Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

Article

Structural comparison of diverse HIV-1 subtypes using Molecular Modelling and Docking analyses of Integrase inhibitors

Darren Isaacs[1,*,+], Sello Given Mikasi[2,+], Adetayo Emmanuel Obasa[2], George Mondinde Ikomey[3], Ruben Cloete[1], Graeme Brandon Jacobs[2,*].

1.South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa. Robert Sobukwe Road, Bellville, 7530. 3433660@myuwc.ac.za, ruben@sanbi.ac.za

2. Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Francie van Zijl Avenue, P.O. Box 241, Cape Town 8000, South Africa;graeme@sun.ac.za, mikasi@sun.ac.za ,obasa@sun.ac.za

3. Centre for the Study and Control of Communicable Diseases (CSCCD), University of Yaoundé 1, Cameroon. PB 8445, Yaoundé Cameroon.mondinde@yahoo.com

* These authors contributed equally to this work

* Correspondence: 3433660@myuwc.ac.za (DI); graeme@sun.ac.za (GBJ); Tel.: +27 21 938 9744 (GBJ)

Abstract: The process of viral integration into the host genome is an essential step of the HIV-1 life cycle. The viral Integrase (IN) enzyme catalyses integration. IN is an ideal therapeutic enzyme targeted by several drugs; raltegravir (RAL), elvitegravir (EVG), dolutegravir (DTG) and bictegravir (BIC) having been approved by the USA Food and Drug Administration (FDA). Due to high HIV-1 diversity, it is not well understood how specific naturally occurring polymorphisms (NOPs) in IN may affect the structure/function and binding affinity of Integrase Strand Transfer Inhibitors (INSTIs). In this study, we applied computational methods of molecular modelling and docking to analyse the effect of NOPs on the full-length IN structure and INSTI binding. We identified 16 NOPs within the Cameroonian derived CRF02_AG IN sequences and further identified 17 NOPs within HIV-1C South African sequences. The NOPs in the IN structures did not show any effect on INSTI binding. INSTIs displayed similar binding affinities to each IN structure. All INSTIs are clinically effective against diverse HIV-1 strains from INSTI treatment naïve populations. This study supports the use of second-generation INSTI DTG as part of first-line combination antiretroviral therapy (cART) regimens, due to DTG possessing a stronger genetic barrier to the emergence of drug resistance.

Keywords: Integrase; Naturally occurring polymorphisms; HIV-1; Molecular Modelling; Molecular docking; diversity

1. Introduction

The HIV/AIDS pandemic continues to be a significant problem worldwide [1]. The viral integration process, which is the insertion of viral DNA into host genomic DNA is an indispensable step of the retroviral life cycle and is catalyzed by the viral Integrase (IN) enzyme [2]. Integration is achieved via two distinct sequential catalytic activities, 3′ processing and strand transfer. IN first processes viral DNA by excising a dinucleotide at the 3′ end, exposing hydroxyl ends. IN then catalyses the introduction of the prepared DNA into genomic DNA by facilitating a nucleophilic attack upon genomic DNA [3,4]. The same active site in IN, which contains a retroviral highly conserved DDE motif and magnesium ions, perform both activities [5,6]. HIV-1 IN is a 32kDa protein that functions as a tetramer or multimer [3,4]. A monomer consists of 3 distinct domains; the N-terminal domain (NTD) comprising residues 1-46, the catalytic core domain (CCD) comprising residues 56 -186 within which the active site DDE motif (Aspartate (D64), Aspartate (D116) and

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

Glutamate (E152) is present and the C-terminal domain (CTD) comprising residues 195 – 288 [7,8]. Several Integrase Strand Transfer Inhibitors (INSTIs) have been developed that targets HIV-1 IN to prevent viral integration into the host genome. The four INSTIs available thus far include; Raltegravir (RAL) and Elvitegravir (EVG) that are considered as first generation inhibitors, while Dolutegravir (DTG) and Bictegravir (BIC), along with the late-phase clinically trial Cabotegravir (CBT), are classified as second generation INSTIs [9].

At present, first-line antiretroviral regimens for HIV-1 are expected to include the INSTI DTG according to World Health Organisation (WHO) recommendations [10], because it has been shown to possess a higher genetic barrier to drug resistance development as compared with RAL and EVG [11]. HIV-1 is a genetically, highly diverse virus, forming different subtypes, recombinant and region specific variants [12]. Development of INSTIs, as with most pharmaceutical agents were primarily conducted by companies in first world nations, where subtype B is the most predominant variant [13,14]. It remains unclear what effects naturally occurring polymorphisms (NOPs) may have upon the IN structure and INSTI susceptibility. This lack of data poses a challenge in drawing conclusions regarding the effects NOPs on the binding of INSTIs to HIV-1 IN subtypes [15-19].

In this study computational methods, which include molecular modelling and docking, were used to determine if NOPs affect INSTI binding to HIV-1C IN subtype C and a circulating recombinant form of HIV-1 IN CRF_02_AG. The recently resolved Cryogenic-Electron Microscopy full-length HIV-1 subtype B IN structure, allowed us to build accurate and complete tetrameric three-dimensional structures of HIV-1C IN and of HIV-1 IN CRF_02_AG. The value of having accurate Protein models allows us to infer the exact mode of interactions formed between active site residues of HIV-1 subtypes and drug atoms.

