

https://dspace.org/

HIV-1 subtype C protease: enzyme kinetics, thermodynamics, and X-ray crystal structure

Dlamini, Nozinhle Precious

2023-09

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/42867

Dlamini, Nozinhle Precious. (2023). HIV-1 subtype C protease: enzyme kinetics, thermodynamics, and X-ray crystal structure. [Master's dissertation, University of the Johannesburg]. https://hdl.handle.net/10539/42867

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



HIV-1 subtype C protease: enzyme kinetics, thermodynamics, and X-ray crystal structure

by

Nozinhle Dlamini

(723159)

Dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Professor Y. Sayed

September 2023

Declaration

I, **Nozinhle Dlamini (723159)**, am a student registered for the degree of Molecular and Cell Biology in the academic year 2023.

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. In this context, I understand that the use of editing services is considered aided work and must be declared.

• I have not submitted this work before for any other degree or examination at this or any other University.

• The information used in the Thesis/Dissertation/Research Report HAS/HAS NOT (delete one) 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_____  _____14_ day of _September__year__ 2023

Abstract

Human immunodeficiency virus (HIV), a precursor for AIDS is still one of the most devastating pandemics in history. In 2021 alone there were 650 000 deaths associated with the virus and the number of people living with the infection was recorded to be 38.4 million globally. Sub-Saharan Africa suffers the most burden of the virus with approximately 8.3 million people living with virus, HIV-1 subtype C is the main driver of the disease in South Africa and accounts for 46% of global infections. Even with these alarming statistics this subtype is not the main focus point for the majority of HIV-1 research which mainly focuses on subtype B though it only accounts for 12% of infections globally. There is no vaccine or cure against HIV; however, great strides have been made in suppressing the virus. Viral suppression drugs have been developed to target different stages of viral replication such as those targeting the three important enzymes (protease, reverse transcriptase and integrase). In this study the focus will be on HIV-1 subtype C protease. This is a homodimeric aspartyl protease with 99 amino acids in each monomer. It plays a crucial role in the replication cycle of HIV-1 by producing mature infectious virions through cleavage of the Gag and Gag-Pol polyproteins. The subtype C protease differs from subtype B protease in that it has eight naturally occurring polymorphisms which are substitution mutations, some occurring in different regions of the protease with some in the fulcrum (T12S, I15V and L19I), others in the hinge region (M36I and R41K), with H69K and L89M found in the loops and I93L in the α helix. In this study, structural and functional characterisation of HIV-1 subtype C protease was carried out. The secondary structure was characterised using far-UV CD, which is a technique that measures the difference in left and right circularly polarised light. The subtype C protease was estimated to be predominantly β-sheeted, with spectra showing a maximum at 195 nm and a minimum between 215-225 nm. Tertiary structure characterisation of protease was performed using fluorescence spectroscopy. The maximum emission at 347 nm close to that of water (350 nm), demonstrated that the tertiary conformation of the HIV-1 protease was conserved, and that the tryptophan residues within the protease are solvent exposed. SE-HPLC was used to characterise the quaternary structure of the protease and the homodimeric size was determined to be approximately 22 kDa. Steady-state enzyme kinetics to assess the catalytic activity of the subtype C protease was performed using a fluorogenic substrate. The activity of the enzyme was confirmed, with the specific activity of 24.22 ± 1.72 µmol. $min^{-1}.mg^{-1}$ and the binding of the substrate to the HIV-1 protease was demonstrated by the K_M value of 79.546 ± 6.491 µM. This correlates to literature indicating that the substrate was weakly bound and that a high substrate concentration will be required to reach the maximum velocity (V_{max}), and V_{max} was determined to be 0.036 ± 0.003 µmol. min^{-1}. Enzyme kinetics was coupled with displacement isothermal titration calorimetry for determination of thermodynamics

