues to make it difficult to control the pandemic. New subtypes are still being identified, with HIV-1 subtype L only being described and characterised in 2019.[6] It is well known that HIV-1 diversity remains a key challenge pertaining to a wide spectrum of fields, such as serological diagnoses, virological follow-up, vaccine development and therapeutic monitoring. Although HIV-1 subtype C is prevalent in southern Africa, the majority of the HIV-1 groups and subtypes, including circulating recombinant forms (CRFs), can be found in Africa.[7] Some mutation pathways clearly differ by subtype variation. For example, a study by Doyle *et al.*[8] comparing major IN RAMs in raltegravir (RAL) recipients at positions 148 and 140 of IN between subtype B and non-B clades found that these mutations were exclusively present in subtype B sequences. The G118R InSTI mutation was only found among individuals infected with HIV-1 subtype C and CRF02_AG. This mutation is rarely present in HIV-1 subtype B.[9] It has been postulated 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.[10] Of note, the majority of group O viruses are naturally resistant to non-nucleoside reverse transcriptase inhibitors owing to the presence of the C181Y mutation in the reverse transcriptase gene.[8] In our studies, we observed low-level RAMs against InSTIs.[9,11,12] The effect of these mutations is yet to be fully understood. Through our structural modelling and docking studies, we observed differences of InSTIs drug-binding interactions to different HIV-1 IN subtypes, but we did not observe any significant differences in binding affinity for each InSTI.[13-15] This finding implies no significant alteration to the binding site in the wild-type IN, which may consequently prevent InSTI drug binding. By using triple therapy, the impact of developing clinical resistance should be limited if patients remain fully adherent. In cases where it is suspected that cART failure is due to resistance development, resistance testing should be done before patients are switched to a new regimen. We do support the full-scale use of DTG in African settings where diverse subtypes are prevalent. Continued close monitoring strategies to ensure a successful switch of regimens is warranted in patients with virological failure and who have developed resistance to their cART regimens.

Sello Given Mikasi
Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
mikasi@sun.ac.za

George Mondinde Ikomey
Centre for the Study and Control of Communicable Diseases, University of Yaoundé 1, Cameroon

Adetayo Emmanuel Obasa
Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

Ruben Cloete
South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa

Graeme Brendon Jacobs
Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
graeme@sun.ac.za

1. World Health Organization. Policy brief: Update of recommendations on first- and second-line antiretroviral regimens (WHO/CDS/HIV/19.15). 2019. https://apps.who.int/iris/handle/10665/325892 (accessed 10 August 2020).
2. Mesplède T, Wainberg MA, Fun A, Nijhuis M, Brenner B, Vandekerckhove L. Clinical use of HIV integrase inhibitors: A systematic review and meta-analysis. PLoS ONE 2013;8(1):e52662. https://doi.org/10.1371/journal.pone.0052662
3. Zash R, Makhema J, Shapiro RL. Neural-tube defects with dolutegravir treatment from the time of conception. N Engl J Med 2018;379(10):979-981. https://doi.org/10.1056/NEJMc1807653
4. Zash R, Holmes L, Diseko M, et al. Neural-tube defects and antiretroviral treatment regimens in Botswana. N Engl J Med 2019;381(9):827-840. https://doi.org/10.1056/NEJMoa1905230
5. Pereira G, Kim A, Safi E, et al.; National Cohort Study of Dolutegravir and Pregnancy Outcomes in Brazil. No occurrences of neural tube defects among 382 women on dolutegravir at pregnancy conception in Brazil. Presented at IAS 2019, Mexico City, 21 - 24 July 2019. http://programme.ias2019.org/Abstract/Abstract/49991 (accessed 23 April 2020).
6. Yamaguchi J, McArthur C, Vallari A, et al. Complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. J Acquir Immune Defic Syndr 2020;83(3):319-322. https://doi.org/10.1097/QAI.0000000000002246
7. Mikasi SG, Gichana JG, van der Walt C, et al. HIV-1 integrase diversity and resistance-associated mutations and polymorphisms among integrase strand transfer inhibitor-naive HIV-1 patients from Cameroon. AIDS Res Hum Retroviruses 2020;36(5):450-455. https://doi.org/10.1089/AID.2019.0264
8. Doyle T, Dunn D, Ceccherini-Silberstein F, et al. Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades. J Antimicrob Chemother 2015;70(11):3080-3086. https://doi.org/10.1093/jac/dkv243
9. Quashie PK, Oliveira M, Veres T, et al. Differential effects of the G118R, H51Y, and E138K resistance substitutions in different subtypes of HIV integrase. J Virol 2015;89(6):3163-3175. https://doi.org/10.1128/JVI.03363-14
10. Trono D, van Lint C, Rouzioux C, et al. HIV persistence and the prospect of long-term drug-free remissions for HIV-infected individuals. Science 2010;329(5988):174-180. https://doi.org/10.1126/science.1191047
11. Obasa AE, Mikasi SG, Brado D, et al. Drug-resistance mutations against protease, reverse transcriptase and integrase inhibitors in people living with HIV-1 receiving boosted protease inhibitors in South Africa. Front Microbiol 2020;11:438. https://doi.org/10.3389/fmicb.2020.00438
12. Mikasi SG, Isaacs D, Ikomey GM, et al. HIV-1 drug resistance mutation analyses of Cameroon-derived integrase sequences. AIDS Res Hum Retroviruses 2020 (epub 10 August 2020). https://doi.org/10.1089/aid.2020.0022
13. Brado D, Obasa AE, Ikomey GM, 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 2018;8:4700. https://doi.org/10.1038/s41598-018-22916-6
14. Chilunga IL, Obasa AE, Mikasi SG, Jacobs GB, Cloete R. Correction: Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C integrase-dolutegravir binding. PLoS ONE 2020;15(4):e0234581. https://doi.org/10.1371/journal.pone.0234581
15. Isaacs D, Mikasi SG, Obasa AE, Ikomey GM. Structural comparison of diverse HIV-1 subtypes using molecular modelling and docking analyses of integrase inhibitors. Preprints 2020 (epub 6 February 2020). https://doi.org/10.20944/preprints202002.0062.v1

