Investigating the structural effect of Raltegravir resistance associated mutations on the South African HIV-1 Integrase subtype C protein structure

By

Rumbidzai Chitongo

Thesis submitted in the fulfilment of the requirements for the degree of Master of Science (Bioinformatics) at the South African National Bioinformatics Institute, Faculty of Natural Sciences,
University of the Western Cape

Supervisor: Dr Ruben Cloete
Faculty of Natural Sciences
South African National Bioinformatics Institute (UWC)

Co-supervisor: Dr Graeme Brendon Jacobs
Faculty of Medicine and Health Sciences
Department of Pathology, Division of Medical Virology
University of Stellenbosch

DECLARATION

I, **Rumbidzai Chitongo**, declare that the contents of this thesis represent my own work, and that the thesis has not been submitted before for any degree or academic examination in any other university. Furthermore, it represents my own opinions and not necessarily those of the University of the Western Cape.

22 September 2020

Signed **Date**

ABSTRACT

Background and Aims

Human Immunodeficiency Virus (HIV) type 1 group M subtype C (HIV-1C) accounts for nearly half of global HIV-1 infections, with South Africa (SA) being one of the countries with the highest infection burden. In recent years, SA has made great strides in tackling its HIV epidemic, resulting in the country being recognized globally as the one sub-Saharan country with the largest combination antiretroviral therapy (cART) programme. Regardless of the potency of cART, the efficacy of the treatment is limited and hampered by the emergence of drug resistance. The majority of research on HIV-1 infections, effect of antiretroviral (ARV) drugs and understanding resistance to ARV drugs has been extensively conducted, but mainly on HIV-1 subtype B (HIV-1B), with less information known about HIV-1C. HIV-1's viral Integrase (IN) enzyme has become a viable target for highly specific cART, due to its importance in the infection and replication cycle of the virus. The lack of a complete HIV-1C IN protein structure has negatively impacted the progress on structural studies of nucleoprotein reaction intermediates. The mechanism of HIV-1 viral DNA's integration has been studied extensively at biochemical and cellular levels, but not at a molecular level. This study aims to use *in silico* methods that involve molecular modeling and molecular dynamic (MD) simulations to prioritize mutations that could affect HIV-1C IN binding to DNA and the IN strand-transfer inhibitor (INSTI) dolutegravir (DTG). The purpose is to help tailor more effective personalized treatment options for patients living with HIV in SA. This study will in part use patient derived sequence data to identify mutations and model them into the protein structure to understand their impact on the HIV-1C IN protein structure folding and dynamics.

Methods

Our sample cohort consisted of 11 sample sequences derived from SA HIV-1 treatment-experienced patients who were being treated with the INSTI raltegravir (RAL). The sequences were submitted to the Stanford HIV resistance database (HIVdb) to screen for any new/novel variants resulting from possible RAL failure. Some of these new variants were analyzed to analyse their effect, if any, on the binding of DTG to the HIV-1C IN protein. Additionally, an HIV-1C IN consensus sequence constructed from SA's HIV-1 infected population was used to model a complete three-dimensional wild type (WT) HIV-1C IN homology model. All samples were sequenced by our collaborators at the Division of Medical Virology, Stellenbosch University together with the National Health Laboratory Services (NHLS), SA. The HIV-$1C_{ZA}$ WT-IN protein enzyme was predicted using SWISS-MODEL, and the quality of the resulting model validated. Various analyses were conducted in order to study and assess

the effect of the selected new variants on the protein structure and binding of DTG to the IN protein. The mutation Cutoff Scanning Matrix (mCSM) program was used to predict protein stability after mutation, while PyMol helped to study any changes in polar contact activity before and after mutation. PyMol was also used to generate four mutant HIV-1C IN complex structures and these structures together with the WT IN were subjected to production MD simulations for 150 nanoseconds (ns). Trajectory analyses of the MD simulations were also conducted and reported.

Results

A total of 21 new variants were detected in our sample cohort, from which only six were chosen for further analyses within the study. A homology model of HIV-1C IN was successfully constructed and validated. The structural quality assessment indicated high reliability of the HIV-1C IN tetrameric structure, with more than 90.0% confidence in modelled regions. Of the six selected variants, only one (S119P) was calculated to be slightly stabilizing to the protein structure, with the other five found to be destabilizing to the IN protein structure. Variant S119P showed a loss in polar contacts that could destabilize the protein structure, while variant Y143R, resulted in the gain of polar contacts which could reduce flexibility of the 140's region affecting drug binding. Similarly, mutant systems P3 (S119P, Y143R) and P4 (V150A, M154I) showed reduced hydrogen bond formation and the weakest non-bonded pairwise interaction energy. These two systems, P3 and P4, also showed significantly reduced to none polar contacts between DTG, magnesium (MG) ions and the IN protein, compared to the WT IN and P2 mutant IN systems. Interestingly, the WT structure and systems P1 (I113V) and P2 (L63I, V75M, Y143R) showed the highest non-bonded interaction energy, compared to systems P3 and P4. This was further supported by the polar interaction analyses of simulation clusters from the WT IN and mutant IN system P2 (L63I, V75M, Y143R), which were the only protein structures that formed polar contacts with DTG, MG ions and DDE motif residues, while P1 only made contacts with DNA and IN residues.

Conclusion

Findings from this study leads to a conclusion that double mutants (S119P, Y143R) and (V150A, M154I) may result in a reduction in the efficacy of DTG, especially when in combination. Furthermore, variants identified in systems P1 and P2 may still allow for effective DTG binding to IN and outcompete viral DNA for host DNA to prevent strand transfer. To the best of our knowledge, this is the first study that uses the consensus WT HIV-

1C IN sequence to build an accurate 3D homology model to understand the effect of less frequently detected/reported variants on DTG binding in a South African context.

ACKNOWLEDGEMENTS

I would like to express my sincere and utmost gratitude to my supervisors Dr Ruben Cloete and Dr Graeme Brendon Jacobs for their guidance and advice throughout the course of this research. I would like to acknowledge and appreciate their immense contributions and unwavering support.

Many thanks to all members of the South African National Bioinformatics Institute, for your encouragement, amazing attitude and assistance during the duration of my study. Many thanks to Mrs Ferial Mullins and Mrs Fungiwe Mphiti for their administrative assistance at all times. I would also like to appreciate Mr Darren Matthew Isaacs, Mr Oluwafemi Peter Abiodun and Miss Maryam Hassan for their encouragement and support throughout my studies, you guys are simply amazing. I would also like to thank Miss Ailsa Prudence Mambinge and Miss Kudzai Chitongo, for their moral support.

My sincere appreciation goes to the National Research Foundation (NRF) and the Poliomyelitis Research Foundation (PRF). Without your financial support, my project would not have been successful.

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

DEDICATION

I dedicate this thesis to my parents, siblings, friends, all health care workers and people working and living with HIV

TABLE OF CONTENTS

DECLARATION ii
ABSTRACT iii
ACKNOWLEDGEMENTS v
DEDICATION vi
TABLE OF CONTENTS vii
LIST OF TABLES x
LIST OF FIGURES xi
SCIENTIFIC CONTRIBUTION xii
ABBREVIATION of TERMS xiii
Chapter 1 Background and Literature Review 1
1.1 Background 1
1.2 HIV-1 Integrase 2
1.2.1 Activity 2
1.2.2 Structure 3
1.3 Integrase Inhibitors 5
1.3.1 Development 5
1.3.2 Function of INSTIs 5
1.4 Resistance to INSTIs 7
1.5 Resistance Pathways of 1st and 2nd Generation INSTIs 8
1.5.1 First-generation INSTIs 8
1.5.2 Second-generation INSTIs 9
1.6 Antiretroviral Therapy and Transmitted Drug Resistance 10
1.7 Transmitted Drug Resistance in South Africa and Africa 12
1.8 Molecular Dynamic Simulations 13
1.8.1 Theory 13
1.8.2 General Applications of MD Simulations 14
1.8.2.1 Applications of MD Simulations in Biology 15
1.8.2.2 Applications of MD Simulations in Virology 16
1.8.3 Limitations of MD Simulations 16
1.9 Computational Studies of HIV-1 IN Drug Resistance 17
1.10 Rationale of Study 19
1.11 Aim of the Study 20
1.12 Objectives of the Study 20
Chapter 2 Materials and Methods 22
2.1 Ethical consideration 22
2.2 Study design and location of research 22
2.3 Inclusion criteria for selection of patient samples 22
2.4 Sequence retrieval 23
2.4.1 Sequence translation 23
2.4.2 Multiple sequence alignment 23
2.4.3 Detection of known RAMs and new mutations 24
2.5 Homology modelling 24
2.5.1 Template search 24
2.5.2 Model construction 24
2.6 Model quality evaluation 25
2.7 Generation of the Complex HIV-1CZA IN Model 26
2.8 Selection criteria of variants 27
2.9 Stability predictions and the loss or gain of polar interactions 27
2.10 MD simulations 28
2.10.1 Preparation of MD simulation input files with CHARMM-GUI 28
2.10.2 Energy minimization 29

2.10.3 System equilibration 29
2.10.4 Production MD 30
2.11 Analysis of MD simulations 30
Chapter 3 Results 32
3.1 Demographic information 32
3.2 Multiple sequence alignment and mutation detection 32
3.3 Homology Modelling 35
3.4 Model quality evaluation 36
3.5 Generation of a DNA-MG-DTG complex 37
3.6 Interaction analysis 38
3.6.1 Interactions of DTG with IN active site 38
3.6.2 Selection of variants for analysis 39
3.6.3 Protein stability and polar interaction predictions 39
3.7 Production MD Simulations 40
3.7.1 Energy minimization of HIV-1CZA IN structures 41
3.7.2 Equilibration of HIV-1CZA IN structures 41
3.8 Simulation trajectory analyses 42
3.8.1 RMSD analysis 42
3.8.2 RMSF analysis 43
3.8.3 Radius of gyration 45
3.8.4 Hydrogen bond analysis 47
3.8.5 Pairwise distance analysis 49
3.8.6 Non-bonded pairwise distance analysis 50
3.8.7 Structural snapshot/conformational analysis 50
Chapter 4 Discussion and Conclusion 52
4.1 Discussion 52
4.2 Conclusion 55
4.3 Limitations of the work 56
REFERENCES 58
Appendices 76
Appendix A 76
Appendix B 77
Appendix C 78
Appendix D 79
Appendix E 80
Appendix F 81
Appendix G 82
Appendix H 87

LIST OF TABLES

Table 1: List of mutations present in SA's patients' samples who are suspected of failing HIV-1C treatment with RAL 33
Table 2: Scores for model quality evaluation parameters 36
Table 3: Stabilizing and destabilizing effects of selected variants on protein stability and assessment of changes in polar contacts before and after mutation 40
Table 4: Energy minimization results for the WT and P1-P4 complex mutant systems 41
Table 5: Non-bonded interaction energy between HIV-1C IN protein and DTG 50
Table 6: Summary of interaction analysis 51

LIST OF FIGURES

Figure 1.1: Schematic illustration of HIV-1 IN organization 4
Figure 3.1: MSA alignment of the 11 patient sequences against the HIV-1C_{ZA} IN consensus 34
Figure 3.2: 3D cartoon representation of the HIV-1C_{ZA} IN tetrameric structure predicted with SWISS-MODEL 35
Figure 3.3: Complete 3D structure of the HIV-1C_{ZA}-IN-MG-DNA-DTG complex homology Model 38
Figure 3.4: HIV-1C IN active site showing interaction with DNA, MG and drug DTG 39
Figure 3.5: Backbone RMSDs of the WT HIV-1C_{ZA} IN protein and its variants P1, P2, P3 and P4 at 300K are shown as a function of time 43
Figure 3.6: RMSF of the backbone atoms of WT HIV-1C_{ZA} IN protein and its variants P1, P2, P3 and P4 at 300K are shown 44
Figure 3.7: Rg of backbone atoms of WT HIV-1C_{ZA} IN protein and its variants P1, P2, P3 and P4 at 300K are shown as a function of time 46
Figure 3.8: Average intramolecullar protein-drug HBs amongst the WT HIV-1C_{ZA} IN, its complexes P1, P2, P3 and P4 and DTG at 300K for the last 50 ns 48
Figure 3.9: Average pairwise distance between MG ions and DTG in the WT and variant systems P1, P2, P3 and P4 at 300K 49

SCIENTIFIC CONTRIBUTION

Accepted publication

- **Chitongo R**, Obasa AE, Mikasi SG, Jacobs GB, Cloete R. Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding [published correction appears in PLoS One. 2020 Jun 5;15(6):e0234581]. *PLoS One*. 2020;15(5):e0223464. Published 2020 May 7. doi:10.1371/journal.pone.0223464

Work submitted for publication

- Sello Given Mikasi, Darren Isaacs, **Rumbidzai Chitongo**, George Mondinde Ikomey, Ruben Cloete, Graeme Brendon Jacobs. Interaction Analysis of Statistically Enriched Mutations Identified in Cameroon Recombinant Subtype CRF02_AG that can Influence the Development of Dolutegravir Drug Resistance Mutations. Under Review with Research Square. https://doi.org/10.21203/rs.3.rs-40608/v1

Conferences and workshops attended

- Attended the 2018 Center for High Performance Computing (CHPC) National Conference workshop which was held at the Century City Convention Centre, Cape Town, SA from 2 - 6 December 2018.

- 9th SA AIDS International Conference, Durban ICC, South Africa. Poster presentation: Structural impact of resistance associated mutations (RAMs) on Dolutegravir binding to HIV-1C Integrase, **Rumbidzai Chitongo**, Graeme Brendon Jacobs, Ruben Cloete

- 63rd Annual Academic Day Stellenbosch University Faculty of Medicine and Health Sciences (Tygerbeg campus): Oral presentation entitled Structural impact of selected Raltegravir resistance variants on Dolutegravir binding to South African HIV-1 Integrase subtype C protein: **Rumbidzai Chitongo**, Adetayo Emmanuel Obasa, Graeme Brendon Jacobs, Ruben Cloete

ABBREVIATION of TERMS

ART – antiretroviral therapy

ARV – antiretroviral

BIC – bictegravir

CAB – cabotegravir

cART – combination antiretroviral therapy

CCD – catalytic core domain

CTD – c-terminal domain

CPU – central processing unit

CryoEM – cryo-electron microscopy

DTG - dolutegravir

EVG - elvitegravir

GPU – graphics processing unit

HAART – highly active antiretroviral therapy

HIV-1 – human immunodeficiency virus type 1

IN - Integrase

INBI – Integrase binding inhibitor

INI – Integrase inhibitor

INSTI - Integrase strand transfer inhibitor

LTR – long terminal repeat

MD – molecular dynamics

NMR – nuclear magnetic resonance

NTD – n-terminal domain

OBT – optimized background therapy

PFV – prototype foamy virus

PR - Protease

RAL - raltegravir

RAM- resistance associated mutation

RNA – ribonucleic acid

RT – Reverse transcriptase

STC – strand transfer complex

TasP – treatment as prevention

tDNA – target DNA

TDR – transmitted drug resistance

vDNA – viral DNA

Chapter 1 Background and Literature Review

1.1 Background

To date, due to difficulties in engineering a cure for the Human Immunodeficiency virus (HIV) type 1 (HIV-1), being infected with the virus is considered a lifelong condition. However, it is possible to control and suppress the virus, thereby prolonging the life of infected individuals, keeping them healthy and dramatically lowering their chance of infecting others. This is possible with the help of medical care, through combination antiretroviral therapy (cART). South Africa (SA) has been reported to have the largest high-profile HIV-1 epidemic worldwide compared to other countries, having the largest population of HIV-1 infected individuals (7.7 million), more than any other country in the world as of 2019 (UNAIDS, 2019). However, the country's HIV prevention initiatives have been showing a significant impact on the population (Sidibé *et al.*, 2016). New HIV infections overall have fallen by half in the last decade, but, there are still too many reported cases. Since 2010, there has been a 49% decrease in new HIV infections reported and a 29% decrease on acquired immunodeficiency syndrome (AIDS)-related deaths (UNAIDS, 2019). SA accounts for approximately 19% of the global population of people living with HIV (PLHIV), 15% of the new infections and 11% of AIDS-related deaths (UNAIDS, 2019). According to Sidibé *et al.* (2016), SA accounts for a third of all the new HIV infections in Southern Africa.

The HIV-1 retroviral ribonucleic acid (RNA) genome encodes for three enzymes essential for virus replication: Reverse Transcriptase (RT), Protease (PR) and Integrase (IN). Of these, the HIV-1 IN, an indispensable viral protein responsible for the integration of the proviral DNA into the human genomic DNA in the cell nucleus, has proven to be a viable target for highly specific HIV-1 therapy (Brado *et al.*, 2018). Integrase strand transfer inhibitors (INSTIs) fall under the novel class of anti-HIV agents showing high activity in inhibiting replication of HIV-1 in patients (You *et al.*, 2016). A few INSTIs have been developed with others still under clinical trials, presenting an important advanced step in treatment of HIV-1 infected patients using a novel class of drugs. Some of the INSTIs that have been approved and licensed for treatment to date include Raltegravir (RAL), Elvitegravir (EVG), Dolutegravir (DTG), Bictegravir (BIC) and Cabotegravir (CAB). First appearing on the market in 2007, INSTIs have fast become staples of the anti-HIV-1 drug arsenal (You *et al.*, 2016). These drugs have played a critical role in HIV-AIDS therapy serving as main weapons in the arsenal for treating HIV-1 infected patients. They are highly potent antiretroviral agents with durable efficacy and minimal toxicity, completely active against viruses resistant to all other classes

of drugs and internationally approved (Heger *et al.*, 2016). The use of INSTIs has been ruled safe and effective for both treatment-naïve and treatment-experienced patients through clinical trials. In patients who could not use other therapeutic agents, INSTIs were observed to inhibit viral reproduction and accelerate plasma HIV-1 RNA level reduction. Moreover, according to You *et al.* (2016), INSTIs exhibit strong activity against multi-drug-resistant HIV-1 strains and have a synergistic effect when co-administered with other drug types.

RAL and EVG have been successful in clinical settings, being highly effective in HIV-1 treatment, but regardless of their potency, they have a relatively low genetic barrier to resistance in patients failing therapy. This has resulted in rapid emergence of resistance associated mutations (RAMs) for these drugs (McGee *et al.*, 2018). Furthermore, they share a high degree of cross-resistance, which necessitated the development of second-generation drugs of this class, that could retain activity against these resistant variants (Anstett *et al.*, 2017). Second-generation INSTIs such as DTG have a high genetic barrier for resistance as well as limited cross-resistance to RAL and EVG. At present, there is still scarcity of any long-term data regarding clinical experience with second-generation INSTI use. However, it has been reported that HIV-1 can become resistant to INSTIs during therapy (McGee *et al.*, 2018). It has been reported elsewhere that the resistance rates of DTG are lower than that of EVG which are lower than that of RAL, in head-to-head comparisons (You *et al.*, 2016). A few researchers have also reported that HIV-1 strains are group-resistant to RAL and DTG. These reports suggest that HIV-1 strains can become resistant to first and second-generation INSTIs, thereby affecting the clinical utility of INSTIs. However, compared to other types of antiretroviral drugs, INSTIs have generally displayed lower resistance rates.

1.2 HIV-1 Integrase

1.2.1 Activity

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) (Passos *et al.*, 2017). 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 (Mouscadet *et al.*, 2010). 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 (Malet *et al.*, 2014). 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 (Mouscadet *et al.*, 2010). During this

process the IN oligomerizes into a higher-order stable synaptic complex (SSC) containing two vDNA ends (Chen *et al.*, 2000). 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 (Passos *et al.*, 2017).

IN functions as a tetramer or higher-order oligomer (Rogers *et al.*, 2018). The enzyme catalyzes two reactions namely 3'-processing and strand transfer; for the successful covalent integration of vDNA into tDNA (Mouscadet *et al.*, 2010). During 3'-processing, the IN first binds to a short sequence located at either end of the long terminal repeats (LTR) of the vDNA. It catalyzes the removal of two nucleotide bases from the blunt-ended 3'end of each LTR strands of the vDNA, leaving a recessed 3'CA dinucleotide. After migration into the nucleus of the infected cell as part of a nucleoprotein complex, IN covalently attaches the 3'processed vDNA end to the host cell DNA and this process is called strand transfer (Maertens *et al.*, 2010). This second reaction occurs simultaneously at both ends of the vDNA molecule. Both reactions are divalent cation-requiring transesterification reactions catalyzed at a single active site (the DDE triad) within the protein's core structure (Chen *et al.*, 2000; Mesplède *et al.,* 2014).

1.2.2 Structure

The HIV-1 IN is a 288-amino acid (32-kDa) protein encoded by the C-terminal part of the *pol* gene of the HIV-1 genome (Mouscadet *et al.*, 2010). It is composed of three structurally independent domains (Cherepanov *et al.*, 2005) as presented in Figure 1. The N-terminal domain (NTD, residues 1-49) is a helical bundle containing a conserved 'HH-CC' zinc-finger motif that promotes protein multimerization. This domain is stabilized by the coordination of a single zinc atom (Cherepanov *et al.*, 2005). The central catalytic core domain (CCD, residues 50-212) contains the essential catalytic site characterized by three invariant essential acidic residues, D64, D116 and E152, collectively referred to as the canonical DDE triad/motif. This domain has been reported to be involved in DNA substrate recognition (Mouscadet *et al.*, 2010). The C-terminal domain (CTD, residues 213-288), which is the least conserved among retroviruses, contributes to non-specific DNA binding, stabilization of the IN-DNA complex and oligomerization necessary for the integration process (Chen *et al.*, 2000; Kessl *et al.*, 2009; Mouscadet *et al.*, 2010) and is linked to the catalytic core by residues 195-220, an extension of the final helix of the core domain.

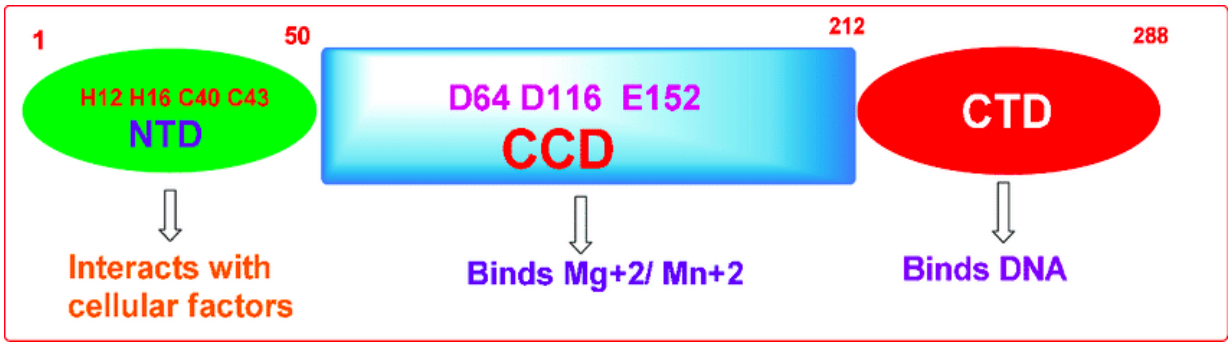


Figure 1.1: Schematic illustration of HIV-1 IN organization. The approximate positions of the different IN domains are shown with each domain coloured differently and the specific residues responsible for the functionality of each domain are highlighted (Wadhwa *et al.*, 2017).

The HIV-1 IN's CCD also belongs to a superfamily of DNA/RNA strand transferases/nucleases that in addition to retroviral INs includes divergent bacterial transposases (Cherepanov *et al.*, 2005). A flexible loop comprising of residues 140-149 (commonly known as the 140s loop region), is encompassed by the CCD for conformational changes that are required for 3'-processing and strand-transfer reactions (Mouscadet *et al.*, 2010). These activities have also been reported to require the presence of a metallic cofactor, which binds to the catalytic DDE triad residues. Each of the IN domains contributes to HIV-1 IN multimerization and is essential for 3'-processing and DNA strand transfer activities (Cherepanov *et al.*, 2005). HIV-1 IN has become a prime target for anti-HIV drug discovery because of the essential nature of integration in HIV-1 replication.

Several partial structural domains of HIV-1 subtype B (HIV-1B) IN have been solved in the past years and are readily available from the RCSB Protein Data Bank (https://www.rcsb.org/) Full intasome structures have also been determined for related retroviruses. An intasome is comprised of a nucleoprotein complex which is active in the integration of DNA from a bacteriophage into a host. Regardless of this development, a complete HIV-1 subtype C (HIV-1C) IN structure is still lacking. However, recently a full length three dimensional structure of the HIV-1B IN strand transfer complex (STC) intasome in 2017 (PDB ID:5U1C) has been determined by Cryo-electron microscopy (CryoEM) methods (Passos *et al.*, 2017; Rogers *et al.*, 2018). This structure (5U1C) has provided a first glimpse of nucleoprotein organization that could be used as a homolog model for the homology 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 resistance 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. However, crystal structures of the prototype foamy virus (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 (Maertens *et al.*, 2010; Hare, Maertens and Cherepanov, 2012; Rogers *et al.*, 2018).

1.3 Integrase Inhibitors

1.3.1 Development

Integrase inhibitors (INIs), which are currently considered as the novel antiretroviral (ARV) drugs in use, are designed to block the action of the HIV-1 IN Enzyme. Blocking the integration process of the HIV-1 virus implemented by the IN enzyme can halt any further spread of the virus. In terms of pharmacological development, two strategies have been considered in designing INIs (Mouscadet *et al.*, 2010). The first targets the free, unbound protein (3'-processing inhibitors), while the second targets the complex resulting from the association of IN with vDNA (strand-transfer inhibitors) (Park *et al.*, 2015). Specific Integrase strand transfer inhibitors (INSTIs) have turned out to be effective antiviral drugs *in vivo*. Another sub-class of INIs known as the integrase binding inhibitors (INBIs) is still experimental (Heger *et al.*, 2016). Due to the central and critical place of integration in HIV-1 replication, ARV drugs under the INSTI class have been approved for clinical use in HIV-infected patients. First appearing on the market in 2007, INSTIs have fast become staples of the anti-HIV-1 drug arsenal. INSTIs which are presently the novel class of ARV drugs for HIV-1, are not a cure for HIV infection, but are used as a means of preventing the HIV virus from multiplying within the host. This class of anti-HIV agents shows high activity in inhibiting replication of HIV-1 in patients (You *et al.*, 2016). Since these INSTI drugs only target a distinct step in the retroviral cycle, they may be taken in combination with other types of HIV drugs to minimize adaptation by the virus. It has also been reported elsewhere that even though INSTIs were engineered for the treatment of HIV infection, they could be used for other retroviruses (Brenner *et al.*, 2006).

1.3.2 Function of INSTIs

The rationale design of INSTIs has incorporated magnesium (Mg) chelating functions resulting from the important role Mg^{2+} plays in the catalysis mediated by IN (Grobler *et al.*, 2002; Mouscadet *et al.*, 2010). INSTIs operate by actively inhibiting the HIV-1 IN strand transfer reaction in order to block the integration of the vDNA into the host cell chromosomal DNA. The active site of the Integrase enzyme binds to the tDNA and it includes two divalent

metal cations that serve as chelation targets for INSTIs. But, when an INSTI is present, the active site of the enzyme is occupied and the integration process is halted (Shah *et al.*, 2014). The inhibitors bind to the IN associated with the ends of the vDNA, by chelating the metal cations present in the active site (Mouscadet *et al.*, 2010). These drugs successfully inhibit this integration process by selectively binding the cleaved stable synaptic complex (SSC) and interfering with the formation of the stand transfer complex (STC) (Passos *et al.*, 2017). INSTI molecules specifically inhibit the strand-transfer reaction due to (a) the chelation properties of Mg^{2+} cations involved in the active site of IN, (b) pi-stacking interactions between their halobenzyl groups and the LTR base immediately upstream of the 3'-CA 3'-OH and (c) interactions with specific IN amino acids (Malet *et al.*, 2014; Brenner *et al.*, 2016). Consequently, these drugs bind to the complex formed by IN and the vDNA (not IN alone) and specifically inhibit the strand transfer reaction by displacing the processed vDNA end, likely by competing with binding to target DNA (Malet *et al.*, 2014).

Due to extensive research efforts, some INSTI drugs have been discovered and developed. These include Raltegravir (RAL), Elvitegravir (EVG), Dolutegravir (DTG), Bictegravir (BIC) and Cabotegravir (CAB) (Hu and Kuritzkes, 2011). RAL, EVG, DTG and BIC have all been approved for clinical use by the United States Food and Drug Association (FDA) in 2007, 2012, 2013 and 2018, respectively (Rogers *et al.*, 2018). RAL and EVG are considered the first-generation INSTIs, while DTG, BIC and CAB are considered as second-generation INSTIs. These drugs have played a critical role in HIV-AIDS therapy, serving now as the current first-line weapons in the arsenal for treating HIV-1 infected patients. INSTIs have demonstrated their efficacy against HIV *in vitro* and in patients, and are particularly useful against virus strains that are resistant to other drug classes. The clinical use of INSTIs represented a milestone that appeared 10 years after cART was introduced to treat AIDS (McGee *et al.*, 2018). They have been reported to be active against both HIV-1B and non-B subtypes both in tissue culture and in patients (Mesplède *et al.,* 2014). INSTIs therefore exhibit strong activity against multi-drug-resistant HIV-1 strains and have a synergistic effect when co-administered with other drug types (Park *et al.*, 2015; Han *et al.*, 2016). Thus, in the context of disclosed therapeutic regimen, INSTIs have been prioritized as a first-line treatment for AIDS (Mondi *et al.*, 2019; Mikasi *et al.*, 2020). Additionally, INSTI-based regimens are favoured also for ART-naive HIV-1 patients, based on improved tolerability, better drug-drug interaction profiles, dosing (once or twice daily) and a high genetic barrier to resistance. Presently INSTIs are the primary option across most HIV-1 treatment guidelines, with the transmission of drug resistance to this class being infrequent (Parczewski *et al.*,

2017). However, regardless of the potency and effectiveness of INSTI drugs in combating the HIV-1 virus, reports of resistance genes are rapidly emerging about how the virus is becoming resistant to these drugs during therapy (Pham *et al.*, 2018; Cecchini *et al.*, 2019).

1.4 Resistance to INSTIs

Although considerable progress has been made in HIV therapy, drug resistance has always emerged for every available drug class, threatening long-term care and the potential for a cure (Anstett *et al.*, 2015). Using drugs that are generally well tolerated has led to far less drug resistance than was the case with most categories of antiretroviral compounds (Brenner and Wainberg, 2017). In multivariate analyses, increased viral suppression has been associated with better adherence to and the expanded use of INSTIs (Brenner *et al.*, 2016). Resistance to INSTIs emerges through the selection of mutations in the IN-coding region of the *pol* gene that affects susceptibility of the virus to the inhibitors (Mouscadet *et al.*, 2010). Drug resistance testing has played an effective role in selecting appropriate antiretroviral regimens and advancing life-long HIV-1 treatment. Brenner and Wainberg (2017) reported that the evolution of drug resistance over time is multifactorial, dependent on the fitness of the mutant variant, the level of phenotypic resistance that a specific mutation confers, the frequency of that mutation within a given viral population and the accumulation of secondary mutations. In addition, reports have been written suggesting HIV-1 can become resistant to INSTIs during therapy (Fourati *et al.*, 2014). More than 60 mutations have been specifically associated with resistance to INSTIs *in vitro* and *in vivo* (Mouscadet *et al.*, 2010). However, INSTIs have been observed to be less prone to transmitted drug resistance problems which are usually a result of a wide variety of substitutions in the HIV genome that can be sexually transmitted from one person to another (Anstett *et al.*, 2015).

