MOLECULAR DYNAMICS STUDY OF HIV-1 PROTEASE INHIBITORS AND THEIR EFFECTS ON THE FLAP DYNAMICS OF THE HIV-1 SUBTYPE-C (C-SA)

SIYABONGA INNOCENT MAPHUMULO

208511728

A thesis submitted to the School of Pharmacy and Pharmacology, Faculty of Health Science, University of KwaZulu-Natal, Westville, for the degree of Master of Medical Science.

2018

MOLECULAR DYNAMICS STUDY OF HIV-1 PROTEASE INHIBITORS AND THEIR EFFECTS ON THE FLAP DYNAMICS OF THE HIV-1 SUBTYPE-C (C-SA)

SIYABONGA INNOCENT MAPHUMULO

208511728

2018

A thesis submitted to the School of Pharmacy and Pharmacology, Faculty of Health Science, University of KwaZulu-Natal, Westville, for the degree of Master of Medical Science.

This is the thesis in which the chapters are written as a set of discrete research publications, with an overall introduction and final summary. Typically, these chapters will have been published in internationally recognized, peer-reviewed journals.

This is to certify that the content of this thesis is the original research work of Mr. Siyabonga Innocent Maphumulo.

As the candidate's supervisor, I have approved this thesis for submission.

Supervisor: Signed: ---------- Name: **Dr. B. Honarparvar** Date:

Co-Supervisor: Signed: ---------- Name: **Prof. H. G. Kruger** Date:

Co-Supervisor: Signed: ---------- Name: **Dr. G. E. M. Maguire** Date:

ABSTRACT

The aspartyl protease human immunodeficiency virus type 1 (HIV-1) is a 99-amino acid-long homodimer responsible for processing the Gag and Gag-Pol polyproteins into functional constituent proteins necessary for development of infectious HIV particles. Of global infections recorded, subSaharan African region is represented by 56 % where nearly 25 million people are living with HIV. South Africa has been shown to carry the heaviest HIV burden in sub-Saharan Africa where the HIV-1 subtype C (C-SA) is the prominent strain. Most of the HIV-1 scientific research has been done specifically for subtype B and this has been highlighted by the weaker binding affinity displayed by the South African HIV-1 subtype C for most of the clinically approved protease inhibitors when compared to the HIV-1 subtype B protease.

The two flaps of the HIV-1 PR are very essential in functioning of the enzyme as their conformations control entry of the substrate into the catalytic site of the enzyme and also to release product. It is very important to explore and understand the dynamics of these flaps in binding of different inhibitors with different binding affinities. In addition, studies have highlighted the focus on inhibiting the cleaving function of HIV-1 PR with protease inhibitors (PIs) by competing with the natural substrate for the enzyme's active or catalytic site and thus rendering its ineffective. It has been shown that in addition to the active site, more regions of the enzyme can be possible targets for inhibition process by developing drugs that can hinder the opening of the flaps or disrupt the stability of the dimer interface.

This study involved the use of computational techniques to explore the major contributing factors other than interactions with the binding site, in binding affinity of FDA approved second generation PIs complexed to HIV-1 C-SA PR. In pursuance of this objective, molecular dynamics simulations, binding-free energy calculations and dynamic analyses were utilized. Several distances, different angles between certain residues were all taken into consideration

Our findings do show that apart from binding free-energy calculations, not one single factor but several factors contribute to the binding affinity of protease inhibitors. It is clear from these results that in the development of new HIV-1 drugs, more emphasis should be made in the

design of drugs with, not only better binding in the active site but also with better interaction with other regions of the enzyme. Another interesting emphasis drawn from this study is that there is still need for drug development targeting HIV-1 PR C-SA as the currently available drugs were modelled around the inhibition of HIV-1 subtype B.

DECLARATION: PLAGIARISM

I, Siyabonga Innocent Maphumulo, declare that

1. The research reported in this thesis, except otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other person's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have quoted, then:
    a. Their words have been re-written but the general information attributed to them has been referenced.
    b. Where their exact words have been used, then their writing have been placed in italics and inside quotation marks, and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.
6. A detailed contribution to publications that form part and/or include research presented in this thesis is stated (include publications submitted, accepted, in press and published)

Signed:_____

TABLE OF CONTENTS

LIST OF PUBLICATIONS vii

DEDICATION viii

ACKNOWLEDGEMENTS ix

LIST OF ABBREVIATIONS x

LIST OF FIGURES xii

LIST OF TABLES xv

CHAPTER ONE 1

1. INTRODUCTION 1

1.1 Background 1

1.2 HIV (HIV-1 & HIV-2) 2

1.3 HIV-1 virus infection –life cycle 3

1.4 Different enzymes (Integrase, transcriptase, protease) 7

1.5 HIV- Inhibitors (RTIs, PIs, IIs, FIs)/ HAART/ side effects 8

1.6 Flap dynamics of HIV-PR 14

1.6.1 The structural information of HIV-1 protease 14

1.6.2 The stability of HIV protease 16

1.6.3 HIV-1 substrate binding and protease catalytic mechanism 16

1.7 Background on relevant studies of HIV-1 PR 18

1.7.1 Experimental studies 18

1.7.2 Computational studies 20

1.7.3 Proposed parameters to measure flap dynamics 21

1.8 Computational background 21

1.9 Objectives of the present study 23

1.10 Outline of this thesis 24

1.11 REFERENCES 25

Exploring the flap dynamics of the South African subtype C HIV protease in presence of FDA-approved inhibitors: MD study 39

Abstract 39

2.1 Introduction 40

2.2 Computational Methods 44

2.2.1 Preparation of inhibitors/enzyme complex 44

2.2.2 Setting up the system for Molecular Dynamics (MD) simulations 45

2.2.3 Post-dynamics analysis 46
2.2.4 Binding free energy calculations 46
2.2.5 Per residue free energy decomposition analysis 47
2.2.6 Principal component analysis (PCA) 47
2.3 Results and discussion 48
2.3.1 MM-GBSA binding energy analyses 48
2.3.2 Stability and flexibility of the HIV-1 PR complexes 50
2.3.3 Cross-correlation analysis 56
2.3.4 Hydrogen bonding interaction 60
2.3.5 Per-residue decomposition analysis 64
2.3.6 Principal Component Analysis (PCA) 67
2.4 Conclusion 69
References 69
Disrupting the *β*-sheet interface of HIV PR to determine the effect of C- and N- terminal residue truncation on C-SA PR's stability: A MD Study 77
Abstract 77
3.1 Introduction 79
3.2 Computational Methods 80
3.2.1 Preparation of inhibitors/enzyme complex 80
3.2.2 Setting up the system for Molecular Dynamics (MD) simulations 81
3.2.3 Post-dynamics analysis 81
3.2.4 Binding free energy calculations 82
3.2.5 Per residue free energy decomposition analysis 82
3.3 Results and discussion 83
3.3.1 Stability and flexibility of the HIV-1 PR complexes 83
3.3.2 Root-Mean Square Fluctuation (RMSF) 84
3.3.3 MM-GBSA binding energy analyses 86
3.3.4 Per-residue free energy decomposition analysis 86
3.3.5 Radius of Gyration (Rg) 90
3.3.6 Distance Gln92-Gln92' 90
3.3.7 Distance Ile50-Ile50' 92
3.3.7 Hydrogen Bond (Arg8-Asp29'/ Asp29-Arg8') 92
3.4 Conclusion 96
References 97

CHAPTER FOUR 100
Summary and conclusion 100
Recommendations 102

LIST OF PUBLICATIONS

1. Exploring the flap dynamics of the South African subtype C HIV protease in presence of FDA-approved inhibitors: MD study (**submitted to Chemical Biology & Drug Design**)

Contributions:

Siyabonga Maphumulo: Principle investigator in the design of this project and first author responsible for writing and preparation of this manuscript.

Amit K. Halder: Provided assistance with outline of the manuscript.

Sibusiso Maseko: Performed experimental work

Thavendran Govender: Provided assistance with technical and experimental input in the overall project.

Glenn E. M. Maguire, Bahareh Honarparvar, Hendrik G. Kruger: Co-supervisors

2. MD Study for dimerization inhibition of C-SA PR resulting from Whisker truncation and F99A mutation. (**submitted to Protein: Structure, Function, and Bioinformatics**)

Contributions:

Siyabonga Maphumulo: Principle investigator in the design of this project and first author responsible for writing and preparation of this manuscript.

Thavendran Govender: Provided assistance with technical and experimental input in the overall project.

Glenn E. M. Maguire, Bahareh Honarparvar, Hendrik G. Kruger: Co-supervisors

DEDICATION

This thesis is dedicated to my late grandparents, Rev Luke Mvuse Maphumulo and Mrs Grace Sebenzile Maphumulo whose teachings are still major pillars of my strength.

ACKNOWLEDGEMENTS

I would like to extend my thanks and gratitude to;

Dr. Bahareh Honarparvar, Prof. Hendrik G. Kruger and Dr. Glenn E. M. Maguire for their guidance and faith in me, I will forever be indebted to all of you.

Prof. T. Govender and Prof. T. Naicker for their mentorship and guidance throughout my academic studies. They made me realize how significant my contribution to science would impact and bring betterment to the society.

Centre for High Performance Computing for their technical platform and support.

The University of KwaZulu Natal Westville Campus, College of Health Sciences and National Research Foundation (NRF) for their funding during the course of this research.

The current and ex-fellow members at Catalysis and Peptide Research Unit are too numerous for me to acknowledge everybody, so I would like to thank the entire group both past and present members. In particular, I would like to thank the group members that I directly shared ideas with on projects; Dr. Amit Halder and Zaynab Fakhar for being present when I needed you the most to resolve all issues pertaining the project, Monsurat Lawal, Gideon Tolufashe, Zainab Sanusi, Thando Ntombela and the whole crew for making the journey worthwhile.

A special thanks to the Free African Congregational Church for your prayers and intercession.

My family, oZikode, oCwalile oDoncabe oMasiyane baseNgoleleni.

LIST OF ABBREVIATIONS

AIDS	Acquired Immune Deficiency Syndrome
Arg	Arginine
Asp	Aspartic acid
ATV	Atazanavir
C-SA	Subtype C South African
DRV	Darunavir
Env	Envelope gene
FDA	Food and Drug Administration (USA)
GAFF	General AMBER Force Field
GB	Generalized Born solvation model
Glu	Glutamic acid
HAART	Highly active antiretroviral therapy
H-B	Hydrogen Bond
Ile	Isoleucine
LPV	Lopinavir
MD	Molecular dynamics
MM-GBSA	Molecular Mechanics Generalized Born Surface Area
MM-PBSA	Molecular Mechanics Generalized Poisson–Boltzmann Surface Area

NS	Natural Substrate
PCA	Principal Component Analysis
PDB	Protein Data Bank
Pol	DNA Polymerase
RMSD	Root-Mean-Square Deviation
RMSF	Root-Mean Square Fluctuation
RNA	Ribonucleic acid
TPV	Tipranavir
UNAIDS	The Joint United Nations Programme on HIV/AIDS
VMD	Visual Molecular Dynamics

LIST OF FIGURES

CHAPTER ONE

Figure 1. Schematic diagram of HIV virus. Reprinted from Open Access literature[33]. (Copyright under Creative Commons Attribution-Share Alike 3.0 (*http://creativecommons.org/licenses/by-sa/4.0*) **4**
Figure 2: Schematic diagram of HIV virus life cycle. Reprinted from Open Access literature[33]. (Copyright under Creative Commons Attribution-Share Alike 3.0 (http://creativecommons.org/licenses/by-sa/4.0) **7**
Figure 3. Structures of HIV PR inhibitors. **12**
Figure 4: Ribbon representation of different flap handedness in HIV-1 protease. The handedness of (A) substrate-bound protease in closed form (e.g., 2BBB), (B) free protease in semi-open form (e.g., 1HHP). The difference in handedness between closed (C) and semi-open (D) flap conformation, with the flaps being highlighted by rotating the structures 90° horizontally behind the plane. **17**

CHAPTER TWO

Figure 1. Chemical structures of four of the FDA approved second generation PIs. **43**
Figure 2. (A) Root Mean Square Deviation and (B) Root Mean Square Fluctuation for the 4 systems over the 60 ns MD trajectory. **50**
Figure 3. Distance and angle analyses for the elucidation of flap dynamics in C-SA PR for the 60 ns MD simulations; (A) distance between flap tips (Ile50 and Ile50′), (B) distance between active site residue and flap tip in chain A (Asp25-Ile50), (C) distance between active site residue and flap tip in chain B (Asp25′ -Ile50′), (D) TriCα angle of chain A (Gly49-Ile50-Gly51), (E) TriCα angle of chain B (Gly49′-Ile50 ′-Gly51′), (F) dihedral angles of chain A (Ile50-Asp25-Ile50 ′ -Asp25′). **53**
Figure 4. Cross-correlation matrices of the fluctuations of the coordinates for Cα atoms around their mean positions for each complex over the 60 ns MD simulations..... **59**
Figure 5. Per-residue decomposition profiles of the ATV and DRV complexes using the MM-GBSA approach. **63**
Figure 6. Per-residue decomposition profiles for the LPV and TPV complexes using the MM/GBSA approach. **64**
Figure 7. Collective motions corresponding to PC1 obtained by performing principal component analysis on a 60 ns MD simulation trajectory after the equilibrium (the protein part is presented in red ribbon, higher fluctuating parts are presented in grey and blue). **66**

CHAPTER THREE

Figure 1. Illustration of three important regions of dimerization interface of HIV-1 PR. **78**
Figure 2. Root Mean Square Deviation of the backbone Cα atoms for the native (untruncated), single and double-truncated systems of (A) Apo enzyme, (B) Natural substrate (NS), (C) ATV complex and (D) TPV complex over the 100 ns MD trajectory. **82**
Figure 3. Root-Mean Square Fluctuation of backbone atoms against residue number for the APO and NS systems in native form over the 100 ns MD trajectory. **83**
Figure 4. Root-Mean Square Fluctuation of backbone atoms against residue number for the

APO and NS systems in single-truncated (ST) form over the 100 ns MD trajectory. **83**
Figure 5. Root-Mean Square Fluctuation of backbone atoms against residue number for the APO and NS systems in double-truncated (DT) form over the 100 ns MD trajectory. **84**
Figure 6. Plot of per-residue decomposition analysis for the native natural substrate complex (NS-Native). **87**
Figure 7. Plot of per-residue decomposition for the single-truncated system of the natural substrate complex (NS-ST). **88**
Figure 8. Per-residue decomposition for the double-truncated system of the NS complex (NS-DT). **88**
Figure 9. The plot of radius of gyration for three different systems; native, ST and DT of (A) apo enzyme, (B) natural substrate complex. **89**
Figure 10. The plot of distance between Gln92 and Gln92' for situated near the β-sheet for the different systems, Native, ST and DT of (A) Apo enzyme, (B) natural substrate complex. ... **90**
Figure 11. The distance (Å) between the flap tip residues (Ile50 and Ile50') for three different systems; native, ST and DT of the (A) apo form and (B) natural substrate complex obtained from a 100 MD simulation. **91**
Figure 12. A 3-D illustration of the two residues (Arg8 and Asp29') that form a salt-bridge in HIV-1 PR. **92**
Figure 13. Distance of Arg8 and Asp29' measured during the 100 ns MD simulation for the native and truncated forms of all systems studied. **95**
Appendix 1. Supplementary material for Chapter 2
Figure S1. Comparison of the RMSD plots of triplicate MD simulation (conformation B, C and D) for (A) ATV and (B) DRV with RMSD of original MD simulation (conformation A). **104**
Figure S2. (A) Root Mean Square Deviation (RMSD) and (B) Root Mean Square Fluctuation for the apo enzyme and inhibitor-bound systems over the 60 ns trajectory. **104**
Figure S3. Distance and angle analyses for the elucidation of flap dynamics in C-SA PR for the 60 ns MD simulations; (A) distance between flap tips (Ile50 and Ile50'), (B) distance between active site residue and flap tip in chain A (Asp25-Ile50), (C) distance between active site residue and flap tip in chain B (Asp25'-Ile50'), (D) TriCα angle of chain A (Gly49-Ile50-Gly51), (E) TriCα angle of chain B (Gly49'-Ile50'-Gly51'), (F) dihedral angles of chain A (Ile50-Asp25-Ile50'-Asp25'). **105**
Figure S4. Per-residue decomposition profiles of the natural substrate complex using the MM/GBSA approach. **106**
Figure S5. Cross-correlation matrices of the fluctuations of the coordinates for Cα atoms around their mean positions after the equilibrium of the 60 ns MD simulation. **106**
Figure S6: 3-D structure representing hydrogen-bond interaction of NS with certain residues of HIV-1 C-SA PR. **107**
Figure S7: 3-D structure representing hydrogen-bond interaction of ATV with certain residues of HIV-1 C-SA PR. **107**
Figure S8: 3-D structures representing (A) hydrogen-bond interaction of DRV with certain residues of HIV-1 C-SA PR and (B) the location of the sulfonamide group inside the complex system. (Note: the black star in the flap tip position denotes the removal of flap tip residues for

illustration purposes) **108**
Figure S9: 3-D structures representing (A) hydrogen-bond interaction of TPV with certain residues of HIV-1 C-SA PR and (B) the location of the sulfonamide group inside the complex system. **108**
Figure S10: Collective motions corresponding to PC1 obtained by performing principal component analysis on MD simulation trajectory after the equilibrium for DRV and LPV (the protein part is presented in red ribbon, higher fluctuating parts are presented in grey and blue. **109**
Appendix 2. Supplementary material for Chapter 3
Figure S5: Comparison of the RMSD plots of triplicate MD simulation (conformation B, C and D) for Apo-native with RMSD of original MD simulation (conformation A). **110**
Figure S2: Root-Mean Square Fluctuation of backbone atoms against residue number for the APO, NS, ATV and TPV systems in Native forms. **110**
Figure S3: Root-Mean Square Fluctuation of backbone atoms against residue number for the APO, NS, ATV and TPV systems in Native forms. **111**
Figure S4: Root-Mean Square Fluctuation of backbone atoms against residue number for the APO, NS, ATV and TPV systems in Native forms. **111**
Figure S5: The plot of radius of gyration for three different systems; native, ST and DT of (A) ATV and (B) TPV complexes..... **112**
Figure S6: The distance (Å) between the flap tip residues (Ile50 and Ile50') for three different systems; native, ST and DT of the (A) ATV and (B) TPV complex obtained from a 100 MD simulation. **112**
Figure S7: The distance (Å) between two Gln residues from chain A and B for the three systems; native, ST and DT for (A) ATV and (B) TPV during the 100 ns MD simulations. **113**
Figure S8: Plot of per residue decomposition energy in kcal/mol for the ATV-Native complex system. **113**
Figure S9: Plot of per residue decomposition energy in kcal/mol for the ATV-ST complex system. **114**
Figure S10: Plot of per residue decomposition energy in kcal/mol for the ATV-DT complex system. **114**

LIST OF TABLES

CHAPTER ONE

Table 1: Chronology of events that took place in the development of HIV-1 therapy.[57]…… **9**

CHAPTER TWO

Table 1. Calculated binding free energies [$\Delta G_{bind(T)}$] and its components for the four second generation HIV-1 protease inhibitors and the natural substrate (NS) on C-SA PR. The experimental binding energies [$\Delta G_{bind(E)}$] (15) is also depicted. The energy components are in kcal/mol and entropy (-TΔS) in kcal/mol…… **48**

Table 2. Average RMSFs observed in different regions of the apo, natural substrate and drug- C-SA protease complexes for the 60 ns MD simulation. ……..…… **52**

Table 3. Hydrogen bond interaction occurring between the ligand and residues in the active site and flap tip region of the protease for different bound systems…….…….. **61**

CHAPTER THREE

Table 1. Calculated binding free energies [ΔG_{bind}] and its components for two HIV-1 protease inhibitors and the natural substrate (NS) on C-SA PR. All data values are in kcal/mol…….. **85**

Table 2. Hydrogen Bond interaction formed between Asp29/29' and Arg8/8' of APO enzyme for the native and truncated systems…… **93**

Table 3. Hydrogen Bond interaction formed between Asp29/29' and Arg8/8' of NS-complex for the native and truncated systems…… **94**

CHAPTER ONE

1. INTRODUCTION

1.1 Background

Since its discovery over three decades ago, when initial reports of the United Center for Diseases Control (CDC) were made [1], Acquired Immunodeficiency Syndrome (AIDS) is still regarded as the most devastating disease outbreak to date [2]. Many studies on the Human immunodeficiency virus (HIV) have been performed over the last few decades, as it is the causative agent for AIDS [3-5]. This has led to epidemic proportions in various parts of the world, whereby more than 36 million people were living with HIV in 2016, while over a million deaths have been recorded as a result of the disease [6, 7]. The progression of HIV infection to AIDS is the leading factor to the extreme suppression of the immune system in infected humans. Of global infections recorded [8], east and southern-African regions represent over 52 % of all HIV infections and nearly 19.5 million people are living with HIV. In 2016 [7], new HIV infections within the east and southern-African region displayed a 29 % decline from 1.1 million to approximately 790 000 infections per year. Within the continent of Africa, South Africa has been shown to be at the epicentre of the epidemic [9] as it counts for over 7 million individuals who are infected with HIV and 110 000 deaths resulting from AIDS-related illnesses [7].

Historically, long-term illness, and ultimately death were strongly associated with AIDS in the early discovery of this epidemic. However, morbidity and mortality due to AIDS have shown to significantly decline to a manageable condition [10] and this is attributed to the introduction and use of the highly effective combination therapy, Highly Active Antiretroviral Therapy (HAART) [11] from the late 1990s [12]. Dramatic improvements were

achieved through the introduction of this therapy for HIV infected people, and also immunological as well as virological response to treatment was experienced by the majority of infected individuals [13].

1.2 HIV (HIV-1 & HIV-2)

Different strains of HIV may be carried by an infected individual in their body. There are two main types of the virus: HIV-1 and HIV-2 and a further classification into groups and subtypes exist. The isolation of HIV-1 was achieved by R.C. Gallo and L. Montagnier in 1983 and this was followed by reports of another type; HIV-2 [4]. HIV-1, the most common type found worldwide, is believed to have originated from the *simian immunodefiency virus* of chimpanzees (SIVcpz) [4, 14] whereas HIV-2, mainly found in Western parts of Africa and some regions in India and some in Europe is believed to genetically related to *sooty mangabey virus* (SIVsmm) [15]. New HIV-1 infections have declined by approximately 33% between the period of 2005 to 2013 [16]. However, it is worth mentioning that the incidence of HIV-1 is still very high despite the decline in new infections [16].

Despite many similarities that can be identified between HIV-1 and HIV-2, they can, however, be clearly distinguished from each other. For example, a region which encompasses the core of mature HIV-1 virions has shown to be more electron-lucent when compared to the one observed in HIV-2 [17]. The transmission of HIV-1 by sexual contact is more efficient than that of HIV-2 [18]. Reduced pathogenicity of HIV-2 in comparison to HIV-1 as well as enhanced immune control of HIV-2 infection and consistent lower viral loads in the plasma of HIV-2 are also other major differences. When these differences are looked into, they can provide necessary insights into the evolution, tropism and pathogenesis of the virus [19]. Despite having considerable phenotypic and sequence differences, HIV-1 and HIV-2 are morphologically quite similar [19, 20]. Basic free

arrangement, intracellular replicate pathways and modes of transmission are some of the similarities shared by HIV-1 and HIV-2 [21].

Group M (main), group O (outlier), group N (non-M/ non-O) and group P [22] are major phylogenetic groups by which HIV-1 variants are classified [22-26]. Group M is further subdivided into 10 phylogenetic subtypes (A to K) and this group constitutes the majority of HIV-1 infections worldwide. Within group M, the average intersubtype genetic variability is 15 % for the gag gene and 25 % for the env gene [22, 25]. Recent studies on HIV1 recorded that subtypes A, B and C, are the most prevalent genetic forms of HIV-1 worldwide, whereby subtype C accounts for approximately 50 % of all HIV-1 infections on a global scale [9]. Countries in central and eastern African regions such as Uganda, Kenya, Rwanda, Tanzania and those countries in Eastern Europe are mostly affected by the subtype A virus, which is found to be predominant in these parts [27-29].

1.3 HIV-1 virus infection –life cycle

HIV is an enveloped, single-stranded RNA retrovirus and is classified as falling within the *lentivirus* subfamily of retroviruses [2, 30] and they're known for their ability to reversetranscribe a genome of a 9,2 kb single-stranded RNA into a linear orientated double stranded DNA by viral transcriptase [2, 4]. Morphologically HIV-mature virions are known to possess a diameter of 100-120 nm and a lipid-bilayer membrane surrounds the viral particle (Figure 1). The outer layer of the virus, the viral envelope (*Env*), acts as the outer shell of the virus and is the only surface protein. Situated on the HIV-1 envelope are two molecules, glycoprotein (gp120) situated on the surface and the transmembrane protein (gp41) and the presence of these leads to the formation of spikes on the virion's surface [31]. There are three major genes carried by HIV-1; (*gag, pol and env*) and these form characteristic of all retroviruses [15, 32].

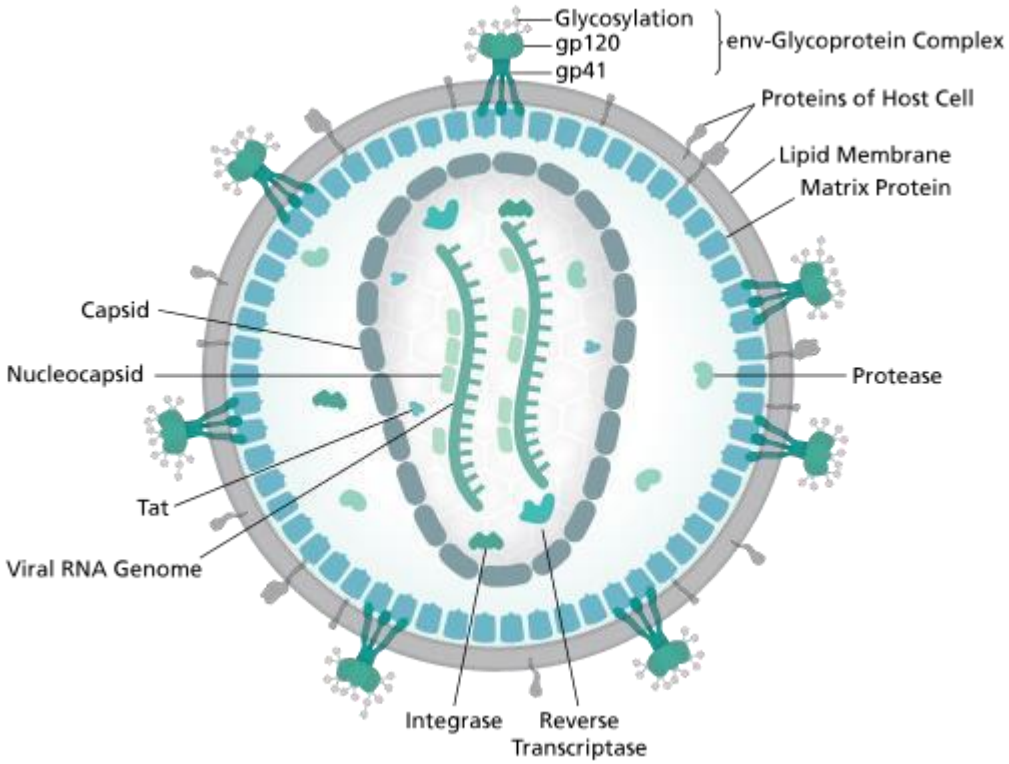


Figure 1. Schematic diagram of HIV virus. Reprinted from Open Access literature[33]. Copyright under Creative Commons Attribution-Share Alike 3.0

The immune system is damaged when the HIV virus attacks the white blood cells (i.e. T-helper cells/ $CD4^+$ cells). The role that $CD4^+$ cells play in the well-functioning of the immune system and its T4 lymphocytes is vital. In order for the HIV virus to achieve growth and reproducibility, it needs to develop new copies of itself and thus damaging and weakening the immune system [34]. This process takes place by a three-phased mechanism initiated by the direct viral killing of infected cells [34, 35]. This is followed by increased rates of apoptosis found in infected cells and lastly, the ultimate killing of infected T-helper cells by CD8 cytotoxic lymphocytes which have the ability to detect infected cells [36]. The continuous development of T-helper cells is termed the HIV life cycle [35] and the stages of how the life cycle of HIV occurs will be discussed next.

1. Binding and Fusion: The life cycle of HIV is initiated by the binding of the viral surface protein Envelope (*Env*) [37, 38]. This attachment of the virus is made possible

by the presence of the spikes (gp41 and gp120) situated on the surface of the virus particle (**Figure** 1) [39]. These proteins bind or stick to the CD4 receptor as well as the two-main chemokine co-receptors, either CCR5 or CXCR4 [40-42]. After viral binding to the host cell, the envelope consisting of gp120 and gp41 heterodimers undergoes an irreversible conformational change [31, 43]. The initiated conformational change leads to fusion of the viral envelope and the $CD4^+$ cell membrane which ultimately leads to the infection of a $CD4^+$ cell [13, 31, 37, 41, 44]. Once the cell membrane of the virus has fused with that of the host cell, it leads to effective infection, whereby the genetic material ribonucleic acid (RNA), is introduced into the host cell [40].

2. Reverse Transcription: The single-stranded RNA of the virus is then converted to a double-stranded DNA when it is inside the host cell by the enzyme referred to as reverse transcriptase in the cytoplasm of the host cell [40, 45]. This step is susceptible to a number of mutations as several errors are reported to occur during the replication process [46].
3. Integration: Once the double-stranded DNA has been introduced into the host cell's nucleus *via* reverse transcriptase, it then integrates into a chromosome of the host cell. The integrase enzyme is responsible for this step and this leads to a provirus being formed, which is basically an integrated HIV DNA [46]. This provirus can exist in an inactive form for several years whereby very few copies of HIV are released [47]. The compounds that are responsible for hindering this process are termed integrase inhibitors.
4. Transcription and translation: Once the virus has been integrated and the host cell has been triggered to be active, the proviral DNA is transcribed into messenger RNA (mRNA). In the nucleus of the host cell, production of full-length RNA molecules

takes place [15, 46]. The newly transcribed mRNA is then translated by host machinery into viral polyprotein precursors consisting of the *env*, *gag* and *gag-pol* [46].

5. Assembly and Budding: After viral polypeptide precursors have been translated, they assemble in the cell membrane into immature (non-infectious) HIV particles [48, 49]. The strands of messenger RNA contain various copies of HIV genetic material and a new virus particle is assembled which results in the formation of a curvature [48]. This curvature leads to the formation of a bud and in the process of budding, part of the host cell's exterior envelope gets taken by the new virus (*via* endocytosis) and the new virus releases HIV protease enzymes, whose main objective is to cleave the long chains of HIV polyproteins to create mature viral proteins [2, 50, 51]. The newly assembled immature (non-infectious) viral particles gather to form mature (infectious) HIV particles and they bud from the host cell to be released for infection of other host cells [46, 52]. The exterior envelope of the new viral particle consists of HIV glycoproteins that are essential for interaction with the CD4 and co-receptors of the host cells. At this stage, the new virus cell selects two copies of genomic RNA and these are encapsulated inside the viral particle [53-55].

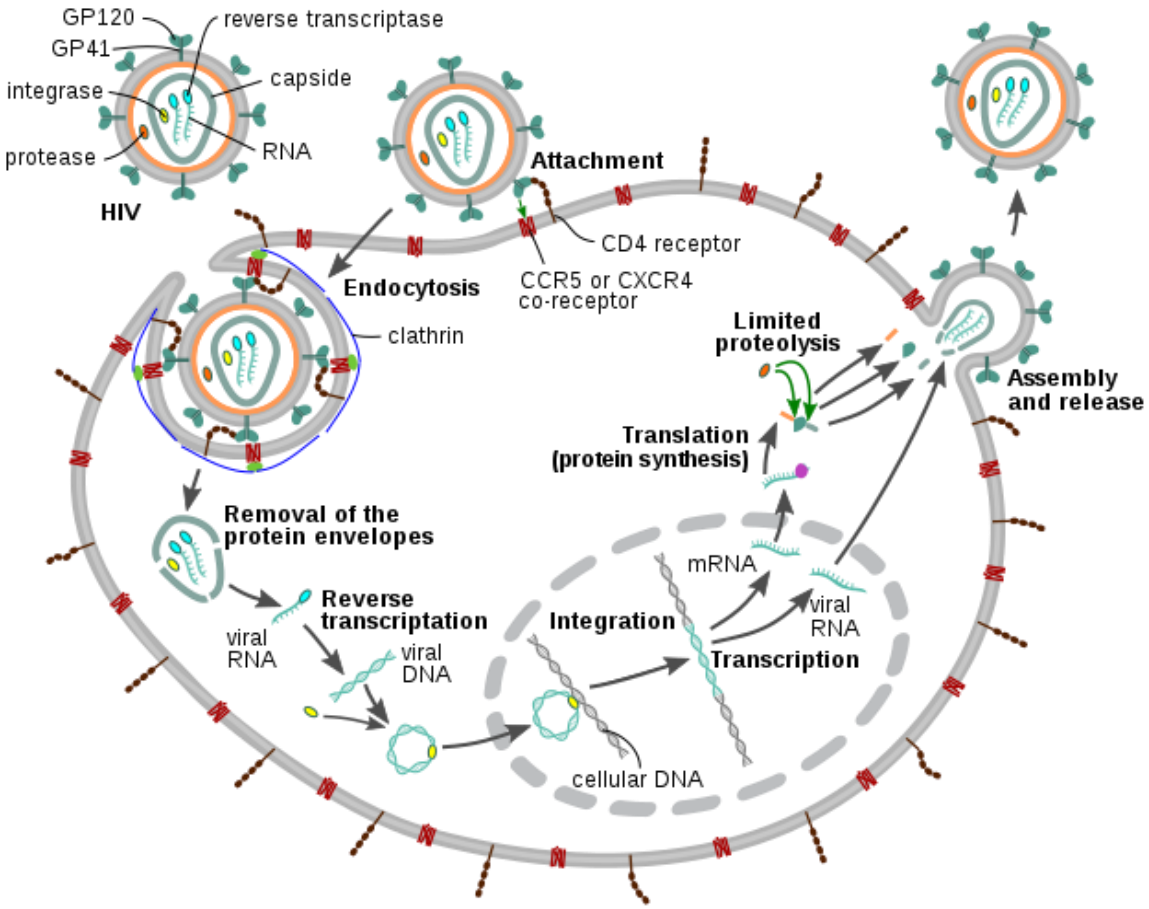


Figure 2: Schematic diagram of HIV virus life cycle. Reprinted from Open Access literature[33]. Copyright under Creative Commons Attribution-Share Alike 3.0

1.4 Different enzymes (Integrase, transcriptase, protease)

Proteins

The identification of the HIV retrovirus and understanding of HIV genome has led researchers around the world to focus largely on three major distinct proteins; integrase, reverse transcriptase, and a protease [56]. Development of inhibitors targeting various steps that constitute the life cycle of the HIV virus has been an area of interest for most scientists [2].

1.5 HIV- Inhibitors (RTIs, PIs, IIs, FIs)/ HAART/ side effects

Since the outbreak of acquired immunodeficiency syndrome (AIDS) epidemic, over three decades ago when its causative agent was first isolated, the understanding and accumulated knowledge about the virus has led to tremendous efforts being made in transforming the HIV1 disease, which was initially regarded as a highly fatal infection, into a treatable chronic infection [12, 57]. Application of antiretroviral therapies has increased in possibility of survival and resulted in a decline in death rates of HIV patients[57]. It is understood that theoretically, the different steps of HIV life cycle are identified as targets for the development of antiretroviral therapy. However, the current highly active antiretroviral therapy (HAART) [10, 58], consists of at least three major groups of inhibitors drawn from four classes which are used in the clinical practice and they are inhibitors of the reverse transcriptase (nucleoside/ nucleotide, NRTIs, and non-nucleoside, NNRTIs), integrase inhibitors (IIs), protease inhibitors (PIs) [12, 59] and the recent first fusion inhibitor (FI). The use of HAART to most patients has shown to result in an increase in $CD4^+$ T cell counts, reduction in morbidity and mortality [60, 61].

While major strides have been made concerning the use and application of HAART, it is well understood that its success is limited by numerous severe side effects associated with the treatment (Table 1) [60, 62, 63]. Insulin resistance hyperlipidemia, subcutaneous fat wasting, lipodystrophy, hypertoxicity, hyperglycemia, diarrhoea etc. [57, 62, 64] are some of recorded side effects associated with long-term use of the antiretroviral therapy. Besides the recorded side effects, there are also numerous drug resistant strains which have developed [50, 65].

Table 1: Chronology of events that took place in the development of HIV-1 therapy. [57]

Year	HIV therapy
1981	Recognition of AIDS [4]
1983	Isolation of HIV-1 [66]
1987	Introduction of first FDA-approved therapy,
1991	Approval of most NRTIs by FDA [57]
1995	Understanding of dynamics on HIV virus [68]
1995/1996	Introduction of the first protease inhibitors
1996	FDA approval of the first NNRTI
2003	Introduction of fusion inhibitors [71]

The first discovery of a therapy targeting HIV was that of a nucleoside analogue called zidovune, designed to inhibit HIV-1 reverse transcriptase and it received FDA approval in 1987 [57, 67]. Inhibitors of this type, nucleoside (NRTIs), interfere with the replication of HIV-1 by blocking the reverse transcriptase, which is used in conversion of the viral RNA into DNA [57, 67]. This results in the slowing down of the replication process and in the case of HIV-1 proviral DNA being terminated [67, 72]. Between the year 1991 and 1995, the period where AIDS was being regarded as the number one killer in the world, additional NRTIs were introduced. The rest of current NRTIs were also introduced after that (**Table** 1).