HIV-1 Subtype C derived from a South African cohort was chosen as one of our IN models, as it represents the most prevalent subtype both globally and for sub-Saharan Africa in particular [14,19]. Our focus on a Cameroonian cohort was spurred on by the previously reported HIV-1 diversity present in Cameroon with all known subtypes/variants found within Cameroon [20,21]. Furthermore, the full-length Cryo-EM HIV-1 IN structure, which was used as the template in our molecular modelling, served additionally as a subtype B IN model in our study, the predominant strain in developed nations.

2. Materials and Methods

2.1. Ethics Statement

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

Ethical permission for this study was obtained from the Health Research Ethics Committee of Stellenbosch University (N14/10/130 and N15/08/071). The study was conducted according to the ethical guidelines and principles of the international Declaration of Helsinki 2013, South African Guidelines for Good Clinical Practice and the Medical Research Council (MRC), Ethical Guidelines for Research. A waiver of consent was awarded to conduct sequence analyses.

2.2. Study design

HIV-1-positive plasma samples were obtained from the Centre for the Study and Control of Communicable Diseases (CSCCD), University of Yaoundé I, Cameroon and In South Africa samples were requested, with permission, through the National Health Laboratory Services (NHLS) within 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 history and patients receiving first-line cART treatment. Patients had their samples sent for HIV-1 genotypic resistance testing to the NHLS. Treatment failure is defined according to the South Africa adult antiretroviral guidelines by a confirmed viral load of >1000 copies/mL on two measurements taken two to three months apart.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

2.3. Nucleic acid extraction

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

2.4. PCR amplification and sequencing

The synthesis of complementary DNA (cDNA) and first-round PCR amplification was performed using the Invitrogen SuperScript® III Reverse Transcriptase (RT) reagents (Invitrogen, Germany), as per the manufacturer's instructions. In-house amplification of the IN region (867 bp, positions 4230–5096, HXB2 strain) by nested RT-PCR was performed as previously described by our laboratory [22,23]. Purified amplicons were sequenced on both strands with conventional Sanger DNA sequencing, using the ABI Prism Big Dye® Terminator sequencing kit version 3.1 and run on the ABI 3130xl automated DNA sequencer (Applied Biosystems, USA), according to manufactures instructions. Primers spanning the full-length integrase (867 bp) were used to sequence the PCR products in both directions. These include sequencing primers, Poli6 and Poli7 and additional sequencing primers were used namely; Poli2 (TAAARACARYAGTACWAATGGCA), relative to position 4745–4766 and KLVO83 (GAATACTGCCATTTGTACTGCTG), corresponding to position 4750– 4772.

2.5. Consensus sequence alignment

We performed a search on the HIV Los Alamos National Library (LANL) database (https://www.hiv.lanl.gov/components/sequence/HIVsearch.com). Our search inclusion criteria included all Cameroonian HIV-1 subtype CRF02_AG IN sequences identified from treatment naïve patients. We selected one sequence per patient and all problematic sequences were excluded from further analyses. Finally, the consensus sequence representing CRF02_AG was generated using the CRF02_AG study sequences as previously reported [24], database-derived HIV-1 CRF02_AG sequences (n = 100), while the consensus sequence for subtype C was derived from cohort sequences (n = 91,) as reported in [23]. Nucleotide sequences were verified for stop codons, insertion and deletions using an online quality control program on the HIVLANL database (https://www.hiv.lanl.gov/content/sequence/QC/index.htm). Multiple sequence alignments were done with MAFFT version 7, from which the consensus sequence was derived [25]. As part of quality control, each of the viral sequences were inferred on a phylogenetic tree in order to eliminate possible contamination. The amino acid sequence alignment was extensively screened for the presence of primary and secondary resistance-associated mutations (RAMs) and NOPs associated with resistance to known INSTIs.

2.6. Protein modelling

A three dimensional model was constructed for HIV-1C IN and recombinant form CRF02_AG using Schrodinger Prime modelling software [26,27]. A suitable homologous template was identified by performing a Blastp search using the consensus amino acid sequences of HIV1C IN and recombinant form CRF02_AG. Prior to modelling, missing residues were fixed by re-modelling the structure of template 5U1C using Schrodinger PRIME modelling software [28]. The Cryo-EM solved IN subtype B intasome structure (ID: 5U1C) was used as the homologous template for comparative modelling, as it shared a high sequence identity and coverage with HIV-1C IN from a South African cohort and with the recombinant form CRF02_AG from a Cameroonian cohort.

2.7. Protein preparation

Processing of protein models was performed using Schrodinger Protein Preparation Wizard, which added hydrogen atoms, created disulphide bonds, assigned bond orders, filled in any missing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

side-chains and optimized the H-Bonds [29]. Magnesium ions were extracted from the Prototype Foamy Virus (PFV) intasome structure by superimposition of the active site of HIV-1 integrase models with the PFV active site (ID:3lV), which was chosen due to high active site conservation between PFV integrase and HIV-1 integrase.