The advantages of INSTIs have generated interest in whether their use could lead to a decline in the number of persistently-infected cells in viral reservoirs (Fourati *et al.*, 2014). The use of INSTIs in ART-naïve and ART-experienced patients raises a greater need to re-evaluate INSTI resistance in the context of large clinical settings (Fourati *et al.*, 2014; Anstett *et al.*, 2015). As the use of INSTIs is increasing in clinical practice, drug resistance to different HIV-1 subtypes have to be carefully monitored and investigated (Han *et al.*, 2016). Understanding the pathways implicated in the evolution of drug resistance is a requirement for the long-term efficacy of INSTIs (Brenner and Wainberg, 2017). Signals of HIV-1 strains that are resistant to first and second-generation INSTIs have emerged, affecting the clinical utility of INSTIs.

But, relative to other types of ARV drugs, INSTIs have generally displayed lower resistance rates (You *et al.*, 2016).

1.5 Resistance Pathways of 1st and 2nd Generation INSTIs

1.5.1 First-generation INSTIs

RAL is now available for more than 10 years and determinants of genotypic resistance are well-described for this drug, and for EVG (Charpentier and Descamps, 2018). Research outputs have reported that RAL and EVG can develop resistance relatively fast, both *in vitro* and in patients who are experiencing treatment failure. These first-generation INSTIs have been found to have a relatively low genetic barrier to resistance, since only a few mutations (1 or 2) in HIV-1 IN are capable of greatly reducing virus susceptibility to either of them. Regardless of their strong efficiency in inhibiting HIV-1 replication, RAL and EVG's low genetic barrier to resistance has led to common resistance associated mutations (RAMs) that threatened the long term efficacy of highly active antiretroviral therapy (HAART). Drug resistance has been selected among virologically failing patients treated with agents characterized by the low genetic barrier to resistance (RAL and EVG) following the introduction of these INSTIs into clinical practice (Wohl *et al.*, 2014; Parczewski *et al.*, 2017).

HIV-1 treatment with RAL where the viral load is not optimally suppressed during treatment have resulted in the development of RAMs against this INSTI, and these RAMs vary on the HIV-1 subtype (Mboumba Bouassa *et al.*, 2019). The effect of RAL on different subtypes has been considered a critical issue for analysis, because patients undergoing this treatment might be at risk of failing INSTI based cART. At least three distinct, but not exclusive, genetic pathways defined by two or more mutations, have been reported from clinical trials to be associated with RAL failure in RAL-experienced patients. These genetic resistance pathways include a signature/major mutation at Y143C/H/R, Q148H/K/R or N155H located in the CCD of IN, together with one or more additional associated secondary/minor mutations at L74M, E92Q, T97A, E138A/K or G140S) that resulted in higher levels of resistance (Hu and Kuritzkes, 2011). These pathways are associated with virological failure and reduced susceptibility to RAL (Shah *et al.*, 2014). The most common mutational pattern is the Q148H in combination with G140S, which confers the greatest loss of drug susceptibility. RAL mutations described in the N155H pathway include this major mutation in combination with minor mutations; L74M, E92Q, T97A, Y143H, G163K/R, V151I, or D232N. Most N155H mutants tend to predominate early in the course of RAL failure, but are gradually replaced by viruses with higher resistance, often bearing mutations G140S and Q148H/R/K, with continuing RAL treatment (Wensing *et al.*, 2017). The Y143R/H/C mutation is uncommon.

Residues associated with catalytic activity and primary resistance to RAL have been found to be highly conserved among all subtypes, whereas variations in residues associated with secondary resistance have been observed through research (Han *et al.*, 2016). Hu and Kuritzkes (2010) have also reported that mutations that confer resistance to RAL are associated with reduced viral fitness when compared to wild-type virus in the absence of drug (Hu and Kuritzkes, 2010). Resistance to RAL can emerge in the setting of incomplete viral suppression as with other antiretroviral drugs (Hu and Kuritzkes, 2010).

EVG specific resistance pathways involve IN mutations at codons T66I/A/K, E92Q/G, and S147G (Doyle *et al.*, 2015; Theys *et al.*, 2015; Brenner and Wainberg, 2017). Mutation E92Q alone reduces susceptibility to EVG by more than 20 fold and causes limited (< 5 fold) cross-resistance to RAL. For first-generation INSTIs, RAMs via the N155H and G140A or G148R/H/Q pathways will confer cross-resistance to RAL and EVG. Polymorphisms and/or secondary integrase RAMs can also increase resistance and cross-resistance to RAL and EVG (Brenner *et al.*, 2011; Doyle *et al.*, 2015; Brenner and Wainberg, 2017). RAL and EVG treatment have suffered an extensive cross-resistance of mutations, highlighting the importance of developing new, improved INSTIs.

1.5.2 Second-generation INSTIs

Second-generation INSTIs, DTG, BIC and CAB, all possess the quality of higher genetic barriers and low likelihood of resistance selection among treatment-naïve cases (Tsiang *et al.*, 2016; Brenner and Wainberg, 2017; Parczewski *et al.*, 2017). These drugs possess a more robust resistance profile than relative to first-generation INSTIs (Mesplède *et al.,* 2014). Currently, most of the research with regards to drug resistance and potency of second generation INSTIs has been conducted on DTG. Very little is known for BIC and CAB.

DTG can be used as a first-line cART in treatment-naïve patients and can also be used by treatment-experienced patients, including those who have been treated with other INSTIs (Osman *et al.*, 2015). The drug was approved for use in a broad population of HIV infected patients due to its effectiveness. Generally, it is believed that the higher genetic barrier of DTG to HIV-1 resistance is partially due to its slow dissociation rate from IN-DNA complexes compared to RAL and EVG (Mesplède *et al.*, 2015; Wainberg and Han, 2015; Han *et al.*, 2016). On the other hand, cross-resistance in patients failing RAL and EVG based regimens may occur, even though DTG would most likely retain at least partial activity against most variants (Anstett *et al.*, 2015; Tsiang *et al.*, 2016; Parczewski *et al.*, 2017).

A general specific pathway of resistance mutations against DTG and/or BIC has not been set out but there are several RAMs that have been selected *in vitro* and *in vivo* against these two INSTIs. Mutation H51Y has been selected *in vitro,* while T66I and S147G were selected *in vivo* by DTG, but, they do not affect the susceptibility of the drug (Abram *et al.*, 2013; Liang *et al.*, 2015). On the other hand, mutations T66K, E92Q, G118R, S230R and R263K have been reported to reduce DTG susceptibility on their own, either *in vitro* or *in vivo* (Kobayashi *et al.*, 2011; Cahn *et al.*, 2013; Vavro *et al.*, 2013; Wijting *et al.*, 2017). Other mutations such as L74M/F, T97A, E138K/A, Q148H/K, G149A and N155H only reduce the susceptibility of DTG and BIC when in combination with other INSTI RAMs, especially at codons 138, 140 and 148 (Abram *et al.*, 2013; Cahn *et al.*, 2013; Eron *et al.*, 2013; Vavro *et al.*, 2013; Hardy *et al.*, 2015; Naeger *et al.*, 2016; Wijting *et al.*, 2017). N155H has been selected by DTG in both INSTI-naïve and RAL-experienced patients. There has been reports of virological failure for patients on DTG harbouring N155H mutation at baseline and the development of additional INSTI RAMs (Hardy *et al.*, 2015; Wijting *et al.*, 2017). According to You *et al.* (2016), the resistance rates of DTG are still lower than those of EVG which are lower than those of RAL, in head-to-head comparisons.

1.6 Antiretroviral Therapy and Transmitted Drug Resistance

Combination antiretroviral therapy (cART) has dramatically reduced HIV infection to a chronic and manageable disease resulting in a near-normal life expectancy (Brenner and Wainberg, 2017; Brado *et al.*, 2018). The quality of life for HIV-1 infected individuals has been significantly improved by the development of cART (Osman *et al.*, 2015). Outstanding advances in cART in the past 4 decades, has markedly improved the prognosis of HIV-1 infection and has resulted in a major decline in HIV-1 related morbidity and mortality thus improving the quality of life, particularly when used in the early stages of HIV infection (Brenner *et al.*, 2016). This has been achieved by the programme contributing to treatment success at the individual level and also reductions in rates of HIV transmission at the population level (Wainberg, Mesplède and Raffi, 2013). Research has reported that the global scale-up of cART has resulted in approximately 25 – 50% decline in generalized HIV epidemics in the middle and low income settings (UNAIDS, 2019). Such findings have raised optimism that Treatment-as-Prevention (TasP) may arrest the global pandemic by 2020 (Brenner and Wainberg, 2017). In an effort to reach this goal, the World Health Organization (WHO) has set ’90-90-90’ targets where 90% of all people living with HIV know their status, 90% of those infected are treated with cART and 90% viral suppression is achieved among those on ART (UNAIDS, 2019).

With more asymptomatic patients being initiated to cART, treatment goals now require more robust and durable first-line regimens. Against this background, the addition of new classes of drugs is important to limit the emergence of resistant strains (Mesplède *et al.,* 2015). Generally, the optimal choice for salvage therapy for HIV-1 infected patients has been shown to require at least 2, and preferably 3, fully active drugs (Yazdanpanah *et al.*, 2009). Standard cART uses a combination of at least 3 ARV drugs that includes agents from more than 1 class (Han *et al.*, 2016; Brenner *et al.*, 2016). Since the introduction of triple ARV therapy, the rates of success of therapy as indicated by suppression of plasma viremia to levels below a cut-off of 50 copies of viral RNA/ml, have increased to almost 90%. This has happened for several major reasons (Brenner *et al.*, 2016):

(a) The drugs used in therapy are now more potent and have longer half-lives than the compounds that were in use 2 decades ago.

(b) Dosing regimens have become simplified, often because of the use of co-formulations, some of which only need to be taken once-daily, and this has greatly enhanced rates of adherence to ARV regimens.

(c) Drug regimens have become far less toxic and more tolerable over time, and this has also promoted adherence as well as diminished the likelihood of development of HIV drug resistance against the components of ARV regimens.

Current cART paradigms typically include two nucleoside reverse transcriptase inhibitors (NRTIs) as part of a regimen backbone with a third core agent from one of the non-nucleoside reverse transcriptase inhibitor (NNRTI), boosted protease inhibitor (PI) or INSTI drug classes (Brenner and Wainberg, 2017). Use of INSTIs is increasing in cART, even though these drugs are not all equal regarding genetic barrier to resistance (Charpentier and Descamps, 2018). HIV-1 INSTIs are the most common drug class recommended by the different international guidelines as a third agent of cART (Charpentier and Descamps, 2018). The INSTI-based regimens are usually favoured for cART-naïve HIV-infected patients, based on improved tolerability, better drug-drug interaction profile, dosing (once or twice daily), excellent side-effect profiles and high genetic barrier to resistance (Miller *et al.*, 2015; Brenner and Wainberg, 2017). Additionally, these drugs were also shown to be effective in cART-experienced HIV infected patients harbouring viruses resistant to other drug classes (Mesplede *et al.*, 2014). However, the widespread introduction of the INSTIs into clinical practice presents a worry about transmission of drug resistance to this class of ARV medications (Parczewski *et al.*, 2017).

Levels of transmitted drug resistance (TDR) vary widely, depending on many factors, such as the population studied, availability, access to cART, duration of the treatment programs, quality of medical care, risk-taking behaviours, mode of transmission, and viral subtype (Bennett *et al.*, 2008). In most developed country settings, the use of mono therapy and dual therapy in the pre-highly active antiretroviral therapy (HAART) era (1980s and 1990s) resulted in rapid emergencies of TDR (Manasa *et al.*, 2012). Inzaule *et al*., (2018) suggested that HIV-1 TDR could reverse the gains of antiretroviral rollout. He also reported that recent evidence suggests that TDR is decreasing with the current levels ranging between 5% and 15% in Europe (2007) and 10% and 18% in the United States (Little *et al.*, 2008; Davy-Mendez *et al.*, 2019). Since transmitted resistance threatens to seriously limit future therapeutic options, North American and European treatment guidelines recommend resistance testing for all treatment-naive patients prior to the initiation of cART (Manasa *et al.*, 2012). This route provides an opportunity to avoid ineffective drug combinations and allows for individualized optimization of first-line cART (Mboumba Bouassa *et al.*, 2019).

1.7 Transmitted Drug Resistance in South Africa and Africa

Until recently, most clinicians in Africa, treating patients infected with multidrug-resistant virus generally added only 1 new agent to the optimized background therapy (OBT) regimens (Yazdanpanah *et al.*, 2009). This approach, often necessary because of limited drug options, put patients at high risk of virological failure and resistance to the new drug agents, as well as to other agents in the same class (Yazdanpanah *et al.*, 2009). Virological failure occurs when HIV therapy fails to suppress and sustain a patient's viral load to less than 200 copies/ml. Several factors such as drug resistance, drug toxicity and poor adherence to cART, can contribute to virological failure (Cecchini *et al.*, 2019; Puertas *et al.*, 2020). In the early days of the development of cART access programs in Africa, concerns have been expressed that the rapid massive scale-up of cART would lead to the widespread emergence and transmission of HIV-1 drug-resistant viruses (Mills *et al.*, 2006; Manasa *et al.*, 2012). Thus, the HIV-1 TDR could significantly reverse the gains of antiretroviral rollout, particularly in Africa. These concerns were largely based on challenges associated with early treatments (high pill burden, frequent dosing, toxicity) as well as financial constraints that would cause drug 'stock-outs' and prohibit essential laboratory monitoring. Pooled data obtained before the widespread roll-out of cART programs globally showed higher adherence levels in Africa compared to those in North America, even though some programs in sub-Saharan Africa did experience difficulty maintaining constant supplies of drugs. However, regardless of the inclined

adherence levels being maintained following the widespread roll-out of ART, reasons such as missed drug pick-ups due to family obligations and travel costs have been the leading cause of viremia and the selection of drug resistance in Africa (Manasa *et al.*, 2012). According to Manasa et al. (2012), formalized cART became available in the public health systems of most African countries only after 2002/2003 with the introduction of first-line treatment with combination drug regimens consisting of two nucleoside combined with one NRTIs, NNRTIs. These simple and affordable treatments have been effective up until now where levels of primary resistance have remained low (< 5%). However, in other reports, several studies in southern Africa (Lusaka, Durban, Cape Town) and east Africa have reported TDR rates ≥ 5% (Price *et al.*, 2011). There has been no significant evidence reported on TDR in rural KwaZulu-Natal so far (Price *et al.*, 2011). These studies were all conducted in large urban centres, either in young primagravidas attending antenatal clinics (ANCs) or in treatment-naïve individuals starting ART (Price *et al.*, 2011; Manasa *et al.*, 2012).

To ensure that current first-line therapies remain effective, TDR levels in recently infected treatment-naive patients need to be monitored (Inzaule *et al.*, 2018; Davy-Mendez *et al.*, 2019). The World Health Organization (WHO) has recommended resistance genotyping of remnant specimens collected from recently infected treatment-naïve individuals, less than 25 years of age primagravida women, individuals consecutively diagnosed with HIV-1 in seroprevalence surveys, or individuals who have laboratory confirmed evidence of recent infection, in all resource-constrained countries (Bennett *et al.*, 2008). TDR levels in South Africa have remained low following a downward trend since 2003 (Price *et al.*, 2011; Manasa *et al.*, 2012). Regardless, there is still a great need for continuous vigilance in the monitoring TDR, as more patients are initiated and maintained onto cART on a daily basis.

1.8 Molecular Dynamic Simulations

1.8.1 Theory

Molecular dynamic (MD) simulation techniques plays an important role in predicting and understanding the properties, structure and function of molecular systems (Braun *et al.*, 2019). Simulation methods are useful for studying structure and dynamics in complex systems that are too complicated for pen and paper theory, aiding in interpreting experimental data in terms of molecular motions (Braun *et al.*, 2019). MD simulations is a powerful computational method widely applied in structural biology for delineating motions of proteins at an atomic-scale via theoretical and empirical principles in physical chemistry (Dodson, Lane and Verma, 2008; Ode *et al.*, 2012). These MD simulations allow for the exploration of the

conformational energy landscape accessible to these molecules (Karplus and Kuriyan, 2005). They represent a well-established method for modelling conformational changes within the macromolecules, the characterization of which provides insight into the workings of bio-molecular systems at spatial and temporal scales that are difficult to access experimentally (Klepeis *et al.*, 2009). The precision and performance of MD has rapidly improved due to recent advances in the hardware and software for bio-molecular simulations. Simulations can now be run on central processing unit (CPU) clusters and graphics processing units (GPUs) and sophisticated algorithms are also being applied in simulating bio-molecules (Lemkul, 2018). Consequently, MD simulations are also quickly extending the range of its applications in biology, helping to reveal unique features of protein structures that would be hard to obtain by experimental methods alone such as macroscopic thermodynamic properties of a protein (Ode *et al.*, 2012). MD simulations can also provide detail concerning individual atomic motions as a function of time; thus, they can be used to answer specific questions about the properties of a model system, often more readily than experiments (Karplus and Kuriyan, 2005).

1.8.2 General Applications of MD Simulations

MD simulations have been applied as one of the major research tools to study wide ranges of biological systems, bridging the gap between biological mechanism, nuclear magnetic resonance (NMR) structures or X-ray crystallography (Kim, Kasprzak and Shapiro, 2017). These simulations have been applied in the field of protein chemistry since 1977. Since then, the performance of this technique has quickly improved both quantitatively (can be applied to a wide group of various systems in biology and chemistry) and qualitatively (more parameters/meaningful information can now be derived from MD simulations) along with the rapid advances in hardware and software on bio-molecular simulations (Lindorff-Larsen *et al.*, 2012; Ode *et al.*, 2012). All results of MD simulations are critically influenced by the force fields (Lindorff-Larsen *et al.*, 2012). In a simple molecular system, all atoms and covalent bonds connecting the atoms are assumed to be the charged spheres and springs, respectively. Parameters of mathematical functions describing the potential energy of a system, termed the “force field,” are set to simulate the movements of atoms and molecules. Potential energy refers to the energy stored in the bonds connecting atoms in molecules (Lodish *et al.*, 2000). Qualities of parameters in the force fields, especially for dihedrals and electrostatic potentials, have been improved quantitatively and qualitatively over time by introducing improved approximation to the quantum ground-state potential energy surface (Ode *et al.*, 2012). Dihedrals also called phi and psi, are torsion angles between two bonds

(Chakrabarti and Pal, 2001). Gupta defined electrostatic potentials in molecules as the energy of interaction of a positive point charge with a molecule (Gupta, 2016). Moreover, explicit introduction of effects of solvation has contributed to the qualitative improvement for the precision and performance of MD simulations (Adcock and McCammon, 2006). However, there is still a need to improve the force field parameters, especially for divalent ions such as magnesium ions (Mg^{2+}) or manganese ions (Mn^{2+}) that are involved in the catalytic activity of the HIV-1 IN catalytic core domain. Several force fields that are commonly used in MD simulations include AMBER (Wang *et al.*, 2004), CHARMM (Brooks *et al.*, 1983) and GROMOS (Christen *et al.*, 2005).

1.8.2.1 Applications of MD Simulations in Biology

Proteins fluctuate spontaneously in solution. Research evidence has indicated that such fluctuations play key roles in the specific functions of proteins, such as catalytic reactions of enzymes (Henzler-Wildman *et al.*, 2007; Abbondanzieri *et al.*, 2008) or interactions with other biomolecules (Thorpe and Brooks III, 2007). Multiple experimental methods are available to characterize the protein's dynamic properties, but, it is usually difficult to delineate motions of proteins at an atomic scale (Ode *et al.*, 2012). MD simulations are computational methods applied to address the above issue. This technique allows us to calculate movements of atoms in a molecular system, such as proteins in water, by numerically solving Newton's equations of motions (Adcock and McCammon, 2006). Frequently used force fields for proteins, such as the "AMBER" (Case *et al.*, 2005) and "CHARMM" (Brooks *et al.*, 2009) force fields, have explicitly included; covalent bonds, angles, dihedrals, van der Waals, and electrostatic potentials within the formulae.

MD simulations have three major applications within the structural biology field. First, it is useful for the refinement of experimentally determined three-dimensional (3D) protein structures (Autore *et al.*, 2010; Özen *et al.*, 2011). Second, it is beneficial in constructing previously undescribed 3-D protein structures in combination with homology modelling techniques (Baker and Sali, 2001) provided a closely related experimental homolog structure is available. Third and most importantly, the technique provides a unique tool to address the structural dynamics of proteins, that is, the time evolution of conformations in solution, at timescales of nanoseconds to microseconds (Henzler-Wildman and Kern, 2007; Dror *et al.*, 2010). Structural snapshots obtained during MD simulations are very helpful in depicting the unique structural features of proteins (Karplus and Kuriyan, 2005; Dodson, Lane and Verma, 2008).

1.8.2.2 Applications of MD Simulations in Virology

Viral enzymes are essential for viral replication making them important targets for anti-viral drug development (Ode *et al.*, 2012). MD simulations have been used to study the basis of the structural dynamics that allow the HIV-1 viral enzyme Reverse-Transcriptase (RT) and its drug to function properly (Zhou *et al.*, 2005; Kirmizialtin *et al.*, 2012). MD simulations have also been applied in studying the reduction in binding affinity of the NRTIs to HIV-1 RT (Carvalho, Fernandes and Ramos, 2006, 2007). Simulations studies performed on the HIV-1 PR enzyme and its substrates have helped to clarify how the viral precursor is processed during viral maturation (Hornak *et al.*, 2006; Deng *et al.*, 2011). Moreover, with antiviral drug resistance becoming a major clinical problem for treatment of virus-infected individuals, MD simulations are used to examine how viral mutations account for drug resistance at the atomic level (Cortez and Maldarelli, 2011). Viral resistance to antiviral drugs is primarily caused by genetic mutations that eventually lead to a reduction in the drug binding affinity to viral target proteins. Using MD simulations, it was observed that a reduction in the binding affinity of the PR inhibitors (PIs) to HIV-1 PR can be as a result of a reduction in hydrophobic interactions (Kagan *et al.*, 2005; Sadiq, Wan and Coveney, 2007; Chen *et al.*, 2010), a reduction in electrostatic interactions (Ode *et al.*, 2007; Chen *et al.*, 2010), changes in flexibility at the flap of the PR (Foulkes-Murzycki *et al.*, 2007) and/or changes in the shape of the inhibitor-binding pocket (Ode *et al.*, 2006, 2007). MD simulations are also used to study how genetic differences of the HIV variants around the world can influence the efficacy of antiviral inhibitors (Ode *et al.*, 2007; Matsuyama *et al.*, 2010; Kar and Knecht, 2012). Thus, MD simulations is a valuable tool to assist in studying drug efficacy when genetic information on drug target proteins is available (Sadiq, Wan and Coveney, 2007).

1.8.3 Limitations of MD Simulations

Besides the successful applications of MD simulations, the utility of MD simulations is still limited by two principal challenges: (i) force fields require further refinement to describe all paramaters in a biological system, and (ii) high computational demands prohibit routine simulations greater than a microsecond in length, in many cases resulting in an inadequate sampling of conformational states (Durrant and McCammon, 2011). Currently, MD simulations only allow us to investigate protein structural dynamics on timescales of nanoseconds to microseconds (Henzler-Wildman and Kern, 2007; Dror *et al.*, 2010). Since the processing speed of computers is doubling approximately every two years according to Moore's law, MD simulations studies will probably be extended to simulations of larger and more complex system at longer timescales of milliseconds in the near future. This will then

lead to a better understanding of the structures and dynamics of macromolecules involved in virus–host interactions. However, there is still a huge limitation to the application of MD simulations especially in countries of low-income. MD simulations require higher-performance-computing systems and in most African countries accessing such resources is challenging.

Furthermore, the force fields used are also approximations of the quantum-mechanical reality that reigns in the atomic regime (Durrant and McCammon, 2011). While simulations can predict many important molecular motions, they cannot accurately approximate the quantum effects when transition metal atoms/ions are involved in binding. To overcome this challenge, some researchers have introduced quantum mechanical calculations into classic molecular-dynamics force fields. With such calculations, the motions and reactions of enzymatic active sites or other limited areas of interest are simulated according to the laws of quantum mechanics, and the motions of the larger system are approximated using molecular dynamics. While far from the computationally intractable 'ideal' of using quantum mechanics to describe the entire system, this hybrid technique has nevertheless been used successfully to study a number of systems such as enzymatic and photobiological systems (Durrant and McCammon, 2011; Groenhof, 2013). Photobiological systems are those that undergo chemical and physical changes as a result of light or any non-ionising radiation (Groenhof, 2013).

1.9 Computational Studies of HIV-1 IN Drug Resistance

Computational methods involving comparative protein structure modelling, molecular docking and MD simulations, have provided fundamental support to HIV research since the initial structural studies, helping to unravel details of HIV biology (Forli and Olson, 2015). Moreover, computational protein models have proven to be useful in analysing and understanding the impact of mutations as well as developing drugs to overcome the structural and functional influence of mutations associated with drug resistance. With the availability of structural data, *in silico* experiments have also been instrumental in exploiting and improving interactions between drugs and viral protein targets, such as HIV-1 IN. Issues, such as viral target dynamics and mutational variability, as well as the role of water and the estimates of binding free energy in characterizing ligand interactions, are areas of active computational research (Forli and Olson, 2015).

MD simulations have been extensively applied in the study of HIV-1 IN, particularly in drug discovery to search for more potent IN inhibitors, understanding molecular mechanisms of

how HIV-1 IN mutations confer drug resistance. Several ground-breaking findings have been achieved through MD simulations in understanding drug resistance within HIV-1 IN. Xue *et al*. (2012) investigated the molecular mechanism of interactions between an HIV-1 IN-vDNA complex and the inhibition action of RAL. In their study, Xue and team were able to show that RAL did not influence interactions between vDNA and HIV-1 IN, but rather the drug targeted a special conformation of the IN to compete with host DNA blocking the function of HIV-1 IN. In a follow up study, in an effort to understand how HIV-1 IN mutations actually confer drug resistance, Xue and his team employed MD simulations to uncover the molecular mechanism behind the resistance of HIV-1 IN to RAL (Xue *et al.*, 2013). A detailed analysis of the results of MD simulations suggested that the Tyrosine143 located in the 140s loop (i.e. residues Glycine140 to Glycine149) is a key anchoring residue that leads to stable RAL binding. A decrease in the interactions at this residue was reported as one of the key reasons responsible for the resistance of HIV-1 IN to RAL. These studies by Xue *et al*. (2012, 2013) provided a structural and energetic understanding of the RAL-resistant mechanism at the atomic level and provided some clues on how to design new drugs that may circumvent the known resistance mutations. Another research group (Horlando Sampaio Araujo da Silva *et al.*, 2018) aimed to predict novel HIV mutations related to DTG resistance using *in silico* approaches performed MD simulations calculating the binding energy along the time-course in a bid to establish a framework of searching for new HIV drug-resistant mutations. They found that the energy of the Y226K mutant complex showed less binding affinity than other mutated complexes compared to the WT. The variant (T226K) was shown to impede the attachment of DTG to INT, indicating this mutant as a possible resistance mutation. MD simulations were also applied to understand the binding mode of the EVG ligand in E92Q/N155H HIV-1 IN double mutant system (Chen *et al.*, 2015). It was observed that even though EVG formed an interaction with the binding site Mg^{2+}, binding affinity decreased relative to the WT. Chen and colleagues also observed interactions being formed between the mutated residues and active site residues in the E92Q/N155H double mutant structure and additionally there were structural changes of the 140s' loop region (Chen *et al.*, 2015). They concluded that such structural changes would eventually cause functional changes of the binding pocket leading to HIV-1 drug resistance (Chen *et al.*, 2015). These results could be helpful for understanding the integration process of the HIV-1 virus and provide some new clues for the rational drug design and discovery of potential compounds that would specifically block HIV-1 virus replication. Ever-increasing computational resources, theoretical and improved algorithms have played a significant role in progress to date, and we

envision a continually expanding role for computational methods in our understanding of HIV biology and structure-based drug design (SBDD) in the future (Forli and Olson, 2015).

1.10 Rationale of Study

HIV-1C accounts for nearly 50% of all global HIV-1 infections, while HIV-1B accounts only for approximately 12%, and approximately 95% of HIV-1 patients in South Africa (SA) are infected with HIV-1C (Wilkinson *et al.*, 2015). South Africa has the highest burden of HIV worldwide compared to any other country, having the largest population of HIV infected individuals (7.7 million) as of 2018 (UNAIDS, 2019) (Levi *et al.*, 2016). In the recent years, South Africa has made great strides in tackling its HIV epidemic. This has resulted in SA being recognized as the one sub-Saharan country with the largest cART programme in the world. On a global scale, SA's antiretroviral (ARV) treatment programme accounts for 19% of people on cART. Although cART is potent and life prolonging for individuals infected with HIV-1, the long-term efficacy of cART is limited by drug resistance. This problem exists in both resource-limited and developed countries. But, a vast majority of research on HIV-1 infections, effect of ARV drugs and understanding any drug resistance to the cART drugs in use has been extensively conducted on HIV-1B, with less information known about HIV-1C. As a result of the extensive research conducted on HIV-1B, all ARVs have been largely developed in relation to subtype B. These drugs have been reported to be effective against a wide range of HIV-1 subtypes (Geretti *et al.*, 2009). However, other studies have revealed very poor cART outcome when associated with HIV-1C infections (Brenner *et al.*, 2006; Häggblom *et al.*, 2016; Sutherland *et al.*, 2016). Although subtype C has not been considered an effective predictor for therapy failure earlier, a recent trial indicated that HIV-1C has independent predictors for viral failure (Sutherland *et al.*, 2016). Recent studies also have identified subtype specific differences in DTG cross-resistance pattern in patients failing the first-generation RAL treatment (Doyle *et al.*, 2015; Brado *et al.*, 2018).