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are a class of drugs which are structurally diverse and bind directly and non-competitively to the enzyme reverse transcriptase. It is interesting to note that despite the chemical differences in their structures, NNRTIs tend to bind at a region distant from the normal binding site of the substrate. The binding of NNRTIs induces conformational change in the substrate-binding site of reverse transcriptase and leading to reduced DNA polymerase activity and thus inhibiting the activity of RT [57]. Unlike nucleoside analogues, these inhibitors do not exhibit activity against RT of HIV-2.

When scientists gained a better understanding of the virus, they recognized other drug targets in addition to reverse transcriptase. Drug development progress was made when a new class of drugs, enfurvitide (T-20), was introduced and was approved in March 2003 by the FDA[57, 58]. Enfurvitide is currently still the only fusion inhibitor approved by the FDA [10, 57]. Structurally, it is a linear, 36-amino-acid synthetic peptide which was designed to inhibit HIV-1 entry into host cells [10]. The fusion portion of the HIV-1 envelope glycoprotein in gp41 is the main target of these drugs as they prevent the full conformational changes of the viral envelope necessary for the fusion of viral membrane and that of the host cell [57, 73]. As is the case for many drugs, enfurvitide also has limitations and these include its requirement to be administered by subcutaneous injection for human administration [74].

Major successes were made in the discovery and development of other HIV targets and these included those of the viral protease. This enzyme is essential and plays a major role in ensuring proper virion maturation and assembly in the post-translational process [75-77]. HIV protease has been a prime target for drug design since the late 1980's when the first crystal structure of HIV1 protease was reported [78, 79]. This is because of its indispensable role in viral maturation which ultimately leads to final morphological rearrangement [2, 80, 81]. The introduction of inhibitors results in production of immature virus particles and thus ensuring that virus replication is hindered while reducing the spread of viral infection in susceptible host cells [2, 76, 82]. It was the discovery of the crystal structure of the HIV-1 protease that made it possible for the design of antiviral protease inhibitors to be achieved [76]. The design of most protease inhibitors mimics the transition state of the actual substrate of the enzyme [83-85].

Ideally protease inhibitors should possess good oral bioavailability in humans, potency against HIV clinical isolates, and specificity for HIV-1 protease [57]. Potency of inhibitors is highly affected by their affinity for the protease, their penetration ability to cells and most importantly, how strongly they are bound to the protein.

Saquinavir (SQV), popularly known by its brand name *Invirase* is a peptidomimetric hydroxyethylamine and was developed by F. Hoffman La Roche Ltd [75, 86, 87]. It was marketed in 1995 and it was the first HIV protease inhibitor used to treat AIDS patients to be approved by FDA.

Ritonavir (RTV), also known as *Norvir* in the market, is a peptidomimetric HIV protease inhibitor and was developed by Abbott Laboratories that was approved in 1996 [83, 87]. It was designed on the basis that it would have a "perfect" fit in the C2-symmetry of the protease binding site [65]. The isopropyl thiazoyl P3 group in ritonavir is found to be much longer than the one found in other HIV protease inhibitors [88]. Initially, there were challenges associated with RTV, for example, the developers of the drug started with the compounds that exhibited activity for the virus but showed poor bioavailability. This then led to some improvements being to the structures of the compounds as pyridal groups were added in place of terminal phenyl residues, which were removed leading to increased water solubility. These improvements resulted in RTV being the final product [75, 83, 88]. In spite of the gastrointestinal side effects associated with it, RTV is still regarded as a strong inhibitor of the cytochrome P_{450} enzyme mediated metabolism[75] as it also prevents the metabolism of other protease inhibitors [83].

Merck & Co, Inc., developed indinavir (IDV) and it was given the brand name *Crixivan* when it was approved in 1996 [52]. It is a peptidomimetric hydroxyethylene HIV protease inhibitor and it was designed on the basis of molecular modelling and the X-ray crystal structure of an earlier drug lead compound bound to the enzyme [89]. An advantage of this inhibitor is the fact that it is effective for both HIV-1 and HIV-2 [83]. A relative short half-life of circulating indinavir and low solubility are some of its main drawbacks [90].

Nelfinavir (NFV), also termed *Viracept*, was the first example of a nonpeptidic, potent HIV protease inhibitor [91]. It was developed by Agouron Pharmaceuticals in USA and was approved in 1997 [83]. This orally bioavailable agent was designed with the help of an iterative protein cocrystal structure analysis of peptidic inhibitors [75]. The structure of nelfinavir has the same DIQ group as saquinavir in one terminus while on the other it possess a 2-methyl-3hydroxybenzamide group [56].

Vertex Pharmaceuticals Incorporated developed Amprenavir (APV) (*Agenerase*) and it was approved by the FDA in 1999 (Figure 3) [87]. APV is an N,N-disubstituted amino sulfonamide nonpeptide HIV protease inhibitor [92]. APV contains a similar core to that of SQV but the functional groups on both ends are somewhat different. There is a tetrahydrofuran carbamate group on one end while an isobutylphenyl sulphonamide is located on another end with an added amide [75]. There are fewer chiral centers present in APV than in other HIV protease inhibitors. In 2004, APV was withdrawn from the market as its prodrug, fosamprenavir, proved to be superior in many aspects. Fosamprenavir (FPV) is the phosphate ester prodrug of APV and was approved for use in 2003. This inhibitor also possesses better solubility and bioavailability than APV [93].

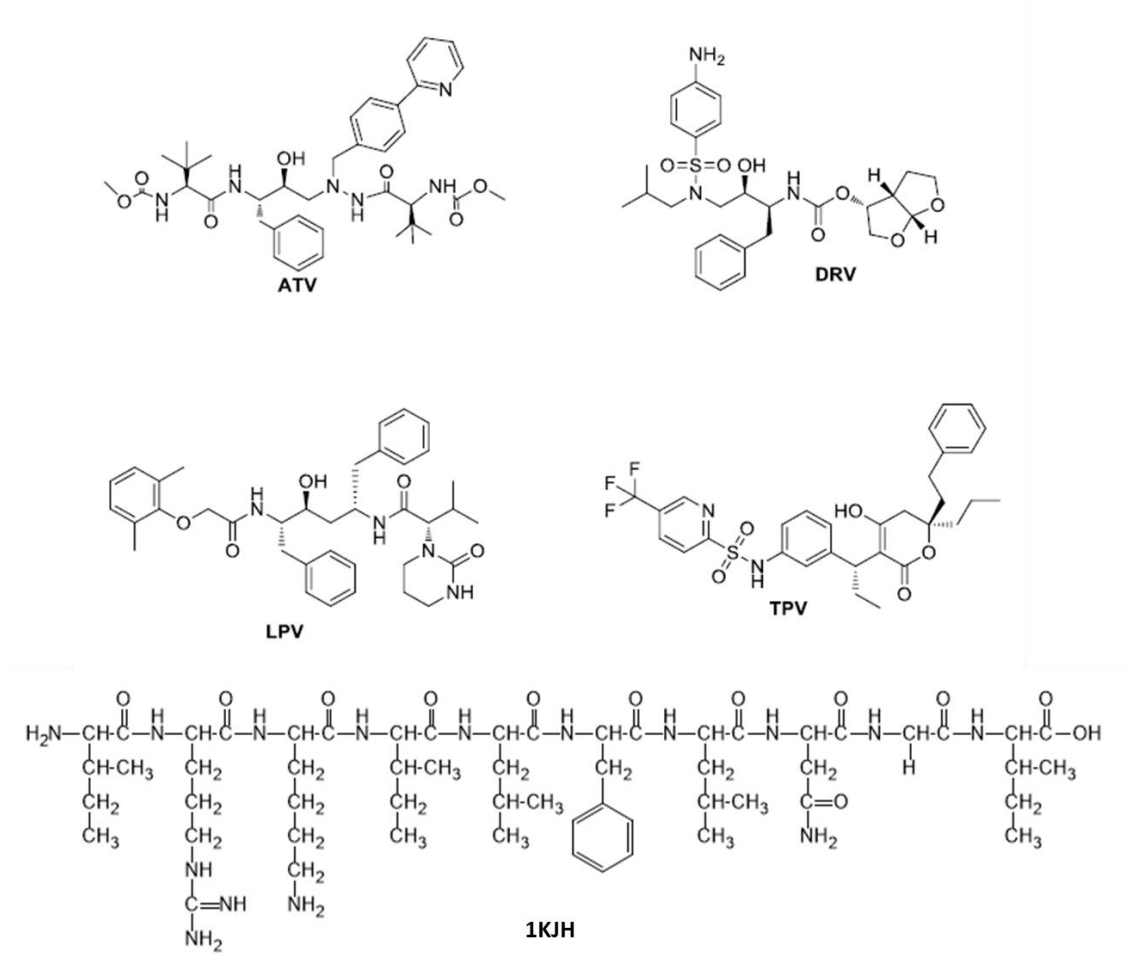


Figure 3. Structures of HIV PR inhibitors and the natural substrate (1KJH).

Lopinavir (LPV) was developed by Abbott Laboratories as ABT-378 with the intention to ensure that the activity against Val 82 mutant HIV protease is retained [84]. It received approval by the FDA as a coformulation with ritonavir under the brand name Kaletra on 15 September 2000 [96]. It was found that there is a substantial increase in LPV drug exposure as a result of coformulation with RTV by inhibiting cytochrome P450 isoenzyme 3A4 [84, 96].

An aza-dipeptide analog, Atazanavir (ATV), (brand name: *Reyataz)* was developed by BristolMyers Squibb Co in 2003 and it has potent anti-HIV activity [83, 87]. ATV has a unique structural characteristic as it has a large phenylpyridyl P1 group which is situated asymmetrical relative to its benzyl P1' group and it was designed to fit the C2-symmetry of the HIV protease binding site [94]. What is interesting about this inhibitor is that it requires an acidic environment

for adequate absorption [95]. Previous reports have shown that fewer side effects are associated with ATV than with other HIV protease inhibitors [96, 97].

Tipranavir (TPV) (brand name: *Aptivus*) was developed in Germany by Boehringer Ingelheim GmbH and was approved in 2005 [83]. It was developed from a nonpeptidic coumarin template and is a sulfonamide which has 5,6-dihydro-4-hydroxy-2-pyrone [98, 99]. TPV uses the lactone oxygen atom situated on the dihydropyrone ring for direct interaction with the Ile50 residues found in the flap region that ensures stability of the complex [83]. The structural orientation of tipranavir in the active site results in seven direct hydrogen bonds with residues of the protease that are conserved and as such, water-mediated hydrogen bonds are fewer when compared to other PIs [100]. Intracranial haemorrhage and decompensated hepatitis are two adverse effects associated with TPV.

Known as *Prezista* in the market, darunavir (DRV) was approved in 2009 by the FDA [87]. It is a nonpeptidic analogue of amprenavir. There are structural and chemical similarities between amprenavir and its analogue, darunavir but they are distinguishable from each other by the number of tetrahydrofuran (THF) groups, APV has one and DRV has two of these groups [83]. Another factor that distinguishes these two inhibitors is the fact that darunavir has a higher binding affinity to PR [101, 102]. The increased binding affinity of darunavir is believed to be linked to the bis-THF moiety at P' [103] which leads to strong hydrogen bonds being formed with the backbone atoms found at residue D29 and D30 [100, 104]. Minor toxicity was observed for darunavir in some of the infected patients. Diarrhoea, rash, nausea and nasopharyngitis are some of the recorded side effects associated with the drug [105].

It is of critical importance for inhibitors to have a high binding affinity for the wild-type HIV protease and possess resilience when mutations occur [106-108]. Despite the side effects that have been observed and reported, it can be stated that HAART has brought dramatic changes in ensuring that morbidity and mortality rates of people infected with HIV are significantly

reduced [57]. Reports now have shown that drug design research on HIV/AIDS is aimed at discovering anti-HIV drugs that will possess low toxicity, reduced resistance, less side effects, improved potency and better oral bioavailability [32, 109]. Requirement for anti-HIV drugs that will act on alternative target areas is of importance. Numerous new anti-retrovirals, which could be potential anti-HIV drugs are currently still under investigation [110-113]. They range from (L-731, 988 and S-1360) [110, 111, 114], gp120-CD4 interaction inhibitors (BMS-378806, FP21399, PRO542, zintevir, BMS806 and PRO140) [113, 115],[116] inhibitors of CXCR4 coreceptor (ALX40-4C; T22, T134, T140; AMD3100; AMD3465) and those of CCR5 coreceptor (AK602, E913, NSC-651016, TAK-220, TAK-779, SCH-C, SCH-D) [61, 117] fusion inhibitors (T1249) [112, 118-120]. In addition to these, many more anti-retrovirals are also still being investigated [112, 121, 122].

1.6 Flap dynamics of HIV-PR

1.6.1 The structural information of HIV-1 protease

HIV-1 protease falls under a class of aspartic proteases which is composed of two flexible βhairpin flaps that are necessary in the control of substrate entry to the binding cavity (Figure 4) [77, 109, 123]. The flexibility of the flaps is said to be necessary for facilitating ligand entry, binding and also product release [75, 124]. The homodimeric HIV-PR normally contains 99 amino acid residues in each monomer and each monomer consists of one of two catalytic aspartyl residues.

Like most other aspartic proteases, HIV-1 PR contains the catalytic triad (Asp25-Thr26-Gly27) at position 25, 26 and 27 respectively [77, 124, 125]. The aspartyl groups found in proteases are found to exist in opposite states of protonation and this is to achieve the catalytic competent form of monoprotonation [32, 77]. The flap tips, residues 46-54 are hydrophobic despite having a high degree of solvent exposure [126].

The HIV-1 protease consists of three main regions which are briefly discussed next.

- Active site region (core region)

The core region contains the conserved catalytic triad, Asp25-Thr26-Gly27 covered by the βhairpin flaps which are also important as they participate in substrate and inhibitor binding and

product release [2, 75, 77]. The inhibitor interacts with the active site of the protease by forming hydrogen bonds and it is this network of bonds which ensures stability of the conserved core region.

- Terminal region (dimerization interface)

The terminal region (residues 1-5 and 95-99) is made up of antiparallel β sheets that result from the four termini in the dimer interface located at residues 1-5 and 95-99 [51, 127]. This particular region is regarded as being very important as it ensures formation of dimer and also stabilization of the active protease enzyme [128].

- The flap domain

  The flap domain consists of residues 32-63 [129] which are found to contain β-hairpin structures (residues 43–58) and they play a significant role in ligand binding interactions. The flap domain has glycine-rich ends, allowing it to be more mobile and flexible. The flaps can adopt many conformational states and they are said to be in a 'handedness' orientation in relation to each other [130]. The stability and movement of the flap domain is highly affected by the hinge region which is defined as the comprising residues 35–42 and 57–61 [131].

1.6.2 The stability of HIV protease

The stability of the protease is highly affected by pH and concentration [132, 133]. It is increased at higher pH values [132]. The presence of the inhibitor stabilizes the closed dimeric structure of the protease in a fashion that can be quantitatively described by a set of thermodynamic equations [134].

1.6.3 HIV-1 substrate binding and protease catalytic mechanism

Aspartic proteases, also known as acidic proteases, fall under the group of endopeptides [135]. Their enzymatic ability is also highly affected by the mechanical fluctuation and conformational changes on the protein structure [136]. In order for the segments of the substrate/ inhibitor to access the active site, the flaps of the protease (residues 33–62) needs to open [126] and for cleavage to occur the flaps need to close while the appropriate region of the substrate is in the active site. The conformational changes that occur in the HIV-1 protease are highly influenced by the flexibility of the flaps and the increased flexibility may lead to drug resistance [137]. This augmented movement is made possible by glycine rich ends of residues 46-56 allowing for easy binding and release of the product [138-140].

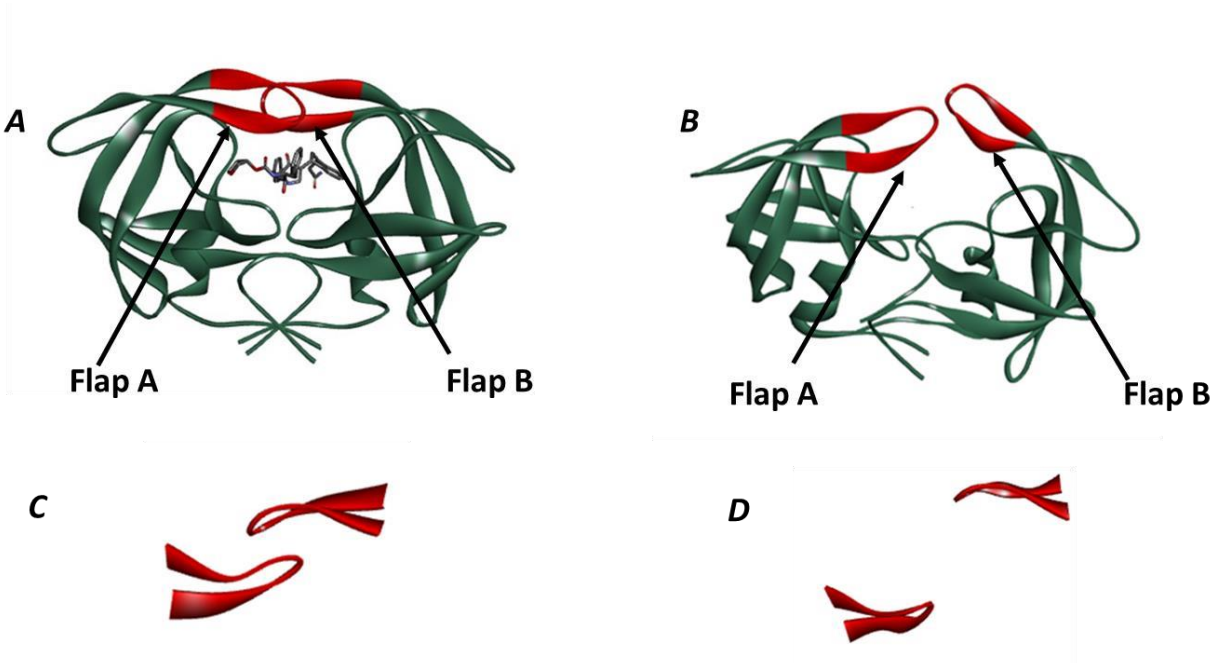


Figure 4: Ribbon representation of different flap handedness in HIV-1 protease. The handedness of (A) substrate-bound protease in closed form (e.g., 2BBB), (B) free protease in semi-open form (e.g., 1HHP). The difference in handedness between closed (C) and semi-open (D) flap conformation, with the flaps being highlighted by rotating the structures 90° horizontally behind the plane.

The two aspartyl β-carboxyl groups situated at the active site of the protease activate the water molecule in between the two residues. A side chain of another amino acid activates the nucleophile which is usually a hydroxyl group and the resulting nucleophilic atom then attacks the carbonyl of the scissile amide bond resulting in the formation of an ester or a thioester acyl intermediate [141]. The activated nucleophile affects the peptide in two ways: first, the carbonyl that is being attacked becomes a gemdiol and second, there is a hybridization change of the amide nitrogen from sp^2 to sp^3 [142]. Eventually, hydrolysis of the acyl enzyme intermediate by the water molecule takes place and the resulting products are formed [2]. Various reports have indicated that a water molecule is located between the active site aspartates[2].

1.7 Background on relevant studies of HIV-1 PR

1.7.1 Experimental studies

There have been several experimental studies that reported on binding affinities of a number of FDA-approved drugs for HIV-1 protease. Velazquez-Campoy *et al*. reported that subtype C HIV-1 protease displayed higher catalytic efficiency against two different substrates in comparison to A and B subtype proteases [143]. This was followed by another study where authors recorded that several protease inhibitors displayed lower binding affinities with factors from 2 to 7.5 against subtype A and B proteases exhibiting naturally occurring amino acid polymorphism[144]. The authors highlighted the fact that the drug-resistant mutation is significantly impacted by the pre-existing lower affinity in subtype A and C proteases.

A study exploring active site mutations in HIV-1 subtype C (C-SA) protease was reported. Their experimental work indicated that the mutations could not affect the catalytic activity of

the C-SA protease [9]. It was demonstrated that the binding affinities of protease inhibitors seemed to decrease as a result of the V82F and I84V double mutation in C-SA protease. A study by our group reported binding free energies for nine clinically approved HIV protease inhibitors against the I36T↑T mutant (100 amino acids in the monomer). This mutant displayed lower binding affinities and weaker inhibition when compared to the wild-type protease [145].

Various experimental techniques have been applied to study the flap opening and closing mechanism as well as dynamics of the flaps covering the active site of the HIV-1 protease. For many years, ^{15}N nuclear magnetic resonance spectroscopy (NMR) relaxation methods have been widely employed to analyse inhibitor-HIV-1 protease interactions [138]. With the use of these experiments, Ishima *et al.*[138] reported measurements of the transverse ^{1}H ^{15}N relaxation rates of backbone amides of HIV-1 in inhibitor-bound and unbound forms. The β-sheet interface of the inhibitor-bound protease was found to have some fluctuation on the millisecond timescale compared to the microsecond timescale for the unbound protease and this implies that the β-sheet interface for the inhibitor-bound form was rigid [138].

Several X-ray diffraction studies [79, 146, 147] on the free enzyme highlighted the fact that the flaps are loosely packed onto each other in their semi-open conformations. Reports on the overall structure of the enzyme showing that the structure is less compact when compared to the crystal structure of the ligand bound protease have also been published [148, 149]. In addition, other studies on the inhibitor-bound HIV-1 protease highlighted the fact that the β-hairpins at the flaps have a well-ordered orientation and tend to have interactions with most inhibitors [5].

Galiano *et al.* [150] reported results from site-directed spin labelling (SDSL) double electronelectron resonance (DEER) studies of the conformations of the flaps of HIV-1. Such results demonstrated that these employed techniques can distinguish conformations of the flaps in the free and inhibitor-enzyme complex forms.

Torbeev *et al.* [151] studied protein conformational dynamics in the mechanism of HIV-1 protease catalysis. The authors used chemical synthesis as well as relatively advanced physical methods to perform thorough analysis of dynamics-function correlations for the HIV-1 protease [151, 152]. Their results showed that the conformational isomerization in the flaps is correlated with structural reorganization of residues in the active site region. They also concluded that this preorganization of the active site is the rate-limiting factor in catalysis [151].

1.7.2 Computational studies

In addition to various experimental studies being reported, a huge impact has been made by computational reports on binding affinities of various protease inhibitors. A study by our group highlighted a comparative analysis of the binding affinities of nine FDA-approved drugs against subtype B as well as the South African subtype C [153]. From the MD study, it was observed that there was an increased flap movement for C-SA PR, which was suggested to be the reason for the weaker affinities. Molecular dynamics simulations were carried out on the complexes to calculate the average binding free energies for each inhibitor using the molecular mechanics generalized Born surface area (MM-GBSA) method. The binding affinities indicated that second generation FDA approved PIs (ATV, DRV, LPV and TPV) displayed a different activity pattern for C-SA protease compared to subtype B [154].

A number of computational techniques have also been reported on the flap opening and closing mechanism of HIV protease [155-157]. A study reported by Collins *et al.* [158] investigated flap opening in HIV protease using activated molecular dynamics simulations in gas phase. The authors used native (PDB: 3HVP) and MVT-101 bound conformations (PDB: 4HVP) of HIV protease, which differ in terms of handedness.

The curling of flap tips was proposed as a key conformational trigger for flap opening in HIVPR by Scott and Schiffer [126]. They highlighted an interesting phenomenon whereby the flap tip residues, Gly-Gly-Ile-Gly-Gly, curled back, thus burying several hydrophobic residues. The conformational flexibility that was highlighted was in good qualitative agreement with reported NMR experiments [138, 139]. However, the flap opening observed in such early studies only occurred within the first few nanoseconds of the calculation. This fact has been suggestive that improper equilibration had occurred [159].

1.7.3 Proposed parameters to measure flap dynamics

Various studies have reported on different parameters used to describe flap motion in aspartic proteases. A molecular dynamics study by Karubiu *et al.* [160]. reported different combined parameters that were proposed and applied to define the flap motion during the opening and closing in plasmepsin proteases [160]. The three combined parameters, the distance between the flap tip residues, d_1, the dihedral angle, Φ, and the TriCα angles, θ_2, were found to be the most appropriate in better defining the flap motion of plasmepsin (Plm) proteases.

Another study also reported proposed parameters in defining flap opening and closing on Plm proteases [161]. The authors considered d_1; Φ, and TriCα, θ_1 as being prime parameters to appropriately define the extent of opening and closing of flap structures in apo plasmepsins [161].

A comparative MD study of flap flexibility on beta-amyloid precursor protein cleavage enzyme 1 (BACE1) and enzyme 2 (BACE2) explored different combined parameters to better describe flap flexibility [162]. The authors found that all four proposed parameters, d_1, the dihedral angle, Φ, and TriCα angles, θ_1 and θ_2 seemed to appropriately define the twisting motion observed during opening and closing of flaps.

1.8 Computational background

Several computational methods have occupied a significant space in the study of biochemistry and physics. Most of computational methods have been used in the exploration of macroscopic and microscopic fluctuations that occur in biomolecular systems, ligand binding, and free energy surface [157, 163-167]. Several of these computational tools have been utilized to better understand the interaction of the substrate/ inhibitor in the active site region of the enzyme (HIV-1 protease). There are two broad areas within computational chemistry devoted to the structure of molecules and their reactivity. Two main methods constitute the area of computational chemistry; *molecular mechanics* and *quantum mechanics* whereby the former is based on the classical laws of physics to make predictions about structures and the properties of molecules [168-170]. The use of molecular mechanics (MM) is advantageous for very large systems such as enzymes. This method of calculation is characterized by a particular force field with embedded parameters, whereby the electronic effects are implicitly included in the force field through parameterization[168, 171]. Consideration of a solvent in MD simulations is vital as this ensures it eliminates artefacts due to a lack of solvent [172, 173].

Quantum mechanics utilise the laws of quantum theory instead of classical physics to describe a chemical system [165, 174]. This branch of computational chemistry solves the Schrödinger equation to determine the energy and other properties of a chemical system or a molecule [175]. MD simulations enable detailed information on the atom vibrations, movements, collisions, fluctuations and conformational changes to be obtained and this ultimately leads to studies on drug design and structure determination to be explored [176-178]. Normally the atoms of molecules are regarded as being in constant motion where they are heated to a specific temperature so as to overcome the potential energy barriers [172].

Explicit solvation includes the effect of water and uses hundreds or even thousands of discrete solvent molecules [173]. The large number of particle involved makes the explicit solvation calculations to converge only slowly [179]. This method is currently only practical for MM calculations. Despite the fact that this method of simulation is widely used its high computational expense is its drawback [179].

Implicit solvation is a model to approximate average effects of solvent by describing the interaction between the solvent and solute as a function of solute coordinates.[180, 181] This allows for faster sampling of conformational space [182, 183]. A number of previous studies have compared explicit and implicit solvent systems and concluded the latter to be much faster [182-185].

Calculating binding free energies [186, 187] is another important area that brings about a defining factor in studying interaction between ligands and enzymes. Binding free energy allows prediction of the ability of particular substrate to associate with the receptor and therefore provides predictions about binding affinities of the substrate [188, 189].

In this study, the binding affinities of various inhibitors ranging from the best, middle, weak, and weakest in terms of their binding energies obtained experimentally were studied to explore flap dynamics.

1.9 Objectives of the present study

This study forms part of a research project aimed at looking at how conformational flexibility of HIV PR, the key structural feature governing the binding mechanism of inhibitors, is affected by the strength of binding affinity. The specific objectives of this study were:

- To make necessary theoretical observations using molecular dynamics calculations as to which of the employed parameters best describe the flap dynamics of HIV-1 C-SA PR as some, flap tip distance for example, are affected by the curling of flap tip residues.

- To calculate the binding affinity of inhibitors for C-SA PR using MD and find the relationship between that and the influence of the respective inhibitors on flap flexibility.
- To determine how binding affinities of the inhibitors correlate to the complimentary/punitive interactions between the inhibitor-enzyme.
- To explore the impact of the dimerization interface on the stability of HIV-1 C-SA PR.

1.10 Outline of this thesis

This thesis consists of four chapters and they are presented as follows:

Chapter 1 presents an in-depth exploration of the background information on the main subject of this study.

Chapter 2 presents the first manuscript submitted for publication and it looks at the flap dynamics of HIV-1 C-SA PR in the bound and substrate-bound form.

Chapter 3 presents a second manuscript which is prepared for submission. This manuscript presents findings from a study on how dimerization inhibition affects the stability of HIV-1 CSA PR.

Chapter 4 provides a summarization, conclusion, and recommendations from the findings of this study.

1.11 REFERENCES

[1] Curran, J. W., and Jaffe, H. W. (2011) AIDS: the early years and CDC's response, *MMWR Surveill Summ 60*, 64-69.

[2] Brik, A., and Wong, C.-H. (2003) HIV-1 protease: mechanism and drug discovery, *Organic & biomolecular chemistry 1*, 5-14.

[3] Ko, G. M., Reddy, A. S., Kumar, S., Bailey, B. A., and Garg, R. (2010) Computational analysis of HIV-1 protease protein binding pockets, *Journal of chemical information and modeling 50*, 1759-1771.

[4] Gallo, R. C., and Montagnier, L. (1988) AIDS in 1988, *Scientific American 259*, 41-48.

[5] Wlodawer, A., and Erickson, J. W. (1993) Structure-based inhibitors of HIV-1 protease, *Annual review of biochemistry 62*, 543-585.

[6] Global, A. (2016) Update: UNAIDS Report, Geneva.

[7] Joint United Nations Programme on HIV/AIDS. (2017) Ending AIDS: progress towards the 90-9090 targets, *Global AIDS Update*. Geneva, Wealth Health Organization.

[8] Joint United Nations Programme on HIV/AIDS. (2015) On the Fast-Track to end AIDS by 2030: Focus on location and population, *UNAIDS 2030*. Geneva, Wealth Health Organization.

[9] Mosebi, S., Morris, L., Dirr, H. W., and Sayed, Y. (2008) Active-site mutations in the South African human immunodeficiency virus type 1 subtype C protease have a significant impact on clinical inhibitor binding: kinetic and thermodynamic study, *Journal of virology 82*, 11476-11479.

[10] Temesgen, Z., Warnke, D., and Kasten, M. J. (2006) Current status of antiretroviral therapy, *Expert opinion on pharmacotherapy 7*, 1541-1554.

[11] Fellay, J., Ledergerber, B., Bernasconi, E., Furrer, H., Battegay, M., Hirschel, B., Vernazza, P., Francioli, P., Greub, G., and Flepp, M. (2001) Prevalence of adverse events associated with potent antiretroviral treatment: Swiss HIV Cohort Study, *The Lancet 358*, 1322-1327.

[12] Carr, A. (2003) Toxicity of antiretroviral therapy and implications for drug development, *Nature Reviews Drug Discovery 2*, 624-634.

[13] Cooley, L. A., and Lewin, S. R. (2003) HIV-1 cell entry and advances in viral entry inhibitor therapy, *Journal of clinical virology 26*, 121-132.

[14] Franchini, G., Gurgo, C., Guo, H.-G., Gallo, R., Collalti, E., Fargnoli, K., Hall, L., Wong-Staal, F., and Reitz, M. (1987) Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses, *Nature 328*, 539-543.

[15] Lever, A. M. (2005) HIV: the virus, *Medicine 33*, 1-3.

[16] Kharsany, A. B., and Karim, Q. A. (2016) HIV infection and AIDS in Sub-Saharan Africa: current status, challenges and opportunities, *The open AIDS journal 10*, 34.

[17] YU, Q.-C., MATSUDA, Z., YU, X., ITO, S., ESSEX, M., and LEE, T.-H. (1994) An electronlucent region within the virion distinguishes HIV-1 from HIV-2 and simian immunodeficiency virus, *AIDS research and human retroviruses 10*, 757-761.

[18] Martinez-Steele, E., Awasana, A. A., Corrah, T., Sabally, S., van der Sande, M., Jaye, A., Togun, T., Sarge-Njie, R., McConkey, S. J., and Whittle, H. (2007) Is HIV-2-induced AIDS different from HIV-1-associated AIDS? Data from a West African clinic, *Aids 21*, 317-324.

[19] Reeves, J. D., and Doms, R. W. (2002) Human immunodeficiency virus type 2, *Journal of general virology 83*, 1253-1265.

[20] Gallo, S. A., Reeves, J. D., Garg, H., Foley, B., Doms, R. W., and Blumenthal, R. (2006) Kinetic studies of HIV-1 and HIV-2 envelope glycoprotein-mediated fusion, *Retrovirology 3*, 90.

[21] Nyamweya, S., Hegedus, A., Jaye, A., Rowland-Jones, S., Flanagan, K. L., and Macallan, D. C. (2013) Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis, *Reviews in medical virology 23*, 221-240.

[22] Hemelaar, J. (2012) The origin and diversity of the HIV-1 pandemic, *Trends in molecular medicine 18*, 182-192.

[23] Simon, F., Mauclère, P., Roques, P., Loussert-Ajaka, I., Müller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barré-Sinoussi, F., and Brun-Vézinet, F. (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O, *Nature medicine 4*.

[24] Ayouba, A., Souquières, S., Njinku, B., Martin, P. M., Müller-Trutwin, M. C., Roques, P., BarréSinoussi, F., Mauclère, P., Simon, F., and Nerrienet, E. (2000) HIV-1 group N among HIV-1seropositive individuals in Cameroon, *Aids 14*, 2623-2625.

[25] Buonaguro, L., Tagliamonte, M., Tornesello, M. L., and Buonaguro, F. M. (2007) Genetic and phylogenetic evolution of HIV-1 in a low subtype heterogeneity epidemic: the Italian example, *Retrovirology 4*, 34.

[26] Gürtler, L. G., Hauser, P. H., Eberle, J., Von Brunn, A., Knapp, S., Zekeng, L., Tsague, J. M., and Kaptue, L. (1994) A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon, *Journal of virology 68*, 1581-1585.

[27] McCormack, G. P., Glynn, J. R., Crampin, A. C., Sibande, F., Mulawa, D., Bliss, L., Broadbent, P., Abarca, K., Pönnighaus, J. M., and Fine, P. E. (2002) Early evolution of the human immunodeficiency virus type 1 subtype C epidemic in rural Malawi, *Journal of virology 76*, 12890-12899.

[28] Salminen, M. O., Johansson, B., Sönnerborg, A., Ayehunie, S., Gotte, D., Leinikki, P., Burke, D. S., and McCUTCHAN, F. E. (1996) Full-length sequence of an Ethiopian human immunodeficiency virus type 1 (HIV-1) isolate of genetic subtype C, *AIDS research and human retroviruses 12*, 1329-1339.

[29] Lihana, R. W., Ssemwanga, D., Abimiku, A. l., and Ndembi, N. (2012) Update on HIV-1 diversity in Africa: a decade in review, *AIDS Rev 14*, 83-100.

[30] Habeshaw, J., Wilson, S., Hounsell, E., and Oxford, J. (1999) How HIV-1 lentivirus causes immune deficiency disease, *Medical hypotheses 52*, 59-67.

[31] Wilen, C. B., Tilton, J. C., and Doms, R. W. (2012) HIV: cell binding and entry, *Cold Spring Harbor perspectives in medicine 2*, a006866.

[32] Vacca, J. P., and Condra, J. H. (1997) Clinically effective HIV-1 protease inhibitors, *Drug Discovery Today 2*, 261-272.

[33] Splettstoesser, T. (2014).

[34] Cunningham, A. L., Donaghy, H., Harman, A. N., Kim, M., and Turville, S. G. (2010) Manipulation of dendritic cell function by viruses, *Current opinion in microbiology 13*, 524529.

[35] Bushman, F., Nabel, G. J., and Swanstrom, R. (2012) *HIV: from biology to prevention and treatment*, Cold Spring Harbor Laboratory Press.

[36] Davenport, M. P., and Petravic, J. (2010) CD8+ T cell control of HIV—a known unknown, *Public Library of Science pathogens 6*, e1000728.

[37] Ray, N. and Doms, R.W., (2006). HIV-1 coreceptors and their inhibitors. In Chemokines and Viral Infection (pp. 97-120). Springer Berlin Heidelberg.

[38] Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., and Berger,

E. A. (1996) CC CKR5: a RANTES, MIP-1α, MIP-1β receptor as a fusion cofactor for macrophage-tropic HIV-1, *Science*, 1955-1958.

[39] Du Toit, A. (2014) The many faces of the HIV-1 spike, *Nature Reviews Microbiology 12*.

[40] Simon, V., Ho, D. D., and Karim, Q. A. (2006) HIV/AIDS epidemiology, pathogenesis, prevention, and treatment, *The Lancet 368*, 489-504.

[41] Schiller, J., and Chackerian, B. (2014) Why HIV virions have low numbers of envelope spikes: implications for vaccine development, *Public Library of Science pathogens 10*, e1004254.

[42] Bachrach, E., Dreja, H., Lin, Y.-L., Mettling, C., Pinet, V., Corbeau, P., and Piechaczyk, M. (2005) Effects of virion surface gp120 density on infection by HIV-1 and viral production by infected cells, *Virology 332*, 418-429.

[43] Korkut, A., and Hendrickson, W. A. (2012) Structural plasticity and conformational transitions of HIV envelope glycoprotein gp120, *Public Library of Science one 7*, e52170.

[44] Eckert, D. M., and Kim, P. S. (2001) Mechanisms of viral membrane fusion and its inhibition, *Annual review of biochemistry 70*, 777-810.

[45] Coffin, J.M., Hughes, S.H. and Varmus, H.E., (1997). Retroviruses, Cold Spring Habour Laboratory Press, Plainview (NY)., DK; Gollackner, B.; Knosalla, C. and Ter how far have we come. *Transplant Immunology*, 9(2-4), pp.251-6.

[46] Makatini, M. M., Petzold, K., Sriharsha, S. N., Soliman, M. E., Honarparvar, B., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E., and Kruger, H. G. (2011) Pentacycloundecane-based inhibitors of wild-type C-South African HIV-protease, *Bioorganic & medicinal chemistry letters 21*, 2274-2277.

[47] Kessl, J. J., McKee, C. J., Eidahl, J. O., Shkriabai, N., Katz, A., and Kvaratskhelia, M. (2009) HIV1 integrase-DNA recognition mechanisms, *Viruses 1*, 713-736.