2.8. Model validation

To assess the quality of the constructed IN Protein models, a variety of structural parameters were tested within each model. The Structural Analysis and Verification Server (SAVES) was used for this purpose and includes the tools; Procheck, Whatcheck, Prove, Verify3D, ERRAT [30-34]. The cut-off values used by the tools were as follows >80% for Verify3D, <1% for Prove, >50 for ERRAT. Whatcheck and Procheck are further subdivided into tests, but both make use of a Ramachrandran plot analysis, which is deemed passed if the majority of residues are within the allowable region.

Furthermore, root mean square deviation (RMSD) analysis was conducted using PYMOL Maestro molecular visualising software to compare backbone structural similarity to the experimentally solved 5U1C template structure [35].

2.9. Energy Minimisation

To reduce any steric clashing within the IN protein models, energy minimisation was performed using Gromacs version 2018.1. This included 5000 steps of energy minimisation using the steepest descent algorithm converging to a maximum force of <10.0 KJ/mol. The CHARMM 36 force field was used because it generates accurate topology parameters for proteins and has been previously applied to the study of IN [36 -38].

2.10. Molecular docking and Interaction analysis

Molecular docking assays was performed using SMINA a fork of AUTODOCK VINA [39,40]. The 3D Structures for the INSTI ligands were acquired from the ZINC database. Conversion of receptor and ligand structures from the respective pdb and sdf formats to pdbqt was done using OBABEL [41]. The docking grid was centred on the active site with a box size of 20Å in all planes. Each of the drugs were docked to each subtype structural intasome HIV-1C, B and AG, respectively. Ranking of generated binding poses was done based on the binding affinity values calculated using the Vinardo Scoring function [42]. Refinement of top ranked docking poses was done using the protein preparation wizard of Schrodinger Maestro software. The ligand interaction diagram 2D viewer within Schrodinger Maestro was used to calculate protein-ligand interactions.

2.11. Binding site analysis

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

3. Results

3.1. Sequence alignments and protein structure prediction

Two IN consensus protein sequences, corresponding to the South African and Cameroonian cohorts respectively, were aligned to the sequence of the subtype B IN structure (ID:5u1c) (Fig. 1) The sequence identity was calculated to 98% for both sequences when compared to subtype B. The alignment revealed 16 NOPs within the Cameroonian cohort derived CRF02_AG IN sequences. The identified polymorphisms being K14R, V31I, V72I, L101I, T112V, V113I, T124A, T125A, G134N, I135V, K136T, I151V, V201I, T206S, V234I, S283G. The alignment also showed 17 NOPs within a South African cohort derived Subtype C consensus sequence namely D25E, V31I, M50I, V72I, F100Y, L101I,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

T112V, V113I, T124A, T125A, K136Q, I151V, V201I, T218I, V234I, R269K, D278A, S283G. The constructed IN models were validated with the following scores obtained with the SAVES server tests. CRF02_AG IN; Verify3D 71%, ERRAT 93/100, Prove 7.8% and for the Ramachandran plot analysis 86.1% of residues are within most favoured region. Subtype C IN; Verify3D 74%, ERRAT 92/100, Prove 7.8% and for the Ramachandran plot analysis 86.1% of residues are within most favoured region. Subtype B IN; Verify3D 71%, ERRAT 92/100, Prove 10.3% and for the Ramachandran plot analysis 82.6% of residues are within most favoured region.

RMSD analysis score was ~0.4Å in difference between all 3 structures. In addition, there was minimal difference in secondary structural makeup (Figure2). However, reference subtype B structure forms an extra helical turn absent in Subtype C and CRF02_AG.