S Afr Med J 2020;110(9):827. https://doi.org/10.7196/SAMJ.2020.v110i9.14848

Chapter 10: Overall discussion and future remarks

Sub-Saharan Africa (SSA) is a developing region with most of the countries listed as resource-limited settings (RLS), with limited access to HIV drug resistance testing (HIVDRT) and infrastructure for proper research. Drug switching to effective regimens remain a major clinical challenge in RLS, especially in patients already experiencing virological failure and who have developed drug resistance. These pose a challenge on meeting the global and regional 90-90-90 target for the year 2020, with the aim that 90% of all people living with HIV (PLHIV) to know their status; 90% of people living with HIV to start cART; and 90% of people on treatment to have fully suppressed the virus. South Africa has successfully met one global target of 90% of all people living with HIV knowing their status, while 62% are on treatment and 54% of these virally suppressed. In Cameroon, 74% of people living with HIV know their status and 52% were on treatment. (UNAIDS, 2019). The high HIV-1 genetic diversity further makes it difficult to control the epidemic. It is well known that HIV diversity remains a key challenge pertaining to a wide spectrum of fields, such as serological diagnosis, virological follow-up, vaccine development and therapeutic monitoring. This is due to the concept that natural occurring polymorphisms (NOPs) from diverse HIV-1 subtypes can affect therapy outcomes. The combined results from (**Chapter 3 - Chapter 8**) discussed in full below, support the hypothesis proposed in this study that HIV-1 IN gene fragment sequences obtained from SSA contains NOPs that might possibly have an effect on the efficacy of second-generation INSTIs, like DTG. The chapters are discussed below.

In **Chapter 3** it was demonstrated that CRF02_AG remains the predominant subtype in Cameroon. This is in agreement with other reported studies (164,165). The study further reported the appearance of accessory mutations and RAMs against naïve INSTI individuals (166). NOPs across diverse subtypes were secondly reported. Some NOPs can play a role in the development of RAMs as some of them are statistically enriched. Given the presence of these mutations in Cameroon, where INSTIs is not yet widely in use, there is the possibility of cross-resistance to occur if drug resistance testing is not done prior to switching cART regimens. This assumption is supported, by the data generated from **Chapter 4** that analysed CRF02_AG sequences data dating back from 1999 to 2004, which also detected major RAMs that cause resistance against INSTIs. Furthermore, we detected Q148H that can cause cross-resistance to other class of INSTIs, if it occurs in combination with other mutations, R263K

(0.3%) mutation identified in our study, is commonly found in patients experiencing therapy failure , especially on incompletely suppressive regimens containing DTG. In South Africa, subtype C, remains the driving force of the epidemic, **Chapter 5**. High level of RAMs against first-line cART in South Africa is increasing and this could be due to poor adherence. The low rate of RAMs against PI and INSTIs was expected as the majority of South African individuals are still on standardised regimens**.** Results from **Chapter 6** showed some interactions between variant residues and DNA, could promote integration of DNA and assumed to have effect on drug binding. NOPs can have two effect on the IN gene, firstly NOPs can interfere with the ability of protein binding caused by primary mutations and also NOPs can help to restore viral enzymatic activity, based on the presence of background mutations and specific subtype. Some mutations that resulted in the change in number of interactions were found to be within the stable alpha-helix secondary structure element and this change could affect the protein conformation. Results from **Chapter 7** give long-term optimism for the use of DTG as part of first-line cART. We explored computational analysis to study the possible impact on protein structure caused by NOPs present in diverse HIV-1 subtypes (CRF02_AG and HIV-1C). NOPs present in IN sequences and *in silico* analysis of IN/DNA complexes predicted the impact of NOPs on the interaction between the DNA and INSTIs (48). We identified 15 NOPs present on CRF02_AG when compared to subtype B using consensus sequence analysis. Based on a model of the HIV-1B integrase/DNA complex, Malet *et al*., suggested that these NOPs might affect IN interaction with DNA or IN susceptibility to INSTIs (101). Moreover, two NOPs V82F and I84V, present in our study were previously reported among subtype A and C viruses to be associated with reduced binding affinity, as compared to subtype B. However, more mechanistic studies are needed to demonstrate the association of NOPs with major INSTI RAMs and evaluate the potency of INSTIs in treating HIV-1 non-B subtypes. In **chapter 8** the potential impact of INSTIs mutations among subtypes C on the protein structure, functions and susceptibility to INSTIs were performed using computational modelling tools. Results suggest that the G140S mutant has the strongest effect on the HIV-1C IN protein structure and DTG binding. However, a combination of mutants still need to be investigated among individuals infected with diverse HIV-1 subtypes. This approach will allow us to have better insight knowledge with the pathways of resistance and drug binding, more especially in SSA, where HIV-1 infections continues to increase and the majority of infections are caused by non-B subtype.

Significance of the study in resource limited settings (RLS)

Wide-scale access to DTG across the SSA region has been given the green light by the WHO to form part as a the preferred first-line and salvage regimens for all HIV infected individuals and the treatment of paediatric patients (167). In SSA few studies have been done on identifying mutational patterns in the HIV-1 IN gene of the infected patients failing first-line and second-line cART (104,168,169). The proposed study had the strength to bridge that gap of lack of data on DRMs against DTG in RLS and able to study genetically diverse HIV-1 strains that might potentially have an impact on the susceptibility to INSTIs of patients' sequences infected with different subtypes. Furthermore, this study will strengthen the collaborative research initiative and skills transfer on HIV-1 drug resistance between South Africa and Cameroon. Data generated from this study will help to guide in tailoring effective treatment strategies in the South African and Cameroonian populations, infected with diverse HIV-1 subtypes. Computational analysis methods used in this study is cost effective, time saving and can be used as a tool to predict therapy failure due to resistance and timely switching of patients to a more efficacious regimen.

Overall strength and limitations of the work

There are several strengths and limitations in the study. There was small samples size of patient sequences analysed (**Chapter 3**). This was as a result of no or poor PCR DNA amplification. This is most likely due to the high sequence variability seen in HIV, thus primer mismatches lead to inefficient PCR amplification and DNA sequencing. In addition, there were no data available, or access to patients, who are currently receiving INSTIs to make comparison between INSTI naïve and treated INSTI individuals. We were only limited to one statistically method for analysis of the mutations that might be statistically enriched. This would have been an interesting and crucial addition to our knowledge of INSTI success in cART regimens. In **Chapter 4** we used online database sequences dating back from 2000-2001 before the roll out of HIV treatment in South Africa and this result provided a baseline resistance in South Africa. In this study we did not screen for resistance associated mutations on the 3′ polypurine tract (PPT) region of HIV-1 that has been suggested as an alternative mechanism to the develop of resistance to INSTIs. **Chapter 5**. 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 study therefore contributed to expanding current HIV-1 knowledge regarding LPV/r and ATV resistance.