Additionally, the lack of a complete IN protein structure has negatively impacted the progress on structural studies of nucleoprotein reaction intermediates. The mechanism of HIV-1 viral DNA's integration has been studied extensively at biochemical and cellular levels, but not at a molecular level. Developing multi-domain structures of the HIV-1C IN would provide insight into several areas including the relative spatial arrangement of the three IN domains and showing how IN binds to the host and viral DNA ends (Chen *et al.*, 2000). Preliminary data published from a recent study by our collaborators in the Division of Medical Virology, Stellenbosch University, Tygerberg has laid the ground work for this study. In their work they

have characterized the HIV-1 IN gene in a South African context, identified 17 natural occurring polymorphisms and no RAMs in treatment-naïve cART patient samples (Brado *et al.*, 2018). Additional, genotypic work on HIV-1C IN identified RAMs that may confer resistance to RAL and also subtype-specific differences that could account for DTG cross-resistance patterns in SA patients failing RAL treatment as previously reported by (Doyle *et al.*, 2015). This emphasizes the need for further interrogation of the RAL identified variants to determine their possible effect on DTG drug binding. Accurate deleterious prediction for non-synonymous variants within the genetics community is crucial for distinguishing pathogenic mutations from background polymorphisms in whole exome sequencing (Dong *et al.*, 2015). Although most laboratories use online prediction tools such as SIFT, PROVEAN or MutationTaster, either alone or in combination, in order to predict pathogenicity; their prediction results are sometimes inconsistent with each other and their relative merits are still unclear in practical applications (Leong *et al*., 2015). These prediction tools are sequence based approaches that do not take the energy change of the protein structure into account. This has presented one of the greatest challenges in whole exome sequencing studies on the ability to distinguish pathogenic mutations from a large number of background variations. Structural approaches allow us to take the immediate amino acid environment into account of the three dimensional protein and simulate different structural conformations. Robust structural *in silico* web servers such as DUET, mCSM and SDM, are computational statistical potential energy functions that employ environment-specific amino-acid substitution frequencies within homologous protein families to calculate a stability score, which is analogous to the free energy difference between the wild type and the mutant protein. Moreover, these *in silico* web servers study the effect of missense mutations by relying on graph-based signatures that encode distance patterns between atoms and are used to represent the protein residue environment. It is for these reasons that the DUET, mCSM and SDM *in silico* web servers are considered superior relative to the pre-mentioned online prediction tools. As a result, this study will apply a structural approach to this problem trying to show the superiority of a structural approach (due to the ability to visualize and asses the effect of each variant on the structure of the protein) to making use of online algorithm prediction tools alone.

1.11 Aim of the Study

This study aims to use *in silico* methods that involves molecular modeling, interaction analyses and molecular dynamics simulation studies to prioritize mutations that will affect HIV-1C IN drug binding to DNA and to DTG. The purpose here is to help tailor more effective personalized treatment options for patients living with HIV in South Africa. This

study will in part use patient derived sequence data to identify mutations and model them into the protein structure to understand their impact on the HIV-1C IN protein structure folding and dynamics.

1.12 Objectives of the Study

a. Perform multiple sequence alignment of HIV-$1C_{ZA}$ sequences from South African HIV-1 infected treatment-experienced patients and identify all the resistance pathways.
b. Perform structural homology modelling of the HIV-$1C_{ZA}$ IN protein to generate a wild-type (WT) structure of the protein.
c. Extract DTG and DNA from homologous structures to construct an HIV-1C IN-DNA-DTG complex.
d. Determine stabilizing and/or destabilizing effects of all recorded mutations/variants from the South African cohort, on the HIV-$1C_{ZA}$ IN structure using mCSM webserver.
e. Identify and select mutations that might affect protein-DNA and drug binding and perform structure preparation for simulation studies.
f. Simulate to determine the effects of the variants/mutations on the HIV-$1C_{ZA}$ IN protein structure in comparison to the WT structure.
g. Determine the loss or gain of crucial interactions between the protein-DNA and the drug DTG.
h. Calculate the change in non-bonded interaction energies for different protein conformations to determine stronger and weaker interaction between the protein and drug.
i. Interrogate structural snapshots/conformations for simulation systems to confirm DTG, MG and DDE motif residue interactions

Chapter 2 Materials and methods

2.1 Ethical consideration

Data published from a study by our collaborators in the Division of Medical Virology, Stellenbosch University, laid the ground work for this study. In their work they characterized the HIV-1 IN gene in a South African context, identified 17 natural occurring polymorphisms and no resistance associated mutations (RAMs) in treatment naïve cART patient derived HIV sequences (Brado *et al.*, 2018). The ongoing study has received ethical approval from the Stellenbosch University Health Research Ethics Committee (HREC); ethics reference number N15/08/071, renewed once a year. Ethics was obtained under the main study "Tracking the molecular epidemiology and resistance pattern of HIV-1 in South Africa". The study is conducted in accordance with the ethical guidelines and principle of the international Declaration of Helsinki 2013, South African Guidelines for Good Clinical Practice and the Medical Research Council (MRC) Ethical Guidelines for Research. Therefore, the patients' personal information will be protected by assigning each patient a research number. For example, patient 1, with a unique episode number will be referred to as P1. A waiver of written informed consent was awarded to conduct sequence analyses on these samples by the HREC of Stellenbosch University, South Africa.

2.2 Study design and location of research

The study conducted was an analytical and descriptive study not statistical, and made use of a small sample set. At the start of the study only 11 HIV patient samples sequences were available. All of the 11 patients were failing RAL treatment and this data is not publicly available. South African patients had their samples submitted for HIV-1 genotypic resistance testing at the National Health Laboratory Service (NHLS). The NHLS provides routine genotypic antiretroviral drug resistance testing for clinics from the Western Cape, Gauteng and Eastern Cape provinces. Notably patients from different clinics and hospitals all over South Africa form part of the cohort.

2.3 Inclusion criteria for selection of patient samples

The inclusion criterion for the study was designed to include any patient who was being treated with RAL after being suspected of failing a second-line regimen. Selected patients' files also had to state which regimen was currently or previously used to treat them to ensure that there was no selection bias during the sample collection process.

2.4 Sequence retrieval

Our collaborators in the Division of Medical Virology, Stellenbosch University and the NHLS, South Africa, have sequenced the HIV-1 subtype C (HIV-$1C_{ZA}$) Integrase (IN) gene and generated a consensus sequence (Brado *et al.*, 2018; Chitongo *et al.*, 2020; Obasa *et al.*, 2020). The consensus sequence was constructed from a South African HIV-1 infected population in DNA nucleotide sequence format. The 11 HIV-$1C_{ZA}$ IN DNA nucleotide sequences for South African HIV-1 treatment-experienced patients, who were suspected of failing RAL treatment, were also sequenced by our collaborators. The wet-lab work was done in the laboratories at the Division of Medical Virology, at Stellenbosch University, Tygerberg Campus by Dr. Obasa (Obasa, 2019). We selected the 11 sequences to identify known or novel mutations that account for RAL resistance, and determine if these mutations would affect DTG drug binding. For this purpose we used a variety of bioinformatics methods and tools to quantify the effect of any mutations on the HIV-$1C_{ZA}$ IN protein structure.

2.4.1 Sequence translation

To perform a multiple sequence alignment (MSA) and homology modelling with the consensus IN sequence, all sequences had to be translated from nucleotide sequences into protein sequences. The online ExPASy – Translate Tool (https://web.expasy.org/translate/) allows the translation of a nucleotide sequence of either DNA or RNA, into a protein sequence. This online tool was used to translate the nucleotide sequences into amino acid (AA)/protein sequences. Sequences were submitted individually to the online tool, while the genetic code and output format were left at default settings. The longest open reading frame (ORF) was selected in a FASTA format for further analysis as it showed more sequence coverage and to avoid the use of truncated sequences.

2.4.2 Multiple sequence alignment

A multiple sequence alignment (MSA) involves aligning 3 or more sequences to identify regions of conservation or to determine mutations in a cluster of sequences. Several programs are available online that perform MSA. For this project, we used the Clustal Omega program (*https://www.ebi.ac.uk/Tools/msa/clustalo/),*because it is freely available on the web, frequently cited, has a fast algorithm and can accommodate more than 1000 sequences (Sievers *et al.*, 2011). We performed our MSA for the 11 sample cohort translated protein sequences against the HIV-$1C_{ZA}$ IN consensus sequence considered here as the wild type (WT) sequence. The aim of the MSA was to identify the location of the mutations/variants among the sequences relative to the HIV-1C IN consensus sequence. The MSA program

employed the global alignment algorithm that involves an end-to-end alignment of the sequences involved. The amino acid sequences were uploaded to the Clustal Omega webserver as a single FASTA file, all other parameters were left at default settings. The output of the MSA was downloaded in a ClustalW format with character counts and saved in FASTA format.

2.4.3 Detection of known IN RAMs and new mutations

In order to assess the type of mutations experienced in South Africa as a result of drug resistance to HIV-1C ART using RAL, the original patient nucleotide sequences were submitted to the Stanford HIVdb database (https://hivdb.stanford.edu/hivdb/by-mutations/) for screening. The HIVdb program is a genotypic resistance interpretation algorithm that operates by accepting nucleotide IN sequences and returns inferred levels of resistance that is hyperlinked to data in the HIV Drug Resistance Database for the most commonly used IN inhibitors (http://hivdb.stanford.edu). The program screened for IN major RAMS, IN accessory RAMs, polymorphic mutations, INSTI RAMs and any new mutations in the sample sequence cohort.

2.5 Homology modelling

The three dimensional (3D) homology model of the tetrameric WT HIV-1C_{ZA} IN protein enzyme was predicted using SWISS-MODEL, an open access web-server dedicated to structural homology modelling of protein structures. Tools for template selection, model building, and structure quality evaluation can be invoked from within the SWISS-MODEL workspace directly (Waterhouse *et al*., 2018). All the steps involved in model construction are discussed below.

2.5.1 Template search and target-template alignment

On the SWISS-MODEL workspace, the target HIV-1C_{ZA} IN consensus sequence was pasted and used to search for homologous protein sequences using the BLAST (Camacho *et al.*, 2009) and HHblits (Remmert *et al.*, 2012) search algorithms, to identify similar structural template(s) in the Protein Data Bank (PDB). The identified potential template(s) were ranked according to the template quality and predicted features of target-template alignment. The highest-ranking template was selected for 3D homology modelling on the basis of high sequence identity, coverage and the global model quality estimation (GMQE) score.

2.5.2 Model construction

The three dimensional (3D) co-ordinates for the protein model was built automatically using the target-template alignment as input for the ProMod3 software in SWISS-MODEL (Waterhouse *et al.*, 2018). Here, the atomic coordinates for the backbone atoms of side chains that are conserved between the target and the template are extracted from the template by satisfaction of spatial restraints and imposed onto the newly built protein model. Insertions and deletions are automatically remodelled using a fragment library within the tool and side chain placements are optimized. Finally, the geometry of the resulting protein model is regularized by using a force field. To resolve problematic loop regions, ProMod3 is used for loop modelling, but, if it fails an alternative protein model is built with PROMOD-II (Guex, Diemand and Peitsch, 1999). The WT tetrameric HIV-1C_{ZA} IN enzyme protein structure was built using the 5U1C homologous template using PROMOD-II and the final output of the resulting model was downloaded in PDB format containing all 3D atomic coordinates.

2.6 Model quality evaluation

Validating the reliability of the generated protein model is very important to infer accurate interactions between the protein-drug and DNA. As part of quality assessment, the generated tetrameric 3D homology model of HIV-1C_{ZA} IN was subjected to various 3D structure assessment programs, including the SWISS-MODEL Structure Assessment program and the Structure Assessment Verification Server (SAVES) (http://servicesn.mbi.ucla.edu/SAVES/) quality checks that included tools like; ERRAT, PROVE and VERIFY3D.

The SWISS-MODEL webserver produces inbuilt quality assessment scores for generated protein models such as: Qualitative Model Energy Analysis (QMEAN) and GMQE scores. QMEAN is a composite scoring function describing the major geometrical aspects (local geometry, torsion angle potential of amino acids, long-range interactions, solvation potential and solvent accessibility) of protein structures. It is able to derive both global (for the entire structure) and local (per residue) absolute quality estimates on the basis of one single model. QMEAN scores are originally in a range of zero to four, with any value less than two considered to be good. GMQE is a quality estimation that combines properties from the target-template alignment and the template search method. The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template and the coverage of the target. Generally, high numbers close to 1 indicate high reliability of the tertiary target model(s) (Benkert, Biasini and Schwede, 2011).

A Ramachandran plot was also calculated and used to assess the structure of the HIV-1C_{ZA} IN homology model. This plot (also known as a Ramachandran diagram or a [φ,ψ] plot), originally developed in 1963 by G. N. Ramachandran, is a diagram to visualize energetically allowed regions for a polypeptide backbone, torsion angles psi (ψ) against phi (φ) of amino acid residues present in a protein structure (Ramachandran, Ramakrishnan and Sasisekharan, 1963). It is used to analyse the structure of a protein, the conformation of the amino acids present in the protein and close contacts between the atoms. It also provides an overview of excluded regions that show which rotations of the polypeptide are not allowed due to steric hindrance (collisions between atoms). The Ramachandran plot for a particular protein may serve as an important indicator of the quality of its 3D structure.

The external quality evaluation of the WT-IN protein model was done by assessing the PDB atomic coordinates of the model through the ERRAT, PROVE and VERIFY3D tools. The ERRAT program assessed the overall quality factor of the model and a good quality structure should correspond to values of 90% and/or higher (Colovos and Yeates, 1993). ERRAT analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function against the position of a 9-residue sliding window which is calculated by a comparison with statistics from highly refined structures (Colovos and Yeates, 1993). Volume of the atoms in the protein model was determined using PROVE. The program calculates the volumes of atoms in macromolecules using an algorithm that treats the atoms like hard spheres and calculates a statistical Z-score deviation for the model and compares it to Z-scores of highly resolved (2.0 Å or better) and refined (R-factor of 0.2 or better) PDB-deposited structures (Pontius, Richelle and Wodak, 1996). Absolute Z-scores of < 3 are considered reasonable for predicted protein models and is comparable to structures of high resolution. VERIFY3D assessed the compatibility of the tertiary atomic protein model by comparing it to its own amino acid sequence (Eisenberg *et al.*, 1992). This program analyses by assigning a structural class based on its location and environment (alpha, beta, loop, polar and nonpolar) and comparing the results to high resolution models. Finally, PyMol (DeLano, 2002) was used to assess the structural similarity between the predicted 3D target model and template structure(s) using the root mean square deviation (RMSD) value (https://pymolwiki.org/index.php/Align). A lower RMSD value indicates higher structural similarity to the structures with respect to the backbone conformation.

2.7 Generation of the complex HIV-1C_{ZA} IN model

A tetrameric protein structure of the modelled HIV-1C_{ZA} IN WT protein in complex with magnesium ions (MG), DNA and the drug DTG was assembled for further analysis. DTG was considered as the drug of choice due to its high genetic barrier to resistance and also its improved efficacy and potency against multiple resistance mutations. To achieve this goal, the HIV-1C_{ZA} IN protein was aligned to the prototype foamy virus (PFV) IN structure, which was in complex with MG and DTG, using PyMol (DeLano, 2002). This was done to determine the accurate location of MG ions and DTG relative to our HIV-1C_{ZA} IN protein structure. Both the MG ions and DTG were extracted from the PFV IN (PDB ID: 3S3M). Similarly, the DNA was extracted from 5U1C onto our HIV-1C_{ZA} IN structure using the PyMol align command.

2.8 Selection criteria of variants of interest and generation of mutant HIV-1C_{ZA} IN complex systems

All the new variants identified using the Stanford HIVdb program in our sample cohort were further interrogated based on their vicinity to the binding pocket and the 140s loop region of the HIV-1C_{ZA} IN. This was done to prioritize mutations that might have a causative effect on the protein structure and in particular DTG drug binding and DNA binding. PyMol was used to measure the distance of each mutation relative to the region of interest. We considered a threshold cut-off distance of < 3 Angstroms to select which variants to consider in our study for further assessment. Mutant IN complex systems that carried selected variants for our study were also prepared from the WT IN structure. For the introduction of mutations into the WT structure, we used the PyMol mutagenesis wizard and created four IN mutant complex structures each containing the mutations and labelled them P1, P2, P3 and P4. These identifiers will be used throughout the study to refer to the different structures.

2.9 Stability predictions and the loss or gain of polar interactions

In this study, the mutation Cutoff Scanning Matrix (mCSM) program was used to predict stability of the protein after mutation. The mCSM program is a structure-based method taking advantage of the protein structural information that has been accumulated on the impact of mutations within the 3D space of natively folded proteins (Pires, Ascher and Blundell, 2014). This program typically attempts to predict either the direction of change in protein stability on mutation (as a classification task) or the actual free energy value (ΔΔG) as a regression task. Before assessing the effect of selected variants on the stability of the protein, we used the Gromacs software (Spoel, 2011) to energy minimize our HIV-1C_{ZA} IN complex structure.

After the energy minimization we went on to upload our WT HIV-1C_{ZA} IN structure in PDB file format into the mCSM online tool followed with a list of variants. We did this in order to test for the effect of these variants on protein stability and protein-DNA affinity. Each mutation was tested separately for its effect on protein stability.

We also used PyMol to study if there were any changes in polar interactions of the residues that underwent mutation within the HIV-1C_{ZA} IN. The loss or gain of polar contacts was calculated for residues in the immediate environment of the mutated residue. Mutations were introduced using the mutagenesis wizard tool of PyMol and the results were tabulated. This was done to determine if the loss or gain of an interaction occurred in the immediate amino acid environment of the protein residue thereby affecting the protein's fold. This was calculated for each of the six mutations. Besides looking at static single changes we also ran simulations to understand the effect of amino acid changes on the dynamics of the protein. This is further described below.

2.10 Molecular dynamics simulations

All input files of the systems to be used for molecular dynamics (MD) simulations were prepared via the CHARMM-GUI platform (Brooks *et al.*, 2009). Each of the steps for structure preparation is discussed in more detail below.

2.10.1 Preparation of MD simulation input files with CHARMM-GUI

CHARMM-GUI is a web-based platform designed to interactively build and prepare complex systems for input to a reproducible bio-molecular simulation protocol using several simulation packages including Gromacs (Brooks *et al.*, 2009). In an attempt to generate a series of input files for MD simulations in explicit solvent conditions, the solution builder interface found under the input generator of CHARM-GUI was employed (Jo *et al.*, 2008). We employed the solution builder interface to solvate the WT and four mutant IN complex systems (P1, P2, P3 and P4) with water molecules. These five systems were prepared separately by uploading the atomic coordinates of each of the Protein-DNA-MG-DTG complexes to the CHARMM-GUI solution builder interface within the input generator (Jo *et al.*, 2008; Lee *et al.*, 2016b). Each system was solvated in a rectangular TIP3 water-box with 10Å distance between the edges of the box. The topology and coordinates for each system was generated using the CHARMM36 all-atom force field (Huang and MacKerell, 2013) and CHARMM general force field (Vanommeslaeghe *et al.*, 2009) for DTG. The CHARMM force field was chosen because it is compatible with Gromacs software, it is a bio-molecular force field containing validated parameters suitable for simulating drug-like molecules and DNA nucleotides in a biological

environment and it primarily targets biological systems (Huang and MacKerell, 2013). Each system was neutralized by adding counter ions to each of the systems. For the WT, P1 and P4 systems; 310 potassium ions (K) and 252 chloride ions (Cl) were added, while for the P2 and P3 systems we added 309 K and 252 Cl ions. Each system was at a final concentration of 0.15M for simulation dynamics.

2.10.2 Energy minimization

The input files for energy minimization of the various complex integrase systems in an aqueous solvent environment were generated by the CHARMM-GUI solution builder interface (Lee *et al.*, 2016a). 50 000 steps of energy minimization using the steepest descent minimization integrator were used to energy minimize the WT system of the solvated complex structure using the CHARMM36 force field (Vanommeslaeghe *et al.*, 2009), and applying constraints to hydrogen bonds using the LINCS constraint algorithm. The number of energy minimization steps employed were to ensure that the systems reached convergence (meaning becomes fully relaxed with no steric hindrance). All this was performed with Gromacs software version 5.1 (Spoel, 2011). Similar parameters were used to energy minimize the WT system and the four mutant systems. Gromacs version 5.1 was used for running all the simulations (Spoel, 2011). At the start of the project Gromacs version 5.1 was available and we note there are newer versions of the software that has been released.

2.10.3 System equilibration

Equilibrating the solvent and ions around the protein is the next step after energy minimization. Attempting unrestrained dynamics without equilibrating the system will most likely result in the system collapsing (Lemkul, 2018). Equilibration is performed to obtain a stable thermodynamic ensemble for any conditions desired. This is because the solvent is mostly optimized within itself and not with solute. It therefore needs to be brought to the desired temperature we use for simulation and establish the proper orientation about the solute (the protein) (Lemkul, 2018). Once the correct temperature is achieved (based on kinetic energies), pressure is then applied to the system until the proper density is reached (Lemkul, 2018). Usually equilibration of bio-molecular simulations is conducted in two phases: the isothermal-isochoric ensemble and the isothermal-isobaric ensemble. It is often considered more robust to first equilibrate the temperature of the system before applying a barostat to control the temperature.

The isothermal-isochoric ensemble also known as the canonical (NVT) ensemble is one where the number of particles (N), volume (V) and temperature (T) of the system are conserved. In NVT, the temperature of the system should reach a plateau at the desired value (Lemkul, 2018). We ran the NVT equilibration of all five systems for 200 picoseconds (ps) to stabilize the temperature of the system. The V-rescale temperature-coupling (Bussi, Donadio and Parrinello, 2007; Wong-ekkabut and Karttunen, 2012) method was used for the NVT ensemble, with constant coupling of 1 ps at 303.15 K.

The next phase of equilibration (NPT ensemble), which involves equilibrating the pressure (and thus also the density) of the system, was executed. Within the isothermal-isobaric (NPT) ensemble, the number of particles (N), pressure (P) and temperature (T) of the system are conserved. The short position restraint NPT was ran for 500 ps for each of the five IN systems to stabilize the pressure by relaxing the systems individually, keeping the proteins restrained. For NPT, the Nose-Hoover pressure coupling (Nosé, 1984; Evans and Holian, 1985; Hoover, 1985) was turned on with constant coupling of 1 ps at 303.15 K under conditions of position restraints (h-bonds) selecting a random seed/starting structure value. Electrostatic forces were calculated for both NVT and NPT using Particle Mesh Ewald method (Essmann *et al.*, 1995).

2.10.4 Production molecular dynamics

Upon completion of the two equilibration phases, the system will be equilibrated at the desired temperature and pressure (Lemkul, 2018). Position restraints can now be released from the protein at this stage and the simulation will proceed in an unbiased manner. This phase of the MD simulation is often referred to as 'unrestrained or production MD'. Initially the WT system was subjected to a 300 nanoseconds (ns) simulation in a bid to determine the time at which the system attains equilibrium. The conditions of the simulation involved no restraints on hydrogen bonds, a simulation integration time step of 0.002 ps and the MD trajectories were recorded every 10 ps. We discovered through analysis that the system reached equilibrium at 90 ns. Therefore all mutant systems were subjected to 150 ns simulations under similar conditions as that of the WT. The production simulations were performed separately in duplicates to validate reproducibility of results.

2.11 Analysis of MD simulations

Thermodynamic parameters such as potential energy, temperature, pressure and density for the systems were calculated using *gmx* energy. Various simulation trajectory analyses for the WT and mutated HIV-1C_{ZA} IN systems were considered which included the root-mean-square-deviation (RMSD); root-mean-square-fluctuation (RMSF); radius of gyration (Rg);

intramolecular hydrogen bonds (HB); pairwise distance measurements, non-bonded pairwise interaction analysis and extraction of structural snapshots/conformations over the last 50 ns for interaction analysis. These analyses of the trajectory files were done using Gromacs utilities.

RMSD and RMSF were calculated using *gmx rmsd* and *gmx rmsf*, respectively. The RMSD analysis shows the average displacement of the atoms relative to a reference structure (usually the first frame of the simulation) over the course of the simulation (Benson and Daggett, 2012; Martínez, 2015). The RMSF measures the average fluctuation of a particular atom, or residues over time (typically the time-averaged position of the particle) (Martínez, 2015). Rg was calculated using *gmx gyrate* to determine if the system reached convergence or remain fully folded over the 150 ns simulation. This analysis is an indicator of protein structure compactness, describing the overall spread of the molecule and is defined as the root-mean-square distance of the collection of atoms from their common center of gravity (Lobanov, Bogatyreva and Galzitskaya, 2008). The total number of hydrogen bonds between the protein and drug was calculated using *gmx hbond*. Moreover, pairwise distance analysis between the drug and MG was done using *gmx pairdist* tool to measure how the distance between the drug and MG varied over the progression of the simulation. The non-bonded pairwise interaction energy analysis was calculated using *gmx energy*. In the calculations, the energy terms considered included the coulombic electrostatic energy and the Lennard-Jones energy. This non-bonded interaction energy is not a free energy or binding energy. The RMSF and Rg analyses were calculated over the last 60ns of the simulation whereas HB, pairwise distance and non-bonded pairwise interaction analyses were calculated over the last 50 ns of the simulation for 1000 frames, for all systems. These specific timeframes of the simulation were considered for the specific analyses because that was the most stable part of the simulation. Interaction analysis was performed on structural snapshots/conformations of the five different IN systems that were extracted over the last 50 ns for every 10 ns. The snapshots were used to calculate polar interactions between DTG, MG and any residues of the protein within the specific timeframes using PyMol.

Chapter 3 Results

3.1 Demographic information

Our study population consisted of 11 individuals that were suspected of failing RAL treatment, after having failed a second-line cART regimen. 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. Further demographic data are available on request from Dr Jacobs.

3.2 Multiple sequence alignment and mutation detection

A multiple sequence alignment (MSA) was performed for the 11 South African (SA) patient sequences obtained from our colleagues (sample cohort), with the HIV-$1C_{ZA}$ consensus IN sequence used as the reference sequence. From the MSA we were able to observe significant similarities among the sample cohort sequences relative to the consensus sequence. Figure 3.1 shows that the DDE motif residues (D64, D116 and E152) in the catalytic-core domain of the IN are conserved in all the sequences. Furthermore, the residues in the 140s loop region (140-149) are also highly conserved (Figure 3.1). We used the automated report of mutations obtained from the HIVdb Program of the Stanford University HIV Drug Resistance Database (last accessed March 2019) to compile a list of all mutations found in a South African cohort possibly failing RAL-treatment for HIV-1C infection. Table 1 shows the list of all mutations identified from the HIVdb program of the Stanford University HIV Drug Resistance Database:

Table 1: List of mutations present in SA's patients' samples who are suspected of failing HIV-1C RAL treatment

Known INSTI resistance associated mutations	Y143R
Different Integrase accessory mutations	T97A, G163E / Q, E157Q, S230R
Naturally occurring polymorphisms	E25D, I31V, I50M, I72V, L74I /M,Y100F, I101L, V112T, A124T, A125T, Q136K, I201V, I218L/ T, I234L, A265V, K269R, A278D, G283D/ S
New mutations observed in our South African sample cohort	S24N, S39C, L63I, V75M, I84M/V, E96D, I113V, S119P, G134N, I135V, V150A, M154I, S195C, T206S, T210A, K211Q/R, K215N, K219Q, I220V, S255G, V281M

Abbreviations of amino acids: A – Alanine, D – Aspartic acid, E – Glutamic acid, G –Glycine, I – Isoleucine, K – Lysine, L – Leucine, M – Methionine, N – Asparagine, P – Proline, Q – Glutamine, R – Arginine, S – Serine, T – Threonine, V – Valine, Y – Tyrosine.

From the list of known primary RAMs associated with RAL failure, we only identified the Y143R mutation in our sample cohort (Table 1). Moreover, all the previously reported IN accessory mutations and naturally occurring polymorphisms (Brado *et al.*, 2018) were detected in our samples. An additional 21 variants were also detected, which are novel in the South African cohort.

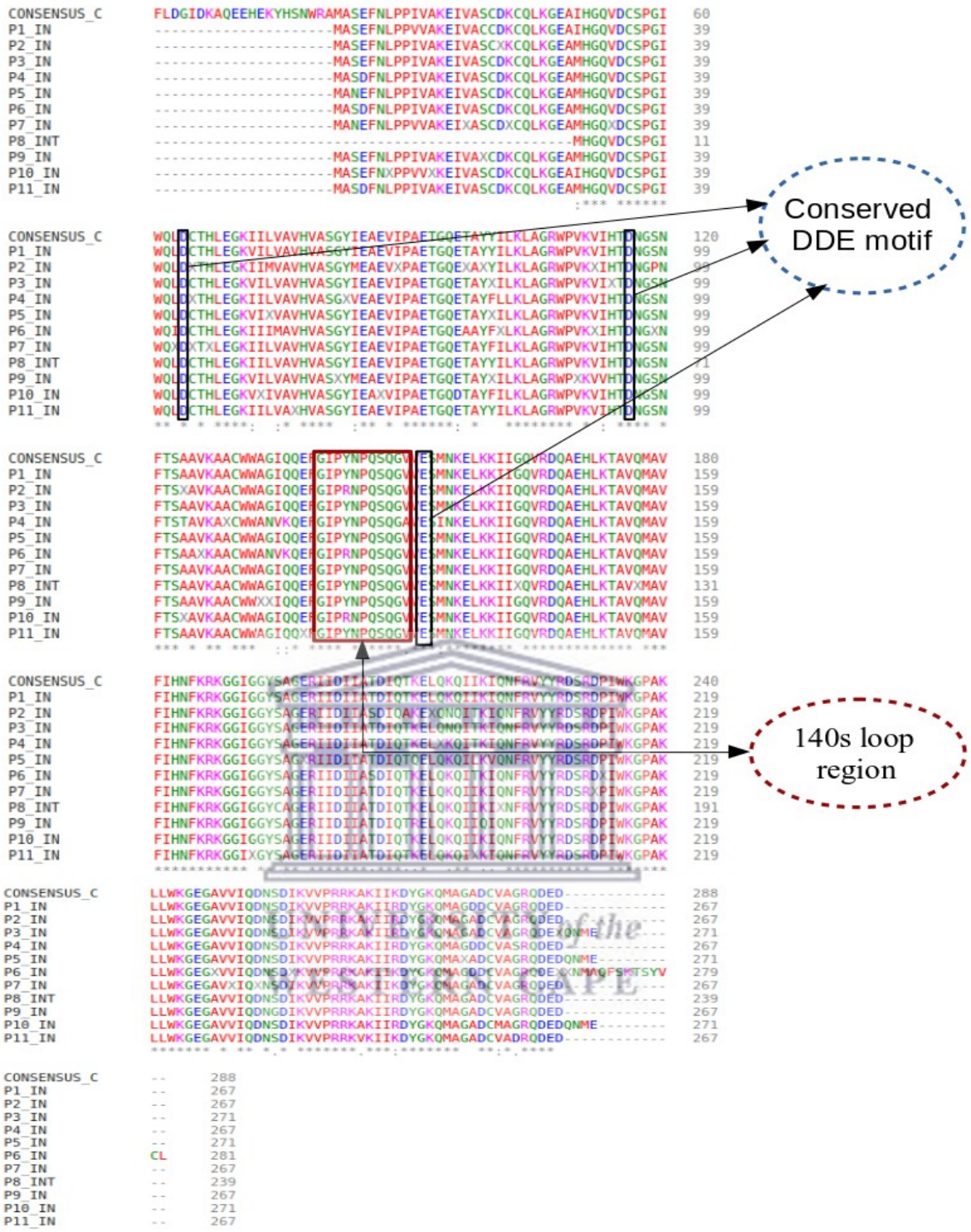


Figure 3.1: MSA alignment of the 11 patient sequences against the HIV-1C$_{ZA}$ IN consensus. Sample sequences were renamed P1 – P11. The numbering on the sides represents the amino acid residue position. Amino acid residues are also assigned colours according to their different properties, with small and hydrophobic residues coloured red; acidic residues coloured blue; basic residues coloured magenta; hydroxyl/sulfhydryl/amine residues coloured green and unusual amino acid residues coloured grey. An increased number of conserved regions within the sequences is observed, symbolized by the asterisk (*) at the bottom of the alignment. However, some residues among the sequences are not conserved as shown by colons (:) and gaps. This is assumed to be due to mutations or deletions of the residues at random positions.

