https://dspace.org/

Structural and enzymatic studies of HIV-1 subtype C protease

Shabangu, Bonginkosi

2019

https://hdl.handle.net/10539/29631

Shabangu, Bonginkosi. (2019). Structural and enzymatic studies of HIV-1 subtype C protease. University of the Witwatersrand, https://hdl.handle.net/10539/29631

Downloaded from a DSpace repository. https://dspace.org/

Structural and enzymatic studies of HIV-1 subtype C protease

By

Bonginkosi Shabangu

(388330)

MSc dissertation

Submitted, in fulfilment of the requirements for the degree of Master of Science

in

Molecular and Cell Biology

to the Faculty of Science, University of the Witwatersrand, Johannesburg

Supervisor: Prof. Yasien Sayed

Co-supervisor: Dr Salerwe Mosebi

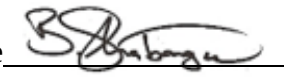
Julu 2019

DECLARATION

I, Bonginkosi Shabangu (388330), am a student registered for the degree of Master of Science (MSc) in the academic year 2019

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Dissertation has not been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature  15th day of July 2019

ABSTRACT

Human immunodeficiency virus (HIV) is a retrovirus that infects T-lymphocytes in the human immune system causing a condition known as Acquired Immunodeficiency Syndrome (AIDS). At the moment, there is an estimated 1.2 million people dying annually from diseases related to AIDS and ~ 36.7 million people are currently infected with HIV across the globe. There are 10 subtypes found within the major group of HIV-1 and subtype C overwhelmingly drives the South African epidemic and accounts for more than 50% of the world infections. The continuous infection and maturation of HIV-1 is yielded by the three viral enzymes; namely, reverse transcriptase, integrase and protease. HIV-1 relies on the catalytic efficacy of the protease enzyme for cleaving precursor polyproteins to yield structural and functional proteins. When HIV-1 protease is inactivated, either by an inhibitor or mutagenically, the virion will remain non-infectious. Thus, the HIV-1 protease is an opportune drug target. It is an important variant in the study of the pathogenesis, treatment and prevention of HIV-1 infections. The study intended to characterise and evaluate the catalytic activity of HIV-1 South African subtype C protease which exhibits high variability compared to subtype B protease. For characterisation properties, far-UV circular dichroism, intrinsic fluorescence spectroscopy and size exclusion high-performance liquid chromatography were used to evaluate secondary, tertiary and quaternary structure, respectively. Secondary structure results indicated a trough at 216 nm which means the protein is predominantly β-sheeted. Using intrinsic tryptophan as a probe, the tertiary structure of protease revealed that the local structural environment had not been perturbed and it was indicated by fluorescence emission intensity peak at 355 nm. The results of the size exclusion-high performance liquid chromatography study revealed that the dimeric molecular size of the wild-type protease was 22 kDa. The proteolytic efficiency of the subtype C wild-type protease was evaluated following the hydrolysis of a fluorogenic substrate resembling the CA/p2 cleavage site in the gag-pol polyprotein precursors. The K_M was determined as 42.07 μM, V_{max} was 0.047 μmol.min^{-1}, specific activity was 76.46 μmol/min/mg, k_{cat} was 24.02 s^{-1} and k_{cat}/K_M was 0.084 s^{-1}μM^{-1}. In the presence of darunavir, the K_i value for protease was determined as 3.4 nM, for saquinavir it was 6.2 nM and for ritonavir it was 9.4 nM. The protease inhibitor GRD 110D had a K_i of 27 nM, showing high susceptibility compared to the other three FDA-approved HIV protease inhibitors (PIs). A molecular docking study was carried out to analyse and compare the binding mode of the FDA-approved HIV PIs. The docking results indicated that the protease recognition site is hydrophobic due to the two-flexible glycine-dense β-sheets and that subtype C protease (PDB code: 3U71) polymorphisms

result in an altered flap-hinge region, thus making it less susceptible to inhibitors as compared to subtype B protease (PDB code: 2P3B). The results of the comparative binding mode analysis of all the FDA-approved drugs could be useful for the design of new potent inhibitors of HIV-1 protease.

ACKNOWLEDGEMENTS

I acknowledge my supervisor, Professor Y. Sayed for his amazing supervision, constant guidance throughout this project and patience throughout.

I also acknowledge my co-supervisor Doctor S. Mosebi, for his advice and always going the extra mile to assist.

To thank my family for their support and unwavering love.

Colleagues and friends at the Protein Structure-Function Research Unit.

I acknowledge the National Research Foundation for funding and Wits PMA for financial assistance.

TABLE OF CONTENTS

DECLARATION i
ABSTRACT ii
ACKNOWLEDGEMENTS iv
TABLE OF CONTENTS v
LIST OF FIGURES viii
LIST OF TABLES x
LIST OF ABBREVIATIONS xi
GMO Clearance Certificate xiii
PSFRU: Wits IBC Clearance Certificate xiv

CHAPTER 1 1
1. INTRODUCTION 1
1.1 Human Immunodeficiency Virus 1
1.1.1 HIV/AIDS background 1
1.1.2 Epidemiology of HIV/AIDS 1
1.1.3 Genetic diversity 4
1.1.4 HIV-1 structure and genome 7
1.1.5 HIV-1 replication cycle 9
1.1.6 Drug therapy 10
1.2 HIV-1 protease 12
1.2.1 The active site 13
1.2.2 The flap region 16
1.2.3 The dimer interface 16
1.2.4 The catalytic mechanism of HIV-1 subtype C protease 16
1.2.5 HIV-1 protease mutations 18
1.3 HIV-1 subtype C protease inhibitors 19
1.3.1 First-generation protease inhibitors 22
1.3.2 Second-generation protease inhibitors 23
1.3.3 Third-generation protease inhibitors 24
1.4 HIV-1 protease inhibitors test compounds 24
1.5 Aim and objectives 29
1.5.1 Aim 29
1.5.2 Objectives 29

CHAPTER 2 30
2. METHODS AND PROCEDURES 30
2.1 Transformation of *E. coli* BL21 (DE3) pLysS with pET-11b 30
2.2 Plasmid DNA purification and verification 31
2.3 Cell growth of *E. coli* BL21 (DE3) pLysS with pET-11b 32
2.4 Expression trials of wild-type HIV-1 C-SA protease 32
2.5 Overexpression and purification of wild-type HIV-1 C-SA protease 33
2.6 Tricine – Sodium dodecyl sulfate-polyacrylamide gel electrophoresis 34
2.7 Determination of protein concentration 35
2.8 Protease active site concentration using an isothermal titration calorimeter 37
2.9 Structural characterisation of HIV-1 wild-type C-SA protease 39
2.9.1 Secondary structural characterisation using far-UV circular dichroism spectroscopy 39
2.9.2 Tertiary structural characterisation using intrinsic fluorescence spectroscopy 41
2.9.3 Quaternary structural characterisation using size exclusion – high-performance liquid chromatography 42
2.10 Protease kinetics 43
2.10.1 Determination of K_M and V_{max} 44
2.10.2 Catalytic turnover number (k_{cat}) 45
2.10.3 Specific activity 46
2.10.4 Catalytic efficiency (k_{cat} / K_M) 46
2.11 Solubility studies for novel inhibitors 47
2.12 Inhibition studies – IC_{50} and K_i values 47
2.13 Molecular docking studies of HIV protease 48
2.13.1 *In silico* studies of HIV-1 protease 49
CHAPTER 3 52
3. RESULTS AND DISCUSSION 52
3.1 Protease cDNA insert sequence verification 52
3.2 Cell growth of *E. coli* BL21 (DE3) pLysS with pET-11b 52
3.3 Expression trials of HIV-1 wild-type C-SA protease 55
3.4 Overexpression and purification of HIV-1 wild-type C-SA protease 55
3.5 Protease concentration determination 58
3.6 Protease active site titration calorimetry 58
3.7 Structural characterisation of HIV-1 wild-type C-SA protease 63

3.7.1 Secondary structural characterisation using far-UV circular dichroism spectroscopy 63
3.7.2 Tertiary structural characterisation using intrinsic fluorescence spectroscopy 63
3.7.3 Quaternary structural characterisation using size exclusion – high - performance liquid chromatography 67
3.8 Protease enzyme kinetics 67
3.8.1 Determination of K_M and V_{max} 69
3.8.2 Determination of specific activity and catalytic turnover number (k_{cat}) 69
3.8.3 Catalytic efficiency (k_{cat} / K_M) 72
3.9 Inhibition kinetics studies using GRD 110D, saquinavir, ritonavir and darunavir against wild-type HIV-1 C-SA protease 76
3.10 Induced-fit docking studies 80

CHAPTER 4 87
4. CONCLUSION 87

CHAPTER 5 89
5. REFERENCES 89
APPENDIX 104
Appendix A. Supplementary material for section 3.3 104
Appendix B. Supplementary material for section 3.10 106

LIST OF FIGURES

Figure 1: The coverage of the number of people receiving antiretroviral therapy out of a total number of people living with HIV, 2016 3

Figure 2: Geographic distribution of HIV-1 subtypes 6

Figure 3: A. Mature HIV-1 virion structure with some of the regulatory proteins. B. HIV-1 schematic representation of gag, gag-pol and nef proteins processing sites showing the 12 individual protease cleavage sites. C. The sequences flanking the HIV-1 protease cleavage sites in the gag and gag-pol 8

Figure 4: HIV-1 replication 11

Figure 5: Cartoon representation of the crystal structure of the homodimer HIV-1 wild-type South African subtype C (C-SA) PR and multiple sequence alignment of different subtypes 14

Figure 6: The arrangement of hydrogen bonds between HIV-1 protease and a modelled substrate 15

Figure 7: The catalytic mechanism for aspartic proteases 18

Figure 8: HIV protease structure depicting mutations associated with resistance to FDA-approved protease inhibitors 20

Figure 9: The chemical structures of FDA-approved HIV protease inhibitors 21

Figure 10: General structures of protease inhibitor compounds 25

Figure 11: Schematic representation of a typical micro-calorimeter with the variables determined directly from isothermal titration calorimetry 38

Figure 12: Protein multiple sequence alignment 53

Figure 13: Cell growth of *E. coli* BL21 (DE3) pLysS with pET-11b 54

Figure 14: Expression trials of HIV-1 wild-type C-SA protease in *E. coli* BL21 (DE3) pLysS at 20 ^{o}C and 37 ^{o}C using varying IPTG concentrations and induction periods. 56

Figure 15: CM-Sepharose cation exchange elution profile of HIV-1 wild-type C-SA protease and SDS-PAGE 57

Figure 16: HIV-1 C-SA protease purity and molecular weight determination 59

Figure 17: HIV-1 wild-type C-SA protease absorption spectrum 60

Figure 18: Active site titration of the protease with acetyl-pepstatin displayed in the ITC thermogram at 20 ºC 61

Figure 19: Far-UV CD spectra of HIV-1 wild-type C-SA protease 64

Figure 20: Three-dimensional structure of C-SA PR (PDB I.D. 3U71) showing the location of tyrosine and tryptophan residues 65

Figure 21: Intrinsic fluorescence emission spectra of tryptophan and tyrosine excited at 280 nm and tryptophan excitation at 295 nm 66

Figure 22: SE-HPLC elution profile of the HIV-1 wild-type C-SA protease 68

Figure 23: Michaelis-Menten plot for the determination of K_M and V_{max} 71

Figure 24: Determination of specific activity for the HIV-1 wild-type C-SA protease 73
Figure 25: Determination of catalytic turnover (k_{cat}) for the HIV-1 wild-type C-SA protease 74
Figure 26: Determination of catalytic efficiency (k_{cat}/K_M) for the HIV-1 wild-type C-SA protease 75
Figure 27: IC_{50} determination of (A) Saquinavir, (B) Ritonavir and (C) Darunavir for the wild-type C-SA protease 77
Figure 28: IC_{50} determination of GRD 110D for the wild-type C-SA protease. 78
Figure 29: Graphical representation of the IC_{50} and K_i values of GRD 110D, saquinavir, ritonavir and darunavir for C-SA PR 79
Figure 30: Bar graph representing the mean – E-model energy (kcal/mol) docking of 3U71 and 2P3B bound to FDA-approved protease inhibitors 81
Figure 31: Induced-fit docking 2D and 3D schematic representation of darunavir bound to HIV-1 protease done using Schrödinger software 83
Figure 32: Structural alignment of HIV-1 protease subtypes bound to darunavir 86

LIST OF TABLES

Table 1. Novel protease inhibitor compounds 21

Table 2. Preparation of 20% tricine-SDS-PAGE..... 36

Table 3. Comparison of the kinetic parameters of C-SA PR obtained experimentally in this study with the literature kinetic parameters of C-SA PR..... 70

LIST OF ABBREVIATIONS

Å	Ångström
AIDS	Acquired immunodeficiency syndrome
ARV	Antiretroviral
°C	degrees Celsius
CA	Capsid protein
CD4	Cluster of differentiation 4
CFRs	Circulating recombinant forms
dH_2O	distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ε	molar extinction coefficient
EDTA	Ethylenediaminetetraacetic acid
Far-UV CD	Far-ultraviolet circular dichroism
FDA	Food and Drug Administration
ΔG	Change in Gibbs free energy
gag	Group-specific antigen
ΔH	Change in enthalpy
HAART	Highly active antiretroviral treatment
HIV-1/2	Human immunodeficiency virus – type 1/2
HIV-1 C-SA PR	HIV-1 South African subtype C protease
IC_{50}	Inhibitor concentration reducing enzyme activity by 50%
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
k_{cat}	Catalytic constant
K_d	Dissociation constant
kDa	kilodalton
K_i	Inhibition constant
K_M	Michaelis constant
k_{cat}/K_M	Catalytic efficiency, specificity constant
ℓ	Litre

LB	Lysogeny Broth
mAu	milli absorbance units
mdeg	milli degrees
MRE	Mean residue ellipticity
n	Stoichiometry
NOPs	Naturally occurring polymorphisms
OD_{600}	Optical density at 600 nm
PDB	Protein Data Bank
PI	Protease inhibitor
PMSF	Phenylmethylsulfonyl fluoride
PR	Protease
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RT	Reverse transcriptase
ΔS	Change in entropy
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SE-HPLC	Size exclusion-high performance liquid chromatography
WHO	World Health Organisation
UNAIDS	United Nations AIDS Organisation
URFs	Unique recombinant forms
V_{max}	Maximum velocity

agriculture,
forestry & fisheries

Department:
Agriculture, Forestry and Fisheries
REPUBLIC OF SOUTH AFRICA

Genetic Resources, Department of Agriculture, Forestry and Fisheries
Private Bag X973, Pretoria 0001

Enquiries: Bathobile Mahlangu • **Tel**: 012 319 6165 • **Fax:** 012 319 6298 • **E-mail**: BathobileM@daff.gov.za •**Ref**: 39.2/University of Witwatersrand -18/001

Prof H. W Dirr
University of Witwatersrand
School of Molecular and Cell Biology
East Campus
Gate House Room 431
Johannesburg
2050

+27 (0) 11 717 6352 **(Tel)**
+27 (0) 86 553 5708 **(Fax)**
Heinrich.dirr@wits.ac.za **(email)**

Dear Prof Dirr

RE: REGISTRATION OF FACILITY

With reference to the application to register a facility, submitted in terms of the Genetically Modified Organisms Act, 1997 (Act No. 15 of 1997). Registration number **39.2/University of Witwatersrand -18/103.**

The facility is hereby registered; please find attached your certificate which serves as proof of registration. Please familiarize yourself with the Standard Operating Procedure approved for Regulation 2(2) to determine whether your current activities or any future activities would require an additional contained use permit or not.

Please consult the website of the Department at www.daff.gov.za (Branches, Agricultural Production, Health & Food Safety/ Genetic Resources/ Biosafety) for the latest application forms and the SOP document referred to in the above paragraph.

If any of the provisions of the Genetically Modified Organisms Act, 1997 (Act No. 15 of 1997), including any condition of any permit issued in terms of the GMO Act, is not complied with at all times, you will be subject to prosecution in terms of Section 21 of the GMO Act, 1997.

Yours sincerely

2018 -01- 31

Ms N L Mkhonza
Registrar: Genetically Modified Organisms Act, 1997 (Act No. 15 of 1997)

UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Research Office

INSTITUTIONAL BIOSAFETY COMMITTEE
(R 14/16)

CLEARANCE CERTIFICATE **PROTOCOL NUMBER: 20180105Lab**

BRIEF DESCRIPTION OF APPLICATION:

Laboratory situated in room 411-434 URC Proten Structure-Fucntion Research Unit Gate House 4th Floor

APPLICANT: Prof Heini Dirr et al

SCHOOL/DEPARTMENT : Molecular and Cell Biology

DATE CONSIDERED: 29 January 2018

DECISION OF COMMITTEE: Approved unconditionally

These laboratories are considered to meet the standards of the Institutional Biosafety Committee (IBC), for BSL1 approval, as specified in the IBC's published Standard Operating procedures.
Any change to the type of biohazardous agents being handled should be reported to the IBC
DAFF- Reg no 39.2/University of the Witwatersrand-18/103 expiry date 31 January 2018 - 31 January 2021

1. This clearance certificate expires on 12 March 2023 and may be renewed on application.

DATE OF APPROVAL: 12 March 2018 **CHAIRPERSON:** _____

Professor Tanki...

Radiation and Health Physics Unit
University of Witwatersrand
Johannesburg, Republic of South Africa
Tel: +27 (0) 11 717 6837

DECLARATION OF APPLICANT:

To be completed in duplicate and **one copy** returned to the University of the Witwatersrand, Faculty of Health Sciences, Research Office, Office 301, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown,2193.

1. I have read, understood and accepted the approval conditions above
2. I note that the University Safety Officer, or his/her representative, may at any reasonable time inspect my laboratory or trial site to ensure compliance with current Health and Safety legislation. I undertake to offer my full co-operation in any such inspection.
3. I have read, understood and will comply with the *recommended standard operating procedures for the handling of biohazardous materials* posted at http://intranet.wits.ac.za/academic/uro/Pages/Institutional-Biosafety-Committee-(IBC).aspx

Signed: _____ **Date:** _____

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

CHAPTER 1

1. INTRODUCTION

1.1 Human Immunodeficiency Virus

1.1.1 HIV/AIDS background

Human Immunodeficiency Virus (HIV) is a member of the genus *Lentivirus* (a subgroup of retrovirus) and belongs to the *Retroviridae* family (Barré-Sinoussi *et al*., 1983; Coffin *et al*., 1986). HIV infects the T-lymphocytes in the human immune system and causes a condition known as Acquired Immunodeficiency Syndrome (AIDS). AIDS is a disease in which there is a severe loss of the body cellular immunity, greatly lowering the resistance to infection and malignancy (Weiss *et al*., 1985; Fauci, 2003). The link between HIV and AIDS was discovered in 1984; however, on the 27th of July 1982, the public health officials began to use the term AIDS to describe the occurrences of infections such as *Pneumocystis jirovecii*, and Kaposi's sarcoma in previously healthy people (Marx, 1982; Safai *et al*, 1984). It was also confirmed that the viral infection does not discriminate based on gender, age, sexual orientation and race (Marx, 1982). In 1983 the first strain of HIV was isolated by Luc Montagnier, Francoise Barré-Sinoussi, and Robert Gallo at the Pasteur Institute in France and they discovered the virus that causes AIDS, which was initially named HTLV-III/LAV (human T-cell lymphotropic virus-type III/lymphadenopathy-associated virus) by an international scientific committee, before being changed to HIV (Barré-Sinoussi *et al*., 1983; Clavel *et al*., 1986; Coffin *et al*., 1986). The earliest known case of HIV infection within humans was detected in 1959 from a blood sample of a man in Kinshasa, the Democratic Republic of the Congo. How the man become infected by the virus is not fully known but many researchers believe that HIV was introduced into the human population when hunters were exposed to infected blood of chimpanzees carrying the Simian Immunodeficiency Virus (SIV) (Sharp *et al*., 2001; Keele *et al*., 2006; Sharp and Hahn, 2011).

1.1.2 Epidemiology of HIV/AIDS

Ever since the discovery of HIV/AIDS in the early 1980s, HIV infection has been reported in every country of the world. It has caused major socioeconomic challenges and global health problems over the past three decades. HIV is spread through certain body fluids from a person who is infected. These fluids are blood, pre-seminal fluid, vaginal fluids, semen, breast milk,

and rectal fluids. Without treatment, the virus damages the immune system and advances to three stages of HIV infection. The first stage is acute HIV infection which generally develops within two to four weeks after infection and during this time, most people have flu-like symptoms. The second stage is known as a chronic (asymptomatic or clinical latency) HIV infection, where HIV multiplies at very low levels as compared to the acute stage. The final stage is AIDS and this is where the body can no longer fight off opportunistic infections, and at this stage, people are diagnosed with AIDS when their CD4 count is less than 200 cells/mm^3 or if they have certain opportunistic infections (Fauci and Lane, 2005). The average survival rate for an infected person with HIV without treatment is ~ 9 – 11 years and with treatment, they can enjoy a near-normal lifespan. However, the drugs must be taken lifelong with long-term side effects (Maartens *et al*., 2014; Poorolajal *et al*., 2016). Despite major advances in HIV/AIDS therapies, there are significant drawbacks to the current treatments. Some of the drawbacks are the drug toxicity, drug-drug interactions, and evolution of different patterns of systematic complications involving bone, kidney, heart and associated neurocognitive disorder are increasing (Fauci and Lane, 2005; Poorolajal *et al*., 2016). One of the greatest challenges that the World Health Organization (WHO) faces today is that there are about 22 million people living with HIV who do not have access to antiretroviral therapy. Many people delay testing for HIV because they fear discrimination and, as a result, they have a late diagnosis of HIV. Late diagnosis remains the substantial barrier to scaling up HIV treatment and thus contributing to HIV transmission (WHO, 2016).

Over the years, a detailed understanding of the HIV epidemic has emerged through data collection, analysis and dissemination. The Joint United Nations Programme on HIV/AIDS (UNAIDS) leads the world's most extensive data collection on HIV epidemiology, programme coverage, financed and published the most authoritative and up-to-date information on the HIV epidemic. UNAIDS latest statistics on the worldwide HIV infection in Figure 1 shows that HIV/AIDS continues to be a global pandemic with 36.7 million [30.8 - 42.9 million] people living with HIV (UNAIDS, 2017). UNAIDS further reported that people who are newly infected with HIV totalled to 1.8 million [1.6 – 2.1 million] in the year 2016. Since the start of the epidemic until 2016, about 76.1 million [65.2 – 88 million] people have become infected and 35 million [28.9 – 41.5 million] people have died from AIDS-related illnesses; although, there has been a 48% decline in the number of deaths from AIDS-related causes due to the HIV treatment. The decline is from a peak of 1.9 million [1.7 - 2.2 million] in 2005 to 1 million [830 000 - 1.2 million] in 2016, and new HIV infections (all ages) have declined by 16% globally

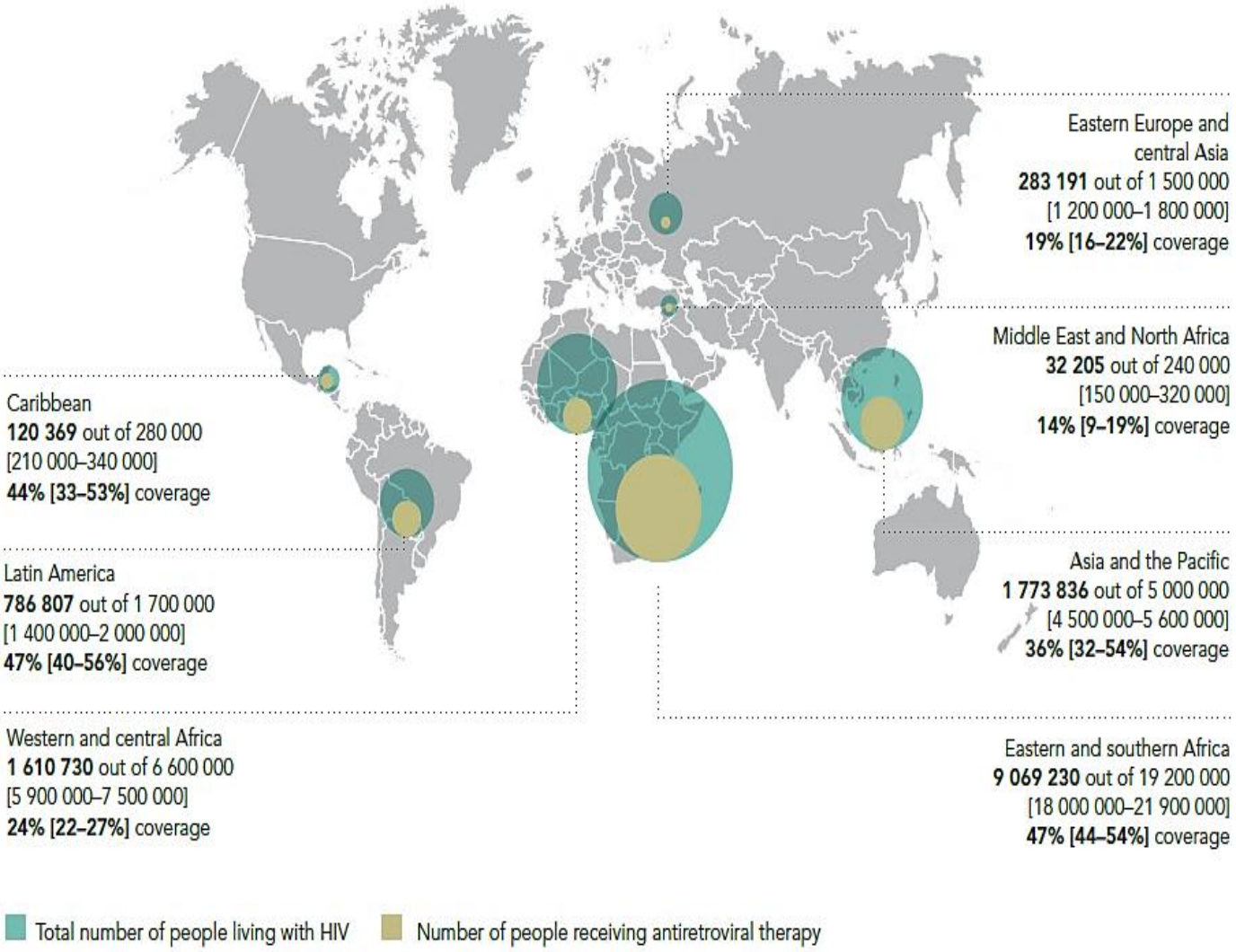


Figure 1: The coverage of the number of people receiving antiretroviral therapy out of a total number of people living with HIV, 2016. North America and Western Europe are omitted due to the lack of data. The image was taken from a UNAIDS report for 2017.

between 2010 and 2016. New HIV infections declined from 500 000 [470 000-530 000] in 2005 to 270 000 [240 000-290 000] in 2016 and AIDS-related deaths declined from 270 000 [240 00-310 000] in 2005 to 110 000 [88 000-140 000] in 2016. In the eastern and southern African regions, the decline was 62% from ~1.1 million in 2004 to ~420 000 in 2016, a trend that is reflected by the rapid scale-up of antiretroviral therapy in the regions. In June 2017, 20.9 million people living with HIV were receiving antiretroviral therapy globally. The decline is seen across the globe, where western and central Africa (30% reduction), western and central Europe and North America (45% reduction), Latin America (16% reduction), Asian and the Pacific (39% reduction), and the Caribbean (52% reduction).

However, the pace of decline for new HIV infections is far too slow to reach the Fast-Track Target agreed upon by the United Nations General Assembly in 2016. The Fast-Track strategy states that new infections per year should be fewer than 500 000. There should be fewer than 500 000 people dying from AIDS and HIV-related discrimination if the disease is to be eliminated by the year 2020. Nonetheless, AIDS-related illnesses remain the leading cause of death, especially among women of reproductive age (15-49) globally, and the second leading cause of death for young women age (15-24) in Africa (Global AIDS monitoring 2017). While new cases have been reported in all regions of the world, the sub-Saharan African region is the most affected, accounting for 43% of new cases in eastern and southern Africa. The UNAIDS global report for 2016 showed that 70% of the sub-Saharan African population is infected with HIV and South Africa is one of the African countries most affected by the epidemic, with an estimated 7.1 million [6.4 - 7.8 million] of its population living with HIV/AIDS. HIV/AIDS still remains problematic in the African continent as there is still a lack of access to care, prevention and treatment, and there are no vaccines to eradicate the virus from an infected patient.

1.1.3 Genetic diversity

HIV has a high genetic diversity; thus, it is classified according to types, groups, subtypes and recombinant forms, based on phylogenetic and genetic distance analysis. HIV is characterised into two distinct groups, HIV-1 and HIV-2. HIV-1 is the more virulent, mutates rapidly, producing variable strains of the virus within a single host cell and is mainly responsible for the worldwide epidemic. HIV-2 is mainly found in Western Africa, and it is less pathogenic than HIV-1 because it does not spread easily, and infected individuals acquire immunodeficiency slower than those infected with HIV-1. HIV-2 has eight known groups (A to H) and of these groups, only A and B are disease causing. It is closely related to simian

immunodeficiency virus endemic in sooty mangabeys (*Cercocebus atys atys*) (SIVsmm) (Chen *et al*., 1997; Sharp *et al*., 2001). HIV-2 has not been explicitly discussed in the study as the research focuses on HIV-1. However, it still causes AIDS and has a similar response to antiretroviral therapy as to HIV-1. HIV-1, which was discovered in 1983 by Francoise Barré-Sinoussi, Luc Montagnier and Robert Gallo, is divided into four groups known as: M (major), O (outlier), N (non-M, non-O), and the newly discovered group, P (pending identification). Group M (comprising 95% of the complete genome sequences of HIV-1) represents the predominant group and is divided into nine subtypes: A (with sub-subtypes; A1, A2), B, C, D, F (with sub-subtypes; F1, F2), G, H, J and K, and a growing number of circulating recombinant forms (CRFs), second generation recombinant (SGRs) and unique recombinant forms (URFs) (Robertson *et al*., 2000; Gonzales *et al*., 2001). The different subtypes have a genetic variation of 25-35% in their nucleotide level, they form a phylogenetic cluster with amino acid variations of 25-40%, and the differences within subtypes usually range from 15-20% (Taylor *et al*., 2008). All HIV subtypes feature naturally occurring genetic variations and these variant forms, differ in genotype and phenotype and they are referred to as naturally occurring polymorphisms (NOPs) (Coman *et al*., 2008).

Different subtypes are geographically distributed across the globe as seen in Figure 2. The classification of subtypes enabled researchers to focus on a specific genetic variation per region. In Australia, North America and Western Europe they focus mainly on the subtype B and in sub-Saharan Africa, China, India and Brazil they research subtype A and C which account for the majority of HIV-1 infection (Figure 2) (Walker *et al*., 2005; Buonaguro *et al*., 2007). Of all the subtypes, subtype B is the most studied subtype because of its occurrence in developed countries, and it accounts for approximately 12% of the global HIV pandemic. As the HIV pandemic continues, more studies are now focusing on subtype C since it accounts for more than 56% of all new infections globally (Esparza and Bhamarapravati, 2000). HIV-1 subtype C is the most prevalent subtype ever since its discovery in 1983, and it accounts for about 95% of HIV-1 infections in Southern Africa (Walker *et al*., 2005). The disproportionate increase in subtype C compared to other HIV-1 subtypes may suggest that subtype C is more easily transmitted, or it has a higher level of "fitness" at the population level (Gordon *et al*., 2003; de Oliveira *et al*., 2003).

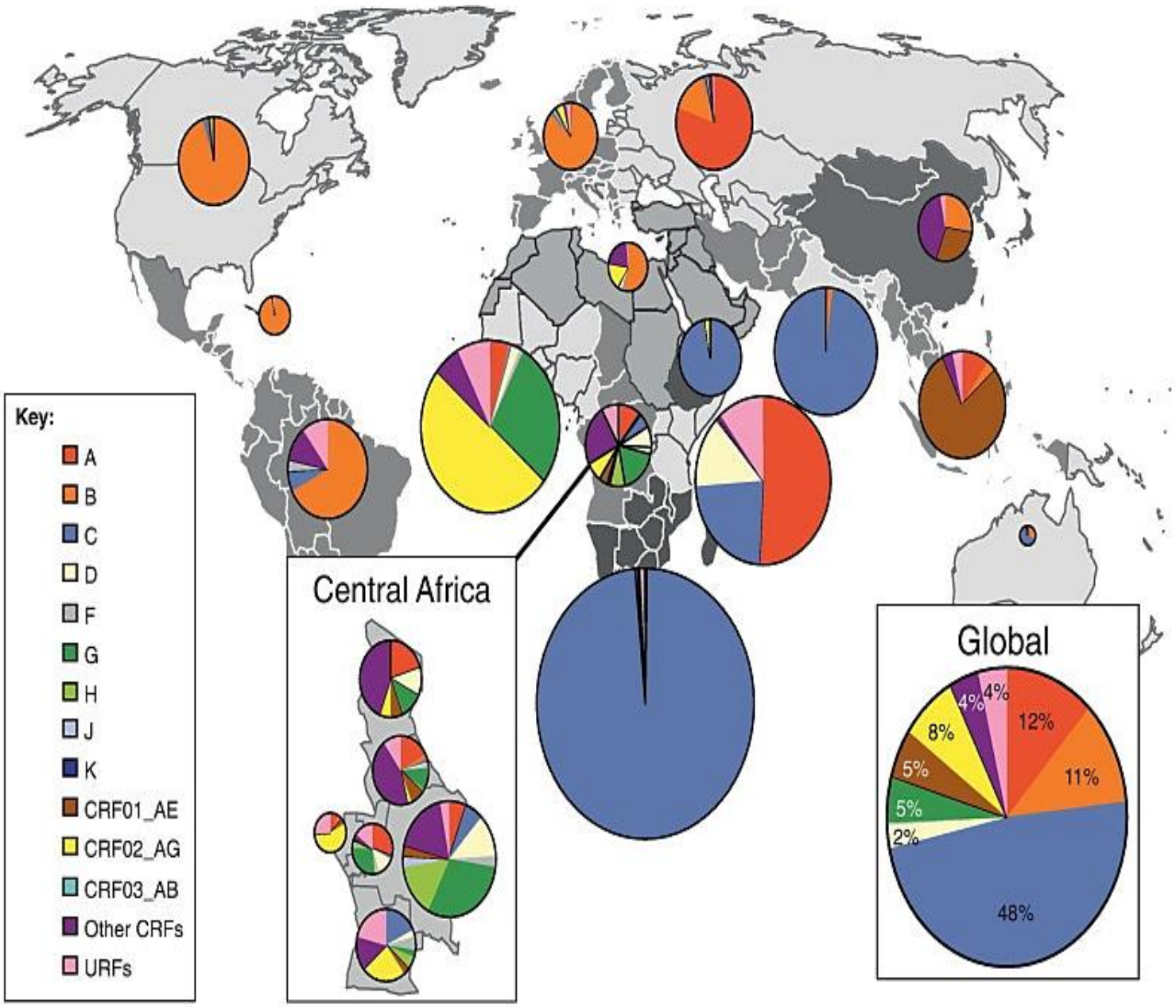


Figure 2: Geographic distribution of HIV-1 subtypes. Nine subtypes of group M are shown with CFRs and URFs. The HIV-1 subtypes vary in their prevalence across the globe and a major variation is seen within the African continent. Subtype C is more prevalent in sub-Saharan Africa and accounts for 48% of all infection of HIV. The picture was taken from Hemelaar *et al*., (2011).

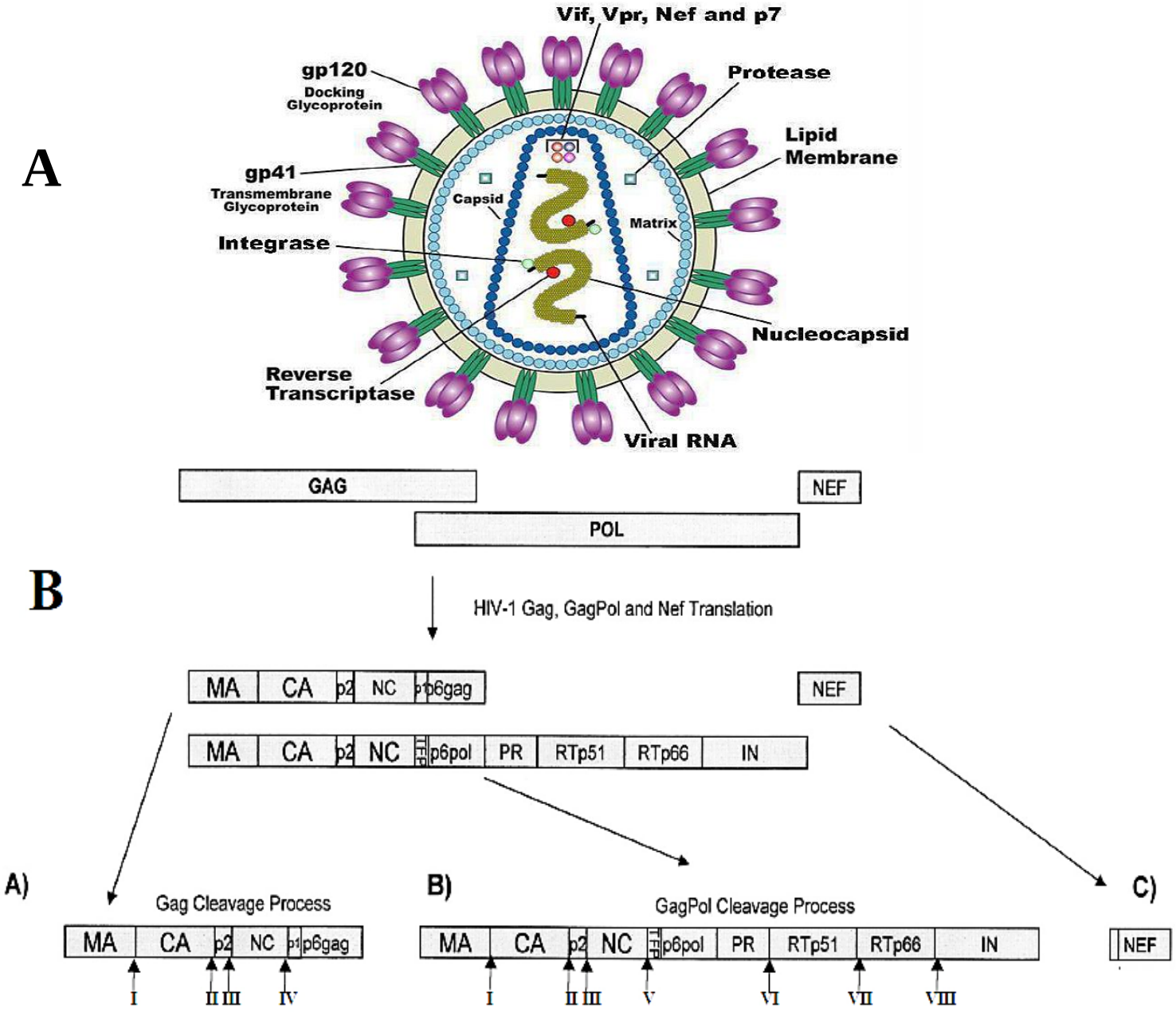
Viral studies indicate that subtype C is differentiated from other HIV-1 subtypes by its distinct phenotypic and genetic properties. Various studies have postulated that an extra NF-$_K$B binding site in the long terminal repeat, a 5-amino-acid-insertion in Vpu or a prematurely truncated Rev protein may influence the viral gene expression, altering the pathogenesis and transmissibility of the subtype C viruses (de Oliveira *et al*., 2003). The increased spread of subtype C may also have been contributed by the factors related to viral entry (such as the CCR5 and non-syncytium-inducing properties of subtype C isolates) and pathogenesis. Another area of research that is receiving consideration is the possibility that subtype C may have a more active, catalytically efficient protease compared to other HIV-1 subtypes (Esparza and Bhamarapravati, 2000; Buonaguro *et al*., 2007; Hemelaar *et al*., 2011).

1.1.4 HIV-1 structure and genome

HIV-1 is roughly spherical in structure, consisting of a cylindrically shaped core capsid (p24 structural protein) and is approximately 100 to 120 nm in diameter (Greene, 1993; Coffin, 1995; Tavassoli, 2011). The HIV-1 envelope is a lipid bilayer obtained from the host cell plasma membrane during the virus budding process. HIV-1 is composed of two single-stranded positive-sense RNA molecules that make up its genetic material. The RNA molecules are bound to p7 proteins forming a nucleocapsid (Engelman and Cherepanov, 2012). The HIV-1 genome together with the three enzymes (reverse transcriptase, integrase and protease) and the six accessory proteins are enveloped by capsid proteins p24 to form a conical capsid (Velazquez-Campoy *et al*., 2003). The mature HIV-1 virion structure is shown in Figure 3 A.

HIV-1, as other retroviruses, synthesises its proteins in the form of polyproteins. The HIV-1 genome is made up of two identical 9.2 kb positive-sense, single-stranded RNA molecules. The RNA molecules are composed of nine genes classified into three functional groups: first is the classification of structural genes which are *gag*, *env* and *pol* genes, second is the classification of regulatory genes which are *tat* and *rev* genes, and third is the classification of accessory genes which are *vpu*, *vpr*, *vif* and *nef* genes. The first three genes encode the structural proteins and the remaining six genes encode regulatory proteins to promote viral transcription, mature virion release and/or viral pathogenesis (de Oliveira *et al*., 2003).

The group-specific antigen (gag), envelope glycoprotein (env) and polymerase (pol) are synthesised as a polyprotein and expressed as zymogens (Figure 3 B) (Coffin, 1984; Ashorn *et*



C

Cleavage site		Sequence							
		P1	P2	P3	P4		P1'	P2'	P3' P4'
I	(MA/CA)	-- Ser-	Gln-	Asn-	Tyr	----	Pro-	Ile-	Val-Gln --
II	(CA/p2)	-- Ala-	Arg-	Val-	Leu	----	Ala-	Glu-	Ala-Met --
III	(p2/NC)	-- Ala-	Thr-	Ile-	Met	----	Met-	Gln-	Arg-Gly --
IV	(NC/p1)	-- Pro-	Gly-	Asn-	Phe	----	Leu-	Gln-	Ser-Arg --
V	(TF/PR)	-- Ser-	Phe-	Asn-	Phe	----	Pro-	Gln-	Ile-Thr --
VI	(PR/RT)	-- Thr-	Leu-	Asn-	Phe	----	Pro-	Ile-	Ser-Pro --
VII	(RT (Int.))	-- Ala-	Glu-	Thr-	Phe	----	Tyr-	Val-	Asp-Gly --
VIII	(RT/IN)	-- Arg-	Lys-	Ile-	Leu	----	Phe-	Leu-	Asp-Gly --

Figure 3: A. Mature HIV-1 virion structure with some of the regulatory proteins. The picture was taken from Sierra *et al*., (2005). B. HIV-1 schematic representation of gag, gag-pol and nef proteins processing sites showing the 12 individual protease cleavage sites. The symbols used are MA; matrix protein, CA; capsid protein, NC; nucleocapsid, TF; transmembrane protein, PR; protease, RT; reverse transcriptase, RN; ribonuclease and IN; integrase (de Oliveira *et al*., 2003). C. The sequences flanking the HIV-1 protease cleavage sites in the *gag* and *gag-pol* (de Oliveira *et al*., 2003).

al., 1990; Luban *et al*., 1993). The zymogens are later hydrolysed at specific sites to produce mature proteins crucial for viral assembly and maturation for the continuation of infection of other host cells (Vogt, 1996; Louis *et al*., 1999; Turner and Summers, 1999). HIV-1 protease mediates the cleavage of these polyproteins to their component proteins, but only after this enzyme has excised itself from *gag-pol*. HIV-1 encodes two polyproteins, *gag* (55 kDa), and *gag-pol* (160 kDa), which are anchored to the plasma membrane via N-terminal myristoylation. There are 12 individual protease cleavage sites in the *gag* and *gag-pol* processing site: five cleavage sites in *gag* (p17/p24, p24/p2, p2/NC, p7/p1 and p1/p6^{gag}), six cleavage sites in *gag-pol* (NC/TFP, TFP/p6^{pol}, p6^{pol}/PR, PR/RT, RT/p66 and p66/IN) and a single site in the *nef* (Figure 3 C) (Debouck *et al*., 1987; de Oliveira *et al*., 2003). The majority of these cleavage sites are significantly more diverse in subtype C viruses when compared with subtype B viruses (de Oliveira *et al*., 2003). These 12 proteolytic reactions are required to generate a mature infectious HIV-1 virion and each reaction occurs at a unique cleavage site that differs in amino acid composition. As shown in Figure 3 B, the main structural proteins are formed by the cleavage of the Pr55^{gag} polyprotein into the matrix (MA; p17), capsid (CA; p24), nucleocapsid (NC; p7), p6^{gag}, and two spacer peptides, p2 and p1. The major phosphoprotein of HIV-1, p6^{gag}, plays an essential role in releasing the virus from the membrane of infected cells (de Oliveira *et al*., 2003).