[48] Sundquist, W. I., and Kräusslich, H.-G. (2012) HIV-1 assembly, budding, and maturation, *Cold Spring Harbor perspectives in medicine 2*, a006924.

[49] Freed, E. O. (2015) HIV-1 assembly, release and maturation, *Nature Reviews. Microbiology 13*, 484.

[50] Parboosing, R., Maguire, G. E., Govender, P., and Kruger, H. G. (2012) Nanotechnology and the treatment of HIV infection, *Viruses 4*, 488-520.

[51] Hoffmann C, R. J. (2010) Goals and principles of therapy., In *Medizin Fokus Verlag, Harmburg*, 234-254.

[52] Monini, P., Sgadari, C., Toschi, E., Barillari, G., and Ensoli, B. (2004) Antitumour effects of antiretroviral therapy, *Nature Reviews Cancer 4*, 861-875.

[53] Mailler, E., Bernacchi, S., Marquet, R., Paillart, J.-C., Vivet-Boudou, V., and Smyth, R. P. (2016) The Life-Cycle of the HIV-1 Gag–RNA Complex, *Viruses 8*, 248.

[54] Moore, M. D., and Hu, W.-S. (2009) HIV-1 RNA dimerization: It takes two to tango, *AIDS reviews 11*, 91.

[55] Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair, *Nature Reviews Microbiology 2*, 461-472.

[56] Rodríguez-Barrios, F., and Gago, F. (2004) HIV protease inhibition: limited recent progress and advances in understanding current pitfalls, *Current topics in medicinal chemistry 4*, 991-1007.

[57] Pomerantz, R. J., and Horn, D. L. (2003) Twenty years of therapy for HIV-1 infection, *Nature medicine 9*, 867-873.

[58] Arts, E. J., and Hazuda, D. J. (2012) HIV-1 antiretroviral drug therapy, *Cold Spring Harbor perspectives in medicine 2*, a007161.

[59] Yeni, P. G., Hammer, S. M., Hirsch, M. S., Saag, M. S., Schechter, M., Carpenter, C. C., Fischl, M. A., Gatell, J. M., Gazzard, B. G., and Jacobsen, D. M. (2004) Treatment for adult HIV infection: 2004 recommendations of the International AIDS Society-USA Panel, *Jama 292(2)*, 251-265.

[60] Monforte, A. d. A., Lepri, A. C., Rezza, G., Pezzotti, P., Antinori, A., Phillips, A. N., Angarano, G., Colangeli, V., De Luca, A., and Ippolito, G. (2000) Insights into the reasons for discontinuation of the first highly active antiretroviral therapy (HAART) regimen in a cohort of antiretroviral naive patients, *Aids 14*, 499-507.

[61] Sierra, S., Kupfer, B., and Kaiser, R. (2005) Basics of the virology of HIV-1 and its replication, *Journal of clinical virology 34*, 233-244.

[62] Qaqish, R. B., Fisher, E., Rublein, J., and Wohl, D. A. (2000) HIV-Associated Lipodystrophy Syndrome, *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy 20*, 13-22.

[63] Duran, S., Savès, M., Spire, B., Cailleton, V., Sobel, A., Carrieri, P., Salmon, D., Moatti, J.-P., Leport, C., and Group, A. S. (2001) Failure to maintain long-term adherence to highly active antiretroviral therapy: the role of lipodystrophy, *Aids 15*, 2441-2444.

[64] Safrin, S., and Grunfeld, C. (1999) Fat distribution and metabolic changes in patients with HIV infection, *Aids 13*, 2493-2505.

[65] Kempf, D. J., Sham, H. L., Marsh, K. C., Flentge, C. A., Betebenner, D., Green, B. E., McDonald, E., Vasavanonda, S., Saldivar, A., and Wideburg, N. E. (1998) Discovery of ritonavir, a potent inhibitor of HIV protease with high oral bioavailability and clinical efficacy, *Journal of medicinal chemistry 41*, 602-617.

[66] Fultz, P. N., McClure, H. M., Anderson, D. C., Swenson, R. B., Anand, R., and Srinivasan, A. (1986) Isolation of a T-lymphotropic retrovirus from naturally infected sooty mangabey monkeys (Cercocebus atys), *Proceedings of the National Academy of Sciences 83*, 5286-5290.

[67] Richman, D. D. (2001) HIV chemotherapy, *Nature 410*, 995.

[68] Benson, C. A. (1995) Viral Dynamics in Human Immunodeficiency Virus Type 1 Infection, *Infectious Diseases in Clinical Practice 4*, 180-181.

[69] Hammer, S. M., Squires, K. E., Hughes, M. D., Grimes, J. M., Demeter, L. M., Currier, J. S., Eron Jr, J. J., Feinberg, J. E., Balfour Jr, H. H., and Deyton, L. R. (1997) A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less, *New England Journal of Medicine 337*, 725-733.

[70] Hirsch, M., Steigbigel, R., Staszewski, S., Mellors, J., Scerpella, E., Hirschel, B., Lange, J., Squires, K., Rawlins, S., and Meibohm, A. (1999) A randomized, controlled trial of indinavir, zidovudine, and lamivudine in adults with advanced human immunodeficiency virus type 1 infection and prior antiretroviral therapy, *The Journal of infectious diseases 180*, 659-665.

[71] Lalezari, J. P., Henry, K., O'hearn, M., Montaner, J. S., Piliero, P. J., Trottier, B., Walmsley, S., Cohen, C., Kuritzkes, D. R., and Eron Jr, J. J. (2003) Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America, *New England journal of medicine 348*, 2175-2185.

[72] Furman, P., and Barry, D. (1988) Spectrum of antiviral activity and mechanism of action of zidovudine. An overview, *The American journal of medicine 85*, 176-181.

[73] Kilby, J. M., Hopkins, S., Venetta, T. M., DiMassimo, B., Cloud, G. A., Lee, J. Y., Alldredge, L., Hunter, E., Lambert, D., and Bolognesi, D. (1998) Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry, *Nature medicine 4*.

[74] Dybul, M., Fauci, A. S., Bartlett, J. G., Kaplan, J. E., and Pau, A. K. (2002) Guidelines for using antiretroviral agents among HIV-infected adults and adolescents: the panel on clinical practices for treatment of HIV, *Annals of internal medicine 137*, 381-433.

[75] Wlodawer, A. (2002) Rational approach to AIDS drug design through structural biology, *Annual review of medicine 53*, 595-614.

[76] Weber, I. T., and Agniswamy, J. (2009) HIV-1 protease: structural perspectives on drug resistance, *Viruses 1*, 1110-1136.

[77] Castro, H. C., Abreu, P. A., Geraldo, R. B., Martins, R. C., dos Santos, R., Loureiro, N. I., Cabral, L. M., and Rodrigues, C. R. (2011) Looking at the proteases from a simple perspective, *Journal of Molecular Recognition 24*, 165-181.

[78] Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, *Nature 337*, 615-620.

[79] Wlodawer, A., Miller, M., JASK6LsKi, M., Sathyanarayana, B. K., BALDWIN, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. (1989) Crystal structure of synthetic HIV-protease, *Science 245*, 6I6.

[80] Lee, T., Laco, G. S., Torbett, B. E., Fox, H. S., Lerner, D. L., Elder, J. H., and Wong, C.-H. (1998) Analysis of the S3 and S3′ subsite specificities of feline immunodeficiency virus (FIV) protease: development of a broad-based protease inhibitor efficacious against FIV, SIV, and HIV in vitro and ex vivo, *Proceedings of the National Academy of Sciences 95*, 939-944.

[81] Pokorná, J., Machala, L., Řezáčová, P., and Konvalinka, J. (2009) Current and novel inhibitors of HIV protease, *Viruses 1*, 1209-1239.

[82] Kaplan, A. H., Zack, J. A., Knigge, M., Paul, D. A., Kempf, D. J., Norbeck, D. W., and Swanstrom, R. (1993) Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles, *Journal of virology 67*, 4050-4055.

[83] Lv, Z., Chu, Y., and Wang, Y. (2015) HIV protease inhibitors: a review of molecular selectivity and toxicity, *HIV/AIDS (Auckland, NZ) 7*, 95-104.

[84] Turk, B. (2006) Targeting proteases: successes, failures and future prospects, *Nature reviews. Drug discovery 5(9)*, 785-799.

[85] Cuccioloni, M., Mozzicafreddo, M., Bonfili, L., Cecarini, V., Eleuteri, A. M., and Angeletti, M. (2009) Natural occurring polyphenols as template for drug design. Focus on serine proteases, *Chemical biology & drug design 74*, 1-15.

[86] Hsu, A., Granneman, G. R., Cao, G., Carothers, L., El-Shourbagy, T., Baroldi, P., Erdman, K., Brown, F., Sun, E., and Leonard, J. M. (1998) Pharmacokinetic interactions between two human

immunodeficiency virus protease inhibitors, ritonavir and saquinavir, *Clinical Pharmacology & Therapeutics 63*, 453-464.

[87] Flexner, C. (2007) HIV drug development: the next 25 years, *Nature reviews. Drug discovery 6*, 959-966.

[88] Zeldin, R. K., and Petruschke, R. A. (2004) Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients, *Journal of antimicrobial chemotherapy 53*, 4-9.

[89] Kurup, A., Mekapati, S. B., Garg, R., and Hansch, C. (2003) HIV-1 protease inhibitors: a comparative QSAR analysis, *Current medicinal chemistry 10*, 1679-1688.

[90] Viraben, R., and Aquilina, C. (1998) Indinavir-associated lipodystrophy, *Aids 12*, F37-F39.

[91] Perez, M., Fernandes, P., and Ramos, M. (2007) Drug design: new inhibitors for HIV-1 protease based on Nelfinavir as lead, *Journal of Molecular Graphics and Modelling 26*, 634-642.

[92] Adkins, J. C., and Faulds, D. (1998) Amprenavir, *Drugs 55*, 837-842.

[93] Chapman, T. M., Plosker, G. L., and Perry, C. M. (2004) Fosamprenavir, *Drugs 64*, 2101-2124.

[94] Becker, S. (2003) Atazanavir: improving the HIV protease inhibitor class, *Expert review of antiinfective therapy 1*, 403-413.

[95] Fulco, P. P., Vora, U. B., and Bearman, G. M. (2006) Acid suppressive therapy and the effects on protease inhibitors, *Annals of Pharmacotherapy 40*, 1974-1983.

[96] Squires, K., Lazzarin, A., Gatell, J. M., Powderly, W. G., Pokrovskiy, V., Delfraissy, J.-F., Jemsek, J., Rivero, A., Rozenbaum, W., and Schrader, S. (2004) Comparison of once-daily atazanavir with efavirenz, each in combination with fixed-dose zidovudine and lamivudine, as initial therapy for patients infected with HIV, *JAIDS Journal of Acquired Immune Deficiency Syndromes 36*, 1011-1019.

[97] Murphy, R. L., Sanne, I., Cahn, P., Phanuphak, P., Percival, L., Kelleher, T., and Giordano, M. (2003) Dose-ranging, randomized, clinical trial of atazanavir with lamivudine and stavudine in antiretroviral-naive subjects: 48-week results, *Aids 17*, 2603-2614.

[98] Larder, B. A., Hertogs, K., Bloor, S., Van den Eynde, C., DeCian, W., Wang, Y., Freimuth, W. W., and Tarpley, G. (2000) Tipranavir inhibits broadly protease inhibitor-resistant HIV-1 clinical samples, *Aids 14*, 1943-1948.

[99] Kashman, Y., Gustafson, K. R., Fuller, R., Cardellina 2nd, J., McMahon, J., Currens, M., Buckheit Jr, R., Hughes, S., Cragg, G., and Boyd, M. (1992) The calanolides, a novel HIV-inhibitory class of coumarin derivatives from the tropical rainforest tree, Calophyllum lanigerum, *Journal of medicinal chemistry 35*, 2735-2743.

[100] Lefebvre, E., and Schiffer, C. A. (2008) Resilience to resistance of HIV-1 protease inhibitors: profile of darunavir, *AIDS reviews 10*, 131-142.

[101] King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., Wigerinck, P., de Béthune, M.-P., and Schiffer, C. A. (2004) Structural and thermodynamic basis for the binding of TMC114, a nextgeneration human immunodeficiency virus type 1 protease inhibitor, *Journal of virology 78*, 12012-12021.

[102] Dierynck, I., Keuleers, I., De Wit, M., Tahri, A., Surleraux, D., Peeters, A., and Hertogs, K. (2005) Kinetic characterization of the potent activity of TMC114 on wild-type HIV-1 protease, In *14th International HIV Drug Resistance Workshop (Vol. 64)*.

[103] Surleraux, D. L., De Kock, H. A., Verschueren, W. G., Pille, G. M., Maes, L. J., Peeters, A., Vendeville, S., De Meyer, S., Azijn, H., and Pauwels, R. (2005) Design of HIV-1 protease inhibitors active on multidrug-resistant virus, *Journal of medicinal chemistry 48*, 1965-1973.

[104] Tie, Y., Boross, P. I., Wang, Y.-F., Gaddis, L., Hussain, A. K., Leshchenko, S., Ghosh, A. K., Louis, J. M., Harrison, R. W., and Weber, I. T. (2004) High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drugresistant clinical strains, *Journal of molecular biology 338*, 341-352.

[105] Molina, J.-M., Cohen, C., Katlama, C., Grinsztejn, B., Timerman, A., de Jesus Pedro, R., Vangeneugden, T., Miralles, D., De Meyer, S., and Parys, W. (2007) Safety and efficacy of darunavir (TMC114) with low-dose ritonavir in treatment-experienced patients: 24-week results of POWER 3, *JAIDS Journal of Acquired Immune Deficiency Syndromes 46*, 24-31.

[106] Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., and Schiffer, C. A. (2003) Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: structural insights for better antiviral therapy, *Journal of virology 77*, 1306-1315.

[107] King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004) Combating susceptibility to drug resistance: lessons from HIV-1 protease, *Chemistry & biology 11*, 13331338.

[108] Ohtaka, H., Velázquez-Campoy, A., Xie, D., and Freire, E. (2002) Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wildtype and drug-resistant mutants of the HIV-1 protease, *Protein science 11*, 1908-1916.

[109] Martinez-Cajas, J. L., and Wainberg, M. A. (2007) Protease inhibitor resistance in HIV-infected patients: molecular and clinical perspectives, *Antiviral research 76*, 203-221.

[110] Yoshinaga, T., Sato, A., Fujishita, T., and Fujiwara, T. (2002) S-1360: in vitro activity of a new HIV-1 integrase inhibitor in clinical development, In *9th conference on retroviruses and opportunistic infections, Seattle, WA*, p 55.

[111] Young, S. (2002) Discovery of a potent antiviral HIV integrase inhibitor with potential clinical utility, In *ANTIVIRAL THERAPY*, pp S4-S4, INT MEDICAL PRESS LTD 2-4 IDOL LANE, LONDON EC3R 5DD, ENGLAND.

[112] Barbaro, G., Scozzafava, A., Mastrolorenzo, A., and Supuran, C. T. (2005) Highly active antiretroviral therapy: current state of the art, new agents and their pharmacological interactions useful for improving therapeutic outcome, *Current pharmaceutical design 11*, 1805-1843.

[113] Trkola, A., Ketas, T. J., Nagashima, K. A., Zhao, L., Cilliers, T., Morris, L., Moore, J. P., Maddon, P. J., and Olson, W. C. (2001) Potent, broad-spectrum inhibition of human immunodeficiency virus type 1 by the CCR5 monoclonal antibody PRO 140, *Journal of virology 75*, 579-588.

[114] Hazuda, D. (2002) A novel HIV-1 integrase inhibitor mediates sustained suppression of viral replication and CD4 depletion in a SHIV rhesus macaque model of infection, In *ANTIVIRAL THERAPY*, pp S3-S3, INT MEDICAL PRESS LTD 2-4 IDOL LANE, LONDON EC3R 5DD, ENGLAND.

[115] Zhu, P., Olson, W. C., and Roux, K. H. (2001) Structural flexibility and functional valence of CD4-IgG2 (PRO 542): potential for cross-linking human immunodeficiency virus type 1 envelope spikes, *Journal of virology 75*, 6682-6686.

[116] Lin, P., Guo, K., Fridell, R., Ho, H., Yamanaka, G., and Colonno, R. (2002) Identification and characterization of a novel inhibitor of HIV-1 entry. II. Mechanism of action, In *Programs and abstracts of the Ninth Conference on Retroviruses and Opportunistic Infections, Seattle*.

[117] Kazmierski, W., Bifulco, N., Yang, H., Boone, L., DeAnda, F., Watson, C., and Kenakin, T. (2003) Recent progress in discovery of small-molecule CCR5 chemokine receptor ligands as HIV-1 inhibitors, *Bioorganic & medicinal chemistry 11*, 2663-2676.

[118] O'Hara, B. M., and Olson, W. C. (2002) HIV entry inhibitors in clinical development, *Current opinion in pharmacology 2*, 523-528.

[119] Kilby, J. M., and Eron, J. J. (2003) Novel therapies based on mechanisms of HIV-1 cell entry, *New England Journal of Medicine 348*, 2228-2238.

[120] Greenberg, M., Cammack, N., Salgo, M., and Smiley, L. (2004) HIV fusion and its inhibition in antiretroviral therapy, *Reviews in medical virology 14*, 321-337.

[121] Goel, A., Mazur, S. J., Fattah, R. J., Hartman, T. L., Turpin, J. A., Huang, M., Rice, W. G., Appella, E., and Inman, J. K. (2002) Benzamide-based thiolcarbamates: a new class of HIV-1 NCp7 inhibitors, *Bioorganic & medicinal chemistry letters 12*, 767-770.

[122] Yu, D., Wild, C. T., Martin, D. E., Morris-Natschke, S. L., Chen, C.-H., Allaway, G. P., and Lee, K.-H. (2005) The discovery of a class of novel HIV-1 maturation inhibitors and their potential in the therapy of HIV, *Expert opinion on investigational drugs 14*, 681-693.

[123] Davies, D. R. (1990) The structure and function of the aspartic proteinases, *Annual review of biophysics and biophysical chemistry 19*, 189-215.

[124] Shafer, R. W., Dupnik, K., Winters, M. A., and Eshleman, S. H. (2001) A guide to HIV-1 reverse transcriptase and protease sequencing for drug resistance studies, *HIV sequence compendium 2001*, 1-51.

[125] Oda, K. (2011) New families of carboxyl peptidases: serine-carboxyl peptidases and glutamic peptidases, *The Journal of Biochemistry 151*, 13-25.

[126] Scott, W. R., and Schiffer, C. A. (2000) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance, *Structure 8*, 1259-1265.

[127] Ishima, R., Ghirlando, R., Tözsér, J., Gronenborn, A. M., Torchia, D. A., and Louis, J. M. (2001) Folded monomer of HIV-1 protease, *Journal of Biological Chemistry 276*, 49110-49116.

[128] Ishima, R., Torchia, D. A., Lynch, S. M., Gronenborn, A. M., and Louis, J. M. (2003) Solution Structure of the Mature HIV-1 Protease Monomer Insight into the tertiary fold and stability of a precursor, *Journal of Biological Chemistry 278*, 43311-43319.

[129] Rose, R. B., Craik, C. S., and Stroud, R. M. (1998) Domain flexibility in retroviral proteases: structural implications for drug resistant mutations, *Biochemistry 37*, 2607-2621.

[130] Deng, N.-j., Zheng, W., Gallicchio, E., and 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*, 9387-9394.

[131] Naicker, P., Achilonu, I., Fanucchi, S., Fernandes, M., Ibrahim, M. A., Dirr, H. W., Soliman, M. E., and Sayed, Y. (2013) Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance, *Journal of Biomolecular Structure and Dynamics 31*, 1370-1380.

[132] Todd, M. J., Semo, N., and Freire, E. (1998) The structural stability of the HIV-1 protease, *Journal of Molecular Biology 283*, 475-488.

[133] Szeltner, Z., and Polgár, L. (1996) Conformational stability and catalytic activity of HIV-1 protease are both enhanced at high salt concentration, *Journal of Biological Chemistry 271*, 5458-5463.

[134] Todd, M. J., and Freire, E. (1999) The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease, *Proteins: Structure, Function, and Bioinformatics 36*, 147156.

[135] Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998) Molecular and biotechnological aspects of microbial proteases, *Microbiology and molecular biology reviews 62*, 597-635.

[136] Piana, S., Carloni, P., and Parrinello, M. (2002) Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease, *Journal of molecular biology 319*, 567-583.

[137] Naicker, P., 2015. Energetic, structural and dynamic evaluation of HIV-1 proteases (Doctoral dissertation).

[138] Ishima, R., Freedberg, D. I., Wang, Y.-X., Louis, J. M., and Torchia, D. A. (1999) Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function, *Structure 7*, 1047-1055.

[139] Freedberg, D. I., Ishima, R., Jacob, J., Wang, Y. X., Kustanovich, I., Louis, J. M., and Torchia, D. A. (2002) Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations, *Protein Science 11*, 221-232.

[140] Chang, C.-e. A., Huang, Y.-m. M., Mueller, L. J., and You, W. (2016) Investigation of Structural Dynamics of Enzymes and Protonation States of Substrates Using Computational Tools, *Catalysts 6*, 82-103.

[141] Silva, A. M., Cachau, R. E., Sham, H. L., and Erickson, J. W. (1996) Inhibition and catalytic mechanism of HIV-1 aspartic protease, *Journal of molecular biology 255*, 321-340.

[142] Parris, K. D., Hoover, D. J., Damon, D. B., and Davies, D. R. (1992) Synthesis and crystallographic analysis of two rhizopuspepsin inhibitor complexes, *Biochemistry 31*, 81258141.

[143] Velazquez-Campoy, A., Todd, M. J., Vega, S., and Freire, E. (2001) Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes, *Proceedings of the National Academy of Sciences 98*, 6062-6067.

[144] Velazquez-Campoy, A., Vega, S., and Freire, E. (2002) Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes, *Biochemistry 41*, 8613-8619.

[145] Maseko, S. B., Padayachee, E., Govender, T., Sayed, Y., Kruger, G., Maguire, G. E., and Lin, J. (2017) I36T↑ T mutation in South African subtype C (C-SA) HIV-1 protease significantly alters protease-drug interactions, *Biological Chemistry*. 398 (10), 1109-1117.

[146] Spinelli, S., Liu, Q., Alzari, P., Hirel, P., and Poljak, R. (1991) The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU, *Biochimie 73*, 1391-1396.

[147] Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., and Geoghegan, K. F. (1989) X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes, *Nature 342*, 299-302.

[148] Louis, J. M., Dyda, F., Nashed, N. T., Kimmel, A. R., and Davies, D. R. (1998) Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease, *Biochemistry 37*, 2105-2110.

[149] Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2000) How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease, *Journal of molecular biology 301*, 1207-1220.

[150] Galiano, L., Bonora, M., and Fanucci, G. E. (2007) Interflap distances in HIV-1 protease determined by pulsed EPR measurements, *Journal of the American Chemical Society 129*, 11004-11005.

[151] Torbeev, V. Y., Raghuraman, H., Hamelberg, D., Tonelli, M., Westler, W. M., Perozo, E., and Kent, S. B. (2011) Protein conformational dynamics in the mechanism of HIV-1 protease catalysis, *Proceedings of the National Academy of Sciences 108*, 20982-20987.

[152] Torbeev, V. Y., Raghuraman, H., Mandal, K., Senapati, S., Perozo, E., and Kent, S. B. (2008) Dynamics of "flap" structures in three HIV-1 protease/inhibitor complexes probed by total chemical synthesis and pulse-EPR spectroscopy, *Journal of the American Chemical Society 131*, 884-885.

[153] Ahmed, S. M., Kruger, H. G., Govender, T., Maguire, G. E., Sayed, Y., Ibrahim, M. A., Naicker, P., and Soliman, M. E. (2013) Comparison of the Molecular Dynamics and Calculated Binding Free Energies for Nine FDA-Approved HIV-1 PR Drugs Against Subtype B and C-SA HIV PR, *Chemical biology & drug design 81*, 208-218.

[154] Lockhat, H. A., Silva, J. R., Alves, C. N., Govender, T., Lameira, J., Maguire, G. E., Sayed, Y., and Kruger, H. G. (2016) Binding Free Energy Calculations of Nine FDA-approved Protease Inhibitors Against HIV-1 Subtype C I36T↑ T Containing 100 Amino Acids Per Monomer, *Chemical biology & drug design 87*, 487-498.

[155] Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations, *Proceedings of the National Academy of Sciences of the United States of America 103*, 915-920.

[156] Tóth, G., and Borics, A. (2006) Flap opening mechanism of HIV-1 protease, *Journal of Molecular Graphics and Modelling 24*, 465-474.

[157] Mahanti, M., Bhakat, S., Nilsson, U. J., and Söderhjelm, P. (2016) Flap dynamics in aspartic proteases: a computational perspective, *Chemical biology & drug design 88*, 159-177.

[158] Collins, J. R., Burt, S. K., and Erickson, J. W. (1995) Flap opening in HIV-1 protease simulated by 'activated'molecular dynamics, *Nature Structural & Molecular Biology 2*, 334-338.

[159] Meagher, K. L., and Carlson, H. A. (2005) Solvation influences flap collapse in HIV-1 protease, *Proteins: Structure, Function, and Bioinformatics 58*, 119-125.

[160] Karubiu, W., Bhakat, S., McGillewie, L., and Soliman, M. E. (2015) Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics, *Molecular BioSystems 11*, 1061-1066.

[161] McGillewie, L., and Soliman, M. E. (2015) Flap flexibility amongst plasmepsins I, II, III, IV, and V: Sequence, structural, and molecular dynamics analyses, *Proteins: Structure, Function, and Bioinformatics 83*, 1693-1705.

[162] Kumalo, H., and Soliman, M. E. (2016) A comparative molecular dynamics study on BACE1 and BACE2 flap flexibility, *Journal of Receptors and Signal Transduction 36*, 505-514.

[163] Fujisaki, H., Moritsugu, K., Matsunaga, Y., Morishita, T., and Maragliano, L. (2015) Extended phase-space methods for enhanced sampling in molecular simulations: a review, *Frontiers in bioengineering and biotechnology 3, 125*.

[164] Steinbrecher, T., and Elstner, M. (2013) QM and QM/MM simulations of proteins, *Biomolecular Simulations: Methods and Protocols*, 91-124.

[165] Paquet, E., and Viktor, H. L. (2015) Molecular dynamics, Monte Carlo simulations, and Langevin dynamics: a computational review, *BioMed research international 2015*, 1-18.

[166] Schlick, T. (2009) Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules, *F1000 biology reports 1-51*.

[167] Abrams, C., and Bussi, G. (2013) Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration, *Entropy 16*, 163-199.

[168] Foresman, J. B., and Frisch, Æ. (1996) Exploring chemistry with electronic structure methods: a guide to using Gaussian. 4-5.

[169] Downs, G. (2004) Molecular descriptors, Marcel Dekker: New York. 515-538,

[170] Rogers, D. W. (2003) *Computational chemistry using the PC*, John Wiley & Sons. 93-100.

[171] Sharma, P., 2011. An assessment of the conformational profile of bombesin and its mammalian analogues using computational chemistry methods (Doctoral dissertation).

[172] Sansom, C. E., and Smith, C. A. (1998) Computer applications in the biomolecular sciences. Part 1: molecular modelling, *Biochemistry and Molecular Biology Education 26*, 103-110.

[173] Bizzarri, A. R., and Cannistraro, S. (2002) Molecular dynamics of water at the protein− solvent interface, ACS Publications. 106 (26), 6617-6633.

[174] Ayers, P., Yang, W., Bultinck, P., de Winter, H., Langenaeker, W., and Tollenaere, J. (2003) Computational Medicinal Chemistry for Drug Discovery, *Bultinck, P*, 571.

[175] Wang, W., Donini, O., Reyes, C. M., and Kollman, P. A. (2001) Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein,
and protein-nucleic acid noncovalent interactions, *Annual review of biophysics and biomolecular structure 30*, 211-243.

[176] Keseru, G. M., and Kolossváry, I. (1999) *Molecular Mechanics and Conformational Analysis in Drug Design*, Wiley-Blackwell. 357.

[177] Dodson, G. G., Lane, D. P., and Verma, C. S. (2008) Molecular simulations of protein dynamics: new windows on mechanisms in biology, *EMBO reports 9*, 144-150.

[178] Karplus, M., and Kuriyan, J. (2005) Molecular dynamics and protein function, *Proceedings of the National Academy of Sciences of the United States of America 102*, 6679-6685.

[179] Anandakrishnan, R., Drozdetski, A., Walker, R. C., and Onufriev, A. V. (2015) Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations, *Biophysical journal 108*, 1153-1164.

[180] Baker, N. A. (2004) Poisson–Boltzmann methods for biomolecular electrostatics, *Methods in enzymology 383*, 94-118.

[181] Dong, F., Wagoner, J. A., and Baker, N. A. (2008) Assessing the performance of implicit solvation models at a nucleic acid surface, *Physical Chemistry Chemical Physics 10*, 48894902.

[182] Tsui, V., and Case, D. A. (2000) Molecular dynamics simulations of nucleic acids with a generalized Born solvation model, *Journal of the American Chemical Society 122*, 2489-2498.

[183] Feig, M. (2007) Kinetics from implicit solvent simulations of biomolecules as a function of viscosity, *Journal of chemical theory and computation 3*, 1734-1748.

[184] Amaro, R. E., Cheng, X., Ivanov, I., Xu, D., and McCammon, J. A. (2009) Characterizing loop dynamics and ligand recognition in human-and avian-type influenza neuraminidases via generalized born molecular dynamics and end-point free energy calculations, *Journal of the American Chemical Society 131*, 4702-4709.

[185] Feig, M., Im, W., and Brooks III, C. L. (2004) Implicit solvation based on generalized Born theory in different dielectric environments, *The Journal of chemical physics 120*, 903-911.

[186] Kamerlin, S. C., Haranczyk, M., and Warshel, A. (2008) Progress in ab initio QM/MM freeenergy simulations of electrostatic energies in proteins: accelerated QM/MM studies of p K a, redox reactions and solvation free energies, *The Journal of Physical Chemistry B 113*, 12531272.

[187] Kollman, P. (1993) Free energy calculations: applications to chemical and biochemical phenomena, *Chemical reviews 93*, 2395-2417.

[188] Essex, J. W., Severance, D. L., Tirado-Rives, J., and Jorgensen, W. L. (1997) Monte Carlo simulations for proteins: Binding affinities for trypsin− benzamidine complexes via free-energy perturbations, *The Journal of Physical Chemistry B 101*, 9663-9669.

[189] Gilson, M. K., and Zhou, H.-X. (2007) Calculation of protein-ligand binding affinities, *Annual review of biophysics and biomolecular structure 36*. 21-42.

CHAPTER TWO

Exploring the flap dynamics of the South African subtype C HIV protease in presence of FDA-approved inhibitors: MD study

Siyabonga I. Maphumulo,[1] Amit K. Halder,[1] Thavendran Govender[1], Sibusiso Maseko[1] Glenn E. M. Maguire,[1,2] Bahareh Honarparvar,[1,*] and Hendrik G. Kruger,[1,*]

[1]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4001, South Africa.

[2] School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4001, South Africa.

***Corresponding authors:** kruger@ukzn.ac.za (Prof. Hendrik G. Kruger), Telephone: + 27 31 2601845, Fax: +27 31 2603091, Honarparvar@ukzn.ac.za (Dr Bahareh Honarparvar), Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.

Abstract

HIV-1 protease (HIV PR) is considered one of the most attractive targets for the treatment of HIV and the impact of flap dynamics of HIV PR on the binding affinities of protease inhibitors (PIs) is one of the crucial ongoing research field. Recently, our research group evaluated the binding affinities of different FDA approved PIs against the South African HIV-1 subtype C (C-SA) protease (PR). The CSA- HIV PR displayed weaker binding affinity for most of the clinical PIs compared to HIV-1 B subtype for West and Central Europe, the Americas. In the current work, the flap dynamics of four different systems of HIV-1 C-SA PR in complex with

FDA approved second generation PIs and its impact on binding was explored over the molecular dynamic trajectories. It was observed that the interactions of the selected drugs with the binding site residues of the protease may not be the major contributor for affinity towards PIs. Various post-MD analyses were performed, also entropic contribution, solvation free energy and hydrophobic core formation interactions were studied to assess how the flap dynamics of C-SA PR is affected by such factors. From these contributions, large van der Waals interactions and low solvation free energies were found to be the major factors for the higher activity of ATV against C-SA HIV PR. Furthermore, a comparatively stable hydrophobic core may be responsible for higher stability of the PR flaps of the ATV complex. The outcome of this study provides significant guidance to how the flap dynamics of C-SA PR is affected by various factors as a results of the binding affinity of various protease inhibitors. It can also leads to the design of potent inhibitors against C-SA HIV PR that apart from binding in the active site of PR can interacts with the flaps to prevent opening of the flaps resulting in inactivation of the protease.

Keywords: HIV-1 protease (HIV PR); South African subtype C (C-SA HIV PR); Molecular dynamics (MD); Flap dynamics.

2.1 Introduction

Acquired Immunodeficiency Syndrome (AIDS) is regarded as one of the most devastating disease outbreaks to date (1-4). Human Immunodeficiency Virus (HIV) is the causative agent for AIDS and approximately 36.7 million people were reported to be suffering from AIDS at the end of 2016 (4). Sub-Saharan Africa faces the most devastating HIV and AIDS epidemic in the world. Of global infections recorded, sub-Saharan African represents 56% with nearly 25 million living with HIV. In the year 2015, 960 000 new HIV infections were experienced by countries within the east and southern Africa regions and close to 470 000 deaths associated

with HIV and AIDS were recorded in the same year (5). Within the continent of Africa, South Africa has been shown to be at the epicentre of the epidemic as an estimated 6.1 million people

(12% of total population) are living with HIV (6). Among different subtypes of HIV-1, subtype C (C-SA), is mostly prevalent in this region and this subtype accounts for approximately 50% of all HIV-1 infections globally (3). C-SA protease has eight point mutations compared to subtype B and these are T12S, I15V, L19I, M36I, R41K, H69K, L89M, and I93L (7). In subtype B, the hinge and cantilever regions (loops) slide over I15 and this interaction with the hydrophobic core allows anchoring of the flaps(7). In C-SA protease the hydrophobic interactions are affected due to polymorphisms observed in I15V, L19I, M36I,H69K, L89M and I93L (7).

The fulcrum, hinge and cantilever regions consist of six mutations (T12S, I15V, L19I, M36I, R41K and H69K) that distinguish C-SA from subtype B (3, 7). The I15V is a part of the hydrophobic core and the L19I mutation is believed to contribute to the rearrangement of hydrophobic interactions in the hydrophobic core (7). This core has 19 amino acid residues with hydrophobic side chains (5L, 11V, 13I, 15I, 22A, 24L, 33L, 36M, 38L, 62I, 64I, 66I, 75V, 77V, 85I, 89L, 90L, 93I and 97L) found in each monomer of HIV-1 protease. These amino acid residues exchange hydrophobic interactions which are essential to maintaining conformational flexibility of the proteases including flap dynamics (7, 8). Furthermore, four of these residues are mutated in C-SA protease and these are I15V, M36I and L89M and I93L. In the C-SA HIV-1 protease, these polymorphisms lead to perturbed hydrophobic sliding that are likely to be one of the major reasons for drug resistance (7, 9).

The HIV-1 protease has been a prime target for drug design because of its indispensable role in viral maturation, which ultimately leads to final morphological rearrangement (1, 10, 11). HIV-1 protease is generally a homodimer with two identical monomers (chain A and chain B)

each consisting of 99 amino acids (12). However, our group recently reported 100 and 101 amino acid mutant variants of C-SA protease that demonstrated little difference in catalytic activity with respect to the natural substrate (2, 13). The first mutant has one extra amino acid

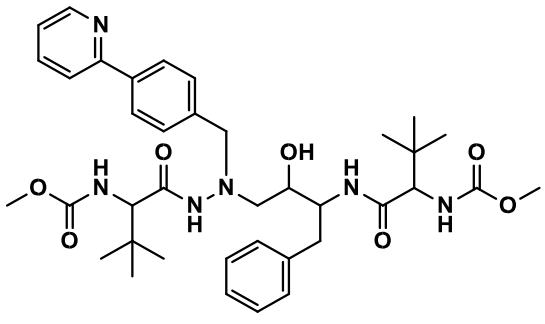
(I36T↑T) and the second mutant has two additional amino acids (L38L↑N↑L) in each monomer (13). Limited data from another study, reported by our group (14) demonstrated that FDA approved protease inhibitors exhibit reduced activities against subtype C because of residue polymorphism.

We have previously reported the theoretical binding affinity of nine FDA approved protease inhibitors (PIs) [amprenavir (APV), atazanavir (ATV), darunavir (DRV), indinavir (INV), lopinavir (LPV), nelfinavir (NFV), ritonavir (RTV), saquanavir (SQV) and tipranavir (TPV)] with subtype B and C-SA on the basis of 2 ns molecular dynamics simulations (14). In addition, we separately reported the theoretical binding affinities of these FDA approved drugs against I36T↑T mutation of subtype C-SA (2). Recently, our group determined the experimental binding affinities of these nine FDA approved drugs against C-SA PR and the mutant (I36T↑T) (15). The binding affinity data (15) (Table 1) indicated that second generation FDA approved PIs (ATV, DRV, LPV and TPV, Figure 1) display a different activity patterns for C-SA protease compared to subtype B. Interestingly, four second generation PIs (ATV, DRV, LPV and TPV) showed considerable variations of binding affinity against C-SA protease. Therefore, the proportion of open conformation of C-SA protease is higher than that of subtype B.