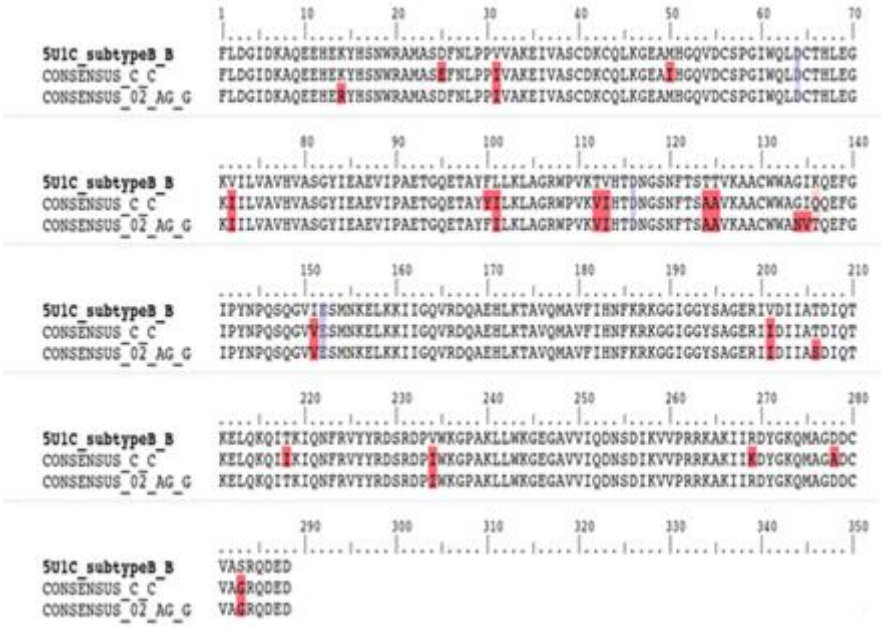
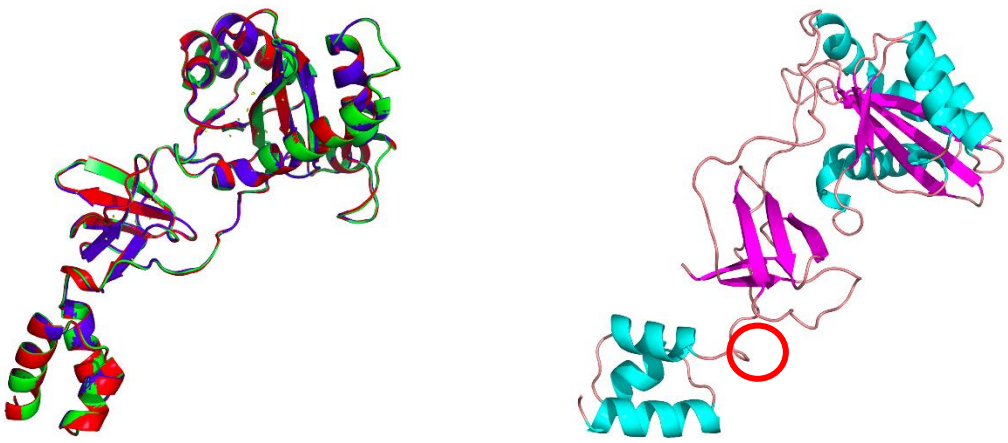
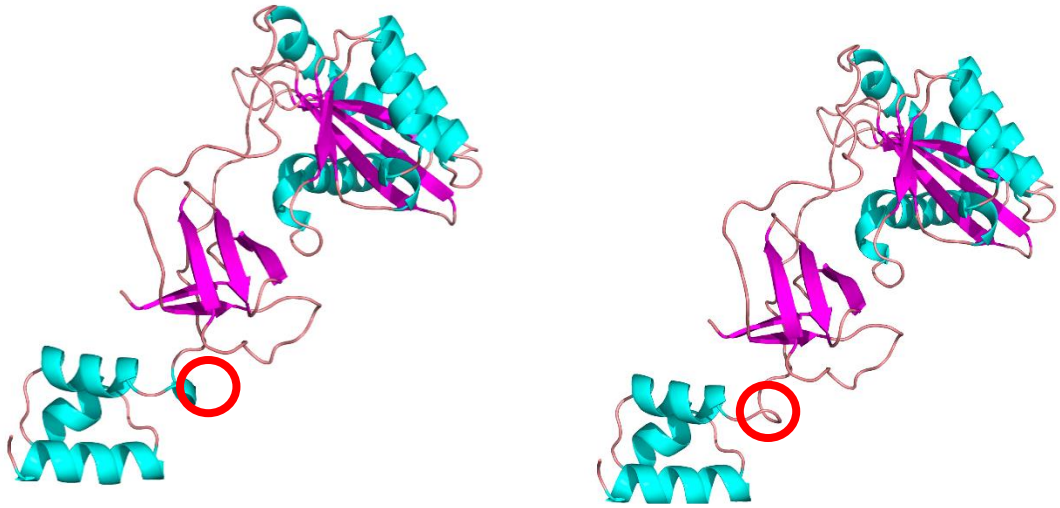


Figure 1. Amino acid sequence alignment of integrase variants Subtype B, subtype Cza and CRF_02_AG respectively. Red highlighted residues indicate polymorphism locations in comparison to the Subtype B template. Blue highlighted residues indicate the DDE motif.



a b

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1



c d

Figure 2. (a): A superimposition of all 3 Integrase models showing high backbone identity and secondary structure conservation. (b): CRF02_AG IN model coloured according to secondary structure and in cartoon depiction. (c): Subtype B IN model coloured according to secondary structure and in cartoon depiction. (d): Subtype C IN model coloured according to secondary structure and in cartoon depiction. Encircled in red is the secondary structure difference observed between the template structure and the generated IN models. Light blue indicates the helices, purple indicates beta-sheets and tint colour indicates loops.

3.2. Molecular docking and Interaction analysis

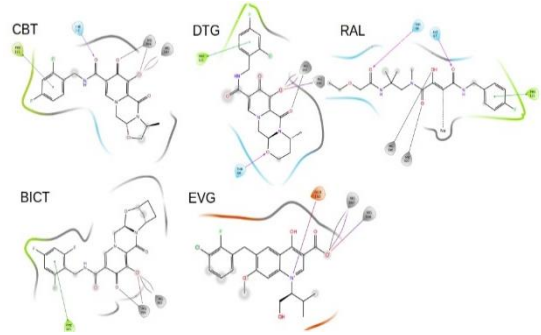
The conducted molecular docking assays of the five FDA approved and/ late phase clinical trial INSTI's showed minimal differences, with less than 3kcal/mol difference observed in predicted binding affinities between each IN structure. (Table 1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

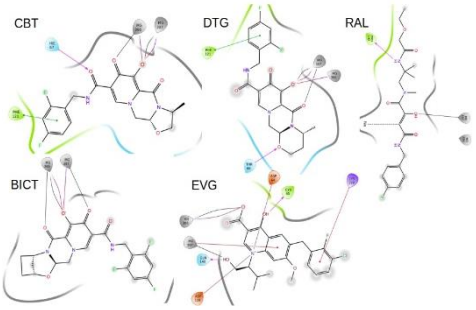
Table 1: Binding affinity scores predicted for each INSTIs bound to the three integrase subtypes CRF02_AG, B and C.