1.1.5 HIV-1 replication cycle

HIV-1 replication is initiated by the virus approaching a target cell (T-lymphocyte). The complex protein (gp120) binds to CD4^{+} receptors antigen of the host T-helper cells (Freed, 2001; Plantier *et al*., 2009). The process is called 'cell attachment' and promotes further binding to the co-receptors (such as CCR5 or CXCR4, depending on the host cell type). High-affinity binding of the trimeric envelope protein results in a conformational change; thus, allowing the viral transmembrane protein to fuse with the cell membrane. The fusion process leads to the subsequent entry of the viral nucleocapsid into the host cell. The viral genomic RNA, integrase, reverse transcriptase and protease are subsequently released into the infected cell (Kramer *et al*., 1986; Mugnaini *et al*., 2005).

Reverse transcriptase (RT) forms a heterodimer (comprising two subunits of 66 kDa (p66) and 51 kDa (p51)) with RT/RNase H and have a high mutation frequency. RT converts the viral genomic RNA into double-stranded DNA in the cytosol once viral entry into the cell has been accomplished. The viral DNA is fully synthesised within 6 hours after the viral entry. RT has

two types of DNA polymerase activity, RNA-dependent DNA and DNA-dependant DNA polymerase activity. In addition to this, RT also possesses ribonuclease H (RNaseH) activity, and these activities are required for the proper and efficient functioning of the RT in reverse transcription, thus, providing a unique target for the antiviral intervention (Kramer *et al*., 1986; Mugnaini *et al*., 2005).

Integrase is responsible for the insertion of the viral DNA into the host cell genome (Brown *et al*., 1987; Burkrinsky *et al.,* 1992). It has exonuclease activity – removes two nucleotides from each 3' ends of linear DNA and endonuclease activity – cleaves host dsDNA at the integration site. The ligase activity of this enzyme is responsible for the formation of a single covalent linkage at each end of the viral DNA. HIV mRNA is translated to HIV polyproteins and the protease (PR) cleaves the synthesised immature viral polypeptide precursors (gag, gag-pol and nef) into individually mature, structural proteins and the process is known as viral maturation (Kramer *et al*., 1986; Mugnaini *et al*., 2005). Unlike most types of retroviruses, HIV-1 eventually kills the cells producing it. The knowledge about HIV as the causative agent and molecular events critical to HIV life-cycle initially identified a number of important biochemical targets for antiviral therapy development. The assembly, budding and maturation of HIV-1 is depicted in Figure 4.

1.1.6 Drug therapy

The identification of HIV-1 and accumulated knowledge about the role of the life cycle led to the development of antiretroviral drugs (ARVs) to transform HIV-1 infection from an inevitably fatal disease into a manageable chronic ailment. To date, there are Food and Drug Administration (FDA) approved drugs available for the treatment of HIV-1 infection. The first drug to be approved by the FDA (in 1987) to fight AIDS was 3'-azido-3'-deoxythymidine (AZT; zidovudine), which was synthesised in 1964 as a possible anti-cancer agent; however, it was ineffective (Vega *et al*., 2004; Olomola *et al*., 2013). AZT is a nucleoside analogue that, on enzymatic conversion to its triphosphate in the cells, inhibits the HIV-1 reverse transcriptase, as do the several other drugs that the FDA had approved prior to 1996 to treat AIDS (Olomola *et al*., 2013). Unfortunately, these agents only slow the progression of an HIV infection but do not stop it. This is partly due to the fact that they are toxic, mainly to the bone marrow cells and hence they cannot be taken in large doses. Unlike other DNA polymerases, the reverse transcriptase cannot correct its mistakes and hence frequently generates mutations (about one per 10^4 bp) (Ohtaka *et al*., 2002). Due to mutations introduced into the viral genome,

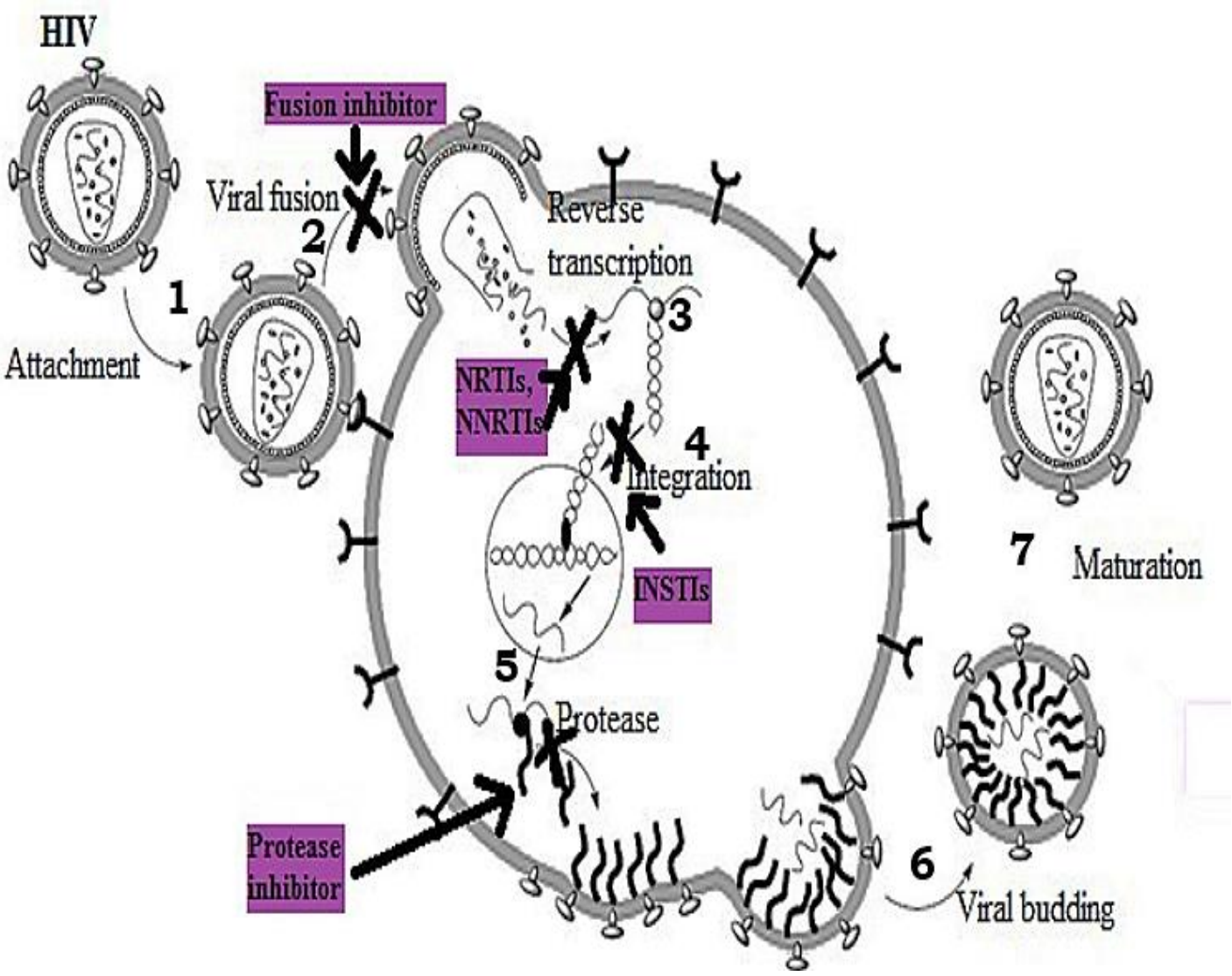


Figure 4: HIV-1 replication. (1) HIV-1 attaches to the CD4+ receptor and (2) viral fusion takes place within the cell membrane of the virus and the host cell. (3) Reverse transcription takes place in the cytosol of the cell via the RT enzyme where it converts the retroviral genomic RNA into dsDNA. (4) Integrase is responsible for the insertion of viral DNA into the host genome. (5) mRNA is translated and PR cleaves the polyprotein into immature viral proteins. (6) Viral budding and (7) maturation via PR activity and the virus is ready to infect other cells. Each step in the replication cycle is a potential target for antiviral intervention (purple boxes). Adapted from Alterman *et al*., (2001).

it provides a challenge for the production of effective antiviral compounds. The occurrence of mutations is advantageous for the virus because they confer drug resistance by lowering the binding affinity to the inhibitors and the virus continues replicating even with treatment. Mutations are characterised by the replacement of amino acids, their deletion, or new insertions into the viral genome or enzymes.

HIV-1 inhibitors are developed to target specific steps in the HIV life cycle, and these drugs are distributed into six distinctive classes based on molecular mechanism and resistance profile: fusion inhibitors, integration inhibitors, nucleoside RT inhibitors (NRTIs) and non-nucleoside RT inhibitors (NNRTIs) that target HIV-1 replication. However, these inhibitors are only partially effective. To treat HIV/AIDS, the inhibitors are used concurrently, and this is known as highly active antiretroviral therapy (HAART) where a combination of ARVs is used for the treatment of HIV-1 infection. HAART began by approval of several HIV-1 protease inhibitors drugs in the mid-1990s and their combination with reverse transcriptase inhibitors. It became evident that the combination of chemotherapy was significantly more effective than dosing the drugs sequentially (Broder, 2010). The inhibitors of HIV-1 protease are among the most effective antiretroviral drugs because the protease is essential for the maturation of immature HIV particles. If HIV-1 protease is inactivated, either by chemical inhibition or genetic mutation, the virus loses its function, thus stopping viral maturation and the spread of HIV. All the protease inhibitors (PIs) in clinical use have been made to target HIV-1 subtype B PR. However, since HIV-1 subtype C overwhelmingly drives the HIV epidemic in sub-Saharan Africa, China, India and Brazil. A better understanding of the structural and enzymatic consequences of HIV-1 protease drug-induced and its natural polymorphisms is required if more effective drugs are to be designed.

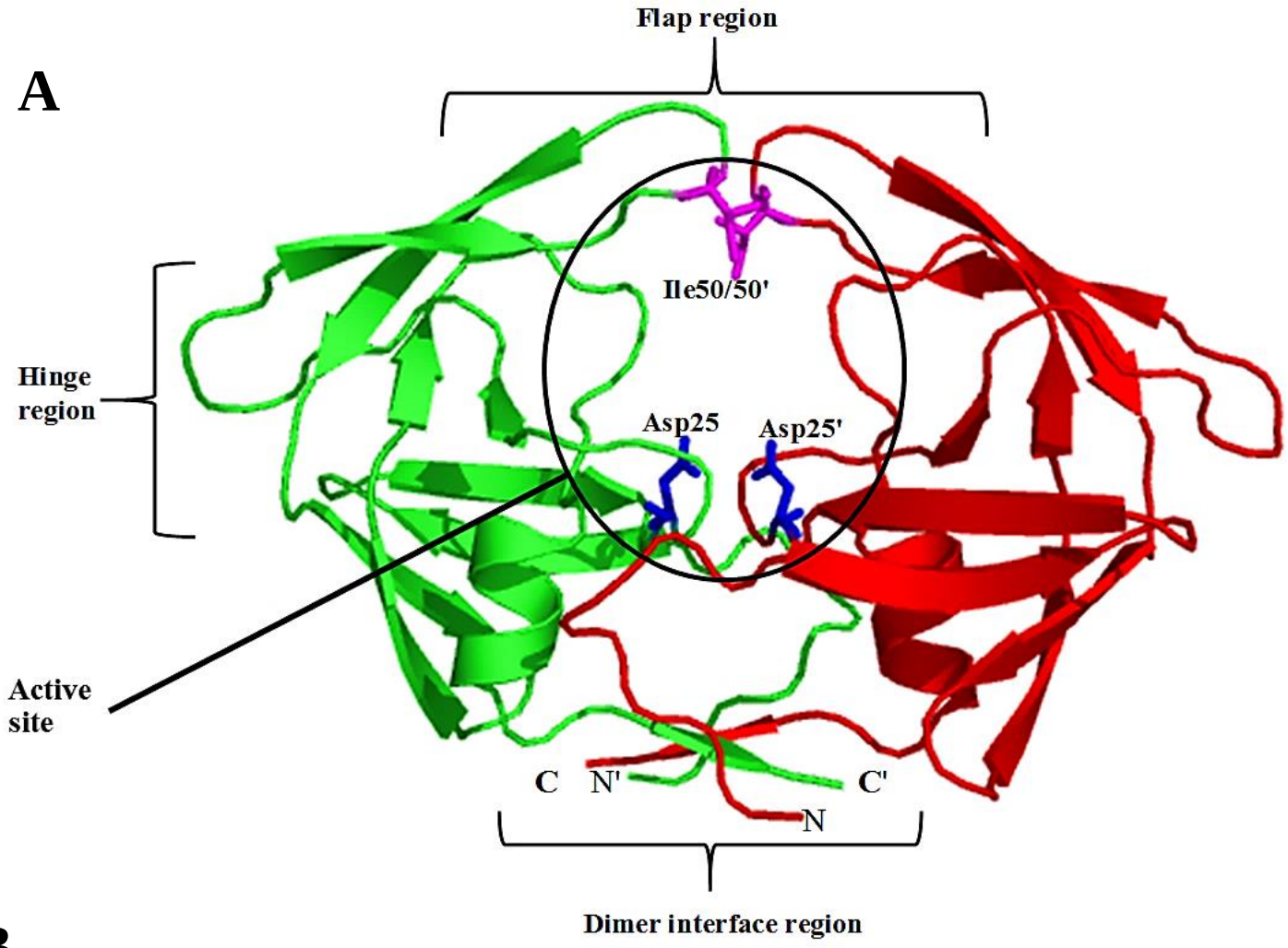
1.2 HIV-1 protease

HIV-1 protease is a C_2-symmetric homodimeric enzyme which folds into a globular protein (Naicker, 2013; PDB ID: 3U71). It consists of 198 amino acids (99 from each monomer), with a subunit molecular mass of 10754.8 Da (~11 kDa) and an isoelectric point of 9.32 (Naicker, 2013; Mosebi, 2008; Gasteiger *et al*., 2005). The enzyme consists of two α helices, nine β strands and predominantly random coils (Toh *et al*., 1985; Richard *et al*., 1989; Trylska *et al*., 2007). The homodimer is formed by interactions between four antiparallel β-strands involving both the amino- and carboxy-termini of each monomer. According to several studies, the HIV-1 protease belongs to the class of the aspartic proteases. HIV protease sequence homology compared to other cellular aspartic protease shows that the enzyme has the sequence Asp-Thr-

Gly, which is conserved among the aspartic protease enzymes. The inhibition of HIV protease *in vitro* by pepstatin (a natural product that selectively inhibits aspartic protease) further supports that HIV protease is a member of the aspartic protease family (Hansen *et al*., 1988; Darke *et al*., 1989; Bessong, 2008). Its primary function is to cleave long polypeptides from translated *gag* and *gag-pol* genes. The protease structure contains distinct regions in each monomer; namely, the active site, the hinge region, flap region and the dimer interface regions (Figure 5). Figure 5 illustrates the HIV-1 South African subtype C protease (C-SA PR), and it also includes the eight positions in each monomer that the C-SA PR differs from the consensus subtype B PR. There are four crystal structures of HIV-1 C-SA PR that have been deposited in the protein data bank (PDB ID: apo form – 3U71, 2R8N, and inhibitor-bound form – 2R5P, 2R5Q).

1.2.1 The active site

The active site of HIV-1 subtype C PR is an asymmetrical cavity that is found below the flap region (Navia *et al*., 1989), as viewed in Figure 5. The dimeric structure comprises of one Asp-Thr-Gly triad to the pseudo-symmetric active site in each monomer positioned at Asp25/25', Thr26/26' and Gly27/27' (Miller *et al.,* 1989; Dash and Rao, 2001). Data related to the roles of Thr26/26' and Gly27/27' is not fully known. It is hypothesised that strong hydrogen-bond forces between Thr26/26' residues stabilise the conformational state of the active site, by forming a hydrogen network known as a 'fireman's grip' (James and Sielecki, 1983). Asp25 hydrogen bonds to the backbone nitrogen Gly27 (this is often referred to as the fireman's grip). Thus, it is hypothesised that the Gly27/27' function is to accommodate the binding of substrate in a position that the catalytic Asp25/25' carboxylate groups can cleave the amide moiety of the substrate (Weber, 1990; Coman *et al*., 2008). Asp25/25' in the active site are in opposite states of protonation and are essential for catalysis of the substrate. Asp25 is protonated and plays a role in hydrogen bonding with substrates. Asp25' exists in its deprotonated form and nucleophilically attacks the carbonyl carbon of the scissile peptide bond that is recognised for hydrolysis (Debouck *et al*., 1987; Walsh, 2003). The active site also spans eight subsites (S4 - S1 and S1' - S4') which are specific pockets in the binding site of the protease and positions with the substrate (P4 - P1 and P1' - P4') (Wlodawer and Vondrasek, 1998; Prabu-Jeyabalan *et al*., 2000). This is depicted in Figure 6 showing Schechter and Berger conventional nomenclature used to designate amino acid residues of peptide substrates. The S2 – S1 and S2' – S1' subsites are hydrophobic except for Asp-29/ 29', Asp30/ 30'. The S3 subsites are mostly hydrophobic and they are adjacent to S1 subsites. The peptide bond that is cleaved by an HIV-



B

```
                                12  15  19                          36      41                                  69                                  89  93
WT C-SA     PQITLWKRPL VSIKVGGQIK EALLDTGADD TVLEEINLPG KWKPKMIGGI GGFIKVRQYD QILIEICGKK AIGTVLVGPT PVNIIGRNML TQLGCTLNF
B           PQITLWKRPL VTIKIGGQLK EALLDTGADD TVLEEMNLPG RWKPKMIGGI GGFIKVRQYD QILIEICGHK AIGTVLVGPT PVNIIGRNLL TQIGCTLNF
C           PQITLWKRPL VTIKIGGQLK EALLDTGADD TVLEEINLPG KWKPKMIGGI GGFIKVRQYD QILIEICGKK AIGTVLVGPT PVNIIGRNML TQIGCTLNF
```

Figure 5: Cartoon representation of the crystal structure of the homodimer wild-type HIV-1 South African subtype C (C-SA) PR and multiple sequence alignment of different subtypes. (A) HIV-1 C-SA PR consists of α-helices and β-strands in each monomer. Each monomer is coloured differently (subunit A in green and subunit B in red). The structure has an open flap conformations. The hinge region is also shown. Dimer formation is stabilised by interface interactions between the four anti-parallel β strands. The figure was generated using PyMOL (DeLano, 2002) with PDB code 3U71 (Naicker *et al.*, 2012). (B) Sequence alignment of the primary structure for different subtypes and eight differences between subtype B and C shown in red.

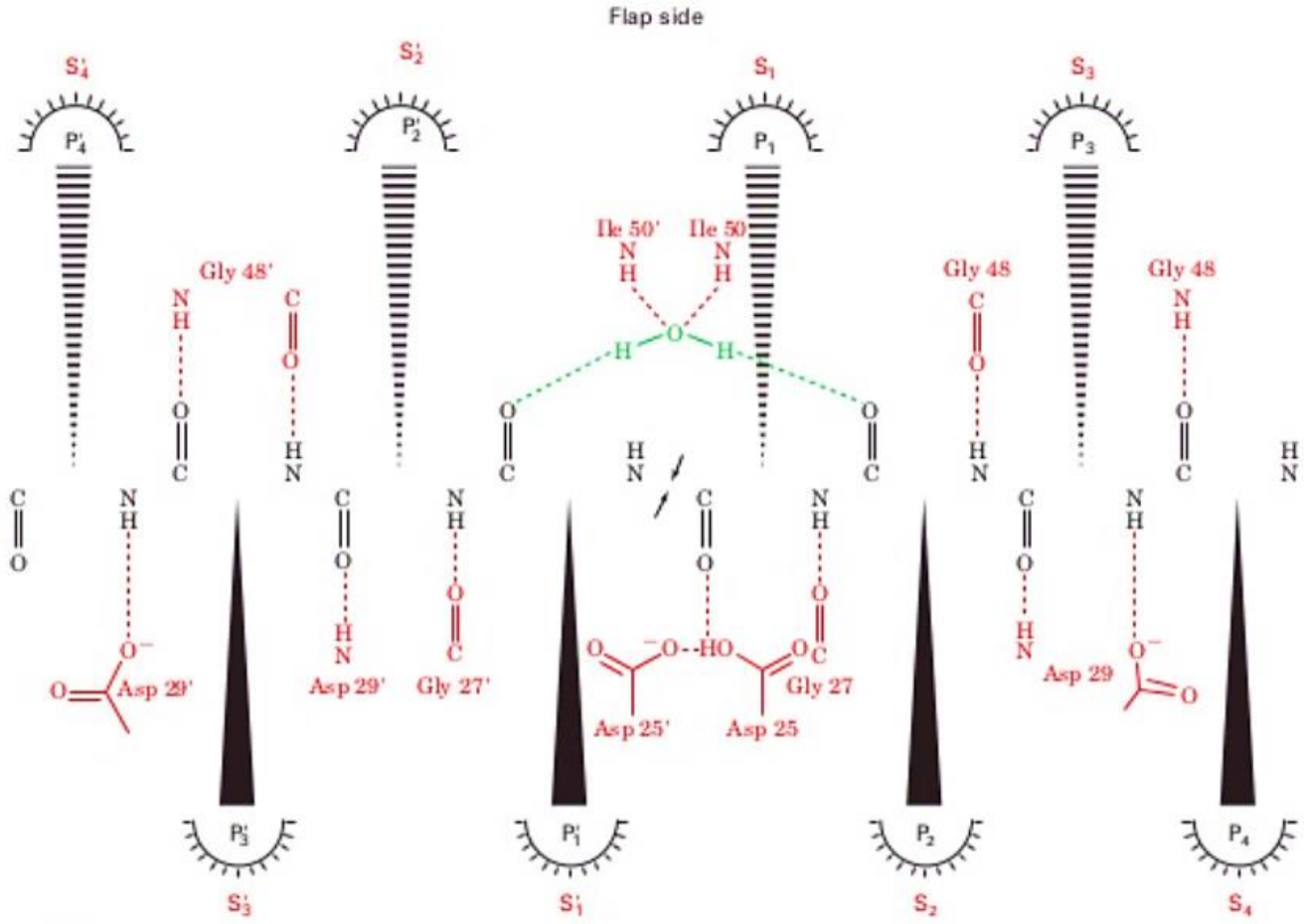


Figure 6: The arrangement of hydrogen bonds between HIV-1 protease and a modelled substrate. In the nomenclature used here, the polypeptide in one subunit is assigned primed numbers to differentiate them from the residues of the other subunit. Substrate residues on the N-terminal side of the scissile peptide bond are designated P1, P2, P3,… , counting towards the N-terminus. Substrate residues on its C-terminal side are designated P'1, P'2, P'3,…, counting towards the C-terminal side. The symbols S1, S2, S3,…, and S'1, S'2, S'3,….., referred to the corresponding residue-binding subsites on the enzyme. The scissile peptide bond is marked by arrows. The picture was taken from Wlodawer and Vondrasek, (1998).

1 protease is known as the “scissile” bond and substrate change occurs between P1 and P1' at the catalytic site (Schechter and Berger, 1967).

1.2.2 The flap region

The HIV-1 protease flap region (formed by two highly flexible glycine-dense β-sheets) plays a crucial role in the exclusion of water from the active site providing a hydrophobic environment. The flaps also control the recognition and the entry of substrates/inhibitor to the active site (Proteaset *et al*., 1994; Prabu-Jeyabalan *et al*., 2006). The flaps have a high glycine content at residues 48-52 within the sequence Met-Ile-Gly-Gly-Ile-Gly-Gly-Phe-Ile, which accounts for the flexibility of the protease (Wlodawer *et al*., 1989; Hong *et al*., 1997). In the flap region, when the enzyme is unliganded, the flaps are held in a closed position by hydrogen bonding from water molecules to the backbone of Ile 50 / Ile 50'. Thus, the flaps are flexible but still maintain a reasonable coverage of the active site in an effort to keep the hydrophobic residues buried. This conformation is referred to as the “semi-open” form. The flap opening and closing movements and stability are closely associated with the hinge region as they function as a single entity (Clement *et al*., 2004; Perryman *et al*., 2006; Coman *et al*., 2008).

1.2.3 The dimer interface

HIV-1 protease also consists of a dimer interface which stabilises the two monomers for dimerisation. The dimer interface is formed by a four-stranded antiparallel β strands. Dimerisation interaction is required for substrate binding and for the formation of the active site (Richards *et al*., 1989; Weber, 1990; Ishima *et al*., 2007; Bessong, 2008). Dimer formation is stabilised by the interface interaction between the N-terminal and C-terminal (residues 1 – 4; and 96 – 99) residues (Todd *et al*., 1998).

1.2.4 The catalytic mechanism of HIV-1 subtype C protease

Proteases catalyse the hydrolysis of peptide bonds with high catalytic proficiency and sequence selectivity. Protease accomplishes catalysis through two different mechanisms; thus, the protease enzyme is divided mechanistically into two broad classes. The first class of the protease uses an activated water molecule as a nucleophile to attack the amide bond carbon of the substrates scissile bond. The water molecule activation can be obtained either through zinc cation (the zinc metallo-proteinases) or the two aspartyl β-carboxy group at the active site. For the second class, a nucleophilic atom of an amino acid side chain is used to commence the hydrolysis of the amide bond. HIV-1 protease belongs to the first class (the aspartate proteases)

with a general acid-base mechanism. The HIV-1 protease mechanism in Figure 7 illustrates that the Asp group that is closer to the nucleophilic water molecule is negatively charged. The carbonyl group in the substrate scissile bond is attacked by the nucleophilic water molecule held between the catalytic aspartates after its activation by the negative aspartate side chain, to generate an oxyanion tetrahedral intermediate. The protonation of scissile amide N atom and rearrangement results in the breakdown of the tetrahedral intermediate to the hydrolysis products. This general acid-base mechanism of the HIV-1 protease precludes the use of a Lewis acid such as Zn^{2+} (the zinc metallo-proteinases) and the formation of covalent acyl intermediates (Pearl and Blundell, 1984; Brik and Wong, 2002). The pH-range profile for the hydrolysis reaction catalysed by HIV protease reveals that substrates and inhibitors only bind to the protease when one of the two active site aspartic acid side chains is protonated in the active pH range (Hyland *et al*., 1991; Smith *et al*., 1996). There are four possible protonation states of the two active site aspartic acids: the dianionic form, the two monoanionic form and the deprotonated or neutral form. It has been proposed that the protonation state of the two Asp groups depends on the local environment near the aspartate and is different for each inhibitor. Meek and co-workers' study on the pH-range profile of a model substrate for HIV-1 protease showed that the side-chains have different pK_a values of 3.1 and 5.2 and in the presence of a pepstatin inhibitor, one aspartic acid is protonated in the pH range of 2.5–6 (Smith *et al*., 1996). The differences in pK_as of these residues is more a function of their proximity to each other than of their differing in the environment (Pearl and Blundell, 1984; Hyland *et al*., 1991).

1.2.5 HIV-1 protease mutations

Mutations can alter and affect the functional kinetics of viral enzymes (Nijhuis *et al*., 1999). More mutations are selected in response to protease inhibitors than any other class of ARVs (Shafer & Schapiro, 2008). In the protease, primary resistance mutations take place at/near the enzyme active site (e.g. the location these mutations are I50L, V32I, I47V, G48V, V82F, and I84V). Although the binding affinity of inhibitors is lowered, it still maintains the catalytic activity. Secondary resistance mutations lower the binding affinity of inhibitors by about three orders of magnitude (Rose *et al*., 1996; Muzammil *et al*., 2003). Viral resistance to treatment in the presence of protease inhibitors is enhanced by factors such as the high mutation rate and the lack of proofreading ability of reverse transcriptase as well as viral replication dynamics.

Mutations in the protease generated drug resistance, which resulting in reduced efficacy of HAART and as such, all current FDA-approved PIs have specific significant mutations associated with them. For example, patients administrated with fosamprenavir commonly have

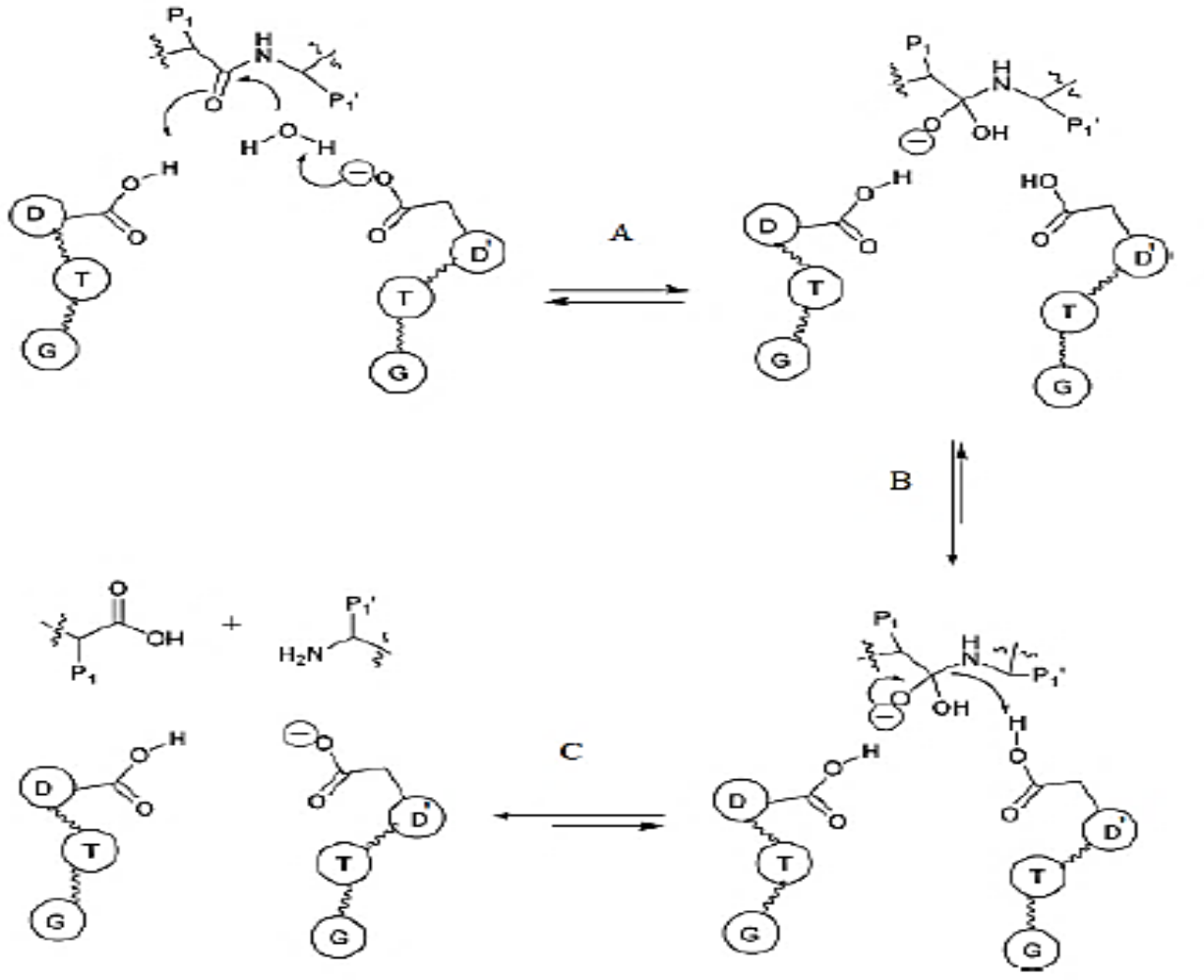


Figure 7: The catalytic mechanism for aspartic proteases. (A) Deprotonation of H_2O by charged Asp', and nucleophilic attack on the carbon involved in the scissile peptide bond by the free hydroxyl group. Thus, the tetrahedral intermediate is formed. (B) Nucleophilic attack on carbonyl carbon by charged carbonyl oxygen, leaving the nitrogen atom charged, which in turn, nucleophilically attacks the hydrogen present on D' (from H_2O). (C) The hydrolysed product is released by protonation of the leaving N-terminus of the cleaved scissile peptide bond (P1'). The protease resets and ready to catalyse again. The figure was adapted from Brik and Wong, (2002).

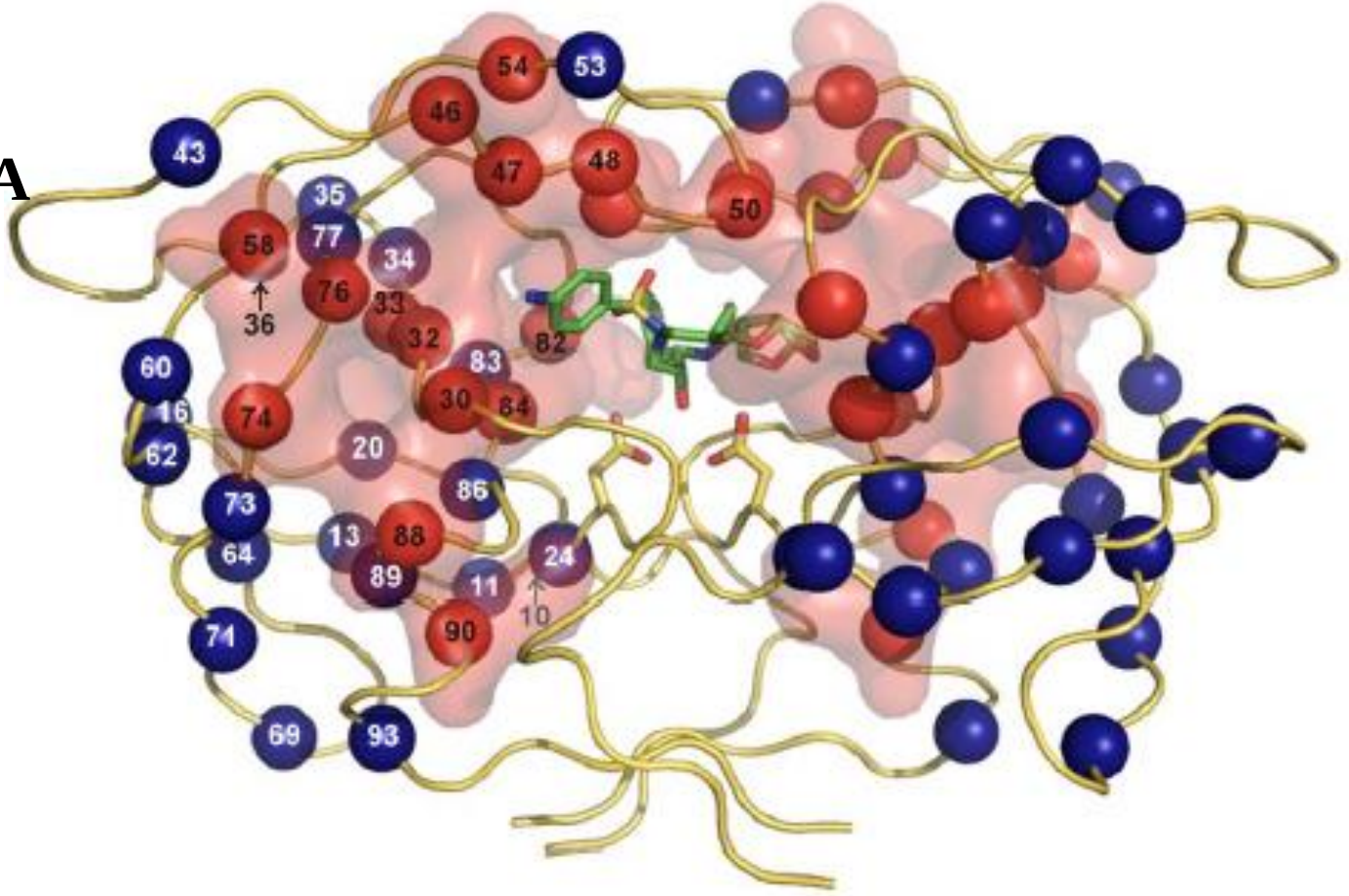
the mutation I84V and resistance to atazanavir, indinavir, darunavir and tipranavir is also seen with this mutation from other viruses. The I84V mutation changes the larger amino acid to a less bulky residue; thus, the van der Waals contacts between the protease and its inhibitor are reduced, thereby, reducing its affinity (Floridia *et al*., 2004; Torres and Arduino, 2007).

The pattern of mutations associated with viral resistance is extremely complex as seen in Figure 8. Mutations are not only found in the substrate binding cleft but also seen outside the active site of PR (such as the hinge region which is prone to insertions) (Todd and Freire, 1999). Major mutations substantially reduce PIs susceptibility and minor mutations emerge later and they improve the replication capacity of viruses containing major mutations. HIV occasionally rescues its fitness under the selection pressure of PIs by changing the PR substrate (the polyprotein cleavage sites), as such, mutations take place without changes in PR and cause viral resistance (Kim *et al*., 2001; Kozisek *et al*., 2008). Also, natural polymorphisms can affect more than 45% of the amino acid content in the protease molecule.

1.3 HIV-1 subtype C protease inhibitors

Ever since the structure of HIV-1 protease became available in 1989, intensive efforts were mounted in numerous laboratories to find therapeutically effective inhibitors of the enzyme. To date, HIV-1 protease inhibitors continue to play an important role in the treatment of HIV/AIDS, transforming this deadly ailment into a more manageable chronic infection. Protease inhibitors are drugs that block the protease from generating mature infectious viral particle. Over the years, intensive research led to a variety of approved protease inhibitors for the treatment of HIV/AIDS. There are currently ten protease inhibitors in clinical use for HIV-1 treatment (Figure 9). These are Amprenavir, Atazanavir, Darunavir, Fosamprenavir, Indinavir, Lopinavir, Nelfinavir, Saquinavir, Ritonavir and Tipranavir (Fitzgerald and Springer, 1991; Wlodawer and Vondrasek, 1998; Piliero, 2002; Broder, 2010). There are structural similarities within these inhibitors – they all possess hydrophobic aliphatic and aromatic groups that mainly make interaction with the lipophilic residues in the protease binding groove. They all also contain an un-cleavable hydroxyl group that comes in contact with the catalytic aspartic acid residues. These protease inhibitors are also in clinical use and active against HIV-2; however, some of these inhibitors are less effective compared to HIV-1 protease (Witvrouw *et al*., 2004).

A



B

Protease inhibitors	Major mutations	Minor mutations
Amprenavir	M46I/L, I47V, I50L/V, I84V	
Atazanavir	I50L, I84V, N88S	1, 16, 20, 24, 32, 33, 34, 36, 46, 48, 53, 54, 60, 62, 64, 71,
+/- ritonavir		73, V82A, 85, 90, 93
Darunavir	I50V, I54L/M, L76V, I84V	V11I, V32I, L33F, I47V, G73S, T74P, L89V
Fosamprenavir	I50V, I84V/A	10, V32I, 46, I47V/A, I54L/M, 73, L76V, 82, 90
Indinavir	46, 82, 84	10, 20, 24, 32, 36, 54, 71, 73, 76, 77, 90
Lopinavir	32, I47A, V82A	10, 2, 24, 33, 46, 50, 53, 54, 63, 71, 3, 76, 84, 90
Nelfinavir	D30N, L90M	10, 36, 46, 71, 77, 82, 84, N88D
Saquinavir	G48V, L90M	10, 24, 54, 62, 71, 73, 77, 82, 84
Tipranavir	33, 47, 58, 74, 82, 84	10, 13, 20, 35, 36, 43, 46, 54, 69, 83, 90

Figure 8: HIV protease structure depicting mutations associated with resistance to FDA-approved protease inhibitors. A. Represents mutated residues with their coloured Cα atoms (spheres) in red (major mutations) and blue (minor mutations). The semi-transparent solvent accessible surface for major mutations is shown in red. The structure is bound to DRV which is shown in stick models (PDB code 3GGU) (Saskova *et al*., 2009) and generated using the program PyMol (DeLano, 2002). B. Major and minor mutations in the protease and the association with protease inhibitors. Major mutations substantially reduce PI susceptibility and minor mutations emerge later and they improve the replication capacity of viruses containing major mutations. (Adapted from an International AIDS society report).

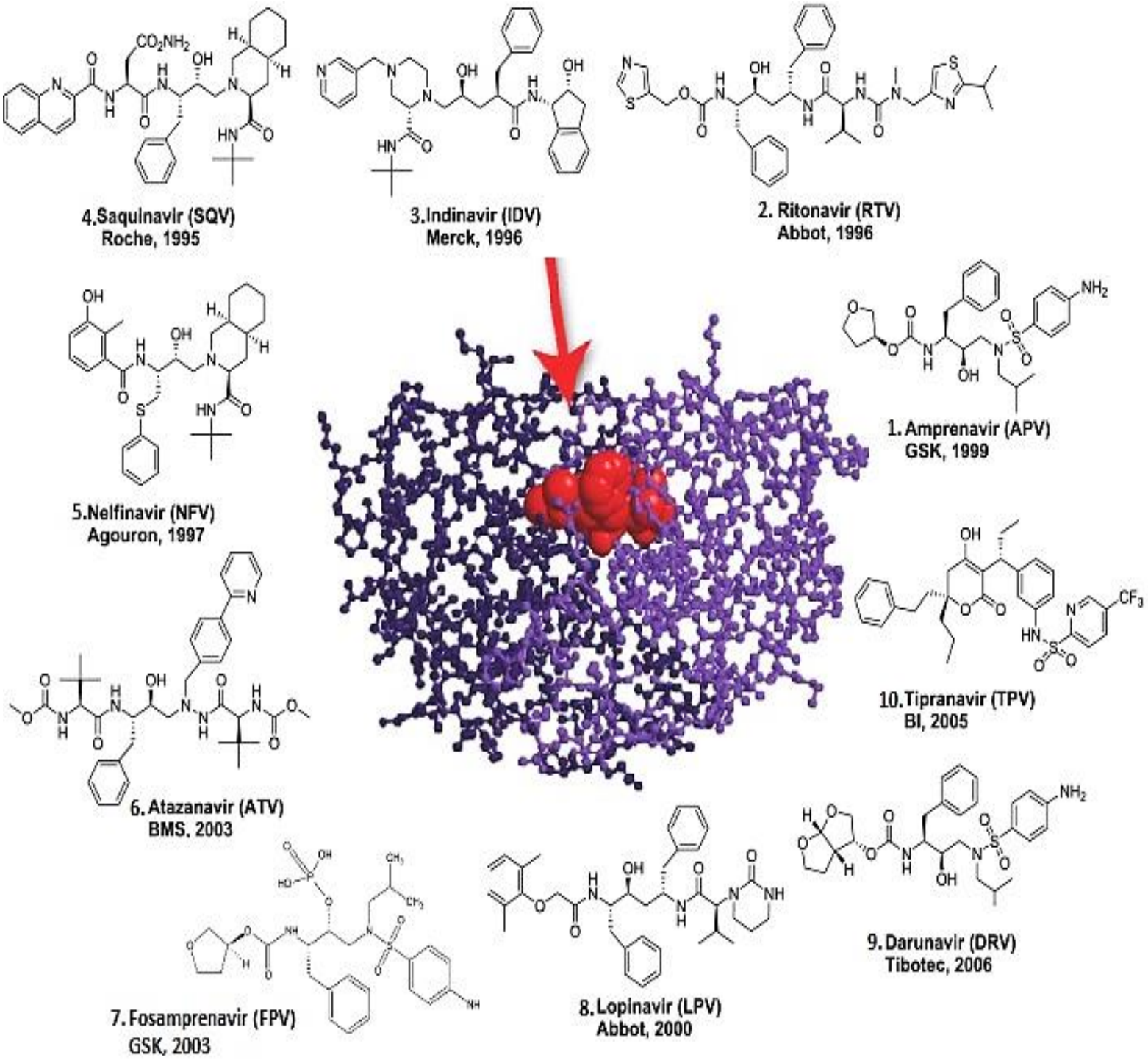


Figure 9: The chemical structures of FDA-approved HIV protease inhibitors. Numbers 1-5 are first-generation, numbers 6-8 are the second-generation and number 9 & 10 are the third-generation HIV protease inhibitors. In the middle is an HIV protease showing protease inhibitor in red binding to the active site. The figure was generated using Swiss-PDB Viewer (Guex and Peitsch, 1997) with PDB code 3GGU.