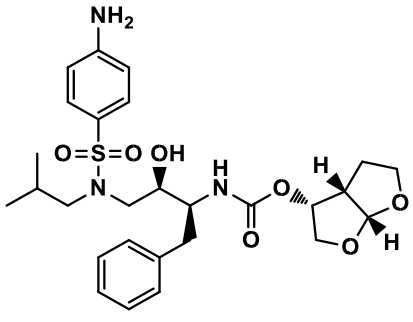
Previously it was reported that overall conformational stability of the C-SA protease is slightly reduced as compared to the subtype B (differential scanning calorimetry experiments) (16). By measuring the magnitude of forces responsible for stabilising the native structure of proteins (17, 18), the authors demonstrated that differences in structural stability could have an influence

in binding affinity of inhibitors. This is as a result of the conformational change that is associated with inhibitor or substrate binding (18). Furthermore, experimental results (7) revealed that the reduced drug susceptibility of C-SA protease may be attributed to altered dynamics around the flaps (hydrophobic core) of the C-SA protease as well as reduced probability of Glu35-Arg57 salt bridge formation. Both of these factors are assumed to have a strong correlation with the region that allows for flap opening and closure (7). Salt bridge formation also affects the conformational stability of the protease. Therefore, apart from binding energy calculations and analyses, it is imperative to understand the flap dynamics and its determinants (such as hydrophobic sliding mechanism and salt bridge formation) in detail to explain the structural requirements of inhibitors against C-SA protease.

Herein, molecular dynamic (MD) simulations, binding free energy calculations and dynamic analyses were combined to understand how the flap dynamics of C-SA protease is affected by binding energy patterns observed in different second-generation FDA approved drugs. It is expected that this work serves as guidance in ensuring that the new potential drugs to be developed do not only have a "strong" binding in the active site but also have interactions with the flaps to prevent opening of the flaps resulting in inactivation of the protease.



Atazanavir (ATV)



Darunavir (DRV)

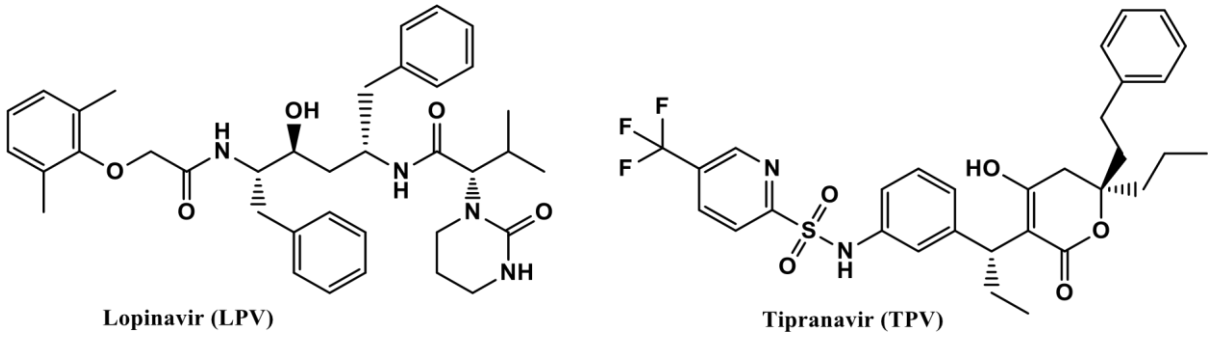


Figure 1. Chemical structures of four of the FDA approved second generation PIs.

Various metrics have been proposed and employed to explore different flap conformational states of HIV-1 PR. The distance between α-carbons of Ile50 and Ile50′ residues is the most commonly used metric parameter (19, 20). Others that were employed are the distance between the α-carbons of Lys55 and Lys55' (21, 22) as well as between those from α-carbons of Ile50 and Asp25 or Ile149 and Asp124 (22, 23). In order to determine the curling behaviour of the flap tip residues, the TriCα angle between Gly49-Ile50-Gly51 for each monomer is measured (23, 24). The flaps of the HIV-1 PR have been shown to be asymmetric in their dynamic motions (25) and the tips of the flaps have been known to display twisting (26). As such, the dihedral angle (Φ) amongst residues ILE50-ASP25-ILE149-ASP124 was taken into account in understanding the flap twisting of HIV-1 PR (27).

2.2 Computational Methods

2.2.1 Preparation of inhibitors/enzyme complex

The crystal structures of the wild-type subtype B HIV-1 protease complexed with four FDA approved drugs were obtained from the Protein Data Bank (PDB). The PDB entries are as follows: 3U71 (apo/free)(28), 3EL1 (ATV)(29), 4DQB (DRV)(9), 1MUI (LPV)(30) and 2O4P (TPV)(31) while the PDB entry for the natural substrate is 1KJH. The C-SA/inhibitor complexes were prepared as described previously (2).

2.2.2 Setting up the system for Molecular Dynamics (MD) simulations

Prior to molecular dynamics simulations, the protonation states of residues of HIV-1 C-SA were assigned at pH=7.0 using the PropKa server (32, 33). The ff99SB (34) and the general AMBER force field (GAFF) (35) were used to describe the inhibitor–protein and inhibitor–solvent interactions, respectively. The ligands were parameterized with Leap (35) implemented in Amber 14, using the general AMBERforce field (GAFF) (35) in Antechamber. MD simulations were performed using Amber 14 with the ff99SB (34) force field with the TIP3P (36) explicit water in a cubic box with 8 Å distance around the complex. The positive charge of the complexes were neutralized with chloride ions. Partial Mesh Ewald (PME) (37) method was used to consider long range electrostatic forces (cutoff 12 Å). Additionally, the SHAKE algorithm (38) was used to constrain all bonds to hydrogen atoms. To reduce the overlapping atom interactions of initially prepared solvated complexes, energy minimization was performed in two stages. In the first stage only ions and water molecules were relaxed by 2000 step minimization process (1000 steps of steepest decent minimization followed by 1000 of conjugated gradient) using a restrained force of 500 kcal/mol on the solute. For the second stage of minimization, the whole system was relaxed by 5000 step minimization process (2500 steps of steepest decent minimization followed by 2500 of conjugated gradient). The minimized systems were gradually heated up from 0 to 300 K with a weak harmonic restraint of 10 kcal/mol to keep the solute fixed for 200 ps. Subsequently, the 2 ns constant pressure equilibration at 300 K were performed. Lastly, 60 ns MD simulations without restriction were run at a constant temperature of 300 K and constant pressure at 1 atm.

In order to validate the results, a set of three different simulations were performed by using different starting structures obtained from 20, 40 and 60 ns snapshots. These structures were retrived from the original 60 ns MD simulation.

2.2.3 Post-dynamics analysis

After the completion of each MD simulation, the analyses for the root mean square deviation (RMSD), root mean square fluctuatuion (RMSF), hydrogen bond analyses as well as distance and angle analyses were performed using PTRAJ and CPPTRAJ module (39) implemented in Amber 14 (Figures 2,3 and Tables 2, 3).

2.2.4 Binding free energy calculations

Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) (40) binding free energies of inhibitors were calculated using MM-PBSA program in Amber 14. For each complex, 400 snapshots were taken from the last 2000 ps of the 60 ns MD trajectory with an interval of 5 ps. In this method the binding energy is represented as:

$$\Delta G = \Delta E_{ele} + \Delta E_{vdW} + \Delta G_{pol} + \Delta G_{nopol} - T\Delta S$$

The first two terms ΔE_{ele} and ΔE_{vdW} are calculated by molecular mechanisms and represents the electrostatic and van der Waals interactions with the proteins in gas phase. The polar solvation free energy ΔG_{pol} is calculated by using the MM-PBSA program (41) and it represents polar interactions with the solvent molecules. The ΔG_{nonpol} signifies non-polar solvation free energy and is obtained from the equation $\Delta G_{nonpolar} = \gamma SASA + \beta$ (42). The solvent accessible surface area (SASA), the surface tension proportionality constant (γ) and the free energy of nonpolar solvation of a point solute (β), were set to 0.00542 kcal mol^{-1} Å^{-2} and 0 kcal mol^{-1}, respectively (43). The entropy contribution to the binding free energies were calculated using normal mode analysis (44-47) for the complexes from changes in the translational, rotational and vibrational entropy components (Table 1).

2.2.5 Per residue free energy decomposition analysis

To evaluate the contributions of individual residues to the total binding free energies of the inhibitors/PR complexes were computed using the MM-GBSA per residue free energy decomposition method (48, 49) in Amber 14 (40). All energy components including van der Waals, electrostatic, polar solvation and nonpolar solvation contributions were calculated using 400 snapshots extracted from the last 2000 ps MD trajectories (Figure 5 and 6) of the 60 ns MD simulation.

2.2.6 Principal component analysis (PCA)

To understand conformational changes of the proteins, PCA analyses were performed using the PTRAJ (39) module of Amber 14. Correlated and anti-correlated motions were studied using the dynamics cross-correlation calculations derived by dynamics trajectories. The crosscorrelation coefficient (c_{ij}) for each pair of alpha carbon (Cα) atoms *i* and *j* were calculated with the following equation:(50) (**1**)

$$c_{ij} = \frac{<\Delta r_i . \Delta r_j>}{(<\Delta r_i^2><\Delta r_j^2>)^{1/2}} \quad (\mathbf{1})$$

The $\Delta r_n$is the displacement of the n^{th}atom mean position. The angle parentheses represent the time average over the entireity of each trajectory. The collective motions were produced using the positional covariance matrix C based on the atomic coordinates and their corresponding eigenvectors. The eigenvalues and eigenvectors stand for extent and direction of motions of atoms respectively (51). The elements of the positional covariance matrix C were computed using the following equation(52):

$$C_i = < (q_i - < q_i >)(q_j - < q_j >) > (i, j = 1,2, \dots ,3N) \quad (\mathbf{2})$$

The Cartesian coordinate Δq_n stands for the $n^{th}C\alpha$ atom, and N is the number of Cα atoms. To remove all translational and rotational movements, the average is calculated after superimposition with a reference structure using a least-square fit procedure to extract the important motion from MD trajectories (53-55). The matrix C is symmetric and is diagonalized by an orthogonal coordinate transformation matrix, which transforms the matrix C into a diagonal matrix of eigenvalues. The eigenvectors corresponds to the direction of motion relative to cartesian coordinate of each atom and each eigenvector is related to an eigenvalue that describes the total mean-square fluctuation of the system along the corresponding eigenvector. The porcupine plot of protein collective motions was created by the NMWiz tool in VMD (56) (Figure 7).

2.3 Results and discussion

2.3.1 MM-GBSA binding energy analyses

To understand the major determinants for the binding affinity of the FDA-approved secondgeneration drugs, MM-GBSA binding free energy analyses were carried out for all four drug-

PR complexes. The calculated binding free energies [$\Delta G_{bind(T)}$] and their components (ΔE_{vdw}, ΔE_{elec}, ΔG_{gas}, ΔG_{polar}, $\Delta G_{nonpolar}$, $\Delta G_{solvation}$, $T\Delta S$) were calculated by MM-GBSA (40) binding free energy method and normal mode analysis (NMA) (44-47) and are presented in Table 1.

Table 1. Calculated binding free energies [$\Delta G_{bind(T)}$] and its components for the four second generation HIV-1 protease inhibitors and the natural substrate (NS) on C-SA PR. The experimental binding energies [$\Delta G_{bind(E)}$] (15) is also depicted. The energy components are in kcal/mol and entropy (-TΔS) in kcal/mol.

Complex[a]	ΔE_{vdw}	ΔE_{elec}	ΔG_{gas}	ΔG_{polar}	$\Delta G_{nonpolar}$	$\Delta G_{solvation}$	-TΔS	ΔH	$\Delta G_{bind\ (T)}$	$\Delta G_{bind\ (E)}$[b]
NS	-89.08	-16.42	-105.05	47.40	-12.13	35.27	33.26	-70.23	-36.97	
ATV	-79.86	-7.75	-87.62	30.81	-10.00	20.80	30.42	-66.81	-36.39	-14.35
DRV	-65.91	-37.15	-103.06	53.46	-8.30	45.15	23.15	-57.92	-34.77	-13.81
LPV	-76.37	-9.40	-85.77	27.83	-9.81	18.02	33.85	-67.74	-33.89	-13.24
TPV	-60.78	-37.97	-98.75	57.79	-8.38	49.41	25.47	-49.34	-23.87	-13.19

[a]The complex systems are ranked in order of theoretical binding energies.

[b]Experimental binding free energies for four inhibitors on C-SA PR reported by Maseko *et al.* (15).

It is evident from Table 1 that the theoretical binding energies [$\Delta G_{bind(T)}$] of the compounds follow the same order as the experimental binding energies [$\Delta G_{bind(E)}$] reported by our group(15) which reveal that the computational model is effective. The purpose of the current work is not necessarily to compare the theoretical and experimental activities but to discover major determinants responsible for different experimental binding affinities. Determinants that will be explored are interatomic interaction occurring in HIV-1 PR complexed with inhibitors as well as distance analysis for the protease. Additional parameters to study flap dynamics will also be explored.

The results demonstrate that the natural substrate attained the highest theoretical binding energy when compared to the inhibitors. The high van der Waals interaction observed -89.08 kcal/mol for the natural substrate is thought to be the major contributing factor towards the high binding affinity of the substrate. Interestingly, ATV exhibited less electrostatic interactions (ΔE_{elec} = -7.75 kcal/mol) among other complexes. However, it still exhibits strong inhibition with a $\Delta G_{bind(T)}$ = -36.39 kcal/mol mainly due to high van der Waals (ΔE_{vdw} = -79.86 kcal/mol)

interactions as well as reduced polar (ΔG_{polar} = 30.81 kcal/mol) and higher non-polar ($\Delta G_{nonpolar}$ = -10.00 kcal/mol) solvation energies.

The TPV—PR complex, on the other hand, demonstrated relatively high electrostatic interactions (ΔE_{elec} = -37.15 kcal/mol) among other inhibitors but its van der Waals interactions (ΔE_{vdw} = -60.78 kcal/mol) showed that TPV exhibited the weakest binding affinity compared to the other inhibitors. Furthermore, this result can be attributed to an increased ΔG_{polar} (= 57.79 kcal/mol) that leads to a high cost of solvation energy ($\Delta G_{solvation}$ = 49.41 kcal/mol). The DRV—PR complex demonstrated similar interactions to TPV—PR as exemplified by the reduced hydrophobic interaction energy (ΔE_{vdw} = -65.91 kcal/mol), increased electrostatic interaction energy (ΔE_{elec} = -37.15 kcal/mol) as well as increased solvation free energy ($\Delta G_{solvation}$ = 45.15kcal/mol). The similar interactions of these two inhibitors can be linked to their structural arrangement as they both contain a sulphonamide moiety, an extremely polar group. This clearly explains the high values of ΔG_{polar} for both DRV and TPV and these inhibitors have to desolvate water around them in order to be accommodated in the active site, leading to paying a desolvation penalty (57) and hence the observed large values of $\Delta G_{solvation}$ for both of these inhibitors.

When comparing LVP—PR with the best inhibitor it can be seen that both demonstrate a similar pattern of interaction energies as exemplified by increased ΔE_{vdw} (= -76.37 kcal/mol), reduced ΔE_{elec}(= -9.40 kcal/mol) and reduced $\Delta G_{solvation}$ (= 18.02 kcal/mol). As far as the entropy of the systems are concerned, comparative increased entropy was observed for ATV and LPV complexes as compared to DRV and TPV complexes.

2.3.2 Stability and flexibility of the HIV-1 PR complexes

In order to explore the dynamic stability of the apo enzyme, natural substrate (NS) and the four

HIV-1 inhibitor complexes, root-mean-square deviation (RMSD) values of the Cα, C and N atoms were calculated and are presented in Figure 2A. All these systems maintained RMSD values of less than 2.5 Å in the whole simulation run whereas all drug/PR complexes displayed better dynamic stability (RMSD < 2.0 Å) than the apo form which was able to attain an RMSD of above 2 Å at around 22 ps. In addition, all these systems (Figure 2A) display an average RMSD values of less than 1.5 Å including those presented in Figure S2 (Apo: 1.408 Å, NS: 1.4436 Å, ATV: 0.944 Å, DRV: 1.173 Å, LPV: 1.378 Å, TPV: 0.838 Å). Interestingly, among all the complexes, ATV—PR and TPV—PR revealed better dynamic stability as the RMSD of the two systems rarely crossed 1.200 Å during the entire run. The RMSD analysis indicates that the fluctuations of the all systems from corresponding crystal structures are similar. Three different MD simulations of ATV and DRV were carried out by varying initial atomic coordinates or initial atomic velocities. RMSD plots of these conformations showed that they displayed relatively similar stability (Figure S1).

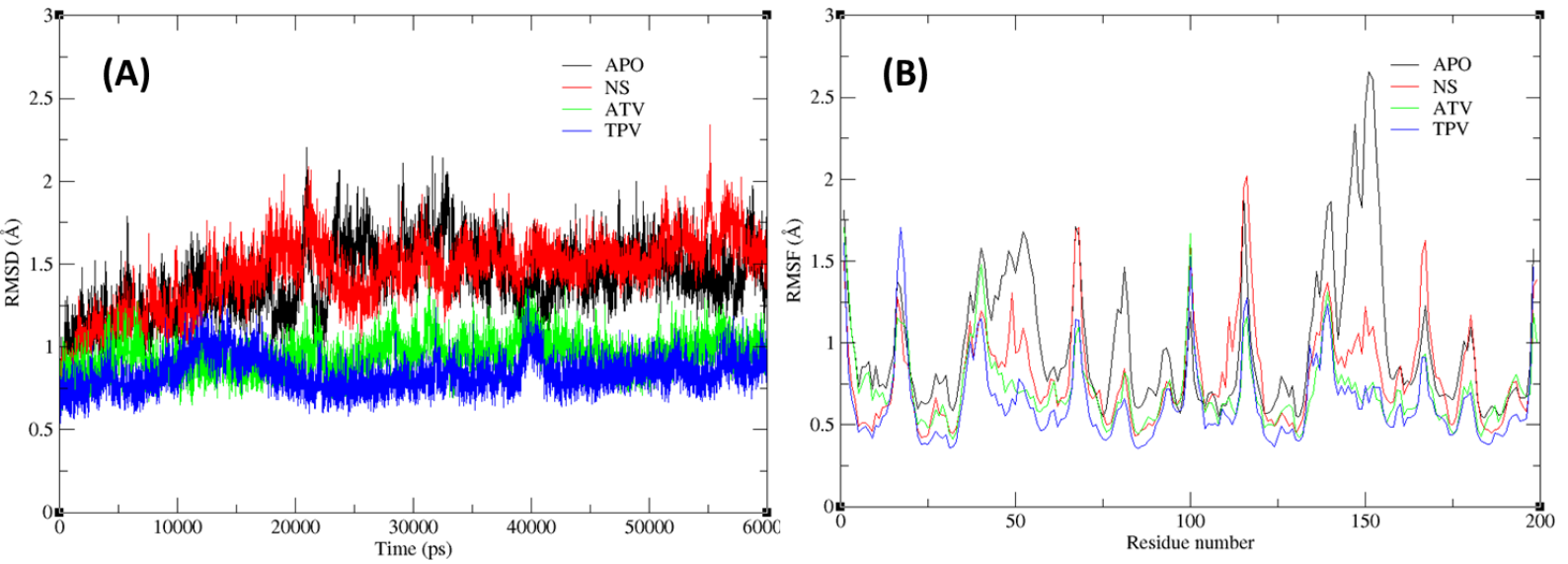


Figure 2. (A) Root Mean Square Deviation and (B) Root Mean Square Fluctuation for the 4 systems over the 60 ns MD trajectory.

To understand the residue fluctuations and flexibilities, the root mean square fluctuation (RMSF) plot was also analysed. This plot (Figure 2B) indicates that the drug/PR complexes

experience much lower fluctuations for the most part of the simulation. Due to the fact that the active site region is known to be rigid (58) our analysis for fluctuation focussed on the flap region. The fluctuations of the drug/PR complexes are extensively reduced in the flap regions (residue 43-56 and residue 42'-55'). Considering that flap dynamics is directly related to substrate stability in the HIV-1 protease (59), it may be assumed that binding of the drugs stabilizes the flap regions (retaining closed/semi-open conformation), although the extent of stabilization varies from compound to compound. The flap opening takes place by concerted downward movements of fulcrum (residues 10-23), hinge (residues 35-42 and 57-61) and cantilever (residues 62-78) regions (60). Simultaneously, these three important regions fulcrum, hinge and cantilever including the flap region were analysed from the RMSF plot (Table 2). Higher stability of residues in these regions were obtained for all complexes in the decreasing order of TPV, ATV, DRV, LPV and NS with the apo enzyme having the least stability (Table S2-B). The RMSF plot indicates that the overall fluctuations of these three regions correspond to the flap regions of the complexes.

Table 2. Average RMSFs observed in different regions of the apo, natural substrate and drugC-SA protease complexes for the 60 ns MD simulation.

Regions[a]	Chain	Average RMSF (Å)					
		Apo	NS	ATV	DRV	LPV	TPV
Fulcrum	A	0.883	0.7307	0.792	0.950	1.05	0.802
	B	0.871	1.0895	0.702	1.026	0.921	0.685
Hinge	A	1.121	1.0242	0.93	0.923	1.035	0.769
	B	1.214	1.0905	0.818	0.870	0.981	0.808
Flap	A	1.421	0.9552	0.737	0.771	0.884	0.639
	B	1.896	0.9490	0.709	0.736	0.898	0.683
Cantilever	A	0.94	0.8286	0.694	0.847	1.026	0.616
	B	0.807	0.8904	0.619	0.761	0.752	0.589

[a] Fulcrum: residues 10-23), hinge: residues 35-42 and 57-61, Flap: residues 45-55 and cantilever: residues 62-78.

Flap dynamics of HIV-1 PR is closely associated with the substrate binding and is therefore a crucial part of drug binding and effectiveness (61). Analyses of the RMSF plots indicated that flap dynamics play an important role and can be used to determine the functionalities of the FDA approved second generation drugs in C-SA PR. The movements of the flap region and other parts of the enzyme were further analysed by the principal component analyses (PCA) analysis in the latter part of the manuscript. To obtain insight on the flap dynamics, some important distances (62, 63) and angle analyses (62, 63) were measured over the MD trajectories. The histogram plots of the distance and angle analyses are presented in Figure 3.

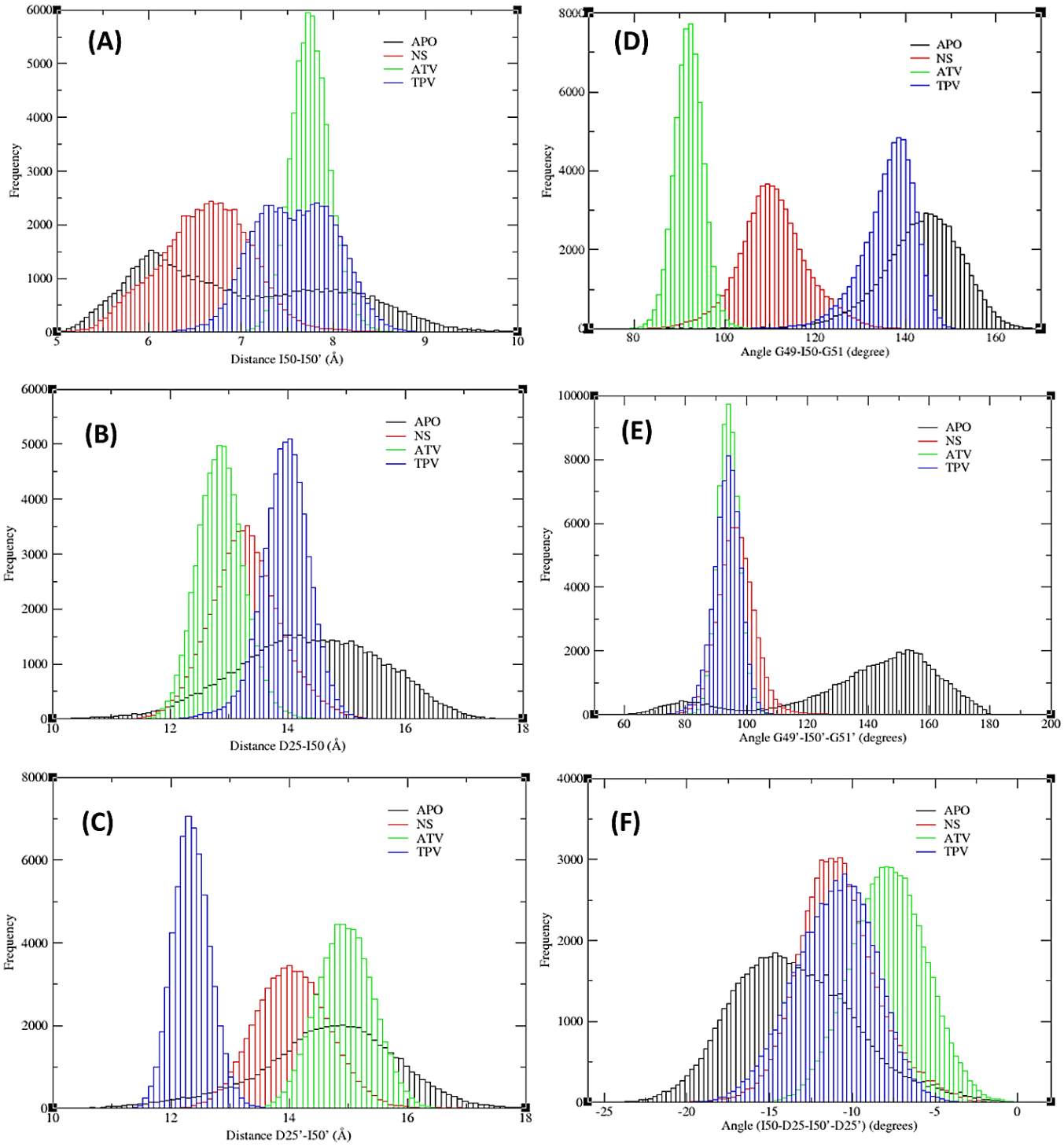


Figure 3. Distance and angle analyses for the elucidation of flap dynamics in C-SA PR for the 60 ns MD simulations; (**A**) distance between flap tips (Ile50 and Ile50′), (**B**) distance between active site residue and flap tip in chain A (Asp25-Ile50), (**C**) distance between active site residue and flap tip in chain B (Asp25′ -Ile50′), (**D**) TriCα angle of chain A (Gly49-Ile50Gly51), (**E**) TriCα angle of chain B (Gly49′-Ile50 ′-Gly51′), (**F**) dihedral angles of chain A (Ile50-Asp25-Ile50 ′ -Asp25′).

Initially the stability of the semi-open flap conformation was monitored by measuring the distance between Ile50 and Ile50′, which are located on the tip of the flaps. The flap tip distance between these two residues is one of the most frequently studied flap dynamics parameter in the HIV-1 protease (59, 64, 65). It is clear from the plot in Figure 3A that the apo form displays prominent flexibility as the flap tip distance ranged from approximately 5 Å to close to 10 Å. The flap tip distance attained by the natural substrate complex ranged between 5.5 and 7.5 Å and this indicates that the presence of the substrate ensured that the flap tip residues are much more rigid. It is observed that the drug ATV complex has a flap tip distance (Ile50 and Ile50′) that moved within a shorter range compared to those of other drug complexes (Figure S3-A), whose flap tip distances displayed a periodic increase from closed (5-6 Å) to semi-open (9-10 Å) form. These figures demonstrate that complex formation significantly reduces fluctuations of the flap residues. Such an observation is supported by experimental evidence (66) where it was reported that inhibitor binding induces a shift to the closed state in C-SA PR. In the current analysis, ATV attained flap tip distance values of slightly less than 8 Å and DRV- and LPVcomplexes showed higher tendency to remain in semi-open conformation (Figure S3) while TPV displayed even smaller flap tip distances. Interestingly, all these drug complexes rarely achieved full closed conformations (5-6 Å) (61) which has been reported as an important characteristic of C-SA protease by experimental evidence (7, 66). The drug complexes displayed a range flap-tip distance of 7-9 Å and did not achieve 5-6 Å attained by Apo as shown in Figure S3-A.

The distance between Asp25 and Ile50 of a given subunit has already been introduced as another an important criterion in understanding flap dynamics (67). A distance of more than 15.8Å is considered to represent a semi-open conformation for the flap gate (68). However, more importantly this parameter helps us to understand the volume of the active site (62). The distances between Asp25 and Ile50 residues of chain A and chain B are presented in Figure 3B

and 3C, respectively. Similar to the flap tip distance analyses, the apo form of the enzyme was found to cover a wide range of values (11-17.5 Å) for these parameters while the natural substrate displayed distances between approximately 12 to 16 Å. The drug complexes, on the other hand, displayed narrow ranges for the corresponding distances. In chain A (Asp25Ile50), the histogram plots for the ATV, DRV, LPV and TPV—PR complexes overlapped with each other with increasing peak values distributed at 12.7, 13.3, 13.7 and 13.8 Å, respectively (Figures 3B and C and Figure S3 B and C). However, in chain B the distance between Asp25′Ile50′ varied to a considerable extent from one complex to another. The histogram peak of ATV—PR was located at 15.0 Å, which is around 2 Å longer than for DRV—PR, having histogram peak value found at 13.2 Å. The Asp25′-Ile50′ peaks for the LPV and TPV complexes are overlapped with larger peak values distributed at 12-12.2 Å. These figures suggest that the ATV—PR has a considerably larger active site volume than other complexes (62). On the other hand, the corresponding distances (Asp25′ -Ile50′) for all complexes were shorter than this value for the apo form suggesting that binding of drugs stabilized the flap fluctuations to a considerable extent. In chain B, the ATV complex showed comparatively longer distances between these two amino acids (Asp25′ and Ile50′), and this clearly shows that the dynamic behaviour of the flaps is unsymmetrical as the ATV complex displayed mostly shorter distances. The results also show that the apo form of HIV-1 PR displayed relative high symmetry as a similar pattern for the distances between the flap tip and active site residue on both chains was almost similar.

Schiffer *et al.* (65) used the term flap curling of the TriCα angles of the residues in the flap tip or nearby region (Gly49-Ile50-Gly51) to investigate the open and closed conformations of flap dynamics. The higher curling of the flap tips produces faster flap opening(68). These results are presented in Figure 3 (D and E) while those that include DRV and LPV are in Figure S3 (D and E). The ATV and DRV complexes reveal more flap curling in of flap tip residues justified

by lower TriCα angles in both chains (Chain A: Gly49-Ile50-Gly51, Chain B: Gly49′Ile50′-Gly51′) in Figure 3D and 3E as well as Figure S3, D and E. The LPV—complex showed comparatively higher TriCα angles in both chains indicating the flap region of this complex appears to exhibit higher flexibility than other complexes (Figure S3D and E). The TPV—complex displayed variable flap curling in two different chains. It seems to suggest that TPV binding leads to anti-correlated movements in two different chains (Figure 3D and 3E). In addition, it is interesting to note that for both D25-50 and the TriCα angle, the TPV complex seemed to reveal smaller distance and angle values in one similar monomer, respectively, and that is due to a very strong hydrogen bond network that is established by TPV in various fundamental regions of the protease. For example, there is a strong hydrogen bond interaction between the backbone carbonyl oxygen of Gly48 and the hydrogen between the sulphonamide and the aromatic ring as can be seen in Figure S9. The hydrogen bond network is also strengthened by the interaction of Gly27 carbonyl oxygen and the aromatic carbon found between the trifluoromethyl group and the pyridine nitrogen (Figure S9).

To study such conformational changes, the dihedral angle, (Φ), was also calculated in the post MD analysis between the Cα atoms of the tip-catalytic dyad-hinge residues (Ile50-Asp25-Ile50′-Asp25′). Evaluating the dihedral angle gave insight into the twisting motion of the Cα atoms between the- flap tip, active site and hinge residues in the flexible region. It is evident that major twisting phenomenon did not occur for the ATV and DRV complexes (Figure S3F). It is evident that a major twisting phenomenon did not occur in all systems as none of them had any shifting in angles from a positive to a negative side or from the positive side to the other side as this would have implied such a motion (69). However, it is worth mentioning that the apo form displayed a wider range of dihedral angle compared to the natural substrate and inhibitor-bound complexes (Figure 3F). Figure S3F does indicate that the two inhibitors with

lower binding affinities (LPV and TPV) seemed to have wider ranges for the dihedral angle compared to the two inhibitors with higher binding affinities (ATV and DRV).

Overall, from all these distance and angle analyses, it may be inferred that the drug complexes achieve neither fully closed nor semi-open conformations. They rather remain in close to semiopen states as observed earlier by Naicker *et al.*(7, 28). The ATV complex showed a larger active site volume compared to other complexes. Furthermore, the DRV—PR and LPV—PR complexes displayed longer flap-tip distances than other two drug complexes. The TriCα analyses as well as dihedral angle analyses revealed that LPV (both chains) and TPV (chain B) complexes exhibit larger flap movements as compared to other two complexes.

2.3.3 Cross-correlation analysis

To understand the alterations of internal dynamics due to drug binding in C-SA PR, crosscorrelation matrices were calculated for apo, natural substrate and four drug bound protein complexes by the CPPTRAJ program (39) on the equilibrated MD trajectories. The results are depicted in Figure 4. The cross-correlation plots for DRV and LPV complexes are provided in the supplementary materials. Strongly correlated and anti-correlated movements are shown by positive (red and yellow) and negative (blue) spots, respectively. The diagonal parts show obvious correlated movements. The other regions rarely show highly correlated movements. The apo form was initially analyzed thoroughly and the most important parts of the matrices are highlighted with four regions R1-R4. R1 and R2 represent the correlations observed with residues 10-40 and residues 60-90 of chain A and chain B, respectively. These regions cover major parts of the fulcrum, hinge and cantilever regions and demonstrate correlation that may exist between fulcrum and cantilever. Higher correlation indicates a possibility of concerted motions leading to flap opening. It is observed from Figure 4 that comparing the apo form of the enzyme, the positive correlations of these regions (R1 and R2) are reduced in all complexes

except for the LPV—complex as indicated in Figure S5. The R3 region points to the correlated movements observed near the flap region of chain A. Although diagonal parts show strong correlated movements for obvious reasons, reduction of flap movements as well as interactions of flap residues with the substrate/ligand should reduce the positive correlation of this region. This phenomenon is observed for most of the analysed complexes. Among these, the natural substrate complex displayed a relative increased correlation when compared to the inhibitor—complexes with the TPV complex showing the most decreased correlation in the R3 region. Therefore, by analysing the R1-R3 parts it can be inferred that higher flap movements are likely for the LPV—complex, where flap movements are considerably reduced for TPV followed by ATV and DRV.

The R4 region represents correlation of residues 30-60 of chain A and 10′-80′ of chain B. The apo form exhibited anti-correlated movements in region R4 whereas for PI complexes the anticorrelated movements are significantly decreased. One explanation for these phenomena may be the fact that all these PIs are known to interact with amino acids of both chains. Therefore, inhibitor binding helped in decreasing the anti-correlated movements in this region. Interestingly, DRV—complex showed the most positive correlation in this region (62, 63).

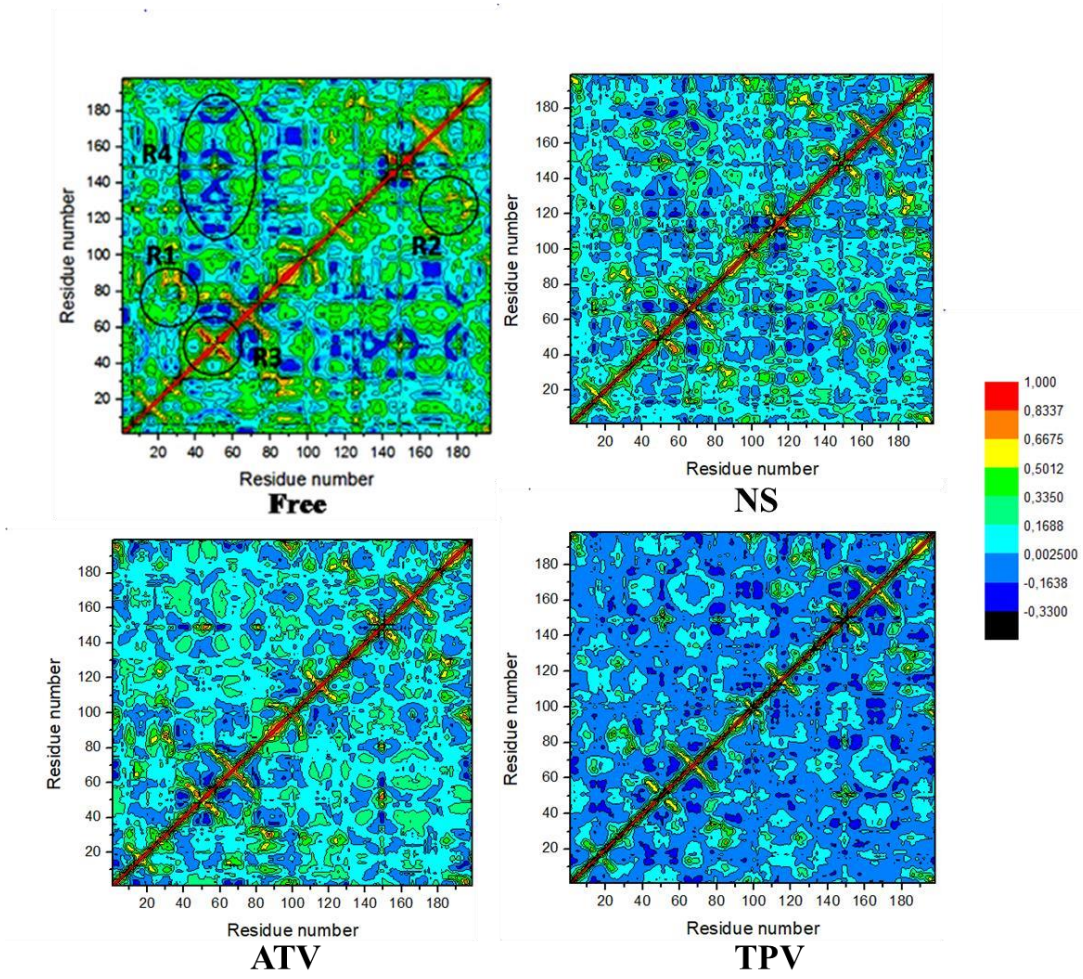


Figure 4. Cross-correlation matrices of the fluctuations of the coordinates for Cα atoms around their mean positions for each complex over the 60 ns MD simulations.