3.3 Homology modelling

The homologous cryoEM intasome structure with PDB ID: 5U1C was selected for the construction of the tetrameric 3D structure of HIV-1C_{ZA} IN protein. The criteria for the template selection were based on several factors that included; sequence identity, coverage and the GMQE score. 5U1C presented the highest sequence coverage (1.0) and sequence identity (93.40%) to our target sequence relative to other templates (Chitongo *et al.,* 2020). The GMQE score of 0.85 was closer to 1.00, indicating a high confidence in the quality of the homology model. The final output of the resulting model was downloaded in a PDB file format containing all 3D atomic coordinates. Figure 3.2 shows the 3D cartoon representation of the tetrameric HIV-1C_{ZA} IN protein homology model.

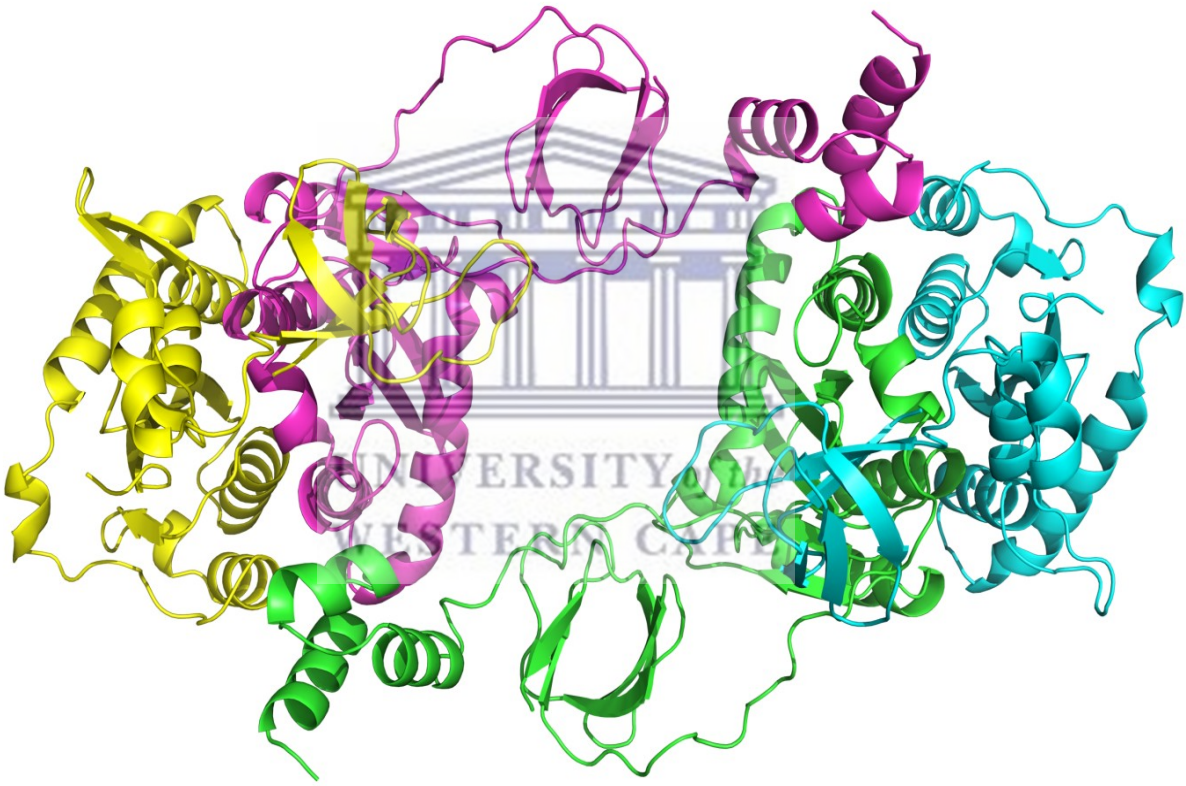


Figure 3.2: 3D cartoon representation of the HIV-1C_{ZA} IN tetrameric structure predicted with SWISS-MODEL. Chain A is coloured green, Chain B is coloured cyan, Chain C is coloured magenta and Chain D is coloured yellow. Two dimer chains of the HIV-1C IN are shown folding and coming into contact to form a tetramer structure as expected from the template. The tetramer folds in such a way that identical chains (Chain A and C) will be on the inside of where they will come in contact with the drug and DNA, while identical outer chains (Chain B and D) will be exposed to the solvent molecules.

3.4 Model quality evaluation

The QMEAN score was found to be -1.47 for the 3D tetrameric structure of HIV-1C_{ZA} IN and is close to its native state. This result indicated a high level of confidence in the modelled structure. Moreover, the GMQE value of 0.79 is close to 1 and indicates a high level of confidence and reliability in the predicted HIV-1C_{ZA} IN homology structure. A very low RMSD score (0.173 Å) was calculated with PyMol align command between the HIV-1C_{ZA} IN homology model and template 5U1C (Table 2). This indicates very little deviation of backbone carbon atoms between target and template, suggesting high homology and structural similarity between the structures. The overall quality factor score, as evaluated by ERRAT for the HIV-1C_{ZA} IN homology model, demonstrated that all residues in the four chains of the tetramer (Chains A-D) had a low error rate for the 3D structure, making it comparable to high resolution crystal structures (1.5-2.5 Å) (Table 2). The PROVE test tool showed that the volume of the atoms in the protein model was carefully assigned, with only a 6.60% margin of error for the predicted protein model. The protein model was also considered compatible based on the assessment by the VERIFY3D tool, where 80.12% of the residues had a 3D-ID score ≥ 0.2, allowing the model to pass the assessment (Table 2).

Table 2: Scores for tetrameric HIV-1C_{ZA} IN protein model quality evaluation parameters

Swiss Model	**PyMOL**	**SAVES**		
Parameters	Superimposition	VERIFY3D	ERRAT	PROVE
GMQE = 0.79 **QMEAN = -1.47**	*RMSD* = 0.173 Angstroms	80.12% of the residues have a 3D-ID score ≥ 0.2 Pass	Overall Quality Factor A: 97.0588 B: 89.5000 C: 94.7581 D: 95.4082	Buried outlier protein atoms total from 1 Model: 6.60% error

The Ramachandran plot for HIV-1C_{ZA} IN shows that the model also satisfied stereo-chemical restraints; the phi-psi dihedral angle distribution of most residues were in favourable conformations. Output from the assessment of the homology model of the IN with the Ramachandran plot indicated that the 3D model contained 92.80% of the residues in favoured regions, with very little outlier residues (0.94%). Additionally, out of 7788 bonds within the structure, none were reported to be bad bonds but 69 bad angles were recorded from a total of 10530 angles which marks an error of approximately 0.65%. Bad angles result from the geometry of amino acids and/or their side chains assigned inappropriately resulting in steric clashes. Based on these results, the generated 3D HIV-1C_{ZA} IN protein model was considered

reliable for any further use within the study. The bad angles and bonds were resolved using the energy minimization procedure.

3.5 Generation of a DNA-MG-DTG complex homology structure

Aligning the PFV IN (PDB ID: 3S3M) amino acid sequence with the HIV-1C_{ZA} IN protein sequence using Clustal Omega pairwise sequence alignment, provided a sequence identity of more than 60% and in addition, all the active site residues were conserved. Based on this result, the PFV IN was considered reliable because it showed good homology to our HIV-1C IN mainly with the catalytic motifs and it was also one of the most accurate IN protein structural templates available which were bound to MG ions and DTG. The sequence conservation allowed us to perform a structural alignment using PyMol between the PFV IN complex and HIV-1C_{ZA} IN structure to extract and obtain the correct position of the MG ions and the drug DTG. The HIV-1C_{ZA} wild type (WT) IN, in complex with MG, DNA and the drug DTG w successfully prepared by extracting MG ions and DTG from the PFV IN and DNA from 5U1C using PyMol. The resulting 3D tetrameric HIV-1C_{ZA} IN protein model structure is presented in Figure 3.3. This arrangement of integrase oligomers bound to the viral/host DNA chimera has been referred to as the intasome core structure (Passos *et al.,* 2017; Rogers *et al.*, 2018).

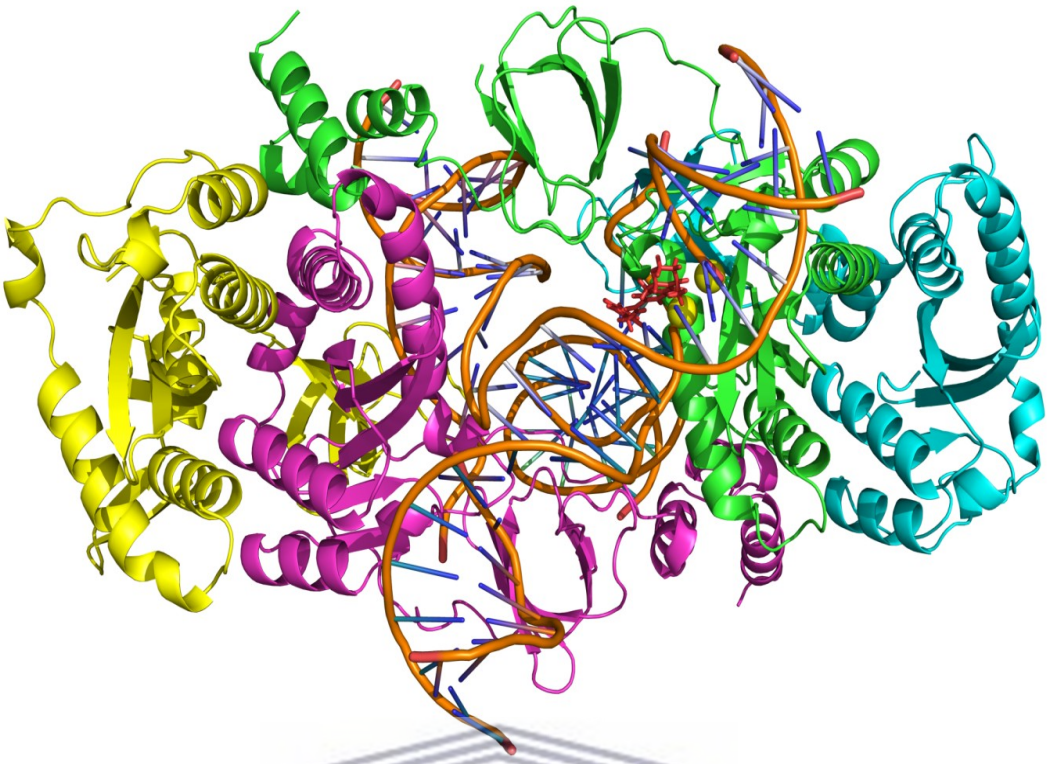


Figure 3.3: Complete 3D structure of the HIV-1C_{ZA}-IN-MG-DNA-DTG complex homology model. Chain A is coloured in green, chain B in cyan, chain C in magenta and chain D is coloured yellow. The MG ions are represented as yellow spheres while DTG is represented as red sticks (Chitongo *et al.,* 2020). A pair of dimers is shown encapsulating the viral/host DNA chimera, in which the two inner molecules (chains A and C) directly interact with DNA, while the outer molecules (chains B and D) have protein-protein interactions with the two inner dimers.

3.6 Interaction Analysis

3.6.1 Interactions of DTG with IN active site

The complete HIV-1C_{ZA} IN complex structure was initially energy minimized successfully using Gromacs. This was done to relax the bond angles of the amino acid residues within the structure before any analysis was performed including calculating polar interactions between the protein and drug. Figure 3.4 (a) shows the 3D representation of the binding pocket of HIV-1C_{ZA} IN, where MG ions and DTG are in contact with the DDE motif. It is observed that DTG makes strong polar contacts with the two Mg^{2+} ions within the active site. The drug also forms polar contacts with DNA residues THY11 and GUA22.

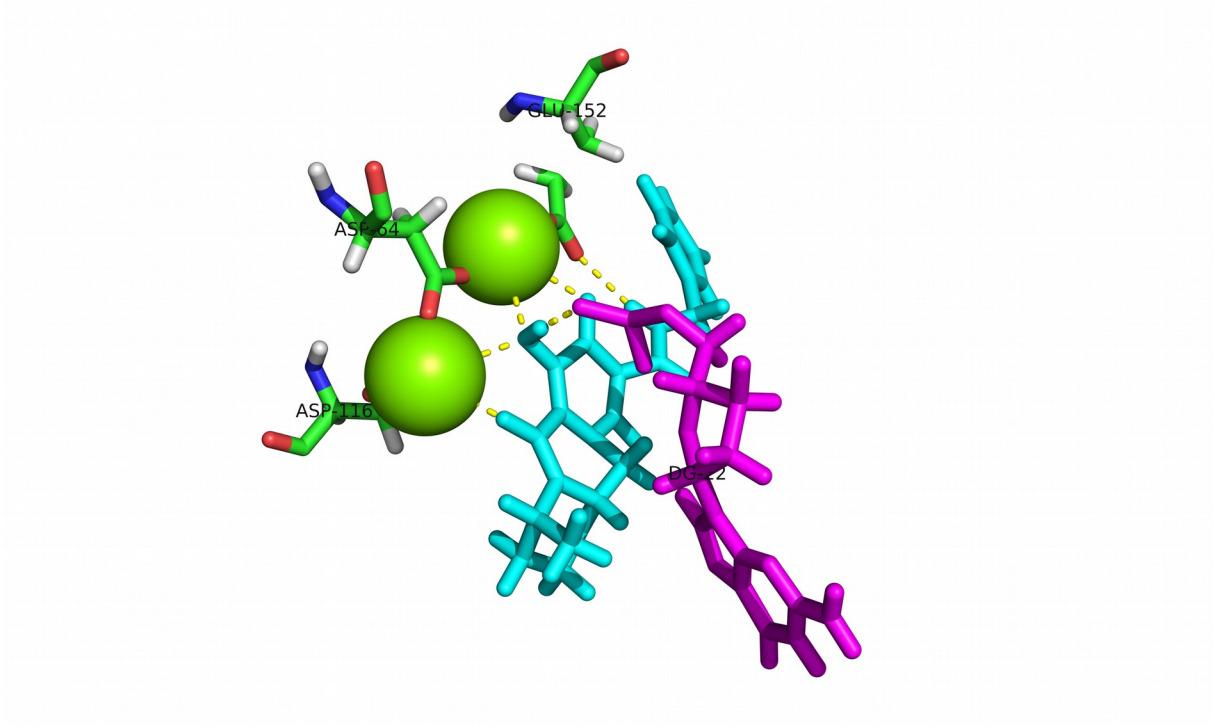


Figure 3.4: HIV-1C IN active site showing interactions with DNA, MG and drug DTG. Magnesium^{2+} ions (green spheres) are shown sitting in close proximity with DTG (cyan) where the binding pocket (DDE motif) residues are labelled and shown as sticks. Two DNA residues: THY11 and GUA22 (magenta sticks) are shown expressing polar interactions with DTG and the drug also interacts with both MG ions as shown. Dashed yellow lines show polar contacts (Chitongo *et al.,* 2020). Polar contacts calculated within 3.5Å or less.

3.6.2 Selection of variants for analysis

After considering the threshold cut-off distance of < 3 Angstroms between the variant and binding pocket residues or the 140s IN loop region, we had six out of the 21 new variants that were found in our sample cohort sequences (from Table 1). Upon further assessment, we analysed all the 11 patient derived viral sequences in our cohort, for the six variants of interest and obtained four separate distinct sequences that carried at least one of the selected six variants. Thus, the four different HIV-1C IN mutant complex systems that we generated were a representative of the four different sequences that contained the variants of interest. Mutant complex structure P1 had a single I113V mutation; structure P2 had three mutations: L63I, V75M and Y143R; structure P3 had two mutations: S119P and Y143R; while structure P4 also had two mutations: V150A and M154I.

3.6.3 Protein stability and polar interaction calculations

Of the six new variants selected, all of them were found to be slightly destabilizing to the protein structure, except for one (S119P), which was calculated to be slightly stabilizing to the

protein structure (Table 3). Furthermore, variant L63I showed increased DNA binding affinity (Table 3), while variants V75M and I113V showed highly increased protein-DNA affinity values. Variants S119P, V150A and M154I demonstrated highly reduced DNA affinity values (Table 3) and mutation Y143R also showed reduced protein-DNA affinity values. Interestingly, only variant S119P showed a loss in polar contacts, while Y143R showed a gain in the number of polar contacts formed upon introduction of the mutation, but the other variants showed no change in the number of interactions with surrounding residues (Table 3). Only variant V75M had a direct interaction with one of the DDE motif residues, D64. Interestingly none of the variants made a direct contact with the drug DTG. A figure showing the localization of all the selected variants relative to the catalytic site (DDE motif) and drug, DTG, is presented in Appendix B

Table 3: Stabilizing and destabilizing effects of selected variants on protein stability, DNA binding affinity and assessment of changes in polar contacts before and after mutation

	Protein Stability change	Protein-DNA affinity change	Polar contacts	
Wild residue	Predicted ΔG	Predicted ΔG	Before mutation	After mutation
L63I	-1.120	1.596	1 (T115)	1 (T115)
V75M	-0.883	4.504	2 (D64)	2 (D64)
I113V	-1.817	2.055	2 (Q136, E138)	2 (Q136, E138)
S119P	0.632	-2.072	3 (N120, THY29[d])	0
Y143R	-0.147	-1.321	0	1 (CYT10[d])
V150A	-1.509	-3.864	2 (S153, M154)	2 (S153, I154)
M154I	-0.736	-2.493	2 (V150, L158)	2 (A150, L158)

Negative ΔG values for protein stability suggest a destabilizing effect while positive ΔG values suggest stabilizing effect. Negative ΔG values for protein-DNA suggest a reduction in the protein's affinity for DNA whereas positive ΔG values suggest an increased affinity for DNA by the protein.

[d]Interactions with DNA nucleotides

Abbreviations of DNA nucleotides: CYT – Cytosine, THY - Thymine

3.7 Production MD simulations

In this study, comparative MD simulations were performed in order to study the difference in protein dynamics of the WT and mutant HIV-1C_{ZA} IN enzymes. Before we could perform molecular dynamics simulations on the WT IN system and the four mutant systems (P1 – P4), the backbone and side chains of the structures had to be relaxed to achieve a low potential energy and be equilibrated to conditions of constant temperature and pressure for both the solvent and protein. After energy minimization and the equilibration processes, the systems were subjected to 150 ns of unrestrained molecular dynamic simulations.

3.7.1 Energy minimization of HIV-1C_{ZA} IN structures

It is always necessary to relax the solvated system to a low energy-state before beginning MD simulations; otherwise any simulation will likely be unstable. This ensures that the system has no steric clashes or inappropriate geometry. The process of energy minimization moves positions of atoms according to the forces acting upon them (Lemkul, 2018). It is impossible to know whether or not a system is at its global energy minimum, but this is not necessarily the goal of energy minimization, rather it aims to find a reasonable starting point for the simulation (Lemkul, 2018). Two very important factors are to be considered when evaluating the success of the energy minimization process. First, the potential energy (E_{pot}) should be negative (for a protein in water) and within the order of 10^5-10^6 kJ mol^{-1}, depending on the system size and number of water molecules (www.mdtutorials; Lemkul, 2018). Second, the maximum force (F_{max}) must be no greater than 1000 kJ mol^{-1} nm^{-1} as targeted within the energy minimization *mdp* file.

All our systems were successfully energy minimized by the steepest descents method converging to F_{max} values of less than 1000 kJ mol^{-1}. The potential energies (E_P) of the five systems were all negative (Table 4) and within the range -4.72 to -4.74 x 10^5 kJ mol^{-1}. This energy minimization process was duplicated for each system for statistical reproducibility and a similar trend with the results was obtained (Table 4).

Table 4: Energy minimization results for the WT and P1-P4 complex mutant systems

	E_{pot} x E05 (kJ mol^{-1})	F_{max} x E02 (kJ mol^{-1} nm^{-1})	Duplicates	E_{pot} x E03 (kJ mol^{-1})	F_{max} x E02 (kJ mol^{-1} nm^{-1})
WT	-4.731	9.831	WT	-4.734	9.252
P1	-4.731	9.154	P1	-4.740	9.559
P2	-4.722	9.239	P2	-4.736	9.537
P3	-4.725	8.271	P3	-4.737	9.248
P4	-4.738	8.072	P4	-4.725	9.320

3.7.2 Equilibration of HIV-1C_{ZA} IN structures

The canonical (NVT) ensemble, where the number of particles (N), volume (V) and temperature (T) of the system are conserved, was successfully executed. The NVT ensemble equilibration process saw the WT and mutant systems successfully subjected to a state of constant temperature between the solvent and protein. Average temperature ranges of 299.83 – 299.87 K were obtained for all systems, including the duplicates. Moreover, the systems were also successfully equilibrated to constant pressure conditions under the isothermal-isobaric (NPT) ensemble, where the number of particles (N), pressure (P) and temperature (T) of the system are conserved. This resulted in very low average pressure (0.608 – 1.169 Pa) for

the individual systems and almost constant average density (1045.52 – 1045.78 kg/m^3) for the protein in the various systems.

3.8 Simulation trajectory analyses

In this study, comparative MD simulations were performed in order to study the difference in protein dynamics of the WT and mutant HIV-1C_{ZA} IN enzymes.

3.8.1 RMSD analysis

The backbone root mean square deviations (RMSDs) for the five HIV-1C_{ZA} IN systems were calculated using the *g_rms* tool, by performing a least square fit to the initial structure. To examine the extent to which mutations affect the protein structure, RMSD values were determined for WT and mutant protein structures. We calculated the RMSD for the backbone atoms from the initial structure to measure the convergence of the protein systems. Trajectory analysis of the RMSD of the backbone indicated that the WT and the P1-P4 mutant systems reached equilibrium after 80 ns (Figure 6). The initial increase in RMSD values before equilibrium time (80 ns) was due to the energy minimization and equilibration steps (Dewdney *et al*., 2013). We report mean and standard deviation values for RMSD values. For the duration of the simulation, the WT structure had an average RMSD value of 0.332 ± 0.038 nm. Among the mutants, structure P3 had slightly larger RMSD deviation values compared to the other structures, attaining a maximum RMSD value of 0.340 ± 0.032 nm followed by P1 and P2, which had values of 0.331 ± 0.038 nm and 0.322 ± 0.026 nm, respectively. P4 had the lowest RMSD value of 0.285 ± 0.029 nm. For the repeat simulations, the RMSD values for the WT, P1, P2, P3 and P4 structures were 0.330 ± 0.031 nm, 0.329 ± 0.023 nm, 0.318 ± 0.044 nm, 0.338 ± 0.045 nm and 0.294 ± 0.034 nm, respectively. The values are similar and follow a similar trend to those obtained in the original simulation run. Throughout the analysis, structure P3 gave the maximum deviation (Figure 6), while the WT and mutants P1 and P2 exhibited intermediated deviation and mutant P4 showed the least amount of RMSD deviation. All the structures had backbone deviations that were similar in range varying by at least 0.02 nm, except for complex P4. The average backbone RMSD values suggest that substitutions at residues 150 and 154 may result in reduced flexibility, as observed with complex structure P4 (carrying mutations V150A and M154I). Variations in the average RMSD values of the P1, P2 and P3 mutant systems lead to the conclusion that these mutations could positively affect the dynamic behaviour of IN, thus providing a suitable basis for further analyses.

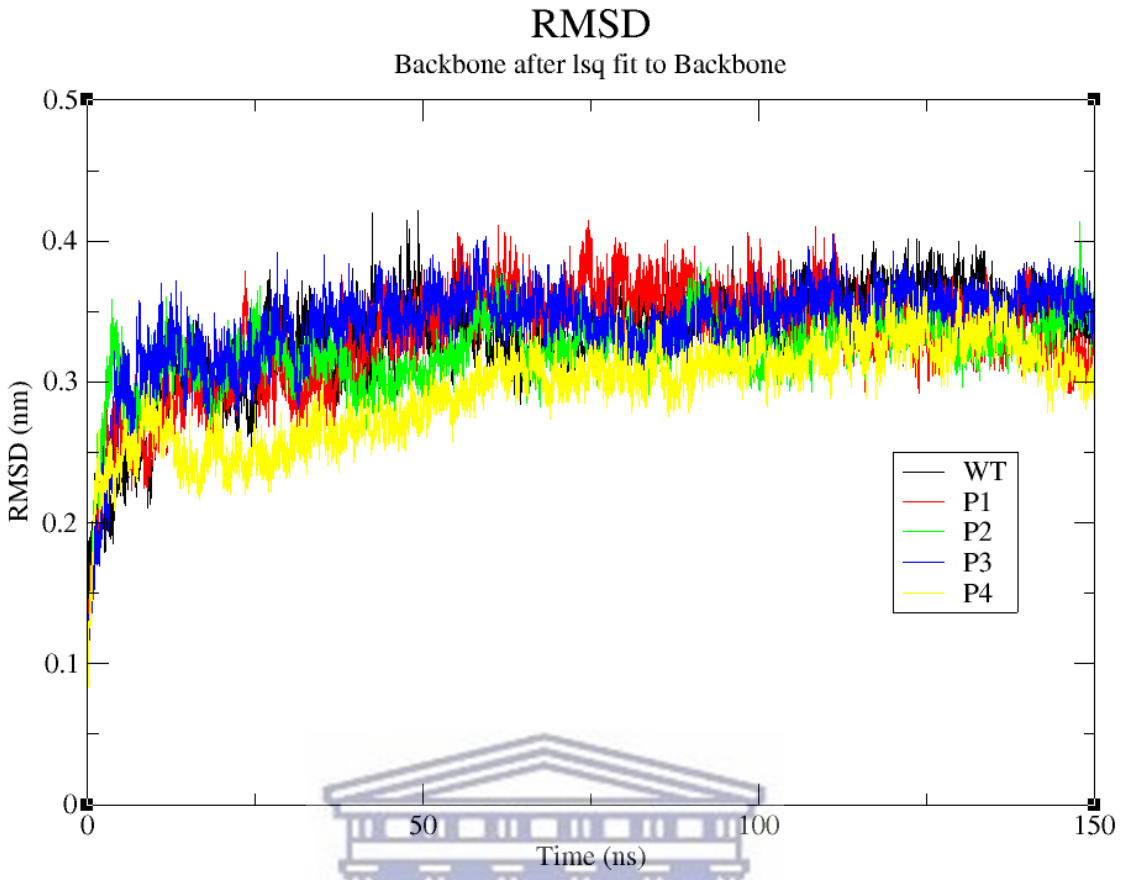


Figure 3.5: Backbone RMSDs of the WT HIV-1C_{ZA} IN protein and variant systems P1, P2, P3 and P4 at 300K are shown as a function of time. All complexes have RMSD values within similar ranges shown by overlapping lines except for P3 that shows a lower RMSD value. At approximately 80 ns all systems reached equilibrium. Black line colour indicates WT IN, red line colour indicates P1 IN (I113V), green line colour indicates P2 IN (L63I, V75M, Y143R), blue line colour indicates P3 IN (S119P, Y143R) and yellow line colour indicates P4 IN (V150A, M154I).

3.8.2 RMSF analysis

In order to determine how the mutations affect the dynamic behaviour of residues, the root mean square fluctuation (RMSF) values for the WT and mutant structures were calculated. For the RMSF trajectory analysis, the single chain A (monomer) of the IN protein in contact with the drug, MG and DNA was considered. The mean RMSF values of WT residues fluctuated within the range of 0.104 ± 0.057 nm in the entire simulation period (Figure 7). Moreover, the P1 system showed a maximum flexibility of 0.136 ± 0.070 nm, while complexes P2, P3 and P4 exhibited flexibility of 0.116 ± 0.061 nm, 0.111 ± 0.065 nm and 0.114 ± 0.061 nm, respectively. Values of 0.111 ± 0.566 nm, 0.133 ± 0.083 nm, 0.118 ± 0.065 nm, 0.109 ± 0.072 nm and 0.116 ± 0.080 nm were obtained during the repeat simulation for the WT, P1, P2, P3 and P4 structures, respectively. RMSF analysis of the fluctuations showed that variant systems P1 and P2 had more flexible regions compared to WT and other mutant

structures (Figure 7). From Figure 7, it is shown that fluctuations of amino acids were largest in the regions 208-257 with all systems. Slight fluctuations were also observed at residues 1-45, 142-143 and 160 for P1; residues127, 146 and 154 for P2 and residues 48, 69 and 131 for P4. The RMSF values for important catalytic triad residues (D64, D116 and E152) for drug binding remains very low, with values less than 0.1 nm (Figure 7), throughout the MD simulations in all structures. Looking at residues 140 – 149, which form the 140s loop region within the IN structure, systems P1 and P2 showed the highest residue fluctuations within that region.