Drug	CRF02_AG	Subtype B	Subtype C
raltegravir	-7.1 kcal/mol	-7.6 kcal/mol	-6.0 kcal/mol
elvitegravir	-5.9 kcal/mol	-6.3 kcal/mol	-6.1 kcal/mol
dolutegravir	-6.5 kcal/mol	-7.4 kcal/mol	-6.5 kcal/mol
bictegravir	-6.5 kcal/mol	-7.1 kcal/mol	-6.0 kcal/mol
cabotagravir	-6.9 kcal/mol	-6.8 kcal/mol	-7.0kcal/mol

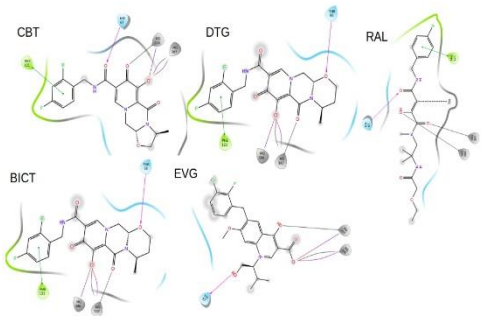
Interaction analysis was done to understand the mode of interaction between the docked ligands and IN protein model active sites. Figures 3a, 3b, and 3c show the predicted binding interactions formed between the IN subtypes and top ranked INSTIs binding pose. The interactions observed between the five INSTIs and three IN subtypes is summarised in Table 2.



(**a**)



(**b**)



(**C**)

Figure 3. (a): 2D diagram of chemical interactions occurring between subtype B and INSTIs. Diagrams are illustrative of top docked Poses generated. In all cases, the important interaction(s) with MG ions takes place; (b) 2D diagram of chemical interactions occurring between subtype C and INSTIs. Diagrams are illustrative of top docked poses generated. In all cases, the important interaction(s) with MG ions occurs; (c) 2D diagram of chemical interactions occurring between recombinant CRF_02_AG and INSTIs. Diagrams are illustrative of top docked poses generated. In all cases the important interaction(s) with MG ions occurs. More interactions are observed with EVG than was observed with subtype B but a lower binding was still calculated by.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

Table 2. Summary of all interactions observed between the five INSTIs and three IN subtypes.

INSTI	HIV-1B (ID:5U1C)	HIV-1C IN	CRF_02_AG IN
RAL	5 (2 MG, 2 Hydrogen bond (T66,H67), 1 pi stacking interaction(F121)	2 (1 MG, 1 Hydrogen Bond(F121))	4(2 MG, 1 Pi stacking(F121), 1 Hydrogen Bond(H67)
DTG	6(4 MG, 1 pi stacking interaction(F121), 1 Hydrogen bond(T66))	6(4 MG, 1Hydrogen Bond(T66),1 Pi stacking interaction(F121))	6(4 MG, 1 Hydrogen Bond(T66), 1 Pi stacking interaction(F121))
EVG	4(3 MG, 1 Hydrogen bond(E152))	9(4 MG,3 Hydrogen bonds(D64, C65, Q148), salt bridge(D116), pi-cation(K159)	5(4 MG, 1 Hydrogen Bond(H67))
BIC	5(4 MG, 1 Hydrogen bond(F121))	4(4MG)	5(4 MG, 1 Hydrogen Bond(T66), 1 pi stacking(F121))
CBT	6(4 MG, 1 Hydrogen bond(H67), 1 Pi-stacking reaction(F121))	6(4MG, Hydrogen Bond(H67), Pi stacking(F121))	6(4 MG, 1 Hydrogen Bond(H67), Pi stacking(F121))

* Number outside bracket indicates total number of reaction.

3.3. Binding site analysis

We calculated differences in total surface area, with 860Å^2 for subtype B IN, 969Å^2 for subtype C IN and 1041Å^2 for CRF_02_AG IN. RMSD analysis show that subtype C IN and CRF_02_AG deviate by 0.309Å and 0.44Å, respectively. The NOP I151V were found to occur within the binding site, but does not directly interact with INSTI's. NOP I151V is present in both subtype C and CRF_02_AG IN's. Both residue types have side-chains, which are highly unreactive and therefore play no functional role, however side-chain constituents and orientations may affect the spatial volume of the binding cavity.