1.3.1 First-generation protease inhibitors

First-generation protease inhibitors are based on renin and pepsin inhibitors (pepstatin), based on the idea to create an inhibitor that can bind in the active site as a substrate analogue but be uncleaved by the protease. Saquinavir, indinavir and nelfinavir were the first HIV-1 protease inhibitors which were derived from the polypeptide sequences being cleaved by the PR. First-generation protease inhibitors were based on hydroxyethylamine and hydroxyethylene isosteres. They are peptide-like and have a poor bioavailability (less than 4%) (Fitzgerald and Springer, 1991). However, HIV-1 rapidly developed resistance against these inhibitors. The free energy of binding is due to an entropy gain associated with filling hydrophobic pockets in the enzyme with hydrophobic substituents on the inhibitors. The enzyme hydrophobic pockets are sensitive to mutation; thus, the virus could easily prevent an inhibitor from binding by mutation of residues forming the hydrophobic pockets (Fitzgerald and Springer, 1991; Wlodawer and Vondrasek, 1998; Piliero, 2002; Xu and Desai, 2009).

Saquinavir (Invirase®, Fortovase®) is a pentapeptide analogue. It is the first FDA-approved protease inhibitor to undergo clinical trials and to be marketed on 6th December 1995 (designed by Hoffman-La Roche), and that marked the beginning of HAART as second-line therapy. Saquinavir combination therapy with RT inhibitors greatly improved the CD4 cell counts, reducing viral loads and halting the progression to AIDS. Saquinavir was co-formulated with ritonavir as a booster (Perry and Noble, 1998; Kilby *et al*., 2000).

Ritonavir (Norvir®) was designed by Abbot Laboratories and FDA-approved on 1st March 1996. The inhibitor was derived from a C_2 symmetric molecule; therefore, its structure is asymmetrical which improved the bioavailability by different modification of the end group. Ritonavir was gradually abandoned as an antiretroviral drug due to the repeated occurrence of associated side effects (such as gastrointestinal symptoms) and high cross-resistance with other protease inhibitors. It was also proven to be a potent inhibitor of cytochrome P-450 3A4, the microsomal enzyme responsible for the bulk metabolism of other protease inhibitors. Thus, it is exclusively used as a pharmacokinetic boosting agent of various protease inhibitors, which increases the plasma concentration of other protease inhibitors; therefore, prolonging their therapeutic effects (Doyon *et al*., 1996; Bold *et al*., 1998).

Indinavir (Crixivan®) was developed by Merck & Co and FDA-approved on 13th March 1996. It comprises a hydroxyaminopentanamide transition state isostere which is derived from the framework of previously developed renin inhibitors. The intake of indinavir consists of side

effects (such as nephrolithiasis and lipodystrophy syndrome) and is taken as a second/third-line cART therapy in case of intolerance or resistance (Patick *et al*., 1997).

Nelfinavir (Viracept®) is manufactured by Agouron Pharmaceuticals and was FDA-approved on 14th March 1997. It was the first protease inhibitor to be optimised by Monte Carlo simulation and the first to be used by people living with HIV (Bardsley-Elliot and Plosker, 2000).

Amprenavir (Agenerase®) is manufactured by Vertex Pharmaceuticals and GlaxoSmithKline and FDA-approved on 15th April 1999. It is derived from saquinavir but the P_2 end group is replaced by tetrahydrofurancarbamate and the P1'-P2' moiety by a sulfonamide derivative. The structure has some advantages over saquinavir such as higher oral bioavailability, easier synthesis and fewer chiral centres. Mutations occurring in the *gag* substrate cleavage site (p1/p6) at positions L449 and P453 lower binding affinity of amprenavir. The intake of amprenavir is harmful and could lead to hypercholesterolemia as well as hypertriglyceridemia and frequently occurring side effects causing gastrointestinal symptoms (such as abdominal pain, loose stools and skin rash) (Adkins and Faulds, 1998; Dube *et al*., 2002).

1.3.2 Second-generation protease inhibitors

In order to circumvent the problem of protease inhibitors only primarily binding through hydrophobic effects, protease inhibitors were designed to have an addition of hydrogen-bonding to the enzyme backbone atoms (amide NH group and carbonyl oxygen atom). These hydrogen-bonding contacts do not involve the side chains but only the backbone atoms; thus, protease inhibitors are less affected by mutations that occur in the protease. The second-generation protease inhibitors are characterised by entropic and enthalpic effects contributing to the free energy of binding, making HIV-1 less likely to develop resistance.

Lopinavir (Kaletra®) was developed by Abbott Laboratories and been in the market since 15th September 2000. It is recommended as the first-line option for protease inhibitor based on the HAART regimen. This PI is used in a fixed-dose combination with ritonavir and it also acts as an inhibitor of cytochrome P-450 CYP3A (Sham *et al*., 1998).

Fosamprenavir (Lexiva®) is a hydrophilic phosphate ester prodrug of amprenavir which was developed in order to overcome the therapeutic potency of amprenavir, improving its safety and pharmacokinetics. It is manufactured by Vertex Laboratories and GlaxoSmithKline and FDA-approved on 20th October 2003. The toxicity and side effects are comparable to those of

amprenavir and it is recommended as the first-line option for PI therapy based on the HAART regimen (Floridia *et al*., 2004; Torres and Arduino, 2007).

Atazanavir (Reyataz®) was designed by Ciba-Geigy and was FDA-approved on 20th June 2003. The structural design was part of a complex exploration of azadipeptide analogues, it is also the bulkiest of all protease inhibitors and it was the first protease inhibitor that was compatible with once-daily dosing. It is recommended as the first-line option for PI therapy based on the HAART regimen (Bold *et al*., 1998; Le Tiec *et al*., 2005).

1.3.3 Third-generation protease inhibitors

The third-generation protease inhibitors have been developed based on the careful optimisation of the structural and energetic contributions to binding. Examples of an HIV-1 PR inhibitor with unique binding characteristics are tipranavir and darunavir.

Tipranavir (Aptivus®) was developed at Pharmacia & Upjohn, later introduced by Boehringer-Ingelheim and FDA-approved on 22nd June 2005. It is a non-peptidic inhibitor which belongs to the 4-hydroxy-5,6-dihydro-2-pyrone sulfonamides. Tipranavir is a highly potent inhibitor which primarily binds to the HIV-1 protease through entropy effects. The release of buried water from the active site most likely caused the unusually high binding entropy. It is recommended as a second-line option in HAART therapy in case of resistance (King and Acosta, 2006; Macias *et al*., 2009).

Darunavir (Prezista®) was developed by Tibotec-Virgo NV and FDA-approved on 23rd June 2006. It is a structural analogue of amprenavir, but it differs by the presence of a bis-tetrahydrofuranyl moiety which allows the forming additional hydrogen bonds with the conserved region of the protease (Koh *et al*., 2003; Kovalevsky *et al*., 2008). These interactions with the backbone are important since there is a minimal change in conformation between wild-type and resistant strains. It is a nonpeptidyl inhibitor and the most potent HIV-1 PI currently that was approved for first-line HAART in 2007 (Ghosh *et al*., 2007).

1.4 HIV-1 protease inhibitor test compounds

The test compounds with a general core structure containing 4-hydroxy-2H-pyran-2-one (inhibitors) were tested in this study. These compounds were mentioned in the paper by Prasad *et al.*, (1995) are similar in nature; thus, the article was useful in the design and analysis of the inhibitors (Figure 10). Prasad *et al.*, (1995) demonstrated that potent enzyme inhibition can be

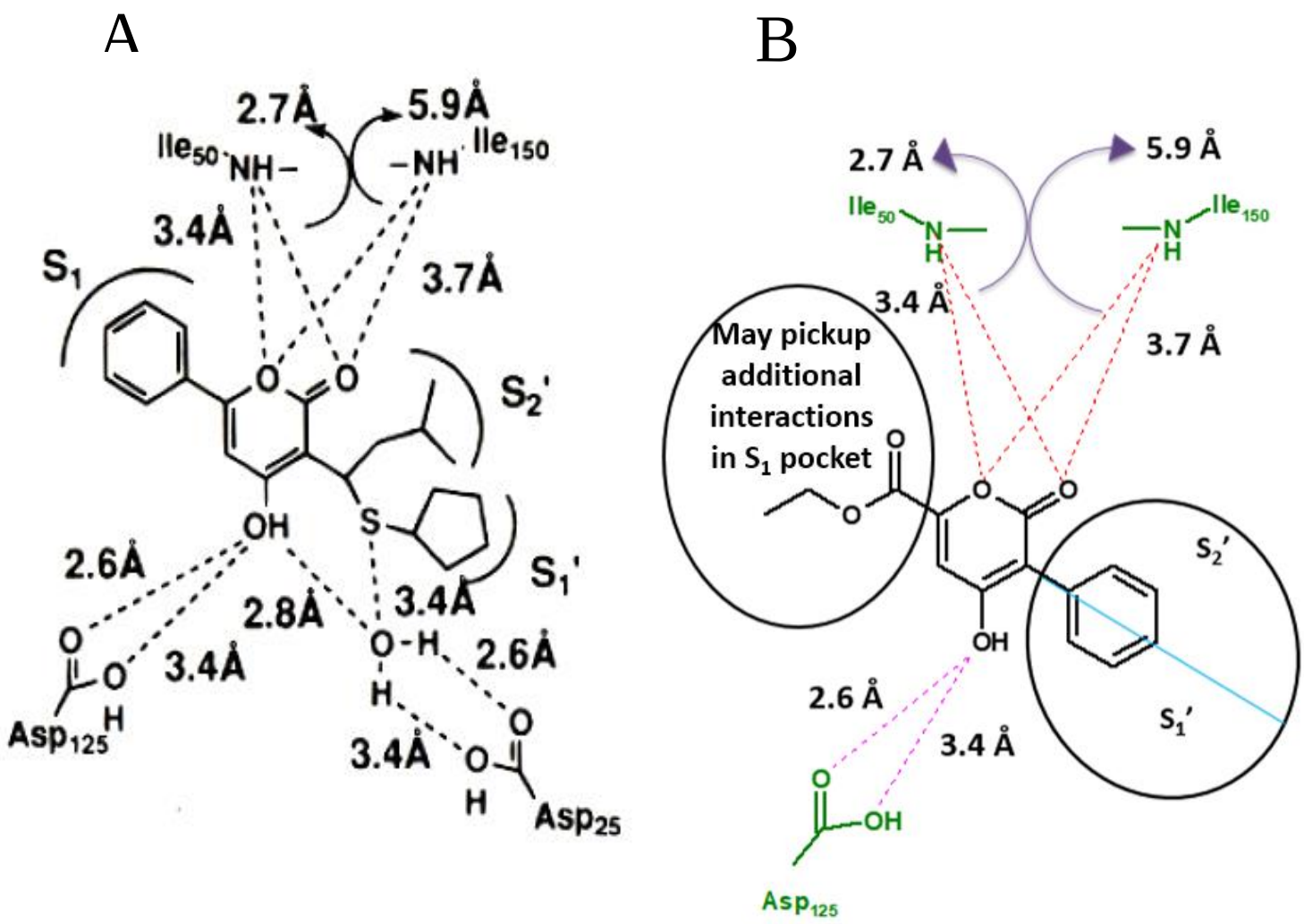


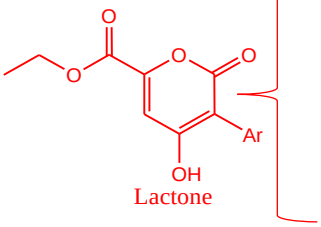
Figure 10: General structures of protease inhibitor compounds. The representative structure is 4-hydroxy-2H-pyran-2-one which is the core structure of the protease inhibitor test compounds. Image A was taken from *J. Med. Chem*. 1995, 38, 898 (Prasad *et al*., 1995) and image B is the test compound's representative structure of the protease inhibitor.

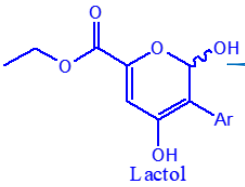
achieved by inhibitors that span only three subsites. They demonstrated that the pyran-2-one template is a versatile nonpeptide scaffold that can be substituted to mimic related peptide-based inhibitors of HIV PR. There work extends on previous studies in the design and discovery of nonpeptide ligands at receptor sites as derived from β-D-glucose, steroid, and pyrrolidinone templates. The compounds were not formulated into clinical drugs as there was no correlation between enzymatic activity and antiviral activity. However, research from the group made it possible for the design of more HIV PR potent inhibitors using iterative structure-based design strategies (Table 1).

The test protease inhibitor compounds are not related to the PR inhibitor ARVs that are currently in clinical use. The test compounds possess a gem-diol mimicking intermediate and the general compound structure is less bulky compared to the FDA-approved PIs. These protease inhibitor compounds that are to be tested are represented in Table 1 and they were a kind gift from Dr Shankar R. Thopate (Department of Chemistry, Ahmednagar College, Ahmednagar, India). They have a general compound structure of 4-hydroxy-2H-pyran-2-one that is heterocyclic and complexed with pyran-2-one, a series of nonpeptidic inhibitors of the HIV-1 C-SA PR (Figure 10). The information derived from the molecular modelling and X-ray crystal structure of a single pyran-2-one-based analog, complexed with HIV PR, led to the design and analysis of a series of second-generation analogs predicted to interact with three binding sites of the target enzyme. The molecular modelling studies and X-ray crystal structure of the nonpeptidic potent protease inhibitors bound to the HIV protease revealed that the 4-hydroxy group of the pyran-2-one moiety interacts with the catalytic Asp25 side chain. The lactone moiety (Table 1) interacts with the Ile50 and Ile50' residues in the homodimeric enzyme (Prasad *et al.*, 1995). The *S*-cyclopentyl and isobutyl groups of the nonpeptidic potent protease inhibitor shown in Figure 10 occupied the S1' and S2' binding sites, respectively. The structural water-301 (inhibitor-flap-bridging water) that is found in HIV protease-peptidomimetic inhibitor complexes (Fitzgerald, 1993) was displaced by the lactone carbonyl of the pyran-2-one ring. It forms a hydrogen bond with the NH of the Ile50. The enol moiety of the inhibitors interacts with Asp25 through a bridging water molecule. The pyran-2-one template is a versatile nonpeptide scaffold that can be substituted to mimic related peptide-based HIV protease inhibitors (Prasad *et al*., 1995).

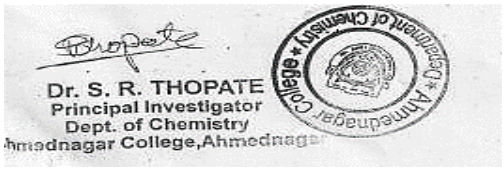
Docking analyses have been done using an HIV protease, pseudo-symmetric inhibitors (PDB ID: 3GGV) as the target protein structure for these novel protease inhibitors. Dr Rakesh Joshi,

Table 1: Novel protease inhibitor compounds. The compounds are from Dr Shankar R. Thopate (Department of Chemistry, Ahmednagar College, Ahmednagar, India).





No	Ar	Code	Molecular Formula	Molecular Weight	Amount (mg)	Nature	Docking Score
1	2,4-Dichloro phenyl-	GRD 100D	$C_{14}H_{10}C\ell_2O_5$	329.14	12	Solid	-7.8
2	4-Fluoro phenyl-	GRD 100J	$C_{14}H_{11}FO_5$	278.24	24	Solid	-7.3
3	2-Naphthyl-	GRD 100N	$C_{18}H_{14}O_5$	310.31	18	Solid	-8.5
4	Phenyl-	GRD 115	$C_{14}H_{14}O_5$	262.25	22	Solid	-8.6
5	3-Methy phenyl-	GRD 115D	$C_{15}H_{15}O_5$	276.28	13	Solid	-7.8



Assistant Professor, Institute of Bioinformatics and Biotechnology (IBB) Savitribai Phule Pune University, assisted Dr Shankar R. Thopate regarding the analysis of the molecular docking could have high potency and improved ability to overcome the diversity of mutation in the protease. They were designed so that they could establish strong hydrogen bonding interactions with conserved amino acids of the HIV-1 PR active site that cannot be mutated/changed without the enzyme losing its function. All PIs that are FDA-approved, except tipranavir, make this interaction through water molecules and they expend energy to mobilise HIV PR. The release of the water molecule is an energetically favourable event (Matsumoto *et al*., 2001). These interactions with water have not been fully observed or investigated with some of the HIV protease crystal structures; thus, the resulting intermolecular hydrogen bond network of the HIV PR complex exemplifies a unique mode of binding (Figure 10). The current studies will examine the novel inhibitor compounds to see if they might have high efficiency to bind the HIV-1 C-SA protease.

An understanding of the HIV-1 protease complexed with an inhibitor on a molecular level is still an ongoing investigation. The HIV-1 protease was studied at the molecular level in order to understand its activity and the comparative study between subtype B and C was done to help understand the nature of interactions between protease inhibitors and proteases from the B and C subtypes. The structural characterisation and enzymatic studies of HIV-1 protease have been extensively based on subtype B protease compared to non-subtype B proteases (Fleury *et al*., 2003; Kantor and Katzenstein, 2003; Martinez-Cajas *et al*., 2009; Wainberg and Brenner, 2010).

The current study addressed the structural characterisation and enzymatic activity of South African wild-type HIV-1 subtype C protease (WT C-SA PR). A comparison with published work was also performed.

1.5 Aim and objectives

1.5.1 Aim

The aim of the study was to assess the structural characteristics and enzymatic activity of wild-type HIV-1 subtype C protease. This would shed light on the drug binding mechanism of HIV-1 protease subtype C as well as contribute to identifying functional differences in comparison to subtype B protease. Perform *in silico* molecular docking studies and comparative binding mode analysis of all ten FDA-approved HIV-1 protease inhibitors to investigate the conformational dynamics of HIV-1 protease.

1.5.2 Objectives

1. Induction studies to optimise the expression of wild-type HIV-1 C-SA protease.
2. Overexpression and purification of wild-type HIV-1 subtype C protease.
3. Protease active site titration using an isothermal titration calorimeter (ITC).
4. The determination of secondary, tertiary and quaternary structure characterisation of subtype C protease in the absence of inhibitors using far-UV circular dichroism, fluorescence spectroscopy and SE-HPLC, respectively.
5. Kinetic characterisation – specific activity for the fluorogenic substrate; K_M; V_{max}; k_{cat}; k_{cat}/K_M.
6. IC_{50} and K_i determination.
7. Induced-fit docking using molecular simulation to analyse the binding of FDA-approved inhibitors to HIV-1 C-SA protease.

CHAPTER 2

2 METHODS AND PROCEDURES

2.1 Transformation of *E. coli* BL21 (DE3) pLysS with pET-11b

*Escherichia col*i BL21 (DE3) pLysS cells were transformed with a plasmid pET-11b encoding the wild-type HIV-1 C-SA protease gene. The plasmid, pET-11b, containing the wild-type HIV-1 C-SA protease insert was purchased from GenScript (Piscataway, NJ). *E. coli* BL21 (DE3) pLysS is a protein expression cell line, particularly useful when expressing cytotoxic proteins like HIV-1 protease. The DE3 designation shows that the strains contain the DE3 lysogen that carries the gene for T7 RNA polymerase under control of the *lac*UV5 promoters (Studier and Moffatt, 1986; Stano and Patel, 2004). Isopropyl-β-D-1-thiogalactopyranoside (IPTG) is required to induce expression of the T7 RNA polymerase. The pLysS plasmid produces the T7 lysozyme which reduces basal level expression of the heterologous gene of interest. Thus, the strain is useful when expressing toxic genes (Studier and Moffatt, 1986).

The pET-11b vector (Novagen®) is used for the expression of recombinant proteins and encodes the T7 promoter upstream from the recombinant gene inserted into the multiple cloning sites. The pET-11b expression vector confers resistance to ampicillin and the pLysS lysogen contains a selectable marker that confers resistance to chloramphenicol. Therefore, the appropriate antibiotics for inoculation and Lysogeny Broth (LB) (Bertani, 2004) agar plating of BL21 (DE3) pLysS transformation with the pET-11b vector used were 100 μg/mℓ ampicillin and 35 μg/mℓ chloramphenicol. The basic transformation protocol for BL21 (DE3) pLysS was provided by Invitrogen™ (One Shoot® BL21 (DE3) pLysS) using the heat-shock method. Briefly, 50 μℓ competent BL21 (DE3) pLysS cells were thawed on ice for 30 minutes, then 5 μℓ of 1 ng/μℓ plasmid DNA was mixed with the competent cells by gently tapping the bottom of the tube. The competent cells/DNA mixture was then placed on ice for 30 minutes and treated using heat-shock by incubating the tube into a 42 °C dry heating block for 30 seconds and immediately placed on ice for 2 minutes. Then 250 μℓ of pre-warmed SOC medium (2% (w/ v) tryptone, 0.5% (w/ v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) was added to improve the transformation efficacy (Hanahan, 1983). The cells were then grown at 37 °C in a shaking incubator at 230 rpm for 1 hour. The cells were then plated onto the agar plates containing sterile LB agar media [1.5% (w/v) agar,

1% (w/v) tryptone, 0.5% (w/v) yeast, 0.5% (w/v) NaCl and 100 mL dH_2O] supplemented with 100 μg/mℓ ampicillin and 35 μg/mℓ chloramphenicol. The remaining volume was concentrated by centrifuging at 13 000 x g for 2 minutes, the supernatant was discarded, and the pellet was resuspended in dH_2O and plated on a different agar plate. The plates were inverted and incubated at 37 °C for 18 hours. A single colony from successfully transformed cells was selected and inoculated into sterile LB medium [1% (w/v) tryptone, 0.5% (w/v) yeast, 0.5% (w/v) NaCl and 100 mℓ dH_2O] supplemented with 100 μg/mℓ ampicillin and 35 μg/mℓ chloramphenicol, and grown at 37 °C in a shaker incubator at 230 rpm. Glycerol stocks were prepared by equal-volume dilution of the overnight culture with sterile 80% glycerol solution (which helps the cells not to die when they are stored at -80 °C because ice crystals can lyse the cell) for future experiments.

2.2 Plasmid DNA purification and verification

The pET-11b plasmid was extracted from overnight culture (T7 Express *E. coli* cells) using an alkaline-lysis method from a GeneJET Plasmid Miniprep Kit (ThermoScientific®). A glycerol stock (1% (w/v)) containing 100 μg/mℓ ampicillin and 35 μg/mℓ chloramphenicol was placed into sterile LB medium and grown in a 37 °C shaker incubator at 230 rpm for the cultivation of bacterial samples containing the plasmid DNA. A 1 mℓ sample of the overnight cell culture was centrifuged at 13 000 x g, the supernatant was discarded and the pellet was resuspended in 250 μℓ of resuspension solution. The resuspension solution (containing Tris and EDTA ions) was used to stabilise the cells and inhibit the action of DNases. The cells were then lysed through the addition of 250 μℓ of lysis solution and mixed by inverting the tube until the solution becomes slightly clear or viscous. Lysis solution is highly alkaline consisting of SDS and NaOH. A 350 μℓ volume of neutralisation solution was added to the sample and mixed thoroughly by inverting the tube repeatedly. This step helps with precipitating out denatured cellular proteins, genomic DNAs and SDS which will stick together to form a white precipitate. The chromosomal DNA and cell debris (white precipitate) were pelleted by centrifugation at 13 000 x g for 5 minutes to separate from the plasmid DNA found in the supernatant. The supernatant was then transferred to a GeneJET spin column and centrifuged at 13 000 x g for 1 minute. Then 500 μℓ of wash solution was added to the column twice to rinse the spin column (essential to get rid of contaminants using ethanol and also to avoid residual ethanol in the plasmid) and centrifuged at 13 000 x g for 30-60 seconds. Ethanol precipitation is a process used to extract the plasmid DNA. The plasmid DNA was eluted by adding 50 μℓ of elution buffer to the centre of the spin column membrane, then incubated and

centrifuge at 13 000 x g for 2 minutes and stored at -20 °C. The purified plasmid DNA was then sent to Inqaba Biotech Industries Pty (Pretoria, RSA) for Sanger sequencing using Universal pGEX3 – T7 promoter primers. The extracted plasmid DNA was then translated (using ExPASy) and compared (using Clustal Omega multiple sequence alignment) with the known wild-type HIV-1 C-SA protease to ensure that there were no changes in the nucleotide sequence that may have occurred during synthesis and transformation.

2.3 Cell growth of *E. coli* BL21 (DE3) pLysS with pET-11b

The growth behaviour of recombinant *E. coli* BL21 (DE3) pLysS cells containing the pET-11b was investigated. The cell culture was inoculated in an LB medium, with 1% (w/v) antibiotics and cultivated at 37 ^{o}C in a shaker incubator at 230 rpm. The shaking of the cell culture in the incubator helps the cells to be uniformly suspended before reading the optical density at 600 nm wavelength (OD_{600}). It also helps by keeping the cell culture aerated; thus, providing adequate oxygen for cell growth. The OD_{600} values were recorded with the time of collection using a 10 mm cuvette to measure OD_{600} on an absorbance SM1 100 UV-VIS spectrophotometer. Ideally, bacterial cultures grow at an exponential rate; thus, it is the change in the number of cells per minute, which is estimated as the change in OD_{600} per minute. At different time intervals, a sample was collected for OD_{600} reading, while continued shaking was done in a 37 °C shaker incubator at 230 rpm to maintain good aeration.

2.4 Expression trials of wild-type HIV-1 C-SA protease

Expression trials were carried out in order to identify the optimal conditions for *E. coli* BL21 (DE3) pLysS expression with the wild-type HIV-1 C-SA PR. Protein solubility was carried out to investigate if wild-type HIV-1 C-SA protein is expressed in the soluble or insoluble fraction. Then, a variation of IPTG concentrations [0.2, 0.4, 0.6, 0.8 and 1.0 mM] and induction period times for overexpression were carried out to investigate their effects on protein expression. For insoluble or soluble protein expression, overnight cell culture was prepared in sterile LB media [1% (w/v) glycerol stock, with 100 µg/mℓ ampicillin and 35 µg/mℓ chloramphenicol] and grown at 37 ^{o}C shaking incubator at 230 rpm. Then, 1% (w/v) overnight cell culture with 100 µg/mℓ ampicillin and 35 µg/mℓ chloramphenicol was pipetted into sterile LB medium, grown at 37 ^{o}C shaker incubator at 230 rpm until OD_{600} of 0.5. A 1 mℓ sample was collected and centrifuged at 13 000 x g at room temperature to harvest the cells. The supernatant (soluble) and the resuspended pellet (insoluble), were then resolved on tricine 20% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to check for expression.

For further induction trials, different IPTG concentrations (0.2 mM, 0.4 mM, 0.6 mM, 0.8 mM and 1.0 mM) were added to different flasks and one flask was used for the control (leaving the cell culture to grow without being induced with IPTG). A 1% (w/v) volume of overnight cell culture was pipetted into fresh sterile LB medium supplemented with 100 µg/mℓ ampicillin and 35 µg/mℓ chloramphenicol making a 100-fold dilution and grown 37 °C in a shaker incubator at 230 rpm until OD_{600} of 0.5. Protein expression was then induced with a different IPTG concentration (0.2 mM-1 mM final concentrations) at different temperatures (20 °C or 37 °C). A control flask was used to monitor cell growth by checking OD_{600} with 10 mm cuvette on an absorbance SM1 100 UV-VIS spectrophotometer. After every hour post induction, 1 mℓ samples were collected from each flask. The cells were then allowed to grow for 6 hours following induction using IPTG. After 6 hours, the cells were allowed to grow overnight and a 1 mℓ sample of the overnight culture was collected. The collected samples were centrifuged for 3 minutes at 13 000 x g at room temperature, and the resuspended pellets resolved on a 20% tricine-SDS-PAGE.

2.5 Overexpression and purification of wild-type HIV-1 C-SA protease

The wild-type HIV-1 C-SA PR was overexpressed as inclusion bodies (Ido *et al*., 1991; Todd *et al*., 1998) in *E. coli* BL21 (DE3) pLysS cells. The T7 Express *E. coli* BL21 (DE3) pLysS cells containing the pET-11b plasmid were grown in LB medium, supplemented with 100 µg/mℓ ampicillin and 35 µg/mℓ chloramphenicol and grown for overnight at 37 ºC shaking incubator at 230 rpm. The overnight culture was diluted 100-fold into fresh LB medium supplemented with 100 µg/mℓ ampicillin and 35 µg/mℓ chloramphenicol and grown at 37 °C in a shaking incubator at 230 rpm. At an OD_{600} of 0.4 - 0.5, overexpression was induced with 1 mM IPTG for 4 hours at 37 °C. The cells were then harvested by centrifugation at 5 000 x g (using Sorval Lynx 4000 Centrifuge, Thermoscientific®) at 4 °C. The cells were resuspended in buffer A (Tris-HCl, EDTA, and PMSF, pH 8.0) and stored at -20 °C overnight.

The cells were thawed at room temperature and 250 µℓ of 2 M $MgCl_2$ with 50 µℓ of 1 mg/mℓ DNAse were added to break genomic DNA so that it does not bind to the ion-exchange column and block sites for protein binding. The cells were homogenised and stirred on ice for 30 minutes to allow the viscosity to decrease. The cells were further lysed and disrupted by sonication (XL-2000 series, MISONIX – Ultrasonic Liquid Processors). The sonication was set through a 30 second (x 10) sonication cycles at 12 Watts, in order to release cellular contents. Protease is expressed as inclusion bodies; therefore, it was found in the insoluble

fraction of the lysate. The sonicated cells were aliquoted into 50 mℓ Beckman centrifuge tubes and centrifuged (using Sorval Lynx 4000 Centrifuge, Thermoscientific®) at 23 000 x g for 30 minutes at 4 °C. The cell debris and protease-containing inclusion bodies formed in a pellet were separated by centrifugation. The pellet was resuspended in ice-cold buffer A^+ (2% Trixon X, 1 mM PMSF) through homogenisation, and centrifuged at 23 000 x g for 30 minutes at 4 °C. The pellet with mostly proteins expressed as inclusion bodies with small amount of cell debris, was then dissolved at room temperature for 30 minutes in buffer B (10 mM Tris-HCl, 8 M urea, 5 mM DTT, pH 8.0). In the solution, the HIV-1 PR was unfolded by 8 M urea and then centrifuged at 23 000 x g for 30 minutes at 20 °C. The supernatant was then dialysed into buffer C (10 mM formic acid, 10% glycerol) using a 3 kDa cut-off snake-skin dialysis tubing to refold the protein. The solution was further dialysed into active buffer D^+ (10 mM sodium acetate, 2 mM DTT, pH 5) overnight at 4 °C. This allowed precipitation of some of the contaminants and also to activate the protease.

The active protease was purified using an AKTA Prime Plus® chromatographic system (GE Healthcare) using a 5 mℓ CM-Sepharose cation exchange column (GE Healthcare). The CM-Sepharose cation exchange column was used to separate molecules based on their net surface charge. Cation exchange chromatography, more specifically, uses a negatively charged ion exchange resin with an affinity for molecules having net positive surface charges. Since the protease has a pI of 9.3, at pH 5 the protease has a net positive charge. This will enable the protease to bind to the CM column and be eluted using a gradient elution (i.e. used to optimize elution conditions). The column was first equilibrated with fresh degassed buffer D^+, and the supernatant was loaded onto the column. The protease was eluted from the column with a 1 M NaCl gradient, pooled and dialysed into buffer D (10 mM sodium acetate, pH 5). Then the protein was concentrated by ultrafiltration, on a 50 mℓ amicon® 8050 pressure cell with a 3 kDa cut-off cell membrane, with constant stirring set at 200 rpm to ~5 mℓ before final storage at -80 ^oC.

2.6 Tricine – Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is a technique that separates proteins according to their molecular weight (Laemmli, 1970). The tricine-SDS-page technique was used to confirm the size and purity of an overexpressed and purified wild-type HIV-1 subtype C protease (Schägger, 2006). Tricine-SDS-PAGE is for the optimal separation of protein <30 kDa (Schägger, 2006). Thus, tricine-SDS-PAGE was used for protease since the

monomeric size of protease is 11 kDa. A standard molecular weight marker from Fermentas™ was used to assess the size and purity of the protein.

Tricine-SDS-PAGE buffers and solutions were prepared as shown in Table 2 (Schagger, 2006). The protein samples collected during protein purification were diluted with an equal volume of sample buffer (10% β-mercaptoethanol, 10% (w/v) glycerol, 10% (w/v) SDS and bromophenol blue). The protein samples were heated on a dry heating block at 95 °C for 5 minutes and centrifuged at 13 000 x g for 30 seconds. Using Bio-Rad™mini-PROTEAN tetra cell electrophoresis system, the cathode and anode buffer were poured into the chamber. The samples were then subjected to electrophoresis at 30 V as the initial voltage and after the samples have reached the separating gel, the voltage was increased to 140 V. The gels were discontinued from the running buffer once the gel front was ~1 cm from the bottom of the gel. Following electrophoresis, the gels were incubated in fixing solution for 1-3 hours, then the staining solution was added for overnight destaining. The staining helped with detecting small amounts of proteins in polyacrylamide gels through visualisation using Bio-Rad™ Gel Doc system (Lab-Image: Software).

2.7 Determination of protein concentration

The concentration of wild-type HIV-1 C-SA PR was determined spectrophotometrically at 280 nm, using the Beer-Lambert Law:

$$A_{280} = \varepsilon.c.l \qquad \text{Equation 1}$$

where 'A_{280}' is the absorbance value of the sample at wavelength (λ) 280 nm, 'ε' is molar extinction coefficient (M^{-1} cm^{-1}) at $\lambda_{280\ nm}$, 'c' is molar concentration (M) value to be determined and '*l*' is the path-length (cm) of the light passing through the solution. The extinction coefficient was determined using ExPASy ProtParam tool (http://web.expasy.org/protparam/) (Gasteiger *et al*., 2005). ProtParam tool uses the amino acid sequence of the protein as an input to compute and deduce its physico-chemical properties (Gasteiger *et al*., 2005). It allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. The extinction coefficient (ε) for the wild-type HIV-1 subtype C protease was determined to be 25 480 M^{-1} cm^{-1} at 280 nm.

Table 2: Preparation of 20% tricine-SDS-PAGE

	4% sample gel (stacker)	16% sample gel (separator)
AB-3 (mℓ)	0.5	5
Gel buffer (3X) (mℓ)	1.5	5
Glycerol (mℓ)		1.5
Add water to final V (mℓ)	6 (4)	15 (3.5)
Polymerise by adding		
10 % APS (ammoniumperoxodisulfate) (μℓ)	45	75
TEMED (tetramethylethylenediamine) (μℓ)	9	15

AB-3 Stock solution (48 g acrylamide, 1.5 g bis-acrylamide, 100 mℓ dH_2O)

Electrode and gel buffers

Anode Buffer (10X) 1 ℓ [121.4 g Tris, 0.3 HCl, pH 8.9]

Cathode Buffer (10X) 400 mℓ [48 g Tris, 1 M Tricine, 1% SDS, pH 8.25]

Gel buffer (3X) 200 mℓ [72.84 g Tris, 1 M HCl, 0.3% SDS, pH 8.5]

Fixing solution (stain) [50% methanol (w/v), 10 % acetic acid (w/v) and 0.25% Coomassie (w/v)]

Staining solution [50% methanol (w/v) and 10% acetic acid (w/v)]

Destaining solution [10% acetic acid (w/v) and 7% ethanol (w/v)]

The protein concentration was determined through an absorbance spectrum of the protein sample in buffer D (10 mM sodium acetate, pH 5). A spectrum measurement was done using a 10 mm path-length quartz cuvette with the use of Jasco V-630 spectrophotometer. An absorbance spectrum of A_{280}-A_{340} was constructed and the absorbance value of 280 nm was subtracted as a blank reading. Light scattering due to aggregation was corrected for by subtracting the absorbance reading at 340 nm.

2.8 Protease active site concentration using an isothermal titration calorimetry

In order to account for the autocatalysis of HIV-1 C-SA PR, an active-site titration was performed using isothermal titration calorimetry (ITC). ITC is a physical technique that directly measures the heat generated or absorbed (ΔH) in an aqueous solution during a biomolecular binding event. ITC consists of a microcalorimeter comprising of the two cells (i.e. reference and sample cells) which are enclosed in an adiabatic jacket (Holdgate and Ward, 2005; Núñez *et al*., 2012) (Figure 11). The reference cell is used as a control and it usually contains the solvent alone which is usually dH_2O or buffer. The sample cell is loaded with the protein that is titrated with the ligand in a series of injections. Both cells are maintained at a constant and identical temperature of 20 ^{o}C throughout the experiment (Wiseman *et al*., 1989). When binding occurs, the heat sensors in the microcalorimeter detect the temperature differences between the two cells and give feedback from the power supplied to sample cell to give an output of raw data. The ligand is usually loaded into a syringe and a series of precisely known aliquots of ligand are injected at regular intervals into the sample cell containing the protein of interest.

When binding occurs between the ligand and the protein, the heat changes in degree Celsius and detected by the sensitive thermopile/thermocouple circuits and measured. As the first injection is made, the microcalorimeter absorbs/releases heat until the binding reaction has reached equilibrium. In an endothermic reaction, the feedback circuit increases the power in order to maintain a constant temperature (isothermal/isothermic operation). Whereas, in an exothermic reaction, the opposite occurs; as the temperature increases through titration in the sample cell, the feedback power decreases in order to maintain the same temperature between the two cells.

The experimental raw data consists of a series of peaks of heat flow (power), with every peak corresponding to one ligand injection. The series of ligand injections gradually increases the molar ratio between the protein and ligand. The protein gets more saturated with every injection

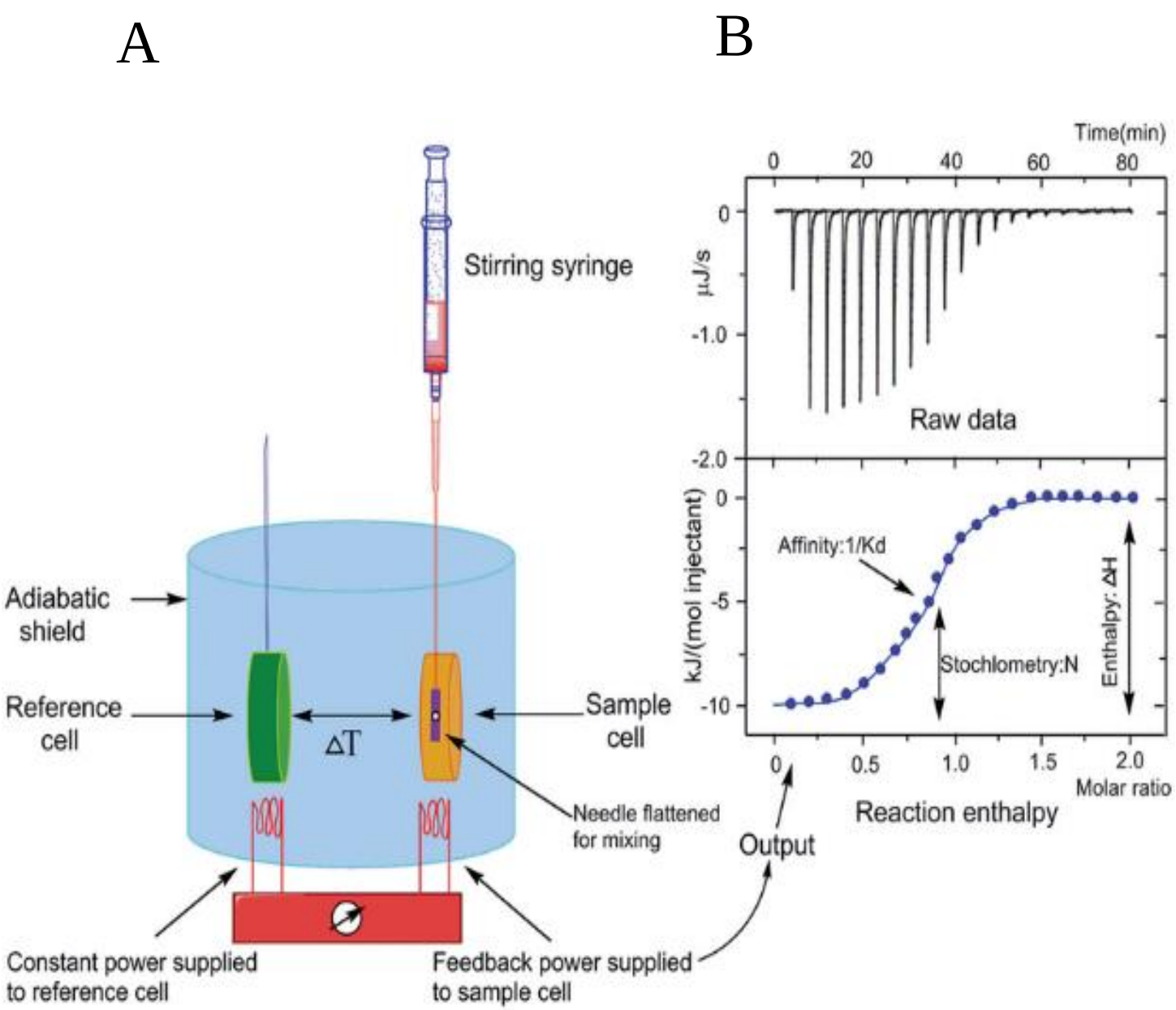


Figure 11: Schematic representation of a typical micro-calorimeter with the variables determined directly from isothermal titration calorimetry. A. Represents a system comprising of two cells isolated by an adiabatic jacket. The syringe loaded with ligand is injected into the sample cell with the protein and rotates to provide continuous mixing of the reactants. B. The image illustrates the parameters determined directly from ITC. The upper right image represents the titration thermogram as heat per unit time released after every injection of the ligand into the protein. The lower right image represents the molar ratio versus the dependence of absorbed/released with each injection. The syringe is inserted in the sample cell and titration with a series of injections is performed (Freyer and Lewis, 2008; Núñez *et al*., 2012).

until there is an excess of ligand versus protein contained in the sample cell, bringing the reaction towards saturation. These heat flow peaks in the raw data are integrated directly with time, giving the total heat exchange per injection. The area of each peak/pulse is then integrated and plotted versus the molar ratio [ligand]/[protein] to give the thermodynamic parameters that provide insight into the nature of interactions involved in the binding process (Freyer and Lewis, 2008). These thermodynamic parameters that can be directly determined are the binding stoichiometry (n), the affinity constant (K_a), and enthalpy changes (ΔH) of the interaction between molecules in the solution (Figure 11). Gibbs free energy (ΔG) and the change in entropy (ΔS) can be calculated using equation 2:

$$\Delta G = \Delta H - T\Delta S = -RTlnK_{eq} \qquad \text{Equation 2}$$

where ΔH is the enthalpy change of the reaction, T is the temperature (kelvin), ΔS is the entropy change of the reaction, R is the universal gas constant (8.314 J/mol/K), and K_{eq} is the equilibrium binding constant.

In the current study, acetyl-pepstatin (Ac-Val-Val-Sta-Ala-Sta) was used as a ligand to be titrated with the wild-type HIV-1 C-SA protease (Luque *et al*., 1998; Velazquez-Campoy *et al*., 2000). A volume of 200 μM acetyl-pepstatin loaded into the syringe was titrated into 2 mℓ of 19.8 μM protease loaded into the sample cell. The default injection volume was 5.72 μℓ and titrated with 35 injection cycles. The buffer (10 mM sodium acetate, pH 5.0), acetyl-pepstatin and the protease were degassed in order to obtain good measurements as the presence of gas bubbles within the sample cell will lead to an unrepresentative data plot. The molar ratio at the mid-point of the binding isotherm gives the reaction stoichiometry of a 1:1 binding reaction. The calculated stoichiometry obtained from the titration data determined the percentage of the active site for the wild-type HIV-1 C-SA PR. If the stoichiometry is 1, it then corresponds that all the wild-type C-SA PR in sample cell binds to the ligand in the active site. If n = 0.8, this corresponds to 80% of the wild-type C-SA PR (titrated with acetyl-pepstatin) with functional active sites. All titrations were carried out using a standard volume in a Nano ITC (TA Instruments) and the data were fitted using NanoAnalyze software (TA Instruments).