Although there was no data of HIV-1 viral load and CD4 count. Other limitations was that we only investigated RAL and EVG and did not include novel DTG mutations. We further excluded entropy effects due to lack of computational resources and this might have led to under or over estimation of the binding free energy. In addition to the study limitations include lack of performing *in vitro* assays to determine the drug resistance against virus isolates that harbour RAMs in relation to WT strains. The study had the strength to study the mutations that might potential have an impact on the protein binding and susceptibility to INSTIs of patients' sequences infected with different subtypes in the context of 3D using computational analysis (**Chapter 6 and Chapter 7**). Thus far, subtype B is the only structure available that is used as a template to study structural difference because there is no structure for non-B subtypes. This is the first study to successfully build an accurate three dimensional model using wild-type HIV-1 subtype C consensus sequence to study the effect of mutations on INSTIs drug binding (**Chapter 8**).

Overall strength of our study, knowledge transfer and strengthening collaboration

Our collaborating partners throughout Africa are using the HIV resistance protocols we have developed in our studies. These include Dr. George Ikomey – University of Yaoundé I in Cameroon, Mr. Henerico Shimba in Mwanza, Tanzania and Dr. Franklin Onyambu from the University of Nairobi, Kenya. This is a real world impact and shows the strength of our work. Other international partners, from the Institute of Virology and Immunobiology, University of Wuerzburg, Germany and the Department of Laboratory Medicine, Karolinska Institute, Sweden, have also used our Methods.

Conclusions

The following conclusions and recommendations can be made from findings in this thesis

- In South Africa, there is a high prevalence of RAMs against first line therapy (NRTI and NNRTI), with increasing number of patients having RAMs against PIs and INSTIs. Thus, close monitoring strategies to ensure a successful switch of regimens is warranted.

- IN Cameroon, there is a low rate of mutations against INSTIs and this could be due to INSTIs not yet fully available. There is a possibility of HIV-1 infected patients to experience virological failure ,when switched to DTG as mutations already appear before the roll out of DTG based regimen in Cameroon

- This study further reports the analysis and interpretation of HIV-1 IN sequences and subtyping among infected individuals from naïve ART Cameroon and South Africa. CRF02_AG continues to increase and becoming the predominant subtype causing the infections in Cameroon, while subtype C also remains the predominant subtype causing the infections in South Africa.

- We further reported the impact of various NOPs and their impact on drug susceptibility. The analyses suggested that NOPs might have impact on IN protein structure and stability, which will in turn affect viral DNA binding and which will subsequently affect drug sensitivity.

- Up to date, the only available crystal structure used to study computational analysis is based on the HIV-1B intasome (PDBID: 5U1C). This is not ideal to assess the viability and long-term efficacy of INSTIs based regimens on non-B subtypes circulating in SSA. Therefore, it is imperative that standard crystal structure to study diverse subtypes mostly affecting sub-Saharan Africa should be investigated.
- Despite the fact that the majority (90%) of infections worldwide are caused by HIV-1 non-B subtypes, knowledge of HIV-1 diversity and drug resistance-associated mutations (RAMs) in non-B HIV-1 and their clinical management in RLS is limited and need to be addressed.

Overall, to the best of our knowledge this is the first study to successfully construct an accurate 3D model structure that ca be used to study further *in silico* studies. Furthermore, our study findings support the use of INSTIs in first line therapy to form part of the long-sought goal of attainment of a functional cure for HIV-1 disease. However, before roll out of DTG in RLS, infrastructure capacity need to be improved, to allow proper and adequate switching of regimens to prevent carrying over resistance mutations.

Future work

- Perform drug resistance analysis of mutations outside IN that can confer resistance against DTG
- Perform phenotypic drug resistance analysis on integrase-based recombinant viral mutants.
- Perform viral fitness assays to determine the effect of known mutants on the HIV-1 virus replication in the presence of DTG.
- Perform viral integration assays to determine if DTG, BIC and CBT can prevent viral integration within plasmid.

Chapter 11: References

1. Zhuang J, Jetzt AE, Sun G, et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. J Virol. 2002;76(22):11273-11282. doi:10.1128/jvi.76.22.11273-11282.2002.

2. Gallo R, Wong-Staal F, Montagnier L, Haseltine WA, Yoshida M. HIV/HTLV gene nomenclature. Nature. 1988;333(6173):504. doi:10.1038/333504a0.

3. Akari H, Sakuragi J, Takebe Y, et al. Biological characterization of human immunodeficiency virus type 1 and type 2 mutants in human peripheral blood mononuclear cells. Arch Virol. 1992;123(1-2):157-167. doi:10.1007/BF01317146.

4. Schwartz O, Maréchal V, Le Gall S, Lemonnier F, Heard JM. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. Nat Med. 1996;2(3):338-342. doi:10.1038/nm0396-338

5. Kim SY, Byrn R, Groopman J, Baltimore D. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. J Virol. 1989;63(9):3708-3713. doi:10.1128/JVI.63.9.3708-3713.1989

6. Estrabaud E, Le Rouzic E, Lopez-Vergès S, et al. Regulated degradation of the HIV-1 Vpu protein through a betaTrCP-independent pathway limits the release of viral particles. PLoS Pathog. 2007;3(7):e104. doi:10.1371/journal.ppat.0030104

7. NAM aidsmap (2017) Factsheet: HIV lifecycle (Accessed January 2020).

8. Regoes RR, Bonhoeffer S. The HIV coreceptor switch: a population dynamical perspective. Trends Microbiol. 2005;13(6):269-277. doi:10.1016/j.tim.2005.04.005.

9. Clapham PR, McKnight Á. Cell surface receptors, virus entry and tropism of primate lentiviruses. J Gen Virol. 2002;83(Pt 8):1809-1829. doi:10.1099/0022-1317-83-8-1809.

10. Doms RW, Moore JP. HIV-1 membrane fusion: targets of opportunity. J Cell Biol.

2000;151(2):F9-F14. doi:10.1083/jcb.151.2.f9.

11. Ambrose Z, Aiken C. HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. Virology. 2014;454-455:371-379. doi:10.1016/j.virol.2014.02.004.

12. Delelis O, Carayon K, Saïb A, Deprez E, Mouscadet JF. Integrase and integration: biochemical activities of HIV-1 integrase. Retrovirology. 2008;5:114. Published 2008 Dec 17. doi:10.1186/1742-4690-5-114.

13. Craigie R, Bushman FD. HIV DNA integration. Cold Spring Harb Perspect Med. 2012 Jul;2(7):a006890. doi: 10.1101/cshperspect.a006890. PMID: 22762018; PMCID: PMC3385939.

14. Guerrero S, Batisse J, Libre C, Bernacchi S, Marquet R, Paillart JC. HIV-1 replication and the cellular eukaryotic translation apparatus. Viruses. 2015;7(1):199-218. Published 2015 Jan 19. doi:10.3390/v7010199.

15. Leblanc J, Weil J, Beemon K. Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. Wiley Interdiscip Rev RNA. 2013;4(5):567-580. doi:10.1002/wrna.1179.