2.3.4 Hydrogen bonding interaction

The general output data generated by hydrogen bonding are; identities of donor and acceptor atoms; frames, which defines the time step of the MD trajectories, fraction (Frac) which defines the portion of the total time during which the hydrogen bond was detected, average bond distance (AvgDist) between the heavy atom acceptor and the hydrogen atom, and the average bond angle (AvgAng) between the donor and acceptor atoms.

The natural substrate appears to exhibits relatively stronger interactions with Arg8 in comparison to other amino acids of the protease, made possible by the proton donation of Arg8

to the oxygen atom of Asp of natural substrate (Table 3). Our observations are in agreement with those reported in an experimental work by Prabu-Jeyabalan *et al(70)* where they observed Arg8/8', in addition to Asp29, Asp30/30' and Gln58', forming direct side chain hydrogen bond interactions with the substrate peptides. Table 3 also reveals that hydrogen bond interactions between the natural substrate and the enzyme was also observed in the flap tip and active site region, with the Gly48/48' and Asp30 residues being responsible for the interaction in these two regions, respectively. Figure S6 properly indicates the interaction of Gly residues in the flap tip region of the protease with the natural substrate.

Hydrogen-bond interactions were also monitored around the regions of the flap tips and the active site for the protease inhibitor—complexes. The most significant hydrogen bonds observed for ATV were with the active site residues, Asp29 and Asp29' (Asp29 of chain B). That is possibly one of the contributing factors to its relatively high binding affinity (Table 1). These results also show that the hydrogen bond interaction between ATV and the flap residue Gly48, also reported by Clemente *et al.* (71) may contribute to the relatively short distance between the flap tip (Ile50) and active site residues (Asp25) in chain A, in comparison to the other inhibitors where Gly48 is located. It can be seen from the 3D structure in Figure S7 that ATV makes a strong hydrogen bond interaction with Gly49 at the flap tip.

Table 3. Hydrogen bond interaction occurring between the ligand and residues in the active site and flap tip region of the protease for different bound systems.

Complex	Acceptor	Donor	Frames	Frac	Avg Dist(Å)	Avg Ang(°)
	ARG_199—OD1[a]	ARG_8—NH2[d]	25284	0.4214	2.8250	156.8562
NS	ARG_199—OD2[b]	ARG_8—NH2	25071	0.4178	2.8252	157.0978
	ARG_199—OD2	ARG_8—NH1[c]	21877	0.3646	2.8444	161.1716
	ARG_199—OD1	ARG_8—NH1	20601	0.3433	2.8446	161.6852
	GLY_148—O	ARG_199—N	18791	0.3132	2.8875	152.5882
	GLY_48—O	ARG_199—N	5391	0.0898	2.9047	157.3447
	ARG_199—O	GLY_148—N	2791	0.0465	2.8940	149.2057
	ASP_30—OD2	ARG_199—NZ	2641	0.0440	2.8040	150.9800
	GLY_27—O	ARG_199—N	2187	0.0365	2.8982	152.0820
	ASP_30—OD1	ARG_199—NZ	1920	0.0320	2.8074	151.9607
	GLY_147—O	ARG_199—N	994	0.0166	2.9076	154.7709
	ARG_199—O	ARG_107—NH2	833	0.0139	2.8824	146.6713
	ARG_199—O	GLY_147—N	824	0.0137	2.8909	157.1457
	ASP_29—OD2	ARG_199—NZ	712	0.0119	2.8125	152.7634
	GLY_49—O	ARG_199—NH1	564	0.0094	2.8646	156.6636
	DR7_199—OAJ[f]	ASP_29—N	10814	0.1802	2.9057	158.4132
ATV	DR7_199—OAI[g]	ASP_128—N	6052	0.1009	2.8997	160.4608
	DR7_199—OAI	DR7_199—OAM[h]	4315	0.0719	2.8923	153.6036

	GLY_48—O	DR7_199—NB[i]	3579	0.0597	2.9161	155.2569
	GLY_147—O	DR7_199—N	494	0.0082	2.9196	159.4964
	AB1_199—O2	ILE_149—N	6073	0.1012	2.9022	150.4994
	AB1_199—O3	GLY_48—N	3899	0.0650	2.9175	149.0338
LPV	GLY_27—O	AB1_199—O4	846	0.0141	2.8719	150.3595
	ASP_124—OD2	AB1_199—O4	388	0.0065	2.8582	156.3095
	TPV_199—O31[j]	GLY_48—N	47862	0.7977	2.8471	161.3949
	TPV_199—O31	GLY_49—N	20977	0.3496	2.8445	149.6769
TPV	ASP_124—OD1	TPV_199—O8	15684	0.2614	2.8294	161.6918
	TPV_199—O7	ILE_149—N	9397	0.1566	2.9152	157.4545
	ASP_124—OD2	TPV_199—O8	3544	0.0591	2.8290	160.8629
	ASH_25—OD1	TPV_199—O8	3473	0.0579	2.8792	148.4075
	ASH_25—OD2	TPV_199—@O8	338	0.0056	2.9178	155.8354

[a]OD1 and [b]OD2: First and Second oxygen treated as proton acceptors in ARG, respectively. [c]NH1 and [d]NH2: First and Second hydrogen donated by the proton donor nitrogen, respectively. [e]NZ: One of the terminal nitrogen of the natural substrate to donate a proton. [f]OAJ, [g]OAI and [h]OAM: Oxygen atoms of ATV treated as hydrogen acceptors. [i]NB: Nitrogen atom of ATV treated as proton donor. [j]O31: One of oxygen atom of TPV treated as proton acceptor.

NOTE: DRV displayed extremely low bond fraction (appr. zero) and as such it was not included in Table 3 but it is included in the supplementary material.(Table S1)

It is interesting to note that TPV, in comparison to ATV, was observed to have a strong interaction with Ile50' (represented as Ile149) in the flap tip region. This result in a shorter distance between TPV and the flap tip residue. This clearly supports the observed results of a short distance between Asp25 and Ile50 of monomer B where Ile50' is located. The RMSF also does supports this observation as it was shown from Figure 2 that Ile50 of chain B

displayed better stability. The 3-D structures reported in Figure S9 also clearly show that TPV has more HB interactions with the flap tip residues than ATV.

Figures S8 and S9 highlight an important factor that distinguishes the interaction of TPV from DRV despite both having sulfonamide moieties. It is noted that the sulfonamide group in TPV has a partially exposure to the solvent while that of DRV is completely buried. The partial exposure of the sulfonamide group in TPV seems to display consistency with the low enthalpy of the TPV—complex (Table 1).

2.3.5 Per-residue decomposition analysis

The binding free energy was decomposed into contributions from the individual residues of the protein to gain insight into the role of particular residues in the binding mechanism of different ligands(72). Post-dynamic binding free energy calculations using the MM/GBSA approach were performed for four drug-complex systems. The results are illustrated in Figure 5 and 6. Figure 5 depicts the decomposition energy of ATV and DRV complexes and it can be seen that ATV demonstrated increased electrostatic interaction energy compared to DRV—PR complex (complying with the results of MM-GBSA binding free energy analyses). DRV—PR exhibited high favourable electrostatic interactions (complying with the theoretical binding energy analyses results, Table 1) with amino acids like Asp29, Asp30, Gly48, Gly49, Ile50, Asp25′, Ala26′, Gly49′ and Ile49′. These interactions either are reduced or absent in the ATV—PR complex. This is in good agreement with theoretical binding energy analyses results (Table 1). However, these strong favourable electrostatic interactions of DRV are compensated by strong unfavourable electrostatic interactions with Arg8′, Asp 30′, Ala8, Ala28, Ile47 and Asp29′.

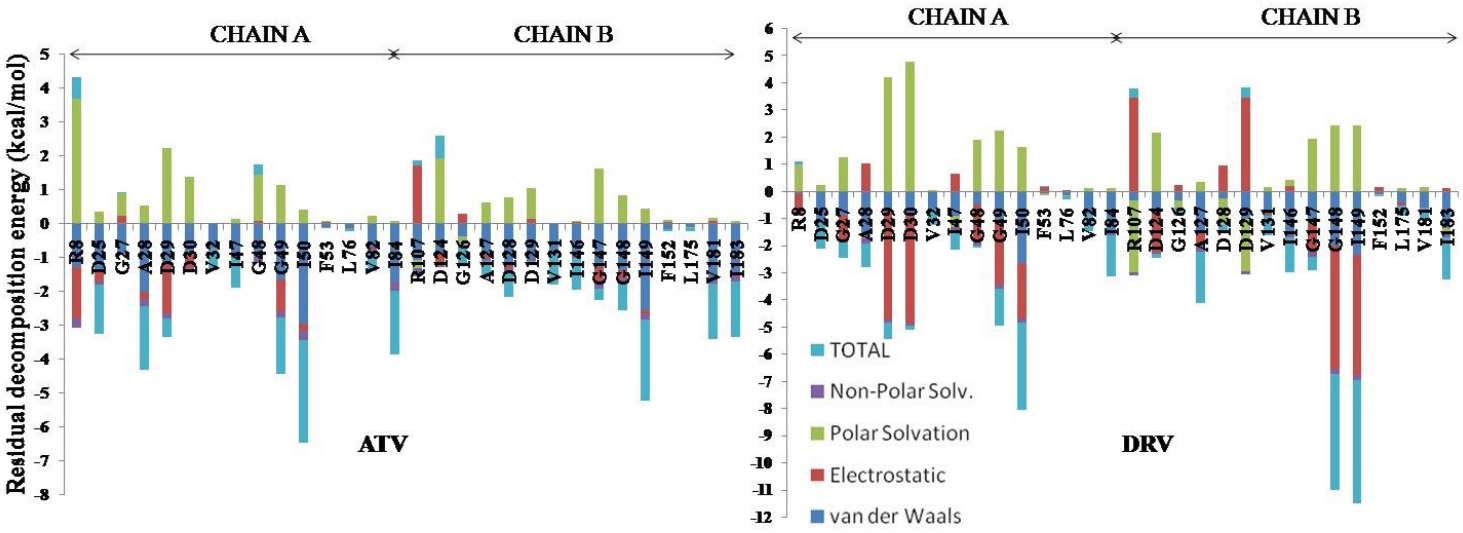


Figure 5. Per-residue decomposition profiles of the ATV and DRV complexes using the MMGBSA approach.

Van der Waals interactions with Ile50, Ile50′ and Ile84 have important contributions in overall binding energy of ATV—PR. Furthermore, ATV depicted a tendency to develop hydrophobic interactions with most of the important binding site residues. In DRV—PR, these van der Waals interactions are reduced causing an overall lowering of binding energy (Table 1). A higher $\Delta G_{solvation}$ was recorded for DRV complex in MM-GBSA analyses (Table 1). High positive solvation energy was also observed for DRV—complex due to interactions with amino acids Asp29, Asp30, Asp25′, Gly48′, Gly49′ and Ile50′ (Figure 5). Negative polar solvation energy as well as positive electrostatic energy for Arg8′, Asp29′ and Asp30′ indicate that these amino acids may be involved in water mediated interactions with DRV.

Figure 6 illustrates per residue decomposition energies of LPV—PR and TPV—PR. Overall, the TPV complex results indicate larger electrostatic interactions compared the LPV—complex, especially around the active site and flap tip residues. This interaction is thought to reduce the flexibility of the flap tip region as seen from the RMSF and flap-tip distance plots (Figures 2 and 3A) whereby TPV—complex displayed quite a reduced flexibility showed by very low fluctuations around the flap region and low flap-tip distances.

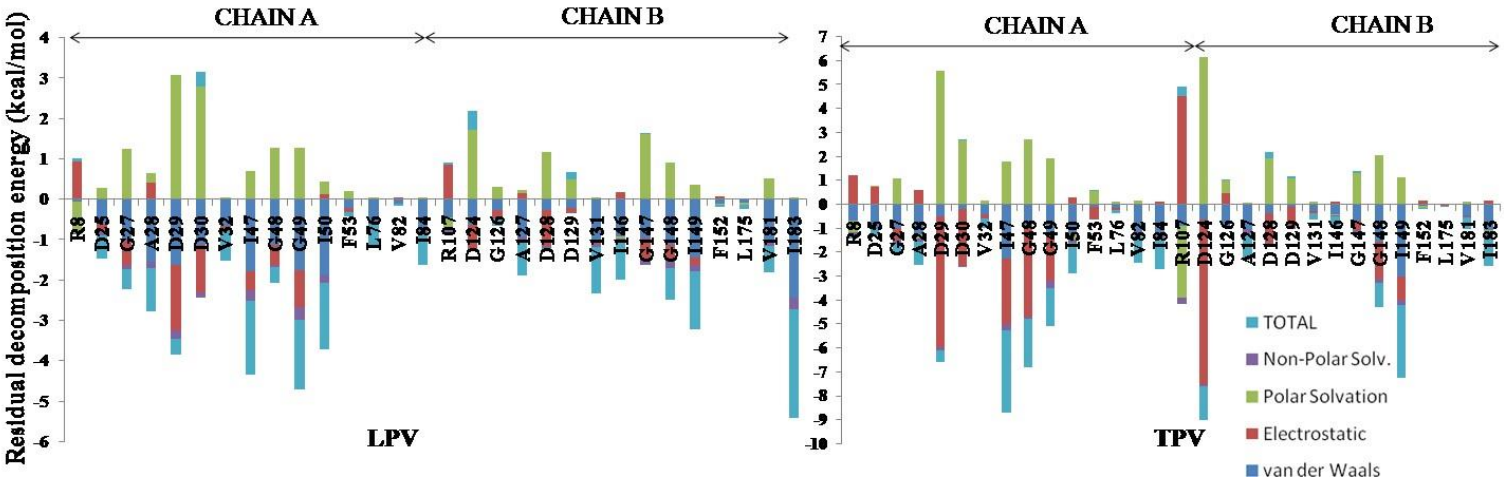


Figure 6. Per-residue decomposition profiles for the LPV and TPV complexes using the MM/GBSA approach.

The LPV complex exhibits favourable electrostatic interactions with amino acids such as Asp25, Gly27, Asp29, Asp30, Gly49, Asp25′, Asp29′ and Gly48′. However, unfavourable electrostatic interactions are also observed with Arg8, Ala28 and Arg8′. Favourable van der Waals interactions with Ile47, Gly49, Ile50, Gly48′, Gly49′ and Ile50′ were found. The TPV complex demonstrated strong electrostatic interactions with residues such as Asp29, Asp30, Ile47, Gly48, Gly49, Asp25′, Asp29′, Gly49′ and Ile50′. Similar to DRV complex, TPV—PR revealed strong positive electrostatic potential as well as negative solvation free energy with Arg8′. However, the major factor that contributed to the lower theoretical binding energy (Table 1) is the positive polar solvation energy. These solvation energies, which extended up to 6-7 kcal/mol, are mainly observed with Asp29 and Asp25′. Furthermore, van der Waals interactions were compromised to a considerable extent for TPV complex especially with the flap residues.

Decomposition energy for the natural substrate complex in Figure S4 clearly shows that Asp29 and Asp30 of the protease exhibit extremely stronger electrostatic interactions with the substrate. Some of the residues surrounding the active site (Asp29 & Thr30) show to display larger electrostatic interaction for the natural substrate compared to those observed

for inhibitors. This particular observation is thought to contribute towards the high binding affinity of the natural substrate compared to the inhibitors.

2.3.6 Principal Component Analysis (PCA)

To explore the nature of the motions in the HIV-1 PR enzyme, PCA was carried out. This analysis highlights in a qualitative manner the effect of the inhibitor's binding on the conformational motion in the HIV-1 PR. Based on the porcupine plots (Figure 7), it is evident that inhibitor binding reduced the overall movement of the enzyme as compared to the apo form. Furthermore, the apo form showed higher fluctuations in the cantilever (residues 66-70) and fulcrum regions. All complexes showed anti-symmetric movement patterns in fulcrum, hinge and cantilever regions.

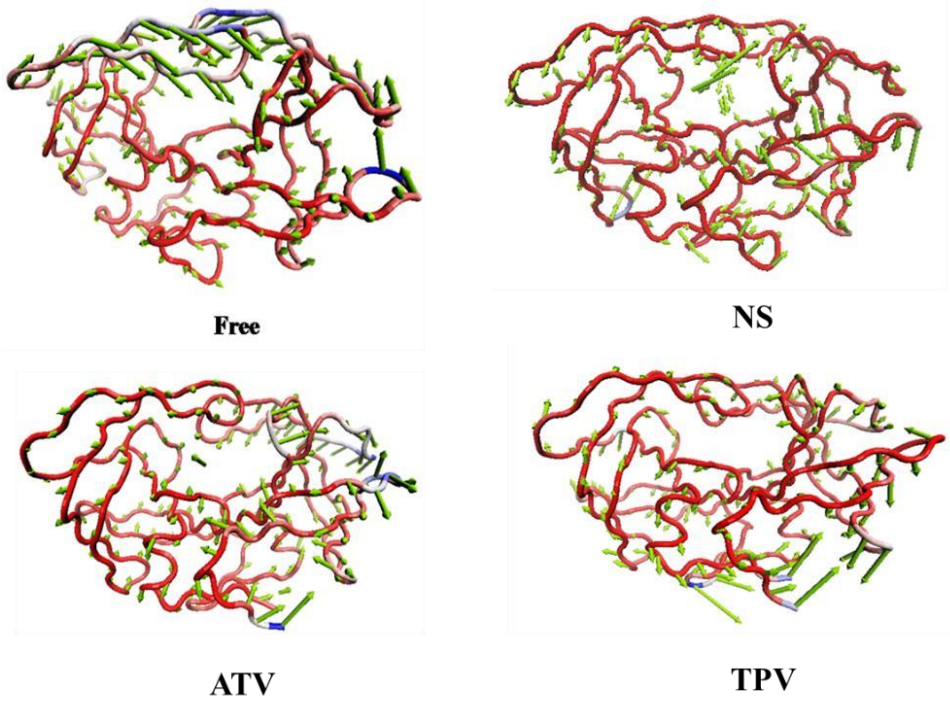


Figure 7. Collective motions corresponding to PC1 obtained by performing principal component analysis on a 60 ns MD simulation trajectory after the equilibrium (the protein part is presented in red ribbon, higher fluctuating parts are presented in grey and blue).

The Figure 7 does indicate that the binding of the natural substrate makes the protease more rigid and it is evident that the complex had mixed patterns of motions in most regions of the enzyme. The ATV—PR showed increased fluctuations in three regions of chain A. The hinge region (residues 35-42), part of fulcrum (residues 15-19) and residues 82-84. Such large fluctuations were not observed in the other chain (chain B) indicating an anti-symmetric movement of the complex. Such anti-symmetric movement was also found in the distance analyses (Figure 3B and 3C, the Asp25-Ile50 distance was shifted to a considerable extent for chain A and chain B). In chain B, the ATV—PR exhibits horizontal movements between fulcrum, hinge and cantilever regions that may facilitate horizontal hydrophobic sliding (73).

Similarly, for DRV the cantilever region of Chain A (residues 66-70) was also found to have larger vertical fluctuations (Figure S10). DRV—PR also demonstrated higher vertical fluctuations in fulcrum regions of both chains (residues 14-18). Similar types of movements were also observed with LPV—PR that depicted significant asymmetric fluctuations at the cantilever region of chain A (residues 66-70) (Figure S10). Both chains of the cantilever region of TPV—PR depicted larger vertical movements than other parts of the enzyme complex. Only ATV-PR showed partial horizontal movement between the fulcrum and cantilever regions (in Chain A) which is indicative of the proper hydrophobic sliding. When combining the PCA results with the correlation matrix analyses (Figure 4) it may be assumed that LVP—PR and TVP—PR may have vertical movements in fulcrum and cantilever regions in both monomers. Such movements are indicative of distorted hydrophobic sliding (73). Such perturbed sliding between fulcrum and cantilever regions reduces hydrophobic interactions (hydrophobic core) in the enzyme system and affects receptor ligand interactions as well (73).

2.4 Conclusion

Several factors affect the binding of protease inhibitors through altering the flexibility of HIV1 protease flaps, the most crucial part of the enzyme in ensuring the catalytic function of the enzyme. It is also interesting to note that this study has been able to bring together two major components in protease complexing with inhibitors; binding of protease inhibitors onto the active site and the interaction of the inhibitors with the flaps of the protease causing it to be inactive. Higher hydrophobic interactions were found to be major contributing factor for higher binding affinity of ATV and this could be attributed to the higher active site volume and higher of flap stability. The hydrophobic interactions are highly compromised for the DRV and TPV complexes.

All complexes showed more stability in the flap dynamics compared to the apo form. However, none of the complexes achieved fully closed form in the entire simulation. Although flap— tip distances of the complexes were found to overlap with each other, LPV and DRV complexes showed comparatively larger flap tip distances versus ATV—PR and DRV—PR. Furthermore, the active site of DRV—PR was predicted to be larger than LPV—PR and TPV—PR.

The distance and angle analyses also highlighted that TPV and LPV and have greater flap curling and movement compared to the other two complexes. One important characteristic of DRV is that it does not interact with residues outside the binding pocket(74). Therefore, the size of the active site volume may not directly affect its overall interactions. Interestingly, ATV and LPV showed almost similar patterns of interactions as the values of enthalpy and entropic contributions were found to follow a comparable trend. Both these complexes showed higher entropy than DRV—PR and TPV—PR. The TVP—PR showed lowered theoretical binding energy is owed to its high solvation free energy as well as lower van der Waals interactions. It was also observed that binding of inhibitors may slightly alter the hydrophobic core of the complexes leading towards variations in overall flap dynamics as well as stabilities of the

inhibitors at the active sites. The correlation matrix analyses, principal component, as well as hydrogen bond analyses, indicated that the possibilities of hydrophobic sliding are slightly higher in ATV complex. The lower binding affinity of TPV may be attributed to high solvation free energy as well as lower hydrophobic interactions. In conclusion, our findings reveal the complexity of drug—PR interactions and that emphasis on protease inhibition shouldn't only be made on the binding affinities of protease inhibitors into the active site but also interactions with the flap region has shown to be critical in ensuring that the flexibility of the flaps is reduced. These should be considered for drug development of protease inhibitors which possess the two characteristics; higher binding affinity and good interactions with the flap region.

Acknowledgments

This work was supported by the University of KwaZulu Natal, South African National Research Foundation and the Centre for High Performance Computing CHPC Cape Town.

Conflict of Interest

The authors declare no financial and academic conflict of interest.

References

1. Brik A, Wong C-H (2003) HIV-1 protease: mechanism and drug discovery. Organic & Biomolecular Chemistry;1: 5-14.

2. Lockhat HA, Silva JR, Alves CN, Govender T, Lameira J, Maguire GE, et al. (2016) Binding Free Energy Calculations of Nine FDA-approved Protease Inhibitors Against HIV-1 Subtype C I36T↑T Containing 100 Amino Acids Per Monomer. Chemical Biology and Drug Design;87: 487-98.

3. Mosebi S, Morris L, Dirr HW, Sayed Y (2008) Active-site mutations in the South african human immunodeficiency virus type 1 subtype C protease have a significant impact on clinical inhibitor binding: kinetic and thermodynamic study. Journal of Virology;82: 11476-9.

4. Carr A (2003) Toxicity of antiretroviral therapy and implications for drug development. Nature Reviews Drug Discovery;2: 624-34.

5. Joint United Nations Programme on HIV/AIDS (2016) Global AIDS update 2016. Geneva: World Health Organization;

6. Walker PR, Pybus OG, Rambaut A, Holmes EC (2005) Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. Infection, Genetics and Evolution;5: 199-208.

7. Naicker P, Stoychev S, Dirr HW, Sayed Y (2014) Amide hydrogen exchange in HIV-1 subtype B and C proteases--insights into reduced drug susceptibility and dimer stability. FEBS J;281: 5395-410.

8. Foulkes-Murzycki JE, Scott WR, Schiffer CA (2007) Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. Structure;15: 225-
33.

9. Mittal S, Cai Y, Nalam MN, Bolon DN, Schiffer CA (2012) Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. The Journal of the American Chemical Society;134: 4163-8.

10. Lee T, Laco GS, Torbett BE, Fox HS, Lerner DL, Elder JH, et al. (1998) Analysis of the S3 and S3' subsite specificities of feline immunodeficiency virus (FIV) protease: development of a broadbased protease inhibitor efficacious against FIV, SIV, and HIV in vitro and ex vivo. Proceedings of the National Academy of Sciences of the United States of America;95: 939-44.

11. Pokorna J, Machala L, Rezacova P, Konvalinka J (2009) Current and Novel Inhibitors of HIV Protease. Viruses;1: 1209-39.

12. Palmisano L, Vella S (2011) A brief history of antiretroviral therapy of HIV infection: success and challenges. Annali dell'Istituto Superiore di Sanità;47: 44-8.

13. Maseko SB, Natarajan S, Sharma V, Bhattacharyya N, Govender T, Sayed Y, et al. (2016) Purification and characterization of naturally occurring HIV-1 (South African subtype C) protease mutants from inclusion bodies. Protein Expression and Purification;122: 90-6.

14. Ahmed SM, Kruger HG, Govender T, Maguire GE, Sayed Y, Ibrahim MA, et al. (2013)

Comparison of the molecular dynamics and calculated binding free energies for nine FDA-approved HIV-1 PR drugs against subtype B and C-SA HIV PR. Chemical Biology and Drug Design;81: 208-18.

15. Maseko SB, Padayachee E, Govender T, Sayed Y, Kruger G, Maguire GEM, et al. (2017) I36T↑T mutation in South African subtype C (C-SA) HIV-1 protease significantly alters protease-drug interactions. Biological Chemistry;398(10):1109-1117

16. Velazquez-Campoy A, Todd MJ, Vega S, Freire E (2001) Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. Proceedings of the National Academy of Sciences of the United States of America;98: 6062-7.

17. Todd MJ, Semo N, Freire E (1998) The structural stability of the HIV-1 protease. Journal of Molecular Biology;283: 475-88.

18. Velázquez-Campoy A, Vega S, Fleming E, Bacha U, Sayed Y, Dirr HW, et al. (2003) Protease inhibition in African subtypes of HIV-1. Aids Review;5: 165-71.

19. Galiano L, Bonora M, Fanucci GE (2007) Interflap distances in HIV-1 protease determined by pulsed EPR measurements. Journal of the American Chemical Society;129: 11004-5.

20. Tóth G, Borics A (2006) Flap opening mechanism of HIV-1 protease. Journal of Molecular Graphics and Modelling;24: 465-74.

21. Karthik S, Senapati S (2011) Dynamic flaps in HIV-1 protease adopt unique ordering at different stages in the catalytic cycle. Proteins: Structure, Function, and Bioinformatics;79: 1830-40.

22. Meher BR, Wang Y (2012) Interaction of I50V mutant and I50L/A71V double mutant HIVprotease with inhibitor TMC114 (darunavir): molecular dynamics simulation and binding free energy studies. Journal of Physical Chemistry B;116: 1884-900.

23. Seibold SA, Cukier RI (2007) A molecular dynamics study comparing a wild-type with a multiple drug resistant HIV protease: Differences in flap and aspartate 25 cavity dimensions. Proteins: Structure, Function, and Bioinformatics;69: 551-65.

24. Perryman AL, Lin JH, McCammon JA (2004) HIV-1 protease molecular dynamics of a wildtype and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs. Protein Science;13: 1108-23.

25. Agniswamy J, Shen C-H, Aniana A, Sayer JM, Louis JM, Weber IT (2012) HIV-1 protease with 20 mutations exhibits extreme resistance to clinical inhibitors through coordinated structural rearrangements. Biochemistry;51: 2819-28.

26. Agniswamy J, Louis JM, Roche J, Harrison RW, Weber IT (2016) Structural Studies of a Rationally Selected Multi-Drug Resistant HIV-1 Protease Reveal Synergistic Effect of Distal Mutations on Flap Dynamics. Public Library of Science one;11: e0168616.

27. Tóth G, Borics A (2006) Closing of the flaps of HIV-1 protease induced by substrate binding: a model of a flap closing mechanism in retroviral aspartic proteases. Biochemistry;45: 6606-14.

28. Naicker P, Achilonu I, Fanucchi S, Fernandes M, Ibrahim MA, Dirr HW, et al. (2013) Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance. Biomolecular Structure and Dynamics;31: 1370-80.

29. King NM, Prabu-Jeyabalan M, Bandaranayake RM, Nalam MN, Nalivaika EA, Ozen A, et al. (2012) Extreme entropy-enthalpy compensation in a drug-resistant variant of HIV-1 protease. ACS Chemical Biology;7: 1536-46.

30. Stoll V, Qin W, Stewart KD, Jakob C, Park C, Walter K, et al. (2002) X-ray crystallographic structure of ABT-378 (lopinavir) bound to HIV-1 protease. Bioorganic & Medicinal Chemistry;10: 2803-6.

31. Muzammil S, Armstrong AA, Kang LW, Jakalian A, Bonneau PR, Schmelmer V, et al. (2007) Unique thermodynamic response of tipranavir to human immunodeficiency virus type 1 protease drug resistance mutations. Journal of Virology;81: 5144-54.

32. Bas DC, Rogers DM, Jensen JH (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. Proteins;73: 765-83.

33. Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. Proteins;61: 704-21.

34. Hornak V, Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics;65: 712-25.

35. Wang JM, 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: 1157–74.

36. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. Chemical Physics;79: 926-35.

37. Harvey M. J. FGD (2009) An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. Journal of Chemical Theory and Computation;5: 2371–7.

38. Ryckaert JP, Ciccotti, G., Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of N-alkanes. Journal of Computational Physics;23: 327–41.

39. Roe D. R. CTE (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation;9: 3084–95.

40. Srinivasan J, Cheatham, T. E., Cieplak, P., Kollman, P. A., Case, D. A. (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate−DNA helices. Journal of the American Chemical Society;120: 9401–9.

41. Onufriev A, Bashford D, Case DA (2000) Modification of the generalized Born model suitable for macromolecules. Journal of Physical Chemistry B;104: 3712-20.

42. Gohlke H, Case DA (2004) Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf. Journal of Computational Chemistry;25: 238-50.

43. Gohlke H, Kiel, C., Case, D. A. (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. Journal of Molecular Biology;330: 891–913.

44. Gohlke H, Case, D. A. (2003) Converging free energy estimates: MMPB(GB)SA studies on the protein−protein complex Ras−Raf. Journal of Computational Chemistry 25: 238−50.

45. Genheden S, Ryde, U. (2012) Will molecular dynamics simulations of proteins ever reach equilibrium? Physical Chemistry Chemical Physics 14: 8662−77.

46. Kopitz H, Cashman, D. A., Pfeiffer-Marek, S., Gohlke, H. (2012) Influence of the solvent representation on vibrational entropy calculations: Generalized Born versus distance-dependent dielectric model. J Comput Chem;33: 1004−13.

47. Xu B, Shen, H., Zhu, X., Li, G. (2011) Fast and accurate computation schemes for evaluating vibrational entropy of proteins. Journal of Computational Chemistry;32: 3188–93.

48. Hou T ZW, Case D. A., Wang W. (2008) Characterization of domain-peptide interaction interface: A case study on the amphiphysin-1 SH3 domain. Journal of Molecular Biology;376: 1201–14.

49. Gohlke H KC, Case D. A. (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes. Journal of Molecular Biology 330: 891-913.

50. Arnold GE, Ornstein RL (1997) Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: cytochrome P450BM-3. The Biophysical Journal;73: 1147-59.

51. Amadei A, Linssen, A.B., Berendsen, H. J. (1993) Essential dynamics of proteins. Proteins;17: 412-25.

52. Case DA, 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: 1668–88.

53. Laberge M, Yonetani, T. (2008) Molecular dynamics simulations of hemoglobin A in different states and bound to DPG: effector-linked perturbation of tertiary conformations and HbA concerted dynamics. Biophysical Journal 94: 2737–51.

54. Ichiye T, Karplus, M. (1991) Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins;11: 205–17.

55. Chen J, Wang, J., Zhu, W. (2014) Binding Modes of Three Inhibitors 8CA, F8A and I4A to AFABP Studied Based on Molecular Dynamics Simulation. Public Library of Science One;9: e99862.

56. Humphrey W, Dalke, A., Schulten, K. (1996) VMD: visual molecular dynamics. Journal of Molecular Graphics;14: 33-8, 27-8.

57. Huggins DJ, Sherman W, Tidor B (2012) Rational approaches to improving selectivity in drug design. Journal of Medicinal Chemistry;55: 1424-44.

58. Sayer JM, Louis JM (2009) Interactions of different inhibitors with active-site aspartyl residues of HIV-1 protease and possible relevance to pepsin. Proteins: Structure, Function, and Bioinformatics;75: 556-68.

59. Huang X, Britto MD, Kear-Scott JL, Boone CD, Rocca JR, Simmerling C, et al. (2014) The role of select subtype polymorphisms on HIV-1 protease conformational sampling and dynamics. Journal of Biological Chemistry;289: 17203-14.

60. Hornak V, Okur A, Rizzo RC, Simmerling C (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. Proceedings of the National Academy of Sciences of the United States of America;103: 915-20.

61. Mahanti M, Bhakat S, Nilsson UJ, Soderhjelm P (2016) Flap Dynamics in Aspartic Proteases: A Computational Perspective. Chemical Biology and Drug Design;88: 159-77.

62. Chen J, Liang Z, Wang W, Yi C, Zhang S, Zhang Q (2014) Revealing origin of decrease in potency of darunavir and amprenavir against HIV-2 relative to HIV-1 protease by molecular dynamics simulations. Scientific Reports;4: 6872.

63. Chen J (2016) Drug resistance mechanisms of three mutations V32I, I47V and V82I in HIV-1 protease toward inhibitors probed by molecular dynamics simulations and binding free energy predictions. RSC Advances;6: 58573-85.

64. Collins JR, Burt SK, Erickson JW (1995) Flap opening in HIV-1 protease simulated by 'activated'molecular dynamics. Nature Structural & Molecular Biology;2: 334-8.

65. Scott WR, Schiffer CA (2000) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. Structure;8: 1259-65.

66. Huang X, de Vera IM, Veloro AM, Blackburn ME, Kear JL, Carter JD, et al. (2012) Inhibitorinduced conformational shifts and ligand-exchange dynamics for HIV-1 protease measured by pulsed

EPR and NMR spectroscopy. Journal of Physical Chemistry B;116: 14235-44.

67. Meher BR, Wang Y (2012) Interaction of I50V mutant and I50L/A71V double mutant HIVprotease with inhibitor TMC114 (darunavir): molecular dynamics simulation and binding free energy studies. Journal of Physical Chemistry B;116: 1884-900.

68. Perryman AL, Lin JH, McCammon JA (2004) HIV-1 protease molecular dynamics of a wildtype and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Protein Science;13: 1108-23.

69. Karubiu W, Bhakat S, McGillewie L, Soliman ME (2015) Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics. Molecular BioSystems;11: 1061-6.

70. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2002) Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. Structure;10: 369-81.

71. Clemente JC, Coman RM, Thiaville MM, Janka LK, Jeung JA, Nukoolkarn S, et al. (2006) Analysis of HIV-1 CRF_01 A/E protease inhibitor resistance: structural determinants for maintaining sensitivity and developing resistance to atazanavir. Biochemistry;45: 5468-77.

72. Fakhar Z, Govender T, Maguire GEM, Lamichhane G, Walker RC, Kruger HG, et al. (2017) Differential flap dynamics in l,d-transpeptidase2 from mycobacterium tuberculosis revealed by molecular dynamics. Molecular Biosystems;13: 1223-34.

73. Goldfarb NE, Ohanessian M, Biswas S, McGee TD, Jr., Mahon BP, Ostrov DA, et al. (2015) Defective hydrophobic sliding mechanism and active site expansion in HIV-1 protease drug resistant variant Gly48Thr/Leu89Met: mechanisms for the loss of saquinavir binding potency. Biochemistry;54: 422-33.

74. Shen Y, Altman MD, Ali A, Nalam MN, Cao H, Rana TM, et al. (2013) Testing the substrateenvelope hypothesis with designed pairs of compounds. ACS Chemical Biology;8: 2433-41.

CHAPTER THREE

Disrupting the *β*-sheet interface of HIV PR to determine the effect of C- and N- terminal residue truncation on C-SA PR's stability: A MD Study

Siyabonga I. Maphumulo[1], Thavendran Govender[1], Glenn E. M. Maguire[1,2], Bahareh Honarparvar[1]*, and Hendrik G. Kruger[1]*

[1]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZuluNatal, Durban 4001, South Africa.

[2]School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4001, South Africa.

***Corresponding authors:** *E-mail: Honarparvar@ukzn.ac.za Fax: +27 31 2603091; Tel: +27 31 2608482*

Abstract

The HIV-1 PR possesses three key dimer interface regions; flap tips, active site and the *β*-sheet interface. Residues located at these regions are known to have a vital role in maintaining the stability of the protease as these residues are highly conserved for most drug-resistant strains. In this study, we attempt to disrupt the dimeric nature of the protease by truncating the C- and N- terminal residues on the South African subtype and thus inducing inactivation of the enzyme. Successful destabilization of HIV binding in this way will provide structural details at the molecular level about the *β*-sheet interface as potential target for HIV inhibition. MMGBSA calculations were performed and showed that truncation led to a substantial drop in binding affinities whereby the natural substrate reduced its binding free energy from -36.15 to -28.13 and -23.48 kcal/mol for single and double truncated systems, respectively. The loss of many key hydrogen bond interactions was seen as a very important factor which contributed to the collapse of the dimeric structure of HIV-1 PR. This study not only compliments several experimental findings recorded on the dimerization inhibition but also provides a detailed analysis on what key factors contribute significantly to the destabilization of the HIV-1 C-SA PR through disruption of the *β*-sheet interface.