2.9 Structural characterisation of wild-type HIV-1 C-SA PR

2.9.1 Secondary structural characterisation using far-UV circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy was used to assess the secondary structural characteristics of the wild-type C-SA PR. CD spectroscopy is observed when optically active

molecules (chromophores) absorb left- and right-handed circularly polarised light differently and result in the formation of an ellipse. The chromophores in proteins are the peptide backbone, aromatic side chains and disulfide bonds (Woody, 1995; Greenfield, 1996). For the CD signal to be generated, the chromophore must either be in an optically asymmetric environment or be intrinsically chiral (Pain, 2004). The secondary structure of proteins can be evaluated in the far-UV region (250 – 190 nm) because of the presence of asymmetric peptide bonds found in the peptide backbone of proteins. The peptide backbone such as β-sheets and α-helices can absorb circularly polarised light in the far-UV region and produce characteristic spectra. The α-helical content display spectra characteristic absorption of a trough at 222 nm and 208 nm, and a large positive peak at 190 nm. Whereas the β-sheets display troughs at 216 and 218 nm and a positive peak at 195 nm. The random coils have a very low ellipticity above 210 nm and a trough at 195 nm (Adler *et al*., 1973; Manning *et al*., 1988; Woody, 1995; Greenfield, 1996). The near-UV CD region (300 – 250 nm) reflects the environments of the aromatic amino acid side chains (Sharon and Nicholas, 2000; Corrêa and Ramos, 2009).

Far-UV CD measurements were conducted with 5 μM wild-type HIV-1 C-SA protease in buffer (10 mM sodium acetate, pH 5.0) and done in triplicate. The experimental parameters for the spectrum measurement were 0.5 nm bandwidth and data pitch of 0.2 nm, with the data interval set to 0.5 nm with a 1-second response. The scan speed was 200 $nm.min^{-1}$ at a constant temperature of 20 ^{o}C and the system was constantly flushed with nitrogen to remove oxygen. A 2 mm path-length cuvette was used and the ellipticity spectrum was taken from 180-250 nm.

All data was collected using a Jasco J-1500 CD spectrometer with data fitting done using SigmaPlot® v.13.0 software. The usage Jasco helped with the absorbance of the protein, it illustrates that if the absorbance is too high, not enough light will reach the detector and hence a meaningful spectrum will not be recorded (i.e. saturation). Likewise, if the absorbance is too low, then the differential absorbance of left and right circularly polarized light will be below the detection threshold. Fortunately, Jasco makes it easy to gauge saturation; thus, it is possible to observe high tension (HT) voltage. HT voltage is roughly proportional to absorbance. If the HT voltage goes above ~600 then the detector is saturated. The amplitude of the spectrum will oscillate wildly. An HT higher than 700 volts was avoided because it impacts on the signal-to-noise ratio. HT voltage is the voltage applied to the photomultiplier tube (PMT) to amplify the CD detector's sensitivity. The CD instrument software can convert this HT window to the UV-VIS spectrum.

The raw CD spectra data was buffer-corrected using Microsoft® Office Excel 2019 and converted to mean residue ellipticity (MRE, deg.cm^2.dmol^{-1}) using equation 3:

$$\text{MRE} = \frac{(100)(\theta)}{\text{cn}l} \qquad \text{Equation 3}$$

where θ is the ellipticity raw data (mdeg), c is the protein concentration (μM), n is the number of residues in the protein (wild-type HIV-1 C-SA PR has a total of 198 residues) and l is the path-length (mm) (Woody, 1995; Sharpe, 2012).

2.9.2 Tertiary structural characterisation using intrinsic fluorescence spectroscopy

Fluorescence spectroscopy is a method used to analyse spectral properties of fluorophores in a compound. Fluorophores are divided into two types – intrinsic and extrinsic. Intrinsic fluorescence (aromatic amino acids, chlorophyll, derivatives of pyridoxyl, flavins and NADH) uses the native compound to exhibit its properties. Extrinsic fluorophores (fluorescein, dansyl, and various other substances) are added to non-fluorescent compounds to change its spectral properties and provide fluorescence (Katritzky and Narindoshvili, 2009). Fluorophores enable a molecule to absorb radiation energy at a specific wavelength and emit the energy again at a longer but specific wavelength. The energy lost between the absorbed and emitted light is known as "Stokes shift". The loss of energy is due to the interaction between the fluorophore and its immediate environment (Lakowicz, 1983; Royer, 1995; Shirley, 1995; Greenfield, 2006).

The protease molecule contains phenylalanine (Phe), tryptophan (Trp) and tyrosine (Tyr) residues. These residues are optically active such that they are easily excited to higher energy states; therefore, used as fluorescent probes. These probes can absorb light at 280 nm. Phe role is neglected in the current study due to its low absorptivity and very low quantum yield, Tyr and Trp are accepted because they have a similar quantum yield. However, the indole ring of Trp can be selectively excited at 280 nm and 295 nm. Trp can lose energy when solvent exposed; thus, its fluorescence is highly sensitive and strongly influenced by its local environment in terms of intensity and wavelength. The greater the exposure of Trp to the polar aqueous environment, the longer its wavelength of maximum emission will be (Royer, 1995).

In the current study, intrinsic fluorescence spectroscopy was used to assess the tertiary structural characteristics of wild-type C-SA protease. The monomeric wild-type HIV-1 C-SA PR has two Trp residues (W6 and W42) and one Tyr residue (Y59). The two Trp residues are completely exposed to the solvent while the Tyr residue protrudes through the flap of the

protease. The Trp and Tyr residues were used to study the local tertiary structure of the proteins using fluorescence spectroscopy.

The fluorescence emission spectra were recorded for 5 μM wild-type HIV-1 C-SA PR samples in a buffer (10 mM sodium acetate, pH 5.0) and done in triplicate. The experimental parameters for spectrum measurement were set as follows: the excitation bandwidth was 2.5 nm, emission bandwidth was 5 nm, data pitch set at 0.5 nm with a 1-second response, and the scan speed was 200 $nm.min^{-1}$. A micro-fluorescence cuvette with a 10 mm path-length was used. The fluorescence emission spectra were generated by selectively exciting Trp residues at 295 nm, and Tyr and Trp residues at 280 nm. The spectrum wavelengths were recorded from 300 – 500 nm (average of three accumulation) at a constant temperature of 20 ^{o}C. The fluorescence measurements were obtained using a Jasco FP-6300 spectrofluorometer and buffer-corrected using Microsoft® Office Excel 2019 with data fitting done using SigmaPlot® v.13.0 software.

2.9.3 Quaternary structural characterisation using size exclusion – high-performance liquid chromatography

Size exclusion – high-performance liquid chromatography (SE-HPLC) was carried out to analyse the quaternary structure of the purified wild-type C-SA protease. SE-HPLC is a method in which components of a mixture are separated based on their molecular weight, diffusion coefficient, hydrodynamic volume and surface properties (Tayyad *et al*., 1991). The Shimadzu UFLC SPD-20A SE-HPLC system and TSK gel SuperSW2000, I.D. 30 cm × 4.6 mm, 4 μm particle size (Tosoh Bioscience, LLC) column were used for the analysis of the protein.

Protein retardation in the column is dependent on the pore/channel size of the stationary phase of the column (Tayyad *et al*., 1991). The column used has a base material of silica with packed porous particles used to separate samples. Larger biomolecules were eluted first as they cannot penetrate the pores of the packed material and the proteins with smaller hydrodynamic size are trapped and retained longer within the column beads (Tayyad *et al*., 1991). High pressure was used and in order to withstand the isocratic pressure of 60 bar at 20 ^{o}C, stainless steel columns were used. The analytical columns used were 15-50 cm in length as the resolving power of a chromatographic column increases with column length. The contaminating materials in the SE-HPLC system were preferentially retained by guard column (3.5 cm) packed with materials similar to that present in the analytical column.

The SE-HPLC system was calibrated with buffer D (10 mM sodium acetate, 0.02% sodium azide, pH 5.0) at a constant flow rate of 2 $m\ell.min^{-1}$. The calibration and operations of the

system were set as per manufacturer's instruction. Eluted fractions were monitored at 280 nm, which is the wavelength at which the side chains of the aromatic residues absorb light maximally. The SE-HPLC column was calibrated using molecular weight standards (Blue Dextran 2000 kDa, albumin 66 kDa, carbonic anhydrase from bovine erythrocytes 29 kDa, ribonuclease A 13.7 kDa, cytochrome C 12.4 kDa, aprotinin from bovine lung 6.5 kDa) in a range 5 000 – 75 000 Da. Then, 30 µℓ of wild-type C-SA protease was injected onto the column with a micro-syringe for separation. The differences in retention time of molecular weight standards and protease were used to estimate the molecular size of the wild-type C-SA protease.

2.10 Protease kinetics

Wild-type HIV-1 C-SA protease enzyme kinetics determinations were performed fluorometrically using a Jasco FP-6300 spectrofluorometer, with a micro-fluorescence 10 mm path-length cuvette. The kinetic parameters of the wild-type HIV-1 C-SA protease were assessed through the cleavage of a fluorogenic substrate (Abz-Thr-Ile-pNO_2Phe-Gln-Arg-NH_2) (Synpeptide Co., Ltd. China), which has a sequence preferred by the HIV-1 C-SA PR (Polgar *et al.*, 1994). The substrate mimics the KARVL cleavage site between the capsid (CA) and nucleocapsid (p2/NC) found within the HIV-1 gag-pol protein precursor chain. To mimic the environment within the cell where HIV protease cleaves the gag and gag-pol polyproteins, 10 mM sodium acetate buffer at pH 5 was used for the enzymatic assays. Though the buffer is not found under physiological conditions, its acidic pH environment mimics the surroundings near the plasma membrane where proteolysis occurs during the viral maturation (Henderson *et al*., 1983; Honig *et al*., 1986; Mervis *et al*., 1988). Also, an arginine residue enhances aqueous solubility at pH 5 of the substrate.

The fluorogenic substrate sequence was recognised and hydrolysed between norleucine (Nle) and nitro-phenylalanine [Phe(NO_2)] by the aspartic protease (Carmel and Yaron, 1978). The fluorophore *p*-aminonitrobenzoic acid (Abz) is excited at 337 nm and emits at 425 nm and its emission chromophore is quenched by the proximal Phe(NO_2) (Toth and Marshall, 1990). Thus, fluorescence spectroscopy was used to monitor the time-dependent cleavage of the scissile peptide bond (product formation) and used to obtain catalytic rates of HIV-1 protease variants (Polgar *et al*., 1994). The kinetic parameters were carried out as follows: the measurement was done for 30 – 60 seconds, with 337 nm excitation wavelength, 425 nm emission wavelength, 2.5 nm excitation bandwidth, 5 nm emission bandwidth, 0.5 seconds response, 1-second data pitch and at a constant temperature of 20 ^{o}C. All kinetic parameters

were performed in triplicate using Jasco FP-63 000 spectrofluorometer and data fitting was done using SigmaPlot® v13.0 software (Systat Software Inc. California, USA).

2.10.1 Determination of K_M and V_{max}

For a majority of enzymes, the initial rate of reaction varies hyperbolically with substrate concentration [S] (Briggs & Haldane, 1925). At low [S], first-order kinetics are observed (i.e. the initial rate is proportional to the [S]), but at high [S], saturation (zero-order) kinetics exist and the initial rate is independent of [S]. The mathematical relationship between initial rate and substrate concentration is expressed by the Michaelis-Menten model, which also describes the kinetic parameters of the wild-type protease. In the model, an enzyme (E) is combined with the substrate (S) to form an enzyme-substrate complex (ES), which can proceed to form a product (P) or dissociate into enzyme and substrate:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

Scheme 1

The equilibrium between the E, S and ES complex is almost instantly maintained and set up, and the breakdown of ES complex to the formation of P is too slow to disrupt the equilibrium. This is known as the Michaelis-Menten assumption, therefore:

$$k_1[S][E] = k_{-1}[ES] \qquad \text{Equation 4}$$

Therefore, the constant can then be removed from the equation; thus, allowing the following:

$$[E][S]/[ES] = k_{-1}/k_1 = K_s \qquad \text{Equation 5}$$

Where the dissociation constant is K_s for ES complex and k_{-1} and k_1 are first-order rate constants for the forward and reverse reactions, respectively.

In 1925, Briggs and Haldane modified the Michaelis-Menten equation. They introduced a more generally valid assumption, that of the steady-state, i.e. the enzyme-substrate complex would be broken down as fast as it is being formed. Thus, in the simplest situation, where the P is formed in a single step and at a rate such that the equilibrium concentration of ES is maintained, it can be shown that the observed K_s is equal to K_M. The result is that the Michaelis-Menten equation becomes:

$$v_o = \frac{V_{max}[S]}{K_M + [S]} \qquad \text{Equation 6}$$

From the equation, the Michaelis constant (K_M) corresponds to the concentration of substrate that induces the wild-type HIV-1 C-SA PR to half maximum velocity and V_{max} will be reached when the wild-type HIV-1 C-SA PR is saturated with the fluorogenic substrate. The equation describes the rate of an enzymatic reaction as a function of substrate concentration. It also describes the enzyme activity towards the end of the reaction, so when [S] >> K_M. K_M is equal to the dissociation constant of the enzyme-substrate (ES) complex if the rate constant k_2 is much smaller than k_{-1}. When these conditions are met, K_M is a measure of the strength of the ES complex: a high K_M indicates weak binding; a low K_M indicates strong binding. An enzyme that has a high affinity for substrate, has a low K_M value.

K_M was used to determine wild-type HIV-1 C-SA protease binding affinity to the fluorogenic substrate. K_M was calculated based on the different catalytic activities measured using 2 µM of wild-type HIV-1 C-SA PR. The concentration of the fluorogenic substrate was gradually increased (5 – 200 µM) until saturation occurred and the substrate range was 0.1 K_M - 10 K_M. The Michaelis-Menten plot was represented as a hyperbolic curve, as the enzymatic activity (µmol.min^{-1}) versus the concentration of the substrate (µM), and the curve was fitted with the Michaelis-Menten equation (Briggs & Haldane, 1925; Johnson & Goody, 2011) using SigmaPlot® v.13 software.

2.10.2 Catalytic turnover number (k_{cat})

The maximum rate, V_{max}, reveals the catalytic turnover number (k_{cat}) of an enzyme, which is defined as the number of completed catalytic cycles by an active site per second (s^{-1}). k_{cat} gives a direct measure for the catalytic production of the product under saturated enzyme conditions. The enzyme is saturated with the substrate in order for specific activity and catalytic turnover to be determined, thus, allowing for steady-state assumption to be made. V_{max} reveals the turnover number of an enzyme if the concentration of the active site $[E]_T$ is known. k_{cat} is calculated using the following equation:

$$V_{max} = k_{cat}\,[E]_{Total} \qquad \text{Equation 7}$$

The catalytic constant (k_{cat}) was calculated under the condition where the [S] >> K_M. The final substrate concentration was 50 µM and the protease concentration was in the range from 0 to 50 nM. k_{cat} was determined by plotting the initial velocity (v_o) in µmol.s^{-1} vs the amount of wild-type HIV-1 C -SA PR in µmol. The catalytic turnover number was determined by following product formation as a function of time. In this case, k_{cat} was monitored

fluorometrically through the formation of product and the data fitted using SigmaPlot® v.13 software.

2.10.3 Specific activity

Specific activity for wild-type HIV-1 C-SA PR gives a measurement of the enzyme purity in a mixture. It is defined as the amount of product formed by an enzyme in a given amount over time, under given conditions per milligrams of the total protein and the units are $\mu mol.min^{-1}.mg^{-1}$ (Dixon & Webb, 1958). The specific activity ([S] >> K_M) was calculated using a final substrate concentration of 50 μM (i.e. steady-state conditions). The protein was added incrementally from 0 to 50 nM. The initial velocity ($\mu mol.min^{-1}$) was recorded against the amount (mg) of HIV-1 C-SA PR. The gradient is equivalent to the specific activity in $\mu mol.min^{-1}.mg^{-1}$. With the molecular weight of the enzyme, the catalytic turnover was determined by specific activity (Wharton & Eisenthal, 1981).

2.10.4 Catalytic efficiency (k_{cat} / K_M)

The catalytic efficiency (k_{cat}/K_M) of the enzyme for a particular substrate is equated as a ratio of the turnover number (k_{cat}) to the Michaelis-Menten substrate dissociation constant (K_M). k_{cat}/K_M describes the ability of an enzyme to increase the reaction rate at low substrate concentration. Equation 8 gives insight into how enzymes operate under typical physiological conditions, that is, when [S] << K_M:

$$v_o = (k_{cat}/K_M)[E]_{Total}[S] \qquad \text{Equation 8}$$

Thus, when [S] << K_M, the enzymatic velocity depends on the values of k_{cat}/K_M, [S], and $[E]_T$. Under these conditions, k_{cat}/K_M is the rate constant for the interaction of substrate and enzyme. The rate constant k_{cat}/K_M is a measure of catalytic efficiency because it takes into account both rate of catalysis with a particular substrate (k_{cat}) and the strength of the enzyme-substrate interaction (K_M). Catalytic efficiency was determined independently from k_{cat} and K_M value and taken from a gradient of k_{cat}/K_M linear plot: catalytic turnover (s^{-1}) versus substrate concentration (μM).

The condition in which k_{cat}/K_M is determined; enzyme concentration is kept constant while the substrate concentration was varied at different increments from 0 μM to 5 μM. The catalytic efficiency compared the effectiveness of wild-type HIV-1 C-SA PR to catalyse the different concentrations of fluorogenic substrate. Data were fitted into a linear plot of catalytic turnover versus the concentration of substrate using SigmaPlot® v 13.0 software.

2.11 Solubility studies for novel inhibitors

The novel protease inhibitors (Table 1) were provided in a lyophilised form and stored as a powder. The novel inhibitor stocks had a 10 mM concentration dissolved in 100% dimethyl sulfoxide (DMSO). All reactions have a final DMSO concentration of 0.002 μM (2%). Each inhibitor was resuspended separately in 100% DMSO and different stock concentration was prepared. An initial stock of 100 μM was prepared in 100% DMSO, the second inhibitor stock was made (100 x dilution) to 1 μM in 100% DMSO and the third stock was made (25 x dilution) to 4 μM in 100% DMSO. The inhibitor stocks had a final of 2% DMSO in each reaction volume mixture. The inhibitors were prepared and stored at -80 ^{o}C until further use.

2.12 Inhibition studies – IC_{50} and K_i values

The half maximum inhibition concentration (IC_{50}) is the concentration of inhibitor that is required to inhibit the activity of an enzyme by 50% *in vitro*. IC_{50} values for GRD 110D, saquinavir (SQV), ritonavir (RTV) and darunavir (DRV) inhibitors were determined in 2% (v/v) DMSO. The buffer (10 mM sodium acetate, pH 5.0) was mixed with the inhibitors, and the wild-type HIV-1 C-SA PR added last. The protease (at a constant concentration of 50 nM) was incubated with the inhibitor in varying concentrations, ranging from 0 M to 200 nM for an hour at 20 °C. The fluorogenic substrate was mixed with the second buffer (10 mM sodium acetate, 1 M NaCl, pH 5.00) and another buffer mixed with the fluorogenic substrate was prepared to initiate the reaction. The standard assays were performed with increasing concentrations of inhibitors in the presence of 50 nM protein concentration, with saturating concentrations of the fluorogenic substrate at 50 μM. The signal from the fluorogenic substrate hydrolysis in the presence of various protease inhibitors was monitored on a fluorometer at 20 °C. As the concentration of inhibitor increases, the signal (i.e. the cleavage of the substrate) decreased. The measured parameters were carried out as follows: measurements done in 30 seconds, with 337 nm excitation wavelength, 425 nm emission wavelength, 2.5 nm excitation bandwidth, 5 nm emission bandwidth, 0.5 seconds response, 1-second data pitch and at a constant temperature of 20 °C. The IC_{50} values for all protease inhibitors were determined from the inhibition percentage vs inhibition concentration plot. The obtained IC_{50} values and standard deviations were used to calculate K_i values. K_i is the dissociation constant of the enzyme-inhibitor complex, in the presence of a substrate. K_i was calculated using the following equation:

$$K_i = \frac{IC_{50} - [E]/2}{[S]/K_M + 1}$$

Equation 9

where IC_{50} is the concentration of the inhibitor required to reduce the activity of the enzyme by 50%, [E] is active enzyme concentration (50 nM), [S] is substrate concentration (500 nM) and K_M is Michaelis constant determined experimentally. The inhibition studies for the determination of IC_{50} and K_i were used to analyse the enzymatic characterisation of wild-type HIV-1 C-SA PR to hydrolyse the fluorogenic substrate in the presence of increasing inhibitor concentration. IC_{5O} and K_i values were performed in triplicate using Jasco FP-63 000 spectrofluorometer and data fitting was done using SigmaPlot® v13.0 software.

2.13 Molecular docking studies of HIV protease

The HIV-1 protease crystal structure combined with activity data gives a good idea of structure-function relations. However, the nature of proteins is highly dynamic as the active site geometry is heavily dependent on the conformational changes induced by the ligand. Thus, in order to identify accurate confirmation of ligands to their protein target and to estimate the strength of the protein-ligand interaction, molecular docking was applied. The dynamics of HIV-1 protease variants have been studied in the past using computational models (Perez *et al*., 1998). A continuous quantitative understanding of HIV-1 protease molecular docking provides continuous insights into the development of structure-based molecular design strategies. Structural information of protease complexed with the inhibitor is being used to an extended degree in the discovery of novel pharmaceutical compounds. Also, the structure of an HIV protease complexed to the inhibitor(s) can assist in the understanding of the biochemical basis underlying the protease inhibitor susceptibility (Perez *et al*., 1998; Rosin *et al*., 1997). Docking in the field of molecular modelling is a method which predicts the preferred orientation of macromolecule when bound to the ligand to form a stable complex. Macromolecules are proteins such as a receptor, enzyme or nucleic acid with known conformations (Vijesh *et al*., 2013). Molecular docking is usually performed through commercial software, such as DOCK (Venkatachalam *et al*., 2003), GOLD (Jones *et al*., 1997), AutoDock (Österberg *et al*., 2002), FRED (McGann *et al*., 2003), Surflex (Jain, 2003), ICM (Schapira *et al*., 2003), FlexX (Rarey *et al*., 1996), UCSF Dock (Allen *et al*., 2015), Glide (Friesner *et al*., 2004) and many others. These programs are mainly grouped into four categories: evolution-based (genetic algorithms), fragment-based, Monte Carlo simulation (LigandFit) and the shape complementary method. Although these classes of methods require different information in addition to the structure of proteins, they all share four common computational processes. The first being, simplified and rigid body search, then selecting the

section(s) of interest, followed by the modification of docked structures and finally selecting the best models.

The category used in the current study was the geometric matching/shape complementarity methods which describe the protein and ligand as a set of features that makes them dockable. There are two types of docking in this method: rigid "lock-and-key" and flexible "induced-fit" docking. In rigid docking, the macromolecule (lock) and ligand (key) are kept rigid as their shapes are geometrically complementary, where no structural optimisation of the binding partners is required (Koshland, 1994). While in flexible docking either or both the macromolecule and ligand allow flexibility, and were adjusted in their conformation to achieve an overall "best-fit" (Caret *et al*., 1993). Docking studies were performed on a commercial program known as Glide from Schrödinger. Glide uses induced-fit docking to exhaustively consider possible binding modes and the associated conformational changes within receptor active sites (Niu *et al*., 2008). It performs a complete systematic search of the conformational, orientational and positional space of the docked ligand with the OPLS (Optimised Potentials for Liquid Simulations) force field.

Docking programs have three main uses. First is the hit identification in which docking combined with a scoring function can be used to quickly screen large databases of potential drugs *in silico* to identify molecules that are likely to bind to a protein target of interest. Second for lead optimisation in which docking can be used to predict where and in which relative orientation a ligand binds to a protein. This information may, in turn, be used to design more potent and selective analogues. Third is the prediction of the binding affinities of related compounds from a known active series.

2.13.1 *In silico* studies of HIV-1 protease

In silico docking techniques were used to investigate the complementarity of a ligand and a protein target at the molecular level. As such, docking studies have been used to identify the structural features that are important for binding and for *in silico* screening. Virtual screening has been used to find ligand hits and assist in lead optimisation in structure-based drug discovery. Thus, structural information of protease complexed with the inhibitor could be used to an extended degree in the discovery of novel protease inhibitor compounds. The structure of an HIV protease complexed with an inhibitor(s) can assist with the examination of the detailed picture of the binding site and to design more active analogues (Perez *et al*., 1998). A number of docking software programmes (such as Dock, AutoDock, GOLD, FlexX, GLIDE, ICM,

PhDOCK, Surflex and others) have been utilized successfully to fulfil the purpose of computer-based drug design (Rosin *et al*., 1997). In the current study is Schrödinger software was used. This program is useful in evaluating the binding mode of the inhibitors to target the protease.

Multiple algorithms implemented in the model of Schrödinger known as Maestro 11.2 (Molecular Modelling Interface) were used to predict the protein-ligand complex geometries. Molecular docking was performed using the Induced-fit^{TM} docking algorithm from Maestro 11.2 software. The Induced-fit docking protocol begins by docking the active ligand with Glide (Grid-based ligand docking with energetics). Glide provides possible binding modes by giving a complete systematic approximation which searches for the orientation, conformational, and positional space of the docking ligand.

The induced-fit docking was initiated through the preparations of the protein and ligand. Maestro 11.2 LigPrep tool was used to construct a 2D sketcher of the ligands into 3D models and to correct bond lengths and angles, correct chirality and perform energy minimisation. LigPrep was then used to produce low energy structures in order to obtain the highest-quality docking results. The protein was prepared using the Protein Preparation Wizard and the energy minimisation was performed for both ligands and protein structures. Induce-fit docking protocol was performed with all ten HIV-1 protease inhibitors (ligands) and two HIV-1 protease crystal structures obtained from RCSB Protein Data Bank (PDB) 3U71 (subtype C protease) and 2P3B (subtype B protease).

After the initial screening was done to select ligand poses, the ligand was refined in torsional spaces in the field of the receptor using optimised potentials for liquid simulations (OPLS) 2005 (Glide HTVS) with a distance-dependent dielectric model. Thereafter, Prime structure prediction was used for each pose in order to accommodate the ligand through the reorientation of side chains. Post-docking minimisation (PDM) was used for a small number of poses which were minimised within the field of the receptor with full ligand flexibility.

Docking methodologies are known as Glide HTVS (High-throughput virtual screening) and SP (Standard-Prediction) which use a series of hierarchical filters to look for possible locations of the binding-site region of the receptor. Induced-fit docking used reduced van der Waals radii and increased Coulomb-van der Waals (vdW) cut-off in order to make a diverse ensemble of ligand poses (Sherman *et al*., 2006). Pose means a complete specification of the ligand: orientation and position relative to the receptor, rotamer-group conformation, and core conformation. For each pose, a Prime structure prediction was then used to accommodate the

ligand by reorienting nearby side chains, and these residues and the ligand were then minimised. Finally, each ligand was docked into its corresponding low energy protein structures and the resulting complexes are ranked according to a scoring function combining GlideScore and Prime energies.

Glide also uses E-model (kcal/mol) to choose the best pose of a ligand and ranks them against one another with GlideScore. GlideScore is an empirical scoring function designed to minimise separation of compounds with strong binding affinity from those with little to no binding ability. The E-model energy (kcal/mol) has a substantial weighting of force field components such as van der Waals and electrostatic energies. The E-model scoring function is primarily defined by the protein-ligand coulomb van der Waals energy with a small contribution from GlideScore. The ranking of the resulting complexes was based on the scoring function Glide E-model (Sherman *et al*., 2006). The results from docking screening hits are subjected to pharmacological validation (e.g. IC_{50}, affinity and potency measurements).

CHAPTER 3

3. RESULTS AND DISCUSSION

3.1 Protease cDNA insert sequence verification

Bacterial expression system was used because of its rapid growth rate, and the fact that it is relatively inexpensive. Transformation was accomplished using a heat shock method as it does not require any equipment other than the dry heating block. The plasmid DNA sequence encoding wild-type HIV-1 C-SA protease was verified by Inqaba Biotechnical Industries (Ltd) (Pretoria, South Africa) using the Universal primers (pGEX3 – T7 promoters). Sanger sequencing was performed on the pET-11b plasmid (Inqaba Biotechnical Industries). The sequencing results confirmed that the gene insert was correct. Thus, the South African HIV-1 subtype C protease consensus sequence shares the same sequence identity as the subtype C protease in group M (Los Alamos HIV sequence database, http://www.huv.lanl.gov/). The Sanger sequencing results were viewed in chromatogram using FinchTV version 1.4.0 (http://www.geospiza.com/FinchTV:GeospizaInc). The DNA sequence was translated to an amino acid sequence using ExPASy translate tool (http://www.web.expasy.org/translate/). The translated sequence was aligned with a known protein sequence (PDB ID: 3U71) using Clustal Omega multiple sequence alignment (https://www.ebi.ac.uk/Tools/msa/clustalo/) to confirm that the sequence obtained corresponded with the protein sequence of wild-type HIV-1 C-SA protease in the protein database (Figure 12).

3.2 Cell growth of *E. coli* BL21 (DE3) pLysS with pET-11b

The growth curve was done to determine the optimal density at which protein expression could be induced with IPTG. Figure 13 reported that high cell concentrations may reduce production of recombinant protein, when nutrient limitations occur. The cell growth of *E. coli* BL21 (DE3) pLysS with pET-11b is illustrated in Figure 13 as a growth curve with three distinct phases (i.e. lag phase, exponential phase and stationary phase). The lag phase (also known as starting phase) is the period where most of the cell culture is in the media. At the exponential phase, cells are growing at an exponential rate and finally, at the stationary phase, there is a decline in cell growth in the media due to resource depletion. However, this phase lasts longer than the other phases.

> Wild-type_HIV-1_C-SA_protease_amino_acid_SEQ

PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEEINLPGKWKPKMIGGIGGFI
KVRQYDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF

Clustal Omega multiple alignments

```
CLUSTAL O(1.2.2) multiple sequence alignment

Seq_WT          PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEEINLPGKWKPKMIGGIGGFIKVRQYD
Known_Seq       PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEEINLPGKWKPKMIGGIGGFIKVRQYD
                ************************************************************

Seq_WT          QILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF
Known_Seq       QILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF
                ***************************************

   #
   #
   #  Percent Identity  Matrix - created by Clustal2.1
   #
   #

        1: Seq_WT       100.00  100.00
        2: Known_Seq    100.00  100.00
```

HIV-1 PROTEASE AMINO ACID SEQUENCES

```
                          7       12  15    19                    35 36 37     41
Seq_WT          PQITLWKRPL VSIKVGGQIK EALLDTGADD TVLEEINLPG KWKPKMIGGI
Known_C-SA      PQITLWKRPL VSIKVGGQIK EALLDTGADD TVLEEINLPG KWKPKMIGGI
B               PQITLWKRPL VTIKIGGQLK EALLDTGADD TVLEEMNLPG RWKPKMIGGI
C               PQITLWKRPL VTIKIGGQLK EALLDTGADD TVLEEINLPG KWKPKMIGGI
A               PQITLWKRPL VTVKIGGQLK EALLDTGADD TVLEDINLPG KWKPKMIGGI

                        57              69                        89    93
Seq_WT          GGFIKVRQYD QILIEICGKK AIGTVLVGPT PVNIIGRNML TQLGCTLNF
Known_C-SA      GGFIKVRQYD QILIEICGKK AIGTVLVGPT PVNIIGRNML TQLGCTLNF
B               GGFIKVRQYD QILIEICGHK AIGTVLVGPT PVNIIGRNLL TQIGCTLNF
C               GGFIKVRQYD QILIEICGKK AIGTVLVGPT PVNIIGRNML TQIGCTLNF
A               GGFIKVKQYD QILIEICGKK AIGTVLVGPT PVNIIGRNML TQIGCTLNF

                ************************************************************
```

Figure 12: Protein multiple sequence alignment. The FASTA format represents an amino acid sequence of wild-type HIV-1 C-SA protease translated with ExPASy translate tool, with the reading frame beginning with PQITL… wild-type HIV-1 C-SA protease (Seq_WT) was aligned with the known sequence (known_Seq_3U71) using Clustal Omega multiple sequence alignment v. 1.2.2. The two sequences were further compared with various subtypes of HIV-1 protease.

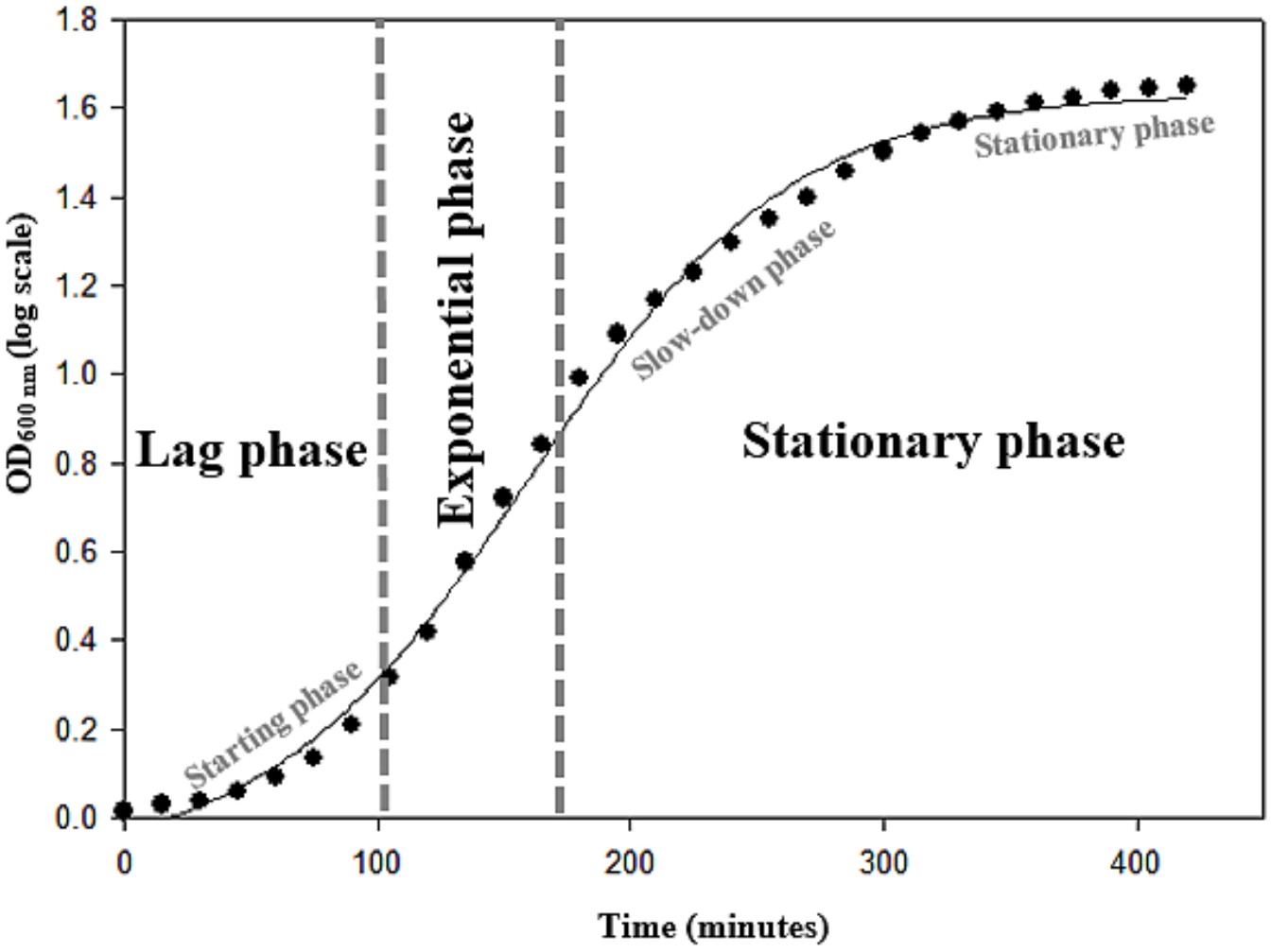


Figure 13: Cell growth of *E. coli* BL21 (DE3) pLysS with pET-11b. Time on the x-axis and number of bacterial growths on the y-axis. During the lag phase, there is almost no growth and starting phase is where growth is increasing at a slow rate. The exponential phase is where the vast majority of cells are alive, healthy and are ideal for protein expression. The stationary phase is where growth begins with the slow-down phase where saturation begins and then slowly cell growth declines due to growth-limiting factor (e.g. depletion of nutrient). Data fitting was performed using SigmaPlot® v13.0 software (Systat software).

During protein expression, a usage of high OD_{600} (>0.8) is not favourable, since the number of dead cells outweighs alive cells; therefore, protein yields are greatly impacted. At low OD_{600} (<0.4) cells density is too low for optimal expression of protein; thus, induction of protein expression was ideal at the exponential phase for protein overexpression. The OD_{600} values represent the amount of light that is transmitted by the sample in the SM1 100 UV-VIS spectrophotometer, which, directly corresponds to cell growth/division. The OD_{600} values recorded were plotted against time points of collection of samples to form a growth curve which usually forms a sigmoidal shape (Figure 13). An increase in the solubility of a recombinant protein is observed when the expression is induced in early-exponential phase. Thus, an OD_{600} value between 0.4 - 0.6 was chosen for induction. Under this condition, the majority of cells exist in a viable state and are ideal for protein expression (Figure 13).

3.3 Expression trials of wild-type HIV-1 C-SA protease

Wild-type HIV-1 C-SA protease was overexpressed and recovered from *E. coli* BL21 (DE3) pLysS cells. Using expression trials, optimal condition for expression of wild-type HIV-1 C-SA protease in *E.coli* BL21 (DE) pLysS were identified. Induction with different IPTG concentrations was carried out at OD_{600} 0.5 (early mid-log phase of the growth curve) and the cells were incubated at 20 °C and 37 °C (Figure 14). The growth temperature was varied because different temperature can affect the expression and the solubility of heterologous (recombinant) proteins expressed in *E. coli* BL21 (DE3) pLysS cells. The highest level of total protein expression was observed at 37 °C after adding IPTG at the middle-exponential growth phase OD_{600} 0.5. The temperature of 37 ^{o}C gave a robust and fast growth compared to 20 ^{o}C. The temperature did not improve the solubility of the protein: as in all cases, the wild-type HIV-1 C-SA PR was expressed as inclusion bodies in *E. coli* BL21 (DE3) pLysS cells (Ido *et al*., 1991; Todd *et al*., 1998) (Figure 14). The highest overexpression was observed after a four-hour period with 1.0 mM IPTG final concentration.

3.4 Overexpression and purification of wild-type HIV-1 C-SA protease

Taking into consideration the results of the growth curve and expression trials, the overexpression of the wild-type HIV-1 C-SA protease in the *E. coli* BL21 (DE3) pLysS cells was induced at OD_{600} of 0.5 with 1 mM IPTG and the cells were incubated for four hours post induction with IPTG. Wild-type C-SA PR was expressed as inclusion bodies in *E. coli* BL21 (DE3) and purified using CM Sepharose cation exchange column chromatography regenerated with 1 M sodium chloride (Figure 15).

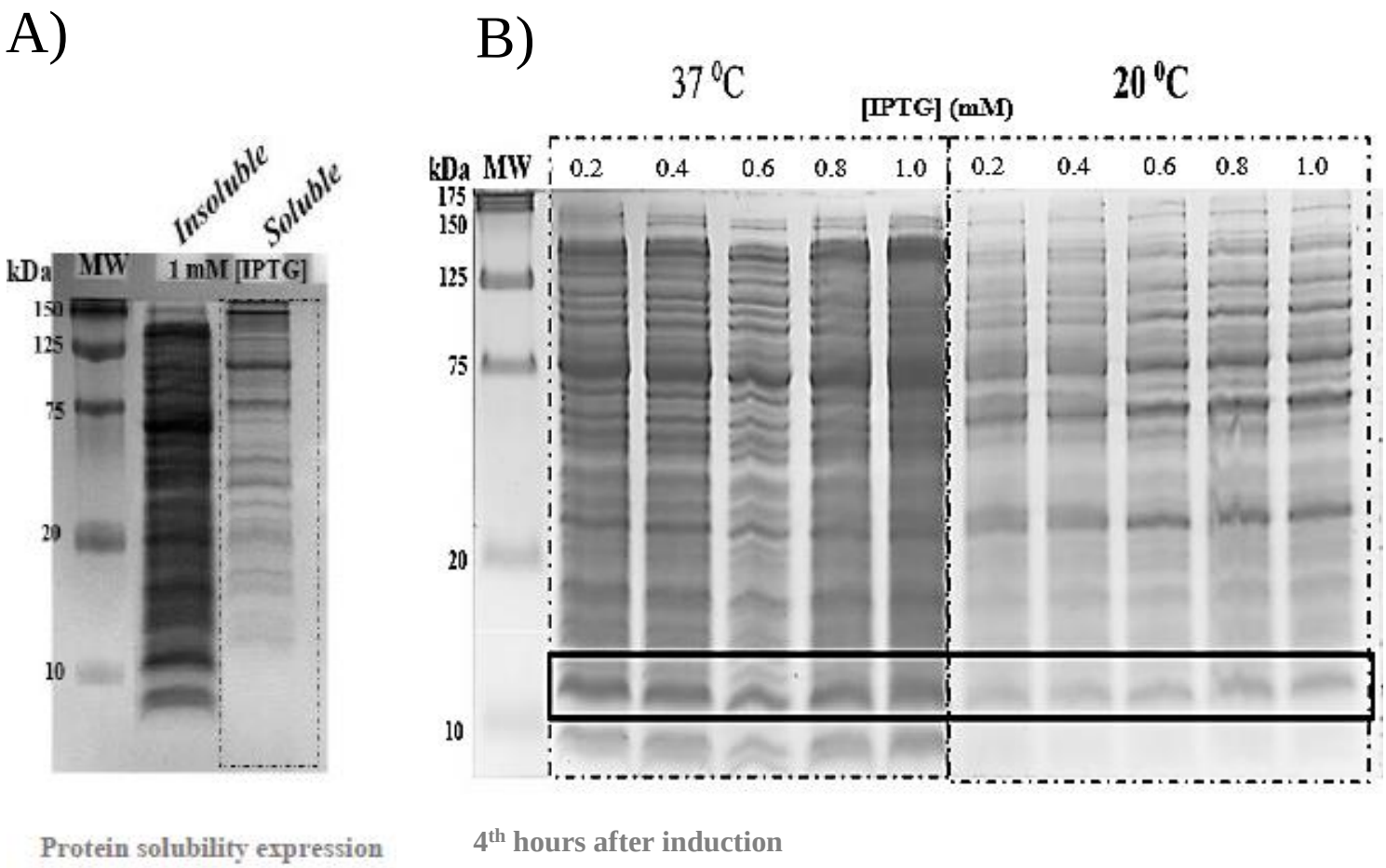


Figure 14: Expression trial of wild-type HIV-1 C-SA protease in *E. coli* BL21 (DE3) pLysS cells at 20 ºC and 37 ºC using varying IPTG concentrations. A. Illustrates that the wild-type HIV-1 C-SA protease is expressed in the insoluble fraction. B. Illustrates a representative of the best conditions for overexpression of the protein (supplementary data on Appendix A). Induction trials were carried out through varying IPTG concentrations (0.2, 0.4, 0.6, 0.8 and 1.0 mM) and temperature, and induced at $O.D_{600}$ 0.5 for 1 – 6 hours and overnight. The high overexpression period chosen for the expression trials was 1.0 mM IPTG concentration for a period of four hours as the yield of the protein was maximised.