16. Checkley MA, Luttge BG, Freed EO. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. J Mol Biol. 2011;410(4):582-608. doi:10.1016/j.jmb.2011.04.042.

17. Ono A, Freed EO. Plasma membrane rafts play a critical role in HIV-1 assembly and release. PNAS; 2001, 98 (24) 13925-13930; DOI: 10.1073/pnas.241320298.

18. Sundquist WI, Kräusslich HG. HIV-1 assembly, budding, and maturation [published correction appears in Cold Spring Harb Perspect Med. 2012 Aug;2(8). doi: 10.1101/cshperspect.a015420]. Cold Spring Harb Perspect Med. 2012;2(7):a006924. doi:10.1101/cshperspect.a006924.

19. Pey SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet P V, Kleinmil CA, et al. The p2 Domain of Human Immunodeficiency Virus Type 1 Gag Regulates Sequential Proteolytic Processing and Is Required To Produce Fully Infectious Virions. Virology. 1994;68(12):8017–27.

20. Wiegers K, Rutter G, Kottler H, Tessmer UWE. Sequential Steps in Human Immunodeficiency Virus Particle Maturation Revealed by Alterations of Individual Gag Polyprotein Cleavage Sites. J Virol. 1998 Apr; 72(4): 2846–2854.

21. Wright ER, Schooler JB, Ding HJ, et al. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. EMBO J. 2007;26(8):2218-2226. doi:10.1038/sj.emboj.7601664.

22. Wang D, Lu W, Li F. Pharmacological intervention of HIV-1 maturation. Acta Pharm Sin B [Internet]. 2015;5(6):493–9. Available from: http://dx.doi.org/10.1016/j.apsb.2015.05.004

23. Pomerantz RJ, Horn DL. Twenty years of therapy for HIV-1 infection. Nat Med. 2003;9(7):867-873. doi:10.1038/nm0703-867.

24. UNAIDS. South Africa 90-90-90 progress (2019).

25. WHO. Update of recommendations on first- and second-line antiretroviral regimens. Geneva, Switzerland: World Health Organization; 2019 (WHO/CDS/HIV/19.15). Licence: CC BY-NC-SA 3.0 IGO. 2019.

26. UNAIDS. UNAIDS Estimates 2019; Global AIDS Monitoring 2019 [Internet]. 2019 [cited 2020 Feb 4]. p. 3. Available from: https://www.unaids.org/en/regionscountries/countries/cameroon.

27. UNAIDS. Ending AIDS: Progress towards 90-90-90 targets. 2017. p. 182–92.

28. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med. 2011;1(1):a006841. doi:10.1101/cshperspect.a006841.

29. Plantier JC, Djemai M, Lemée V, et al. Census and analysis of persistent false-negative results in serological diagnosis of human immunodeficiency virus type 1 group O infections. J Clin Microbiol. 2009;47(9):2906-2911. doi:10.1128/JCM.00602-09.

30. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK,et al. (2000) 'HIV-1 Nomenclature Proposal', Science, 288(5463), pp. 55 LP – 55. doi: 10.1126/science.288.5463.55d.

31. Kalish ML, Robbins KE, Pieniazek D, et al. Recombinant viruses and early global HIV-1 epidemic. Emerg Infect Dis. 2004;10(7):1227-1234. doi:10.3201/eid1007.030904.

32. Serwadda D, Sewankambo N, Gray RH. NIH Public Access. 2009;23(14):1–14.

33. Yamaguchi J, Vallari A, McArthur C, et al. Brief Report: Complete Genome Sequence of CG-0018a-01 Establishes HIV-1 Subtype L. J Acquir Immune Defic Syndr. 2020;83(3):319-322. doi:10.1097/QAI.0000000000002246.

34. Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. J Infect [Internet]. 2013;66(5):391–400. Available from: http://dx.doi.org/10.1016/j.jinf.2012.10.026

35. Abram ME, Hluhanich RM, Goodman DD, et al. Impact of primary elvitegravir resistance-associated mutations in HIV-1 integrase on drug susceptibility and viral replication fitness. Antimicrob Agents Chemother. 2013;57(6):2654-2663. doi:10.1128/AAC.02568-12.

36. Jacobs GB, Loxton AG, Laten A, Robson B, van Rensburg EJ, Engelbrecht S. Emergence and diversity of different HIV-1 subtypes in South Africa, 2000-2001. J Med Virol. 2009;81(11):1852-1859. doi:10.1002/jmv.21609.

37. Novitsky V, Wester CW, DeGruttola V, et al. The reverse transcriptase 67N 70R 215Y genotype is the predominant TAM pathway associated with virologic failure among HIV type 1C-infected adults treated with ZDV/ddI-containing HAART in southern Africa. AIDS Res Hum Retroviruses. 2007;23(7):868-878. doi:10.1089/aid.2006.0298.

38. Vallari A, Bodelle P, Ngansop C, et al. Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low. AIDS Res Hum Retroviruses. 2010;26(1):109-115. doi:10.1089/aid.2009.0178.

39. Vallari A, Holzmayer V, Harris B, et al. Confirmation of putative HIV-1 group P in Cameroon. J Virol. 2011;85(3):1403-1407. doi:10.1128/JVI.02005-10.

40. Powell R, Barengolts D, Mayr L, Nyambi P. The Evolution of HIV-1 Diversity in Rural Cameroon and its Implications in Vaccine Design and Trials. Viruses. 2010;2(2):639-654. doi:10.3390/v2020639.

41. Neogi U, Prarthana BS, Gupta S, et al. Naturally occurring polymorphisms and primary drug resistance profile among antiretroviral-naive individuals in Bangalore, India. AIDS Res Hum Retroviruses. 2010;26(10):1097-1101. doi:10.1089/aid.2010.0092.

42. Lataillade M, Chiarella J, Kozal MJ. Natural polymorphism of the HIV-1 integrase gene and mutations associated with integrase inhibitor resistance. Antivir Ther. 2007;12(4):563-570.

43. Paraschiv S, Otelea D, Dinu M, Maxim D, Tinischi M. Polymorphisms and resistance mutations in the protease and reverse transcriptase genes of HIV-1 F subtype Romanian strains. Int J Infect Dis. 2007;11(2):123-128. doi:10.1016/j.ijid.2005.11.006.

44. Bar-Magen T, Sloan RD, Donahue DA, et al. Identification of novel mutations responsible for resistance to MK-2048, a second-generation HIV-1 integrase inhibitor. J Virol. 2010;84(18):9210-9216. doi:10.1128/JVI.01164-10.

45. Malet I, Fourati S, Charpentier C, et al. The HIV-1 integrase G118R mutation confers raltegravir resistance to the CRF02_AG HIV-1 subtype. J Antimicrob Chemother. 2011;66(12):2827-2830. doi:10.1093/jac/dkr389.