Keywords: HIV-1 protease, Dimer interface, Molecular dynamics, Mutation, Truncation.

3.1 Introduction

Dimerization of HIV-1 PR monomers to form a homodimer has been found to be a key factor in ensuring that the protease enzyme achieve its catalytic activity[1, 2]. The HIV-1 PR monomers are held together by several polar and nonpolar interactions which contributes to its stability[1].

The dimer interface constitutes of the N- and C-terminal anti-parallel β-strands (residues 1-5 and 94-99, the β-sheet interface), active site (residues 24-29), helices (residues 84-93) and the flap tips (residues 48-54), which are found to be highly conserved regions of the protease[3-5].

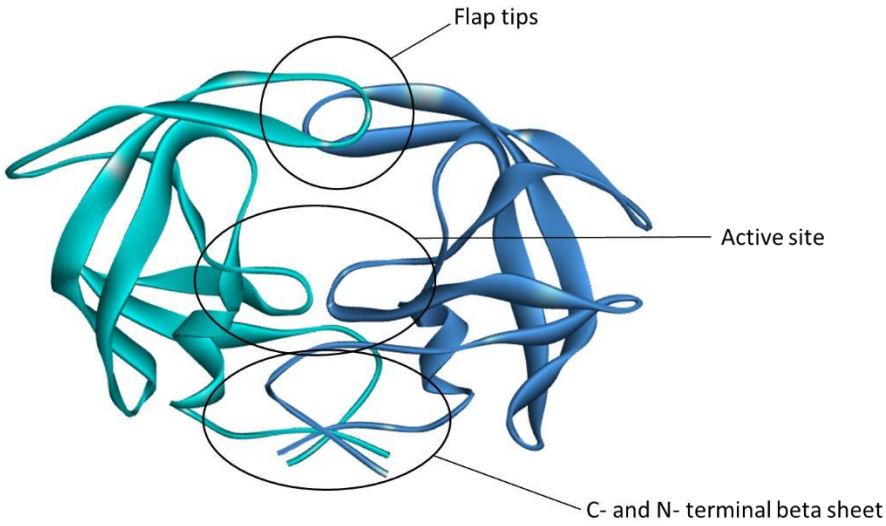


Figure 1. Illustration of three important regions of dimerization interface of HIV-1 PR.

Despite the fact that the monomeric state of HIV-1 PR is believed to possess structural stability, it is however inactive[6-8] indicating the significance of dimerization of HIV-1 PR subunits for proteolytic activity of the enzyme. Several studies have reported disruption of the dimeric state as a possible target for protease inhibition[4, 9-12]. These studies have highlighted alternative strategies for inhibitory mechanisms to active site–directed inhibition. There have been several studies reported on drug resistance of several active site inhibitors as a result of mutations occurring in the binding pocket of the protease[13]. These studies have shown that destabilization of a dimer interface can be a viable inhibition mechanism for HIV-1 PR.

Earlier reported studies on C- and N-terminally cross-linked peptides were designed to mimic the intertwined dimer interface of HIV PR[9, 10]. Despite their success in dimerization inhibitory activity, these agents displayed quite high molecular complexities[14]. Developments were later achieved when inhibitors of low molecular weight were eventually proposed, and these were

based on short 'interface' peptide segments. Some of these agents showed high inhibitory power against HIV PR as shown by their low nanomolar K_i values[14, 15].

Bannwarth *et al.*[4] developed improved lipopeptides that had several chains of different amino acid sequence with different lengths. The authors demonstrated that all of the cross-linked peptides with flexible spacers displayed less efficiency compared to the alkyl peptides[4, 10].

Another exploration of the dimer interface was performed by Naicker *et al.*[5] and also highlighted the disruption of the dimer interface for inactivation of the enzyme. The residue Phe99 situated at the terminal end of the chain was mutated to Ala to determine the contribution brought upon by the residue to the South African HIV-1 subtype C protease. The study demonstrated that no activity was attained by the Phe99Ala mutated protease and as such, the report suggested that the mutant displayed a monomeric nature because of a reduced β-sheet content.

It has been reported that one possible way of disrupting the dimer interface of HIV-1 PR is by truncation of the terminal residues from the β-sheet interface region[16]. The authors used a molecular dynamics approach to simulate different states of the subtype B wild-type complex and showed that by truncating four residues from each terminal ends of the HIV-1 PR subunits, a less compact structure is achieved, which they proposed should ultimately lead to HIV-1 PR deactivation.

Given these findings, our study utilized a molecular dynamics approach to analyse the effectiveness of removing the five terminal residues from both ends of one and both subunits of HIV-1 PR. The primary objective for the MD simulation was to determine the effect that truncation has on the binding affinity of the South African HIV-1 subtype C PR. Secondly, the interest of this study was also to further explore how other dimer interface regions get affected by the disruption at the β-sheet dimer interface. In addition to the apo and natural substrate bound systems, inhibitor-bound systems of atazanavir (ATV) and tipranavir (TPV) were also considered so as to explore how their inhibitory strengths are affected by truncation.

3.2 Computational Methods

3.2.1 Preparation of inhibitors/enzyme complex

The following PDB codes were used to obtain complex crystal structures from the Protein Data

Bank; 3U71 (apo/free), 3EL1 (ATV), 2O4P (TPV) and 1KJH (natural substrate). The same computational methods as employed for our recent paper[17] were also followed. Five residues from N- and C- terminals were selectively removed from subunit A to obtain single truncation and five residues were also removed from N and C terminals from both subunits to obtain a double-truncated system.

3.2.2 Setting up the system for Molecular Dynamics (MD) simulations

Prior to molecular dynamics simulations, the protonation states of residues of HIV-1 C-SA were assigned at pH=7.0 using the PropKa server[18, 19]. The ff99SB[20] and the general AMBER force field (GAFF)[21] were used to describe the inhibitor–protein and inhibitor–solvent interactions, respectively. The ligands were parameterized with Leap[21] implemented in Amber 14, using the general AMBERforce field (GAFF)[21] in Antechamber. MD simulations were performed using Amber 14 with the ff99SB[22] force field with the TIP3P[23] explicit water in acubic box with 8 Å distance around the complex. The positive charge of the complexes were neutralized with chloride ions. Partial Mesh Ewald (PME) [20] method was used to consider long range electrostatic forces (cutoff 12 Å). Additionally, the SHAKE algorithm[24] was used to constrain all bonds to hydrogen atoms. To reduce the overlapping atom interactions of initially prepared solvated complexes, energy minimization was performed in two stages. In the first stage only ions and water molecules were relaxed by 2000 step minimization process (1000 steps of steepest decent minimization followed by 1000 of conjugated gradient) using a restrained force of 500 kcal/mol on the solute. For the second stage of minimization, the whole system was relaxed by a 5000 step minimization process (2500 steps of desent minimization followed by 2500 of conjugated gradient). The minimized systems were gradually heated up from 0 to 300 K with a weak harmonic restraint of 10 kcal/mol to keep the solute fixed for 200 ps. Subsequently, the 2 ns constant pressure equilibrations at 300 K were performed. Lastly, 100 ns MD simulations without restriction were run at a constant temperature of 300 K and constant pressure of 1 atm.

3.2.3 Post-dynamics analysis

After all 100 ns MD simulations had finished, analyses for the root-mean square deviation (RMSD), root-mean square fluctuation (RMSF) and other various distance analyses were performed using the CPPTRAJ and PTRAJ modules[25] implemented in Amber 14.

3.2.4 Binding free energy calculations

Molecular Mechanics/Generalized Born Surface Area (MM-GBSA)[26] binding free energies of inhibitors were calculated using MM-PBSA program in Amber 14. For each complex, 400 snapshots were taken from the last 2000 ps of the 100 ns MD trajectory with an interval of 5 ps. The binding energy is represented as:

$$\Delta G = \Delta E_{ele} + \Delta E_{vdW} + \Delta G_{pol} + \Delta G_{nopol} - T\Delta S$$

The first two terms ΔEele and ΔEvdW are calculated by molecular mechanisms and represents the electrostatic and van der Waals interactions with the proteins in gas phase. The polar solvation free energy ΔGpol is calculated by using the MM-PBSA program[27] and it represents polar interactions with the solvent molecules. The ΔG_{nonpol} signifies non-polar solvation free energy and is obtained from the equation $\Delta G_{nonpolar} = \gamma SASA + \beta$ [28] . The solvent accessible surface area (SASA), the surface tension proportionality constant (γ) and the free energy of nonpolar solvation of a point solute (β), were set to 0.00542 kcal mol^{-1} $Å^{-2}$ and 0 kcal mol^{-1}, respectively[29]. The entropy contribution to the binding free energies were calculated using normal mode analysis[29-32] for the complexes from changes in the translational, rotational and vibrational entropy components (Table 1).

3.2.5 Per residue free energy decomposition analysis

To evaluate the contributions of individual residues to the total binding free energies of the inhibitors/PR complexes were computed using the MM-GBSA per residue free energy decomposition method[29, 33] in Amber 14[26]. All energy components including van der Waals, electrostatic, polar solvation and nonpolar solvation contributions were all calculated using 400 snapshots extracted from the last 2000 ps MD trajectories (Figure 5 and 6) of the 100 ns MD simulation.

3.3 Results and discussion

3.3.1 Stability and flexibility of the HIV-1 PR complexes

In order to gain insight into the structural drift of the simulated systems from the initial coordinates, the root mean square deviation (RMSD) of the Cα backbone was monitored throughout the 100 ns MD simulations.

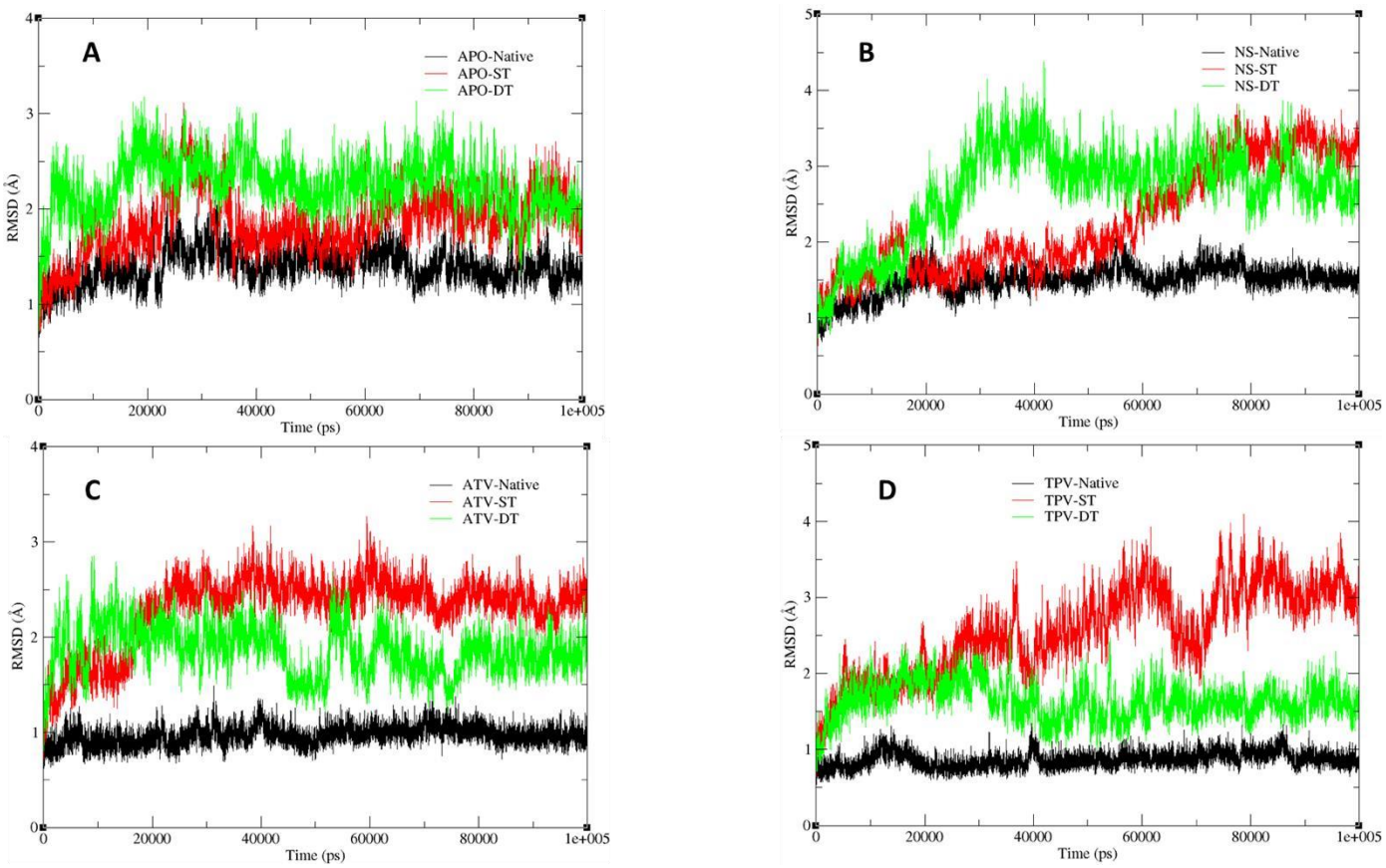


Figure 2. Root Mean Square Deviation of the backbone Cα atoms for the native (untruncated), single and double-truncated systems of (A) Apo enzyme, (B) Natural substrate (NS), (C) ATV complex and (D) TPV complex over the 100 ns MD trajectory.

The RMSD plot of the apo enzyme shows that the truncated systems displayed slightly higher RMSD values compared to the native form which displayed approximately 1.5 Å of RMSD on average. The RMSD displayed by the truncated systems of the natural substrate complex (Figure 2 B) and inhibitor-bound complexes (Figure 2 C and D) were higher than their native counterparts were. Three different MD simulations of Apo enzyme were carried out by varying initial atomic coordinates or initial atomic velocities. RMSD plots of these conformations showed that they displayed relatively similar stability (**Figure S1, Supplementary Material**).

3.3.2 Root-Mean Square Fluctuation (RMSF)

To understand the fluctuation of each residue over the 100 ns MD simulation, the root-mean square fluctuation (RMSF) was analysed. The analysis provided insight into the flexible regions of the protein for the native and truncated forms.

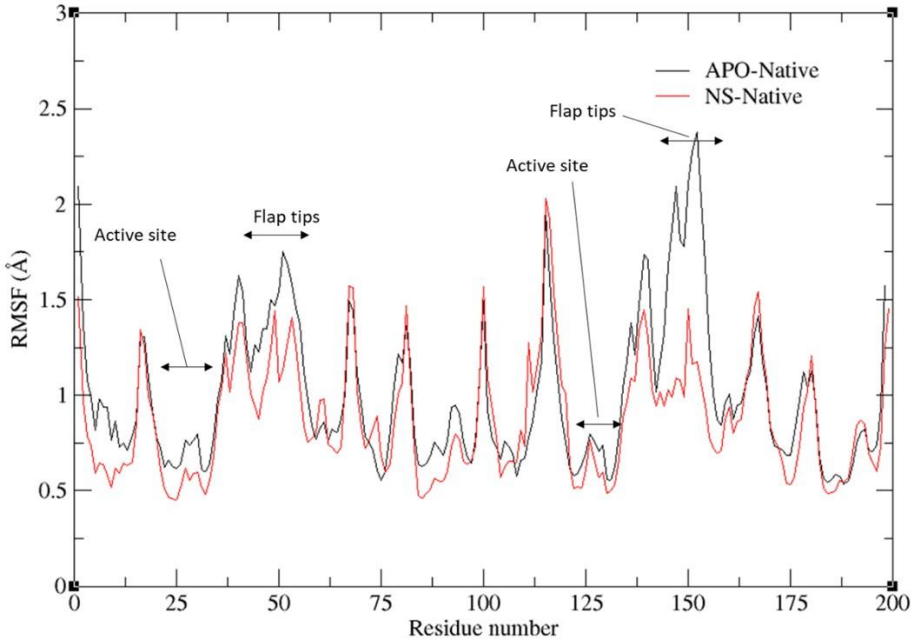


Figure 3. Root-Mean Square Fluctuation of backbone atoms against residue number for the APO and NS systems in native form over the 100 ns MD trajectory.

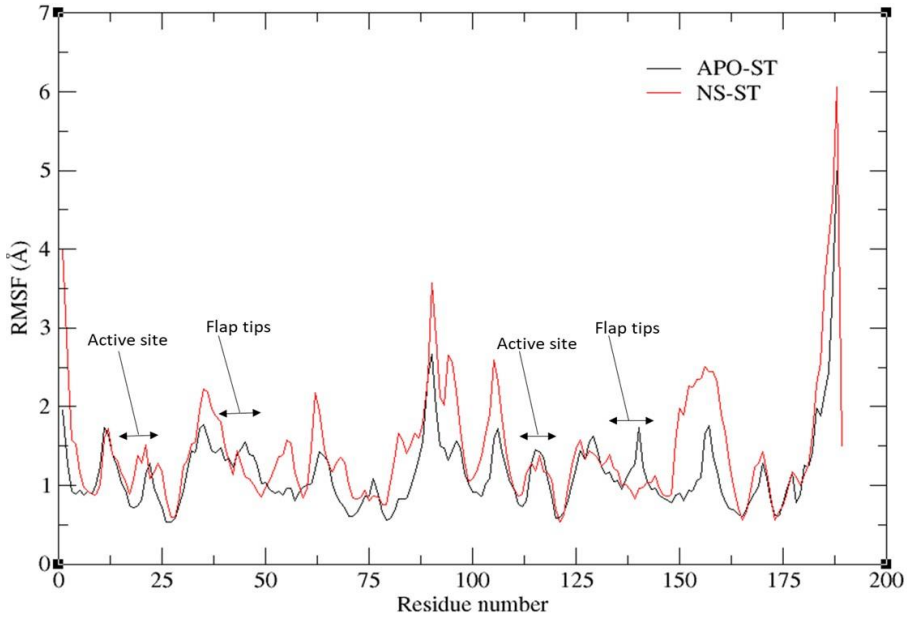


Figure 4. Root-Mean Square Fluctuation of backbone atoms against residue number for the APO and NS systems in single-truncated (ST) form over the 100 ns MD trajectory.

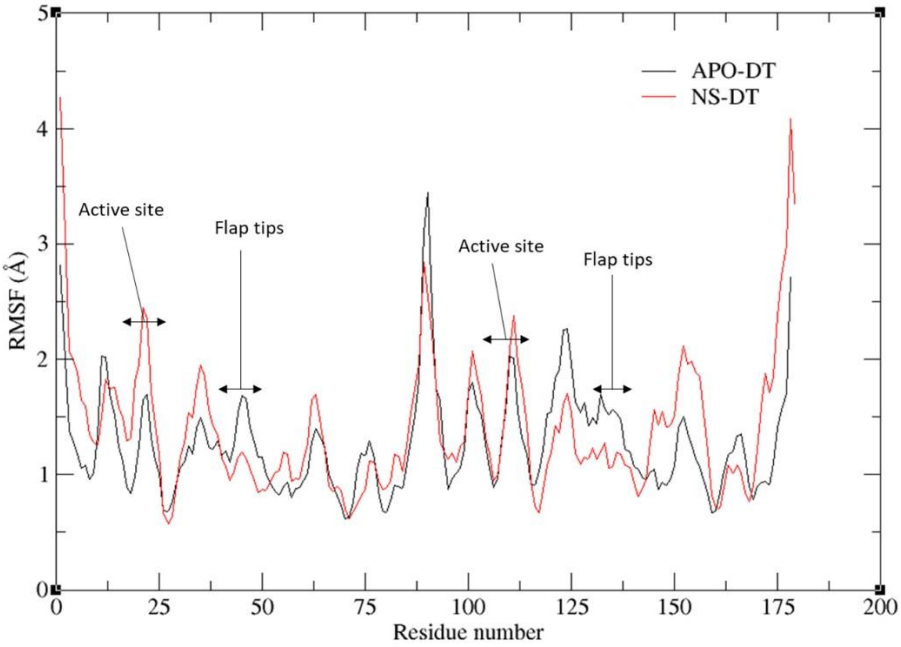


Figure 5. Root-Mean Square Fluctuation of backbone atoms against residue number for the APO and NS systems in double-truncated (DT) form over the 100 ns MD trajectory.

Plot of native forms for systems of APO and NS demonstrate that both systems displayed higher fluctuations for most residues particularly around the flap region. Figures 4 and 5 demonstrate that the active site residues gained significant fluctuations as a result of truncation at the N- and C-terminal residues. The RMSF values attained by the active site residues of the ST and DT forms of NS complex reached 1.69 and 2.49 Å, respectively, while that for the native enzyme (untruncated) reached approximately 0.6 Å. This indicates that the truncated forms experience much more movement at the active site.

The TPV-ST result shows that when only chain A is truncated, most residues in chain A displayed relatively higher fluctuations as indicated in Figure S3 compared to Figure S2 of **Supplementary Material**. The increased fluctuation observed for the truncated systems corresponds well with the decrease in binding free energies recorded in Table 1. Fluctuations of the Cα atom for each residue of the double-truncated systems are depicted in Figure S4 (**Supplementary Material**). According to the RMSF plot in figure S4, it is evident that the flap tip residues displayed a decrease in fluctuations and this supports the idea that the flap tip residues became more rigid as a result of double truncation. The loss in flexibility (increase in rigidity) at the flap tips ensured that the residues in this region supplement for the loss of dimerization at the β-sheet interface.

3.3.3 MM-GBSA binding energy analyses

To understand the effects of truncation of the N and C-terminal residues on binding inhibitors to the C-SA truncated PR, the binding free energies and their components (ΔE_{vdw}, ΔE_{elec}, ΔG_{gas}, ΔG_{polar}, $\Delta G_{nonpolar}$ $T\Delta S$ and ΔG_{bind}) were calculated for all systems using the MM-GBSA approach and normal mode analysis (NMA).

As shown in Table 1, the calculated binding free energies of the truncated systems demonstrate a decrease compared to their respective native systems. The double-truncated system of the natural substrate (NS) seemed to display a greater decrease in binding free energy (ΔG_{bind}), with an energy difference of 12.67 kcal/mol while that of ST differed by approximately 8.02 kcal/mol. It is thought that the negative value (-68.27 kcal/mol) observed for polar solvation free energy (ΔG_{polar}) of the double-truncated NS Native case was due to the highly unfavourable intermolecular electrostatic interactions of this substrate complex.

Table 1. Calculated binding free energies [ΔG_{bind}] and its components for two HIV-1 protease inhibitors and the natural substrate (NS) on C-SA PR. All data values are in kcal/mol.

Complex[a]	ΔE_{vdw}	ΔE_{elec}	ΔG_{polar}	$\Delta G_{nonpolar}$	$-T\Delta S$	ΔH	ΔG_{bind}
NS Native	-86.49	4.69	24.15	-11.76	33.26	-69.41	-36.15
ST	-83.65	12.01	23.35	-11.50	31.66	-59.79	-28.13
DT	-75.70	97.26	-68.27	-10.122	33.39	-56.84	-23.48
ATV Native	-76.27	-6.948	29.25	-9.66	29.26	-63.63	-34.37
ST	-57.54	-4.067	21.90	-6.99	25.21	-46.70	-21.49
DT	-71.78	-10.36	33.41	-9.11	29.65	-57.84	-28.20
TPV Native	-59.96	-37.93	58.49	-8.30	27.79	-47.70	-19.91
ST	-45.93	-6.54	25.14	-6.50	24.06	-33.83	-9.77
DT	-55.14	-25.53	46.50	-7.31	26.32	-41.49	-15.17

Since the electrostatic interactions were unfavourable (97.26 kcal/mol) it makes sense that an overcompensation by the desolvation of polar groups[34-36] was necessary.

The truncated systems of the inhibitor complexes (ATV and TPV) also demonstrated a decrease in the calculated binding energy compared to their native counterparts. The single truncated system of the ATV complex displayed a binding free energy of -21.49 kcal/mol, which is an

energy difference of approximately 12.88 kcal/mol and the energy difference attained by the double-truncated system was 6.17 kcal/mol.

A decline of the binding free energy for the truncated systems was also displayed for TPV as Table 1 clearly demonstrates an energy difference of 10.14 and 4.74 kcal/mol for ST and DT, respectively. The observed decrease in binding free energy for the truncated systems of the natural substrate, ATV and TPV complexes is related to the disruption of the β-sheet dimer interface. These results also correlate with the distance analysis carried out (Figure 10) for two residues (Gln92 and Gln92') situated close to the terminal residues at the β-sheet and this distance was found to increase when truncation of terminal residues was done. A decrease in binding free energy for these systems shows that the reduction of interactions known to occur at the β-sheet dimer interface has an important impact in destabilizing the protease dimer.

3.3.4 Per-residue free energy decomposition analysis

In order to gain a further insight into the binding energies for the studied systems and the role of particular residues involved in the binding mechanism, binding free energy was decomposed into contributions from various residue pairs of certain regions. Figures 6, 7 and 8 show contributions from residues in the active site as well as the flap tip region, these being the two other dimer interface regions.

According to Figure 6, the two catalytic aspartates for chain A and B display different energy contributions as the aspartate of chain B aspartate offers far greater electrostatic contributions compared to that of chain A. There seems to be a symmetrical contribution from these two aspartates of HIV-1 PR when the natural substrate-complex is truncated and also the electrostatic contribution is greater compared to the untruncated (native) system. The substantial rise in electrostatic contribution by the catalytic aspartates is confirmed by the high values displayed in the RMSF plot (Figure 5) with fluctuation reaching values of approximately 2.5 Å compared to approximately 0.7 Å of the native form.

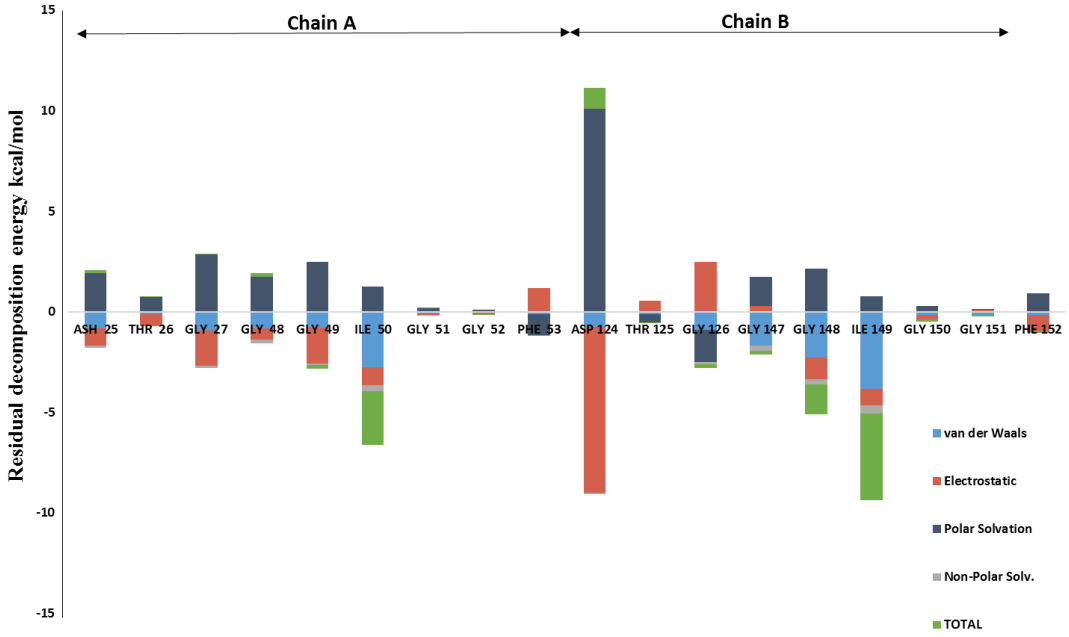


Figure 6. Plot of per-residue decomposition analysis for the native natural substrate complex (NS-Native).

It is clear that only Ile50 for chain A had a significant contribution to the binding of the natural substrate to the HIV-1 PR in the native form. This effect was also displayed by Ile50 of chain B (Ile149) with a total energy contribution of approximately -9.5 kcal/mol as indicated in Figure 7 and 8. The two Ile50 residues also made notable contribution to the van der Waals interactions in the native form. Finally, the destabilization effect of both Asp 25 and Asp 124 for NS-ST and NS-DT explains the decrease in binding energies (Table 1) for these systems.

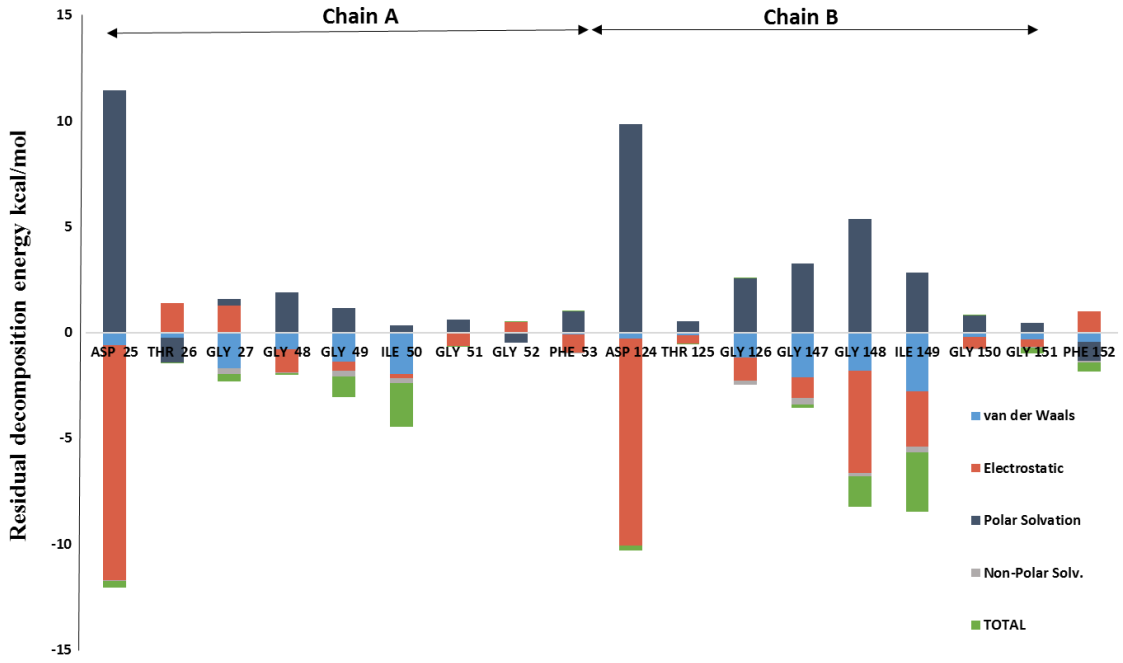


Figure 7. Plot of per-residue decomposition for the single-truncated system of the natural substrate complex (NS-ST).

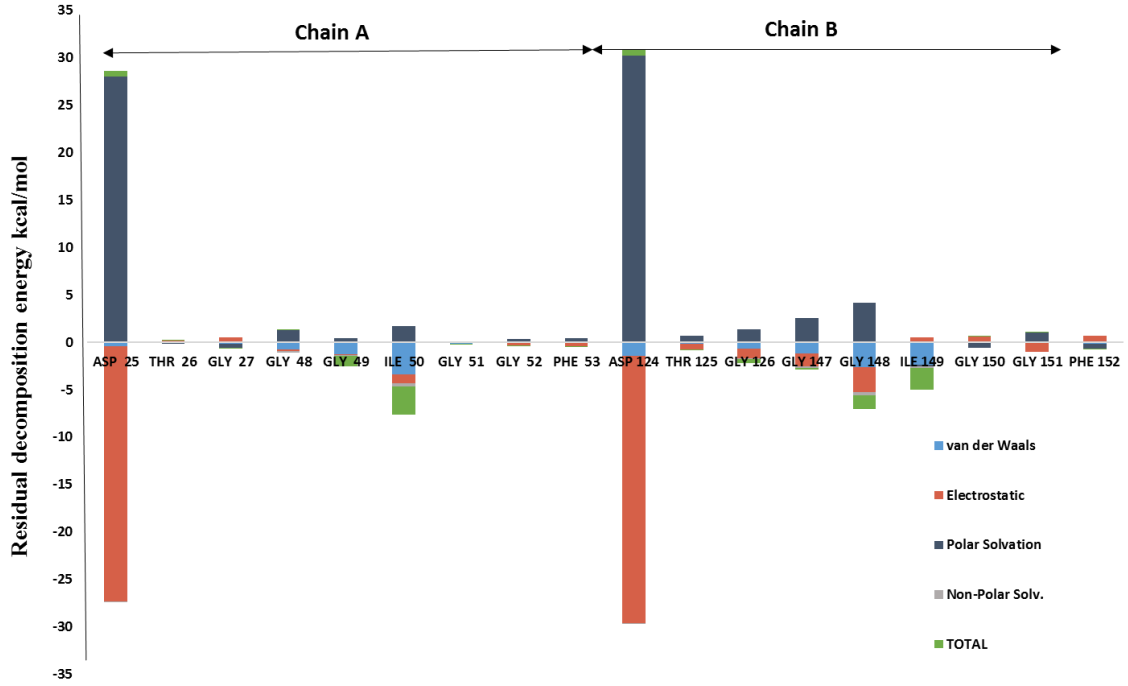


Figure 8. Per-residue decomposition for the double-truncated system of the NS complex (NSDT).

3.3.5 Radius of Gyration (Rg)

Radius of Gyration (Rg) over the 100 ns MD simulation was used to better explain the compactness of the native and truncated systems as a result of the overall folding or unfolding. The distribution of the radius of gyration values is represented in Figure 9 A and B.

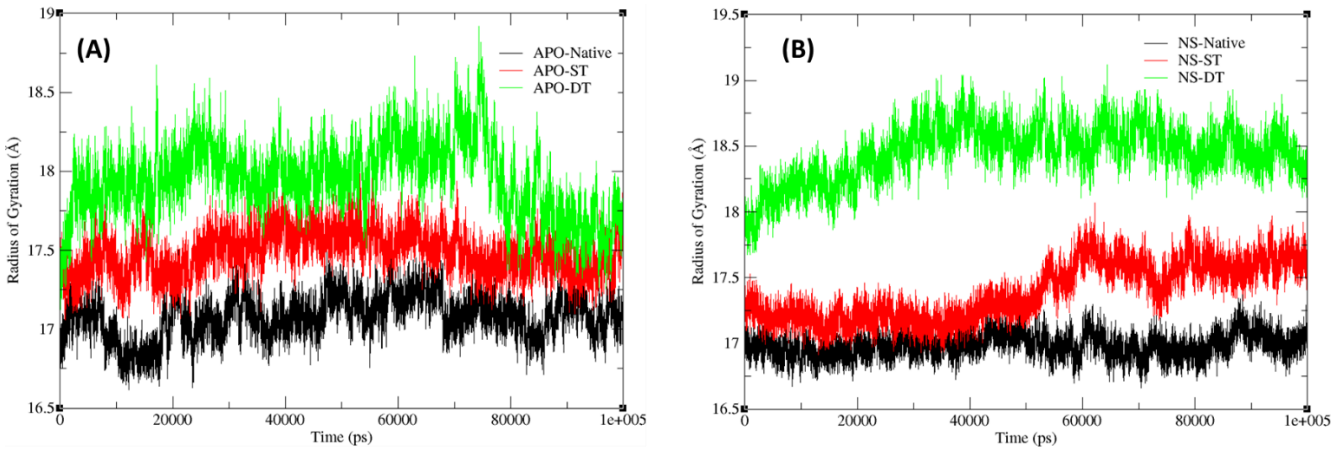


Figure 9. The plot of radius of gyration for three different systems; native, ST and DT of (A) apo enzyme, (B) natural substrate complex.

It is quite evident from Figure 9 that the truncated systems of all prepared complexes demonstrated significantly higher Rg values when compared to the native systems. The average Rg values for the APO systems increase from 17.08 Å for the native form to 17.48 and 17.95 for ST and DT forms, respectively. A similar trend was also observed for the NS complex systems whereby average Rg values of 16.98; 17.39; and 18.45 for native, ST and DT systems, respectively. The increase in values attained by the truncated systems indicates less stability and as a result, less compactness. These results correspond with the binding energies reported in Table 1 whereby the truncated systems displayed reduced binding energies compared to the native system. The reduced compactness of truncated systems may be attributed to the collapse of the dimer interface which influenced the destabilization in the whole structure of the complexes by breaking of essential hydrogen bonds in the dimer interface. The truncated systems of inhibitor-bound complexes also displayed less compactness as highlighted by an increase in Rg values for ATV and TPV in Figure S5 (**Supplementary Material**).

3.3.6 Distance Gln92-Gln92'

The distance analysis was carried out to determine the extent to which residues near the β-sheet drift apart as result of truncation. Glutamine at position 92 and 92' were chosen as these were conserved even for the truncated systems.

Figure 10 A shows that the two residues started to drift apart for the native form of APO system after 20 ns when a single truncation was performed. The distance was observed to increase with even greater fluctuation when truncation was performed in both chains of the protease. Figure 10 B does demonstrate an increase in the distance after 50 ns for the single truncated system of the natural substrate-bound complex. An even greater separation of the residues was observed when double truncation of the C- and N-terminal residues was done whereby the residues separated by a distance of above 36 Å.