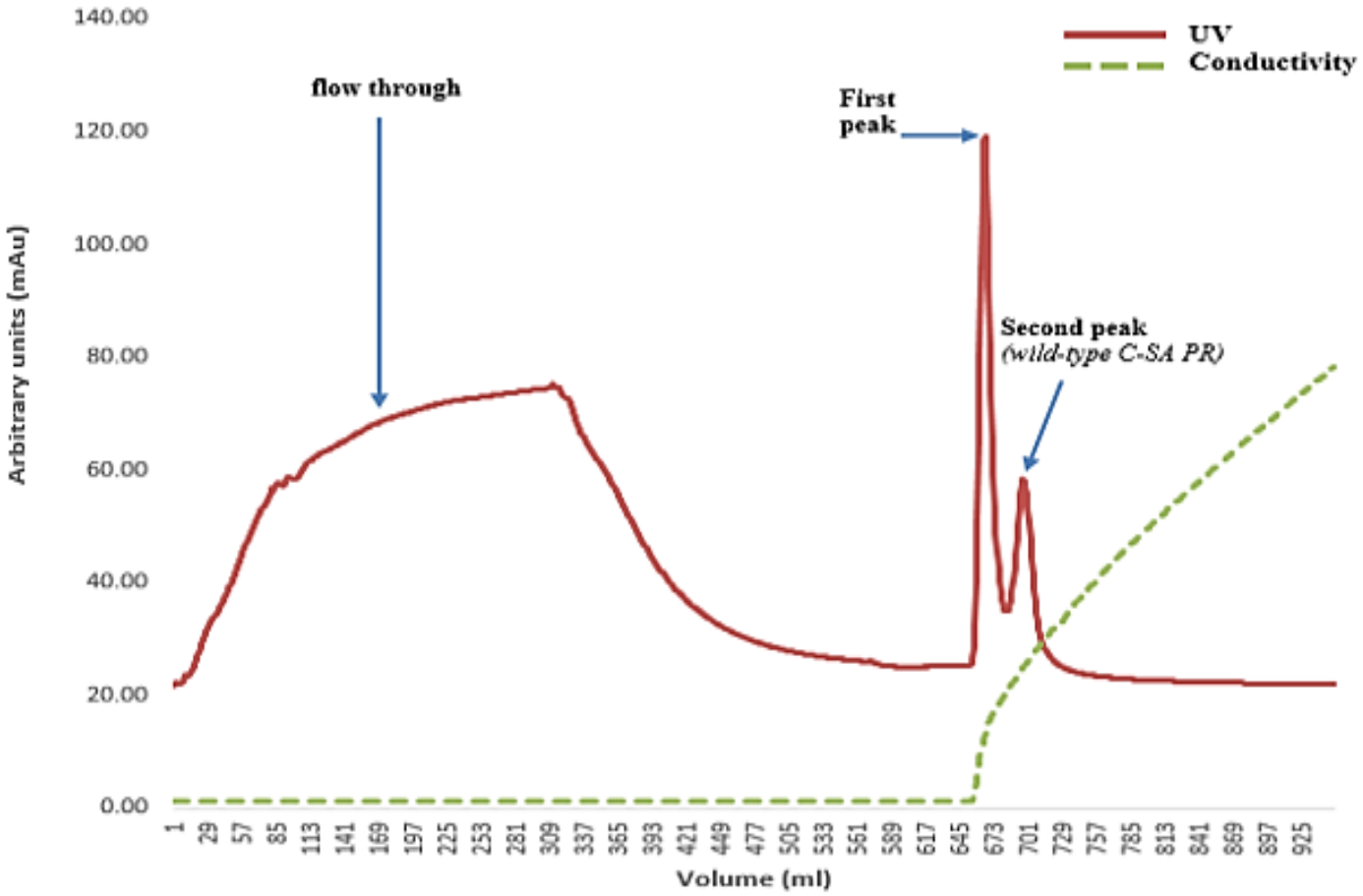


Figure 15: CM-Sepharose cation exchange elution profile of wild-type HIV-1 C-SA protease and SDS-PAGE. Elution profile of HIV-1 C-SA wild-type protease. Unbound proteins and other cellular components eluted in the first peak. The first peak represents all other proteins except for the desired protease that was eluted with 0-1 M salt gradient. The second peak shows the elution of the protease.

The protease is represented in the chromatogram as the second peak. However, it is possible that other proteins with a similar charge to the protease were eluted with the second peak and those would have been seen during resolution with a 20% tricine SDS-PAGE gel. The single band observed on the 20% tricine SDS-PAGE gel corresponds to the monomeric size (~11 kDa) of the protease, suggesting that the protein was pure, since no other bands were observed (Figure 16). These results are in agreement with previous finding of HIV-1 protease (Ido *et al*., 1991; Mosebi *et al*., 2008).

3.5 Protease concentration determination

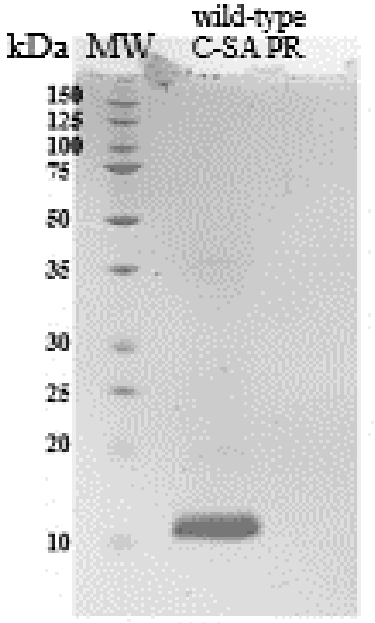
Wild-type HIV-1 C-SA protease concentration was determined using the Beer-Lambert Law following measurements of absorbance of light in the ultra-violet region of the spectrum. The absorption spectrum was used to assess the optimum absorbance intensity at 280 nm (Figure 17) which was due to the chromophores (i.e. aromatic amino acid residues) in the HIV-1 protease. Using the protein absorption spectrum (Figure 17), the protease concentration was calculated as 19.8 μM and aggregation was accounted for by monitoring the absorbance at 340 nm. The protease concentration determination was performed in triplicate and the calculations were directly from this single spectrum. A dilution series was also performed for improved accuracy for protease concentration. Based on the fact that the protease auto-catalytic (i.e. catalysis of a reaction by one of its products), the actual protease concentration was determined through utilising ITC to get the active site concentration.

3.6 Protease active site titration calorimetry

Usually, protein concentration would be determined using spectroscopic methods but because of the protease auto-catalysis (meaning it cleaves itself) characteristic, any peptides would contribute towards the total concentration determination regardless of activity. Thus, any concentration calculated would not be the representation of the true active concentration of this protein. So, in order to determine how much of the protease is active, isothermal titration calorimetry (ITC) was used to titrate the wild-type HIV-1 C-SA PR with acetyl pepstatin (ligand) which is an aspartic PR inhibitor.

Isothermal titration calorimetry (ITC) was used to determine the concentration of active wild-type HIV-1 C-SA protease by measuring its binding to acetyl-pepstatin. HIV-1 protease was found to be inhibited *in vitro* by acetyl-pepstatin, a general inhibitor of aspartic proteinases (Luque *et al*., 1998; Velazquez-Campoy *et al*., 2000). The titration on the top panel (Figure 18) is represented by the peak which occurs when acetyl-pepstatin binds to the wild type protease.

A



B

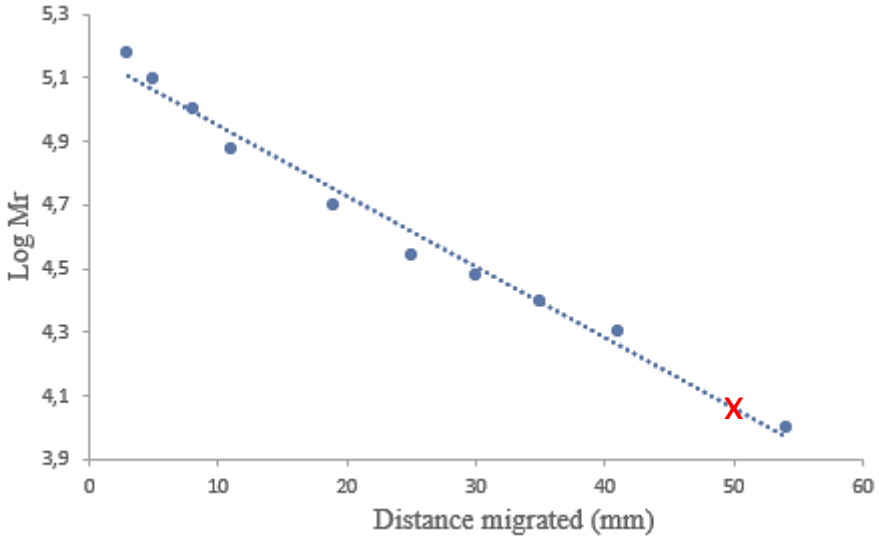


Figure 16: HIV-1 C-SA protease purity and molecular weight determination. A) A 20% Tricine (4% stacking and 16% separating gel) SDS-PAGE gel. Lane 1: molecular weight marker; lane 2: purified protease. The protease had an expected monomeric size of approximately 11 kDa. B) Calibration curve for determination of the molecular weight of the monomeric HIV-1 C-SA proteases. The protein standard is a mixture of 10 unstained recombinant proteins of a molecular weight range from 10 to 200 kDa (separation on an SDS Tris-glycine gel). Wild-type HIV-1 C-SA protease had an approximate molecular weight of 11.4 kDa. Position of the protease is marked (**x**). The correlation coefficient is 0.98.

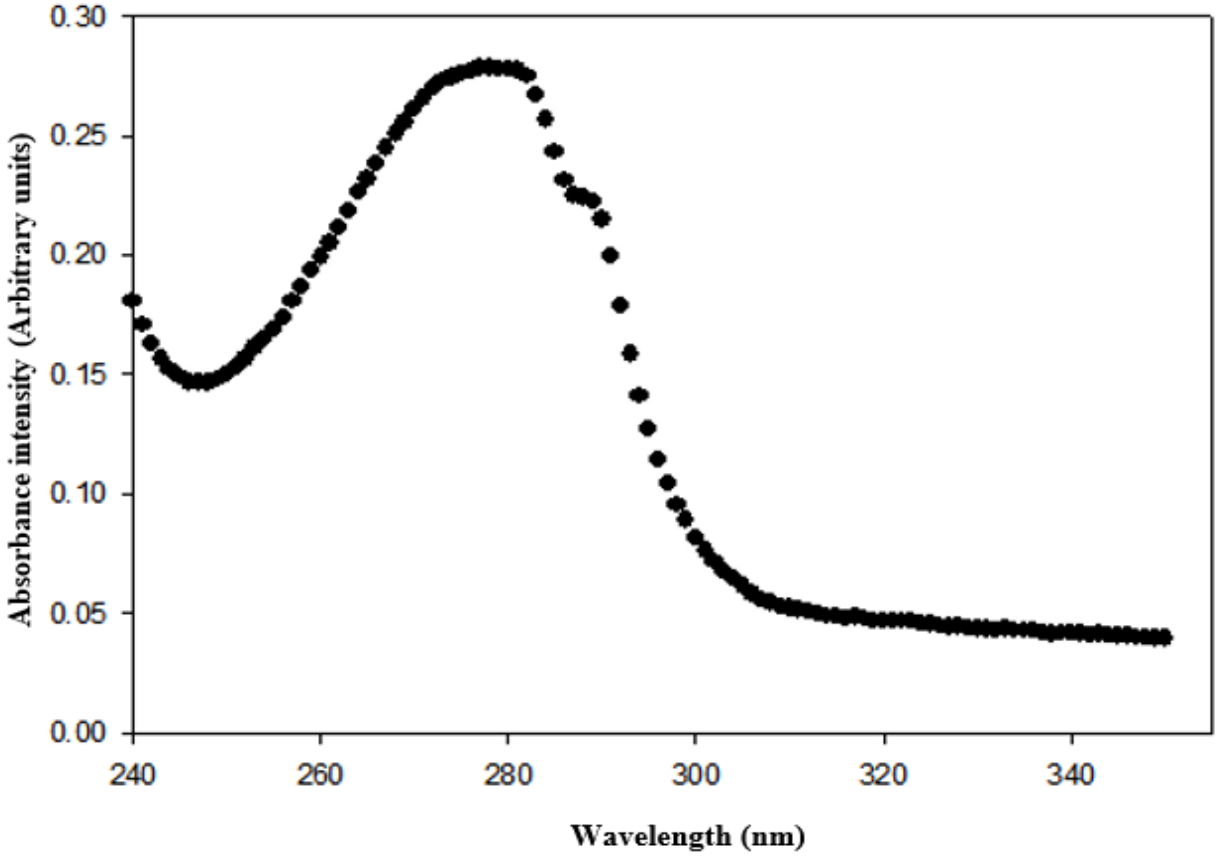


Figure 17: HIV-1 C-SA wild-type protease absorbance spectrum. Absorption spectrum reading was from a wavelength of 240 – 340 nm. The 19.8 μM protein concentration was calculated using the Beer-Lambert law using a dilution series. The experiment was performed in triplicate at 20 ºC in 10 mM sodium acetate buffer, pH 5.0. The image was constructed using SigmaPlot® v13.0 software (Systat software).

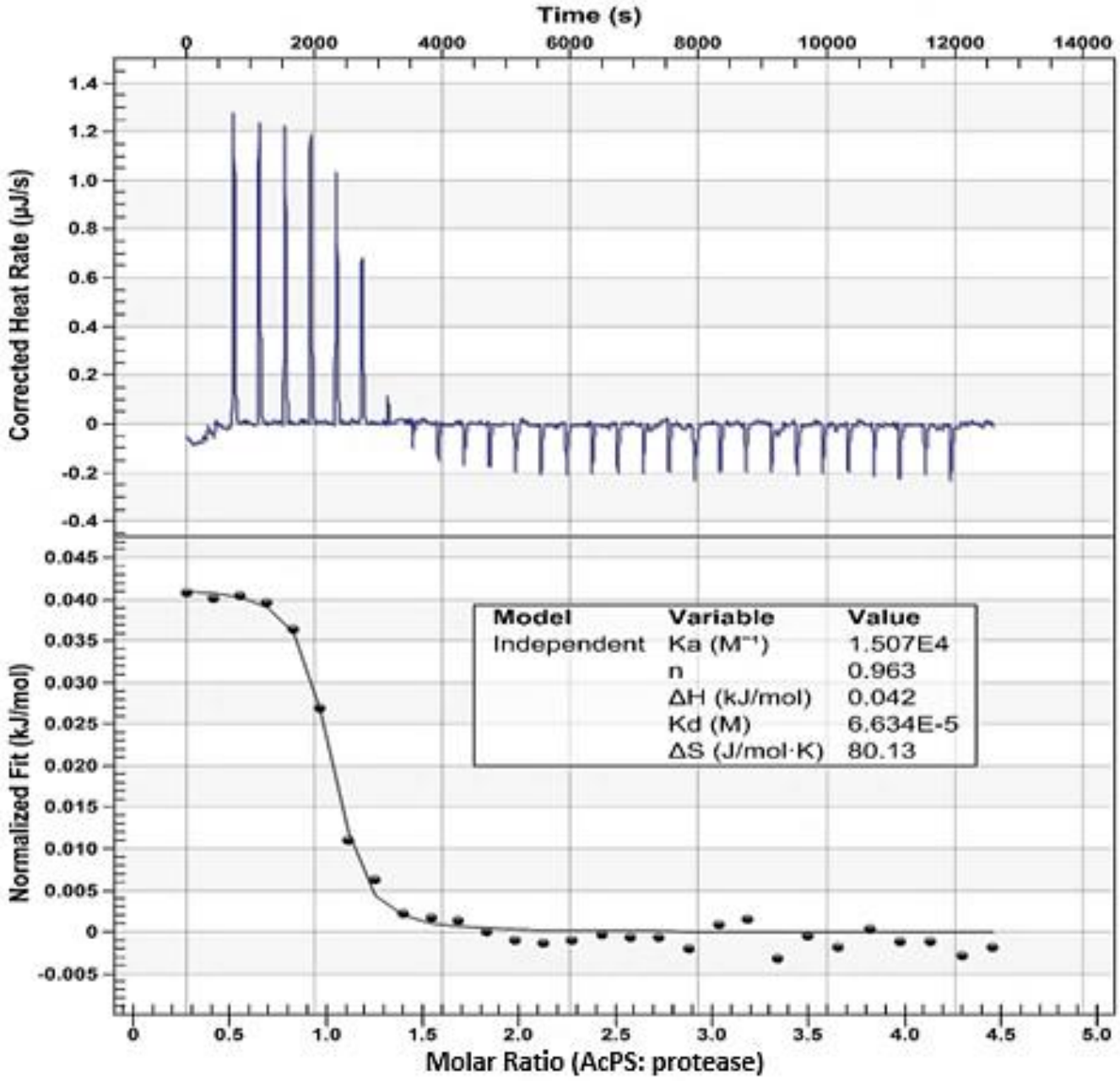


Figure 18: Active site titration of the protease with acetyl-pepstatin displayed in the ITC thermogram at 20 ºC. The stoichiometry of binding (n) is 0.963 (96.3% active protease). Thus, the corrected wild-type HIV-1 C-SA protease concentration is 19.1 µM. The data was fitted with a one-to-one fitting model using NanoAnalyze (TA Instruments).

As the binding reaction carries on, the area under each peak eventually decreases due to increased occupancy of available binding sites on HIV-1 protease and, thus, saturation occurs. The bottom panel represents the binding isotherm obtained by plotting the integrated heat obtained after each injection versus the ratio of concentration of the inhibitor and protein added, corrected for heats of dilution. The binding isotherms are best described as monophasic with a sigmoidal fit in the data plot representing the decrease in available binding sites on the protein due to ligand binding. The heat of binding of the acetyl pepstatin to wild-type HIV C-SA PR was used to determine the stoichiometry. It is known that acetyl pepstatin binds to the wild-type protease in a 1:1 ratio; as such, one molecule of acetyl pepstatin binds to a molecule of protease. The molar ratio of the binding isotherm gives us the stoichiometry and the stoichiometry obtained from the titration data was used to calculate the active site concentration for the wild-type C-SA PR. For example, a stoichiometry of 1 indicates that 100% of the PR molecules are in the active conformation and are capable of binding to the ligand. On the other hand, if, for example, n = 0.9, this would correspond to 90% of the wild-type C-SA PR sample existing in the active conformation. Thus, the result indicated how much of the wild-type C-SA PR enzyme is active in the sample cell. The thermodynamic studies using ITC revealed the percentage of active enzyme had a stoichiometry value (n) of 0.963 (Figure 18). As seen in figure 18, the binding affinity, stoichiometry and the enthalpy and entropy of biomolecular interactions were determined in a single experiment. The thermodynamic parameters from the analysis of titration data for the wild-type HIV-1 C-SA protease yielded a ΔS of 80.13 J/molK, ΔH of 0.042 kJ/mol, K_d of 6.634×10^{-5} M and a K_a of 1.51×10^{4} M and the values are similar to the published data (Mosebi *et al*., 2008; Naicker *et al.*, 2012). All titrations were carried out using a standard volume in a Nano ITC (TA Instruments) and the data were fitted using NanoAnalyze (TA Instruments).

3.7 Structural characterisation of wild-type HIV-1 C-SA PR

Structural characterisation of HIV-1 C-SA wild-type protease was investigated using an array of biophysical techniques to probe various aspects of the enzyme such as the secondary, tertiary and quaternary structure.

3.7.1 Secondary structural characterisation using far-UV circular dichroism spectroscopy

Far-UV CD spectroscopy was used to characterise the secondary structure of wild-type protease in the far-UV range of 250 nm to 190 nm. The spectrum in Figure 19 shows that HIV-1 C-SA wild-type protease has a trough at 216 nm indicating that protease is predominantly β-sheeted. The trough is due to the n-π^* transition of the peptide backbone (Hennessey and Johnson, 1981; Modi *et al*., 1993). Figure 20 shows that homodimeric protease consists of nine β-strands, two α helices, and random coils and that can also be confirmed by studying the crystal structure of the HIV-1 protease (Figure 5) (Adler *et al*., 1973; Woody, 1995). The reported results in the study in Figure 19 are consistent with crystal structure evidence (Figure 20) and previous CD studies on HIV-1 protease, which report that the protein is about 4 % alpha-helical, 48 % ß-sheeted and 48 % aperiodic structures (Navia *et al.*, 1989; Dash and Rao, 2001; Foulkes *et al.*, 2006). The far-UV CD spectrum confirms that the wild-type HIV-1 C-SA protease secondary structure is intact, and the spectrum agrees with published literature (Navia *et al*., 1989; Woody, 1995; Kelly & Price, 2000; Mosebi *et al*., 2008; Noel *et al.*, 2009).

3.7.2 Tertiary structural characterisation using intrinsic fluorescence spectroscopy

Intrinsic fluorescence spectroscopy was used to study the local tertiary structure of wild-type HIV-1 C-SA protease. The fluorescence of the indole ring of tryptophan and tyrosine and their locations were used as probes to monitor the localised conformational structure induced in the protein as they are sensitive to changes in the local environment. Figure 21 shows the fluorescence spectra of wild-type HIV-1 C-SA protease at 280 nm and 295 nm excitation.

The dimeric HIV-1 C-SA wild-type protease has two Trp residues (W6 and W42) and one Tyr residue (Y59) in each monomer. These residues were used to monitor only local structure (Figure 20). The probes (Trp and Tyr residues) absorb light at 280 nm; however, the indole ring of Trp was excited both at 280 nm and 295 nm. The fluorescence emission spectra of both tryptophan and tyrosine residues (Trp 6/6', Trp 42/42', and Tyr 59/59') were monitored after excitation at 280 nm. The tryptophan residues were selectively excited at 295 nm. It is known that free Trp fluoresces maximally at 350 – 355 nm in water (Lakowicz, 1999) and the spectrum in Figure 21 shows a maximum emission obtained at ~355 nm. Thus, the four tryptophan residues in the homodimer are completely exposed to the solvent between the hinge and flap region, while the tyrosine residue protrudes through the flap of the protease. Trp can lose energy when solvent exposed; thus, it is highly sensitive and strongly influenced by its local

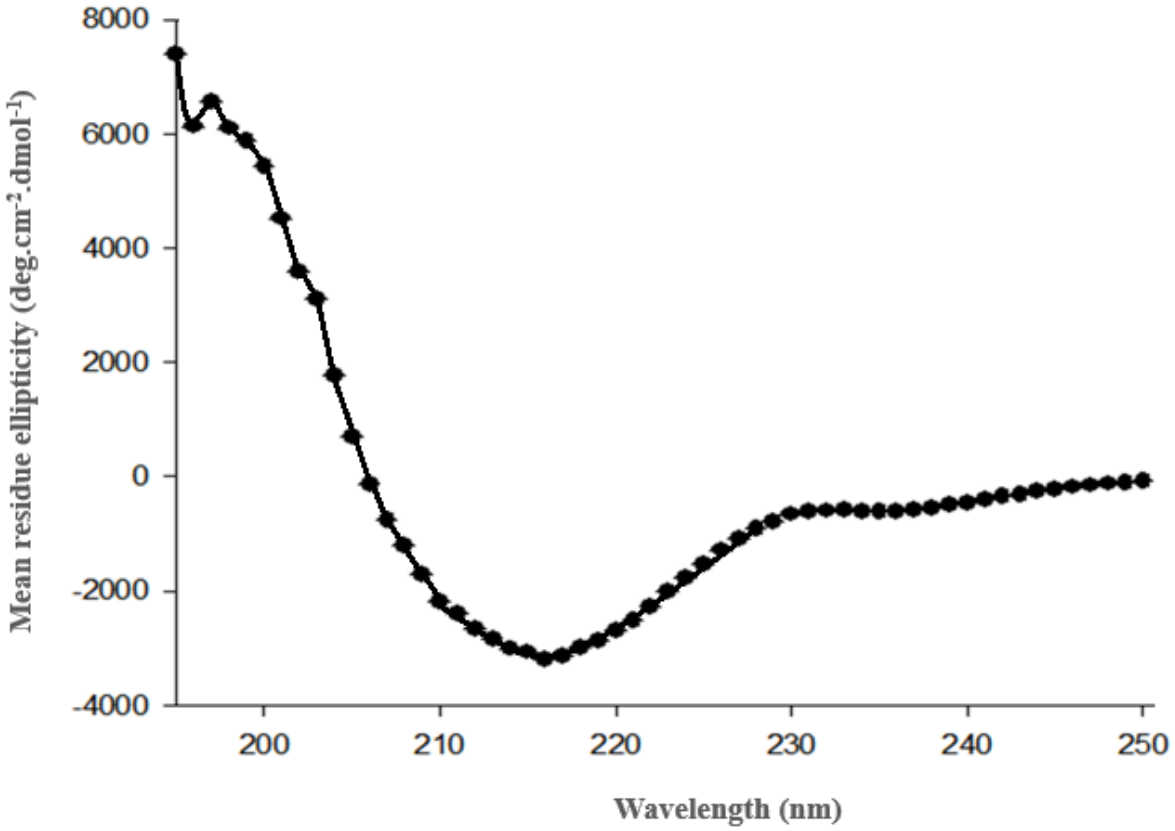


Figure 19: Far-UV CD spectrum of wild-type HIV-1 C-SA protease. The secondary structure was characterised in a far-UV range of 250 nm to 190 nm. The experiment was performed in triplicate at 20 ºC using 5 µM protease in 10 mM sodium acetate buffer, pH 5. The far-UV CD spectrum of the WT C-SA PR has a mean residue ellipticity trough at 216 nm and a peak at 201 nm. The data fitting was carried out using SigmaPlot® v13.0 software (Systat software).

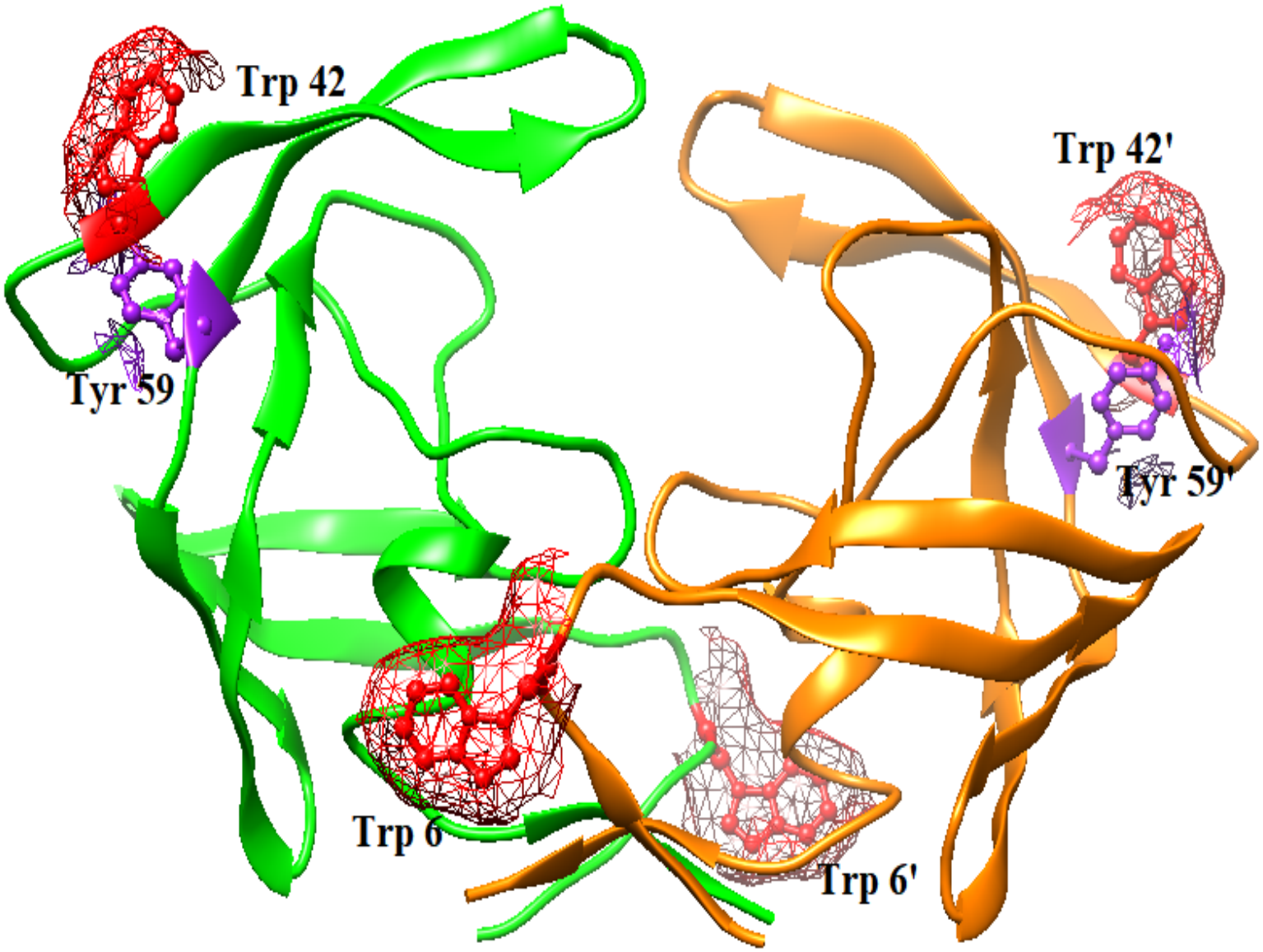


Figure 20: Three-dimensional structure of C-SA PR (PDB I.D. 3U71) showing the location of tyrosine and tryptophan residues. The two monomers are represented in orange and green and the fluorophores in red (W6/6' and W42/42') and purple (Y59/59'). Tryptophan is solvent exposed; thus, the surface covering the residues is a red mesh form and tyrosine is protruded through the flap of the protease. The figure was generated using UCSF Chimera v 1.12 with PDB code 3U71 (Naicker *et al.*, 2012).

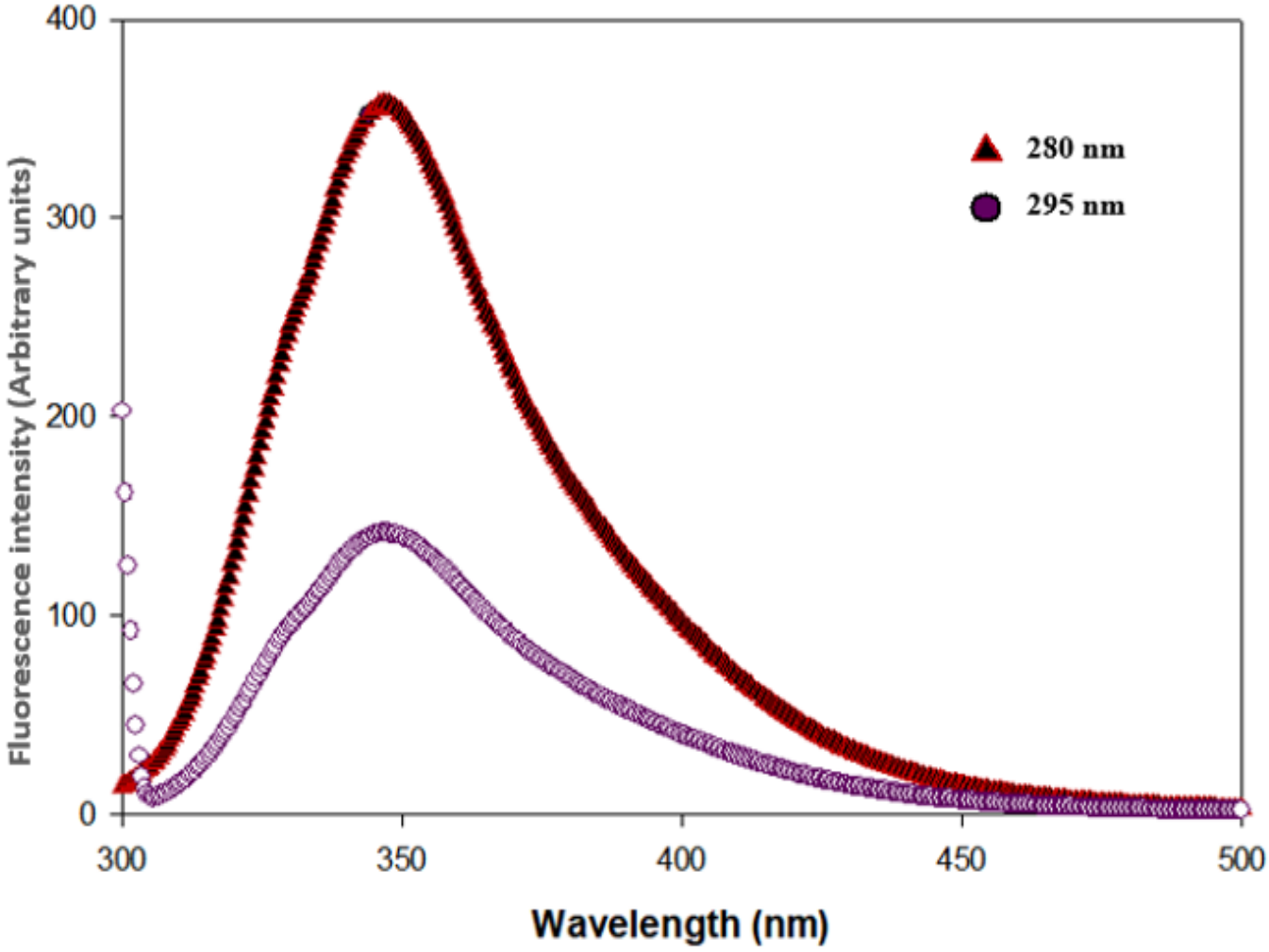


Figure 21: Intrinsic fluorescence emission spectra of tryptophan and tyrosine excited at 280 nm and tryptophan excitation at 295 nm. The emission maximum is at 355 nm at both excitation wavelengths. The excitation at 280 nm produced an emission signal approximately two times higher than the excitation at 295 nm. The above spectra was generated using SigmaPlot® v13.0 software (Systat software).

environment in terms of intensity and wavelength. The greater the exposure of Trp to the polar aqueous environment, the longer its wavelength of maximum emission will be (Royer, 1995). The fluorescence spectrum does not suggest any structural changes in the local environment of the tryptophan and tyrosine residues of the wild-type protease.

3.7.3 Quaternary structural characterisation using size exclusion – high-performance liquid chromatography

To assess the oligomeric state of the protease, size exclusion-high performance liquid chromatography (SE-HPLC) was performed under non-denaturing conditions. SE-HPLC is used to determine the molecular weight, diffusion coefficient, hydrodynamic volume and surface properties of protein molecules. The calibration curve of SE-HPLC in Figure 22 shows that larger biomolecule eluted first from the gel filtration column as they cannot penetrate the pores of the packed material. The proteins with smaller hydrodynamic size was trapped and retained longer within the column beads. The molecular standards used had a range of 5 – 150 kDa and the oligomeric size of the protease falls within the size range. The mark (x) in Figure 22 denotes the position of protease and the calibration curve of SE-HPLC molecular weight markers: albumin 66 kDa, carbonic anhydrase from bovine erythrocytes 29 kDa, ribonuclease A 13.7 kDa, aprotinin 6.5 kDa. The elution profile of the standards (Figure 22) was used to draw a standard curve in order to determine the molecular weight of the protease. The standard curve was plotted using the log of the molecular weight of the standards and their retention time. The standard curve straight line equation was used to calculate the molecular weight of the wild-type HIV-1 C-SA protease by interpolation. Analysis of the standard curve indicated that the homodimeric integrity of the protease was maintained and with an apparent oligomeric molecular mass of 22 kDa, which correlates with the homodimer molecular mass of protease (Gasteiger *et al*., 2005; Mosebi *et al.*, 2008; Mpye, 2010, Naicker, 2013).

3.8 Protease enzyme kinetics

The K_M, $V_{max,}$ specific activity, k_{cat} and k_{cat}/K_M of C-SA PR determined in this study did not differ greatly from that of published studies (Naicker *et al*., 2013; Naicker *et al.,* 2014) with wild-type C-SA PR and subtype B PR. The kinetic parameters were comparable for the fluorogenic substrate binding and catalytic activity between this study and that in the literature (Naicker *et al*., 2013; Naicker *et al.,* 2014). The catalytic parameters for wild-type HIV-1 C-SA protease were determined by monitoring the hydrolysis of the fluorogenic substrate (Abz-Thr-Ile-pNO_2Phe-Gln-Arg-NH_2). The hydrolysis of the fluorogenic substrate was conducted

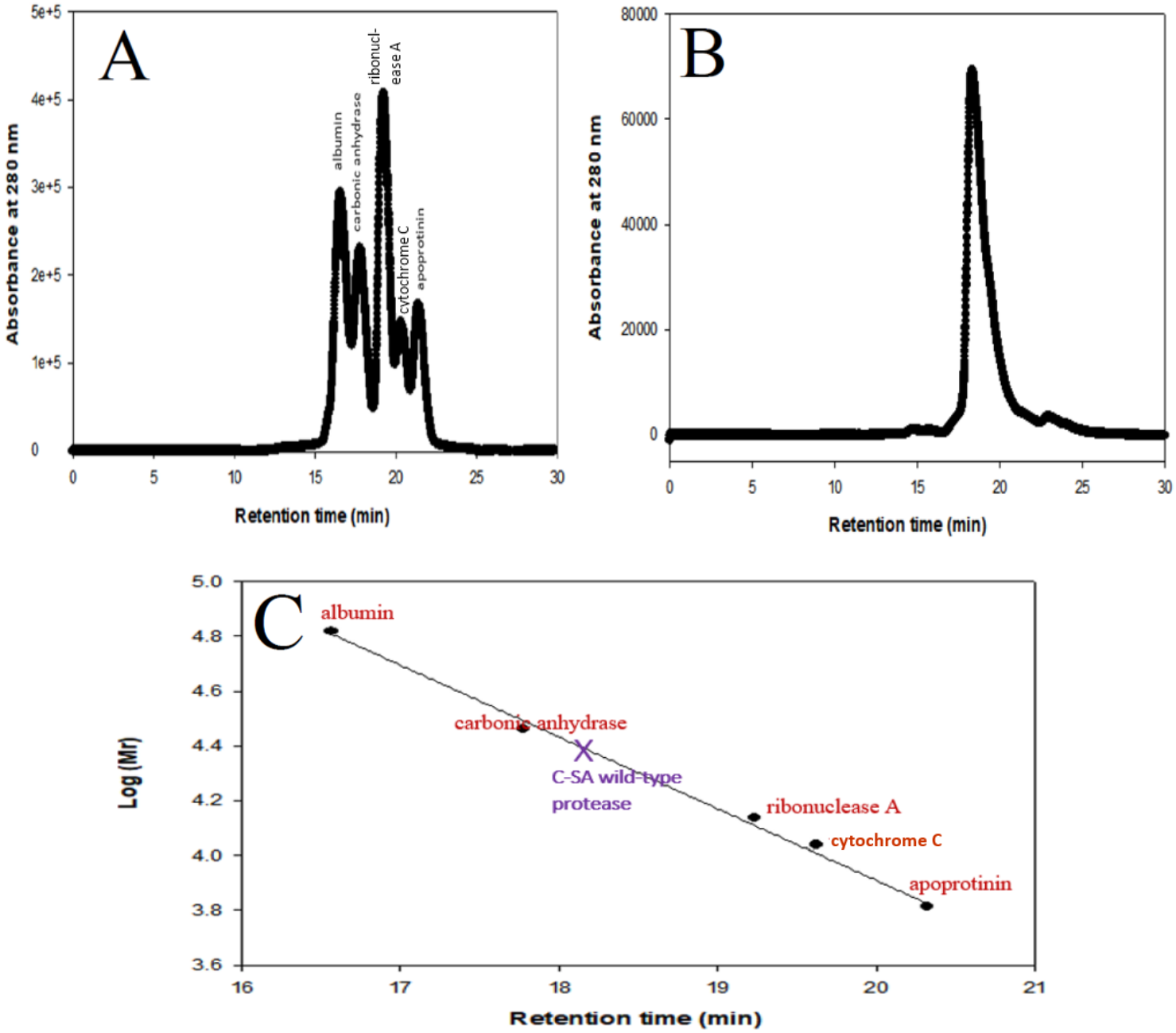


Figure 22: SE-HPLC elution profile of the wild-type HIV-1 C-SA protease. (A) SE-HPLC column elution of molecular standards and (B) wild-type HIV-1 C-SA protease. The column was equilibrated with a 10 mM sodium acetate buffer, pH 5 at a flow rate of 2 mℓ.min^{-1} with an isocratic pressure of 60 bar at 20 ^{o}C, and the eluted samples were monitored at 280 nm. Wild-type HIV-1 C-SA protease was eluted at 18.3 minutes. (C) Calibration curve of SE-HPLC molecular weight markers: albumin 66 kDa, carbonic anhydrase from bovine erythrocytes 29 kDa, ribonuclease A 13.7 kDa, cytochrome C 12.4 kDa, aprotinin 6.5 kDa. The standard curve was from the log of the molecular weight of the standards and their retention time. The (x) indicates the position of the eluted wild-type protease with a corresponding molecular mass of 22 kDa. The data fitting was carried out using SigmaPlot® v13.0 software (Systat software).

in triplicate for each kinetic assay in order to generate reproducible data and the data was reported as the mean ± standard deviation (SD).

3.8.1 Determination of K_M and V_{max}

The Michaelis-Menten constant, K_M, and the maximal rate, V_{max}, were derived from the rates of catalysis measured at different [S] since the enzyme operated according to the simple scheme given in equation 6. The K_M and V_{max} were determined through gradually increasing concentrations of the fluorogenic substrate (5 to 200 μM) which was hydrolysed by 2 μM wild-type HIV-1 C-SA protease until saturation occurs. The enzyme assays were conducted at 20 ^oC in 10 mM sodium acetate buffer, pH 5.0. The K_M and V_{max} values were determined using a curve-fitting computer program (SigmaPlot v.13). Figure 23 represented as a hyperbolic curve known as Michaelis-Menten plot which is for the determination of K_M and V_{max} values using an equation for the hyperbolic relationship. K_M value was found to be 42.07 μM and V_{max} was 0.047 μmol/min and the correlation coefficient R^2 was 0.99. The K_M was higher than that shown in literature (Naicker *et al.*, 2013; Naicker *et al.*, 2014), suggesting that the fluorogenic substrate binds slightly weaker to the wild-type protease compared to that published in the literature (Table 3). The difference in K_M and V_{max} could be that the conditions for determining the kinetic parameters were slightly different from that shown in literature or that the protease lost some of its activity due to freezing and thawing. The difference in wild type subtype C protease in this study and subtype B in the literature (Naicker *et al*., 2013; Naicker *et al.*, 2014) in the stronger binding of subtype B could be due to the closed conformation of subtype B, which makes it more stable (i.e. it has a higher melting temperature) than the wild-type protease in this study. If the flaps are in an open conformation, they lower the binding affinity of substrates or inhibitors because of the increased energy required to bring the flaps into a closed conformation (Naicker *et al*., 2013; Naicker *et al.*, 2014).

3.8.2 Determination of specific activity and catalytic turnover (k_{cat})

When substrate concentration is much greater than K_M ([S] >> K_M), the rate of catalysis is equal to V_{max}, which is a function of k_{cat}. k_{cat} is depicted by the activity units, which is the fluorogenic substrate hydrolysed by wild-type protease over time. Specific activity is the unit of enzyme divided by the concentration of enzyme. Specific activity assays confirmed that the protease was catalytically active since it was recovered and refolded from inclusion bodies. The specific activity and turnover number (k_{cat}) of wild-type HIV-1 C-SA protease were determined from the same data set under steady-state conditions (enzyme saturated with the substrate).

Table 3. Comparison of the kinetic parameters of C-SA PR obtained experimentally in this study with the literature kinetic parameters of C-SA PR. The kinetic parameters were obtained using fluorogenic substrate (Abz-Thr-Nle-pNO2Phe-Gln-Arg-NH2).