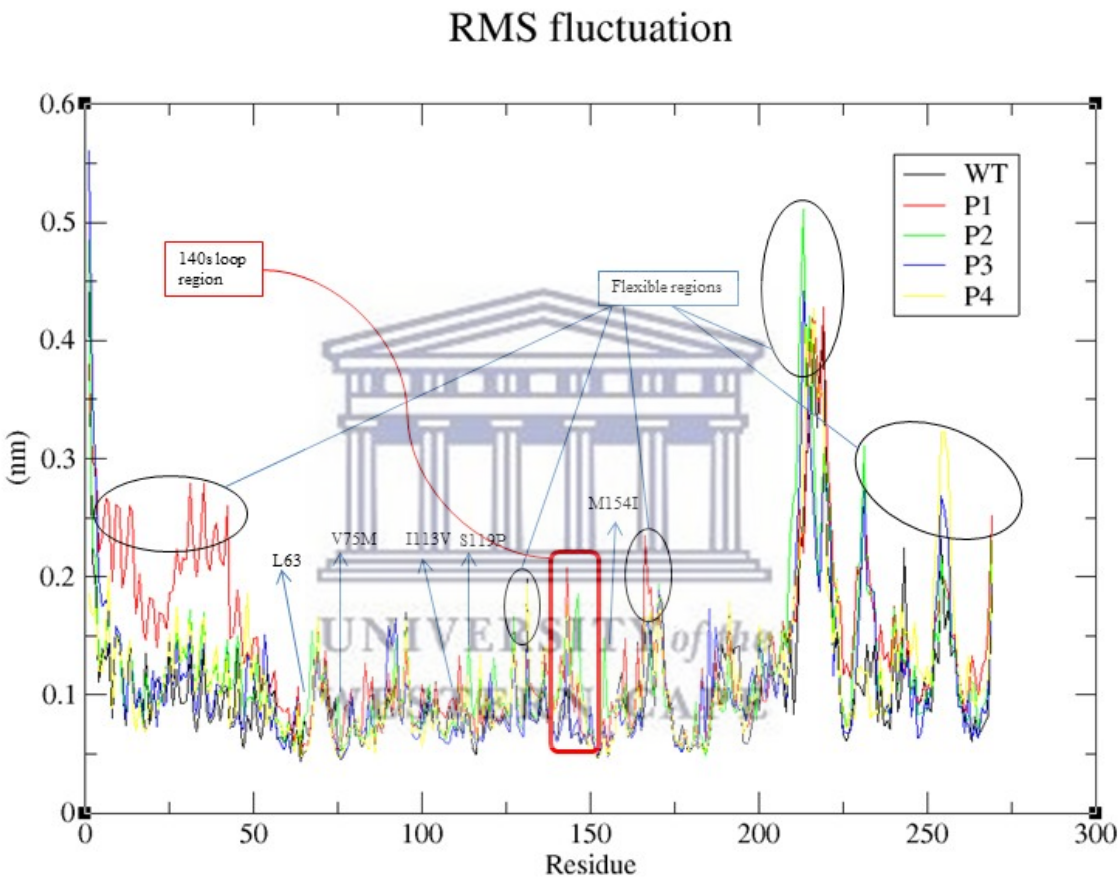


Figure 3.6: RMSF of the protein atoms of WT HIV-1C$_{ZA}$ IN protein and variant systems P1, P2, P3 and P4 at 300K are shown. Structures P1 and P2 showing flexibility of residues in the 140s loop and an increased overall number of flexible regions compared to the WT, P3 and P4 structures. Black line colour indicates WT IN, red line colour indicates P1 IN (I113V), green line colour indicates P2 IN (L63I, V75M, Y143R), blue line colour indicates P3 IN (S119P, Y143R) and yellow line colour indicates P4 IN (V150A, M154I).

3.8.3 Radius of gyration

Radius of gyration (Rg) analysis was performed to determine the compactness of the WT and mutant structures. This type of analysis provides insight into the overall dimensions of the protein structure. Figure 8 illustrates the plot of Rg for backbone atoms of the WT and complexes P1-P4 from 90 - 150 ns simulation time at 300K. The mean and standard deviation Rg value for the WT was found to be 3.612 ± 0.012 nm (Figure 8), with complex P3 having the highest mean Rg value of 3.689 ± 0.015 nm. Structures P1, P2 and P4 had average Rg values of 3.615 ± 0.014 nm, 3.678 ± 0.015 nm and 3.620 ± 0.012 nm, respectively. These results suggest that the WT, P1 and P4 structures are more compact compared to other mutant structures. The repeat simulations also followed a similar trend giving the Rg values for the WT, P1, P2, P3 and P4 structures as 3.617 ± 0.012 nm, 3.620 ± 0.012 nm, 3.673 ± 0.013 nm, 3.680 ± 0.013 nm and 3.624 ± 0.016 nm, respectively. After the relaxation period, the Rg values for all the systems were observed to show a stable trend. From the Rg graph, it is revealed that variants in complexes P2 and P3 affect the compactness of the IN protein enzyme.

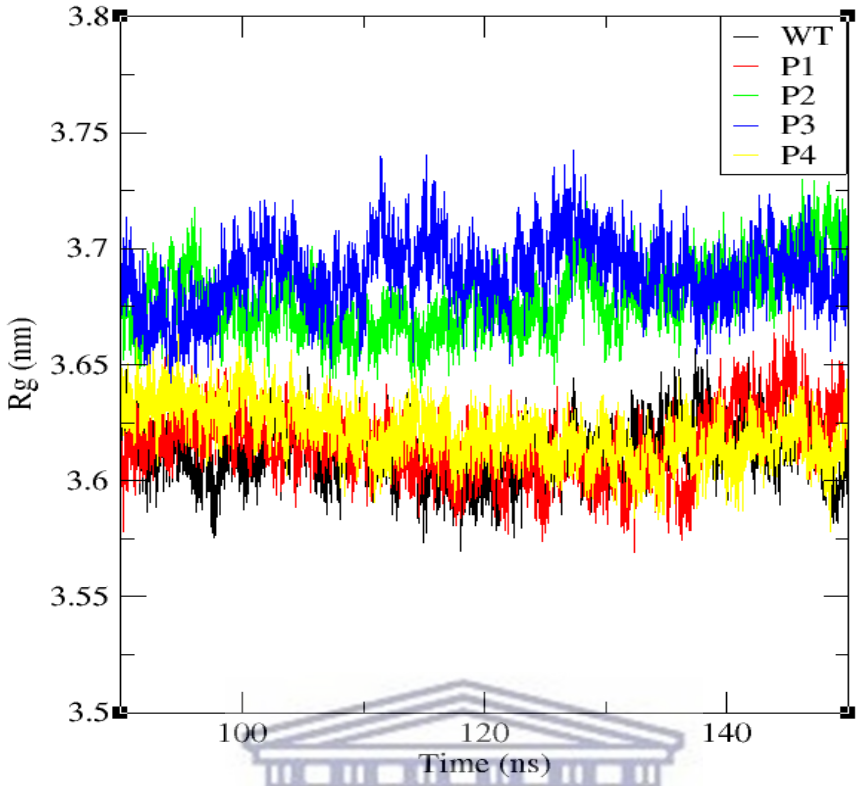


Figure 3.7: Rg for backbone atoms of WT HIV-1C$_{ZA}$ IN protein and variant systems P1, P2, P3 and P4 at 300K are shown as a function of time. Measure of compactness for the different complex systems plotted over the last 60ns. Structures P2 and P3 are less compact in comparison with structures P1, P4 and WT. Black line colour indicates WT IN, red line colour indicates P1 IN (I113V), green line colour indicates P2 IN (L63I, V75M, Y143R), blue line colour indicates P3 IN (S119P, Y143R) and yellow line colour indicates P4 IN (V150A, M154I).

3.8.4 Hydrogen bond analysis

The average number of hydrogen bonds (HBs) formed between the protein and drug was calculated for the last 50 ns of the simulations. This timeframe was chosen because we considered it to be the most stable part of the simulation trajectory. The average number of HBs formed between the protein and drug were calculated to be 0.482, 0.092, 0.249, 0.007 and 0.011 for the WT, P1, P2, P3 and P4 structures, respectively (Figures 9A-E). From the HB bar diagram it is shown that the HBs formed within the WT structure between protein and drug diminished towards the end of the simulation, while P1 formed the highest number of HBs within the structure between 110 and 140 ns (Figures 9A and 9B). Structure P2 showed HB contacts throughout the last 50 ns of the simulation and was the only mutant complex that showed an increased average number of HB formation between the protein and drug (Figure 9C). With structure P3, HBs were sparsely formed at irregular intervals and it was the structure with the least number of HBs (Figure 9D). As previously observed from the polar interaction analysis (Table 3), no variant made direct hydrogen bond contacts with DTG. The variants only made hydrogen bond contacts with neighbouring protein residues to either form more or less polar contacts.

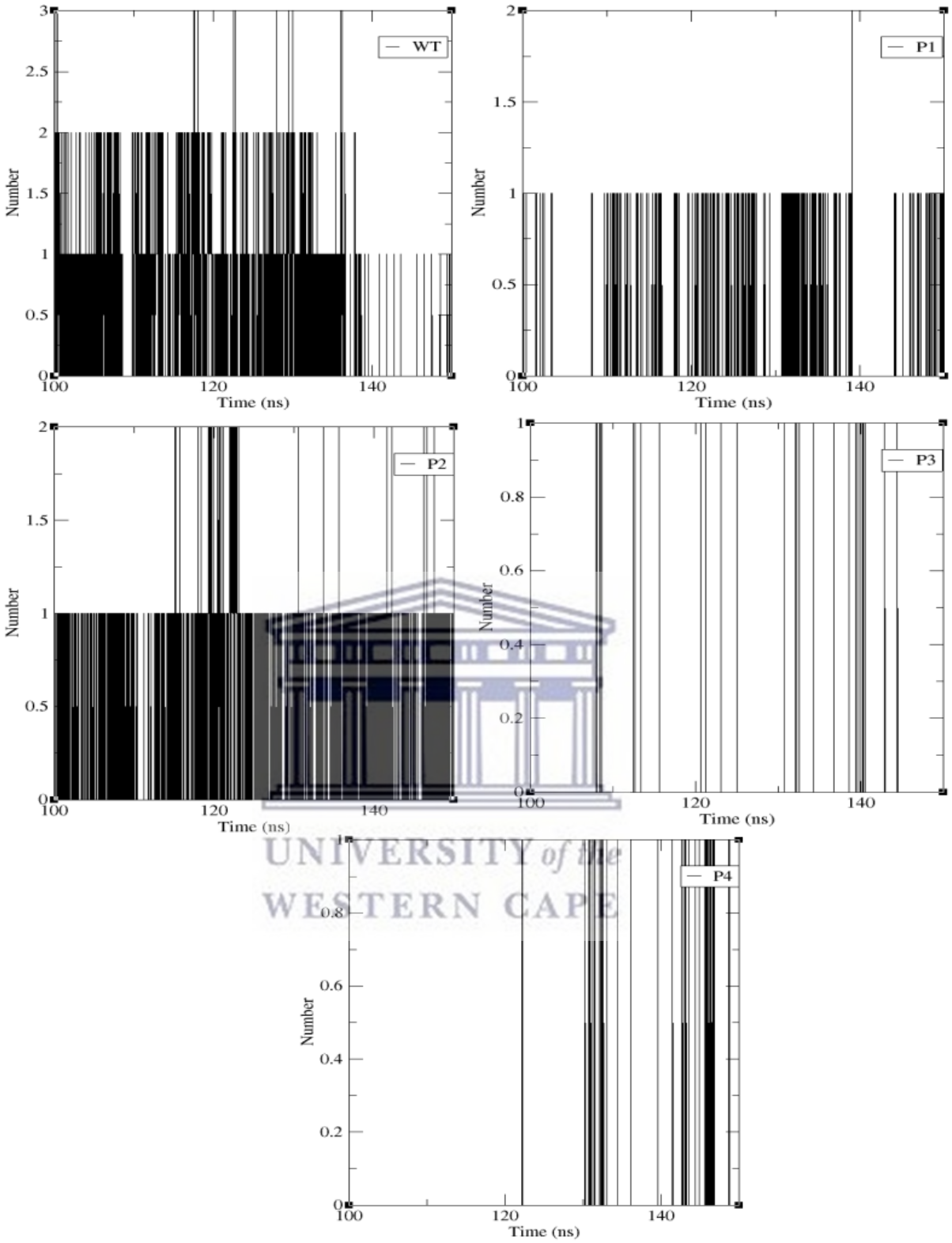


Figure 3.8: Average intramolecular protein-drug HBs amongst the WT HIV-1C_{ZA} IN, its variant systems P1, P2, P3 and P4 and DTG at 300K for the last 50 ns of the simulations. (A) Average number of HBs formed between WT IN and DTG. (B) Average number of HBs formed between mutant P1 IN and DTG. (C) Average number of HBs formed between mutant P2 IN and DTG. (D) Average number of HBs formed between mutant P3 IN and DTG. (E) Average number of HBs formed between mutant P4 IN and DTG.

3.8.5 Pairwise distance analysis

The pairwise distances of the MG ion and the drug DTG in the various HIV-1 IN proteins systems were calculated for the last 50 ns of the simulations. The analysis indicated smaller average distances of 0.208 ± 0.001 nm and 0.397 ± 0.028 nm between the MG ions and DTG in the P2 and WT structures, respectively compared to P1, P3 and P4, each having 0.570 ± 0.039 nm, 0.490 ± 0.040 nm and 0.587 ± 0.041 nm, respectively (Figure 10). From the pairwise analysis of the repeat simulations we obtained average distances of 0.383 ± 0.036 nm, 0.562 ± 0.041 nm, 0.201 ± 0.008 nm, 0.483 ± 0.030 nm and 0.601 ± 0.055 nm for WT, P1, P2, P3 and P4 structures, respectively.

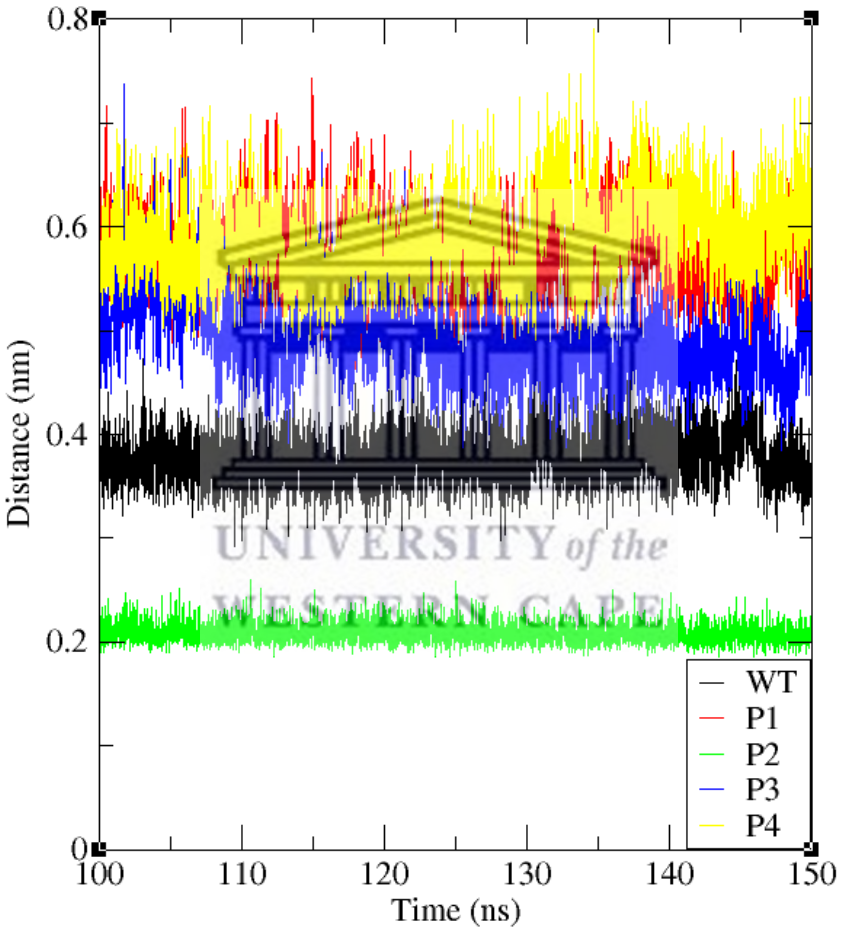


Figure 3.9: Average pairwise distance between MG ions and DTG in the WT and variant systems P1, P2, P3 and P4 at 300K. Minimum distance calculated for the last 50ns of the simulations**.** Black line colour indicates WT IN, red line colour indicates P1 IN (I113V), green line colour indicates P2 IN (L63I, V75M, Y143R), blue line colour indicates P3 IN (S119P, Y143R) and yellow line colour indicates P4 IN (V150A, M154I).

3.8.6 Non-bonded pairwise interaction analysis

In an effort to quantify the strength of the interaction between the HIV-1C IN protein and drug DTG, we computed the non-bonded interaction energy between these two molecules. The total pairwise non-bonded interaction energy (which is not a free energy or a binding energy) between the IN protein and DTG was calculated for the WT and mutant IN complex systems over the full duration of the simulation run time of 150 ns. We included the short-range Coulombic electrostatic interaction energy term and the short-range Lennard-Jones energy term for our analysis, giving the total non-bonded pairwise interaction energy as the sum of the two energy terms considered. The calculation does not include entropy changes. The total non-bonded pairwise interaction energies between the HIV-$1C_{ZA}$ protein and DTG were found to be higher for the WT (-103.215 kJ/mol), compared to the four mutant structures (P1, P2, P3 and P4) each having -66.179, -86.059, -12.728 and -46.780 kJ/mol, respectively. Of all the complex mutant systems, structure P2 had the highest total non-bonded pairwise interaction energy while structure P3 had the lowest (Table 5).

Table 5: Non-bonded interaction energy between each of the five HIV-1C IN protein systems and DTG.

Energy (kJ/Mol)	WT	P1	P2	P3	P4
Coul-SR	-22.512	-7.908	-15.982	+12.044	+3.429
LJ-SR	-80.703	-58.271	-70.077	-24.772	-50.209
Total Interaction energy	-103.215	-66.179	-86.059	-12.728	-46.780

Coul-SR: short-range Coulombic electrostatic interaction energy, LJ-SR: short-range Lennard-Jones interaction energy.

3.8.7 Structural snapshot/conformational analysis

We performed interaction analyses of six snapshots (every 10 ns) for each of the simulations systems to determine which residues played a role in the binding of DTG to the IN protein in the WT and mutant IN protein complex structures. For the WT complex, we observed interactions between DTG, MG, DDE motif resides, DNA nucleotides and also with some amino acid residues (Table 6). Interactions were also observed between DTG, DNA nucleotides and amino acid residues for the P1 system. Interestingly, structure P2 showed the most interactions between DTG and DNA nucleotides and amino acid residues. Additionally, the structure showed DTG interacting with the DDE motif residues D64, D116 and E152 and MG ions (Table 6). The mutation complexes P3 and P4 had interactions between DTG and DNA nucleotides, with P3 having the least number of polar interactions (Table 6).

Table 6: Summary of interaction analysis calculated for each of the six different snapshots of the five systems.

Structure	Cluster	Interactions	
		Hydrogen bonds	**Ionic**
WT	1 (100 ns)	9 (D64, C65, P142, E152, THY11[a], GUA22[a])	MG
	2 (110 ns)	8 (D64, C65, E152, THY11[a], GUA22[a])	MG
	3 (120 ns)	7 (D64, C65, THY11[a], GUA22[a])	MG
	4 (130 ns)	8 (D64, C65, P142, E152, THY11[a])	MG
	5 (140 ns)	9 (D64, C65, E152, THY11[a], GUA22[a])	MG
	6 (150 ns)	6 (C65, THY11[a], GUA22[a])	MG
P1	1 (100 ns)	2 (N144)	None
	2 (110 ns)	1 (P142)	None
	3 (120 ns)	2 (ADE21[a], THY24[a])	None
	4 (130 ns)	3 (ADE21[a], P142, Y143)	None
	5 (140 ns)	3 (ADE21[a]; Y143)	None
	6 (150 ns)	0	None
P2	1 (100 ns)	8 (CYT10[a], THY11[a], D64, D116, P142)	MG
	2 (110 ns)	7 (CYT10[a], THY11[a], GUA22[a], D64, D116, P142)	MG
	3 (120 ns)	5 (THY11[a], GUA22[a], D116, Q148, E152)	MG
	4 (130 ns)	5 (THY11[a], GUA22[a], D64, D116)	MG
	5 (140 ns)	12 (CYT10[a], THY11[a], GUA22[a], D64, D116, P142, E152)	MG
	6 (150 ns)	10 (CYT10[a], THY11[a], D64, D116, P142, E152	MG
P3	1 (100 ns)	0	None
	2 (110 ns)	0	None
	3 (120 ns)	1 (ADE21[a])	None
	4 (130 ns)	1 (ADE27[a])	None
	5 (140 ns)	0	None
	6 (150 ns)	0	None
P4	1 (100 ns)	2 (CYT20[a], ADE21[a])	None
	2 (110 ns)	1 (CYT20[a])	None
	3 (120 ns)	1 (CYT20[a])	None
	4 (130 ns)	2 (GUA19[a], CYT20[a])	None
	5 (140 ns)	3 (CYT20[a], ADE21[a])	None
	6 (150 ns)	3 (CYT20[a], ADE21[a], GUA22[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; N-Asparagine; P-Proline; Q-Glutamine; S - Serine; Y - Tyrosine.

CHAPTER 4 Discussion and Conclusion

4.1 Discussion

The assessment of mutations/variants from our sample cohort of viral sequences led to the discovery of newly identified variants in a South African population that we considered to be novel within the country and continent. Out of the 21 new variants that we identified, a few of these variants have been identified in patients from other countries and were reported in literature. A study done in German HIV-1 non-B subtype infected patients, frequently observed mutations S119P, M154I and T206S in HIV-1 IN for patients that were failing RAL treatment (Sichtig *et al.*, 2009). They classified these mutations as natural polymorphisms. Another laboratory in France also observed mutations S39C, I113V, S119P, I135V, I220L/V, S195G and T206S from four HIV-1B patients who were failing RAL treatment and also classified them as naturally occurring polymorphisms (Malet *et al.*, 2008). Based on these findings and their reports, we may also classify our new variants as naturally occurring polymorphisms associated with RAL failure in HIV-1C patients.

The emergence of cross-resistance to first-generation INSTIs has become a serious problem in the therapy of HIV-1 infection. Several *in silico* studies involving MD simulations, have been performed to investigate and analyze the structural impact of known primary RAMs from RAL and EVG treatment on the HIV-1 IN structure as single or double mutants (Chen *et al*., 2013, 2015; Dewdney *et al*., 2013; Xue *et al*., 2013). These studies have helped to give insight to the mechanisms of resistance imposed by these mutations on the IN structure, but, some of the findings were inconclusive. This was due to the use of incomplete HIV-1 IN structures or poor quality of the protein models used. All of these studies also only considered HIV-1B IN and protein models of low sequence identity. However, this data from these previous studies has helped to build towards the successful modelling of the HIV-1C IN tetramer. In our study, we have successfully modelled a reliable and complete 3D homology model of the HIV-1C IN protein and have made the structure to be available upon request (Chitongo *et al*., 2020). 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 90.0% 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.

Mutations play a fundamental role in evolution by introducing diversity into genomes. Missense mutations in structural genes may become either selectively advantageous or disadvantageous to the organism by affecting protein stability and/or interfering with interactions between partners, with the big challenge being to identify and characterize those genetic mutations that have functional consequences (Pires, Ascher and Blundell, 2014). Predicting the impacts of mutations on protein stability and interactions is fundamental to the understanding of various biological processes, including disease and drug resistance. One of the known primary RAMs associated with RAL failure (Y143R) was selected in our study, together with six other variants new to the SA population (L63I, V75M, I113V, S119P, V150A, M154I), to analyze their effect on HIV-1C IN protein stability. Stability predictions showed contradicting results, whereby an amino acid substitution that resulted in loss of interactions (S119P) was predicted to stabilize the protein structure, but reduce protein-binding affinity, and vice versa. Moreover, the mCSM stability predicting tool had a limitation of its inability to simultaneously consider the effect of multiple mutations on the protein IN structure, which could further assist to have a better understanding of the impact of multiple mutations on the protein stability.

In an effort to fully comprehend the effects of individual, double or triple mutations, we opted to use MD simulations to understand the effect of selected variants and Y143R on protein movement and drug interactions. MD analyses have been shown to be successful in quantifying small changes in protein structures that can affect overall drug binding (Nair and Miners, 2014). Interaction analyses from the various HIV-1C IN complex clusters from the MD simulations verified the mCSM stability prediction results. Variants L63I, V75M and I113V were all calculated to cause instability within the protein structure, but they favoured the protein-DNA affinity. Interaction analysis of clusters for mutant IN complexes P1 (I113V) and P2 (L63I, V75M, Y143R) showed a higher number of polar contacts between the protein with DTG and DNA nucleotides with DTG, suggesting DTG outcompetes viral DNA for host DNA. Pairwise interaction analysis between the MG ions and DTG also showed that structure P2 had the smallest average distance (0.208 ± 0.001 nm), a finding that further supports the increased polar contacts between DTG, MG ions and the DDE motif residues of the HIV-1C IN. Structure P3 contained variants S119P and Y143R, both of which reduced the protein-DNA affinity of the complex, not stabilizing the viral DNA complex. However, S119P also

resulted in a loss of residue interactions, destabilizing the protein structure which can affect DTG drug binding. Furthermore, variants V150A and M154I (mutant system P4) showed reduced protein-DNA affinity, not favouring viral DNA integration, although having polar contacts between DNA nucleotides and DTG meaning these variants have no impact on drug binding.

Additionally, to the interaction analysis of the clusters, hydrogen bonds (HBs) and non-bonded pairwise interaction analysis supported the findings reported above. The WT HIV-1C IN complex had the highest number of possible HBs formed between the protein and DTG, which in turn was confirmed by high non-bonded interaction energy. HIV-1C mutant IN complex P2 also had a high probability of forming HBs between the protein and DTG, a result that was also supported by the high non-bonded interaction energy amongst the mutant complexes. Based on these findings, we can conclude that variants that destabilize the protein structure using the mCSM stability prediction tool, are favourable to the binding of the drug to the IN protein. This is because such variants result in increased HB formation between the IN protein and drug, and they also lead to high non-bonded interaction energies within the structure.

Various trajectory analyses were performed with the WT and mutant IN complexes, with all results being reported relative to the WT structure. From the analyses we found that mutant complex structures P1 (I113V) and P2 (L63I, V75M, Y143R) had increased flexibility in various regions including the 140s loop region. Structure P2 also showed a reduction in structure compactness, implying a possibility of the structure unfolding during simulation but this can only be confirmed by performing long microsecond simulations. Interestingly, analyses of IN structure P3 (S119, Y143R) showed a high deviation of its backbone atoms from the starting structure, suggesting that a combination of mutations S119P and Y143R could result in increased rigidity. Additionally, the same structure had reduced flexibility of residues overall and an increased loss in the compactness of the structure amongst the mutant complexes. On the contrary, mutant complex P4 remained tightly folded and had the lowest deviation of its backbone atoms compared to the WT and other mutant complexes. The role to which variants V150A and M154I play in this observation, still remains unknown.

By virtue of observation, HIV-1C IN mutant complex structures containing mutation Y143R (P2 and P3) were observed to have reduced compactness. The reason for this is unknown, but other reports have also confirmed a similar effect of Y143R on the compactness of the IN, but

only when the mutation is accompanied by another mutation (Kobayashi *et al*., 2011; Cahn *et al*., 2013). Of interest was how complexes P1 and P4 had large distances between their MG ions and DTG relative to other structures. The reason for this is assumed to be due to the increased instability of the IN protein from the mutations in these complexes as indicated by the mCSM webserver results, where I113V (ΔG = -1.817) and V150A (ΔG = -1.509) highly destabilize the IN protein. Another finding from a study on the impact of variant M154I (which is in P4) to the effectiveness of EVG binding showed that the variant mediated resistance to EVG (Hutapea and Widodo, 2018). In their work, Hutapea and Widodo (2018) showed that the variant M154I resulted in EVG binding more strongly to the mutant IN compared to the WT. This contradicts our findings where in the P4 IN complex, the drug is far from the binding pocket compared to the WT IN complex. Additionally, mutant IN complex P4 also had relatively fewer HBs forming between the protein and DTG, which could be due to the increased distance between the binding pocket and DTG. Mutant IN complex P3 had a significant distance also between MG ions and DTG and in turn the least HBs. The reason for this finding is assumed to be a result of the variant S119P that was found to lose all polar contacts from the IN protein after mutation when we analyzed polar contact interactions and therefore destabilize the IN-DNA complex (Table 3). Even though Y143R was found to result in one increased polar contact in the IN structure after mutation (Table 3), it can reduce flexibility of the 140s loop, directly affecting drug binding in combination with S119P which destabilizes the IN structure. To support this effect of variant S119P, an *in vitro* study by Santoro *et al*. (2020) reported high-level resistance to all INSTIs, including bictegravir (BIC) and DTG for isolates that contained S119P and Y143R. Hachiya and team (Hachiya *et al*., 2015) also conducted *in vitro* studies on the impact of S119P to the susceptibility of INSTIs. In their work they observed that variant S119P does not confer much resistance on its own but it rather reduced the susceptibility of INSTIs (mainly RAL) in combination with known primary resistance associated mutations.

4.2 Conclusion

In this study, a complete 3D HIV-1C IN tetrameric protein model of good quality and reliability was reproduced, that can be used for further *in silico* studies for HIV-1C subtype. We also managed to identify new variants from our South African sample cohort which we selectively assessed their impact on the HIV-1C IN protein. From our findings, we can report and conclude that variant S119P in combination with any known RAM might reduce the efficacy of DTG treatment in HIV-1C patients, and this needs to be validated experimentally. Additionally, variants V150A and M154I should also be investigated further, as they too

result in poor drug binding from reduced polar interactions and weaker non-bonded interaction energy. Overall, DTG is susceptible of being effective when used by HIV-1C treatment experienced patients who will be failing cART treatment regimens containing RAL.

4.3 Limitations of the work

Even though we successfully modelled a complete tetrameric homology model of HIV-1C IN and assessed the effects of some new selected polymorphic mutations on the stability of the IN protein and binding of DTG, there are several limitations to this work. Our study was limited to the number of samples that were analysed. More sequences need to be screened, for any new resistance mutations that could be associated with RAL failure from ART treatment-experienced patients. The genetic diversity of HIV-1 continues to complicate the control of the pandemic. This is an evolving epidemic; as such it would be critical to keep monitoring patients for therapy outcomes, when switched to new INSTIs treatment regimen. Future computational work will include binding free energy calculations to include entropic effects on drug binding.