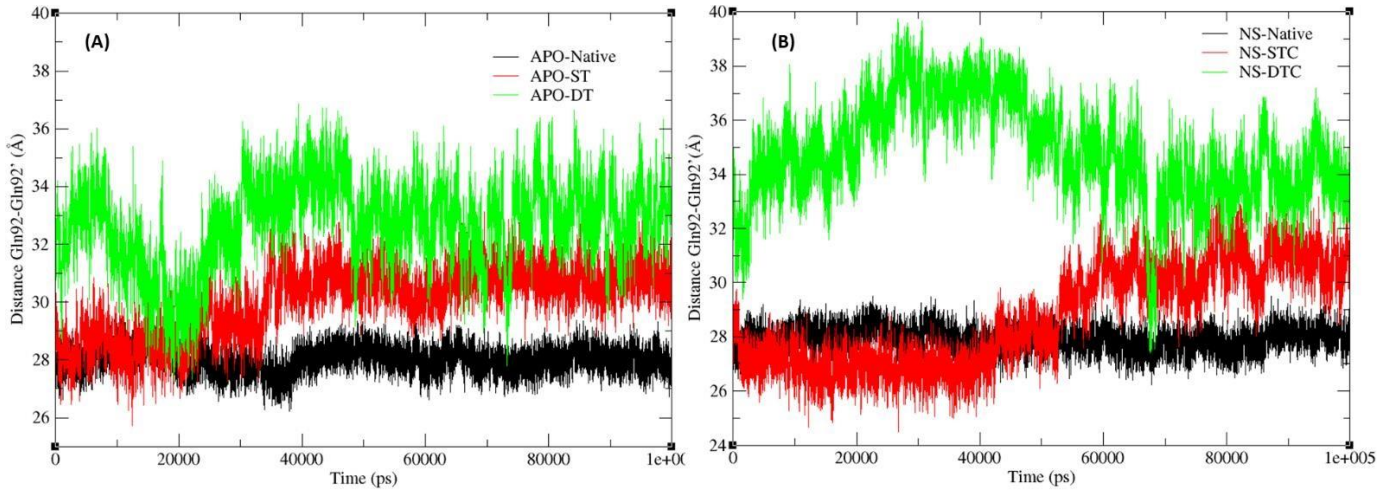


Figure 10. The plot of distance between Gln92 and Gln92' for situated near the β-sheet for the different systems, Native, ST and DT of (A) Apo enzyme, (B) natural substrate complex.

The inhibitor-bound systems also showed a similar display of the increase in distance between the two residues (**Figure S7, Supplementary Material**). This analysis confirms the trend displayed in the MM-GBSA binding energies recorded in Table 1 where the double-truncated systems had reduced energy values compared to the untruncated ones.

3.3.7 Distance Ile50-Ile50'

A distance between the Cα atoms of the flap tip residues (I50 & I50') from both subunits was examined in order to determine the extent of flap opening. The distances obtained for all systems did not display any particular trend as there were some considerable overlaps demonstrated. The flap tip distances displayed by all systems of the natural substrate complex overlapped throughout the 100 ns MD simulation. It is demonstrated in Figure 11 B that the double truncated system showed the lowest distance between the flap tips for most part of the simulation.

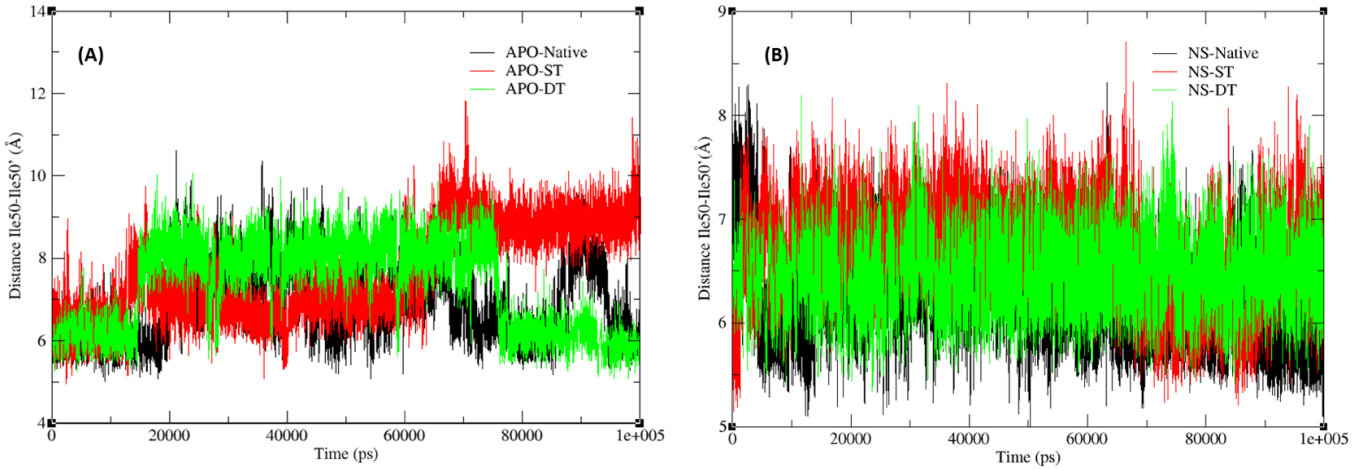


Figure 11. The distance (Å) between the flap tip residues (Ile50 and Ile50') for three different systems; native, ST and DT of the (A) apo form and (B) natural substrate complex obtained from a 100 MD simulation.

As was pointed out by previous studies,[37, 38] this metric may sometimes not be an adequate parameter for analysis of flap opening as flap tip distance is affected by the curling in of the flap tips and also the asymmetry of the flaps. However, the flap tip distance measured for the inhibitor bound complexes for ATV and TPV show that the double truncated systems attained reduced values when compared to those of native systems as indicated in Figure S6 (**Supplementary Material**). Plots of flap tip distance are represented in This is as a result of the flap tips of the PR coming even closer to compensate for the loss of dimerization on one of it dimer interface regions as the inhibitors used in this study are those that are designed to bind onto the active site region and induce a closed conformation. Other supporting evidence for this result is the per residue decomposition analysis of the truncated systems of ATV (**Figure S8-S10, Supplementary Material**) where despite the loss of significant binding contributions

by the catalytic aspartate residues, the flap tip residues still exhibited significant per residue contributions towards the overall binding process.

3.3.8 Hydrogen Bond (Arg8-Asp29'/ Asp29-Arg8')

Residues Arg8 of subunit A and Asp29 of subunit B have been shown to form a salt-bridge interaction and positioned at the outer edge of the cavity[16] and as such, the distance between them was monitored throughout the simulation to determine the effect of truncation on the interaction of these residues. Figure 12 shows where the two residues are located and gives a clear indication of how the separation of these residues can represent the separation of the two subunits. To better analyse the effect of this salt-bridge, we also performed HB analysis on the other edge of the cavity using Asp29-Arg8' as interacting residues. The strength of the hydrogen bond interaction at this salt bridge is recorded in Table 2 and 3.

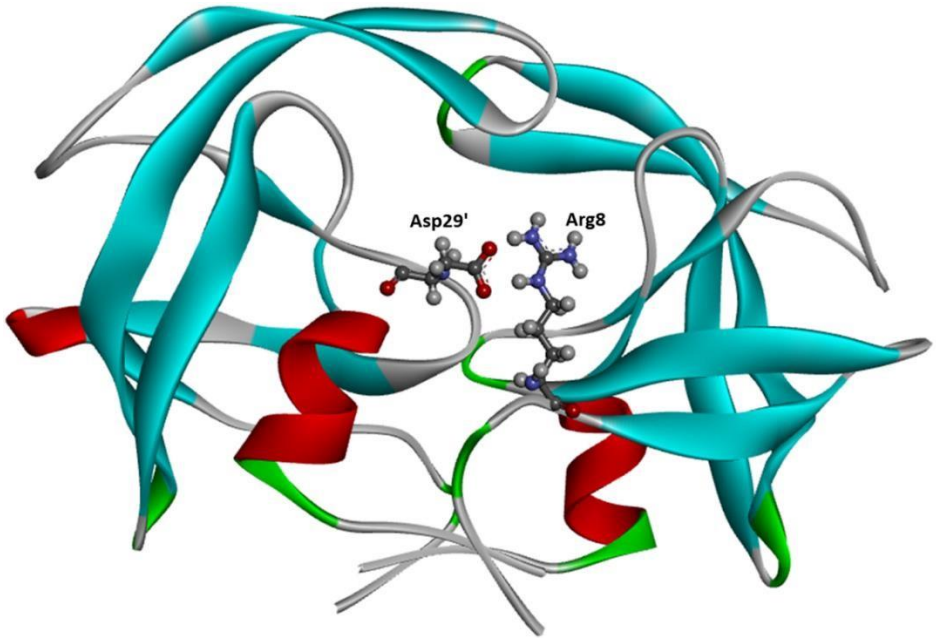


Figure 12. A 3-D illustration of the two residues (Arg8 and Asp29') that form a salt-bridge in HIV-1 PR.

Table 2. Hydrogen Bond interaction formed between Asp29/29' and Arg8/8' of APO enzyme for the native and truncated systems.

#Acceptor	Donor	Frames	Frac	AvgDist	AvgAng
APO Native					
ASP_29@OD2	ARG_107@NH2	44408	0.4441	2.7822	160.9139
ASP_29@OD1	ARG_107@NH2	43006	0.4301	2.7811	160.9609
ASP_29@OD1	ARG_107@NE	30597	0.306	2.8664	160.1509
ASP_29@OD2	ARG_107@NE	29955	0.2995	2.8674	160.0251
ASP_128@OD1	ARG_8@NH2	5393	0.0539	2.8028	160.4957
ASP_128@OD2	ARG_8@NE	4831	0.0483	2.8495	160.111
ASP_128@OD2	ARG_8@NH2	3337	0.0334	2.8207	159.1025
ASP_128@OD1	ARG_8@NE	2977	0.0298	2.8514	159.4549
ASP_128@OD2	ARG_8@NH1	741	0.0074	2.8011	152.4173
ASP_128@OD1	ARG_8@NH1	669	0.0067	2.8181	153.0649
ASP_128@OD1	ARG_8@NH2	644	0.0064	2.82	158.0249
ASP_29@OD1	ARG_107@NH1	620	0.0062	2.8123	158.2965
APO ST					
ASP_24@OD2	ARG_97@NH2	807	0.0079	2.798	160.1247
ASP_24@OD1	ARG_97@NE	688	0.0067	2.8376	158.7441
APO DT					
ARG_92@O	ARG_92@NH1	919	0.009	2.8445	160.8413

Frac (/100), AvgDist (Å), AvgAng (°)

Table 2 shows that the native form of APO native system displayed a relatively high number of hydrogen bonds between Arg8/8' and Asp29/Asp29'. However, both truncated systems showed fewer hydrogen bonding interactions. The percentage of these bonds forming during the simulation is represented by the fraction (Frac) which indicates the proportion of the total time during which the hydrogen bond was detected. The truncated systems displayed lower fractions when compared to the native forms for APO and NS**.** The double truncated system of NS complex completely lost this salt bridge interaction as shown in Table 3 and also confirmed by the increased distance between the two residues.

Table 3. Hydrogen Bond interaction formed between Asp29/29' and Arg8/8' of NS-complex for the native and truncated systems.

#Acceptor	Donor	Frames	Frac	AvgDist	AvgAng
NS Native					
ASP_29@OD1	ARG_107@NH2	46673	0.4667	2.77	162.0756
ASP_128@OD1	ARG_8@NH2	46435	0.4643	2.7593	160.1163
ASP_128@OD2	ARG_8@NH2	44957	0.4496	2.7581	161.162
ASP_29@OD2	ARG_107@NH2	43667	0.4367	2.7737	161.6128
ASP_128@OD2	ARG_8@NE	37259	0.3726	2.856	159.3066
ASP_29@OD2	ARG_107@NE	33066	0.3307	2.8606	158.9933
ASP_128@OD1	ARG_8@NE	32525	0.3252	2.8576	160.6622
ASP_29@OD1	ARG_107@NE	32449	0.3245	2.8607	158.8056
ASP_128@OD1	ASP_128@N	558	0.0056	2.8868	139.3175
NS ST					
ASP_24@OD2	ARG_97@NH2	54297	0.543	2.7776	160.2575
ASP_24@OD1	ARG_97@NE	33885	0.3388	2.8606	161.787
ASP_24@OD1	ARG_97@NH2	32513	0.3251	2.8082	160.8726
ASP_24@OD1	ARG_97@NH1	4468	0.0447	2.8425	150.5733
ASP_24@OD2	ARG_97@NE	1933	0.0193	2.8924	145.3985
ASP_24@OD1	ARG_97@NH2	1406	0.0141	2.8272	155.0197

Frac (/100), AvgDist (Å), AvgAng (°)

To gain a better insight into the interaction of Asp29 and Arg8' we measured the distance between the Cα atoms of these residues to determine if there is any drifting occurring as a result of truncation. Figure 13 shows that the distance between Arg8 and Asp29' for the apo enzyme seemed to slightly increase due to truncation. The observed distance for the native form was between 6 and slightly over 8 Å while upon truncation the distance was observed to increase to values of 12 Å and above. The double-truncated system of the NS complex displayed very high distances as compared to its native state and this clearly shows that truncation led to the residues (Arg8 and Asp29') drifting apart.

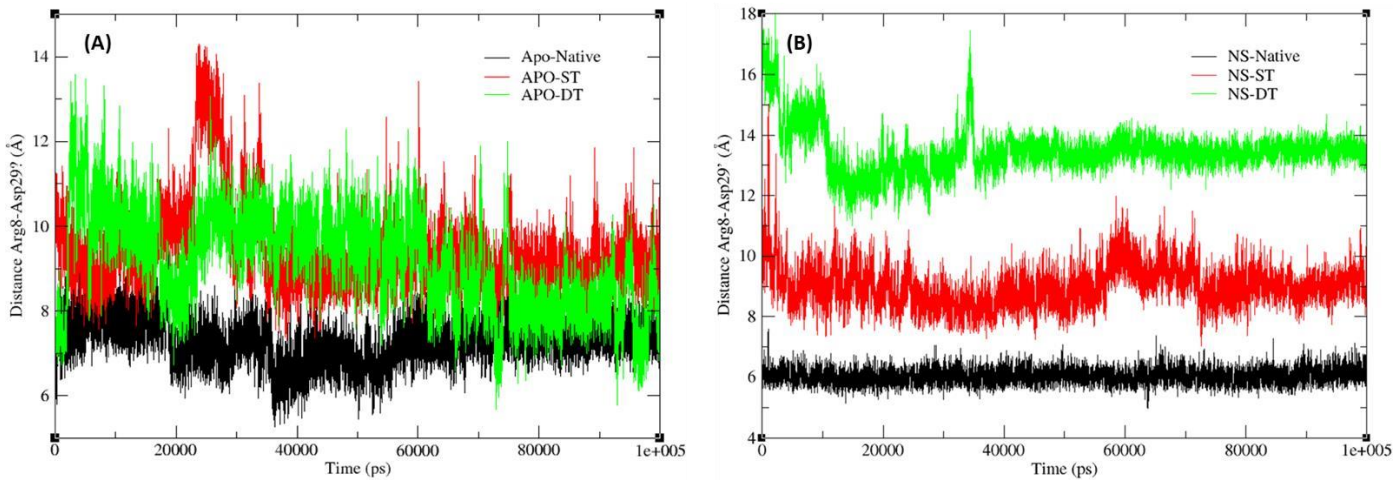


Figure 13. Distance of Arg8 and Asp29' measured during the 100 ns MD simulation for the native and truncated forms of all systems studied.

It is quite evident that upon truncation, the distance between the two residues tends to increase as a result of reduced flexibility of the HIV-1 PR. The increase in the Arg8-Asp29' distance give an indication that the dimeric state of the protease collapses and also this could very well be associated with the decrease in binding energy values for the truncated systems as observed in Table 1.

3.4 Conclusion

Although many studies have looked into the β-sheet interface region as being a possible target for HIV-1 PR inhibition, we believe that not sufficient evaluation of what impact the disruption of this region has on other dimer interface regions has been recorded. The detailed analyses performed in this study demonstrate that truncation of C- and N-terminal residues brings about destabilization to the dimeric structure of HIV-1 C-SA PR, which is the first study done for this particular subtype. Destabilization of the dimeric form of HIV-1 PR led to many key hydrogen bond interactions to break which then led to poor binding pattern for the substrate as well as inhibitors. The results clearly indicate that less compactness is achieved as a result of truncation of β-sheet residues.

This study revealed that disruption of the β-sheet interface due to truncation induces fluctuation at the active site region, but the flap tips still managed to keep their distance close particularly when truncation was performed on both chains. We have also been able to demonstrate that the inhibitors bound to a double-truncated protease displayed a larger increase in binding free

energy (approximately 6 kcal/mol) relative to single truncation. This means that the inhibitorbound complexes prefer a symmetrical conformation within the protease. Truncation leads to a drop in Gibbs free binding energy as result of loss of key molecular interactions in the active site region as was shown in the per-residue decomposition analysis. The other contributing factor also discovered is the high fluctuation of the catalytic residues after truncation of the C- and N-terminal residues which led to the substrate and inhibitors not being able keep ideal binding interactions within the active site region. This study therefore provides insight at the molecular level on what happens when dimerization of HIV is disrupted at the β-sheet interface.

In terms of compensating for the loss of one key dimer interface region, our results showed that despite the increase in fluctuation of catalytic aspartates, they seem to have a high electrostatic contribution towards binding free energy as a result of truncation. Future studies should involve MD of a suitable dimerization inhibitor positioned at the β-sheet region to determine the nature of the interactions involved.

Acknowledgments

This work was supported by the University of KwaZulu Natal, South African National Research Foundation and the Centre for High Performance Computing CHPC Cape Town.

Conflict of Interest

The authors declare no financial and academic conflict of interest.

References

[1] Koh, Y., Matsumi, S., Das, D., Amano, M., Davis, D. A., Li, J., Leschenko, S., Baldridge, A., Shioda, T., and Yarchoan, R. (2007) Potent inhibition of HIV-1 replication by novel non-peptidyl small molecule inhibitors of protease dimerization, Journal of Biological Chemistry 282, 28709-28720.
[2] Bowman, M. J., and Chmielewski, J. (2002) Novel strategies for targeting the dimerization interface of HIV protease with cross-linked interfacial peptides, Peptide Science 66, 126-133.
[3] Zutshi, R., Shultz, M. D., Ulysse, L., Lutgring, R., Bishop, P., Schweitzer, B., Vogel, K., Franciskovich, J., Wilson, M., and Chmielewski, J. (1998) Inhibiting the dimerization of HIV-1 protease, Synlett 1998, 1040-1044.
[4] Bannwarth, L., Rose, T., Dufau, L., Vanderesse, R., Dumond, J., Jamart-Grégoire, B., Pannecouque, C., De Clercq, E., and Reboud-Ravaux, M. (2008) Dimer disruption and

monomer sequestration by alkyl tripeptides are successful strategies for inhibiting wild-type and multidrug-resistant mutated HIV-1 proteases, Biochemistry 48, 379387.
[5] Naicker, P., Seele, P., Dirr, H. W., and Sayed, Y. (2013) F99 is critical for dimerization and activation of South African HIV-1 subtype C protease, The protein journal 32, 560-567.
[6] Huang, D., and Caflisch, A. (2012) How does darunavir prevent HIV-1 protease dimerization?, Journal of chemical theory and computation 8, 1786-1794.
[7] Levy, Y., Caflisch, A., Onuchic, J. N., and Wolynes, P. G. (2004) The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations, Journal of molecular biology 340, 67-79.
[8] Levy, Y., and Caflisch, A. (2003) Flexibility of monomeric and dimeric HIV-1 protease, The Journal of Physical Chemistry B 107, 3068-3079.
[9] Babé, L. M., Rosé, J., and Craik, C. S. (1992) Synthetic "interface" peptides alter dimeric assembly of the HIV 1 and 2 proteases, Protein Science 1, 1244-1253.
[10] Zutshi, R., Franciskovich, J., Shultz, M., Schweitzer, B., Bishop, P., Wilson, M., and Chmielewski, J. (1997) Targeting the dimerization interface of HIV-1 protease: inhibition with cross-linked interfacial peptides, Journal of the American Chemical Society 119, 4841-4845.
[11] Davis, D. A., Tebbs, I. R., Daniels, S. I., Stahl, S. J., Kaufman, J. D., Wingfield, P., Bowman, M. J., Chmielewski, J., and Yarchoan, R. (2009) Analysis and characterization of dimerization inhibition of a multi-drug-resistant human immunodeficiency virus type 1 protease using a novel size-exclusion chromatographic approach, Biochemical Journal 419, 497-506.
[12] Hayashi, H., Takamune, N., Nirasawa, T., Aoki, M., Morishita, Y., Das, D., Koh, Y., Ghosh, A. K., Misumi, S., and Mitsuya, H. (2014) Dimerization of HIV-1 protease occurs through two steps relating to the mechanism of protease dimerization inhibition by darunavir, Proceedings of the National Academy of Sciences 111, 12234-12239.
[13] Gustchina, A., and Weber, I. T. (1991) Comparative analysis of the sequences and structures of HIV-1 and HIV-2 proteases, Proteins: Structure, Function, and Bioinformatics 10, 325-339.
[14] Hwang, Y. S., and Chmielewski, J. (2005) Development of low molecular weight HIV-1 protease dimerization inhibitors, Journal of medicinal chemistry 48, 2239-2242.
[15] Schramm, H. J., Rosny, E. D., Reboud-Ravaux, M., Büttner, J., Dick, A., and Schramm, W. (1999) Lipopeptides as dimerization inhibitors of HIV-1 protease, Biological chemistry 380, 593-596.
[16] Dayer, M. R., and Dayer, M. S. (2013) Whiskers-less HIV-protease: a possible Way for HIV-1 deactivation, Journal of biomedical science 20, 67.
[17] Maphumulo, S., Halder, A., Govender, T., Maseko, S., Maguire, G., Honarparvar, B., and Kruger, G. (2018) Exploring the flap dynamics of the South African subtype C HIV protease in presence of FDA-approved inhibitors: MD study, Chemical Biology and Drug Design. (submitted)
[18] Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) Very fast prediction and rationalization of pKa values for protein–ligand complexes, Proteins: Structure, Function, and Bioinformatics 73, 765-783.
[19] Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values, Proteins: Structure, Function, and Bioinformatics 61, 704-721.

[20] Harvey, M., and De Fabritiis, G. (2009) An implementation of the smooth particle mesh Ewald method on GPU hardware, Journal of Chemical Theory and Computation 5, 2371-2377.
[21] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, Journal of computational chemistry 25, 1157-1174.
[22] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins: Structure, Function, and Bioinformatics 65, 712-725.
[23] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, The Journal of chemical physics 79, 926-935.
[24] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of nalkanes, Journal of Computational Physics 23, 327-341.
[25] Roe, D. R., and Cheatham III, T. E. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data, Journal of chemical theory and computation 9, 3084-3095.
[26] Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and Case, D. A. (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate− DNA helices, Journal of the American Chemical Society 120, 9401-9409.
[27] Onufriev, A., Bashford, D., and Case, D. A. (2000) Modification of the generalized Born model suitable for macromolecules, The Journal of Physical Chemistry B 104, 37123720.
[28] Gohlke, H., and Case, D. A. (2004) Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf, Journal of computational chemistry 25, 238-250.
[29] Gohlke, H., Kiel, C., and Case, D. A. (2003) Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes, Journal of molecular biology 330, 891-913.
[30] Genheden, S., and Ryde, U. (2012) Will molecular dynamics simulations of proteins ever reach equilibrium?, Physical Chemistry Chemical Physics 14, 8662-8677.
[31] Kopitz, H., Cashman, D. A., Pfeiffer-Marek, S., and Gohlke, H. (2012) Influence of the solvent representation on vibrational entropy calculations: Generalized born versus distance-dependent dielectric model, Journal of computational chemistry 33, 10041013.
[32] Xu, B., Shen, H., Zhu, X., and Li, G. (2011) Fast and accurate computation schemes for evaluating vibrational entropy of proteins, Journal of computational chemistry 32, 3188-3193.
[33] Hou, T., Zhang, W., Case, D. A., and Wang, W. (2008) Characterization of domain–peptide interaction interface: a case study on the amphiphysin-1 SH3 domain, Journal of molecular biology 376, 1201-1214.
[34] Fiorucci, S., and Zacharias, M. (2010) Prediction of protein-protein interaction sites using electrostatic desolvation profiles, Biophysical journal 98, 1921-1930.
[35] Kar, P., Lipowsky, R., and Knecht, V. (2013) Importance of polar solvation and configurational entropy for design of antiretroviral drugs targeting HIV-1 protease, The Journal of Physical Chemistry B 117, 5793-5805.

[36] Kannt, A., Lancaster, C. R. D., and Michel, H. (1998) The role of electrostatic interactions for cytochrome c oxidase function, Journal of bioenergetics and biomembranes 30, 81-87.
[37] Karthik, S., and Senapati, S. (2011) Dynamic flaps in HIV-1 protease adopt unique ordering at different stages in the catalytic cycle, Proteins: Structure, Function, and Bioinformatics 79, 1830-1840.
[38] Scott, W. R., and Schiffer, C. A. (2000) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance, Structure 8, 12591265.

CHAPTER FOUR

Summary and conclusion

There are a reasonable number of studies that have calculated the binding free energies of commercial drugs in complex with HIV PR. The nature of the interactions between the inhibitors of these drugs and the HIV PR is not well studied. The fundamental aim of this work was to explore how flap dynamics is affected by various molecular interactions induced by the differing in binding affinities of inhibitors. The study also looked into the factors that seem to influence the affinity of protease inhibitors towards the protease. In addition, disruption of the dimeric nature of the protease was performed to induce less stability on the overall conformation and thus leading to inactivation of the enzyme.

In chapter two (publication 1), a molecular dynamics study was successfully carried out to explore the flap dynamics of the South African subtype C HIV-1 PR over 60 ns simulation period in explicit solvent. The MM-GBSA method showed a distinguishing factor in terms of binding affinities of stronger and weaker inhibitors by highlighting the largest van der Waals energy value observed for the ATV-enzyme complex as compared to that of the weaker value for the TPV-complex. The stronger affinity of ATV towards the PR was also displayed by the per-residue decomposition whereby the active site residues of the C-SA PR displayed fairly significant contributions to the overall binding of ATV. The per-residue decomposition analysis showed that the high contribution by the flap tip residues of the TPV-complex enables the flap tip region to be rigid and the complexed system to be reasonable stable, despite TPV lacking strong hydrogen bonding and key interactions at the active site region.

The RMSD analysis demonstrated well equilibrated systems whereby the inhibitor-complexes maintained better dynamic stability. The inhibitor-complexes attained less residual fluctuations compared to the unbound system. The PCA analysis demonstrated that there seems to be a

reduction in overall movement of the HIV-1 PR as a result of complexation of the inhibitor to the enzyme. Distance analysis highlighted the flap-tip distance attained by the ATV-complex as being very fixed compared to the complexed system of the weaker binder (TPV). In addition, the active site-flap tip distance analysis revealed an unsymmetrical nature of the inhibitor-protease complex as was displayed by the differing trend in distances obtained for chain A and B of the PR.

The angles measured between the flap tip residues demonstrate that the angle values were lowered when there a substrate or an inhibitor bound to the protease and this analysis also showed the flap tip residues being held at relative fixed positions. The bound complexes displayed less twisting compared to the apo system as highlighted by the dihedral angle analysis. Interestingly, the hydrogen bond analysis revealed several other factors causing the flap tip residues of the PR to be rigid more especially when complexed with a weaker TPV. This analysis also confirmed the stability of the TPV-complex as was displayed from the RMSF plot and distance analysis.

The flap dynamics study highlighted several factors as being vital in determining how inhibitors achieve binding to the PR active site. This study also provided deep insight into how the flap dynamics is affected due to binding of inhibitors differing in their binding affinities and how this can be used to the design of potential drugs with more potency. This study highlighted key factors associated with effective binding for protease inhibitors; their good binding into the active site and favourable interactions with the flap region.

In chapter three (publication 2), the stability of the dimeric form of HIV-1 C-SA PR and the impact of C- and N-terminal residue truncation on C-SA PR was monitored throughout 100 ns MD simulations. RMSD analysis depicted a less stable pattern for the truncated systems when compared to the native forms of all other native systems. The Rg analysis showed that truncation of C- and N-terminal residues led to less compact complex structures. The

MMGBSA method displayed a reduction in binding free energies for the truncated complexed systems. This is possibly due to the disruption of one of the most conserved interface regions (β-sheet interface) and this has been shown to have resulted in less binding affinity for the substrate and also the inhibitors.

The flap tip distance analysis showed that the inhibitor bound complexes compensate for the loss of dimerization at the β-sheet dimer interface by ensuring that the flap tips come closer. The hydrogen bond interaction that occurs at the salt bridge between ARG8 of chain A and ASP29 of chain B or ARG8 of chain B and ASP29 of chain A was also monitored. Truncation led to a collapse of this salt bridge as the truncated systems displayed less and fewer hydrogen bond interactions. The outcome of this study has shown that disruption of the β-sheet dimer interface enables the dimerization to collapse and reduces the binding affinity of inhibitors designed to bind onto the active site of C-S PR. It has also been shown that the active site residues displayed significant fluctuations which ultimately led to loss in binding affinity for active site-directed inhibitors

Recommendations

There are several lines of research arising from our study which we believe should be pursued. Firstly, the design of drugs which do not only have good binding into the active site of the HIVPR but also possess good interactions with other regions of the enzyme is a great area of interest for future research. In other words, intelligent design that optimize optimum interactions between the inhibitor and the enzyme.

From the results of chapter 3, a molecular modelling study is recommended that will explore a number of inhibitors that were experimentally shown to inhibit HIV PR dimerization. This will enable the development of a suitable computational model. It should be determined if these inhibitors lead to similar disruption of the β-sheet interface and subsequent interactions of the

rest of the PR complex. If that is the case, new inhibitors can be proposed and tested theoretically, based on the results obtained. Developing new potential inhibitors focussing on their interactions with the flap tips in addition to the active site region is of interest as we believe that when interactions of inhibitors with several regions of the protease are taken into considerations, potential drugs are can be developed with better potency. We are also interested in exploring the significance of key residues in the dimerization interface region and how mutation in this region can affect the conformational stability of the HIV-1 PR.

Appendix 1. Supplementary material for Chapter 2

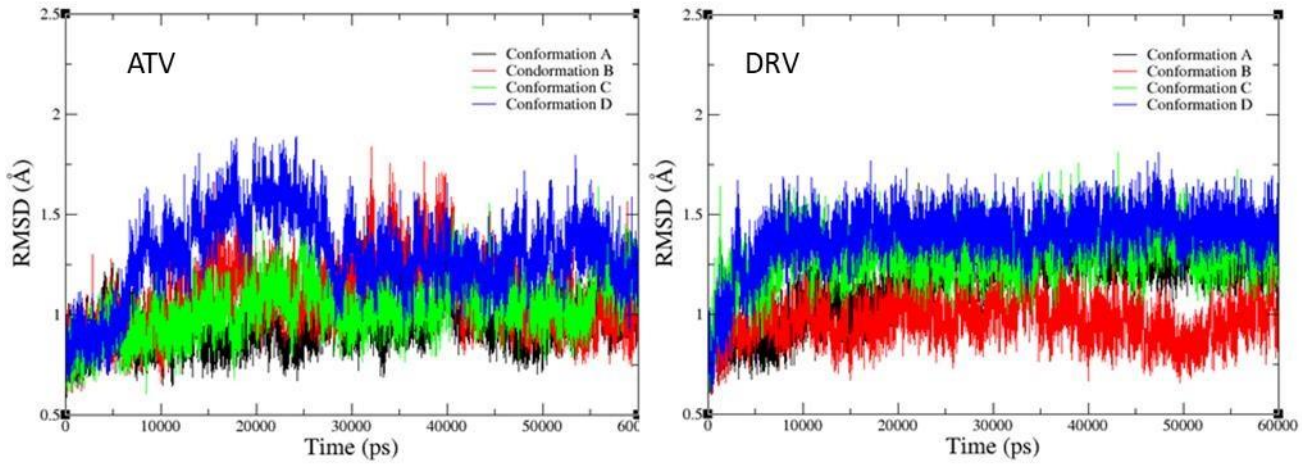


Figure S1. Comparison of the RMSD plots of triplicate MD simulation (conformation B, C and D) for (A) ATV and (B) DRV with RMSD of original MD simulation (conformation A).

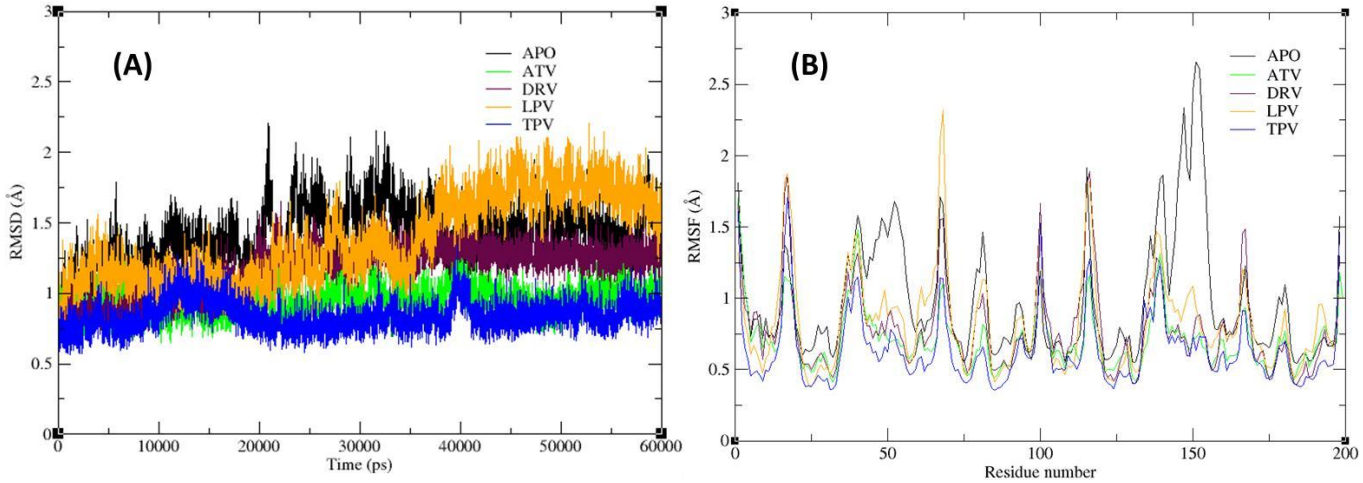


Figure S2. (A) Root Mean Square Deviation (RMSD) and (B) Root Mean Square Fluctuation for the apo enzyme and inhibitor-bound systems over the 60 ns trajectory.

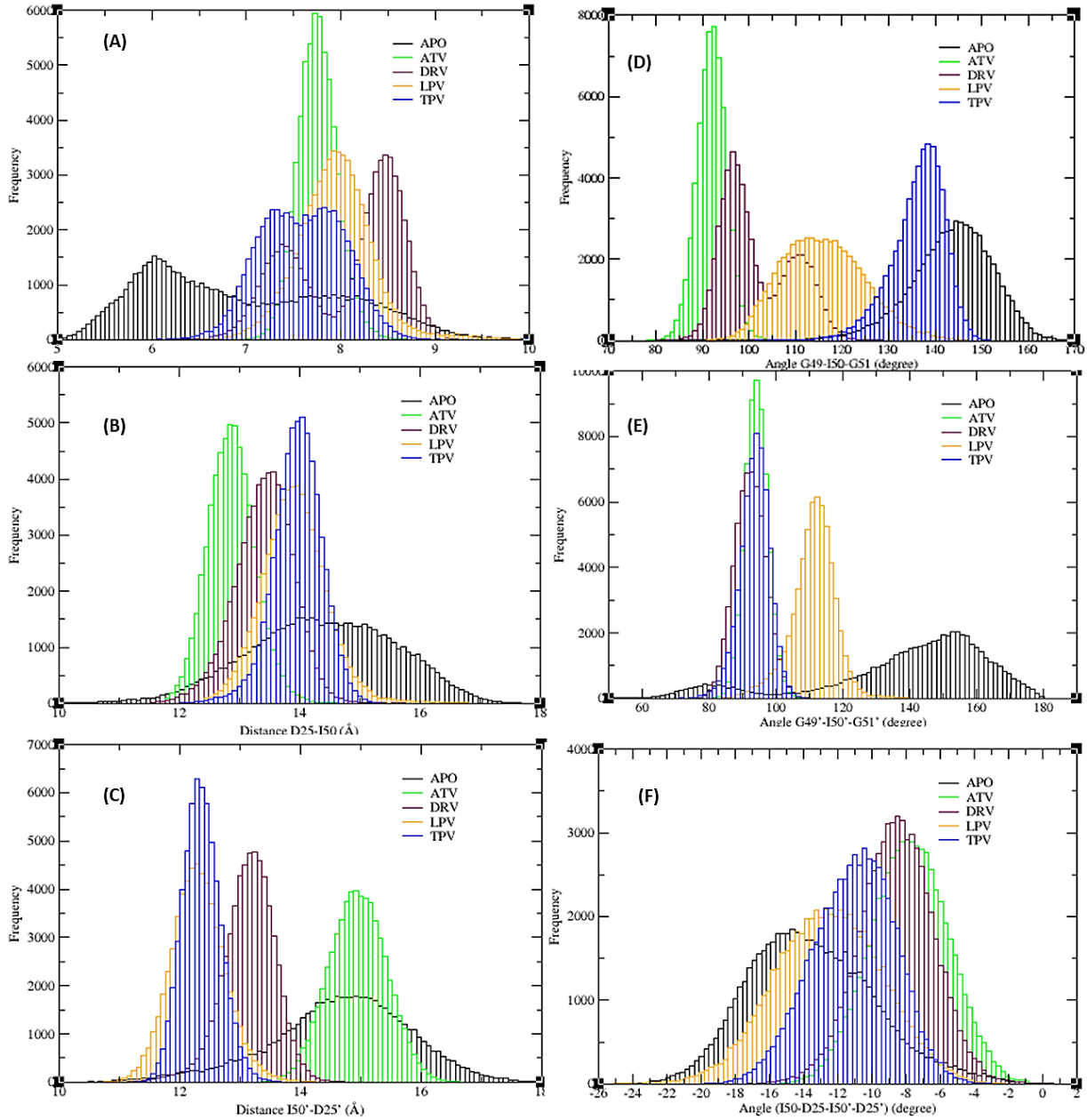


Figure S3. Distance and angle analyses for the elucidation of flap dynamics in C-SA PR for the 60 ns MD simulations; (A) distance between flap tips (Ile50 and Ile50'), (B) distance between active site residue and flap tip in chain A (Asp25-Ile50), (C) distance between active site residue and flap tip in chain B (Asp25'-Ile50'), (D) TriCα angle of chain A (Gly49-Ile50Gly51), (E) TriCα angle of chain B (Gly49'-Ile50'-Gly51'), (F) dihedral angles of chain A (Ile50-Asp25-Ile50'-Asp25').