Kinetic parameter	**This study**	**C-SA PR (Naicker *et al.*, 2013)**	**C-SA PR (Naicker *et al.*, 2014)**	**Subtype B PR (Naicker *et al.*, 2014)**
Specific activity (μmol/min/mg)	76.46	79	-	-
Turnover number (k_{cat}) (s^{-1})	24.02±0.4	-	20.6 ±0.8	9.0 ±0.2
K_M (μM)	42.07±3.8	-	29.4 ±4.4	38.7 ±6.0
V_{max} (μmol/min)	0.0465±0.1	-	-	-
k_{cat}/K_M (s^{-1} μM^{-1})	0.084±0.5	-	0.46 ±0.05	0.32 ±0.01

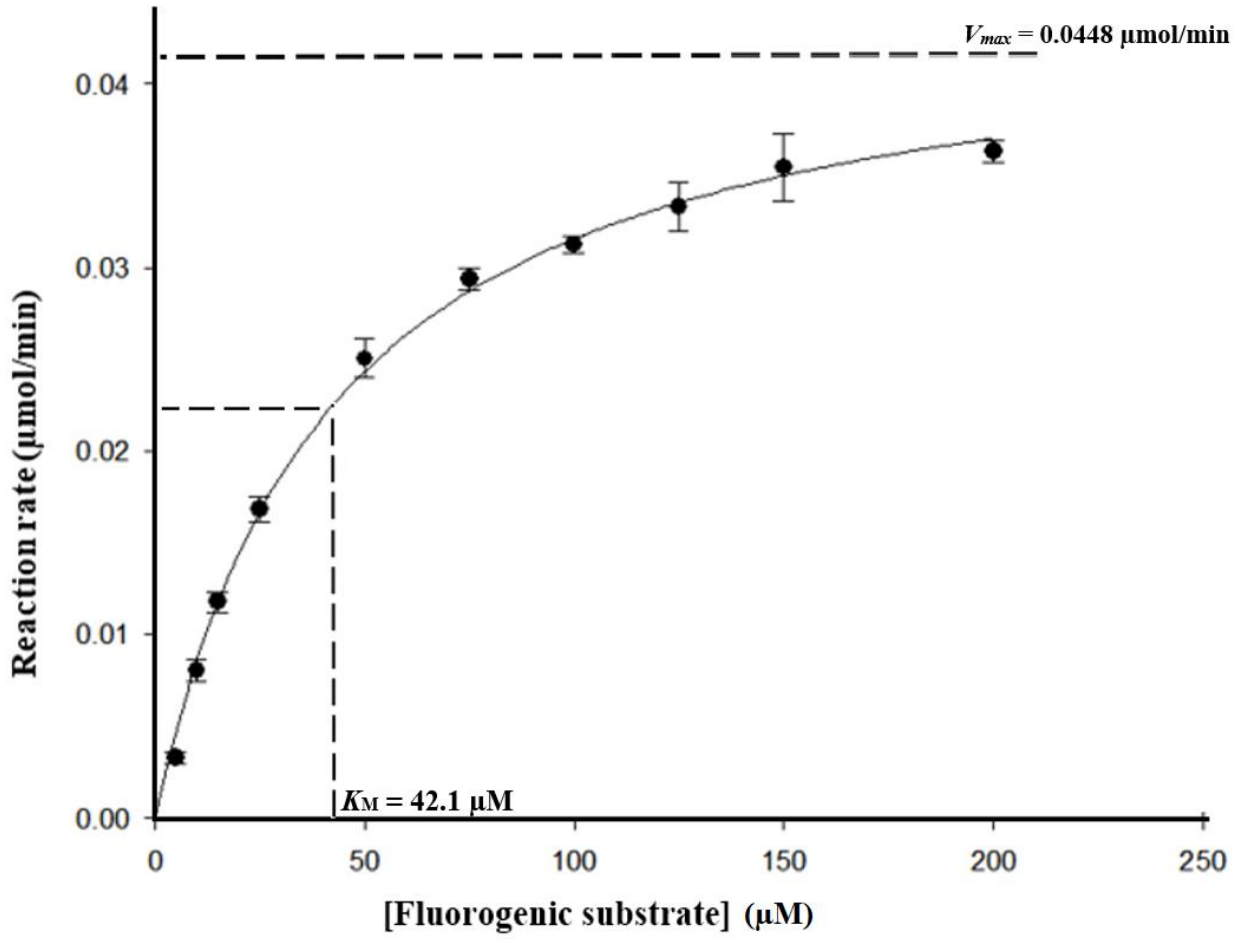


Figure 23: Michaelis-Menten plot for the determination of K_M and V_{max}. The Michaelis-Menten plot of the hydrolysis of fluorogenic substrate catalysed by wild-type C-SA protease is represented in a hyperbolic curve. The data fitting was carried out using SigmaPlot® v13.0 software (Systat software), and K_M and V_{max} were determined using Michaelis-Menten equation (equation 6). K_M value was determined as 42.1 µM and V_{max} was 0.0448 µmol/min. The correlation coefficient (R^2) was 0.99.

The units of specific activity were converted to catalytic turnover. Figure 24 illustrates specific activity graphs and Figure 25 shows the turnover number graph of wild-type HIV-1 C-SA protease. The specific activity value was 65.51 μmol/min/mg and turnover number value was determined as 24.02 s^{-1}, and both their correlation coefficient R^2 values were 0.99. The values are from the gradient of the linear equation fitted using SigmaPlot® v 13.0. Each type of assays was conducted at 20 ºC in 10 mM sodium acetate buffer, pH 5.0. The protease concentration was varied between 10 – 50 nM while the substrate concentration was kept constant at 500 μM.

The specific activity and k_{cat} value in the study and that of C-SA PRs in literature are comparable (Table 3). However, the wild-type protease in both literature and the current study show a 2-fold increase in substrate turnover per second (higher k_{cat}). This improved rate of substrate processing may stem from a 23% increase in the proportion of fully-open conformers evident in the C-SA PR population relative to that of the subtype B protease (Kear *et al*., 2009). Importantly, structure-based calculations reveal that semi-open conformers do not permit the entry of substrate/inhibitor to the active site (Rick *et al*., 1998). Entry of a substrate or inhibitor to the active site of the PR requires substantial movement of the flaps and flap flexibility is a requirement for substrate binding and product release (Tozser et al., 1997). Therefore, only the fully-open conformer may allow a substrate/inhibitor access to the active site. An increased preference for the fully-open conformer may improve the rate substrate entry and/or product release, evidenced by the increase in k_{cat} for the wild-type subtype C protease (Tabale 3).

3.8.3 Catalytic efficiency (k_{cat} / K_M)

The ratio of k_{cat}/K_M gave the index of the catalytic efficiency. The catalytic efficiency (k_{cat}/K_M) of wild-type HIV-1 C-SA protease was calculated based on different catalytic activities measured for sixty seconds using a 2 μM wild-type HIV-1 C-SA PR and with gradually increasing concentrations of the fluorogenic substrate (1 – 5 μM). The assay was conducted at 20 ºC in 10 mM sodium acetate buffer, pH 5.0. Figure 26 represents a linear curve and the slope of the graph from the linear equation was used to calculate k_{cat}/K_M. k_{cat}/K_M value was determined as 0.084 $\mu M^{-1}.s^{-1}$. The wild-type protease in this study was not as efficient as the literature wild-type as it seems to have a much lower k_{cat}/K_M which is about five-fold. The lower catalytic efficiency must be a result of some altered dynamics in the wild-type protease in the study as compared to that of the literature. Alterations could be due to the freezing and thawing of the protein, thus; the protease lost some of its activity.

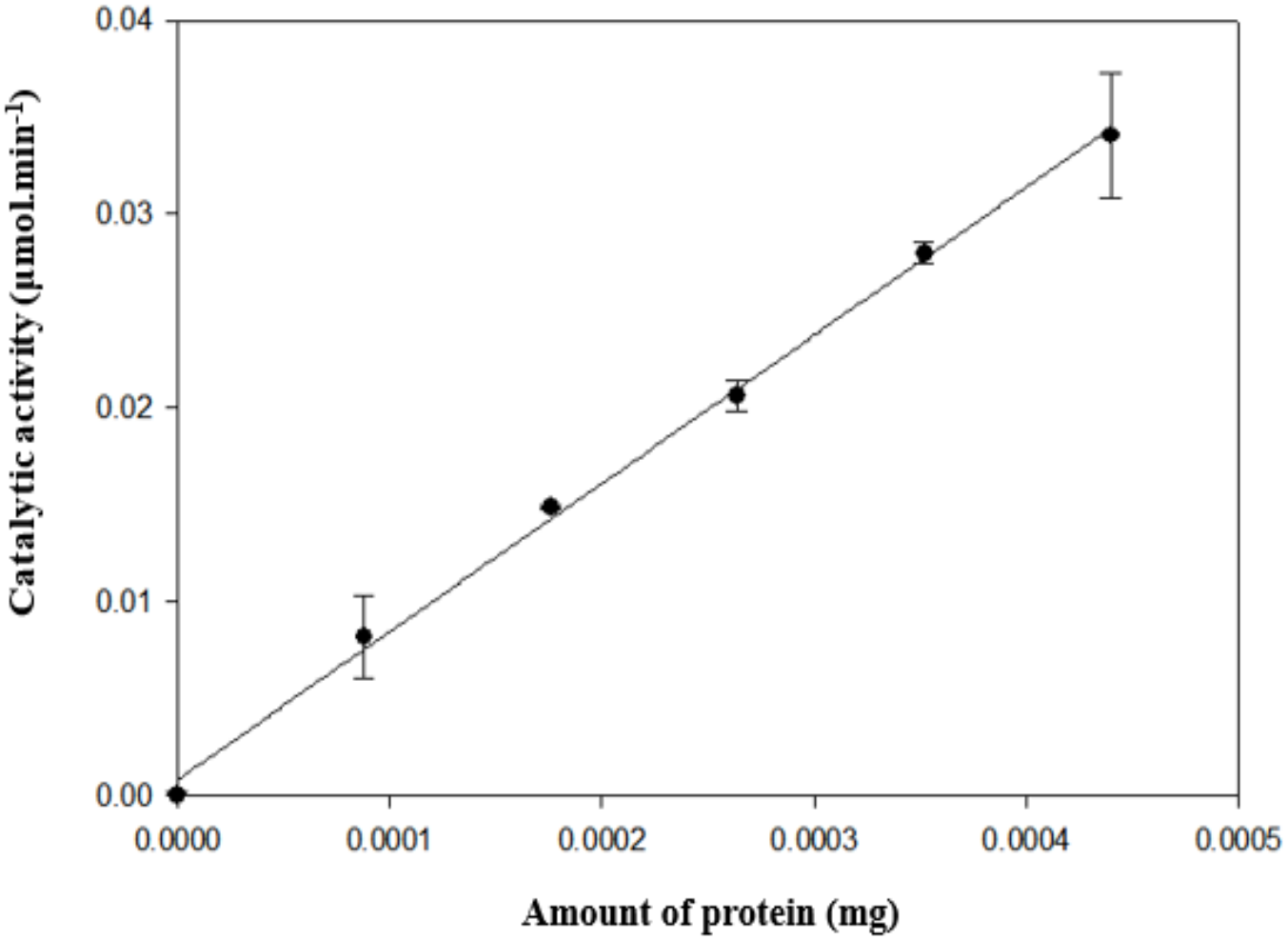


Figure 24: Determination of specific activity for the wild-type HIV-1 C-SA protease. The specific activity of the wild-type protease was calculated from the slope of the linear curve using a standard curve model from SigmaPlot® v13.0 and the value was determined to be 76.46 µmol/min/mg. The correlation coefficient of the linear fit was 0.9978.

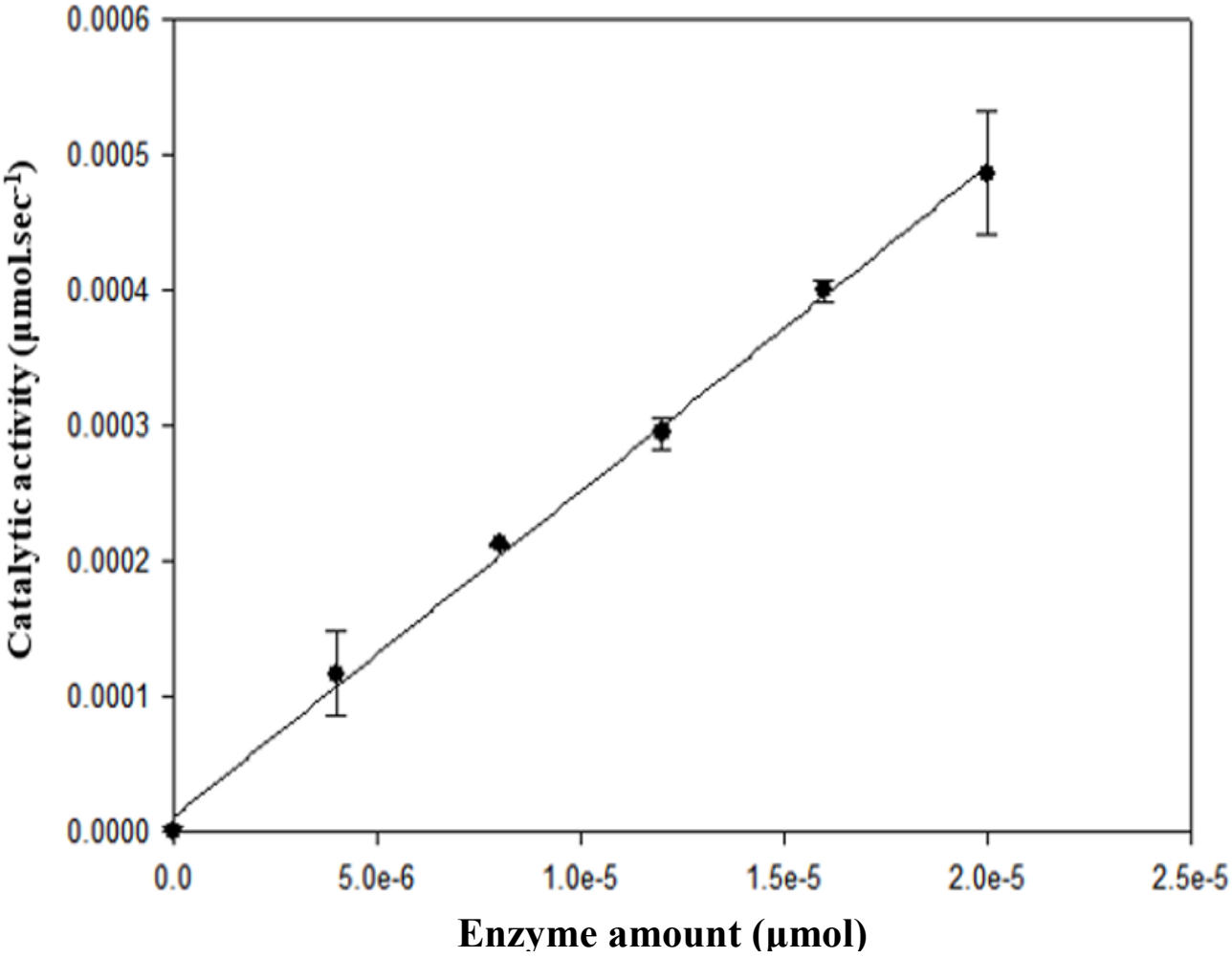


Figure 25: Determination of catalytic turnover (k_{cat}) for wild-type HIV-1 C-SA protease. k_{cat} was determined from the slope of the linear curve using a standard curve model from SigmaPlot® v13.0 and was found to be 24.02 s^{-1}, and the correlation coefficient was 0.9978.

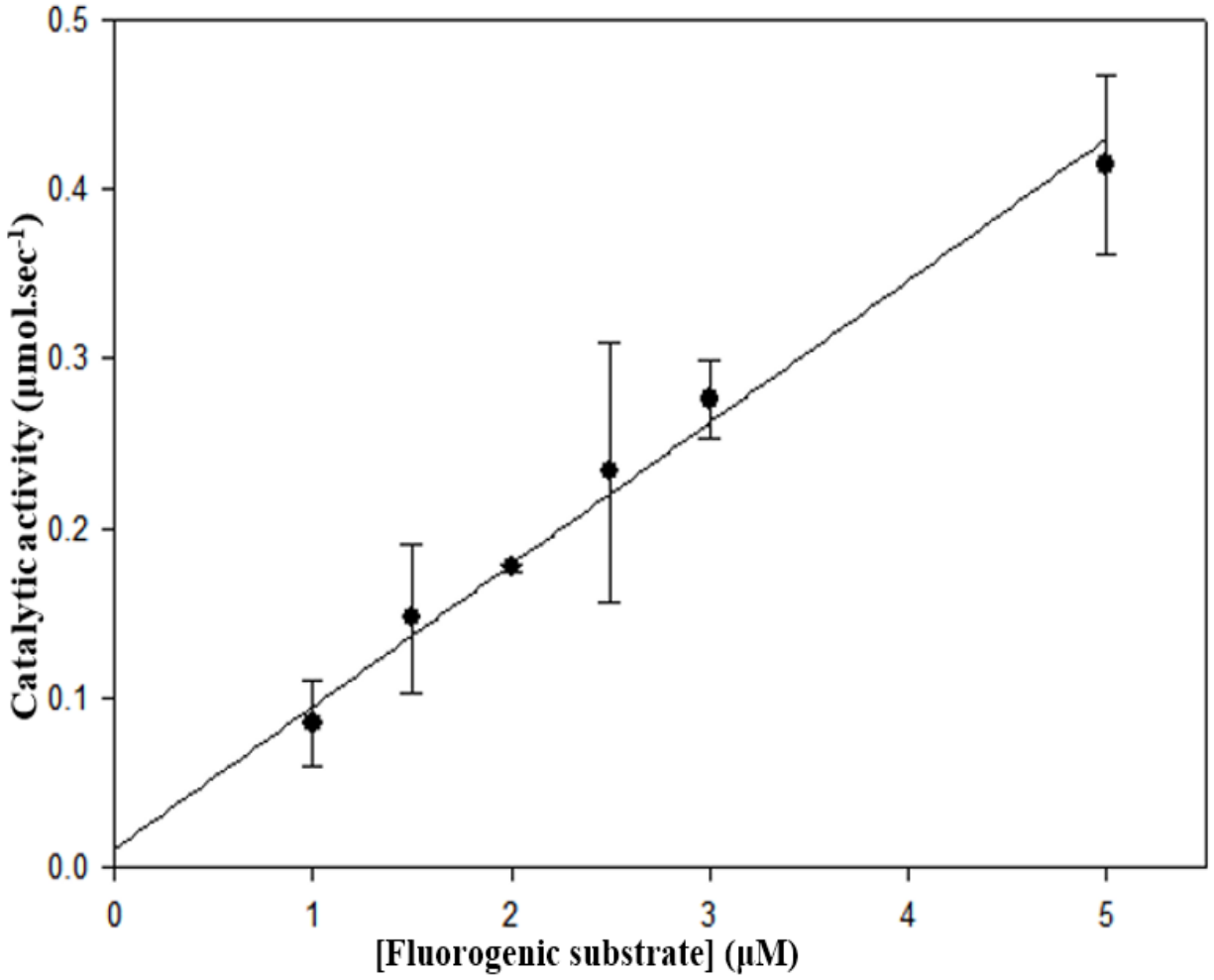


Figure 26: Determination of catalytic efficiency (k_{cat}/K_M) for wild-type HIV-1 C-SA protease. k_{cat}/K_M was determined from the slope of the linear curve using a standard curve model from SigmaPlot® v13.0 and the value was found to be 0.08 $\mu M^{-1}.s^{-1}$. The correlation coefficient of the linear fit was 0.9869.

3.9 Inhibition kinetics studies of GRD 110D, saquinavir, ritonavir and darunavir against wild-type HIV-1 C-SA protease

Enzymatic characteristics of wild-type protease were also assessed in the presence of the novel inhibitor compound and three clinically approved inhibitors. The concentration of each protease inhibitor that reduces the catalytic activity of HIV-1 C-SA wild-type protease by 50% was determined. Thus, IC_{50} results determined that darunavir is a significantly more potent C-SA wild-type protease inhibitor than saquinavir and ritonavir. The GRD 110D inhibitor tested did not indicate good binding with the wild-type protease (Figure 28). The fitted data sets for the IC_{50} determination for the protease inhibitors are displayed in Figure 27 A, B and C. The standard deviation is shown as error bars obtained from the experiment carried out in triplicate and the data was fitted with a three-parameter hyperbolic decay model using SigmaPlot® v. 13.0. IC_{50} values that were obtained from Figure 27 A, B and C were determined to be 32.5 nM for darunavir, 45.6 nM for ritonavir and 38.5 nM for saquinavir.

The dissociation constant (K_i) was calculated using equation 9 and the determined K_M was used to calculate the K_i of the four inhibitors (Copeland *et al*., 1995). The stronger the binding of protease inhibitors, the lower the K_i value; therefore, lower concentrations of inhibitor needed to achieve IC_{50}. The K_i values for darunavir was 3.4 nM, for ritonavir was 9.4 nM and for saquinavir was 6.2 nM. The IC_{50} and K_i values for the HIV-1 C-SA wild-type protease inhibitors were in the nanomolar range, in agreement with published data (Klabe *et al*., 1998; Velazquez-Campoy *et al*., 2001; Naicker *et al*., 2014). The novel protease inhibitor GRD 110D was soluble at 2% DMSO and the determined IC_{50} value was 85 nM, with a K_i value of 27 nM (Figure 28). The data fitted best to a three-parameter hyperbolic decay model and provided the IC_{50} values. Figure 29 represents the summary of the IC_{50} and K_i values for all inhibitors that were determined following the catalysis of the same fluorogenic substrate. The determination was done following the hydrolysis rate of a fluorogenic substrate in the presence of protease and increasing protease inhibitor concentrations.

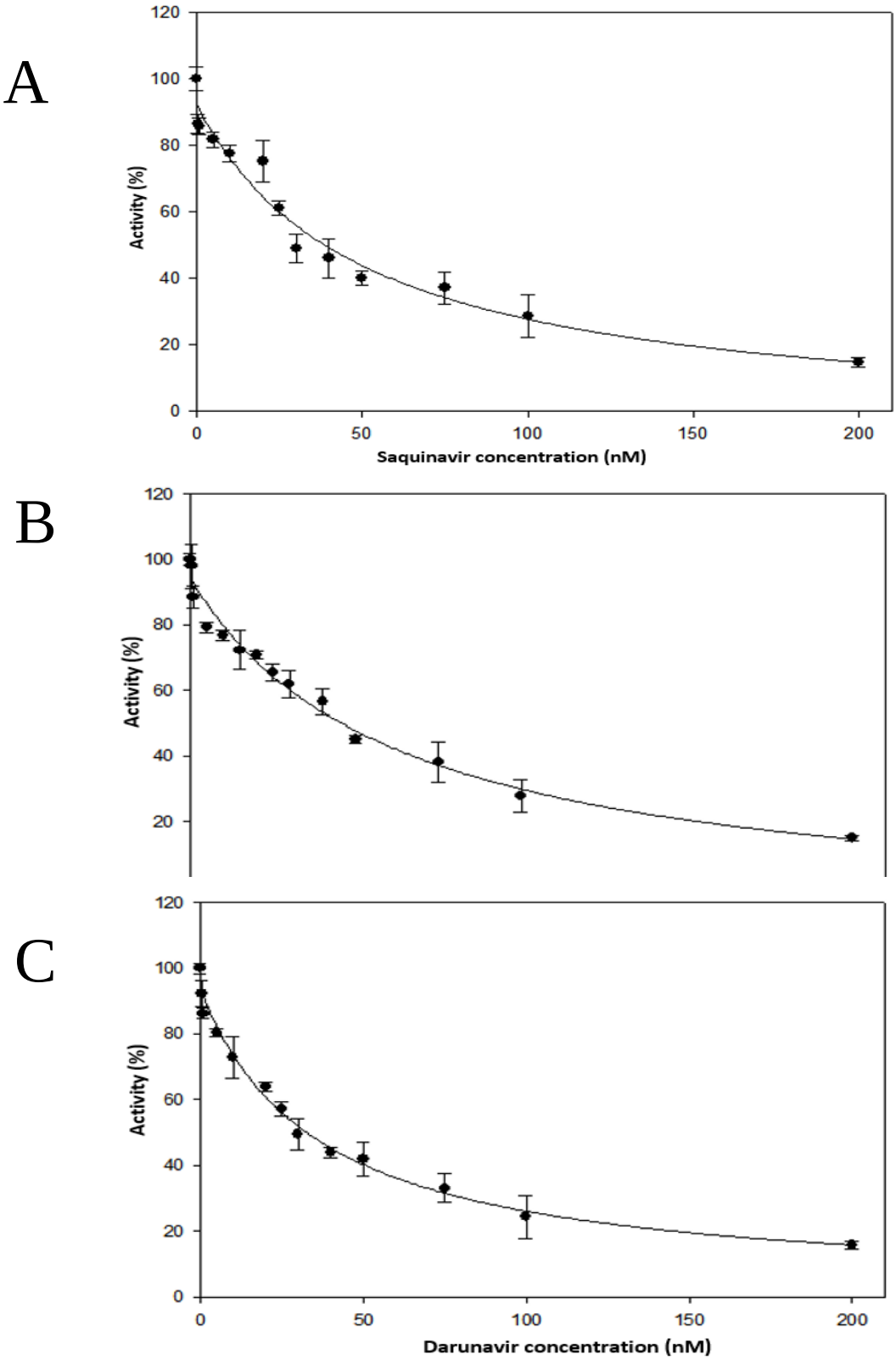


Figure 27: IC_{50} determination of (A) saquinavir, (B) ritonavir and (C) darunavir for the C-SA wild-type protease. IC_{50} value for darunavir was 32.5 nM, for ritonavir was 45.6 nM and for saquinavir was 38.5 nM. The correlation coefficient R^2 were 0.9633, 0.9804 and 0.9888 for saquinavir, ritonavir and darunavir, respectively. The data were fitted with a three-parameter hyperbolic decay model using SigmaPlot® v. 13.0.

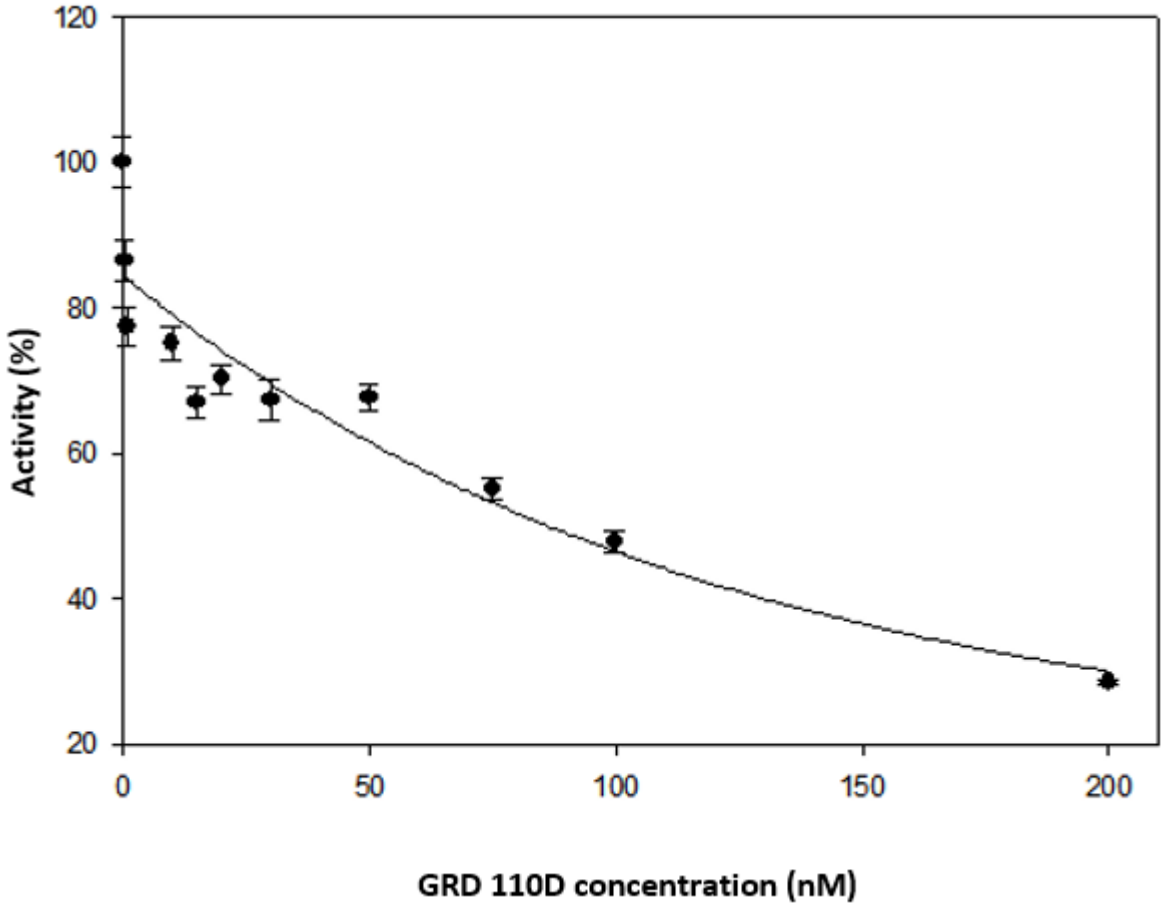


Figure 28: IC_{50} determination of GRD 110D for the C-SA wild-type protease. IC_{50} value was determined as 85 nM and the correlation coefficient R^2 was 0.8791. The data was fitted with a three-parameter hyperbolic decay model using SigmaPlot® v. 13.0.

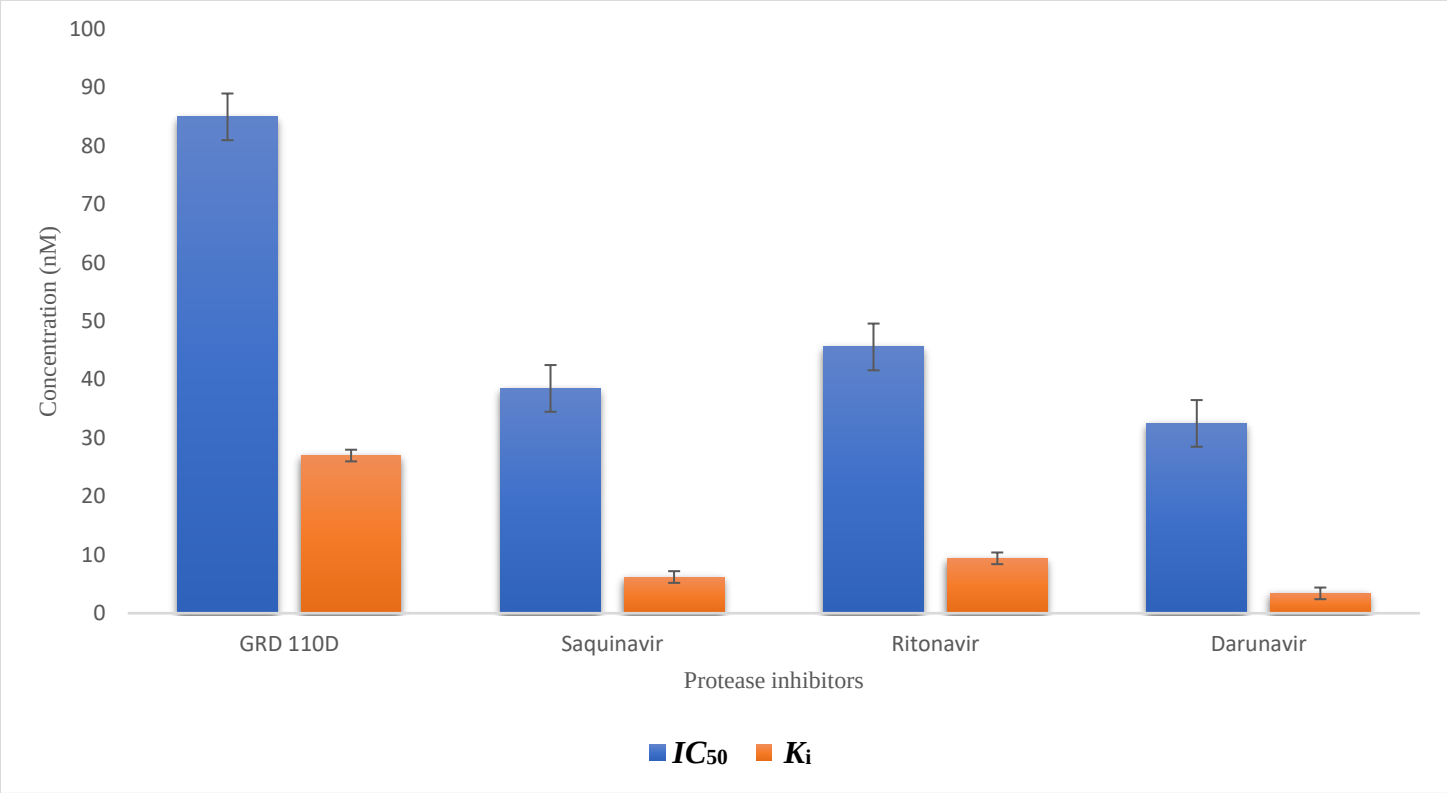


Figure 29: Graphical representation of the IC_{50} and K_i values of GRD 110D, saquinavir, ritonavir and darunavir for C-SA PR. Saquinavir, ritonavir have comparable inhibitory potencies. However, ritonavir is a stronger inhibitor of C-SA PR than saquinavir. Darunavir inhibits C-SA PR with the highest potency between the four inhibitors. The binding of darunavir is significantly tighter, having an IC_{50} value close to 35 nM (value expected for irreversible inhibitors). The error bars are standard deviation values obtained from the experiments performed in triplicate. The graph was produced using Microsoft® Office Excel 2019.

3.10 Induced-fit docking studies

There have only been a few studies that directly compared the different subtype B and subtype C protease based on their genetic differences, polymorphisms, particularly drug resistance mutations that give subtype C protease inadequate responses to the binding of PIs (Naicker *et al*., 2013; Naicker *et al.,* 2014). The Wits/PSFRU (Protein Structure-Function Research Unit) group conducted studies that suggested an increased catalytic efficiency and vitality in subtype C protease. The Wits/PSFRU and Freire group suggested that, in comparison to subtype B protease, there is a decrease in binding affinity in subtype A, subtype C, and subtype G in the presence of inhibitors (Velazquez-Campoy *et al*., 2003). Their research correlates with the results from induced-fit docking, concluding that PIs were designed and optimised to be more effective toward subtype B rather that subtype C protease. The computational investigation was aimed at understanding the binding mechanism of HIV PIs to subtype C and subtype B.

Molecular docking is regularly used in structure-based drug design to accurately identify the conformations of ligands to proteins and estimate the strength of the protein-ligand interaction. The commercial software Glide uses an E-model scoring function to choose between the protein-ligand complexes of a given ligand and the GlideScore function. The GlideScore function uses a rank-order for compounds in order to separate those that bind with high-affinity (actives) and from those that do not (inactive). A low E-model energy (kcal/mol) indicates a higher binding affinity of the ligand to the protein. The binding energy is from van der Waal interactions between the protein and the ligands. Ten ligands were docked into two protein structures 3U71 and 2P3B to determine the binding affinities of the ligands to the protein (APPENDIX A). The data for the E-model energy (kcal/mol) was plotted as a bar graph and reported as the mean ± standard deviation (Figure 30).

The Schrödinger Induced-fit docking algorithm related the molecular characteristics with biological activities of protein-ligand interaction and gave the results in a scoring function. The scoring function evaluated the binding affinity and the E-model graph (Figure 30) showed a reduction of binding affinity as binding energy (kcal/mol) for wild-type C-SA protease (PDB: 3U71) relative to the wild-type subtype B protease (PDB: 2P3B). Thus, wild-type C-SA protease has a lower drug susceptibility in comparison to wild-type subtype B to some of the FDA-approved protease inhibitors (Del Amo *et al*., 1998; Alaeus *et al*., 1999; Alexander *et al*., 2002; Pillay *et al*., 2002).

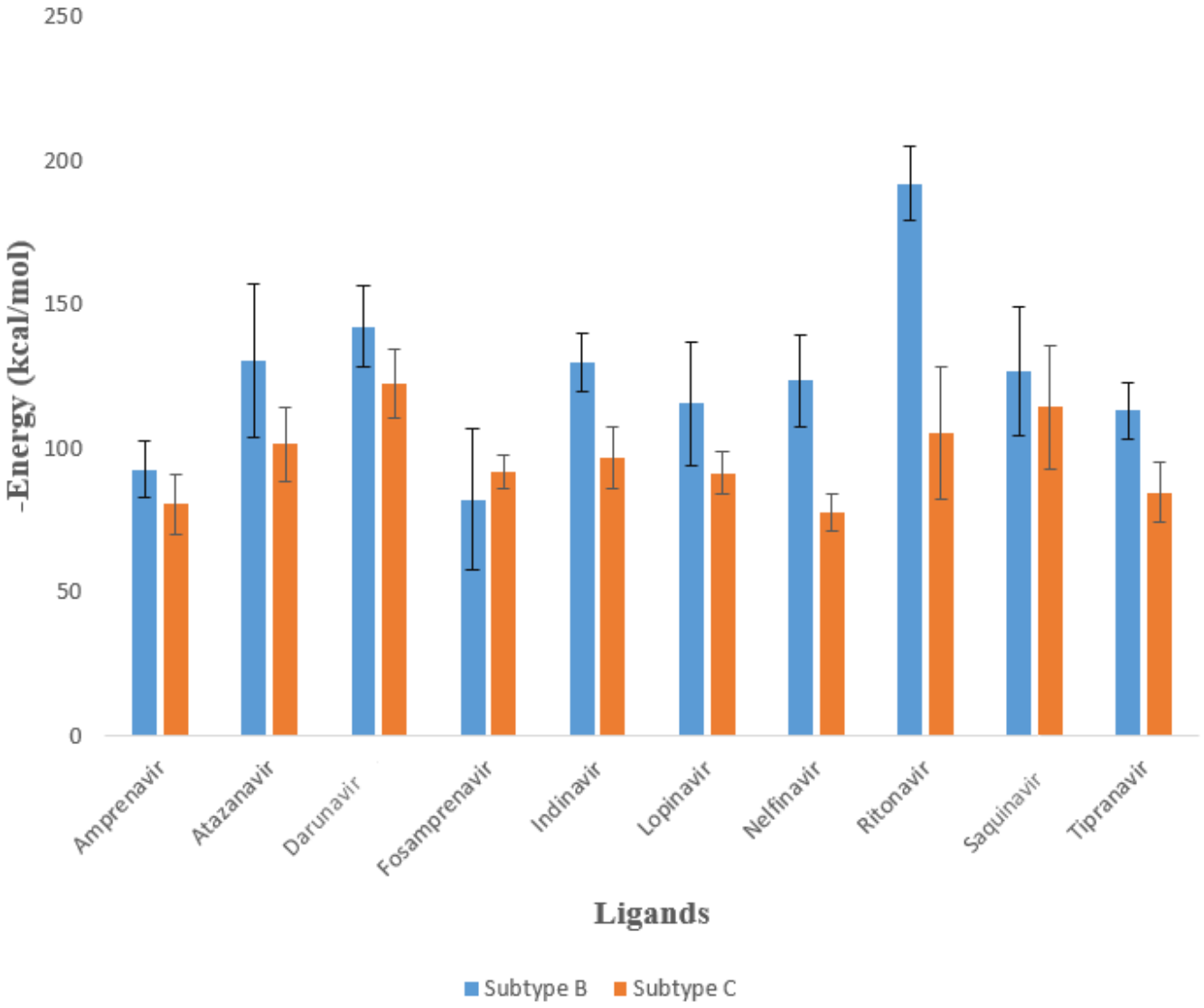
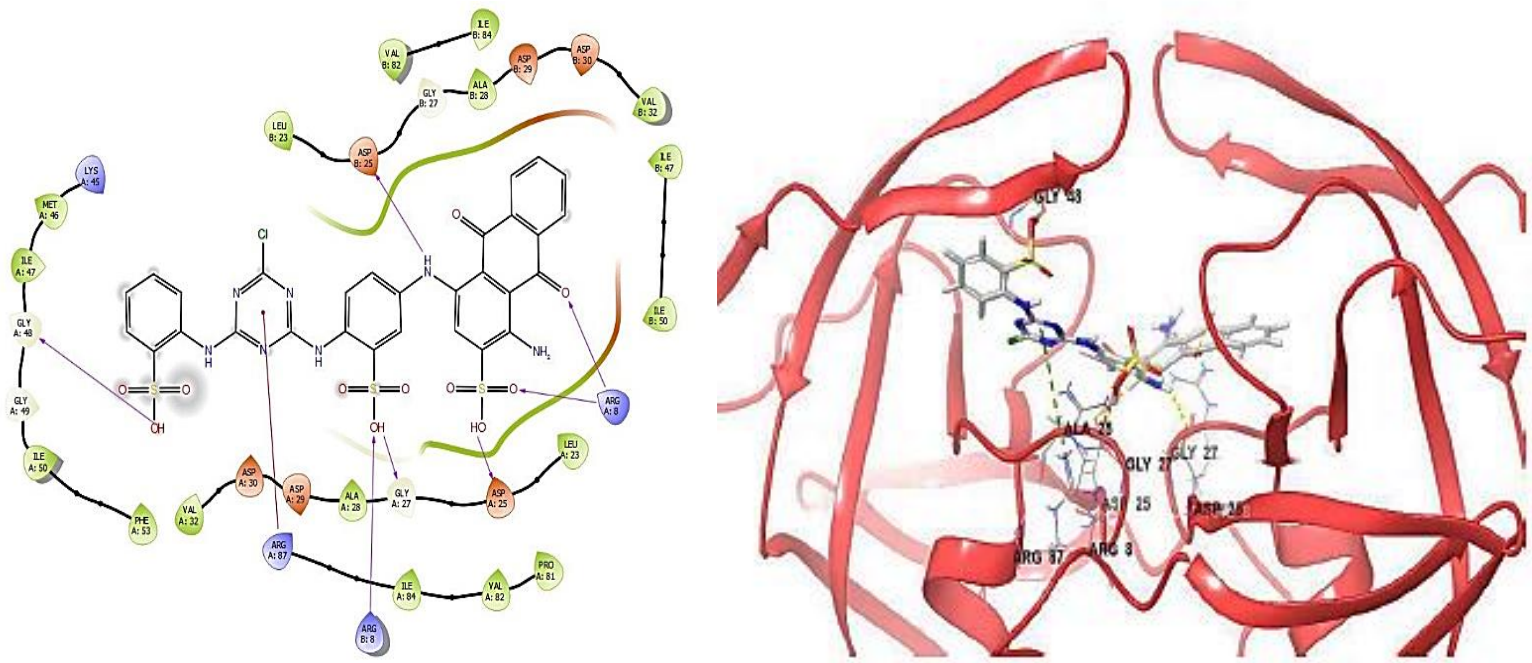


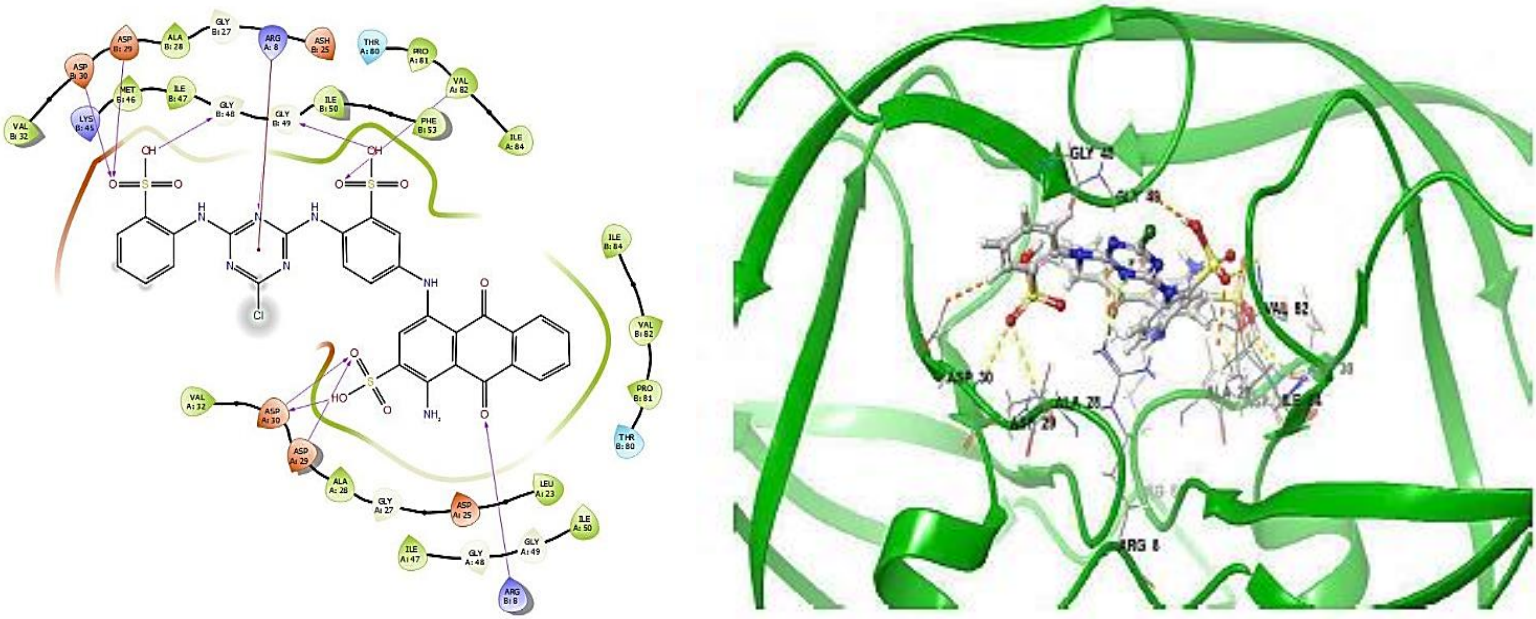
Figure 30: Bar graph representing the mean–E-model energy (kcal/mol) docking of 3U71 and 2P3B bound to FDA-approved protease inhibitors. All protease inhibitors with an exception of fosamprenavir are more potent inhibitor of subtype B PR than of C-SA PR Each inhibitor was docked with subtype B (blue) and subtype C (orange).