REFERENCES

- Abbondanzieri, E. A., Bokinsky, G., Rausch, J. W., Zhang, J. X., Le Grice, S. F. J. & Zhuang, X. 2008. Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature*, **453**: pp 184–189. https://doi.org/10.1038/nature06941
- Abram, M. E., Hluhanich, R. M., Goodman, D. D., Andreatta, K. N., Margot, N. A., Ye, L., Niedziela-Majka, A., Barnes, T. L., Novikov, N., Chen, X. *et al*. 2013. Impact of Primary Elvitegravir Resistance-Associated Mutations in HIV-1 Integrase on Drug Susceptibility and Viral Replication Fitness. *Antimicrobial Agents and Chemotherapy*, **57** (6): pp 2654-2663. https://doi.org/10.1128/AAC.02568-12
- Adcock, S. A. & McCammon, J. A. 2006. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews,* **106** (5): pp 1589-1615. https://doi.org/10.1021/cr040426m
- Anstett, K., Fusco, R., Cutillas, V., Mesplède, T. & Wainberg, M. A. 2015. Dolutegravir-Selected HIV-1 Containing the N155H and R263K Resistance Substitutions Does Not Acquire Additional Compensatory Mutations under Drug Pressure That Lead to Higher-Level Resistance and Increased Replicative Capacity. *Journal of Virology*, **89** (20): pp 10482-10488. https://doi.org/10.1128/JVI.01725-15
- Anstett, K., Brenner, B., Mesplede, T & Wainberg, M. A. 2017. HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology,* **14**, 36. https://doi.org/10.1186/s12977-017-0360-7
- Autore, F., Bergeron, J. R. C., Malim, M. H., Fraternali, F. & Huthoff, H. 2010. Rationalisation of the Differences between APOBEC3G Structures from Crystallography and NMR Studies by Molecular Dynamics Simulations. *PLoS ONE,* **5** (7): pp e11515. https://doi.org/10.1371/journal.pone.0011515
- Baker, D. & Sali, A. 2001. Protein structure prediction and structural genomics. *Science*, **294** (5540): pp 93-96. https://doi.org/10.1126/science.1065659
- Benkert, P., Biasini, M. & Schwede, T. 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, **27** (3): pp 343–350 https://doi.org/10.1093/bioinformatics/btq662
- Bennett, D.E., Bertagnolio, S., Sutherland, D. & Gilks, C.F. 2008. The World Health Organization's global strategy for prevention and assessment of HIV drug resistance. *Antiviral therapy*, **13** (2): pp 1–13. PMID: 18578063
- Benson, N.C. & Daggett, V. 2012. A comparison of multiscale methods for the analysis of molecular dynamics simulations. *The Journal of Physical Chemistry B*, **116** (29): pp 8722-8731. https://doi.org/10.1021/jp302103t

- Brado, D., Obasa, A. E., Ikomey, G. M., Cloete, R., Singh, K., Engelbrecht, S., Neogi, U. & Jacobs, G. B. 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:pp 4709. https://doi.org/10.1038/s41598-018-22914-5
- Braun, E., Gilmer, J., Mayes, H. B., Mobley, D. L., Monroe, J. I., Prasad, S., & Zuckerman, D. M. 2019. Best Practices for Foundations in Molecular Simulations. *Living journal of computational molecular science*, **1** (1): pp 5957. https://doi.org/10.33011/livecoms.1.1.5957
- Brenner, B.G., Oliveira, M., Doualla-Bell, F., Moisi, D.D., Ntemgwa, M., Frankel, F., Essex, M. & Wainberg, M.A. 2006. HIV-1 subtype C viruses rapidly develop K65R resistance to tenofovir in cell culture. *Aids*, **20** (9): pp F9-F13. https://doi.org/10.1097/01.aids.0000232228.88511.0b
- Brenner, B.G., Thomas, R., Blanco, J.L., Ibanescu, R.I., Oliveira, M., Mesplède, T., Golubkov, O., Roger, M., Garcia, F., Martinez, E. & Wainberg, M.A. 2016. Development of a G118R mutation in HIV-1 integrase following a switch to dolutegravir monotherapy leading to cross-resistance to integrase inhibitors. *Journal of Antimicrobial Chemotherapy*, **71** (7): pp 1948-1953. https://doi.org/10.1093/jac/dkw071
- Brenner, B.G. & Wainberg, M.A. 2017. Clinical benefit of dolutegravir in HIV-1 management related to the high genetic barrier to drug resistance. *Virus research*, **239**: pp 1-9. https://doi.org/10.1016/j.virusres.2016.07.006.
- Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S.A. & Karplus, M. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, **4** (2): pp 187-217. https://doi.org/10.1002/jcc.540040211
- Brooks, B.R., Brooks III, C.L., Mackerell Jr, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S. and Caflisch, A. 2009. CHARMM: the biomolecular simulation program. *Journal of computational chemistry*, **30** (10): pp 1545-1614. https://doi.org/10.1002/jcc.21287
- Bussi, G., Donadio, D. & Parrinello, M. 2007. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, **126** (1): pp 014101. https://doi.org/10.1063/1.2408420
- Cahn, P., Pozniak, A.L., Mingrone, H., Shuldyakov, A., Brites, C., Andrade-Villanueva, J.F., Richmond, G., Buendia, C.B., Fourie, J., Ramgopal, M. & Hagins, D.

2013. Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naive adults with HIV: week 48 results from the randomised, double-blind, non-inferiority SAILING study. *The Lancet*, **382** (9893): pp 700-708. https://doi.org/10.1016/S0140-6736(13)61221-0

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. 2009. BLAST+: architecture and applications. *BMC bioinformatics*, **10** (1): pp 1-9. https://doi.org/10.1186/1471-2105-10-421
- Carvalho, A.T., Fernandes, P.A. & Ramos, M.J. 2006. Insights on resistance to reverse transcriptase: the different patterns of interaction of the nucleoside reverse transcriptase inhibitors in the deoxyribonucleotide triphosphate binding site relative to the normal substrate. *Journal of medicinal chemistry*, **49** (26): pp 7675-7682. https://doi.org/10.1021/jm060698c
- Carvalho, A.T., Fernandes, P.A. & Ramos, M.J. 2007. The excision mechanism in reverse transcriptase: pyrophosphate leaving and fingers opening are uncoupled events with the analogues AZT and d4T. *The Journal of Physical Chemistry B*, **111** (41): pp 12032-12039. https://doi.org/10.1021/jp0746594
- Case, D.A., Cheatham III, T.E., Darden, T., Gohlke, H., Luo, R., Merz Jr, K.M., Onufriev, A., Simmerling, C., Wang, B. & Woods, R.J. 2005. The Amber biomolecular simulation programs. *Journal of computational chemistry*, **26** (16): pp 1668-1688. https://doi.org/10.1002/jcc.20290
- Cecchini, D. M., Castillo, S., Copertari, G., Lacal, V., Rodriguez, C. G., & Cassetti, I. 2019. Resistance to HIV integrase strand transfer inhibitors in Argentina: first interim survey. *Revista espanola de quimioterapia: publicacion oficial de la Sociedad Espanola de Quimioterapia*, **32** (3): pp 263–267. PMID: 31037930; PMCID: PMC6609941.
- Chakrabarti, P. & Pal, D. 2001. The interrelationships of side-chain and main-chain conformations in proteins. *Progress in biophysics and molecular biology*, **76** (1-2): pp 1-102. https://doi.org/10.1016/S0079-6107(01)00005-0.
- Charpentier, C. & Descamps, D. 2018. Resistance to HIV integrase inhibitors: about R263K and E157Q mutations. *Viruses*, **10** (1): pp 41. https://doi.org/10.3390/v10010041
- Chen, J.C.H., Krucinski, J., Miercke, L.J., Finer-Moore, J.S., Tang, A.H., Leavitt, A.D. & Stroud, R.M. 2000. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proceedings of the National Academy of Sciences*, **97** (15): pp 8233-8238. https://doi.org/10.1073/pnas.150220297

- Chen, J., Zhang, S., Liu, X. & Zhang, Q. 2010. Insights into drug resistance of mutations D30N and I50V to HIV-1 protease inhibitor TMC-114: free energy calculation and molecular dynamic simulation. *Journal of molecular modeling*, **16** (3): pp 459-468. https://doi.org/10.1007/s00894-009-0553-7
- Chen, Q., Buolamwini, J.K., Smith, J.C., Li, A., Xu, Q., Cheng, X. & Wei, D. 2013. Impact of resistance mutations on inhibitor binding to HIV-1 integrase. J*ournal of chemical information and modeling*, **53** (12): pp 3297-3307. https://doi.org/10.1021/ci400537n
- Chen, Q., Cheng, X., Wei, D. & Xu, Q. 2015. Molecular dynamics simulation studies of the wild type and E92Q/N155H mutant of Elvitegravir-resistance HIV-1 integrase. *Interdisciplinary Sciences: Computational Life Sciences*, **7** (1): pp 36-42. https://doi.org/10.1007/s12539-014-0235-8
- Cherepanov, P., Sun, Z.Y.J., Rahman, S., Maertens, G., Wagner, G. & Engelman, A. 2005. Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nature structural & molecular biology*, **12** (6): pp 526-532. https://doi.org/10.1038/nsmb937
- Chitongo, R., Obasa, A.E., Mikasi, S.G., Jacobs, G.B. & 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): pp e0223464. https://doi.org/10.1371/journal.pone.0234581
- Christen, M., Hünenberger, P.H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D.P., Heinz, T.N., Kastenholz, M.A., Kräutler, V., Oostenbrink, C. & Peter, C. 2005. The GROMOS software for biomolecular simulation: GROMOS05. *Journal of computational chemistry*, **26** (16): pp 1719-1751. https://doi.org/10.1002/jcc.20303
- Colovos, C. & Yeates, T.O. 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, **2**: pp 1511-1519. https://doi.org/10.1002/pro.5560020916
- Cortez, K.J. & Maldarelli, F. 2011. Clinical Management of HIV Drug Resistance. *Viruses,* **3**: pp 347-378. https://doi.org/10.3390/v3040347
- DeLano, W.L. 2002. Unraveling hot spots in binding interfaces: progress and challenges. *Current opinion in structural biology*, **12** (1): pp 14-20. https://doi.org/10.1016/S0959-440X(02)00283-X.
- Deng, N.J., Zheng, W., Gallicchio, E. & Levy, R.M. 2011. Insights into the dynamics of HIV-1 protease: a kinetic network model constructed from atomistic simulations.

Journal of the American Chemical Society, **133** (24): pp 9387-9394. https://doi.org/10.1021/ja2008032

- Dewdney, T.G., Wang, Y., Kovari, I.A., Reiter, S.J. & Kovari, L.C. 2013. Reduced HIV-1 integrase flexibility as a mechanism for raltegravir resistance. *Journal of structural biology*, **184** (2): pp 245-250. https://doi.org/10.1016/j.jsb.2013.07.008
- Dodson, G.G., Lane, D.P. & Verma, C.S. 2008. Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO reports*, **9** (2): pp 144-150. https://doi.org/10.1038/sj.embor.7401160
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., & Liu, X. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human molecular genetics*, **24** (8): pp 2125–2137. https://doi.org/10.1093/hmg/ddu733
- Doyle, T., Dunn, D.T., Ceccherini-Silberstein, F., De Mendoza, C., Garcia, F., Smit, E., Fearnhill, E., Marcelin, A.G., Martinez-Picado, J., Kaiser, R. & Geretti, A.M. 2015. Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades. *Journal of Antimicrobial Chemotherapy*, **70** (11): pp 3080-3086. https://doi.org/10.1093/jac/dkv243
- Dror, R.O., Jensen, M.Ø., Borhani, D.W. & Shaw, D.E. 2010. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *Journal of General Physiology*, **135** (6): pp 555-562. https://doi.org/10.1085/jgp.200910373
- Durrant, J.D. & McCammon, J.A. 2011. Molecular dynamics simulations and drug discovery. *BMC biology*, **9** (1): pp 1-9. https://doi.org/10.1186/1741-7007-9-71
- Eisenberg, D., Bowie, J.U., Lüthy, R. & Choe, S. 1992. Three-dimensional profiles for analysing protein sequence–structure relationships. *Faraday discussions*, **93**: pp 25-34. https://doi.org/10.1039/FD9929300025
- Eron, J.J., Cooper, D.A., Steigbigel, R.T., Clotet, B., Gatell, J.M., Kumar, P.N., Rockstroh, J.K., Schechter, M., Markowitz, M., Yeni, P. & Loutfy, M.R. 2013. Efficacy and safety of raltegravir for treatment of HIV for 5 years in the BENCHMRK studies: final results of two randomised, placebo-controlled trials. *The Lancet Infectious Diseases*, **13** (7): pp 587-596. https://doi.org/10.1016/S1473-3099(13)70093-8
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. & Pedersen, L.G. 1995. A smooth particle mesh Ewald method. *The Journal of chemical physics*, **103** (19): pp 8577-8593. https://doi.org/10.1063/1.470117

- Evans, D.J. & Holian, B.L. 1985. The nose–hoover thermostat. *The Journal of chemical physics*, **83** (8): pp 4069-4074. https://doi.org/10.1063/1.449071
- Forli, S. & Olson, A.J. 2015. Computational challenges of structure-based approaches applied to HIV. *The Future of HIV-1 Therapeutics:* pp 31-51. https://doi.org/10.1007/82_2015_432
- Foulkes-Murzycki, J.E., Scott, W.R.P. & Schiffer, C.A. 2007. Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure*, **15** (2): pp 225-233. https://doi.org/10.1016/j.str.2007.01.006
- Fourati, S., Lambert-Niclot, S., Soulie, C., Wirden, M., Malet, I., Valantin, M.A., Tubiana, R., Simon, A., Katlama, C., Carcelain, G. & Calvez, V. 2014. Differential impact of APOBEC3-driven mutagenesis on HIV evolution in diverse anatomical compartments. *Aids*, **28** (4): pp 487-491. https://doi.org/10.1097/QAD.0000000000000182
- Geretti, A.M., Harrison, L., Green, H., Sabin, C., Hill, T., Fearnhill, E., Pillay, D., Dunn, D. & UK Collaborative Group on HIV Drug Resistance and the UK Collaborative HIV Cohort Study. 2009. Effect of HIV-1 subtype on virologic and immunologic response to starting highly active antiretroviral therapy. *Clinical infectious diseases*, **48** (9): pp 1296-1305. https://doi.org/10.1086/598502
- Grobler, J.A., Stillmock, K., Hu, B., Witmer, M., Felock, P., Espeseth, A.S., Wolfe, A., Egbertson, M., Bourgeois, M., Melamed, J. & Wai, J.S. 2002. Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proceedings of the National Academy of Sciences*, **99** (10): pp 6661-6666. https://doi.org/10.1073/pnas.092056199
- Groenhof, G. 2013. Introduction to QM/MM simulations. *Biomolecular Simulations:* pp 43-66. https://doi.org/10.1007/978-1-62703-017-5_3
- Guex, N., Diemand, A. & Peitsch, M.C. 1999. Protein modelling for all. *Trends in biochemical sciences*, **24** (9): pp 364-367. https://doi.org/10.1016/S0968-0004(99)01427-9
- Gupta, V.P. 2016. Electron density analysis and electrostatic potential. *Principles and Applications of Quantum Chemistry; Elsevier: Amsterdam, The Netherlands*: pp 195-214.
- Hachiya, A., Ode, H., Matsuda, M., Kito, Y., Shigemi, U., Matsuoka, K., Imamura, J., Yokomaku, Y., Iwatani, Y. & Sugiura, W. 2015. Natural polymorphism S119R of HIV-1 integrase enhances primary INSTI resistance. *Antiviral research*, **119**: pp 84-88. https://doi.org/10.1016/j.antiviral.2015.04.014.

- Häggblom, A., Svedhem, V., Singh, K., Sönnerborg, A. & Neogi, U. 2016. Virological failure in patients with HIV-1 subtype C receiving antiretroviral therapy: an analysis of a prospective national cohort in Sweden. *The lancet HIV*, **3** (4): pp e166-e174. https://doi.org/10.1016/S2352-3018(16)00023-0.
- Han, Y.S., Mesplède, T. & Wainberg, M.A. 2016. Differences among HIV-1 subtypes in drug resistance against integrase inhibitors. *Infection, Genetics and Evolution*, **46**: pp 286-291. https://doi.org/10.1016/j.meegid.2016.06.047.
- Hardy, I., Brenner, B., Quashie, P., Thomas, R., Petropoulos, C., Huang, W., Moisi, D., Wainberg, M.A. & Roger, M. 2015. Evolution of a novel pathway leading to dolutegravir resistance in a patient harbouring N155H and multiclass drug resistance. *Journal of Antimicrobial Chemotherapy*, **70** (2): pp 405-411. https://doi.org/10.1093/jac/dku387
- Hare, S., Maertens, G. N. & Cherepanov, P. 2012. 3′-Processing and strand transfer catalysed by retroviral integrase in crystallo. *The EMBO Journal*, **31**: pp 3020-3028. https://doi.org/10.1038/emboj.2012.118
- Heger, E., Theis, A.A., Remmel, K., Walter, H., Pironti, A., Knops, E., Di Cristanziano, V., Jensen, B., Esser, S., Kaiser, R. & Lübke, N. 2016. Development of a phenotypic susceptibility assay for HIV-1 integrase inhibitors. *Journal of virological methods*, **238**: pp 29-37. https://doi.org/10.1016/j.jviromet.2016.10.002.
- Henzler-Wildman, K.A., Lei, M., Thai, V., Kerns, S.J., Karplus, M. & Kern, D. 2007. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, **450** (7171): pp 913-916. https://doi.org/10.1038/nature06407
- Henzler-Wildman, K. & Kern, D. 2007. Dynamic personalities of proteins. *Nature*, **450** (7172): pp 964-972. https://doi.org/10.1038/nature06522
- Hoover, W.G. 1985. Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*, **31** (3): pp 1695. https://link.aps.org/doi/10.1103/PhysRevA.31.1695
- Hornak, V., Okur, A., Rizzo, R.C. & Simmerling, C. 2006. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, **103** (4): pp 915-920. https://doi.org/10.1073/pnas.0508452103
- Hu, Z. & Kuritzkes, D.R. 2010. Effect of raltegravir resistance mutations in HIV-1 integrase on viral fitness. *Journal of acquired immune deficiency syndromes*, **55** (2): pp 148. https://doi.org/10.1097/QAI.0b013e3181e9a87a

- Hu, Z. & Kuritzkes, D.R. 2011. Interaction of reverse transcriptase (RT) mutations conferring resistance to lamivudine and etravirine: effects on fitness and RT activity of human immunodeficiency virus type 1. *Journal of virology*, **85** (21): pp 11309-11314. https://doi.org/10.1128/JVI.05578-11
- Huang, J. & MacKerell, A. D. 2013. CHARMM36 all atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.*, **34**: pp 2135–2145. https://doi.org/10.1002/jcc.23354
- Hutapea, H.M.L. & Maladan, Y. 2018. Relationship between HIV integrase polymorphisms and integrase inhibitor susceptibility: An in silico analysis. *Heliyon*, **4** (12): pp e00956. https://doi.org/10.1016/j.heliyon.2018.e00956.
- Inzaule, S.C., Hamers, R.L., Noguera-Julian, M., Casadellà, M., Parera, M., Rinke de Wit, T.F. & Paredes, R. 2018. Primary resistance to integrase strand transfer inhibitors in patients infected with diverse HIV-1 subtypes in sub-Saharan Africa. *Journal of Antimicrobial Chemotherapy*, **73** (5): pp 1167-1172. https://doi.org/10.1093/jac/dky005
- Jo, S., Kim, T., Iyer, V.G. & Im, W. 2008. CHARMM GUI: a web based graphical user interface for CHARMM. *Journal of computational chemistry*, **29** (11): pp 1859-1865. https://doi.org/10.1002/jcc.20945
- Kagan, R.M., Shenderovich, M.D., Heseltine, P.N. & Ramnarayan, K. 2005. Structural analysis of an HIV-1 protease I47A mutant resistant to the protease inhibitor lopinavir. *Protein science*, **14** (7): pp 1870-1878. https://doi.org/10.1110/ps.051347405
- Kar, P. & Knecht, V. 2012. Energetic basis for drug resistance of HIV-1 protease mutants against amprenavir. *Journal of computer-aided molecular design*, **26** (2): pp 215-232. https://doi.org/10.1007/s10822-012-9550-5
- Karplus, M. & Kuriyan, J. 2005. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences*, **102** (19): pp 6679-6685. https://doi.org/10.1073/pnas.0408930102
- Kessl, J.J., McKee, C.J., Eidahl, J.O., Shkriabai, N., Katz, A. & Kvaratskhelia, M. 2009. HIV-1 integrase-DNA recognition mechanisms. *Viruses*, **1** (3): pp 713-736. https://doi.org/10.3390/v1030713
- Kim, T., Kasprzak, W.K. & Shapiro, B.A. 2017. Protocols for Molecular Dynamics Simulations of RNA Nanostructures. In *RNA Nanostructures, Humana Press, New York, NY*: pp 33-64. https://doi.org/10.1007/978-1-4939-7138-1_3

- Kirmizialtin, S., Nguyen, V., Johnson, K.A. & Elber, R. 2012. How conformational dynamics of DNA polymerase select correct substrates: experiments and simulations. *Structure*, **20** (4): pp 618-627. https://doi.org/10.1016/j.str.2012.02.018.
- Klepeis, J.L., Lindorff-Larsen, K., Dror, R.O. & Shaw, D.E. 2009. Long-timescale molecular dynamics simulations of protein structure and function. *Current opinion in structural biology*, **19** (2): pp 120-127. https://doi.org/10.1016/j.sbi.2009.03.004.
- Kobayashi, M., Yoshinaga, T., Seki, T., Wakasa-Morimoto, C., Brown, K.W., Ferris, R., Foster, S.A., Hazen, R.J., Miki, S., Suyama-Kagitani, A. & Kawauchi-Miki, S. 2011. In vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrobial agents and chemotherapy*, **55** (2): pp 813-821. https://doi.org/10.1128/AAC.01209-10
- Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y. & Jo, S. 2016. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of chemical theory and computation*, **12** (1): pp 405-413. https://doi.org/10.1021/acs.jctc.5b00935
- Lemkul, J. 2019. From proteins to perturbed Hamiltonians: A suite of tutorials for the GROMACS-2018 molecular simulation package [article v1. 0]. *Living Journal of Computational Molecular Science*, **1** (1): pp 5068.
- Leong, I. U., Stuckey, A., Lai, D., Skinner, J. R. & Love, D.R. 2015. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC medical genetics*, **16** (1): pp 1-13. https://doi.org/10.1186/s12881-015-0176-z
- Levi, J., Raymond, A., Pozniak, A., Vernazza, P., Kohler, P. & Hill, A. 2016. Can the UNAIDS 90-90-90 target be achieved? A systematic analysis of national HIV treatment cascades. *BMJ global health*, **1** (2). https://doi.org/10.1136/bmjgh-2015000010
- Liang, J., Mesplède, T., Oliveira, M., Anstett, K. & Wainberg, M.A. 2015. The combination of the R263K and T66I resistance substitutions in HIV-1 integrase is incompatible with high-level viral replication and the development of high-level drug resistance. *Journal of virology*, **89** (22): pp 11269-11274. https://doi.org/10.1128/JVI.01881-15
- Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M.P., Dror, R.O. & Shaw, D.E. 2012. Systematic validation of protein force fields against experimental data. *PloS one*, **7** (2): pp e32131. https://doi.org/10.1371/journal.pone.0032131

- Little, S.J., Frost, S.D., Wong, J.K., Smith, D.M., Pond, S.L.K., Ignacio, C.C., Parkin, N.T., Petropoulos, C.J. & Richman, D.D. 2008. Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection. *Journal of virology*, **82** (11): pp 5510-5518. https://doi.org/10.1128/JVI.02579-07
- Lobanov, M.Y., Bogatyreva, N.S. & Galzitskaya, O.V. 2008. Radius of gyration as an indicator of protein structure compactness. *Mol Biol,* **42**: pp 623–628. https://doi.org/10.1134/S0026893308040195
- Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D. & Darnell, J. 2000. Biochemical Energetics. In *Molecular Cell Biology. 4th edition*. WH Freeman.
- Maertens, G., Hare, S. & Cherepanov, P. 2010. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**: pp 326–329. https://doi.org/10.1038/nature09517
- Malet, I., Delelis, O., Valantin, M.A., Montes, B., Soulie, C., Wirden, M., Tchertanov, L., Peytavin, G., Reynes, J., Mouscadet, J.F. & Katlama, C. 2008. Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. *Antimicrobial agents and chemotherapy*, **52** (4): pp 1351-1358. https://doi.org/10.1128/AAC.01228-07
- Malet, I., Gimferrer Arriaga, L., Artese, A., Costa, G., Parrotta, L., Alcaro, S., Delelis, O., Tmeizeh, A., Katlama, C., Valantin, M.A. & Ceccherini-Silberstein, F. 2014. New raltegravir resistance pathways induce broad cross-resistance to all currently used integrase inhibitors. *Journal of Antimicrobial Chemotherapy*, **69** (8): pp 2118-2122. https://doi.org/10.1093/jac/dku095
- Manasa, J., Katzenstein, D., Cassol, S., Newell, M.L. & de Oliveira, for the Southern Africa Treatment and Resistance Network, T. 2012. Primary drug resistance in South Africa: data from 10 years of surveys. *AIDS research and human retroviruses*, **28** (6): pp 558-565. http://doi.org/10.1089/aid.2011.0284
- Martínez, L. 2015. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PloS one*, **10** (3): pp e0119264. https://doi.org/10.1371/journal.pone.0119264
- Matsuyama, S., Aydan, A., Ode, H., Hata, M., Sugiura, W. and Hoshino, T., 2010. Structural and energetic analysis on the complexes of clinically isolated subtype C HIV-1 proteases and approved inhibitors by molecular dynamics simulation. *The Journal of Physical Chemistry B*, **114** (1): pp 521-530. https://doi.org/10.1021/jp908314f

- Mboumba Bouassa, R.S., Mossoro-Kpinde, C.D., Gody, J.C., Veyer, D., Péré, H., Matta, M., Robin, L., Grésenguet, G., Charpentier, C. & Bélec, L. 2019. High predictive efficacy of integrase strand transfer inhibitors in perinatally HIV-1-infected African children in therapeutic failure of first-and second-line antiretroviral drug regimens recommended by the WHO. *Journal of Antimicrobial Chemotherapy*, **74** (7): pp 2030-2038. https://doi.org/10.1093/jac/dkz099
- McGee, K.S., Okeke, N.L., Hurt, C.B. & McKellar, M.S. 2018, November. Canary in the coal mine? Transmitted mutations conferring resistance to all integrase strand transfer inhibitors in a treatment-naive patient. In *Open forum infectious diseases*, **5** (11): pp 294). US: Oxford University Press. https://doi.org/10.1093/ofid/ofy294
- Mesplede, T., Osman, N., Wares, M., Quashie, P.K., Hassounah, S., Anstett, K., Han, Y., Singhroy, D.N. & Wainberg, M.A. 2014. Addition of E138K to R263K in HIV integrase increases resistance to dolutegravir, but fails to restore activity of the HIV integrase enzyme and viral replication capacity. *Journal of Antimicrobial Chemotherapy*, **69** (10), pp 2733-2740. https://doi.org/10.1093/jac/dku199
- Mesplède, T., Quashie, P.K., Hassounah, S., Osman, N., Han, Y., Liang, J., Singhroy, D.N. & Wainberg, M.A. 2015. The R263K substitution in HIV-1 subtype C is more deleterious for integrase enzymatic function and viral replication than in subtype B. *Aids*, **29** (12): pp 1459-1466. https://doi.org/10.1097/QAD.0000000000000752
- 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. & Ikomey, G.M. 2020. HIV-1 integrase diversity and resistance-associated mutations and polymorphisms among integrase strand transfer inhibitor-naive HIV-1 patients from Cameroon. *AIDS research and human retroviruses*, **36** (5): pp 450-455. http://doi.org/10.1089/aid.2019.0264
- Miller, J.C., Zhang, L., Xia, D.F., Campo, J.J., Ankoudinova, I.V., Guschin, D.Y., Babiarz, J.E., Meng, X., Hinkley, S.J., Lam, S.C. & Paschon, D.E. 2015. Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nature methods*, **12** (5): pp 465-471. https://doi.org/10.1038/nmeth.3330
- Mills, E.J., Nachega, J.B., Buchan, I., Orbinski, J., Attaran, A., Singh, S., Rachlis, B., Wu, P., Cooper, C., Thabane, L. & Wilson, K. 2006. Adherence to antiretroviral therapy in sub-Saharan Africa and North America: a meta-analysis. *Jama*, **296** (6): pp 679-690. https://doi.org/10.1001/jama.296.6.679
- Mondi, A., Cozzi□Lepri, A., Tavelli, A., Rusconi, S., Vichi, F., Ceccherini□Silberstein, F., Calcagno, A., De Luca, A., Maggiolo, F., Marchetti, G. & Antinori, A.