46. Fransen S, Gupta S, Danovich R, et al. Loss of raltegravir susceptibility by human immunodeficiency virus type 1 is conferred via multiple nonoverlapping genetic pathways. J Virol. 2009;83(22):11440-11446. doi:10.1128/JVI.01168-09.

47. Ceccherini-Silberstein F, Malet I FL. Specific mutations related to resistance to HIV-1 integrase inhibitors are associated with reverse transcriptase mutations in HAART-treated patients [abstract 4]. Antivir Ther 12. 2007;12:S6 (abstract no. 4).

48. Brenner BG, Lowe M, Moisi D, et al. Subtype diversity associated with the development of HIV-1 resistance to integrase inhibitors. J Med Virol. 2011;83(5):751-759. doi:10.1002/jmv.22047.

49. Roxana M. Coman, Arthur H. Robbins, Marty A. Fernandez, C. Taylor Gilliland, Anthony A. Sochet, Maureen M. Goodenow, Robert McKenna and BMD. The Contribution of Naturally Occurring Polymorphisms in Altering the Biochemical and Structural Characteristics of HIV-1 Subtype C Protease†. Biochemistry. 2008;47, 2:731–43. doi.org/10.1021/bi051886s.

50. Engelman A, Bushman FD, Craigie R. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. EMBO J. 1993;12(8):3269-3275.

51. 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):13659-13664. doi:10.1073/pnas.93.24.13659.

52. Engelman A, Craigie R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. J Virol. 1992;66(11):6361-6369. doi:10.1128/JVI.66.11.6361-6369.1992.

53. Esposito D, Craigie R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. EMBO J. 1998;17(19):5832-5843. doi:10.1093/emboj/17.19.5832.

54. Goldgur Y, Craigie R, Cohen GH, et al. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. Proc Natl Acad Sci U S A. 1999;96(23):13040-13043. doi:10.1073/pnas.96.23.13040.

55. Reigadas S, Masquelier B, Calmels C, et al. Structure-analysis of the HIV-1 integrase Y143C/R raltegravir resistance mutation in association with the secondary mutation T97A. Antimicrob Agents Chemother. 2011;55(7):3187-3194. doi:10.1128/AAC.00071-11.

56. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci U S A. 2005;102(48):17308-17313. doi:10.1073/pnas.0506924102.

57. Maignan S, Guilloteau JP, Zhou-Liu Q, Clément-Mella C, Mikol V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. J Mol Biol. 1998;282(2):359-368. doi:10.1006/jmbi.1998.2002.

58. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. Proc Natl Acad Sci U S A. 1998;95(16):9150-9154. doi:10.1073/pnas.95.16.9150.

59. Passos DO, Li M, Yang R, et al. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. Science. 2017;355(6320):89-92. doi:10.1126/science.aah5163.

60. Mouscadet JF, Delelis O, Marcelin AG, Tchertanov L. Resistance to HIV-1 integrase inhibitors: A structural perspective. Drug Resist Updat. 2010;13(4-5):139-150. doi:10.1016/j.drup.2010.05.001.

61. Malet I, Gimferrer Arriaga L, Artese A, et al. New raltegravir resistance pathways induce broad cross-resistance to all currently used integrase inhibitors. J Antimicrob Chemother. 2014;69(8):2118-2122. doi:10.1093/jac/dku095.

62. Ferris AL, Wu X, Hughes CM, et al. Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. Proc Natl Acad Sci U S A. 2010;107(7):3135-3140. doi:10.1073/pnas.0914142107.

63. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci U S A. 2005;102(48):17308-17313. doi:10.1073/pnas.0506924102.

64. Rogers L, Obasa AE, Jacobs GB, et al. Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. Front Microbiol. 2018;9:1754. Published 2018 Aug 2. doi:10.3389/fmicb.2018.01754.

65. Hare S, Smith SJ, Métifiot M, et al. Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). Mol Pharmacol. 2011;80(4):565-572. doi:10.1124/mol.111.073189.

66. Inzaule SC, Hamers RL, Noguera-Julian M, et al. Primary resistance to integrase strand transfer inhibitors in patients infected with diverse HIV-1 subtypes in sub-Saharan Africa. J Antimicrob Chemother. 2018;73(5):1167-1172. doi:10.1093/jac/dky005.

67. Craigie R. The molecular biology of HIV integrase. Future Virol. 2012 Jul; 7(7): 679–686. doi: 10.2217/FVL.12.56.

68. Boyer PL, Clark PK, Hughes SH. HIV-1 and HIV-2 reverse transcriptases: different mechanisms of resistance to nucleoside reverse transcriptase inhibitors. J Virol. 2012;86(10):5885-5894. doi:10.1128/JVI.06597-11.

69. Rosenblum LL, Patton G, Grigg AR, et al. Differential susceptibility of retroviruses to nucleoside analogues. Antivir Chem Chemother. 2001;12(2):91-97. doi:10.1177/095632020101200202.

70. De Clercq E. The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. Antiviral Res. 1998;38(3):153-179. doi:10.1016/s0166-3542(98)00025-4.

71. C de Clercq E. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) for the treatment of human immunodeficiency virus type 1 (HIV-1) infections: strategies to overcome drug resistance development. Med Res Rev. 1996;16(2):125-157.

doi:10.1002/(SICI)1098-1128(199603)16:2<125::AID-MED1>3.0.CO;2-2.

72. Pedersen OS, Pedersen EB. Non-nucleoside reverse transcriptase inhibitors: the NNRTI boom. Antivir Chem Chemother. 1999;10(6):285-314. doi:10.1177/095632029901000601.

73. Espeseth AS, Felock P, Wolfe A, 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 U S A. 2000;97(21):11244-11249. doi:10.1073/pnas.200139397.

74. Kawasuji T, Fuji M, Yoshinaga T, Sato A, Fujiwara T, Kiyama R. A platform for designing HIV integrase inhibitors. Part 2: a two-metal binding model as a potential mechanism of HIV integrase inhibitors. Bioorg Med Chem. 2006;14(24):8420-8429. doi:10.1016/j.bmc.2006.08.043.

75. Adamson CS. Protease-Mediated Maturation of HIV: Inhibitors of Protease and the Maturation Process. Mol Biol Int. 2012;2012:604261. doi:10.1155/2012/604261.

76. Wlodawer A, Erickson JW. Structure-based inhibitors of HIV-1 protease. Annu Rev Biochem. 1993;62:543-585. doi:10.1146/annurev.bi.62.070193.002551

77. Flexner C. HIV-protease inhibitors. N Engl J Med. 1998;338(18):1281-1292. doi:10.1056/NEJM199804303381808.

78. McQuade TJ, Tomasselli AG, Liu L, et al. A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. Science. 1990;247(4941):454-456. doi:10.1126/science.2405486.