4. Discussion

The NOPs that have been identified in subtype C and CRF_02_AG IN have not been previously associated with HIV-1 IN drug resistance, with the exception of the polymorphism M50I. M50I was identified in our subtype C IN sequence and this polymorphism has been reported to reduce DTG susceptibility when found in combination with the mutation R263K in HIV-1 subtype B IN [43,44] However, in our study the mutation R263K was not present. Furthermore, M50I is not able to cause drug resistance on its own, but increases the effect of resistance exhibited by R263K [43] In our study M50I had no effect on INSTI's binding to IN. A comparison of the backbone structure of the modelled IN structures to one another showed high similarity with less than 0.4Å backbone deviation found between the 3 variants calculated from RMSD analysis. Our IN homology modelling showed one slight secondary structural feature alteration within the N-terminal domain. Subtype B displayed a helical turn absent in Subtype C and CRF02_AG IN. Secondary structural features, such as helices may influence the efficacy of drugs upon proteins, either by directly changing binding pocket properties or through affecting stability of the whole protein [45]. No alteration to the binding site properties implicating the helix was noted in our study. This might be attributed to the distance away from the active site, which is within the Catalytic core domain.

Molecular docking showed that all INSTIs are able to bind to either of the tested IN subtype structures with plausible binding poses. No significantly reduced binding affinity was observed for each of the INSTIs, implying no alteration to the binding site, which may prevent INSTI drug binding. However, differences in binding affinity are present. The Binding affinity is an indication of how strong the ligand is binding to the active site. The tested INSTIs bound consistently stronger to subtype B IN than either CRF_02_AG IN or subtype C IN, albeit that the differences in binding affinity were minimal. This may be attributable to INSTI development largely having taken place in subtype B predominant populations. The implication being that we predict INSTIs to likely stay bound for a shorter amount of time to CRF_02_AG or Subtype C as compared with subtype B IN. Experimental binding affinity assays need to be conducted to support this finding. EVG susceptibility reducing NOPs associated with reduction to susceptibility to EVG have been previously reported to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

occur with a relatively high frequency within CRF02_AG IN, however those NOPs were not identified in our study [46 – 48].

Magnesium ions are responsible for the binding of DNA, INSTIs competitively inhibit this process by binding to the magnesium ions. It was therefore expected that interaction analysis would reveal that an interaction(s) takes place between the docked INSTIs and magnesium ions present within integrase active sites. The interaction(s) with magnesium ions are considered to be essential for inhibition to take place whereas other interactions are considered non-essential for the activity of INSTIs to take place. Therefore, in this study we considered an INSTI to be successfully bound if interaction analysis predicted interactions occurring with magnesium ions, it may however be likely that NOPs favour or reduce the likelihood of additional interactions occurring enhancing binding affinity. It should be emphasised that our molecular docking assays were not performed on IN-DNA complexes, it has been widely reported that INSTI binding is strengthened in the presence of viral DNA. Omission of viral DNA from our reported integrase structures allowed for the direct assessment of the impact the NOPs have on INSTI drug binding [49].

Binding site analysis further support the results from the molecular docking. The analyses conducted show that binding sites between each IN structures are identical. One NOP I151V is present within the binding sites of our study but does not induce any effect beyond a slight spatial change due to its different side chain orientation. Future work will include molecular dynamic simulation studies of the different protein-MG-drug complexes to estimate free energy of ligand binding to different IN conformations to include polar solvation and entropy effects.

5. Conclusion

Our study show that unique polymorphisms within geographically distinct HIV-1 infected populations with different variants do not prevent INSTI binding. However, polymorphisms may affect the strength of the binding of INSTI's and possibly favour drug resistance mutations. We therefore put forward that second generation DTG should be added to anti-viral regimens as part of first-line regimens, this would account for cross-resistance which may occur between EVG and RAL as DTG has a higher genetic barrier to resistance.

Author Contributions: Conceptualization, R.C.; methodology, D.I, S.G.M, AEAO.; formal analysis, D.I., S.G.M., and R.C.; investigation, D.I and S.G.A. .; data curation, D.I.; writing—original draft preparation, D.I and S.G.M.; writing—review and editing, D.I, S.G,M, G.B.J, R.C, G.M.I, A.E.A.O and R.C.; supervision, R.C and G.B.J.; project administration, R.C.; funding acquisition, R.C.

Acknowledgments: The National Research Foundation (NRF) of South Africa and the Poliomyelitis Research Foundation (PRF) of South Africa funded this study. Grant support for this study was received from the National Research Foundation (NRF) of South Africa, the Poliomyelitis Research Foundation (PRF) of South Africa, Harry Crossley Foundation and National Health Laboratory Services (NHLS) research trust

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

References

1. Piot, P.; Quinn, T. C. Response to the AIDS Pandemic — A Global Health Model. *N Engl J Med* 2013, *368* (23), 2210–2218.

2. Lesbats, P.; Engelman, A. N.; Cherepanov, P. Retroviral DNA Integration. *Chem. Rev.* 2016, *116* (20), 12730–12757.

3. Craigie, R. The Molecular Biology of HIV Integrase. *Future Virology* 2012, *7* (7), 679–686.

4. Craigie, R.; Bushman, F. D. HIV DNA Integration. *Cold Spring Harbor Perspectives in Medicine* 2012, *2* (7), a006890–a006890.