Figure S4. Per-residue decomposition profiles of the natural substrate complex using the MM/GBSA approach.

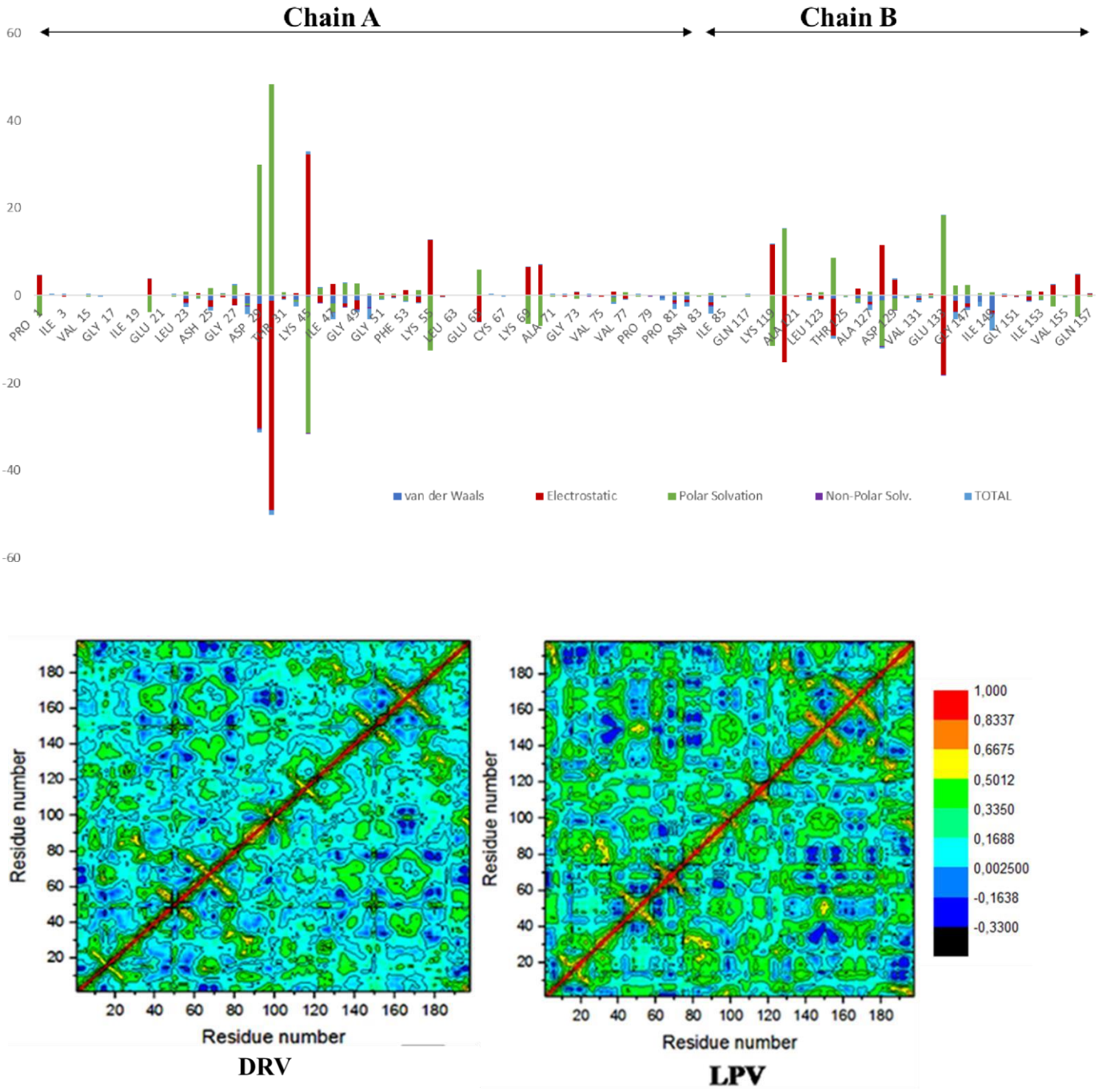


Figure S5. Cross-correlation matrices of the fluctuations of the coordinates for Cα atoms around their mean positions after the equilibrium of the 60 ns MD simulation.

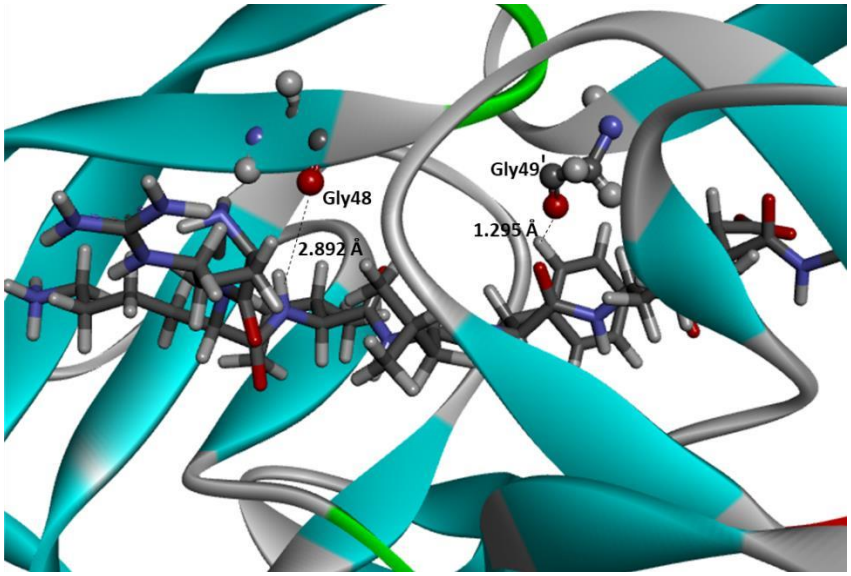


Figure S6: 3-D structure representing hydrogen-bond interaction of NS with certain residues of HIV-1 C-SA PR.

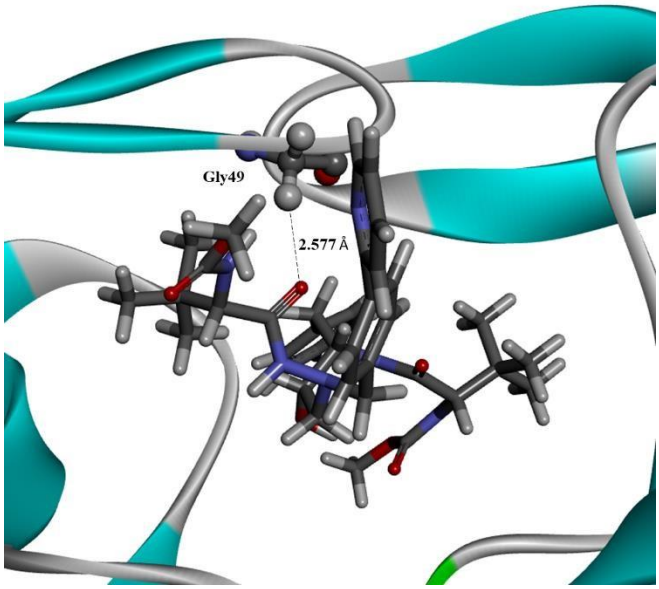


Figure S7: 3-D structure representing hydrogen-bond interaction of ATV with certain residues of HIV-1 C-SA PR.

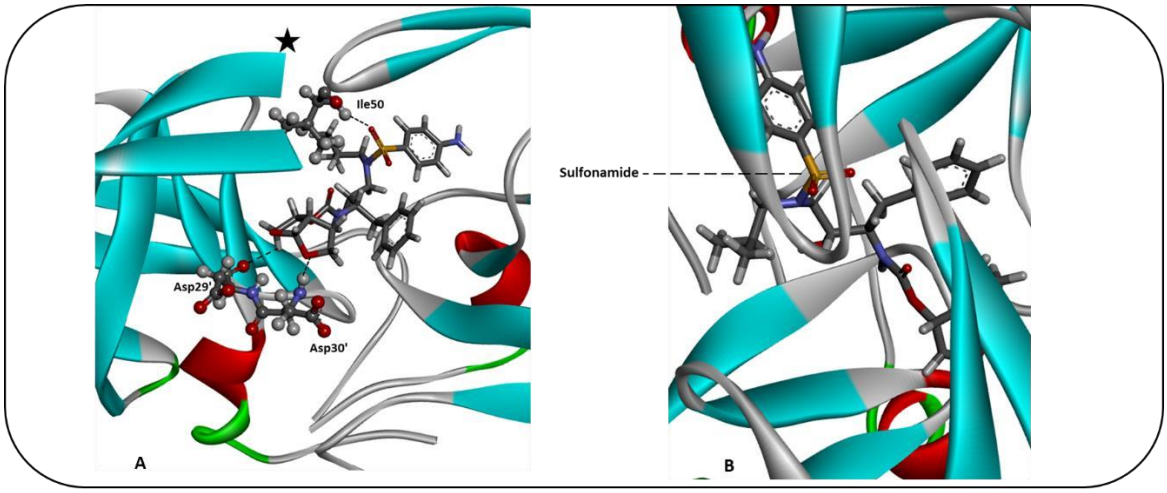


Figure S8: 3-D structures representing (A) hydrogen-bond interaction of DRV with certain residues of HIV-1 C-SA PR and (B) the location of the sulfonamide group inside the complex system. (**Note**: the black star in the flap tip position denotes the removal of flap tip residues for illustration purposes)

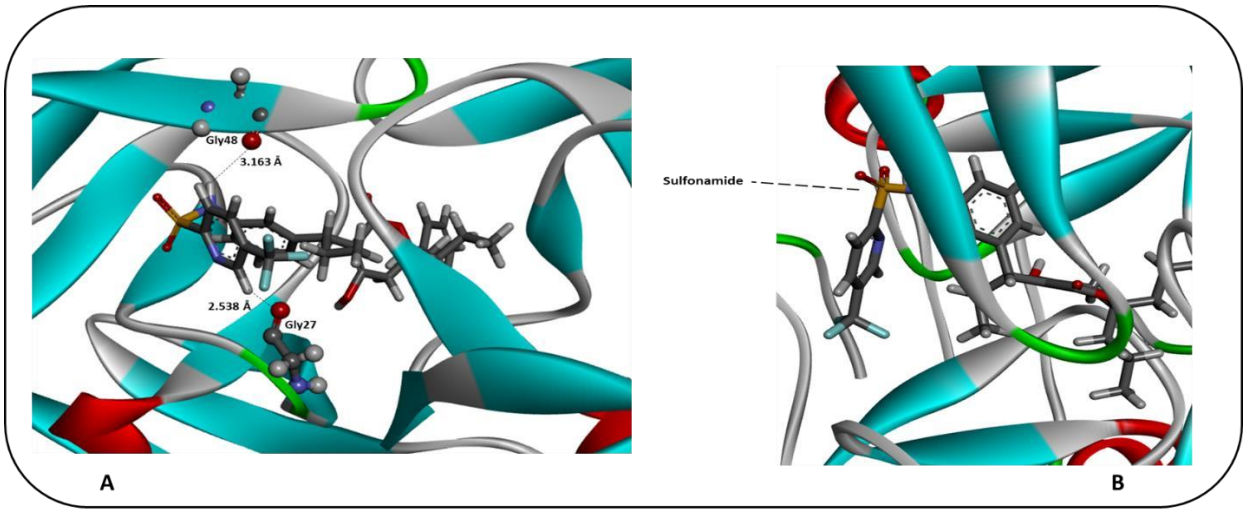


Figure S9: 3-D structures representing (A) hydrogen-bond interaction of TPV with certain residues of HIV-1 C-SA PR and (B) the location of the sulfonamide group inside the complex system.

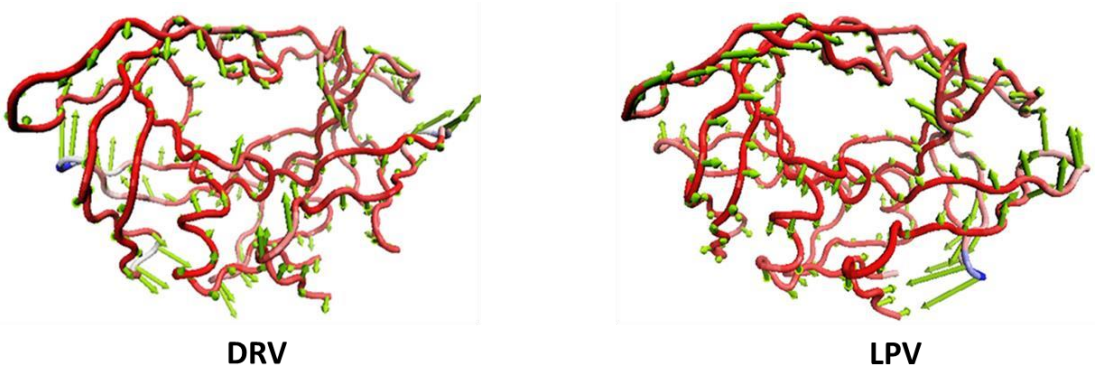


Figure S10: Collective motions corresponding to PC1 obtained by performing principal component analysis on MD simulation trajectory after the equilibrium for DRV and LPV (the protein part is presented in red ribbon, higher fluctuating parts are presented in grey and blue.

Appendix 2. Supplementary material for Chapter 3 Supplementary materials

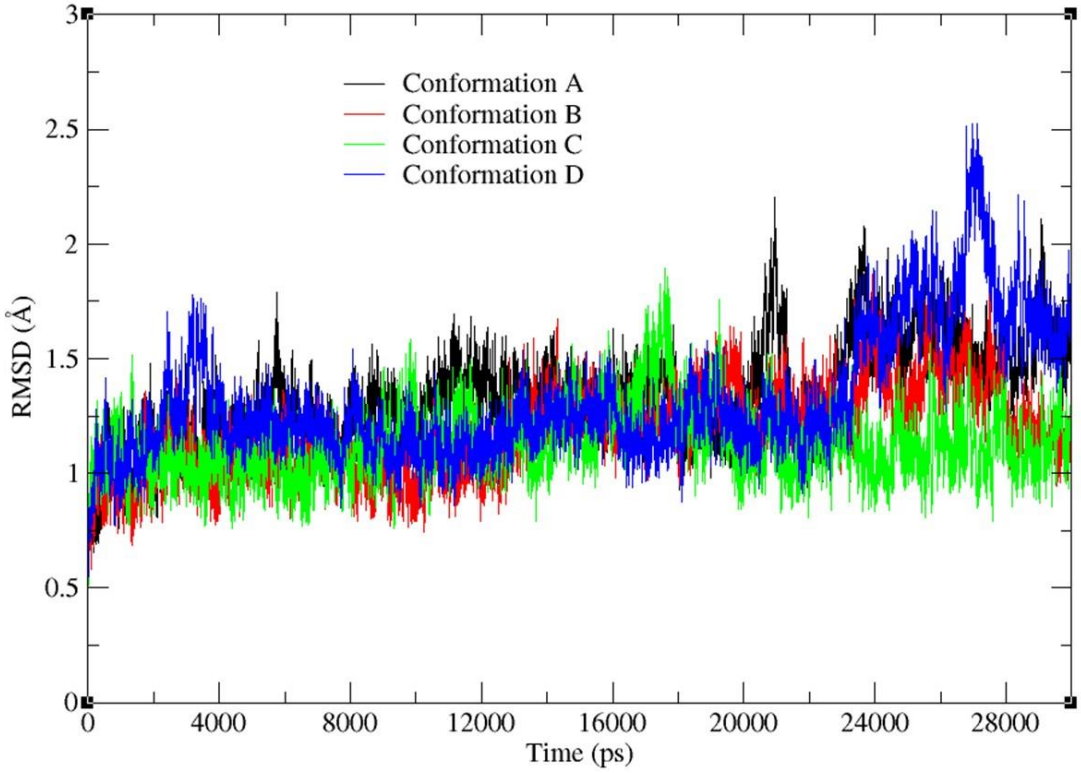


Figure S4: Comparison of the RMSD plots of triplicate MD simulation (conformation B, C and D) for Apo-native with RMSD of original MD simulation (conformation A).

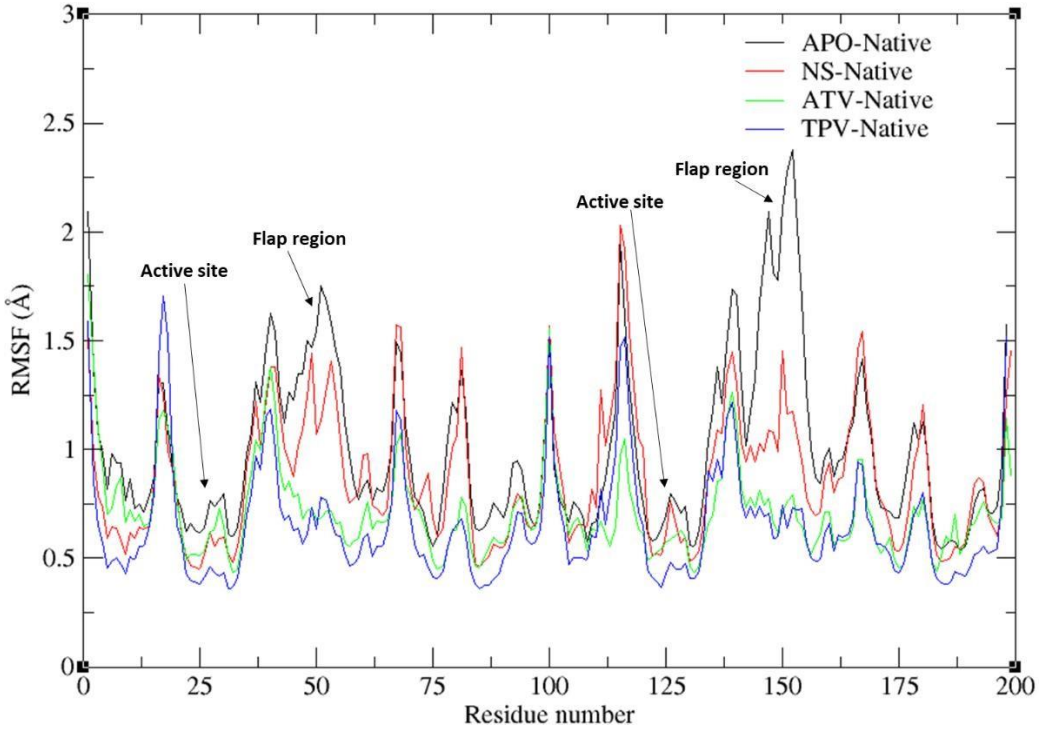


Figure S2: Root-Mean Square Fluctuation of backbone atoms against residue number for the APO, NS, ATV and TPV systems in Native forms.

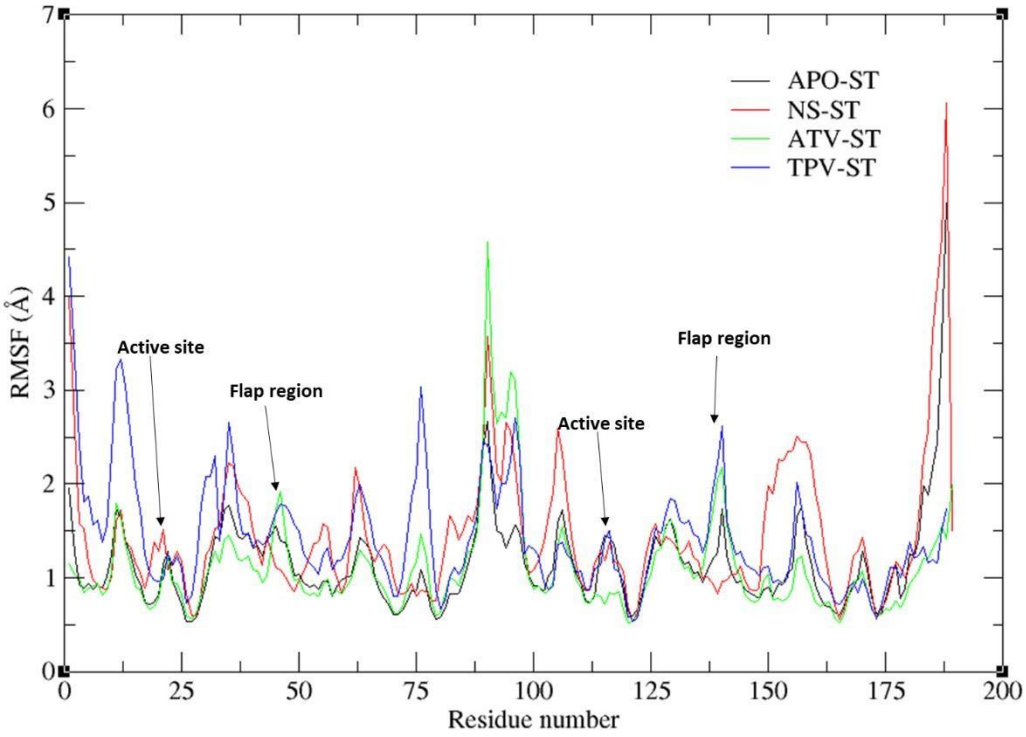


Figure S3: Root-Mean Square Fluctuation of backbone atoms against residue number for the APO, NS, ATV and TPV systems in Native forms.

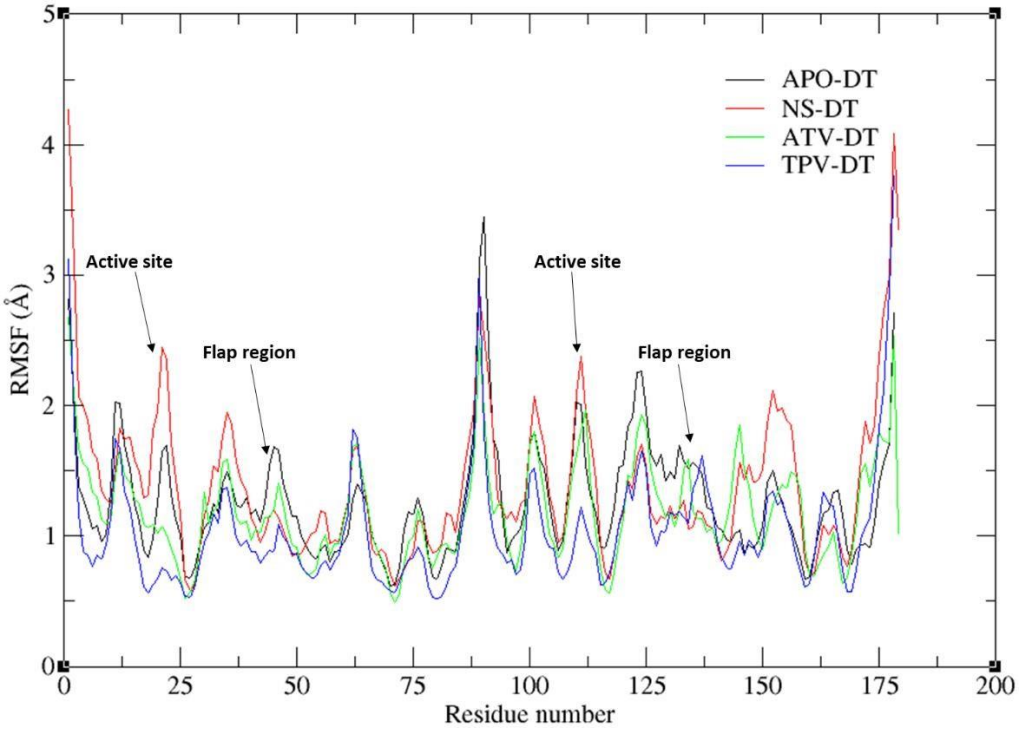


Figure S4: Root-Mean Square Fluctuation of backbone atoms against residue number for the APO, NS, ATV and TPV systems in Native forms.

Figure S5: The plot of radius of gyration for three different systems; native, ST and DT of (A) ATV and (B) TPV complexes.

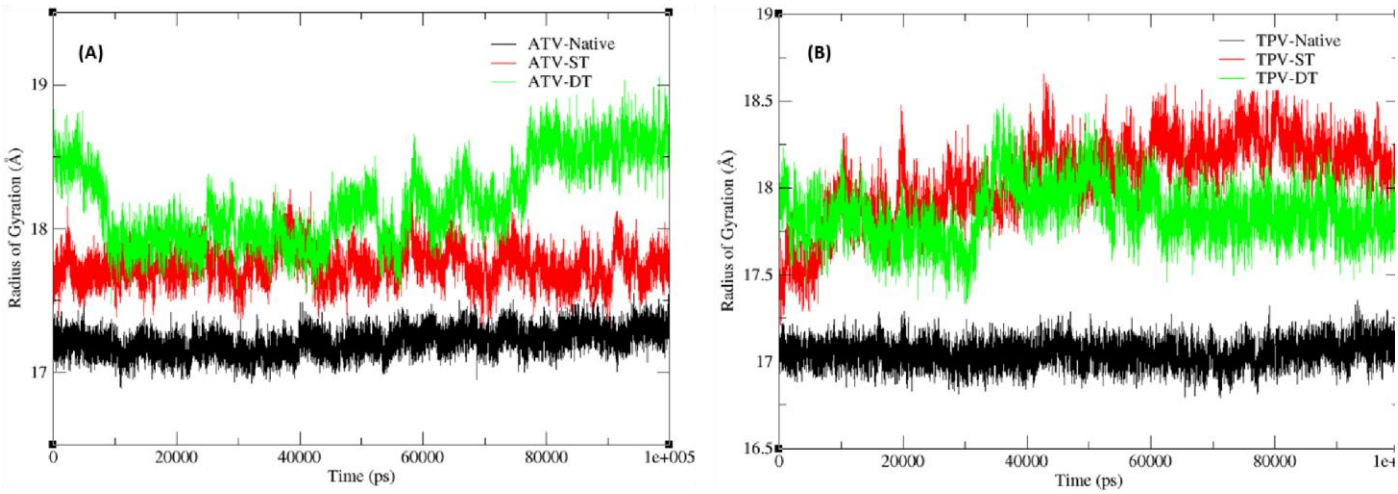


Figure S6: The distance (Å) between the flap tip residues (Ile50 and Ile50') for three different systems; native, ST and DT of the (A) ATV and (B) TPV complex obtained from a 100 MD simulation.

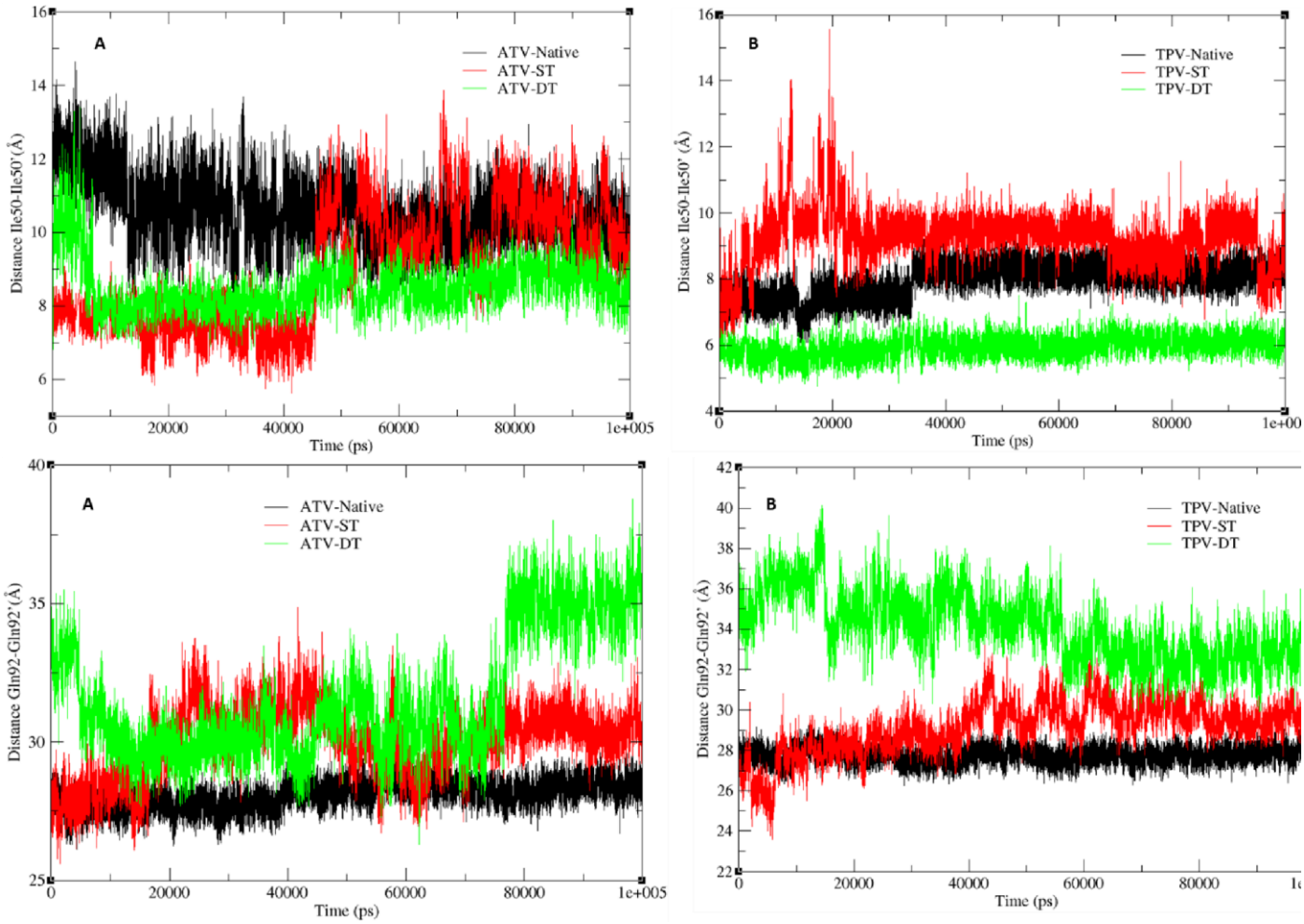


Figure S7: The distance (Å) between two Gln residues from chain A and B for the three systems; native, ST and DT for (A) ATV and (B) TPV during the 100 ns MD simulations.

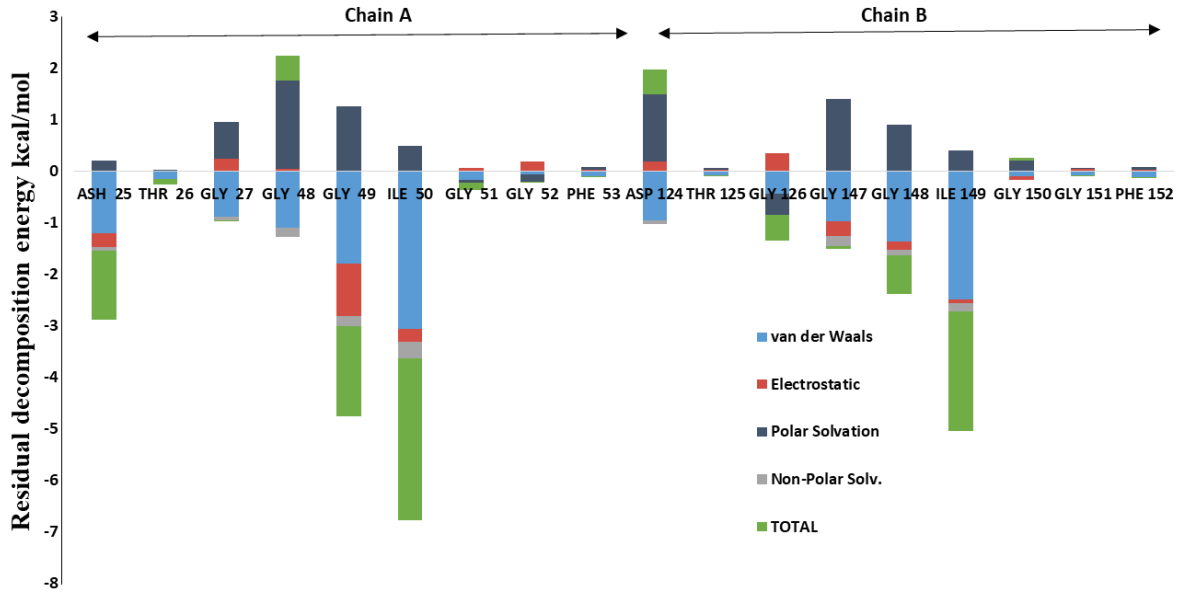


Figure S8: Plot of per residue decomposition energy in kcal/mol for the ATV-Native complex system.

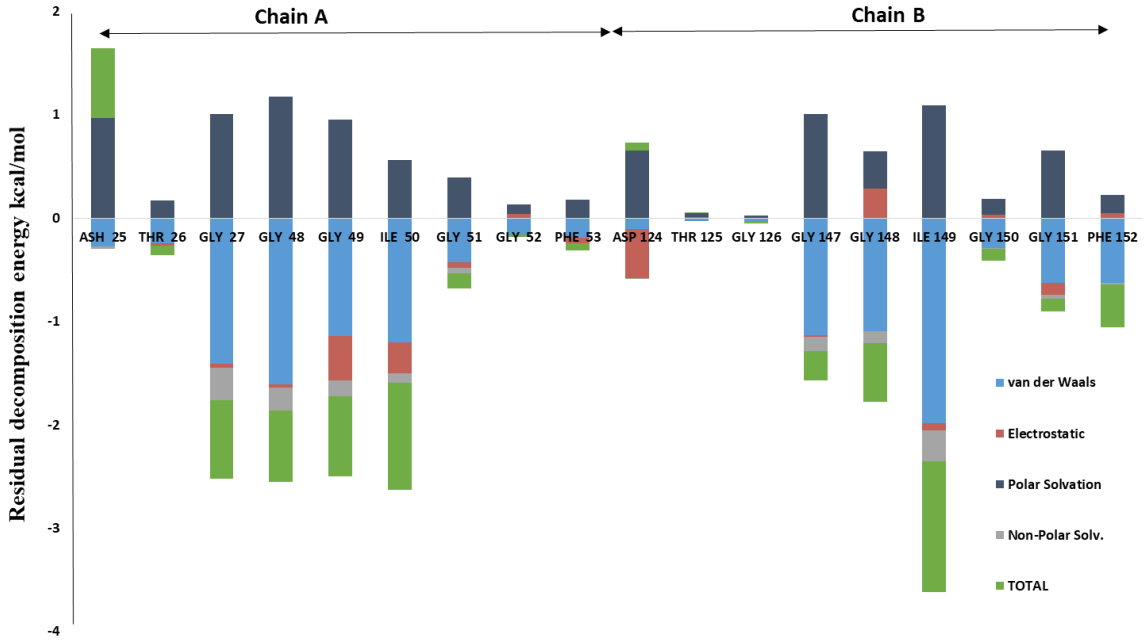


Figure S9: Plot of per residue decomposition energy in kcal/mol for the ATV-ST complex system.

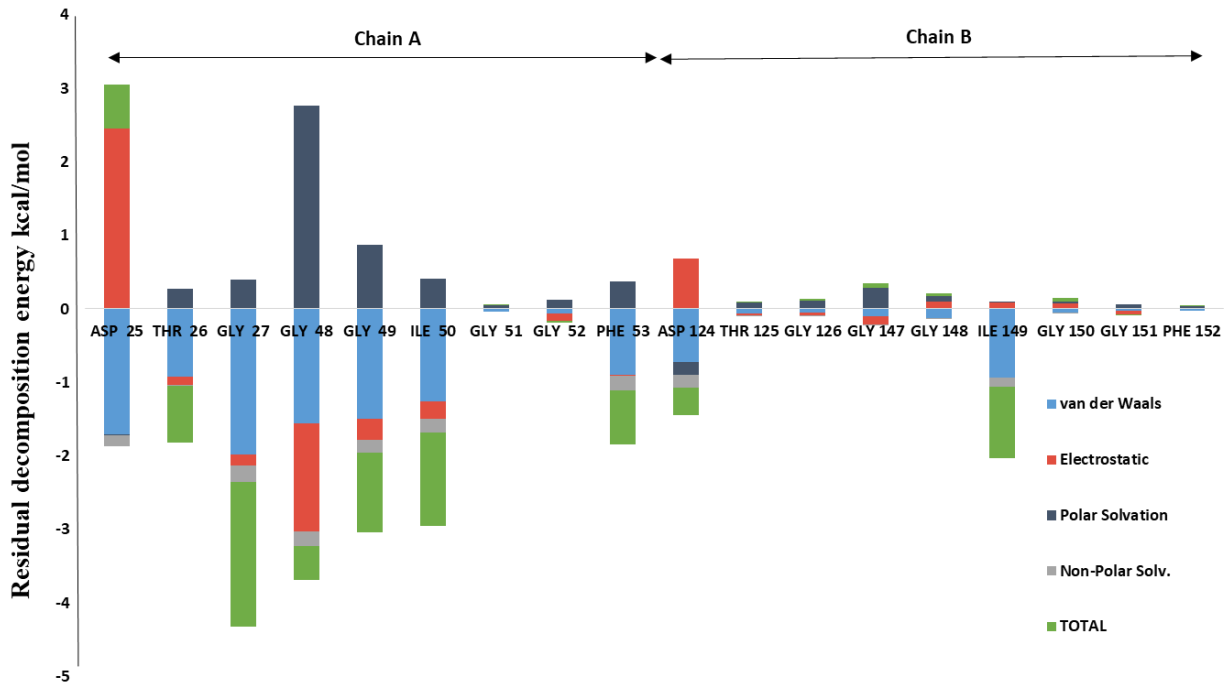


Figure S10: Plot of per residue decomposition energy in kcal/mol for the ATV-DT complex system.