Figure 30 indicates that amprenavir, darunavir, lopinavir, saquinavir shows a 3-6 fold weaker binding, atazanavir, indinavir, nelfinavir, tipranavir shows an 8-24-fold weaker binding, ritonavir displays a 40-fold weaker binding to the C-SA protease in comparison to subtype B protease. However, fosamprenavir shows a 2-fold stronger binding in comparison to the two subtypes. Darunavir displays a higher binding affinity to HIV-1 protease in comparison with other inhibitors due to its high affinity and induced-fit mechanism (Naicker *et al*., 2013; Naicker *et al.*, 2014). Most inhibitors have a weaker binding affinity towards subtype C in comparison with subtype B protease. These could most possibly emerge from computational error from the lack of data on the experimental free energy binding. The experimental results of IC_{50} and K_i values correlate with the induced-fit docking results; thus, the combination of experimental data and a combination of computational methods improves the reliability and accuracy of the results.

The molecular docking results (Figure 31) indicate the difference in drug efficacy between the two subtypes. Subtype B binding is better due to the key difference in the hinge region (amino acid residues at positions 35-42 and 57-67). The consensus subtype B protease differs from South African subtype C protease at eight positions in each monomer (T12S, I15V, L19I, M36I, R41K, H69K, L89M and I93L). These differences influence the stability of the hinge region which further contributes to the movement of the flap region (amino acid residue positions at 46-54). Thus, subtype B flaps are less flexible with smaller inter-flap distances and this is associated with stronger interactions with the inhibitors. The extent of flap opening for both subtype B and subtype C protease are very much similar; however, subtype B flaps show a much smaller range of conformations, which indicates lesser flexibility around the flap tips (Figure 31). The differences in the hinge region may be caused by the presence of the E35-R57 salt bridge in subtype B protease as the formation certainly facilitated by M36, whereas, I36 in subtype C protease does not. The salt bridge generated through the E35-R57 ionic interactions in the flap-hinge region could be an essential requirement for the maintenance of the rigidity of the hinge region and is most likely to contribute to the flexibility of the flap tips (Naicker *et al*., 2013). Third generation protease inhibitors seem to have a strong interaction with the flap residues showing a stronger binding affinity than those with limited interactions. Induced-fit docking does not fully elucidate the mechanism for the reduced drug susceptibility shown by subtype C-SA wild-type protease compared to subtype B protease. However, the differences in the flap-hinge region prove that it may be vital for the interpretation of other dynamic comparisons for the two subtypes. Figure 31 illustrates that darunavir fits into the active site



A. HIV-1 subtype C protease (3U71; red) complexed with the inhibitor darunavir



B. HIV-1 subtype B protease (2P3B; green) complexed with the inhibitor darunavir

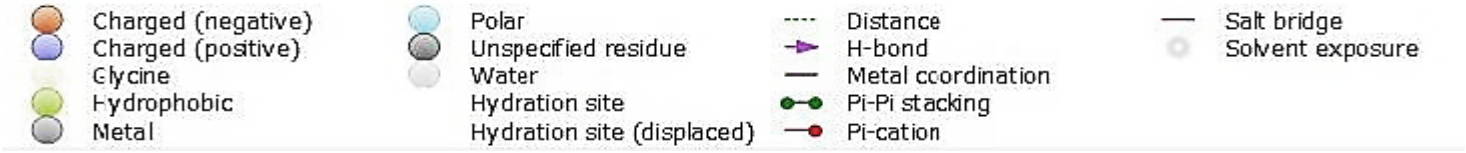


Figure 31: Induced-fit docking 2D and 3D schematic representation of darunavir bound to HIV-1 protease done using Schrödinger. Darunavir was only illustrated as it inhibits the HIV1 protease with the highest potency compared to other inhibitors. The images on the left show a 2-D ligand interaction with highlights of the differences between the different subtypes bound to darunavir. The image on the right shows a 3-D image represented in a ribbon structure form and the interaction between the ligand the protein is shown in dotted lines. Schrödinger Maestro v.11.2 was used to predict the protein-ligand complex geometries.

of HIV-1 protease without excessively protruding outwards. Hence, it binds to the backbone of HIV-1 proteases which makes it challenging for HIV-1 proteases to acquire drug resistance against the protease inhibitor (Ghosh *et al*., 2007). Moreover, darunavir makes contacts with a coordinated water molecule at the conserved S1 and S1' subsites. The oxygen atom in the hydroxyethylamine isostere is a hydrogen bond acceptor for the donor hydrogen on the protonated catalytic aspartic acid (D25 or D25'). The hydrogen bond partners on darunavir are circled and annotated with the contacted residues. Darunavir makes 11 hydrogen bonds with the protease subsites in this instance. Darunavir being a smaller molecule (MW = ~547 g/mol) binds enthalpically because it makes more contacts than it disrupts water molecules bonded to the protein (Kožíšek *et* al., 2014; Mittal *et al*., 2013).

Figure 31 shows that the binding of the HIV-1 protease to the inhibitors has a formation of protein-ligand hydrogen bonds within the region of the hydrophobic enclosure. The hydrophobic enclosed regions are the regions that displace water molecules through inhibitor binding. The majority of residues forming the active site are hydrophobic with the exception of catalytic Asp 25/29 which form hydrogen bonds with the peptide main chain group. These polar interactions with the backbone atoms in the active site enable inhibitors such as darunavir to mimic the conserved hydrogen bond of the natural substrates. The polar and charged residues are also implicated in the binding specificity as they form 25% (polar) and 20% (charged) of the accessible surface area. Reviewing the protein-ligand interaction it can be concluded that the accessible surface area of the protease is hydrophobic, and charged residues of the protein-ligand interaction drives the stability of the protein and responsible for the binding recognition and specificity (Janin and Chothia, 1990; Xu *et al*., 1997; Dey *et al*., 2010). In comparison to other protease inhibitors, darunavir makes enough hydrogen bonds and is small enough to fit in the active site without protruding outwards and displaying too many water molecules. The larger size of saquinavir and ritonavir may be the contributing factors that compromise the binding of these two inhibitors to HIV-1 protease. The binding orientation of the inhibitors to HIV protease revealed that in subtype B there is good interaction with the active site residues and in subtype C the interaction is weaker as it binds to non-conserved amino acids.

Figure 31 shows 2D and 3D schematic representations of protease inhibitors (darunavir) interacting with the amino acid residues of HIV-1 subtype C and B proteases (APPENDIX A.). The induced-fit docking results indicated that recognition sites carry amino acids with electrostatic groups which are mostly hydrophobic in nature and are water soluble. Hence,

hydrophobic/non-polar drugs are able to interact efficiently with the active site of the protease to form a complex which maintain the drugs' properties and mechanisms while allowing it to

be soluble. Figure 31 illustrates that protein-ligand complexes include hydrogen bonds, salt bridges, polar charged groups and van der Waals interactions. The magnitude and combination of these interactions are unique for different inhibitors of protease.

Figure 32 illustrates the statistical measure of the similarity between the two sets of the subtypes through Root-Mean-Square Deviation (RMSD). The proteins is structurally aligned through the method of least-squares; which rotates one protein around its geometric centre to match its orientation to the other protein. Then the protein was translated such that the sum of the distances between homologous pairs of atoms between the proteins was minimised. Once aligned, the RMSD was performed as described in the following equation:

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} d_i^{\,2}} \qquad \text{Equation 10}$$

where N is the number of superimposed atoms in each set of values, and d_i is the Euclidean distance between the i^{th} pair of superimposed values (Westhead *et al*., 2002). The RMSD value was the atomic coordinator of homologous proteins and provided a measure of structural similarities. The C_α atoms from the backbones of each protein was considered for the RMSD values. This is because proteins that do not share exact sequence similarity do not necessarily share the same number of side-chain atoms. However, all amino acids share the same core backbone atoms (NH-CH-CO_2), as long as the RMSD is performed over homologous subsets of the proteins' structures, it will calculate a measure of structural similarity. The RMSD value between the two subtype's backbone C_α atoms was 1.223 Å which means that the two subtypes have almost exactly the same quaternary structures. Figure 32 shows a schematic diagram of protease inhibitors interacting with the amino acid residues of HIV-1 subtype C and B protease. The recognition sites carry electrostatic groups which are predominantly hydrophobic and water soluble through the addition of polar molecules. Hence, hydrophobic/non-polar drugs will be able to interact efficiently with the active site of the protease when they are complexed with external chemical structures which maintain the drugs' properties and mechanisms of inhibition while allowing them to be soluble. Generally, protein-ligand interaction includes hydrogen bonds, salt bridges, polar, charged groups and van der Waals interactions. The magnitude and combination of these interactions are unique for different inhibitors of protease (APPENDIX B).

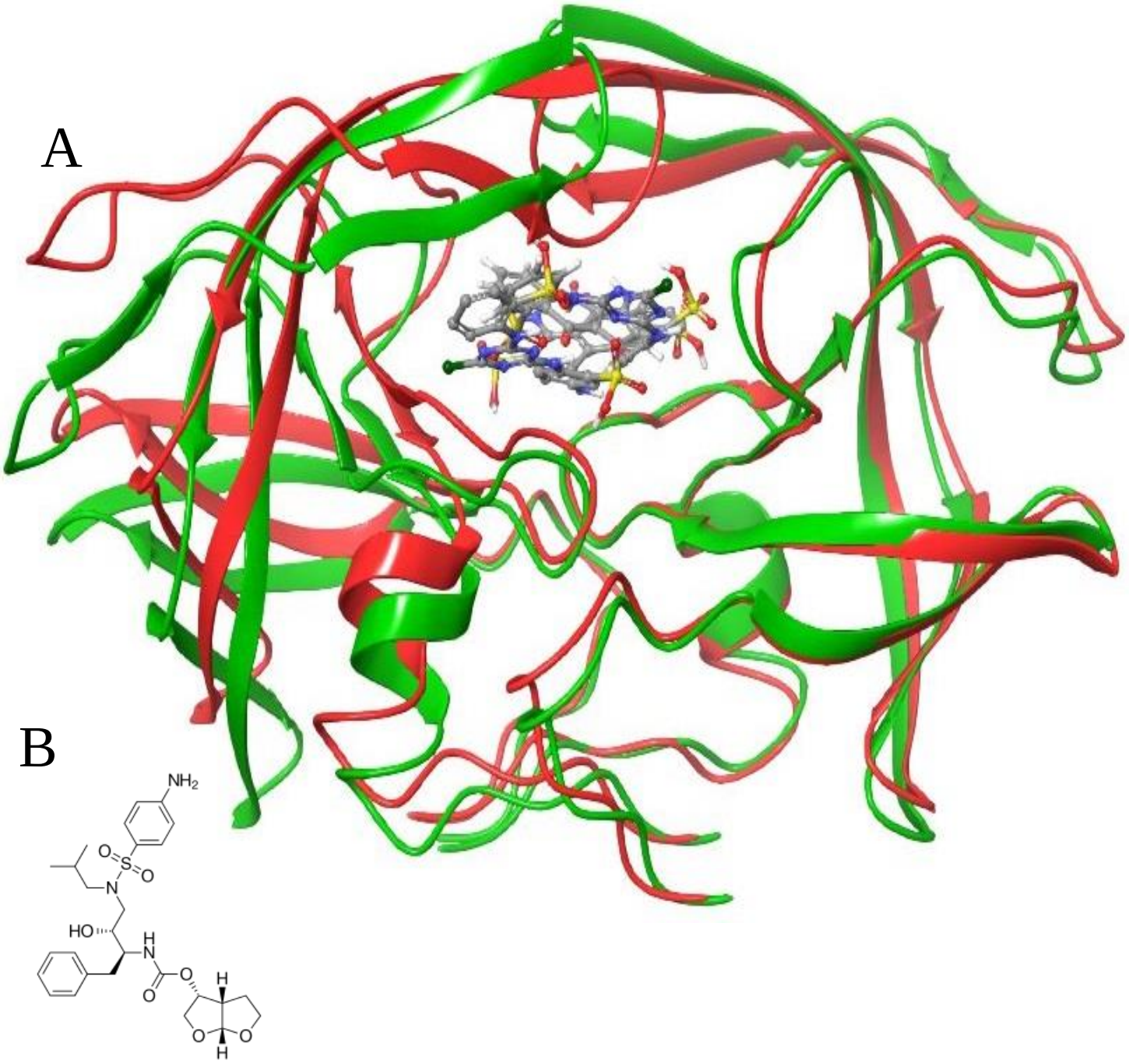


Figure 32: Structural alignment of HIV-1 protease subtypes bound to darunavir. A. The structure of subtype C (3U71; red) was superimposed with subtype B (2P3B; green). The alignment score was 0.080 (smaller is better), RMSD was 1.223 Angstrom. B. The 2D structure of darunavir protease inhibitor. The figures were generated using Schrödinger Maestro v.11.2.

CHAPTER 4

4. CONCLUSION

The current study reported that the protease was capable of cleaving its substrate in the presence of various protease inhibitors. However, due to the lack of research done in the hydrolysis of the fluorogenic substrate, it is difficult to conclude on the significance of these differences in the kinetic parameters determined. Most research in enzyme kinetics of HIV-1 protease is done using a chromogenic substrate (Lys-Ala-Arg-Val-Nle-*p*NO_2Phe-Glu-Ala-Nle-NH_2) which mimic the CA/p2 cleavage site. The specificity of the HIV-1 protease is mainly attributed to the shape complementarity induced upon substrate binding. Thus, different substrates can have a different residue interaction with the protein and that can alter the effects on the specific/catalytic activity of the HIV-1 protease (Prabu-Jeyabalan *et al*., 2000).

Deeper insight into the structural and enzymatic studies of HIV-1 subtype C protease could further improve the knowledge of the enzyme dynamics, drug-enzyme interactions and conformational changes. Other novel protease inhibitors such as GRD 110D in the current study could have inhibitory activity against the protease. However, the binding affinity of GRD 110D is lower than the level required for effective inhibition as compared to the known inhibitors (Figure 29). For future experiments, this study could be improved by studying the antiviral activities of the novel protease inhibitor and that could improve the next-generation inhibitors for protease. The results from the binding mode analysis of FDA-approved drugs could be useful for the design of new potent inhibitors of HIV-1 protease. Computational methods employed in the current study provided an efficient and cost-effective tool for the determination of the energetic interactions between the protease and the ligands, inhibitor stability in the enzyme-binding site and verification of docking results. In this study, molecular docking was performed to fill up the gap between the theoretical and experimental information of the relative inhibitory activity against HIV-1 C-SA wild-type protease, a comparative study with subtype B protease was performed using molecular docking. In conclusion, the natural polymorphisms found in subtype C-SA protease have a significant effect on binding affinity of inhibitors compared to subtype B protease. Some protease inhibitors do not interact only with the active site but also interact with the flaps (flexible loops covering the enzyme active site) of the protease (third generation). An example of the protease inhibitor that is an uncompetitive

inhibitor is darunavir which binds not only to the active site cavity but also occupy a surface pocket formed by one of the flaps. The flap region has a remarkable influence on the binding energy and these influences can have serious implications for the long-term viability of protease inhibition therapy in the clade C-SA virus.

CHAPTER 5

5. REFERENCE

Adler, A. J., Greenfield, N. J. and Fasman, G. D. (1973) Circular dichroism and optical rotator dispersion of proteins and polypeptides. *Methods Enzymol*. **27**, 675-735.

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

Alaeus, A., Lidman, K., Bjorkman, A., Giesecke, J. and Albert, J. (1999) Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A – D. *AIDS* **13**, 901 – 907.

Alexander, C.S., Montessori, V., Wynhoven, B., Dong, W., Chan, K., O ' Shaughnessy, M.V., Mo, T., Piaseczny, M., Montaner, J.S. and Harrigan, P.R. (2002) Prevalence and response to antiretroviral therapy of non-B subtypes of HIV in antiretroviral-naïve individuals in British Columbia. *Antivir. Ther.* ***7,*** 31 – 35.

Allen, W. J., Balius, T. E. and Mukherjee S. (2015) DOCK 6: impact of new features and current docking performance. *J Comput Chem.* **36**, 1132–1156.

Ashorn, P. A., Berger, E. A. and Moss, B. (1990) Human immunodeficiency virus envelope glycoprotein/CD4-mediated fusion of nonprimate cells with human cells. *J. Virol*. **64,** 2149-2156.

Bardsley-Elliot, A. and Plosker, G.L. (2000) Nelfinavir: an update on its use in HIV infection. *Drugs*, **59,** 581-620.

Bertani, G. (2004). Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *Journal of Bacteriology.* **186** (3): 595–600.

Bessong, P. O. (2008) Polymorphisms in HIV-1 subtype C proteases and the potential impact on protease inhibitors. *Science.* **13(**2), 144–151.

Bold, G., Fassler, A., Capraro, H. G., Cozens, R., Klimkait, T., Lazdins, J., Mestan, J., Poncioni, B., Rosel, J., Stover, D., Tintelnot-Blomley, M., Acemoglu, F., Beck, W., Boss, E., Eschbach, M., Hurlimann, T., Masso, E., Roussel, S., Ucci-Stoll, K., Wyss, D. and Lang, M. (1998) New aza-dipeptide analogues as potent and orally absorbed HIV-1

protease inhibitors: candidates for clinical development. *J. Med. Chem.* **41**, 3387-3401.

Briggs, H. and Haldane, J. B. S. (1925) A note on the kinetics of enzyme action. *Biochem. J.* **19**, 338−339.

Brik, A. and Wong, C. (2002) HIV-1 protease: mechanism and drug discovery. *Organic & Biomolecular Chemistry*, **1**. 5-14.

Buonaguro, L., Tornesello, M. L. and Buonaguro, F. M. (2007) MINIREVIEW Human Immunodeficiency Virus Type 1 Subtype Distribution in the Worldwide Epidemic: Pathogenetic and Therapeutic Implications. *Science.* **81**(19), 10209–10219.

Carmel, A. and Yaron, Y. (1978) An intramolecularly quenched fluorescent tripeptide as a fluorogenic substrate of angiotensin-i-converting enzyme and of bacterial dipeptidyl carboxypeptidase. *European Journal of Biochemistry*, **87**, 265–273.

Chen, Z., Luckay, A., Sodora, D. L., Telfer, P., Reed, P., Gettie, A., Kanu, J. M., Sadek, R. F., Yee, J. and Ho, D. D. (1997) Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *J. Virol.* **71**, 3953-60.

Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M. A., Santos-Ferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C. and Rouzioux, C. (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science.* **233**, 343-6.

Clemente, J. C., Moose, R. E., Hemrajani, R., Whitford, L. R., Govindasamy, L., Reutzel, R., McKenna, R., Agbandje-McKenna, M., Goodenow, M. M. and Dunn, B. M. (2004) Comparing the accumulation of active- and nonactive-site mutations in the HIV-1 protease. *Biochemistry* **43**, 12141-51.

Coffin, J. (1984) Structure of the retroviral genome. In Weiss, R., Teich, N., Varmus, H. and Coffin, J (ed.), RNA tumor viruses. *Cold Spring Harbor Laboratory*, *Cold Spring Harbor, N.Y.* 261-368.

Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., and Vogt, P. (19861) What to call the AIDS virus? *Nature*, **321:** 10.

Coffin, J. M. (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* **267,** 483-489.

Coman, R. M., Robbins, A. H., Goodenow, M. M., Dunn, B. M. and McKenna, R. (2008) High-resolution structure of unbound human immunodeficiency virus 1 subtype C protease: implications of flap dynamics and drug resistance. *Acta. Crystallogr. D. Biol. Crystallogr.* **D64**, 754-63.

Darke, P. L., Leu, C. T., Davis, L. J., Heimbach, J. C., Diehl, R. E., Hill, W. S., Dixon, R. A. and Sigal, I. S. (1989) Human immunodeficiency virus protease. Bacterial expression and characterization of the purified aspartic protease. *J. Biol. Chem.* **264**, 2307-12.

Dash, C. and Rao, M. (2001) Interactions of a novel inhibitor from an extremophilic *Bacillus sp.* with HIV-1 protease. *J. Biol. Chem.* **276**, 2487-2493.

Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W. and Rosenberg M. (1987) Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the *gag* precursor. *Proc. Natl. Acad. Sci. USA.* **84**, 8903-8906.

De Lano, W. L. (2002) The PyMOL Molecular Graphics System. *DeLano Scientific, LLC, San Carlos, CA.*

Del Amo, J., Petruckevitch, A., Phillips, A., Johnson, A.M., Stephenson, J., Desmond, N., Hanscheid, T., Low, N., Newell, A. and Obasi, A. (1998) Disease progression and survival in HIV-1-infected Africans in London. *AIDS **12***, 1203 – 1209.

Dey, S., Pal, A., Chakrabarti, P. and Janin, J. (2010) The subunit interfaces of weakly associated homodimeric proteins. *J Mol Biol,* **398,** 146-60.

Dixon, M. F. and Webb, E. C. (1958) *Enzymes.*, London: Longmans, Green and Co.

de Oliveira, T., Engelbrecht, S., van Rensburg, E. J., Gordon, M., Bishop, K., zur Megede, J., Barnett, S. W. and Cassol, S. (2003) Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? *J. Virol.* **77**, 9422-9430.

Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L. and Lamarre, D. (1996) Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J. Virol*, **70**, 3763-3769.

Doyon, L., Tremblay, S., Bourgon, L., Wardrop, E. and Cordingley, M. G. (2005) Selection and characterization of HIV-1 showing reduced susceptibility to the nonpeptidic protease inhibitor tipranavir. *Antiviral. Res*. **68**, 27-35.

Dube, M. P., Qian, D., Edmondson-Melancon, H., Sattler, F. R., Goodwin, D., Martinez, C., Williams, V., Johnson, D. and Buchanan, T. A. (2002) Prospective, intensive study of metabolic changes associated with 48 weeks of amprenavir-based antiretroviral therapy. *Clin. Infect. Dis.35, 475-481.*

Engelman, A. and Cherepanov, P. (2012) The structural biology of HIV - 1 : mechanistic and therapeutic insights. *Nature Publishing Group,* **10**(4), pp.279–290.

Esparza, J. and Bhamarapravati, N. (2000) Accelerating the development and future availability of HIV-1 vaccines: why, when, where and how? *Lancet*. **355**, 2061-2066.

Fauci, A. S. (2003) HIV SPECIAL HIV and AIDS: 20 years of science. *Adv. Virus. Res.* **9**(7), 839–843.

Fauci, A. S. and Lane, H. C. (2005) Human Immunodeficiency Virus Disease: AIDS and Related Disorders. Pages 1076-1139 in Kasper, D.L., Braunwald, E., Fauci, A.S., Hauser, S.L., Longo, D.L., Jameson, J.L., eds. *Harrison's Principles of Internal Medicine. 16th Edition.* New York: McGraw-Hill Professional.

Fitzgerald, P. M. and Springer, J. P. (1991) Structure and function of retroviral proteases. *Annu Rev Biophys Biophys Chem,* **20,** 299-320.

Fitzgerald, P. M. D. (1993) HIV protease-ligand complexes. *Curr. Opin. Struct. Biol.* **3**, 868-874.

Fleury, H., Recordon-Pinson, P., Caumont, A., Faure, M., Roques, P., Plantier, J. C., Couturier, E., Dormont, D., Masquelier, B. and Simon, F. (2003) HIV type 1 diversity in France, 1999-2001: molecular characterization of non-B HIV type 1 subtypes and potential impact on susceptibility to antiretroviral drugs. *AIDS. Res. Hum. Retroviruses.* **19**, 41-7.

Floridia, M., Bucciardini, R., Fragola, V., Galluzzo, C. M., Giannini, G., Pirillo, M. F., Amici, R., Andreotti, M., Ricciardulli, D., Tomino, C. and Vella, S. (2004) Risk factors and occurrence of rash in HIV positive patients not receiving nonnucleoside reverse transcriptase inhibitor: data from a randomized study evaluating use of protease inhibitors in nucleoside-experienced patients with very low CD4 levels (<50 cells/microL). *HIV Med.* **5**, 1-10.

Freed, E. O. (2001) HIV-1 replication. *Somat Cell Mol Genet,* **26,** 13-33.

Freyer, M. W., and Lewis, E. A. (2008) Isothermal titration calorimetry: experimental design,

data analysis, and probing macromolecule/ ligand binding and kinetic interactions. *Methods in Cell Biology*, **84:** 79-113.

Friesner, R. A., Banks, J. L. and Murphy, R. B. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem.* **47**, 1739–1749.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, 571-607. Humana Press.

Ghosh, A. K., Dawson, Z. L. and Mitsuya, H. (2007) Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drugresistant HIV. *Bioorg. Med. Chem.* **15,** 7576–7580.

Gonzales, M. J., Machekano, R. N. and Shafer, R. W. (2001) HIV-1 reverse transcriptase and protease subtypes: Classification, amino acid mutation patterns, and prevalence in a Northern California clinic-based population. *J. Infect. Dis.* **184**, 998-1006.

Gordon, M., De Oliveira, T., Bishop, K., Coovadia, H. M., Madurai, L., Engelbrecht, S., Jansen van Rensburg, E., Mosam, A. and Cassol, S. (2003) Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies. *J. Virol.* **77**, 2587-2599.

Greene, W. C. (1993) AIDS and the immune system. *Sci Am,* **269,** 98-105.

Greenfield, N. J. (1996) Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal. Biochem.* **235**, 1-10.

Greenfield, N.J. 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 1: 2876-2890.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723.

Hansen, J., Billich, S., Schulze, T., Sukrow, S. and Moelling, K. (1988) Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. *EMBO. J.* **7**, 1785-91.

Hemelaar, J., Gouws, E., Ghys, P. D. and Osmanov, S. (2011) Global trends in molecular

epidemiology of HIV-1 during 2000–2007. *AIDS* **25**, 679–689.

Henderson, L. E., Krutzsch, H. C. and Oroszlan, S. (1983) Myristyl amino-terminal acylation of murine retrovirus proteins: an unusual post-translational proteins modification. *Proc. Natl. Acad. Sci. U S A* **80**, 339-43.

Hong, L., Zhang X., Foundling, S., Hartsuck, A. N. and Tang, J. (1997) Structure of a g48h mutant of HIV-1 protease explains how glycine-48 replacements produce mutants resistant to inhibitor drugs, *FEBS J.* **420**, 11-16.

Honig, B. H., Hubbell, W. L. and Flewelling, R. F. (1986) Electrostatic interactions in membranes and proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 163-93.

Hennessey, J. P. Jr. and Johnson, W. C. Jr. (1981) Information content in the circular dichroism of proteins. *Biochemistry*. **20,** 1085-1094.

Hyland, L. J., Tomaszek, T. A., Jr., Roberts, G. D., Carr, S. A., Magaard, V. W., Bryan, H. L., Fakhoury, S. A., Moore, M. L., Minnich, M. D. and Culp, J. S. (1991) Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by 18O isotope exchange. *Biochemistry* **30**, 8441-53.

Ido, E., Han, H. P., Kezdy, F. J. and Tang, J. (1991) Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S. *J. Biol. Chem.* **266**, 24359-66.

Ishima, R., Torchia, D. A. and Louis, J. M. (2007) Mutation and Structural Studies Aimed at Characterizing the Monomer of HIV-1 Protease and its Precursor. *J. Biol. Chem.* **282**, 17190-17199.

Jain, A. N. (2003) Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem*. **46**, 499–511.

James, M. N. and Sielecki, A. R. (1983) Structure and refinement of penicillopepsin at 1.8 A resolution. *J. Mol. Biol.* **163**, 299-361.

Jamin, J. and Chothia, C. (1990) The structure of protein-protein recognition sites. *J. Biol. Chem.* **265**, 16027-16030.

Johnson, K. A., and Goody, R. S. (2011) The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry,* **50:** 8264-826.

Jones, G., Willett, P., Glen, R. C., Leach, A. R. and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol.* **267**, 727–748.

Kantor, R. and Katzenstein, D. (2003) Polymorphism in HIV-1 Non-subtype B Protease and Reverse Transcriptase and its Potential Impact on Drug Susceptibility and Drug Resistance Evolution. *AIDS Rev.* **5**, 25-35.

Katritzky, A. R. and Narindoshvili, T. (2009) Fluorescent amino acids: advances in proein-extrinsic flourophores. *Org. Biomol. Chem*., **7**, 627-634.

Kear, J. L., Blackburn, M. E., Veloro, A. M., Dunn, B. M. & Fanucci, G. E. (2009). Subtype polymorphisms among HIV-1 protease variants confer altered flap conformations and flexibility. *J. Am. Chem. Soc.* **131**, 14650-1.

Keele, B. F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M. L., Bibollet-Ruche, F., Chen, Y., Wain, L. V., Liegeois, F., Loul, S., Ngole, E. M., Bienvenue, Y., Delaporte, E., Brookfield, J. F. Y., Sharp, P. M., Shaw, G. M., Peeters, M. and Hahn, B. H. (2006) Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, **313:** 523-526.

Kelly, S. M., and Price, N. C. (2000) The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science*, **1**: 349-384.

Kilby, J. M., Sfakianos, G., Gizzi, N., Siemon-Hryczyk, P., Ehrensing, E., Oo, C., Buss, N. and Saag, M. S. (2000) Safety and pharmacokinetics of once-daily regimens of soft-gel capsule saquinavir plus minidose ritonavir in human immunodeficiency virus-negative adults. *Antimicrob. Agents Chemother.* **44**, 2672-2678.

Kim, E., Winters, M. A., Kagan, R. M. and Merigan, T. C. (2001) Functional correlates of insertion mutations in the protease gene of human immunodeficiency virus type 1 isolates from patients. *J. Virol.* **75**, 11227-11233.

King, J. R., and Acosta, E. P. (2006) Tipranavir: a novel nonpeptidic protease inhibitor of HIV. *Clin. Pharmacokinet. 45*, 665-682.

Klabe, R. M., Bacheler, L. T., Ala, P. J., Erickson-Viitanen, S. and Meek, J. L. (1998) Resistance to HIV protease inhibitors: a comparison of enzyme inhibition and antiviral potency. *Biochemistry* **37**, 8735-8742.

Koh, Y., Nakata, H., Maeda, K., Ogata, H., Bilcer, G., Devasamudram, T., Kincaid, J. F.,

Boross, P., Wang, Y. F., Tie, Y., Volarath, P., Gaddis, L., Harrison, R. W., Weber, I. T., Ghosh, A. K. and Mitsuya, H. (2003) Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro, *Antimicrob Agents Chemother.* **47**, 3123-3129. 171.

Kovalevsky, A. Y., Ghosh, A. K. and Weber, I. T. (2008) Solution kinetics measurements suggest HIV-1 protease has two binding sites for darunavir and amprenavir, *J Med Chem.* **51**, 6599-6603.

Kozisek, M., Saskova, K. G., Rezacova, P., Brynda, J., van Maarseveen, N. M., De Jong, D., Boucher, C. A., Kagan, R. M., Nijhuis, M. and Konvalinka, J. (2008) Ninety-nine is not enough: molecular characterization of inhibitor-resistant human immunodeficiency virus type 1 protease mutants with insertions in the flap region. *J. virol.* **82**, 5869-78.

Kramer, R. A., Schaber, M. D., Skalka, A. M., Ganguly, K., Wong-Staal, F. and Reddy, E. P. (1986) HTLV-III gag protein is processed in yeast cells by the virus pol-protease. *Science,* **231,** 1580-4.

Laemmli, U. K. (1970) Cleavage of structural protein during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Lakowicz, J. (1983) *Principles of fluorescence spectroscopy*, New York, Plenum press.

Lakowicz, J. R. (1999) Principles of fluorescence spectroscopy. Plenum Press, New York, USA.

Le Tiec, C., Barrail, A., Goujard, C. and Taburet, A. M. (2005) Clinical pharmacokinetics and summary of efficacy and tolerability of atazanavir. *Clin. Pharmacokinet.* **44,** 1035-1050.

Louis, J. M., Wondrak, E. M., Kimmel, A. R., Wingfield, P. T. and Nashed, N. T. (1999) Proteolytic processing of HIV-1 protease precursor: Kinetics and mechanism. *J. Biol. Chem.* **274**, 23437-23442.

Luban, J., Lee, C. and Goff, S. P. (1993) Effect of linker insertion mutations in the human immunodeficiency virus type 1 *gag* gene on activation of viral protease expressed in bacteria. *J. Virol*. **67**, 3630-3634.

Maartens, G., Celum, C. and Lewin, S. R. (2014) HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet.* **384**, 258-271.

Macias, J., Orihuela, F., Rivero, A., Viciana, P., Marquez, M., Portilla, J., Rios, M. J., Munoz, L., Pasquau, J., Castano, M. A., Abdel-Kader, L. and Pineda, J. A. (2009) Hepatic safety of tipranavir plus ritonavir (TPV/r)-based antiretroviral combinations: effect of hepatitis virus co-infection and preexisting fibrosis. *J. Antimicrob. Chemother.* **63,** 178-183.

Manning, M. C., Illangasekare, M. and Woody, R. W. (1988) Circular dichroism studies of distorted alpha-helices, twisted beta-sheet, and beta turns. *Biophys. Chem.* **31**, 77-86.

Marx, J. L. (1982) New disease baffles medical community. *Science*, **217:** 618-621.

McGann, M. R., Almond, H. R., Nicholls, A., Grant, J. A. and Brown, F. K. (2003) Gaussian docking functions. *Biopolymers.* **68**, 76–90.

Mervis, R. J., Ahmad, N., Lillehoj, E. P., Raum, M. G., Salazar, F. H., Chan, H. W. and Venkatesan, S. (1988) The gag gene products of human immunodeficiency virus type 1: alignment within the gag open reading frame, identification of posttranslational modifications, and evidence for alternative gag precursors. *J. Virol.* **62**, 3993-4002.

Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. and Wlodawer, A. (1989) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science.* **246**, 1149-52.

Modi, S., Madan, A., Behere, D. V. and Mitra, S. (1993) Circular dichroism studies of acid alkaline transition in hemeproteins. *Proc. Indian Acad. Sci.* (Chem. Sci.). **105**, 167-172.

Mosebi, S., Morris, L., Dirr, H. W. and Sayed, Y. (2008) Active Site Mutations in the South African HIV-1 subtype C protease impact significantly on clinical inhibitor binding: a kinetic and thermodynamic study. *J. Virol*. **82**, 11476-11479.

Mpye, K. L. (2010) Structural and functional effects of an insertion I36TT in the South African HIV-1 subtype C protease. *Biochemistry*, vol. MSc. Johannesburg: University of the Witwatersrand.

Mugnaini, C., Petricci, E., Corelli, F. and Botta, M. (2005) Combinatorial chemistry as a tool for targeting different stages of the replicative HIV-1 cycle. *Comb Chem High Throughput Screen,* **8,** 387-401.

Muzammil, S., Ross, P. and Freire, E. (2003) A major role for a set on non-active site mutations in the development of HIV-1 protease drug resistance. *Biochemistry* **42**, 631-638.

Naicker, P., Fanucchi, S., Achilonu, I.A., Fernandes, M.A., Dirr, H.W. and Sayed, Y. (2012) Crystal structure analysis of South African wild type HIV-1 subtype C apo protease. *Escherichia coli*. DOI: 10.2210/pdb3u71/pdb.

Naicker, P. and Sayed, Y. (2014) Non-B HIV–1 subtypes in sub-Saharan Africa: impact of subtype on protease inhibitor efficacy. *Biol Chem.* **395**, 1151–1161.

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**, 615-20.

Nijhuis, M., Schuurman, R., de Jong, D., Erickson, J., Gustchina, E., Albert, J., Schipper, P., Gulnik, S. and Boucher, C.A. (1999) Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS.* **13**, 2349–2359. 96

Niu, B., Lu, L., Liu, L., Gu, T. H., Feng, K., Lu, W. and Cai, Y. (2008) HIV-1 protease cleavage site prediction based on amino acid property. *J. Comp. Chem.* **30**, 33-39.

Noel, A. F., Bilsel, O., Kundu, A., Wu, Y., Zitzewitz, J. A. and Matthews, C. R. (2009) The folding free-energy surface of HIV-1 protease: insights into the thermodynamic basis for resistance to inhibitors. *J Mol Biol*, **387,** 1002-16.

Núñez, S., Venhorst, J. and Kruse, C. G. (2012) Target–drug interactions: first principles and their application to drug discovery. *Drug discovery today*, **17:** 10-22.

Ohtaka, H., Schon, A. 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 Sci.* **11**, 1908-1916.

Olomola, T.O., Klein, R., Mautsa, N., Sayed, Y. and Kaye, P.T. (2013) Synthesis and evaluation of coumarin derivatives as potential dual-action HIV-1 protease and reverse transcriptase inhibitors. Bioorg. Med. Chem. **2**, 1964 – 1971.

Österberg, F., Morris, G. M., Sanner, M. F., Olson, A. J. and Goodsell, D. S. (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins*. **46**, 34–40.

Pain, R. (2004) Determining the CD spectrum of a protein. In: Current protocols in Protein Science. Vol. 1, pp7.6.1-7.7.20. John Wiley and Sons, Inc., New York, USA.

Patick, A. K., Boritzki, T. J. and Bloom, L. A. (1997) Activities of human immunodeficiency virus type 1 (HIV-1) protease inhibitor nelfinavir mesylate in combination with reverse transcriptase and protease inhibitors against acute HIV-1 infection in vitro. *Antimicrob. Agents Chemother.* **41**, 2159-2164.

Perez, C., Pastor, M., Ortiz, A. R. and Gago, F. (1998) Comparative Binding energy analysis of HIV -1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptorbased drug design, *J. Med. Chem*. **41**, 836-852.

Perry, C. M. and Noble, S. (1998) Saquinavir soft-gel capsule formulation. A review of its use in patients with HIV infection. *Drugs* **55**, 461-486.

Perryman, A. L., Lin, J. H. and McCammon, J. A. (2006) HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and 97 a potential new target site for drugs. *Biopolymers* **82**, 272-284.

Piliero, P. J. (2002) Atazanavir: a novel HIV-1 protease inhibitor. *Expert Opin Investig Drug*. **11**, 1295-1301.

Pillay, D., Walker, A. S., Gibb, D. M., de Rossi, A., Kaye, S., Ait-Khaled, M., Munoz-Fernandez, M. and Babiker, A. (2002) Impact of human immunodeficiency virus type 1 subtypes on virologic response and emergence of drug resistance among children in the Paediatric European Network for Treatment of AIDS (PENTA) 5 trial. *J Infect. Dis*. **186**, 617 – 625.

Plantier, J. C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemee, V., Damond, F., Robertson, D. L. and Simon, F. (2009) A new human immunodeficiency virus derived from gorillas. *Nat. Med.* **15**, 871-872.

Polgár, L., Szeltner, Z. and Boros, I. (1994) Substrate-dependent mechanisms in the catalysis of human immunodeficiency virus protease. *Biochemistry.* **33**, 9351-9357.

Poorolajal, J., Hooshmand, E., Mahjub, H., Esmailnasabd, N. and Jenabie, E (2016) Survival rate of AIDS disease and mortality in HIV-infected patients: a meta-analysis. *Public health*, 139, 3-12.

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. *J Mol Biol*, **301,** 1207-20.

Prasad, J., Para, K., Tummino, P., Ferguson, D., McQuade, T., Lumney, E., Rapundalo, S., Batley, B. and Hingorani, G. (1995) Nonpeptidic Potent HIV-1 Protease Inhibitors: (4-Hydroxy-6-phenyl-2-oxo-2H-pyran-3-yl) thiomethanes That Span P1-P2' Subsites in a Unique Mode of Binding.J. *Med. Chem*., **38**(6).898-905.

Rarey, M., Kramer, B., Lengauer, T. and Klebe, G. (1996) A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*. 261:470–489.

Richards, A. D., Roberts. R., Dunn, B. M., Graves, M. C. and Kay J. (1989) Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic proteinases. *FEBS Lett.* **247**, 113-117.

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, 98

F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leimer, T., McCutchan, F., Osmanov, S., Peeters, M., Pieiazek, D., Salminen, M., Sharp, P. M., Wolinsky, S. and Korber, B. (2000) HIV-1 nomenclature proposal. *Science.* **288**, 55-56.

Rick, S. W., Erickson, J. W. & Burt, S. K. (1998). Reaction path and free energy calculations of the transition between alternate conformations of HIV-1 protease. *Proteins* **32**, 7-16.

Rose, R. E., Gong, Y.F., Greytok, J.A., Bechtold, C.M., Terry, B.J., Robinson, B.S., Alam, M., Colonno, R.J. and Lin, P.F. (1996) Human immunodeficiency virus type 1 viral background plays a major role in development of resistance to protease inhibitors. *Proc. Natl. Acad. Sci. USA.* **93**, 1648–1653.

Royer, C. A. (1995) in Methods in Molecular Biology, Vol 40: Protein Stability and Folding: Theory and Practice (Shirley, B. A., ed) Pp. 65-89, Humana Press Inc., Totowa, NJ.

Safai, B., Groopman, J. E., Popovic, M., Schurpbach, J., Sarangadharan, M. G., Arnett, K., Sliski, A. and Gallo, R. C. (1984) Seroepidemiological studies of human T-lymphotropic retrovirus type III in acquired immunodeficiency syndrome. *The Lancet*, **323**. 1438-1440.

Schägger, H. (2006) Tricine – SDS-PAGE. *Science.* **1**(1), 16–23.

Schapira, M., Abagyan, R. and Totrov, M. (2003) Nuclear hormone receptor targeted virtual screening. *J Med Chem.* **46**, 3045–3059.

Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain.

Biochemical and biophysical research communications, **27:** 157-162.

Shafer, R. W. and Schapiro, J.M. (2008). HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev*. **10**, 67 – 84.

Sham, H. L., Kempf, D. J., Molla, A., Marsh, K. C., Kumar, G. N., Chen, C. M., Kati, W., Stewart, K., Lal, R., Hsu, A., Betebenner, D., Korneyeva, M., Vasavanonda, S., McDonald, E., Saldivar, A., Wideburg, N., Chen, X., Niu, P., Park, C., Jayanti, V., Grabowski, B., Granneman, G. R., Sun, E., Japour, A. J., Leonard, J. M., Plattner, J. J. and Norbeck, D. W. (1998) ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrob. Agents Chemother.* **42**, 3218-3224.

Sharp, P. M., Bailes, E., Chaudhuri, R. R., Rodenburg, C. M., Santiago, M. O. and Hahn, B. H. (2001) The origins of acquired immune deficiency syndrome viruses: where and when?. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **356:** 867-876. 99

Sharp, P. M., and Hahn, B. H. (2011). Origins of HIV and the AIDS pandemic. Cold Spring Harbor perspectives in medicine. **1**, 1-23.

Shirley, B. A. (1995) Urea and guanidine hydrochloride denaturation curves. *Methods Mol. Biol.* **40**, 177-190.

Smith, R., Brereton, I. M., Chai, R. Y., and Kent, S. B. H. (1996) Ionization states of the catalytic residues in HIV-1 protease. Nature Structural Biology, 3, 946–950.

Stano, N. M and Petel, S. S. (2004) T7 lysozyme represses T7 RNA polymerase transcription by destabilizing the open complex during initiation. *J. Biol. Chem*. **279**, 16,136-16,143.

Studier, F. W., and Moffatt, B. A. (1986) *J. Mol. Biol*. 189, 113-130

Tavassoli, A. (2011) Targeting the protein-protein interactions of the HIV life-cycle. *Chem Soc Rev.* **40**, 1337-46.

Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E. and Hammer, S. M. (2008) The challenge of HIV-1 subtype diversity. *N. Engl. J. Med.* ***358,*** 1590–1602.

Todd, M. J., Semo, N. and Freire E. (1998) The structural stability of the HIV-1 protease. *J. Mol. Biol*. **283**, 475–488.

Todd, M. and Freire, E. (1999) The effect of inhibitor binding on the structural stability and

cooperativity of the HIV-1 protease. *Proteins* **36**, 147-156.

Toh, H., Ono, M. and Miyata, T. (1985) Retroviral gag and DNA endonuclease coding sequences in IgE-binding factor gene. *Nature* **318**, 388-9.

Torres, H. A. and Arduino, R. C. (2007) Fosamprenavir calcium plus ritonavir for HIV infection. *Expert Rev. Anti Infect. Ther.* **5,** 349-363.

Tozser, J., Yin, F. H., Cheng, Y. S., Bagossi, P., Weber, I. T., Harrison, R. W. & Oroszlan, S. (1997). Activity of tethered human immunodeficiency virus 1 protease containing mutations in the flap region of one subunit. *Eur. J. Biochem.* **244**, 235-41.

Trylska, J., Tozzini, V., Chang, C. A. and McCammon, J. A. (2007) HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophys. J.* **92**, 4179-4187.

Turner, B. G. and Summers, M. F. (1999) Structural biology of HIV. *J. Mol. Biol.* **285**, 1-32.

UNAIDS (2017) Global report: UNAIDS report on the global epidemic 2017. *Joint United Nations Programme on HIV/AIDS (UNAIDS)*

Velazquez-Campoy, A., Todd, M. J., Vega, S. and Freire, E. (2001) Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc. Natl. Acad. Sci.* **98**, 6062-6067.