5. Grobler, J. A.; Stillmock, K.; Hu, B.; Witmer, M.; Felock, P.; Espeseth, A. S.; Wolfe, A.; Egbertson, M.; Bourgeois, M.; Melamed, J.; et al. Diketo Acid Inhibitor Mechanism and HIV-1 Integrase: Implications for Metal Binding in the Active Site of Phosphotransferase Enzymes. *Proceedings of the National Academy of Sciences* 2002, *99* (10), 6661–6666.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

6. Neamati, N.; Lin, Z.; Karki, R. G.; Orr, A.; Cowansage, K.; Strumberg, D.; Pais, G. C. G.; Voigt, J. H.; Nicklaus, M. C.; Winslow, H. E.; et al. Metal-Dependent Inhibition of HIV-1 Integrase. *J. Med. Chem.* 2002, *45* (26), 5661–5670.
7. Ceccherini-Silberstein, F.; Malet, I.; D'Arrigo, R.; Antinori, A.; Marcelin, A.-G.; Perno, C.-F. Characterization and Structural Analysis of HIV-1 Integrase Conservation. *AIDS Reviews.* 13.
8. Malet, I.; Soulie, C.; Tchertanov, L.; Derache, A.; Amellal, B.; Traore, O.; Simon, A.; Katlama, C.; Mouscadet, J.-F.; Calvez, V.; et al. Structural Effects of Amino Acid Variations between B and CRF02-AG HIV-1 Integrases. *J. Med. Virol.* 2008, *80* (5), 754–761.
9. Anstett, K.; Brenner, B.; Mesplede, T.; Wainberg, M. A. HIV Drug Resistance against Strand Transfer Integrase Inhibitors. *Retrovirology* 2017, *14* (1), 36.
10. World Health Organisation.Update of recommendations of first-and second-line antiretroviral regimens. *World Health Organisation* 2019
11. Rhee, S.-Y.; Grant, P. M.; Tzou, P. L.; Barrow, G.; Harrigan, P. R.; Ioannidis, J. P. A.; Shafer, R. W. A Systematic Review of the Genetic Mechanisms of Dolutegravir Resistance. *Journal of Antimicrobial Chemotherapy* 2019, *74* (11), 3135–3149.
12. Santoro, M. M.; Perno, C. F. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiology* 2013, *2013*, 1–20.
13. Keyhani, S.; Wang, S.; Hebert, P.; Carpenter, D.; Anderson, G. US Pharmaceutical Innovation in an International Context. *Am J Public Health* 2010, *100* (6), 1075–1080.
14. Bbosa, N.; Kaleebu, P.; Ssemwanga, D. HIV Subtype Diversity Worldwide: *Current Opinion in HIV and AIDS* 2019, *14* (3), 153–160.
15. Bar-Magen, T.; Sloan, R. D.; Faltenbacher, V. H.; Donahue, D. A.; Kuhl, B. D.; Oliveira, M.; Xu, H.; Wainberg, M. A. Comparative Biochemical Analysis of HIV-1 Subtype B and C Integrase Enzymes. *Retrovirology* 2009, *6* (1), 103.
16. Lessells, R.; Katzenstein, D.; de Oliveira, T. Are Subtype Differences Important in HIV Drug Resistance? *Current Opinion in Virology* 2012, *2* (5), 636–643.
17.Depatureaux, A.; Quashie, P. K.; Mesplède, T.; Han, Y.; Koubi, H.; Plantier, J.-C.; Oliveira, M.; Moisi, D.; Brenner, B.; Wainberg, M. A. HIV-1 Group O Integrase Displays Lower Enzymatic Efficiency and Higher Susceptibility to Raltegravir than HIV-1 Group M Subtype B Integrase. *Antimicrob. Agents Chemother.* 2014, *58* (12), 7141–7150.
18. Han, Y.-S.; Mesplède, T.; Wainberg, M. A. Differences among HIV-1 Subtypes in Drug Resistance against Integrase Inhibitors. *Infection, Genetics and Evolution* 2016, *46*, 286–291.
19. Llácer Delicado, T.; Torrecilla, E.; Holguín, Á. Deep Analysis of HIV-1 Natural Variability across HIV-1 Variants at Residues Associated with Integrase Inhibitor (INI) Resistance in INI-Naive Individuals. *J. Antimicrob. Chemother.* 2016, *71* (2), 362–366.
20. Courtney, C. R.; Agyingi, L.; Fokou, A.; Christie, S.; Asaah, B.; Meli, J.; Ngai, J.; Hewlett, I.; Nyambi, P. N. Monitoring HIV-1 Group M Subtypes in Yaoundé, Cameroon Reveals Broad Genetic Diversity and a Novel CRF02_AG/F2 Infection. *AIDS Research and Human Retroviruses* 2016, *32* (4), 381–385.
21. Abongwa, L. E.; Nyamache, A. K.; Torimiro, J. N.; Okemo, P.; Charles, F. Human Immunodeficiency Virus Type 1 ((HIV-1) Subtypes in the Northwest Region, Cameroon. *Virol J* 2019, *16* (1), 103.
22. Jacobs, G. B.; Laten, A.; van Rensburg, E. J.; Bodem, J.; Weissbrich, B.; Rethwilm, A.; Preiser, W.; Engelbrecht, S. Phylogenetic Diversity and Low Level Antiretroviral Resistance Mutations in HIV Type 1 Treatment-Naive Patients from Cape Town, South Africa. *AIDS Research and Human Retroviruses* 2008, *24* (7), 1009–1012.
23. Brado, D.; Obasa, A. E.; Ikomey, G. M.; Cloete, R.; Singh, K.; Engelbrecht, S.; Neogi, U.; Jacobs, G. B. Analyses of HIV-1 Integrase Sequences Prior to South African National HIV-Treatment Program and Availability of Integrase Inhibitors in Cape Town, South Africa. *Sci Rep* 2018, *8* (1), 4709.
24 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. and Ikomey, G.M. . 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*. 2020
25. Katoh, K.; Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 2013, *30* (4), 772–780.
26. Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *Journal of Molecular Biology* 2002, *320* (3), 597–608.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0062.v1