2019. Effectiveness of dolutegravir□based regimens as either first□line or switch antiretroviral therapy: data from the Icona cohort. *Journal of the International AIDS Society*, **22** (1): pp e25227. https://doi.org/10.1002/jia2.25227

- Mouscadet, J.F., Delelis, O., Marcelin, A.G. & Tchertanov, L. 2010. Resistance to HIV-1 integrase inhibitors: a structural perspective. *Drug Resistance Updates*, **13** (4-5): pp 139-150. https://doi.org/10.1016/j.drup.2010.05.001.
- Naeger, L.K., Harrington, P., Komatsu, T. & Deming, D. 2016. Effect of dolutegravir functional monotherapy on HIV-1 virological response in integrase strand transfer inhibitor resistant patients. *Antivir Ther*, **21** (6): pp 481-488. https://doi.org/10.3851/IMP3033
- Nair, P.C. & Miners, J.O. 2014. Molecular dynamics simulations: from structure function relationships to drug discovery. *In silico pharmacology*, **2** (1): pp 1-4. https://doi.org/10.1186/s40203-014-0004-8
- Nosé, S. 1984. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, **52** (2): pp 255-268. https://doi.org/10.1080/00268978400101201
- Obasa, A.E., Mikasi, S.G., Brado, D., Cloete, R., Singh, K., Neogi, U. & Jacobs, G.B. 2020. Drug resistance mutations against protease, reverse transcriptase and integrase inhibitors in people living with HIV-1 receiving boosted protease inhibitors in South Africa. *Frontiers in microbiology*, **11**: pp 438. https://doi.org/10.3389/fmicb.2020.00438
- Obasa, A.E.A. 2019. Multidisciplinary viral analyses in People Living with HIV-1C and receiving second-line combination antiretroviral therapy (cART) in South Africa *(Doctoral dissertation, Stellenbosch: Stellenbosch University)*.
- Ode, H., Neya, S., Hata, M., Sugiura, W. & Hoshino, T. 2006. Computational simulations of HIV-1 proteases multi-drug resistance due to nonactive site mutation L90M. *Journal of the American Chemical Society*, **128** (24): pp 7887-7895. https://doi.org/10.1021/ja060682b
- Ode, H., Matsuyama, S., Hata, M., Hoshino, T., Kakizawa, J. & Sugiura, W. 2007. Mechanism of drug resistance due to N88S in CRF01_AE HIV-1 protease, analyzed by molecular dynamics simulations. *Journal of medicinal chemistry*, **50** (8): pp 1768-1777. https://doi.org/10.1021/jm061158i
- Ode, H., Nakashima, M., Kitamura, S., Sugiura, W. & Sato, H. 2012. Molecular dynamics simulation in virus research. *Frontiers in microbiology*, **3**: pp 258. https://doi.org/10.3389/fmicb.2012.00258

- Osman, N., Mesplede, T., Quashie, P.K., Oliveira, M., Zanichelli, V. & Wainberg, M.A. 2015. Dolutegravir maintains a durable effect against HIV replication in tissue culture even after drug washout. *Journal of Antimicrobial Chemotherapy*, **70** (10): pp 2810-2815. https://doi.org/10.1093/jac/dkv176
- Özen, A., Haliloğlu, T. & Schiffer, C.A. 2011. Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. *Journal of molecular biology*, **410** (4): pp 726-744. https://doi.org/10.1016/j.jmb.2011.03.053.
- Parczewski, M., Leszczyszyn-Pynka, M., Witak-Jędra, M., Szetela, B., Gąsiorowski, J., Knysz, B., Bociąga-Jasik, M., Skwara, P., Grzeszczuk, A., Jankowska, M. & Barałkiewicz, G. 2017. Expanding HIV-1 subtype B transmission networks among men who have sex with men in Poland. *PLoS One*, **12** (2): pp e0172473. https://doi.org/10.1371/journal.pone.0172473
- Park, T.E., Mohamed, A., Kalabalik, J. & Sharma, R. 2015. Review of integrase strand transfer inhibitors for the treatment of human immunodeficiency virus infection. *Expert review of anti-infective therapy*, **13** (10): pp 1195-1212. https://doi.org/10.1586/14787210.2015.1075393
- Passos, D.O., Li, M., Yang, R., Rebensburg, S.V., Ghirlando, R., Jeon, Y., Shkriabai, N., Kvaratskhelia, M., Craigie, R. & Lyumkis, D. 2017. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science*, **355** (6320): pp 89-92. https://doi.org/10.1126/science.aah5163
- Pham, H.T., Labrie, L., Wijting, I.E., Hassounah, S., Lok, K.Y., Portna, I., Goring, M.E., Han, Y., Lungu, C., Van Der Ende, M.E. & Brenner, B.G. 2018. The S230R integrase substitution associated with virus load rebound during dolutegravir monotherapy confers low-level resistance to integrase strand-transfer inhibitors. *The Journal of infectious diseases*, **218** (5): pp 698-706. https://doi.org/10.1093/infdis/jiy175
- Pires, D.E., Ascher, D.B. & Blundell, T.L. 2014. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, **30** (3): pp 335-342. https://doi.org/10.1093/bioinformatics/btt691
- Pontius, J., Richelle, J. & Wodak, S.J. 1996. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *Journal of molecular biology*, **264** (1): pp 121-136. https://doi.org/10.1006/jmbi.1996.0628
- Price, M.A., Wallis, C.L., Lakhi, S., Karita, E., Kamali, A., Anzala, O., Sanders, E.J., Bekker, L.G., Twesigye, R., Hunter, E. & Kaleebu, P. 2011. Transmitted HIV type 1 drug resistance among individuals with recent HIV infection in East and Southern

Africa. *AIDS research and human retroviruses*, **27** (1): pp 5-12. http://doi.org/10.1089/aid.2010.0030

- Puertas, M.C., Ploumidis, G., Ploumidis, M., Fumero, E., Clotet, B., Walworth, C.M., Petropoulos, C.J. & Martinez-Picado, J. 2020. Pan-resistant HIV-1 emergence in the era of integrase strand-transfer inhibitors: a case report. *The Lancet Microbe*, **1** (3): pp e130-e135. https://doi.org/10.1016/S2666-5247(20)30006-9
- Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configurations. *J. mol. Biol*, **7**: pp 95-99.
- Remmert, M., Biegert, A., Hauser, A. & Söding, J. 2012. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment, *Nature Methods,* **9**: pp 173–175. https://doi.org/10.1038/nmeth.1818
- Rogers, L., Obasa, A.E., Jacobs, G.B., Sarafianos, S.G., Sönnerborg, A., Neogi, U. & Singh, K. 2018. Structural implications of genotypic variations in HIV-1 integrase from diverse subtypes. *Frontiers in microbiology,* **9**: pp 1754. https://doi.org/10.3389/fmicb.2018.01754
- Sadiq, S.K., Wan, S. & Coveney, P.V. 2007. Insights into a mutation-assisted lateral drug escape mechanism from the HIV-1 protease active site. *Biochemistry*, **46** (51): pp 14865-14877. https://doi.org/10.1021/bi700864p
- Santoro, M.M., Fornabaio, C., Malena, M., Galli, L., Poli, A., Menozzi, M., Zazzi, M., White, K.L., Castagna, A. & PRESTIGIO Study Group. 2020. Susceptibility to HIV-1 integrase strand transfer inhibitors (INSTIs) in highly treatment-experienced patients who failed an INSTI-based regimen. *International Journal of Antimicrobial Agents*, **56** (1): pp 106027. https://doi.org/10.1016/j.ijantimicag.2020.106027
- Shah, S.M., Heylen, E., Srinivasan, K., Perumpil, S. & Ekstrand, M.L. 2014. Reducing HIV stigma among nursing students: a brief intervention. *Western journal of nursing research*, **36** (10): pp 1323-1337. https://doi.org/10.1177/0193945914523685
- Sichtig, N., Sierra, S., Kaiser, R., Däumer, M., Reuter, S., Schülter, E., Altmann, A., Fätkenheuer, G., Dittmer, U., Pfister, H. & Esser, S. 2009. Evolution of raltegravir resistance during therapy. *Journal of Antimicrobial Chemotherapy*, **64** (1): pp 25-32. https://doi.org/10.1093/jac/dkp153
- Sidibé, M., Loures, L. & Samb, B. 2016. The UNAIDS 90–90–90 target: a clear choice for ending AIDS and for sustainable health and development. *Journal of the International AIDS Society*, **19** (1). https://doi.org/10.7448/IAS.19.1.21133
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J. & Thompson, J.D. 2011. Fast, scalable

generation of high□quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, **7** (1): pp 539. https://doi.org/10.1038/msb.2011.75

- da Silva, H.H.S.A., Pereira, N., Brandão, L., Crovella, S. & Moura, R. 2019. Prediction of HIV integrase resistance mutation using in silico approaches. *Infection, Genetics and Evolution*, **68**: pp 10-15. https://doi.org/10.1016/j.meegid.2018.11.014.
- van der Spoel, D. & Hess, B. 2011. GROMACS—the road ahead. *WIREs Comput Mol Sci*, **1**: pp 710-715. https://doi.org/10.1002/wcms.50
- Sutherland, K.A., Collier, D.A., Claiborne, D.T., Prince, J.L., Deymier, M.J., Goldstein, R.A., Hunter, E. & Gupta, R.K. 2016. Wide variation in susceptibility of transmitted/founder HIV-1 subtype C Isolates to protease inhibitors and association with in vitro replication efficiency. *Scientific reports*, **6** (1): pp 1-7. https://doi.org/10.1038/srep38153
- Theys, K., Camacho, R.J., Gomes, P., Vandamme, A.M., Rhee, S.Y. & Portuguese HIV-1 Resistance Study Group. 2015. Predicted residual activity of rilpivirine in HIV-1 infected patients failing therapy including NNRTIs efavirenz or nevirapine. *Clinical Microbiology and Infection*, **21** (6): pp 607-e1. https://doi.org/10.1016/j.cmi.2015.02.011
- Thibaut, D.M., Napravnik, S., Zakharova, O., Wohl, D.A. & Farel, C.E. 2019. Effectiveness of integrase strand transfer inhibitors among treatment-experienced patients in a clinical setting. *AIDS (London, England)*, **33** (7): pp 1187. https://doi.org/10.1097/QAD.0000000000002194
- Thorpe, I.F. & Brooks, C.L. 2007. Molecular evolution of affinity and flexibility in the immune system. *Proceedings of the National Academy of Sciences*, **104** (21): pp 8821-8826. https://doi.org/10.1073/pnas.0610064104
- Tsiang, M., Jones, G.S., Goldsmith, J., Mulato, A., Hansen, D., Kan, E., Tsai, L., Bam, R.A., Stepan, G., Stray, K.M. & Niedziela-Majka, A. 2016. Antiviral activity of bictegravir (GS-9883), a novel potent HIV-1 integrase strand transfer inhibitor with an improved resistance profile. *Antimicrobial agents and chemotherapy*, **60** (12): pp 7086-7097. https://doi.org/10.1128/AAC.01474-16
- Unaids, G.H., AIDS statistics—2019 fact sheet (2019).
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I. & Mackerell Jr, A.D. 2010. CHARMM general force field: A force field for drug□like molecules compatible with

the CHARMM all□atom additive biological force fields. *Journal of computational chemistry*, **31** (4): pp 671-690. https://doi.org/10.1002/jcc.21367

- Vavro, C., Hasan, S., Madsen, H., Horton, J., DeAnda, F., Martin-Carpenter, L., Sato, A., Cuffe, R., Chen, S., Underwood, M. & Nichols, G. 2013. Prevalent polymorphisms in wild-type HIV-1 integrase are unlikely to engender drug resistance to dolutegravir (S/GSK1349572). *Antimicrobial agents and chemotherapy*, **57** (3): pp 1379-1384. https://doi.org/10.1128/AAC.01791-12
- Wadhwa, P., Jain, P., Rudrawar, S. & Jadhav, H.R. 2018. Quinoline, coumarin and other heterocyclic analogs based HIV-1 integrase inhibitors. *Current drug discovery Technologies*, **15** (1): pp 2-19. https://doi.org/10.2174/1570163814666170531115452
- Wainberg, M.A. & Han, Y.S. 2015. Will drug resistance against dolutegravir in initial therapy ever occur?. *Frontiers in pharmacology*, **6**: pp 90. https://doi.org/10.3389/fphar.2015.00090
- Wainberg, M.A., Mesplède, T. & Raffi, F. 2013. What if HIV were unable to develop resistance against a new therapeutic agent?. *BMC medicine*, **11** (1): pp 1-6. https://doi.org/10.1186/1741-7015-11-249
- Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. & Case, D.A. 2004. Development and testing of a general amber force field. *Journal of computational chemistry*, **25** (9): pp 1157-1174. https://doi.org/10.1002/jcc.20035
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L. & Lepore, R. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, **46** (W1): pp W296-W303. https://doi.org/10.1093/nar/gky427
- Wensing, A.M., Calvez, V., Günthard, H.F., Johnson, V.A., Paredes, R., Pillay, D., Shafer, R.W. & Richman, D.D. 2016. 2017 update of the drug resistance mutations in HIV-1. *Topics in antiviral medicine*, **24** (4): pp 132. PMID: 28208121; PMCID: PMC5677049.
- Wijting, I., Rokx, C., Boucher, C., de Vries-Sluijs, D., Schurink, K., Andrinopoulou, E. & van Gorp, E.C.M. 2017, February. Dolutegravir as maintenance monotherapy for HIV-1: a randomized clinical trial. In *Conference on Retroviruses and Opportunistic Infections*.
- Wilkinson, A.L., El-Hayek, C., Spelman, T., Fairley, C., Leslie, D., McBryde, E., Hellard, M. & Stoové, M. 2015. "Seek, test, treat" lessons from Australia: a study of HIV testing patterns from a cohort of men who have sex with men. *JAIDS Journal of*

Acquired Immune Deficiency Syndromes, **69** (4): pp 460-465. https://doi.org/10.1097/QAI.0000000000000613

- Wohl, D.A., Cohen, C., Gallant, J.E., Mills, A., Sax, P.E., DeJesus, E., Zolopa, A., Liu, H.C., Plummer, A., White, K.L. & Cheng, A.K. 2014. A randomized, double-blind comparison of single-tablet regimen elvitegravir/cobicistat/emtricitabine/tenofovir DF versus single-tablet regimen efavirenz/emtricitabine/tenofovir DF for initial treatment of HIV-1 infection: analysis of week 144 results. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, **65** (3): pp e118-e120. https://doi.org/10.1097/QAI.0000000000000057
- Wong-Ekkabut, J. & Karttunen, M. 2012. Assessment of common simulation protocols for simulations of nanopores, membrane proteins, and channels. *Journal of chemical theory and computation*, **8** (8): pp 2905-2911. https://doi.org/10.1021/ct3001359
- Xue, W., Liu, H. & Yao, X. 2012. Molecular mechanism of HIV-1 integrase–vDNA interactions and strand transfer inhibitor action: A molecular modeling perspective. *Journal of computational chemistry*, **33** (5): pp 527-536. https://doi.org/10.1002/jcc.22887
- Xue, W., Jin, X., Ning, L., Wang, M., Liu, H. & Yao, X. 2013. Exploring the molecular mechanism of cross-resistance to HIV-1 integrase strand transfer inhibitors by molecular dynamics simulation and residue interaction network analysis. *Journal of chemical information and modeling*, **53** (1): pp 210-222. https://doi.org/10.1021/ci300541c
- Yazdanpanah, Y., Fagard, C., Descamps, D., Taburet, A.M., Colin, C., Roquebert, B., Katlama, C., Pialoux, G., Jacomet, C., Piketty, C. & Bollens, D. 2009. High rate of virologic suppression with raltegravir plus etravirine and darunavir/ritonavir among treatment-experienced patients infected with multidrug-resistant HIV: results of the ANRS. *Clinical Infectious Diseases*, **49** (9): pp 1441-1449. https://doi.org/10.1086/630210
- You, J., Wang, H., Huang, X., Qin, Z., Deng, Z., Luo, J., Wang, B. & Li, M. 2016. Therapy-emergent drug resistance to integrase strand transfer inhibitors in HIV-1 patients: A subgroup meta-analysis of clinical trials. *PLoS One*, **11** (8): pp e0160087. https://doi.org/10.1371/journal.pone.0160087
- Zhou, Z., Madrid, M., Evanseck, J.D. & Madura, J.D. 2005. Effect of a bound non-nucleoside RT inhibitor on the dynamics of wild-type and mutant HIV-1 reverse transcriptase. *Journal of the American Chemical Society*, **127** (49): pp 17253-17260.

https://doi.org/10.1021/ja053973d

APPENDICES

Appendix A: Ethics Letter

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Approval Letter
Progress Report

07/04/2020

Project ID: 2215

Ethics Reference No: N15/08/071

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

Dear Dr. Graeme Jacobs

We refer to your request for an extension/annual renewal of ethics approval dated 28/03/2020.

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

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

Approval date: 07 April 2020

Expiry date: 06 April 2021

1. Kindly note that although the study has been granted ethics approval, the study may not proceed during the current national lockdown as an embargo has been placed on studies that require interaction with research participants in order to prevent potential harm to participants.
2. HREC will publish on the HREC website a date when the said embargo is to be lifted taking into consideration the best interest of participants and national interests around COVID-19.

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

Where to submit any documentation

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

Please remember to use your Project Id 2215 and ethics reference number N15/08/071 on any documents or correspondence with the HREC concerning your research protocol.

Yours sincerely,

Mrs. Brightness Nxumalo
Coordinator: Health Research Ethics Committee 2

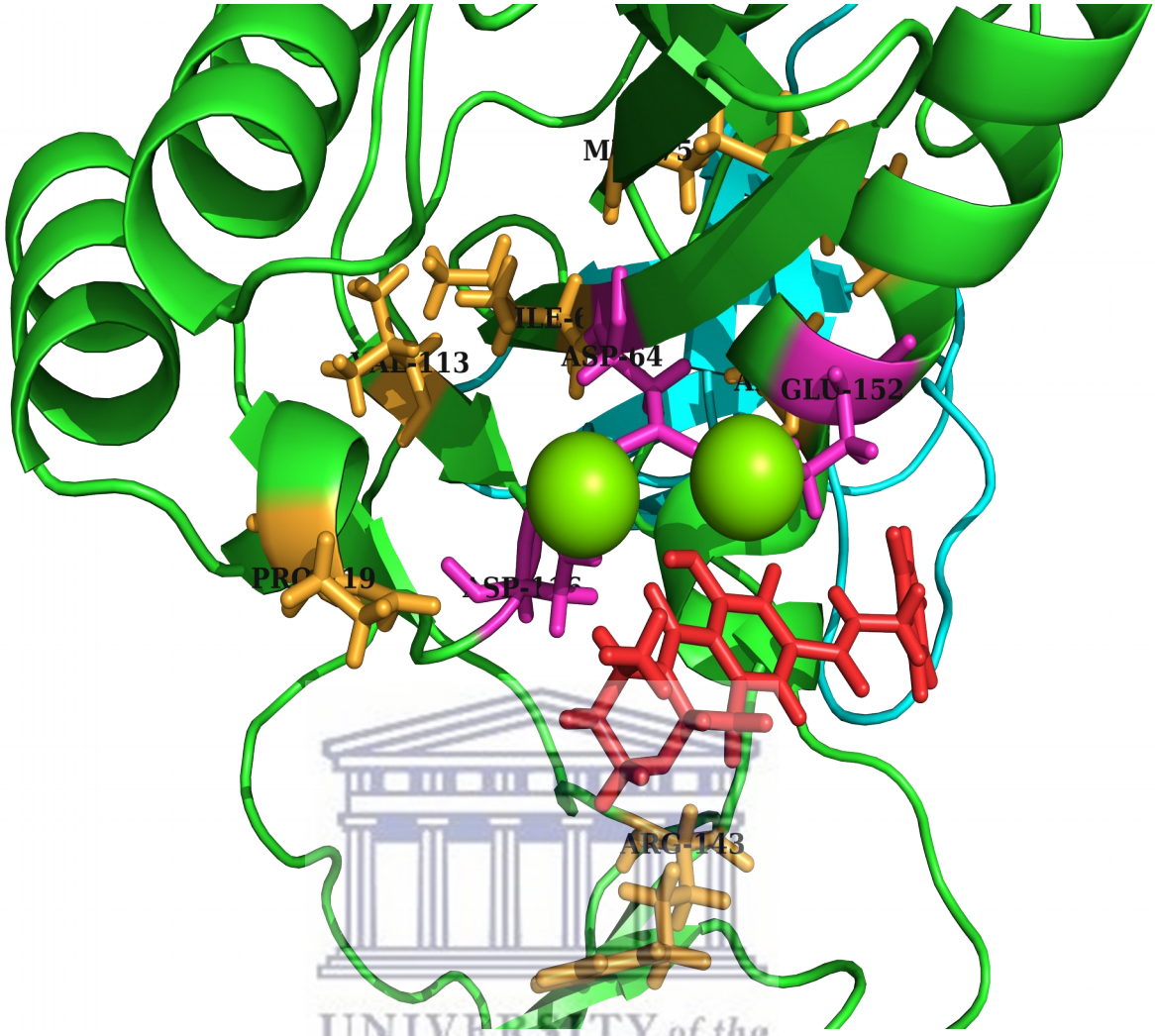
National Health Research Ethics Council (NHREC) Registration Number:
REC-130408-012 (HREC1)•REC-230208-010 (HREC2)

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

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

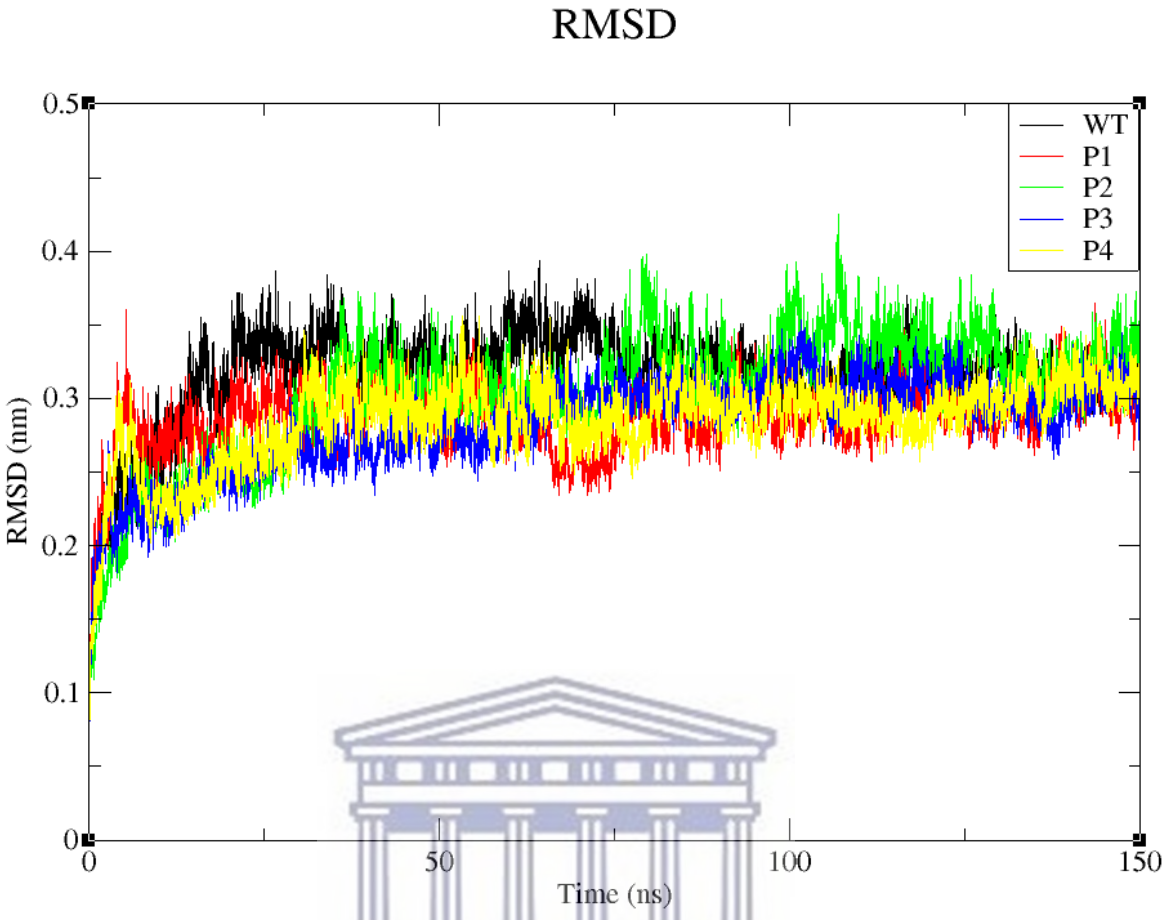
The Health Research Ethics Committee reviews research involving human subjects conducted or supported by the Department of Health and Human Services, or other federal departments or agencies that apply the Federal Policy for the Protection of Human Subjects to such research (United States Code of Federal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

Appendix B: Figure showing localization of variants



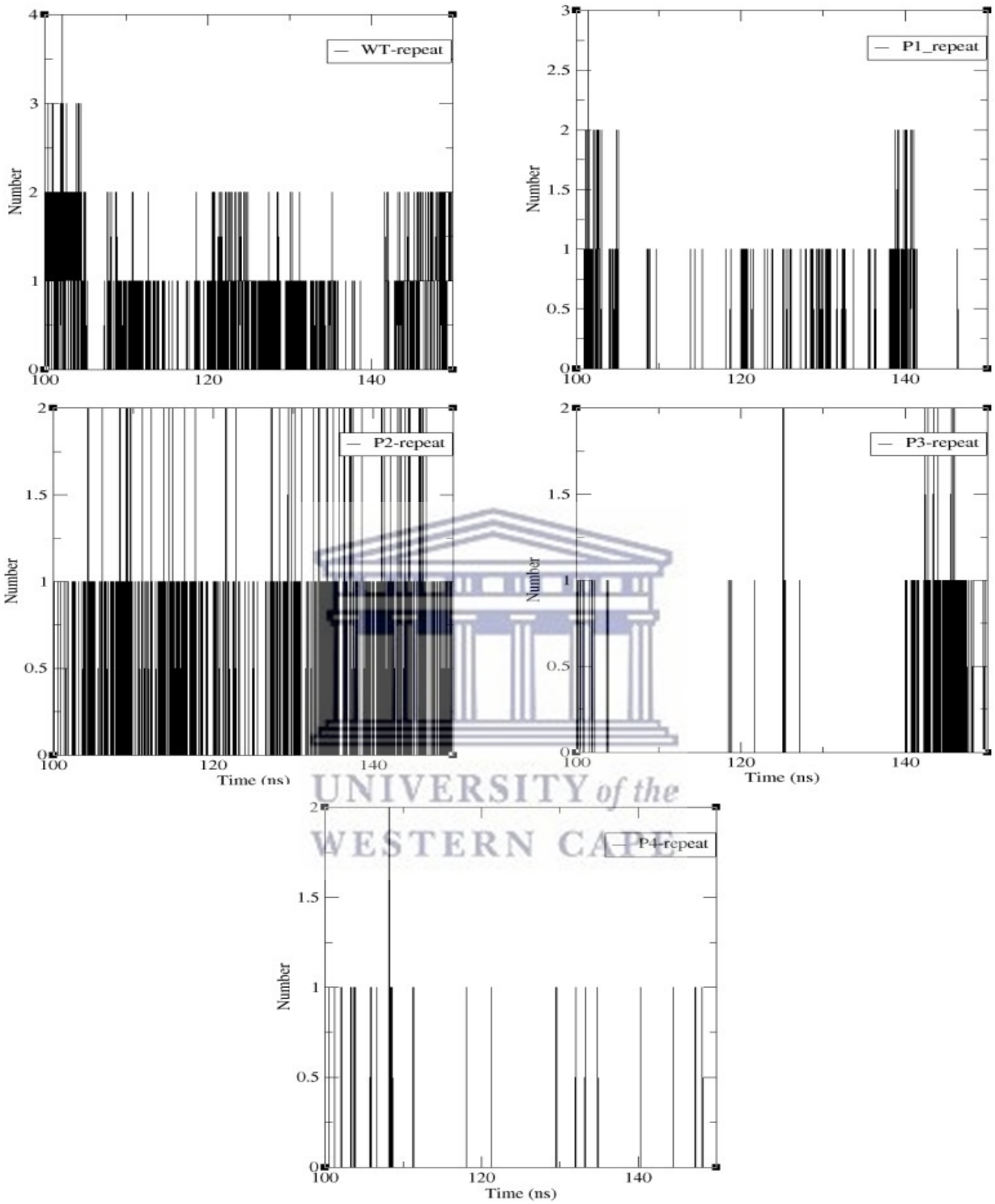
HIV-1C IN showing the localization of mutations and their overall positions relative to the catalytic actuve site (DDE motif) and drug DTG. Drug DTG is coloured red, DDE motif residues (D64, D116, E152) are coloured magenta and selected variants (I63, M75, V113, P119, R143, A150 and I154) are coloured orange.

Appendix C: Repeated backbone RMSDs



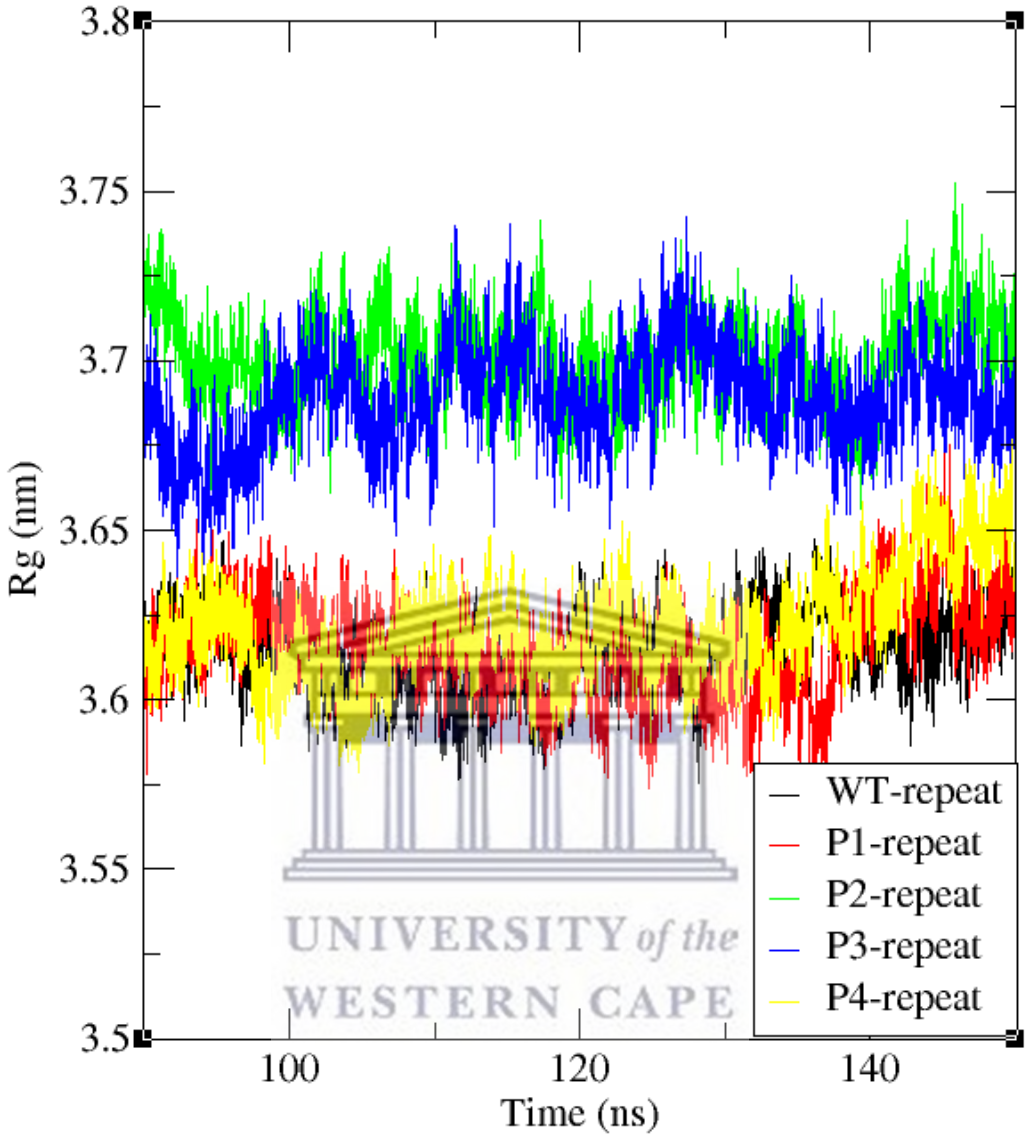
Backbone RMSDs of the WT HIV-1C$_{ZA}$ IN protein and variant systems P1, P2, P3 and P4 at 300K are shown as a function of time. At approximately 80 ns all systems reached equilibrium. Black line colour indicates WT IN, red line colour indicates P1 IN (I113V), green line colour indicates P2 IN (L63I, V75M, Y143R), blue line colour indicates P3 IN (S119P, Y143R) and yellow line colour indicates P4 IN (V150A, M154I).