Velazquez-Campoy, A., Vega, S., Fleming, E., Bacha, U., Sayed, Y., Dirr, H. W. and Freire, E. (2003) Protease inhibition in Africa subtypes of HIV-1. *AIDS Rev.* **5**, 165-171.

Venkatachalam, C. M., Jiang, X., Oldfield, T. and Waldman, M. (2003) LigandFit: a novel 100 method for the shape-directed rapid docking of ligands to protein active sites. *J Mol Graph Model*. **21**, 289–307.

Vega, S., Kang, L. W., Velazquez-Campoy, A., Kiso, Y., Amzel, L. M. and Freire, E. (2004) A structural and thermodynamic escape mechanism from a drug resistant mutation of the HIV-1 protease. *Proteins* **55**, 594-602.

Vogt, V. M. (1996) Proteolysis and particle maturation. *Curr. Top. Microbiol. Immunol.* **214**, 95-132.

Walker, P. R., Pybus, O. G., Rambaut, A. & Holmes, E. C. 2005. Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic

growth. *Infect Genet Evol*, **5,** 199-208.

Weber, I. T. (1990) Comparison of the crystal structures and intersubunit interactions of human immunodeficiency and Rous Sarcoma virus proteases. *J. Biol. Chem.* **265**, 10492-10496.

Weiss, A., Hollander, H. and Stobo, J. (1985) Acquired immunodeficiency syndrome: epidemiology, virology, and immunology. *Annu. Rev. Med.* **36**, 545-62.

Westhead, D. R., Parish, J. H. and Twyman, R. M. (2002) Instant Notes in Bioinformatics. Oxford: BIOS Scienti_c Publishers Ltd.

Williston, S. and Brandts, J. F. (1989) Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. *Science.* **137**, 131–137.

Wiseman, T., Williston, S., Brandts, J. F. and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131-37.

Witvrouw, M., Pannecouque, C., Switzer, W. M., Folks, T. M., de Clercq, E. and Heneine, W. (2004) Susceptibility of HIV-2, SIV and SHIV to various anti-HIV-1 compounds: implications for treatment and postexposure prophylaxis. *Antivir. Ther.* **9,** 57-65.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. and Kent, S. B. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, **245:** 616-621.

Wlodawer, A. and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: a major success of 101 structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249-284.

Woody, R. W. (1995). Circular dichroism. *Methods Enzymol*. **246**, 34-71.

World Health Organization. (2016) Global HIV/AIDS statistics. http://www.who.org.

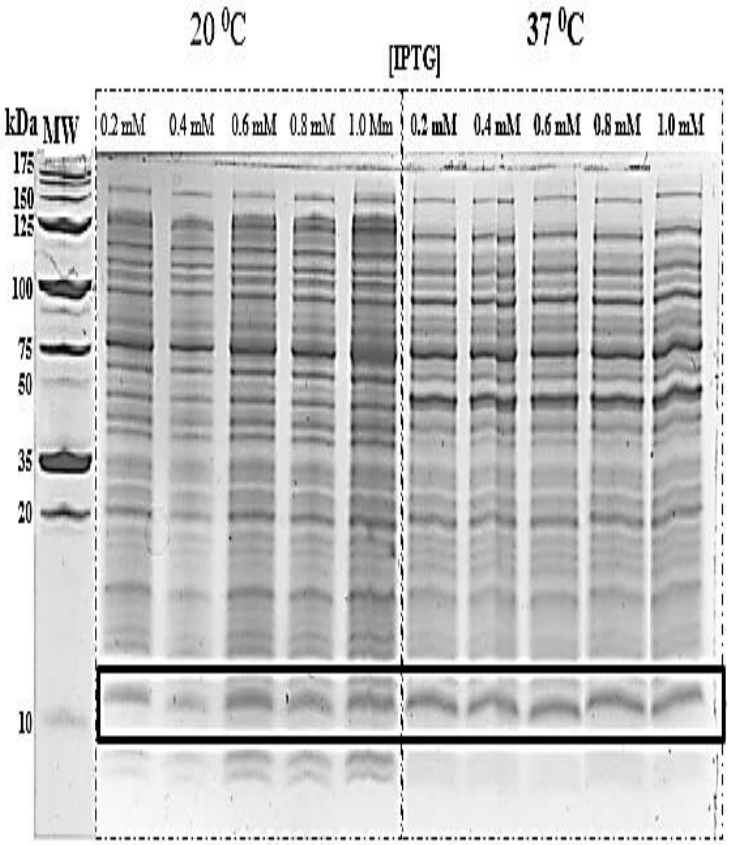
Xu, D., Lin, S. L., and Nussinov, R. (1997) Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. *J. Mol. Biol.* **265**, 68-84.

Xu, L. and Desai, M. C. (2009) Pharmacokinetic enhancers for HIV drugs. *Curr. Opin. Investig. Drugs*, **10,** 775-786.

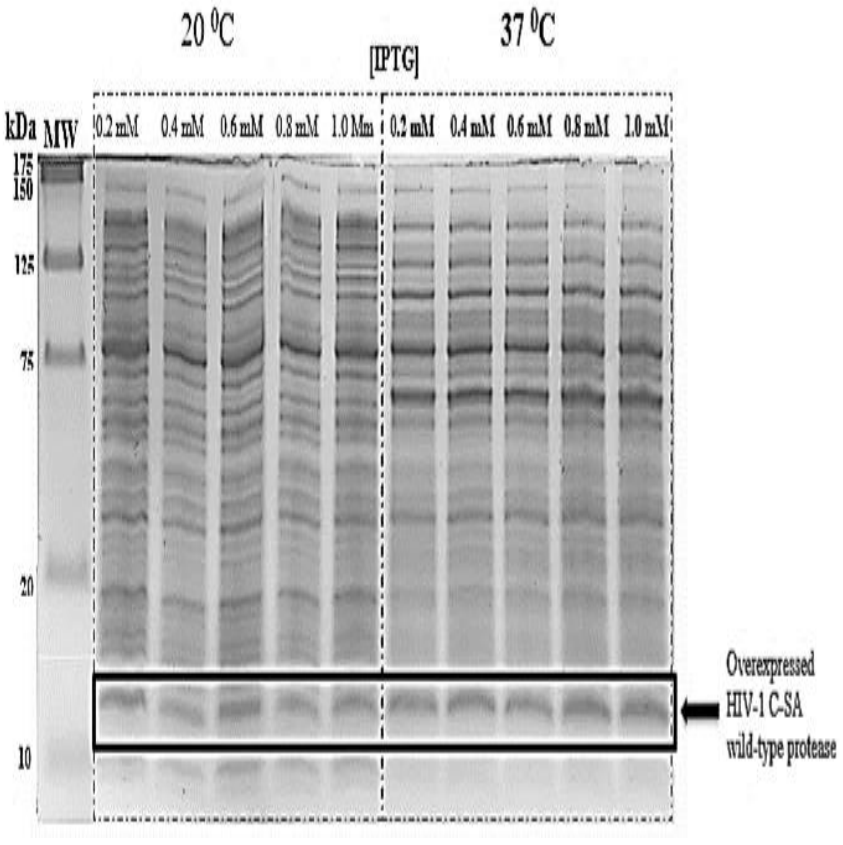
APPENDIX

Appendix A. Supplementary material for section 3.3

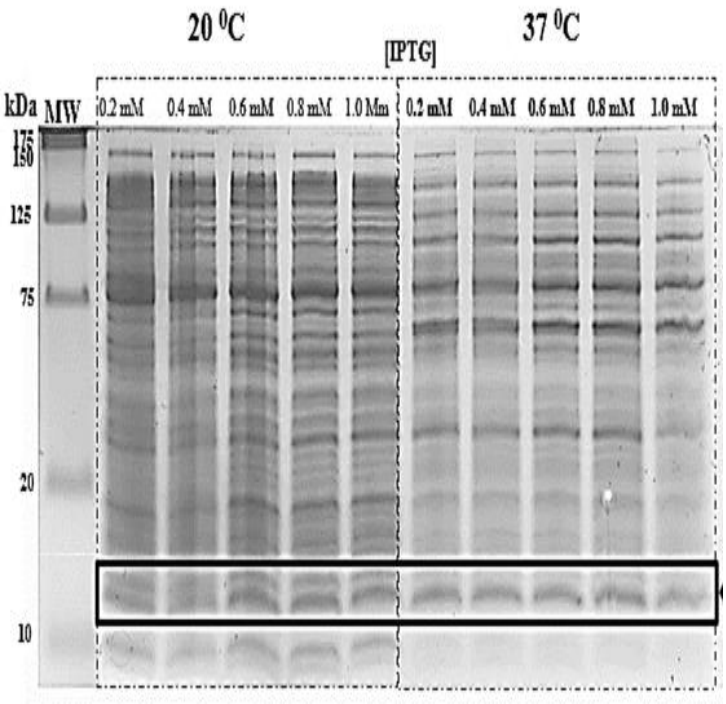
Supplementary Figures. Expression trial of wild-type HIV-1 C-SA protease in *E. coli* BL21 (DE3) pLysS cells at 20 ^{o}C and 37 ^{o}C using varying IPTG concentrations (0.2, 0.4, 0.6, 0.8 and 1.0 mM).



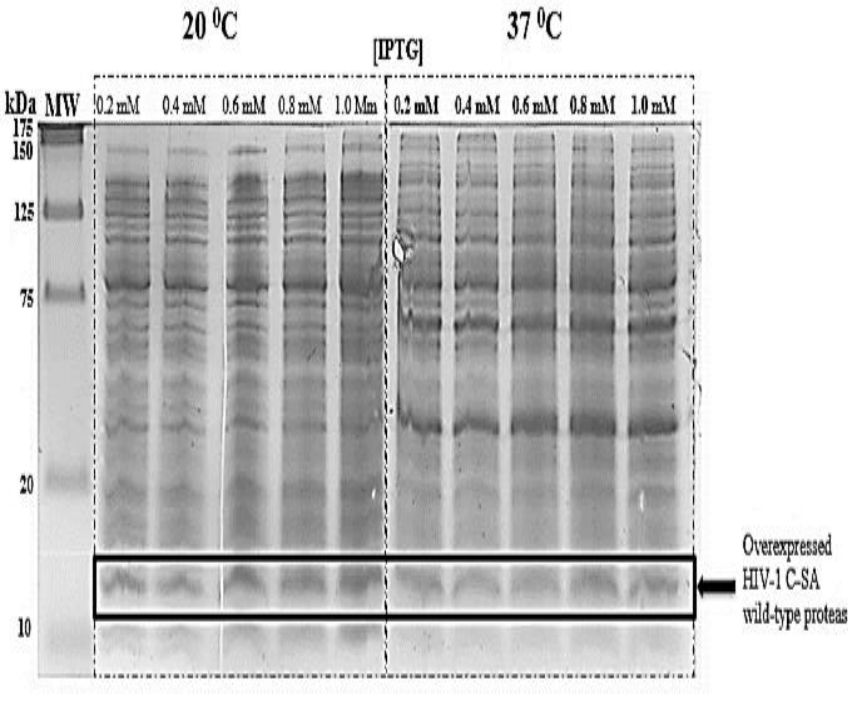
1st hour after induction



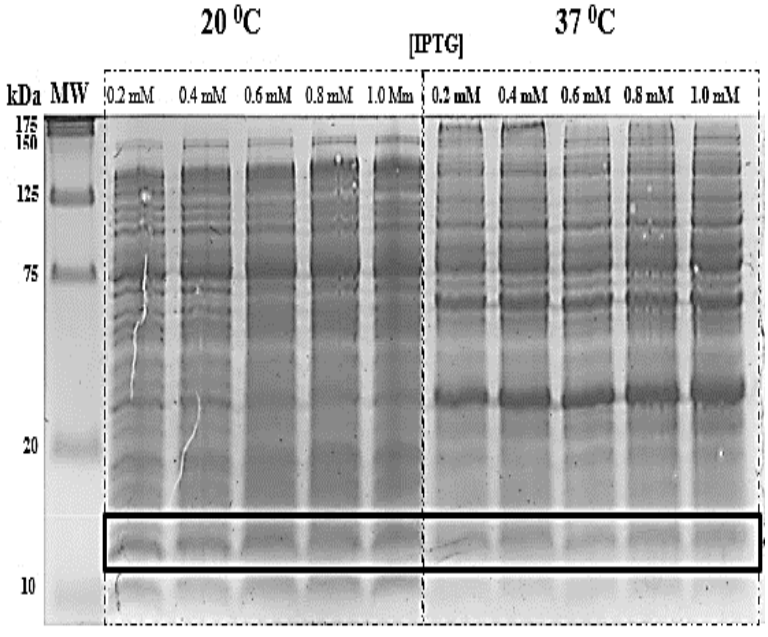
2nd hour after induction



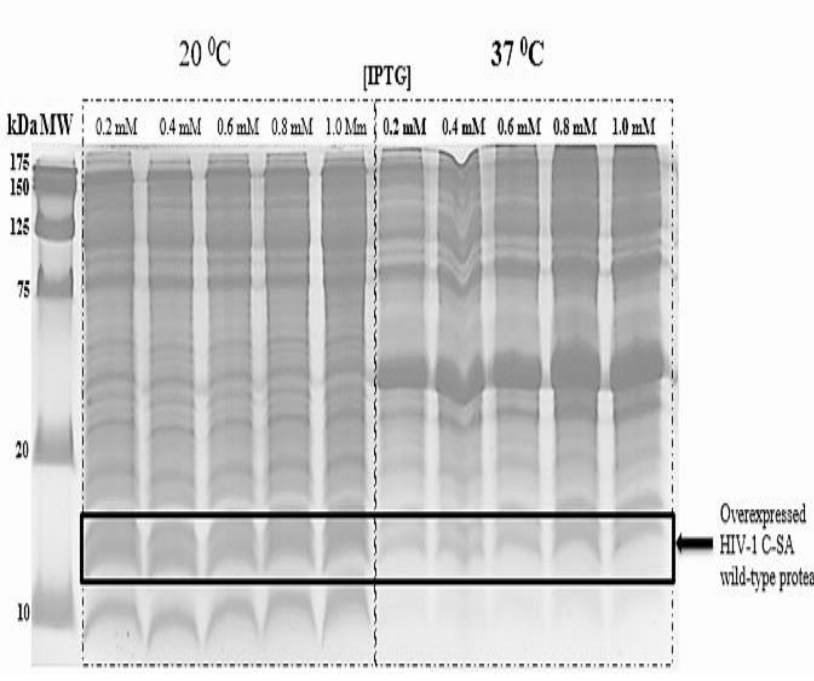
3rd hour after induction



5th hour after induction



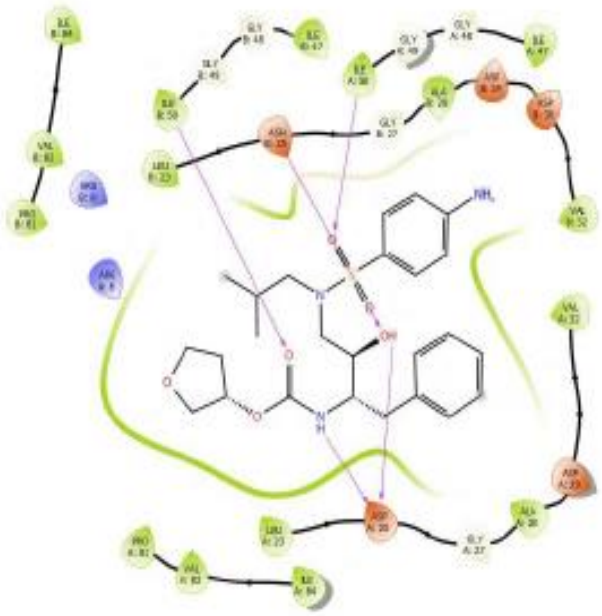
6th hour after induction



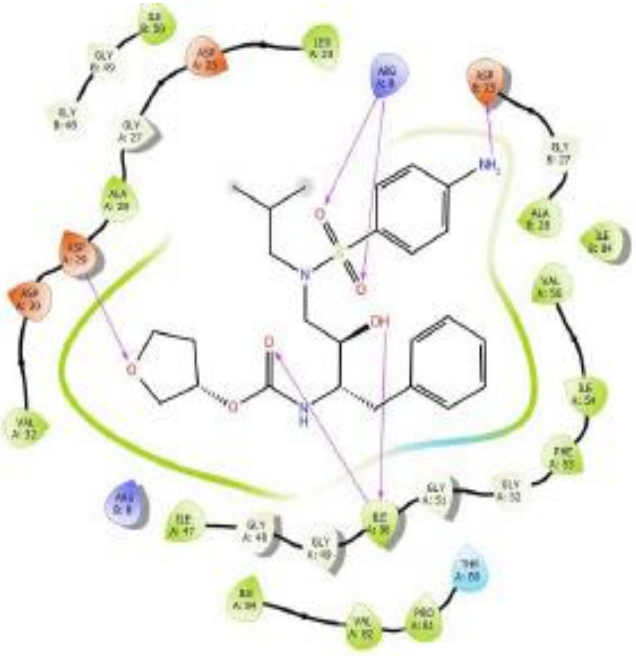
overnight after induction

Appendix B. Supplementary material for section 3.10

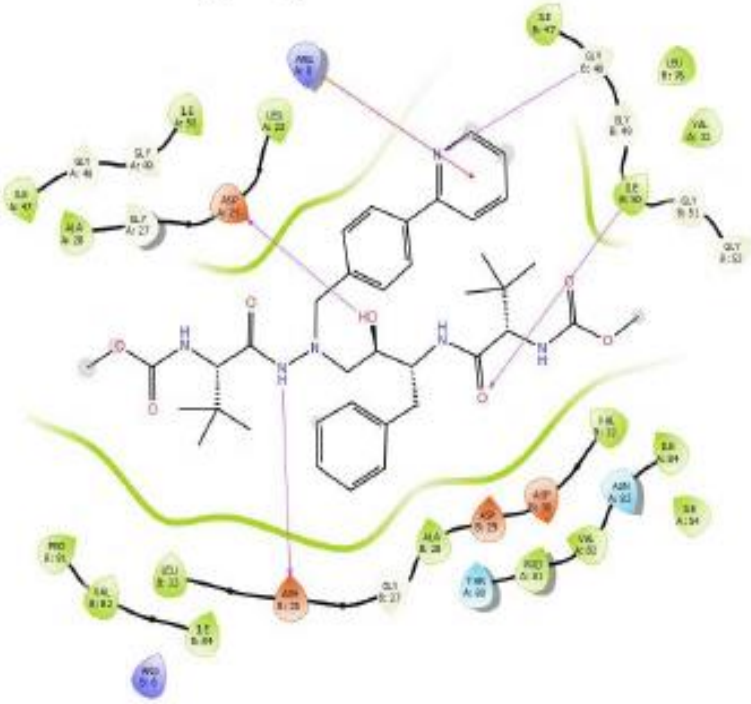
Supplementary Figures. Electrostatic and hydrogen bonding interaction between the inhibitors and the nearby residues of subtype B (A) and C HIV PR (B) in the active sites.



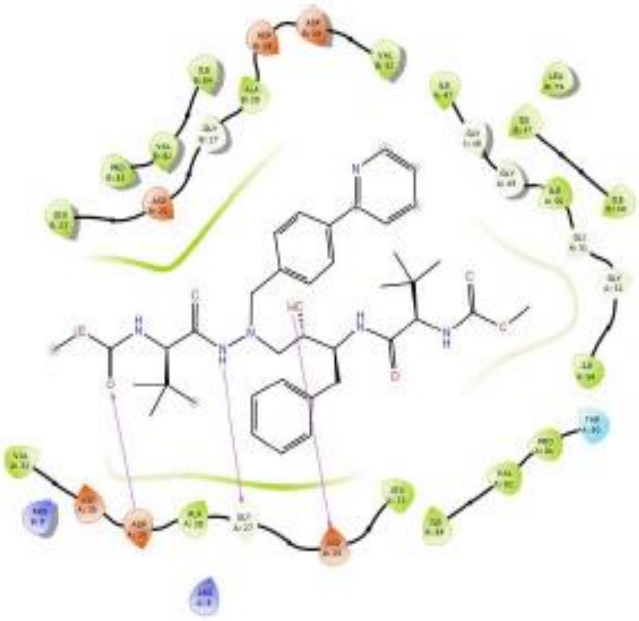
Subtype B protease bound to APV



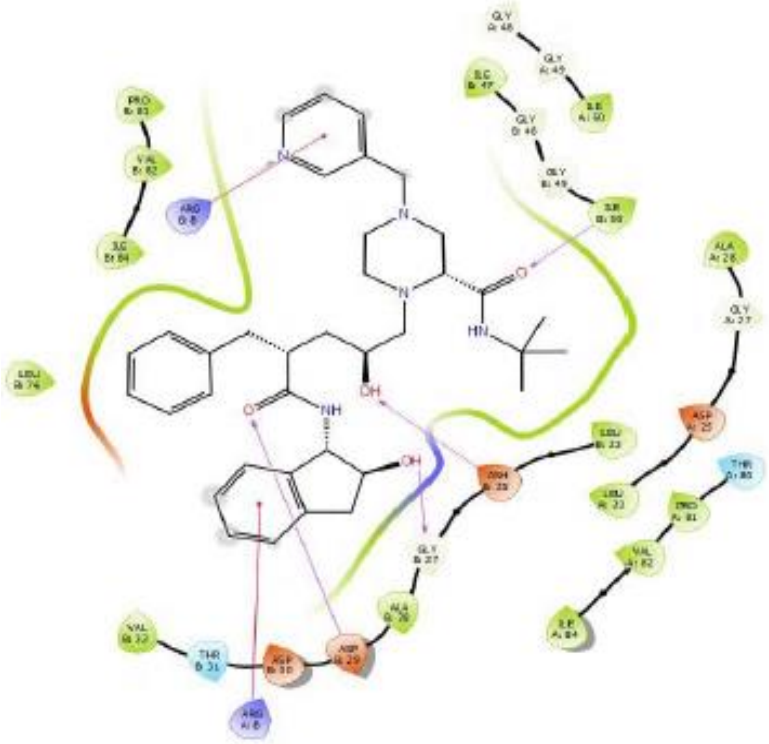
Subtype C protease bound to APV



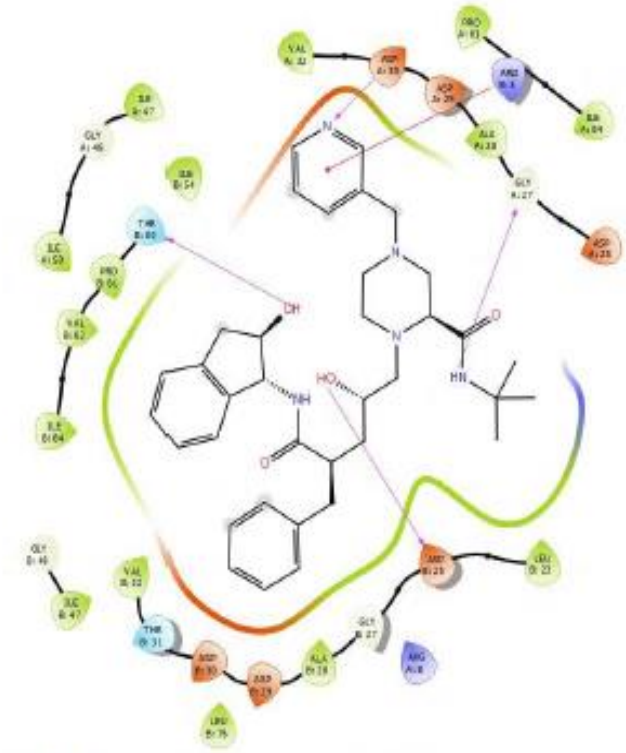
Subtype B protease bound to ATV



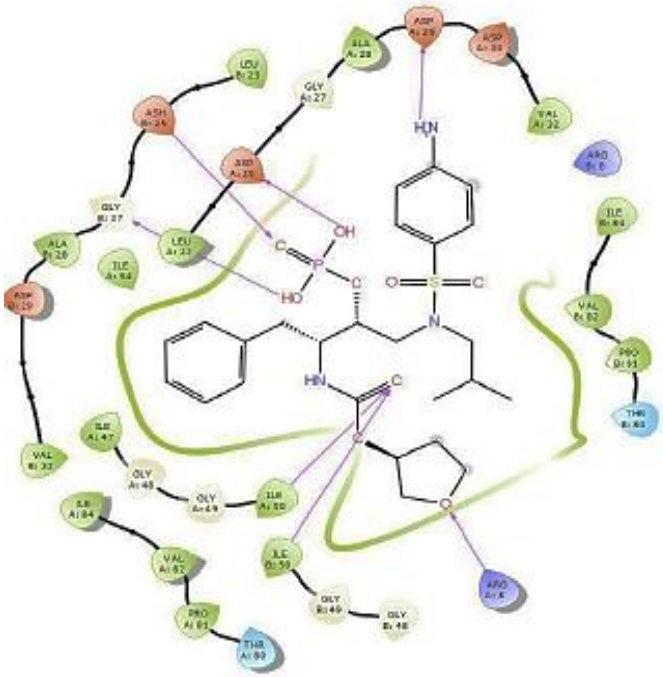
Subtype C protease bound to ATV



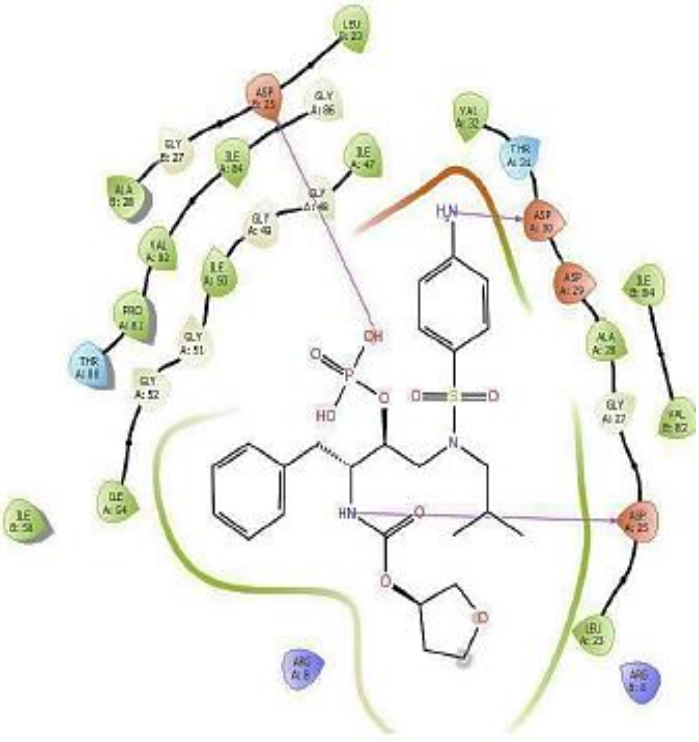
Subtype B protease bound to IDV



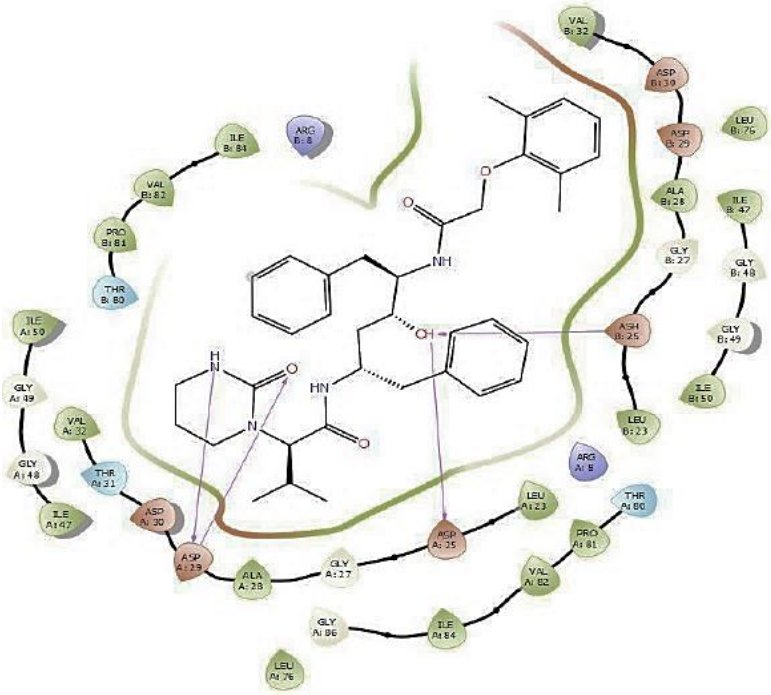
Subtype C protease bound to IDV



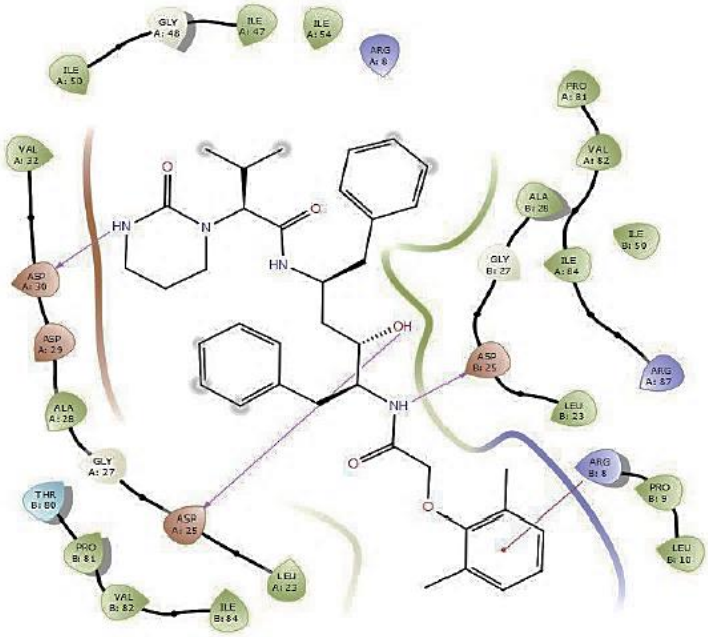
Subtype B protease bound to FPV



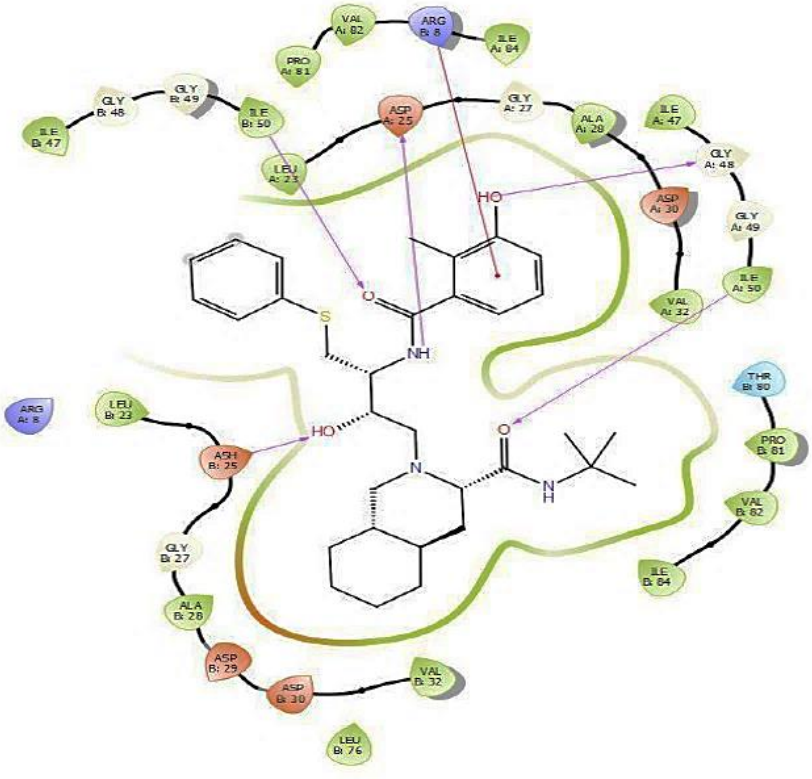
Subtype C protease bound to FPV



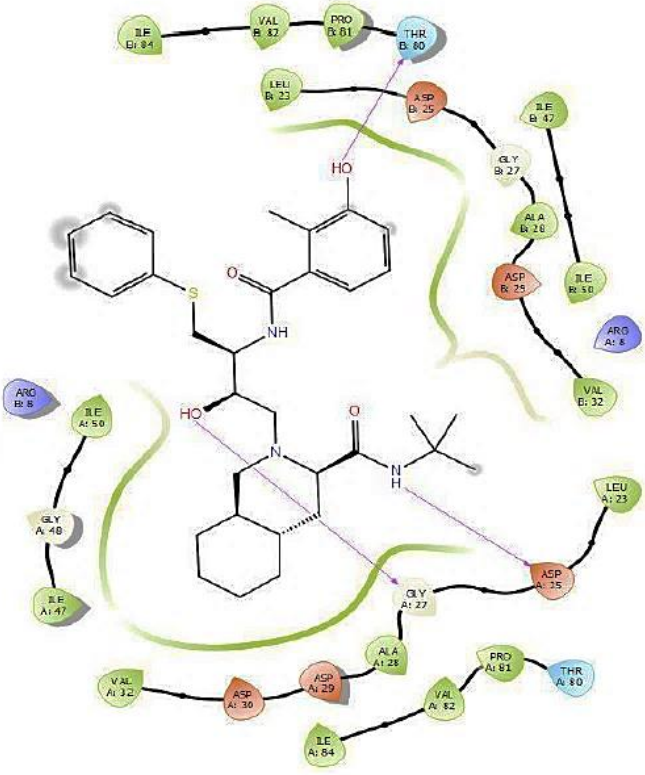
Subtype B protease bound to LPV



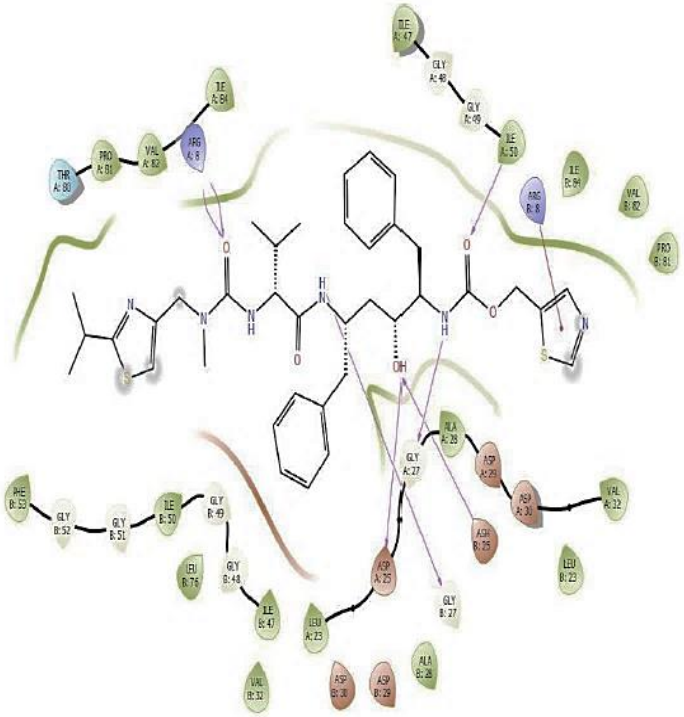
Subtype C protease bound to LPV



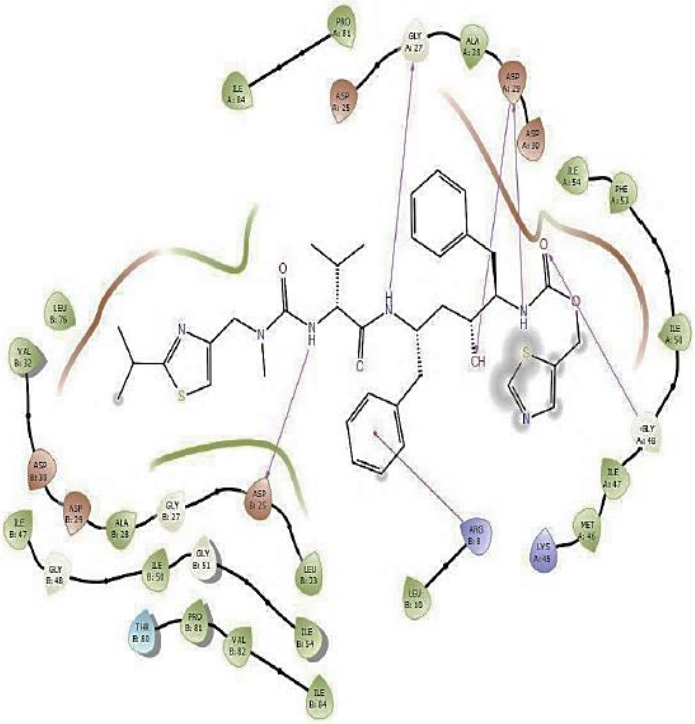
Subtype B protease bound to NFV



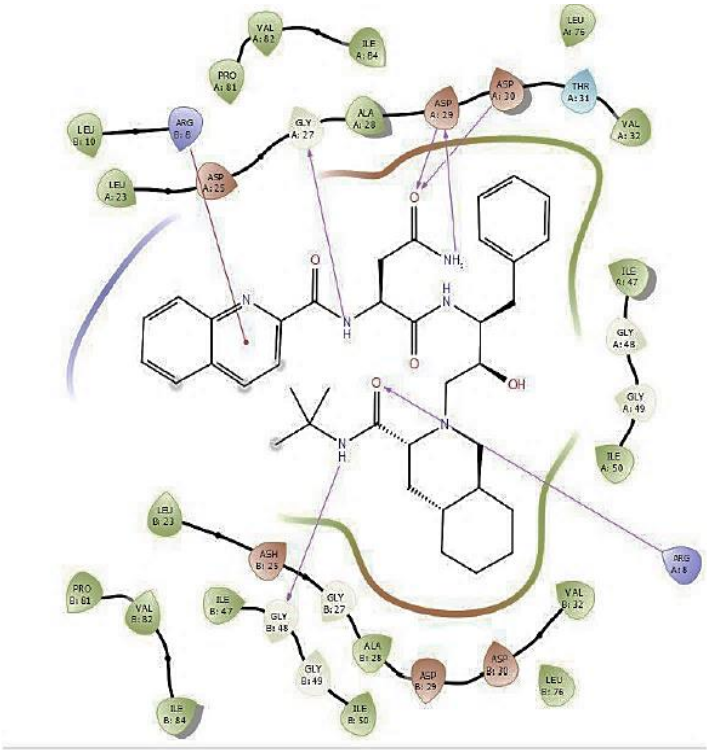
Subtype C protease bound to NFV



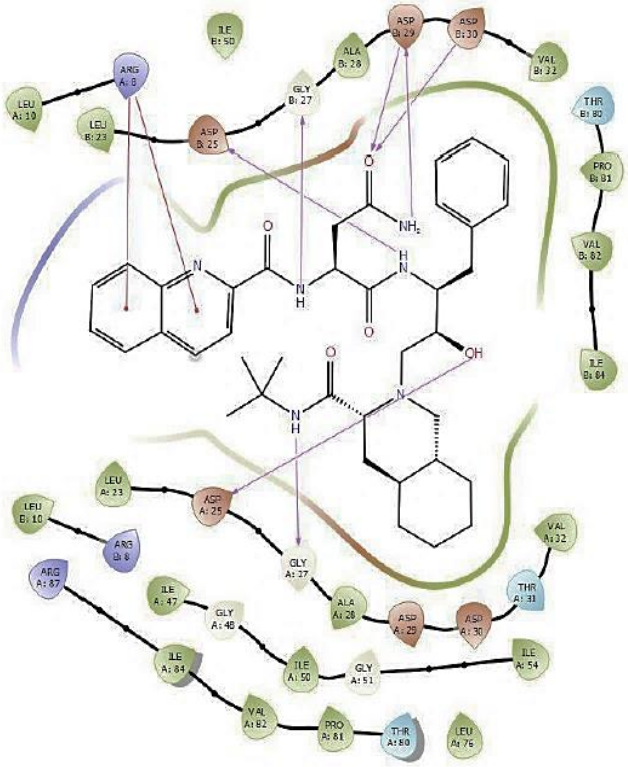
Subtype B protease bound to RTV



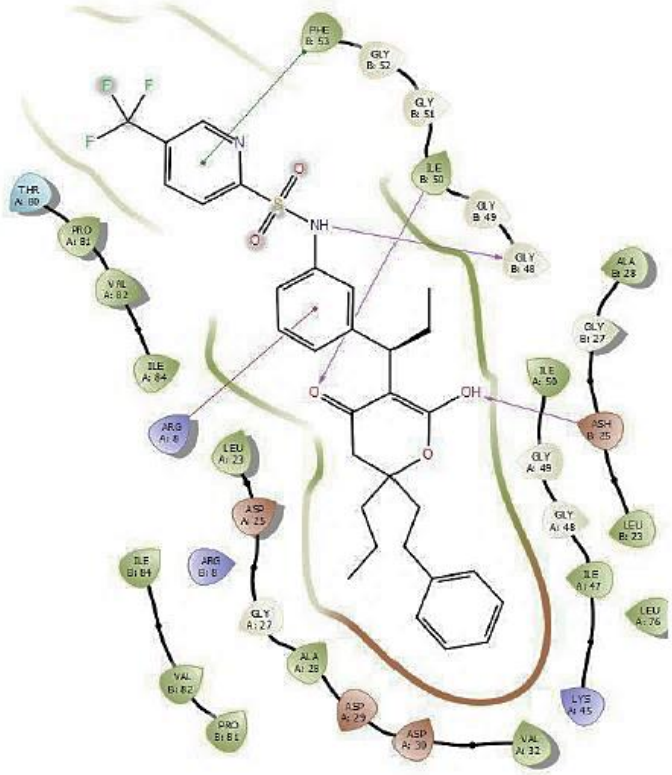
Subtype C protease bound to RTV



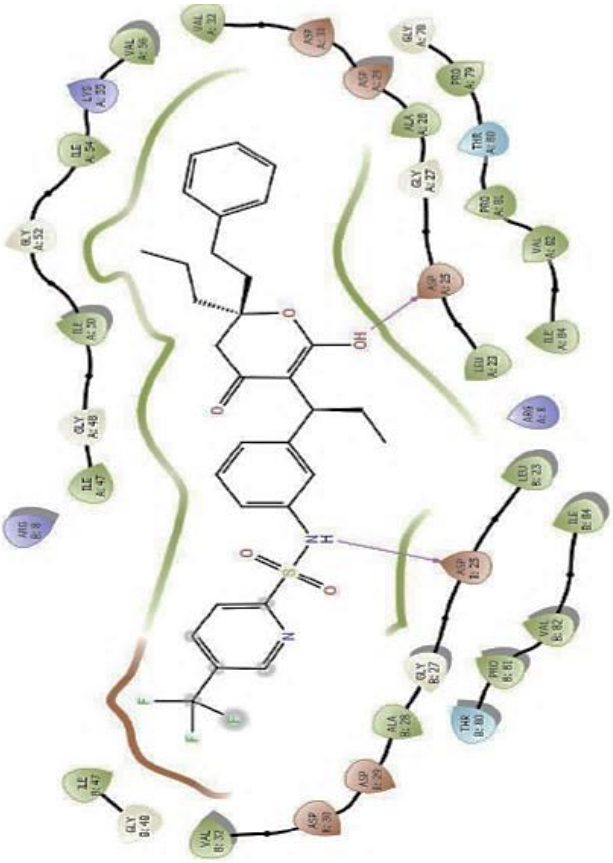
Subtype B protease bound to SQV

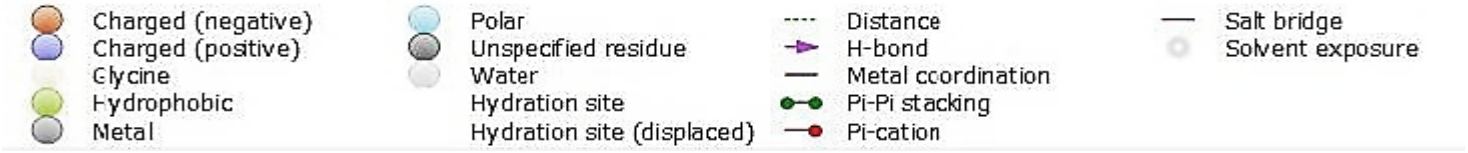


Subtype C protease bound to SQV



Subtype B protease bound to TPV



Subtype C protease bound to TPV

The legends for the ligand-enzyme interaction generated from Ligplot