79. Trifilo MJ, Bergmann CC, Kuziel WA, Lane TE. CC chemokine ligand 3 (CCL3) regulates CD8(+)-T-cell effector function and migration following viral infection. J Virol. 2003;77(7):4004-4014. doi:10.1128/jvi.77.7.4004-4014.2003.

80. Dittmar MT, McKnight A, Simmons G, Clapham PR, Weiss RA, Simmonds P. HIV-1 tropism and co-receptor use. Nature. 1997;385(6616):495-496. doi:10.1038/385495a0.

81. Tsamis F, Gavrilov S, Kajumo F, et al. Analysis of the mechanism by which the small-molecule CCR5 antagonists SCH-351125 and SCH-350581 inhibit human immunodeficiency virus type 1 entry. J Virol. 2003;77(9):5201-5208. doi:10.1128/jvi.77.9.5201-5208.2003.

82. Boyd M, Pett S, Physicians ID. Experimental and clinical pharmacology HIV fusion inhibitors : a review. Aust Prescr. 1983;66–9. DOI: 10.18773/austprescr.2008.040.

83. Malik T, Chauhan G, Rath G, Murthy RS, Goyal AK. "Fusion and binding inhibition" key target for HIV-1 treatment and pre-exposure prophylaxis: targets, drug delivery and nanotechnology approaches. Drug Deliv. 2017;24(1):608-621. doi:10.1080/10717544.2016.1228717.

84. Rezende LF, Kew Y, Prasad VR. The effect of increased processivity on overall fidelity of human immunodeficiency virus type 1 reverse transcriptase. J Biomed Sci. 2001;8(2):197-205. doi:10.1007/BF02256413.

85. Tang MW, Shafer RW. HIV-1 antiretroviral resistance: scientific principles and clinical applications. Drugs. 2012;72(9):e1-e25. doi:10.2165/11633630-000000000-00000

86. Ramirez BC, Simon-Loriere E, Galetto R, Negroni M. Implications of recombination for HIV diversity. Virus Res. 2008;134(1-2):64-73. doi:10.1016/j.virusres.2008.01.007

87. Crumpacker CS, Thr L, Val L, Val M. Decreased Processivity of Human Immunodeficiency Virus Type 1 Reverse Transcriptase (RT) Containing Didanosine-Selected Mutation Leu74Val : a Comparative Analysis of RT Variants. Journal of Virology. 1999, 73 (10) 8448-8456; DOI: 10.1128/JVI.73.10.8448-8456.1999.

88. Dapp MJ, Heineman RH, Mansky LM. Interrelationship between HIV-1 fitness and mutation rate. J Mol Biol. 2013;425(1):41-53. doi:10.1016/j.jmb.2012.10.009.

89. Xu HT, Asahchop EL, Oliveira M, et al. Compensation by the E138K mutation in HIV-1 reverse transcriptase for deficits in viral replication capacity and enzyme processivity associated with the M184I/V mutations. J Virol. 2011;85(21):11300-11308.

doi:10.1128/JVI.05584-11.

90. Huang KH, Goedhals D, Carlson JM, et al. Progression to AIDS in South Africa is associated with both reverting and compensatory viral mutations. PLoS One. 2011;6(4):e19018. Published 2011 Apr 22. doi:10.1371/journal.pone.0019018.

91. Hu Z, Kuritzkes DR. Interaction of reverse transcriptase (RT) mutations conferring resistance to lamivudine and etravirine: effects on fitness and RT activity of human immunodeficiency virus type 1. J Virol. 2011;85(21):11309-11314. doi:10.1128/JVI.05578-11

92. Johnson VA, Brun-Vézinet F, Clotet B, et al. Drug resistance mutations in HIV-1. Top HIV Med. 2003;11(6):215-221.

93. Shirasaka T, Kavlick MF, Ueno T, et al. Emergence of human immunodeficiency virus type 1 variants with resistance to multiple dideoxynucleosides in patients receiving therapy with dideoxynucleosides. Proc Natl Acad Sci U S A. 1995;92(6):2398-2402. doi:10.1073/pnas.92.6.2398

94. Ueno T, Shirasaka T, Mitsuya H. Enzymatic characterization of human immunodeficiency virus type 1 reverse transcriptase resistant to multiple 2',3'-dideoxynucleoside 5'-triphosphates. J Biol Chem. 1995;270(40):23605-23611. doi:10.1074/jbc.270.40.23605

95. Saini S, Bhalla P, Gautam H, Baveja UK, Pasha ST, Dewan R. Resistance-associated mutations in HIV-1 among patients failing first-line antiretroviral therapy. J Int Assoc Physicians AIDS Care (Chic). 2012;11(3):203-209. doi:10.1177/1545109711421217.

96. Zhao Y, Lu X, Yang CY, et al. Computational modeling toward understanding agonist binding on dopamine 3. J Chem Inf Model. 2010;50(9):1633-1643. doi:10.1021/ci1002119.

97. Richman D, Shih CK, Lowy I, et al. Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture. Proc

Natl Acad Sci U S A. 1991;88(24):11241-11245. doi:10.1073/pnas.88.24.11241.

98. Melikian GL, Rhee SY, Varghese V, et al. Non-nucleoside reverse transcriptase inhibitor (NNRTI) cross-resistance: implications for preclinical evaluation of novel NNRTIs and clinical genotypic resistance testing. J Antimicrob Chemother. 2014;69(1):12-20. doi:10.1093/jac/dkt316.

99. Hu Z, Kuritzkes DR. Altered viral fitness and drug susceptibility in HIV-1 carrying mutations that confer resistance to nonnucleoside reverse transcriptase and integrase strand transfer inhibitors. J Virol. 2014;88(16):9268-9276. doi:10.1128/JVI.00695-14.

100. Shafer RW .Rationale and uses of a public HIV drug-resistance database. J Infect Dis. 2006; 194 Suppl 1S51–S58. doi: 10.1097/01.qai.0000232609.99705.a2.

101. Malet I, Soulie C, Tchertanov L, et al. Structural effects of amino acid variations between B and CRF02-AG HIV-1 integrases. J Med Virol. 2008;80(5):754-761. doi:10.1002/jmv.21169.

102. Elliott JH, Lynen L, Calmy A, et al. Rational use of antiretroviral therapy in low-income and middle-income countries: optimizing regimen sequencing and switching. AIDS. 2008;22(16):2053-2067. doi:10.1097/QAD.0b013e328309520d.