Appendix D: Repeated average intramolecular protein-drug HBs



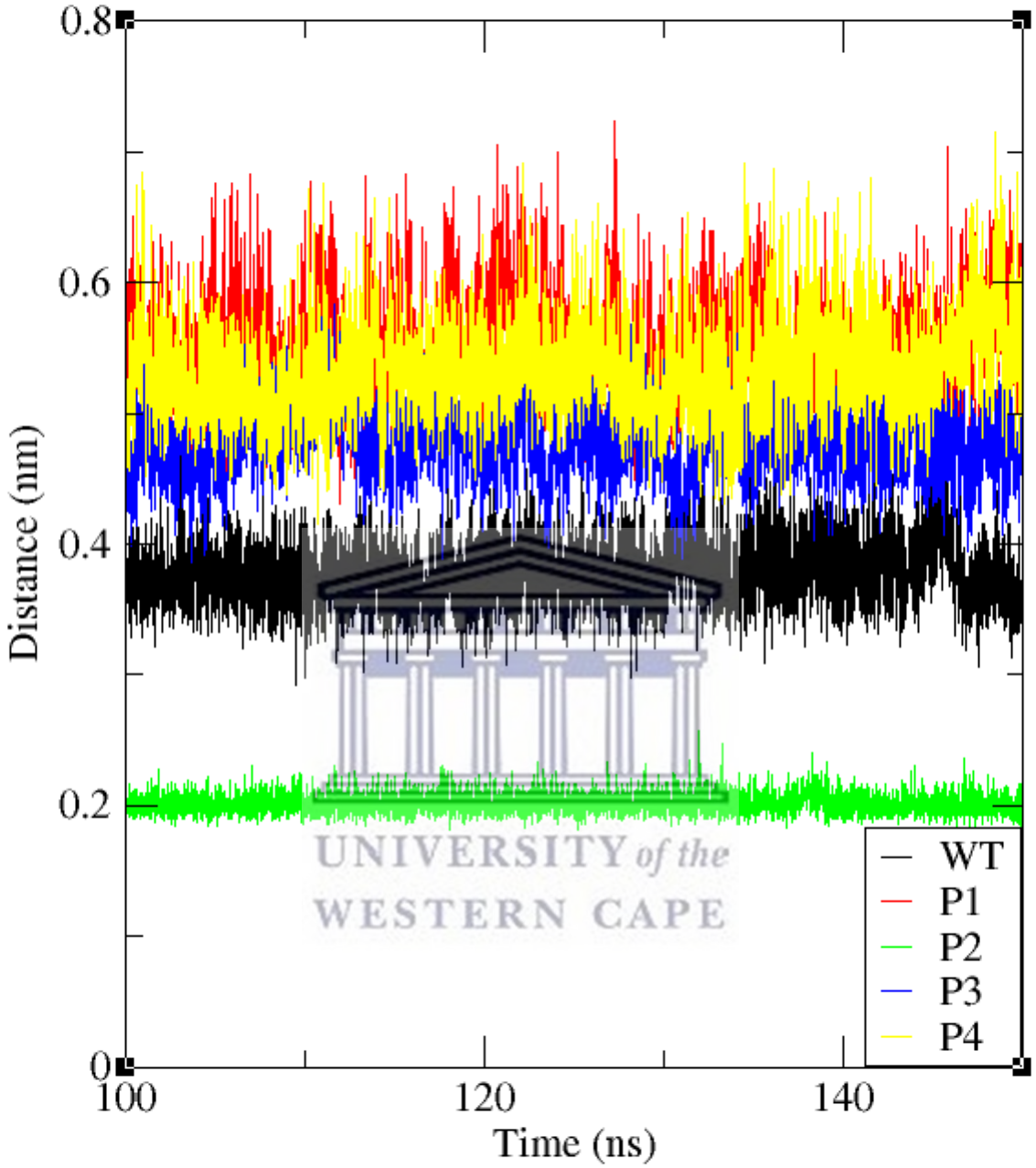
Average intramolecular protein-drug HBs amongst the WT HIV-1C_{ZA} IN, its variant systems P1, P2, P3 and P4 and DTG at 300K for the last 50 ns of the simulations. (A) Average number of HBs formed between WT IN and DTG. (B) Average number of HBs formed between mutant P1 IN and DTG. (C) Average number of HBs formed between mutant P2 IN and DTG. (D) Average number of HBs formed between mutant P3 IN and DTG. (E) Average number of HBs formed between mutant P4 IN and DTG.

Appendix E: Repeated Rg of backbone atoms



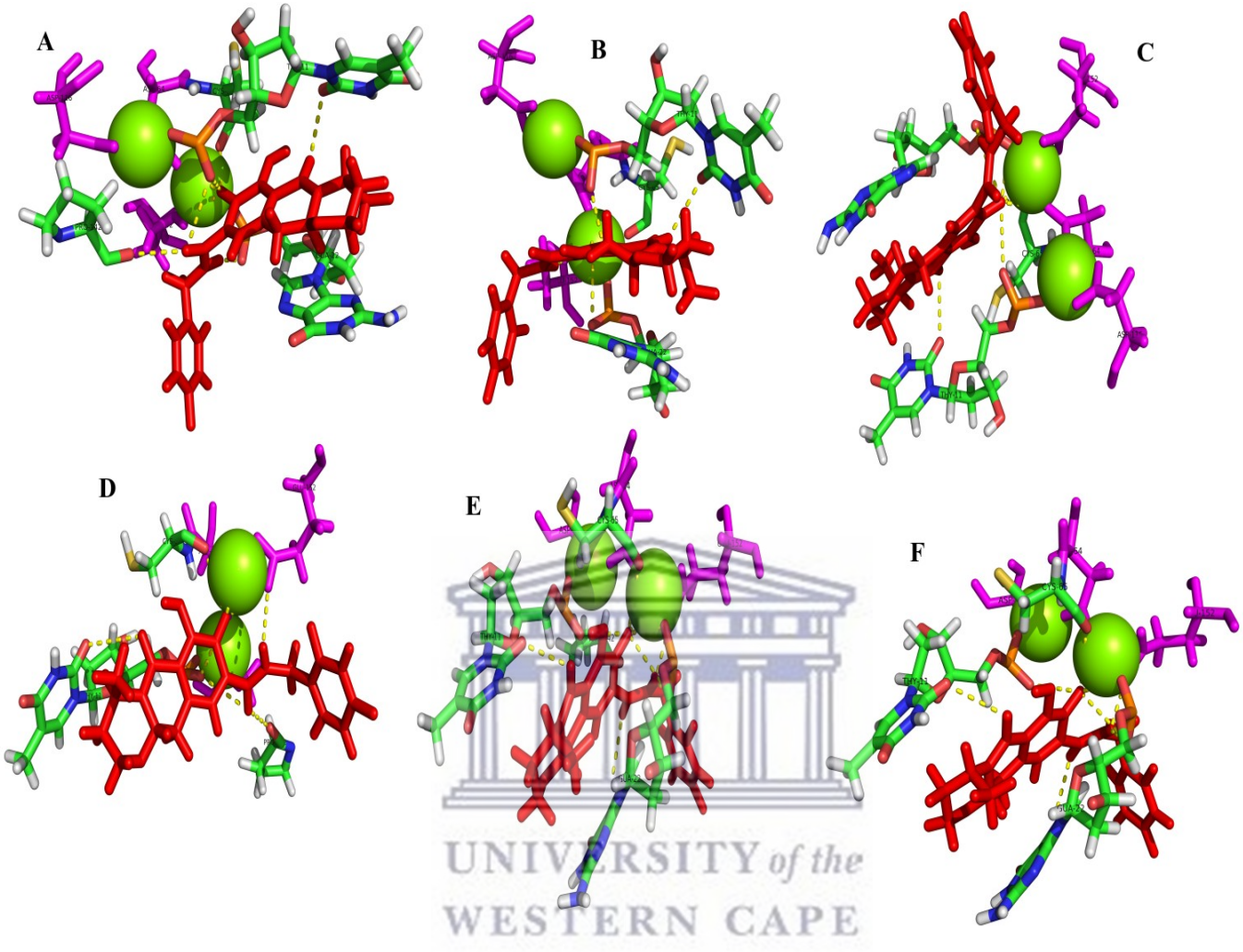
Rg for backbone atoms of WT HIV-1C$_{ZA}$ IN protein and variant systems P1, P2, P3 and P4 at 300K are shown as a function of time. Measure of compactness for the different complex systems plotted over the last 60ns. Black line colour indicates WT IN, red line colour indicates P1 IN (I113V), green line colour indicates P2 IN (L63I, V75M, Y143R), blue line colour indicates P3 IN (S119P, Y143R) and yellow line colour indicates P4 IN (V150A, M154I).

Appendix F: Repeated average pairwise distance

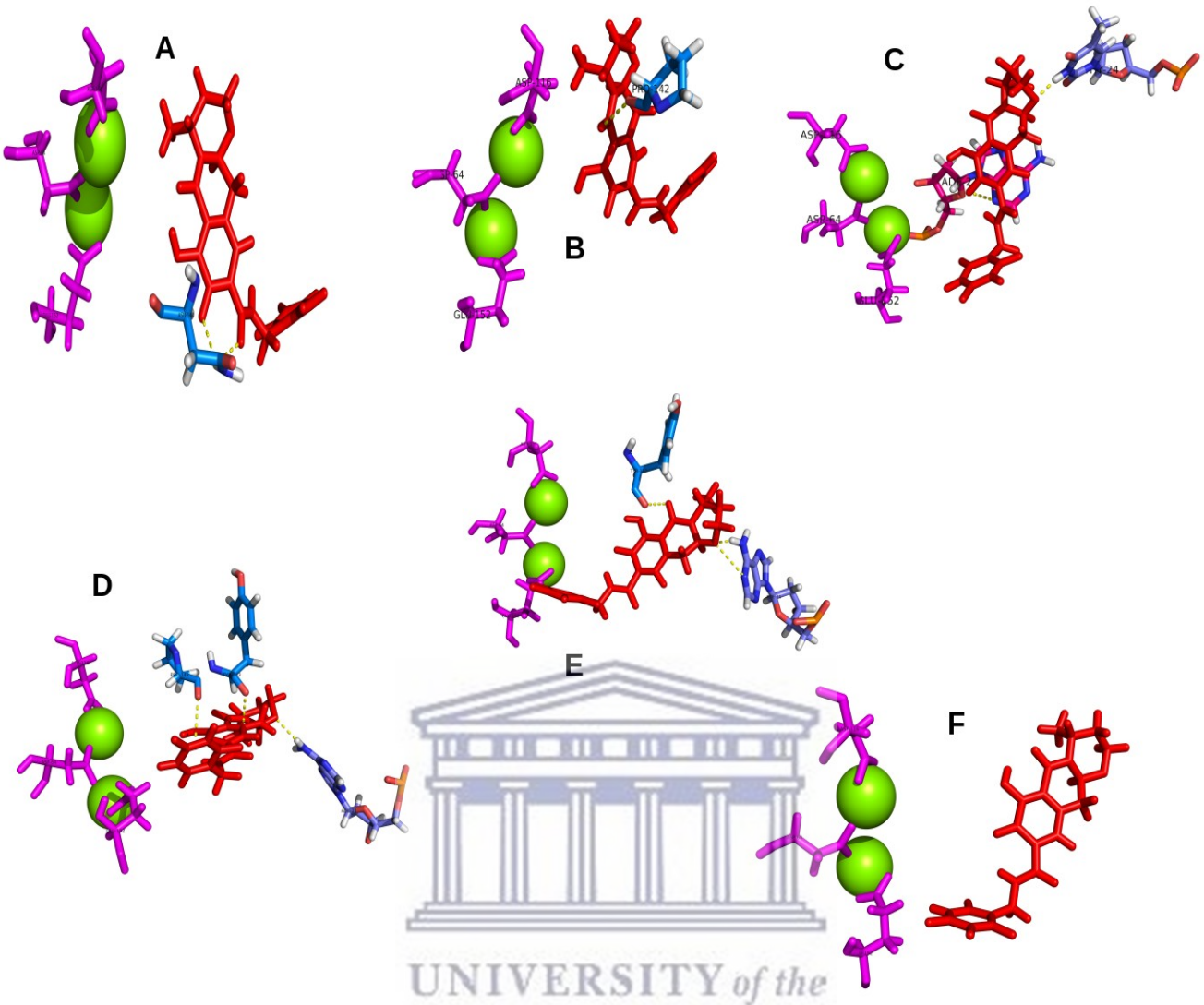


Average pairwise distance between MG ions and DTG in the WT and variant systems P1, P2, P3 and P4 at 300K. Minimum distance calculated for the last 50ns of the simulations. Black line colour indicates WT IN, red line colour indicates P1 IN (I113V), green line colour indicates P2 IN (L63I, V75M, Y143R), blue line colour indicates P3 IN (S119P, Y143R) and yellow line colour indicates P4 IN (V150A, M154I).

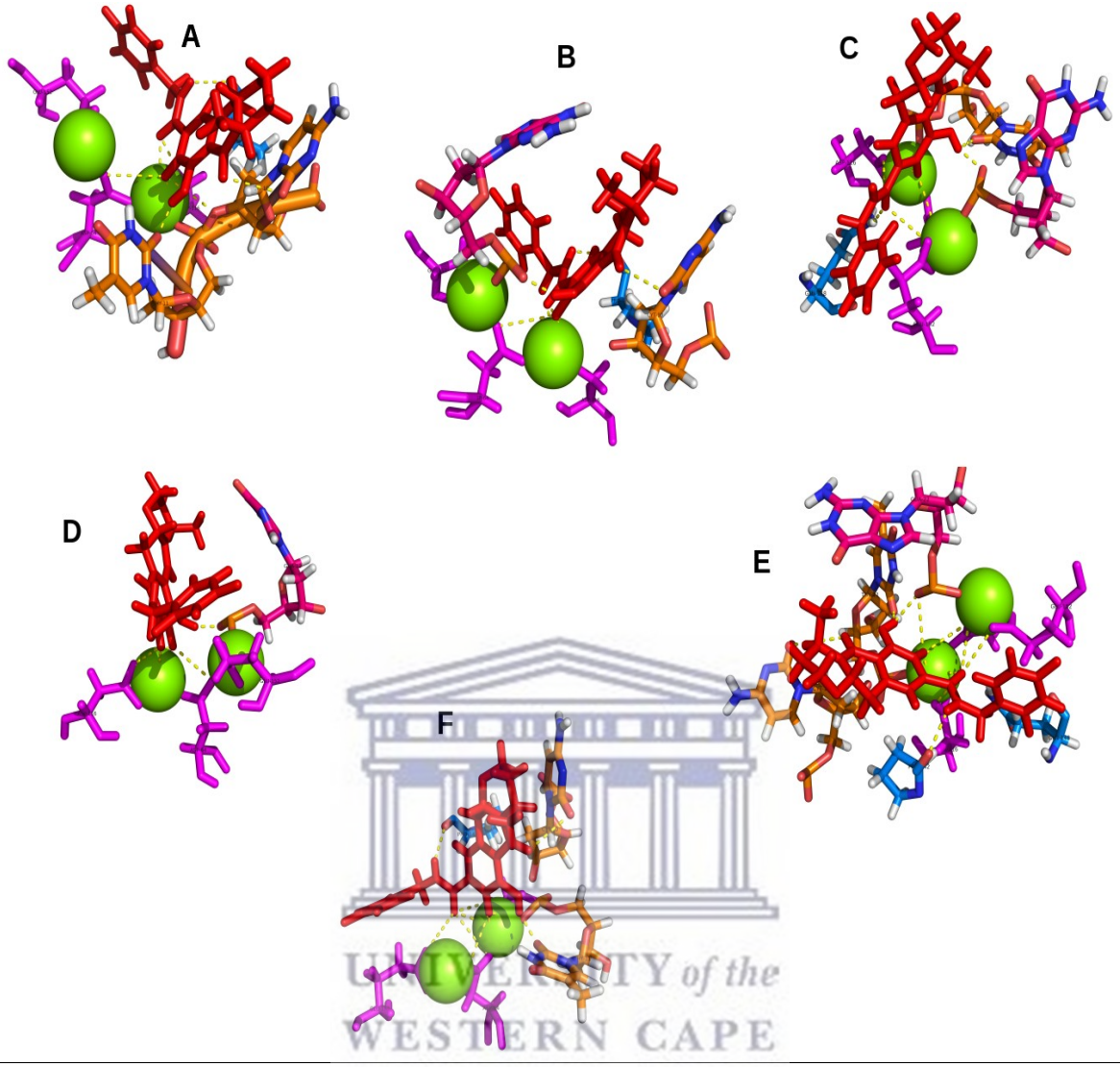
Appendix G: Polar interaction analyses on extracted simulation clusters for the WT HIV-$1C_{ZA}$ IN and variant systems P1, P2, P3 and P4



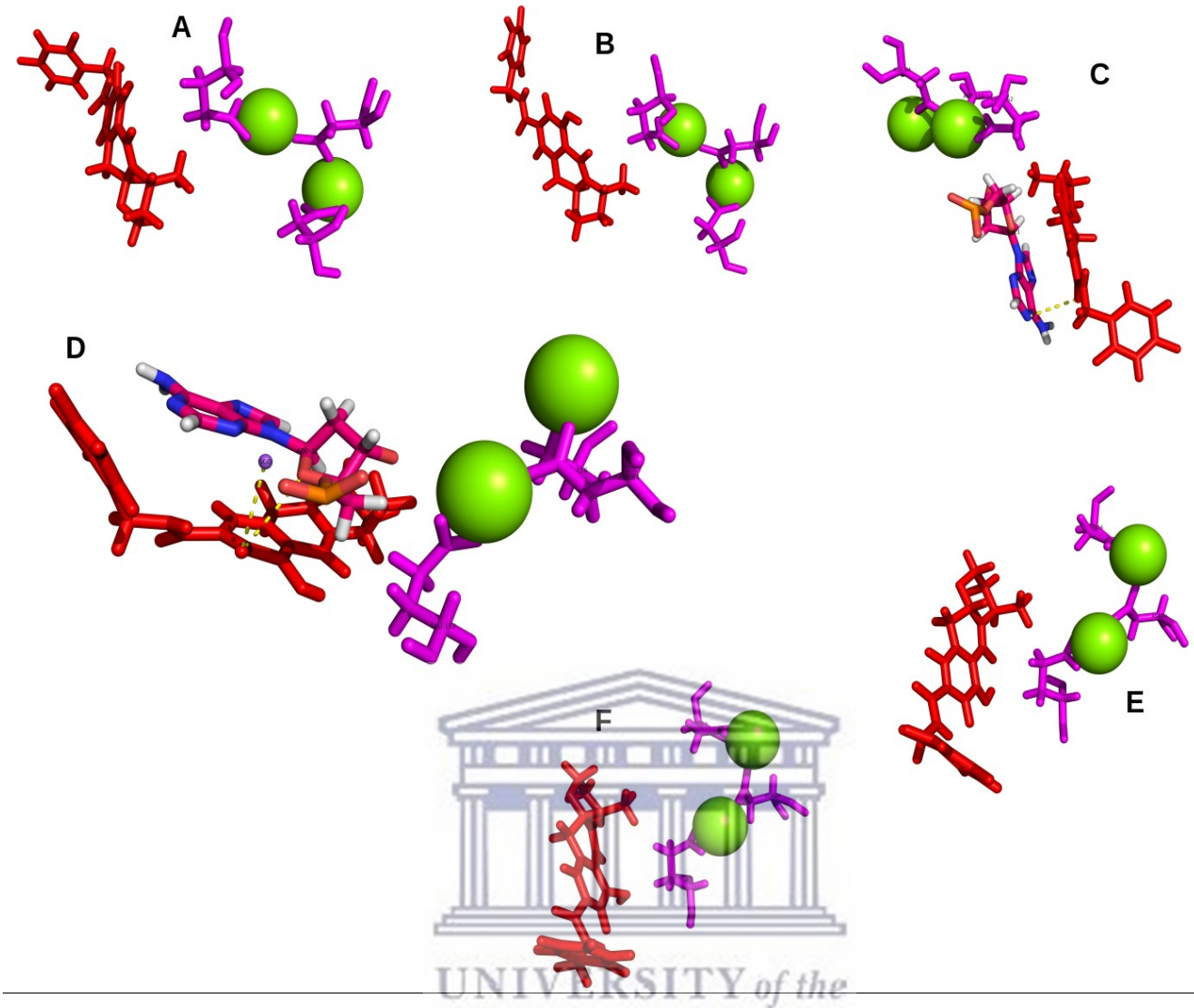
WT IN system's polar interactions between DTG and protein residues for the last 50 ns of the simulation. (A) 100 ns (B) 110 ns (C) 120 ns (D) 130 ns (E) 140 ns (F) 150 ns. DTG is coloured in red, MG ions coloured green, DDE motif residues (D64, D116, E152) are coloured magenta and DNA nucleotides or protein resides are coloured by element. The polar contacts between DTG and the neighbouring residues are represented by the yellow dashes.



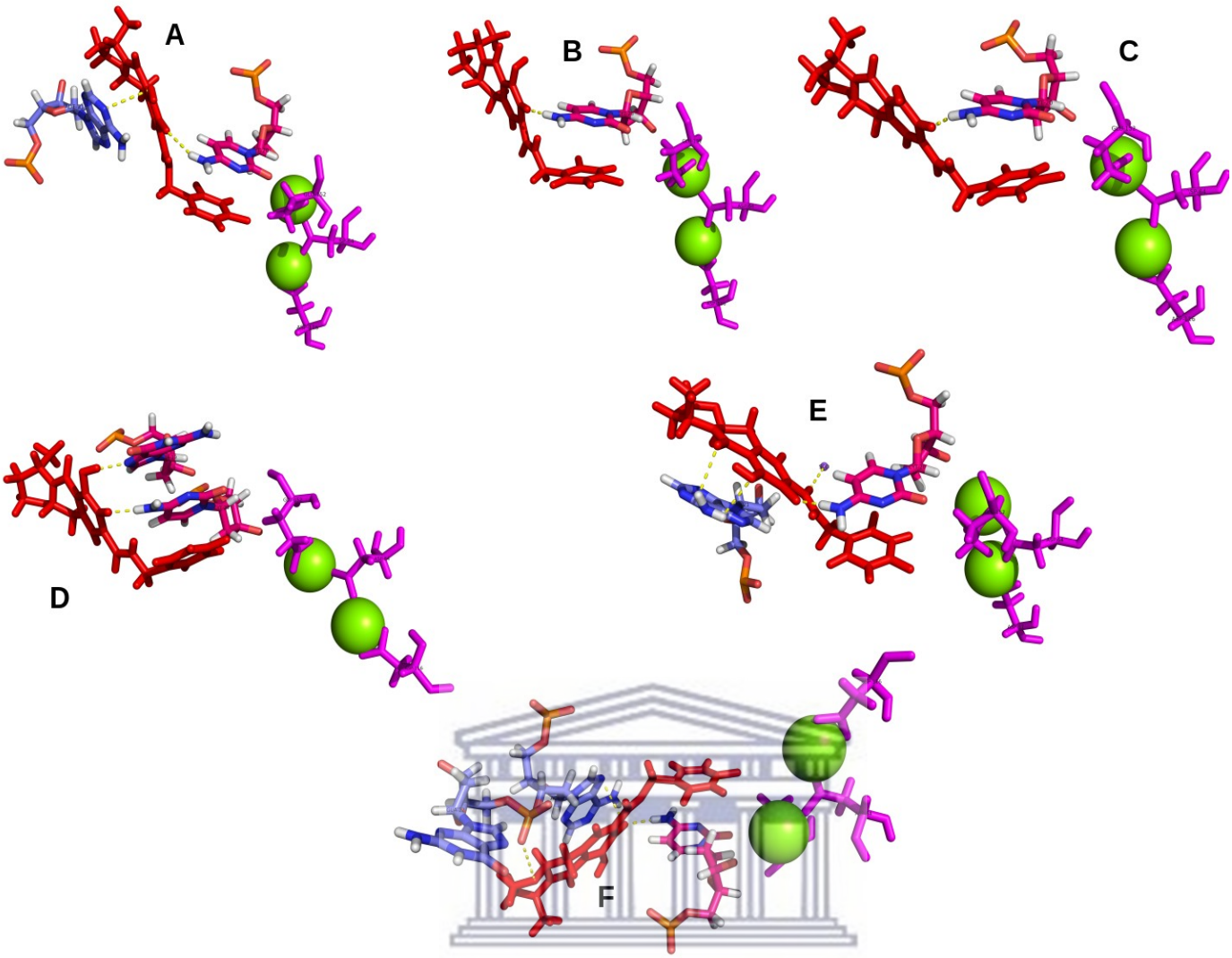
P1 variant IN system's polar interactions between DTG and protein residues for the last 50 ns of the simulation. (A) 100 ns (B) 110 ns (C) 120 ns (D) 130 ns (E) 140 ns (F) 150 ns. DTG is coloured in red, MG ions coloured green, DDE motif residues (D64, D116, E152) are coloured magenta and DNA nucleotides or protein resides are coloured by element. The polar contacts between DTG and the neighbouring residues are represented by the yellow dashes.



P2 variant IN system's polar interactions between DTG and protein residues for the last 50 ns of the simulation. (A) 100 ns (B) 110 ns (C) 120 ns (D) 130 ns (E) 140 ns (F) 150 ns. DTG is coloured in red, MG ions coloured green, DDE motif residues (D64, D116, E152) are coloured magenta and DNA nucleotides or protein resides are coloured by element. The polar contacts between DTG and the neighbouring residues are represented by the yellow dashes.



P3 variant IN system's polar interactions between DTG and protein residues for the last 50 ns of the simulation. (A) 100 ns (B) 110 ns (C) 120 ns (D) 130 ns (E) 140 ns (F) 150 ns. DTG is coloured in red, MG ions coloured green, DDE motif residues (D64, D116, E152) are coloured magenta and DNA nucleotides or protein resides are coloured by element. The polar contacts between DTG and the neighbouring residues are represented by the yellow dashes.



P4 variant IN system's polar interactions between DTG and protein residues for the last 50 ns of the simulation. (A) 100 ns (B) 110 ns (C) 120 ns (D) 130 ns (E) 140 ns (F) 150 ns. DTG is coloured in red, MG ions coloured green, DDE motif residues (D64, D116, E152) are coloured magenta and DNA nucleotides or protein resides are coloured by element. The polar contacts between DTG and the neighbouring residues are represented by the yellow dashes.

Appendix H: Published work from this study

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

OPEN ACCESS

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 identi-

PLOS ONE

Molecular dynamic simulations to investigate the structural impact of resistance mutations on HIV-1C Integrase

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 1ps 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

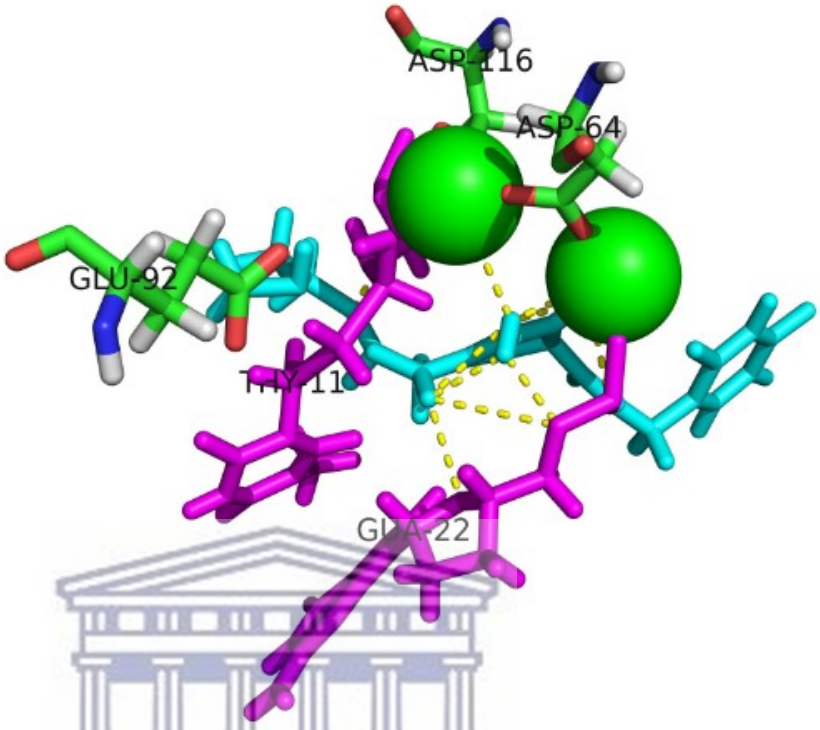


Fig 1. HIV-1C Integrase active site showing interactions with DNA, MG and drug Dolutegravir. Magnesium^{2+} ions (green spheres) are shown sitting in close proximity with Dolutegravir (cyan) within the binding pocket (DDE motif) residues are labelled and shown as sticks. Two DNA residues; THY11 and GUA22 (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).

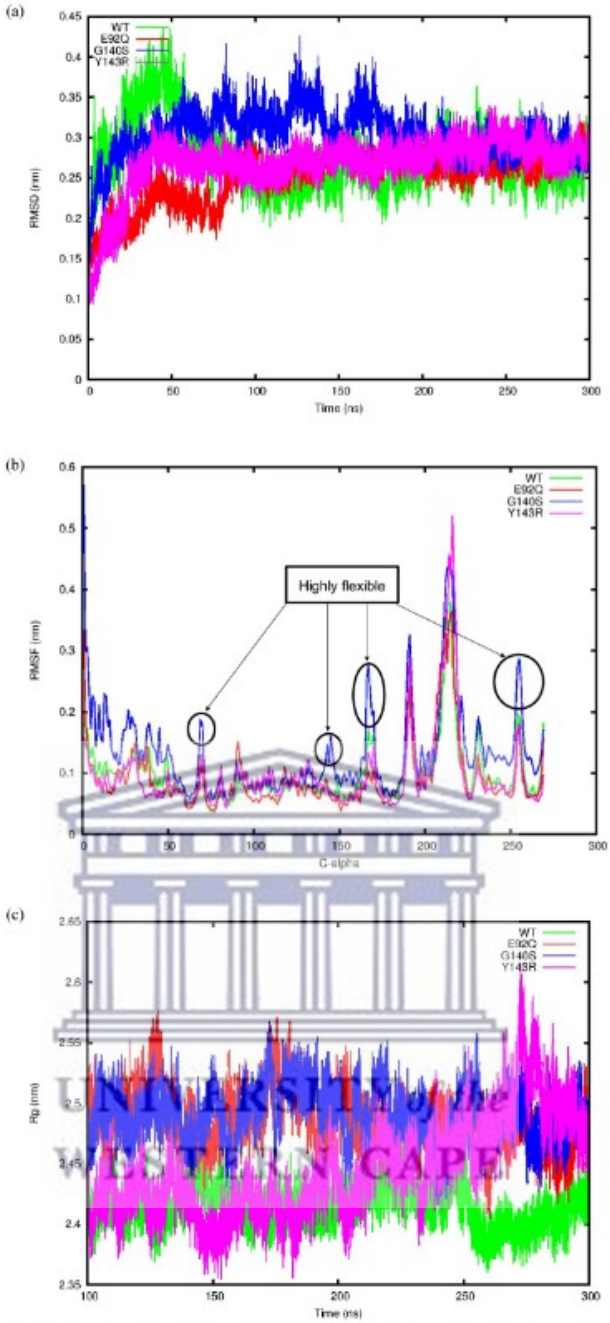


Fig 2. Trajectory analysis of the four simulation systems. (A) Change in backbone RMSD for the WT, E92Q, G140S and Y143R systems plotted over 300 ns. (B) Change in RMSF for the C-alpha residues for the WT, E92Q, G140S and Y143R systems plotted over the last 200ns. (C) Measure of compactness for the WT,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