27.Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins* 2004, *55* (2), 351–367.
28. Passos, D. O.; Li, M.; Yang, R.; Rebensburg, S. V.; Ghirlando, R.; Jeon, Y.; Shkriabai, N.; Kvaratskhelia, M.; Craigie, R.; Lyumkis, D. Cryo-EM Structures and Atomic Model of the HIV-1 Strand Transfer Complex Intasome. *Science* 2017, *355* (6320), 89–92.
29. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J Comput Aided Mol Des* 2013, *27* (3), 221–234.
30. Colovos, C.; Yeates, T. O. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci.* 1993, *2* (9), 1511–1519.
31. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J Appl Crystallogr* 1993, *26* (2), 283–291.
32. Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. Errors in Protein Structures. *Nature* 1996, *381* (6580), 272–272.
33.Eisenberg, D.; Lüthy, R.; Bowie, J. U. [20] VERIFY3D: Assessment of Protein Models with Three-Dimensional Profiles. In *Methods in Enzymology*; Elsevier, 1997; Vol. 277, pp 396–404.
34. Pontius, J.; Richelle, J.; Wodak, S. J. Deviations from Standard Atomic Volumes as a Quality Measure for Protein Crystal Structures. *Journal of Molecular Biology* 1996, *264* (1), 121–136.
35. Schrodinger LLC. The PyMOL molecular graphics system, version 1.8, 2015
36. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* 2005, *26* (16), 1701–1718.
37. Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* 2013, *34* (25), 2135–2145.
38. Musyoka, T.; Tastan Bishop, Ö.; Lobb, K.; Moses, V. The Determination of CHARMM Force Field Parameters for the Mg2+ Containing HIV-1 Integrase. *Chemical Physics Letters* 2018, *711*, 1–7.
39. Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* 2009, NA-NA.
40. Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* 2013, *53* (8), 1893–1904.
41. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J Cheminform* 2011, *3* (1), 33.
42. Quiroga, R.; Villarreal, M. A. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLoS ONE* 2016, *11* (5), e0155183.
43. Wares, M.; Mesplède, T.; Quashie, P. K.; Osman, N.; Han, Y.; Wainberg, M. A. The M50I Polymorphic Substitution in Association with the R263K Mutation in HIV-1 Subtype B Integrase Increases Drug Resistance but Does Not Restore Viral Replicative Fitness. *Retrovirology* 2014, *11* (1), 7.
44. Rogers, L.; Obasa, A. E.; Jacobs, G. B.; Sarafianos, S. G.; Sönnerborg, A.; Neogi, U.; Singh, K. Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. *Front. Microbiol.* 2018, *9*, 1754.
45. Lahti, J. L.; Tang, G. W.; Capriotti, E.; Liu, T.; Altman, R. B. Bioinformatics and Variability in Drug Response: A Protein Structural Perspective. *Journal of The Royal Society Interface* 2012, *9* (72), 1409–1437.
46. Kobayashi, M.; Nakahara, K.; Seki, T.; Miki, S.; Kawauchi, S.; Suyama, A.; Wakasamorimoto, C.; Kodama, M.; Endoh, T.; Oosugi, E. Selection of Diverse and Clinically Relevant Integrase Inhibitor-Resistant Human Immunodeficiency Virus Type 1 Mutants. *Antiviral Research* 2008, *80* (2), 213–222.
47. Hatano, H.; Lampiris, H.; Fransen, S.; Gupta, S.; Huang, W.; Hoh, R.; Martin, J. N.; Lalezari, J.; Bangsberg, D.; Petropoulos, C.; et al. Evolution of Integrase Resistance During Failure of Integrase Inhibitor-Based Antiretroviral Therapy: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2010, *54* (4), 389–393.
48. Abram, M. E.; Ram, R. R.; Margot, N. A.; Barnes, T. L.; White, K. L.; Callebaut, C.; Miller, M. D. Lack of Impact of Pre-Existing T97A HIV-1 Integrase Mutation on Integrase Strand Transfer Inhibitor Resistance and Treatment Outcome. *PLoS ONE* 2017, *12* (2), e0172206.
49. Hare, S.; Vos, A. M.; Clayton, R. F.; Thuring, J. W.; Cummings, M. D.; Cherepanov, P. Molecular Mechanisms of Retroviral Integrase Inhibition and the Evolution of Viral Resistance. *Proceedings of the National Academy of Sciences* 2010, *107* (46), 20057–200.