103. Hassounah SA, Mesplède T, Quashie PK, Oliveira M, Sandstrom PA, Wainberg MA. Effect of HIV-1 integrase resistance mutations when introduced into SIVmac239 on susceptibility to integrase strand transfer inhibitors. J Virol. 2014;88(17):9683-9692. doi:10.1128/JVI.00947-14

104. 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 [published correction appears in Sci Rep. 2018 Apr 16;8(1):6262]. Sci Rep. 2018;8(1):4709. Published 2018 Mar 16. doi:10.1038/s41598-018-22914-5.

105. Hachiya A, Ode H, Matsuda M, et al. Natural polymorphism S119R of HIV-1 integrase

enhances primary INSTI resistance. Antiviral Res. 2015;119:84-88. doi:10.1016/j.antiviral.2015.04.014

106. Malet I, Subra F, Charpentier C, et al. Mutations Located outside the Integrase Gene Can Confer Resistance to HIV-1 Integrase Strand Transfer Inhibitors. mBio. 2017;8(5):e00922-17. Published 2017 Sep 26. doi:10.1128/mBio.00922-17.

107. Delelis O, Malet I, Na L, et al. The G140S mutation in HIV integrases from raltegravir-resistant patients rescues catalytic defect due to the resistance Q148H mutation. Nucleic Acids Res. 2009;37(4):1193-1201. doi:10.1093/nar/gkn1050.

108. Wensing AM, van Maarseveen NM, Nijhuis M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. Antiviral Res. 2010;85(1):59-74. doi:10.1016/j.antiviral.2009.10.003.

109. Logsdon BC, Vickrey JF, Martin P, et al. Crystal structures of a multidrug-resistant human immunodeficiency virus type 1 protease reveal an expanded active-site cavity. J Virol. 2004;78(6):3123-3132. doi:10.1128/jvi.78.6.3123-3132.2004

110. Bally F, Martinez R, Peters S, Sudre P, Telenti A. Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors. AIDS Res Hum Retroviruses. 2000;16(13):1209-1213. doi:10.1089/08892220050116970.

111. Abecasis AB, Deforche K, Snoeck J, et al. Protease mutation M89I/V is linked to therapy failure in patients infected with the HIV-1 non-B subtypes C, F or G [published correction appears in AIDS. 2006 Aug 22;20(13):1789]. AIDS. 2005;19(16):1799-1806. doi:10.1097/01.aids.0000188422.95162.b7

112. Deforche K, Camacho R, Grossman Z, et al. Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. Infect Genet Evol. 2007;7(3):382-390. doi:10.1016/j.meegid.2006.09.004.

113. Martinez-Picado J, Savara AV, Sutton L, D'Aquila RT. Replicative fitness of protease

inhibitor-resistant mutants of human immunodeficiency virus type 1. J Virol. 1999;73(5):3744-3752. doi:10.1128/JVI.73.5.3744-3752.1999.

114. Xue W, Jin X, Ning L, Wang M, Liu H, Yao X. Exploring the molecular mechanism of cross-resistance to HIV-1 integrase strand transfer inhibitors by molecular dynamics simulation and residue interaction network analysis. J Chem Inf Model. 2013;53(1):210-222. doi:10.1021/ci300541c.

115. Xue W, Liu H, Yao X. Molecular modeling study on the allosteric inhibition mechanism of HIV-1 integrase by LEDGF/p75 binding site inhibitors. PLoS One. 2014;9(3):e90799. Published 2014 Mar 5. doi:10.1371/journal.pone.0090799.

116. Santos LH, Ferreira RS, Caffarena ER. Computational drug design strategies applied to the modelling of human immunodeficiency virus-1 reverse transcriptase inhibitors. Mem Inst Oswaldo Cruz. 2015;110(7):847-864. doi:10.1590/0074-02760150239.

117. Safi M, Lilien RH. Efficient a priori identification of drug resistant mutations using Dead-End Elimination and MM-PBSA. J Chem Inf Model. 2012;52(6):1529-1541. doi:10.1021/ci200626m.

118. Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. Drug Discov Today. 2009;14(13-14):676-683. doi:10.1016/j.drudis.2009.04.006.

119. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J. 1986;5(4):823-826.

120. Westbrook J, Feng Z, Chen L, Yang H, Berman HM. The Protein Data Bank and structural genomics. Nucleic Acids Res. 2003;31(1):489-491. doi:10.1093/nar/gkg068.

121. Boeckmann B, Bairoch A, Apweiler R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 2003;31(1):365-370. doi:10.1093/nar/gkg095.

122. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. 2003;31(13):3381-3385. doi:10.1093/nar/gkg520.

123. Szilagyi A, Zhang Y. Template-based structure modeling of protein-protein interactions. Curr Opin Struct Biol. 2014;24:10-23. doi:10.1016/j.sbi.2013.11.005.

124. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. Protein Sci. 2000;9(9):1753-1773. doi:10.1110/ps.9.9.1753.

125. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality [published correction appears in Proteins. 2007 Sep 1;68(4):1020]. Proteins. 2004;57(4):702-710. doi:10.1002/prot.20264.

126. Walker JM. Methods IN M OLECULAR B IOLOGY ™ Series Editor. 852 p.

127. Nayeem A, Sitkoff D, Krystek S Jr. A comparative study of available software for high-accuracy homology modeling: from sequence alignments to structural models. Protein Sci. 2006;15(4):808-824. doi:10.1110/ps.051892906.

128. Ginalski K. Comparative modeling for protein structure prediction. Curr Opin Struct Biol. 2006;16(2):172-177. doi:10.1016/j.sbi.2006.02.003.

129. Ghosh S, Gadiyaram V, Vishveshwara S. Validation of protein structure models using network similarity score. Proteins. 2017;85(9):1759-1776. doi:10.1002/prot.25332.

130. Hillisch A, Pineda LF, Hilgenfeld R. Utility of homology models in the drug discovery process. Drug Discov Today. 2004;9(15):659-669. doi:10.1016/S1359-6446(04)03196-4.

131. Munsamy G, Soliman MES. Homology Modeling in Drug Discovery-an Update on the Last Decade Homology Modeling in Drug Discovery-An Update on the Last Decade. Letters in Drug Design & Discovery . 2017; 14(9). DOI: 10.2174/1570180814666170110122027 71.

132. Rohit Arora. Molecular mechanism of HIV-1 integrase inhibition by Raltegravir proposed by using of molecular modeling approaches. Agricultural sciences. École normale supérieure de Cachan - ENS Cachan, 2012. English. ffNNT : 2012DENS0055f.

133. Adcock, S. A. and McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. Chemical Reviews. American Chemical Society.2006; 106(5), pp. 1589–1615. doi: 10.1021/cr040426m.

134. Jorgensen, W. L., Maxwell, D. S. and Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, Journal of the American Chemical Society. American Chemical Society.1996; 118(45), pp. 11225–11236. doi: 10.1021/ja9621760.

135. MacKerell AD, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998;102(18):3586-3616. doi:10.1021/jp973084f.

136. Darian E, Gannett PM. Application of molecular dynamics simulations to spin-labeled oligonucleotides. J Biomol Struct Dyn. 2005;22(5):579-593. doi:10.1080/07391102.2005.10507028.

137. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem. 2004;25(13):1656-1676. doi:10.1002/jcc.20090.

138. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26(16):1701-1718. doi:10.1002/jcc.20291.

139. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., Ferguson, D., Seibel, G., & Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications.1995; 91(1-3), 1-41. https://doi.org/10.1016/0010-

4655(95)00041-D.

140. Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., & Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. Journal of Computational Physics.1996; 151(1), 283-312. https://doi.org/10.1006/jcph.1999.6201.

141. Karplus M, Kuriyan J. Molecular dynamics and protein function. Proc Natl Acad Sci U S A. 2005;102(19):6679-6685. doi:10.1073/pnas.0408930102.

142. Karplus M. Molecular dynamics simulations of biomolecules. Acc Chem Res. 2002;35(6):321-323. doi:10.1021/ar020082r.

143. Wang W, Donini O, Reyes CM, Kollman PA. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. Annu Rev Biophys Biomol Struct. 2001;30:211-243. doi:10.1146/annurev.biophys.30.1.211.

144. Galeazzi R. Molecular dynamics as a tool in rational drug design: current status and some major applications. Curr Comput Aided Drug Des. 2009;5:225–40.

145. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical free energy methods for drug discovery: progress and challenges. Curr Opin Struct Biol. 2011;21(2):150-160. doi:10.1016/j.sbi.2011.01.011.

146. Kumari I, Sandhu P, Ahmed M, Akhter Y. Molecular Dynamics Simulations, Challenges and Opportunities: A Biologist's Prospective. Curr Protein Pept Sci. 2017;18(11):1163-1179. doi:10.2174/1389203718666170622074741.

147. Salmaso V, Moro S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. Front Pharmacol. 2018;9:923. Published 2018 Aug 22. doi:10.3389/fphar.2018.00923.

148. Chaudhary KK, Mishra N. A Review on Molecular Docking: Novel Tool for Drug

Discovery. JSM Chem . 2016; 4(3): 1029.

149. Abagyan R, Totrov M. High-throughput docking for lead generation. Curr Opin Chem Biol. 2001;5(4):375-382. doi:10.1016/s1367-5931(00)00217-9.

150. Huang SY, Zou X. Advances and challenges in protein-ligand docking. Int J Mol Sci. 2010;11(8):3016-3034. Published 2010 Aug 18. doi:10.3390/ijms11083016.

151. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004;3(11):935-949. doi:10.1038/nrd1549.

152. Lahti JL, Tang GW, Capriotti E, Liu T, Altman RB. Bioinformatics and variability in drug response: a protein structural perspective. J R Soc Interface. 2012;9(72):1409-1437. doi:10.1098/rsif.2011.0843.

153. Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit. 1996;9(1):1-5. doi:10.1002/(sici)1099-1352(199601)9:1<1::aid-jmr241>3.0.co;2-6.

154. Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des. 2001;15(5):411-428. doi:10.1023/a:1011115820450.

155. Kramer B, Rarey M, Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins. 1999;37(2):228-241. doi:10.1002/(sici)1097-0134(19991101)37:2<228::aid-prot8>3.0.co;2-8.

156. Halgren TA, Murphy RB, Friesner RA, et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem. 2004;47(7):1750-1759. doi:10.1021/jm030644s.

157. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins. 2003;52(4):609-623. doi:10.1002/prot.10465.

158. Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. J Med Chem. 2006;49(11):3315-3321. doi:10.1021/jm051197e.

159. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-461. doi:10.1002/jcc.21334.

160. Spitzer R, Jain AN. Surflex-Dock: Docking benchmarks and real-world application. J Comput Aided Mol Des. 2012;26(6):687-699. doi:10.1007/s10822-011-9533-y.

161. Guimarães CR, Cardozo M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. J Chem Inf Model. 2008;48(5):958-970. doi:10.1021/ci800004w.

162. Santos LH, Ferreira RS, Caffarena ER. Computational drug design strategies applied to the modelling of human immunodeficiency virus-1 reverse transcriptase inhibitors. Mem Inst Oswaldo Cruz. 2015;110(7):847-864. doi:10.1590/0074-02760150239.

163. Chaudhary KK, Mishra N. A Review on Molecular Docking: Novel Tool for Drug Discovery. JSM Chem. 2016; 4(3): 1029.

164. Courtney CR, Agyingi L, Fokou A, et al. Monitoring HIV-1 Group M Subtypes in Yaoundé, Cameroon Reveals Broad Genetic Diversity and a Novel CRF02_AG/F2 Infection. AIDS Res Hum Retroviruses. 2016;32(4):381-385. doi:10.1089/aid.2015.0286.

165. Abongwa LE, Nyamache AK, Torimiro JN, Okemo P, Charles F. Human immunodeficiency virus type 1 ((HIV-1) subtypes in the northwest region, Cameroon. Virol J. 2019;16(1):103. Published 2019 Aug 15. doi:10.1186/s12985-019-1209-6.

166. Tongo M, Martin DP, Zembe L, Mpoudi-Ngole E, Williamson C, Burgers WA. Characterization of HIV-1 gag and nef in Cameroon: further evidence of extreme diversity at the origin of the HIV-1 group M epidemic. Virol J. 2013;10:29. Published 2013 Jan 22. doi:10.1186/1743-422X-10-29.

167. INDICATIONS AND USAGE TIVICAY and TIVICAY PD are indicated in combination with other antiretroviral agents for the treatment of human immunodeficiency virus type 1 (HIV-1) infection in adults (treatment-naïve or -experienced) and in pediatric patients. 2020; Available from: https://www.gsksource.com/pharma/content/dam/GlaxoSmithKline/US/en/Prescribing_Information/Tivicay/pdf/TIVICAY-PI-PIL-IFU.PDF.

168. Obasa AE, Mikasi SG, Brado D, Cloete R, Singh K, Neogi U and Jacobs GB (2020) Drug Resistance Mutations Against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living With HIV-1 Receiving Boosted Protease Inhibitors in South Africa. Front. Microbiol. 11:438. doi: 10.3389/fmicb.2020.00438.

169. Steegen K, Carmona S, Bronze M, et al. Moderate Levels of Pre-Treatment HIV-1 Antiretroviral Drug Resistance Detected in the First South African National Survey. PLoS One. 2016;11(12):e0166305. Published 2016 Dec 1. doi:10.1371/journal.pone.0166305.