parameters using second generation PIs (atazanavir, darunavir and lopinavir). Thermodynamic studies indicated that the HIV-1 protease has a high affinity for LPV (K_d = 1 nM), compared to ATV (K_d = 18.57 nM) and DRV (K_d = 42.26 nM) and binding reactions were all spontaneous with ΔG values (ATV = -43.39 kJ/mol, DRV = -41.39 kJ/mol and LPV = -50.51 kJ/mol). The values also indicated that LPV complexed with HIV-1 is more a stable complex. Also, all the binding reactions were exothermic as indicated by the negative ΔH values of ATV = -45.54 kJ/mol, DRV = -55.62 kJ/mol and LPV = -54.71 kJ/mol. The entropy of all the reactions were determined to be unfavourable with the $-T\Delta S$ of DRV = 14.23 kJ/mol followed by LPV: 4.2 kJ/mol and ATV: 2.15 kJ/mol. Overall this suggested that all the binding reactions were enthalpically driven. Furthermore, the three-dimensional structure of the HIV-1 subtype C protease was elucidated using X-ray crystallography. The three-dimensional structure the HIV-1 CSA (PDB ID: 8CI7) was solved at a 2.4 Å resolution which is better than the 2.7 Å (PDB ID: 3U71) initially solved in our lab. The high-resolution three-dimensional structure of the protease will provide precise information about the arrangement of atoms within the protease molecule, thus enabling the design and development of protease inhibitors that will be specific for the subtype C protease. This study emphasised the significance of investigating subtype C protease in the context of enzyme kinetics, thermodynamics and detailed X-ray crystallography.

Acknowledgements

To God Almighty with whom all things are possible I am thankful.

I would like to thank my supervisor for his guidance, and patience thank you Prof. Y. Sayed for the opportunity you have given me I will forever be grateful.

To Dr. Pandian, I am thankful for the assistance you have given me.

To my family (my mom especially) thank you for your strength and my friends (Mpho Setshedi, Aasiya Lakhi thank you for all your support you made it seem possible even in the face of defeat

To the protease group thank you for the good working environment you provided.

To the PSFRU members I am grateful to have been part of team and I have learned through you

To my best friend, my constant support Dr Sibusiso Maseko thank you

Table of contents

Abstract……i

Acknowlegments……iii

List of Abbreviations……vii

List of Figures……ix

List of Tables……x

1. Introduction……1

1.1. A brief view of HIV/AIDS ……1

1.2. HIV-1 replication cycle ……3

1.3. Antiretroviral therapy of HIV-1……6

1.4. The structure and function of HIV-1 protease……7

1.4.1. The HIV-1 protease structure ……7

1.4.2. The catalytic mechanism and specificity of HIV-1 protease……9

1.5. HIV-1 protease inhibition ……14

1.6. Resistance to protease inhibitors ……16

1.7. South African subtype C protease……16

1.8. Gene cloning……17

1.9. Overexpression, purification, and characterisation……17

1.10. Enzyme kinetics……18

1.11. Thermodynamics studies……19

1.12. Three-dimensional structure of proteins……20

1.13. Techniques in structural and functional characterisation of protein……20

1.13.1. Spectrophotometry……21

1.13.2. Spectrofluorimetry ……21

1.13.3. Far-UV circular dichroism spectroscopy ……22

1.13.4. Size-exclusion high performance liquid chromatography ……22

1.13.5. ITC……24

1.14. X-ray crystallography ……24

1.15. Aim.....27
1.16. Objectives.....27
2. Materials and Methods.....28
2.1. HIV-1 protease expression construct and strain.....28
2.2. Expression construct transformed into host strain.....28
2.3. HIV-1 protease growth curve.....29
2.4. HIV-1 protease induction trials.....29
2.5. Tricine SDS-PAGE.....29
2.6. Overexpression of HIV-1 protease.....30
2.7. Inclusion body isolation and purification.....30
2.8. HIV-1 protease characteristics.....31
2.8.1. 1 UV-Vis spectroscopy for protein quality and quantity.....31
2.8.2. ITC for active site titration.....32
2.8.3. Protein secondary structure via Far-UV CD Active.....32
2.8.4. Tertiary structure using Fluorescence spectroscopy.....33
2.8.5. Determination of quaternary structure using SE-HPLC.....33
2.9. HIV-1 protease functionality.....34
2.9.1. Enzyme kinetics via fluorescence spectroscopy.....34
2.9.1.1. Michaelis-Menten kinetics.....34
2.9.1.2. Specific activity and catalytic turnover number.....35
2.9.1.3. Catalytic efficiency.....35
2.9.2. Thermodynamics by displacement ITC.....35
2.10. Three-dimensional structure using X-ray crystallography.....36
2.10.1. Crystallisation and collection of data.....36
2.10.2. Model building and refinement of structure.....37
3. Results.....38
3.1. HIV-1 protease overexpression and purification.....38
3.1.1. Expression construct and sequence verification.....38
3.1.2. Target expression and growth curve.....38
3.1.3. Induction trials.....38
3.1.4. Protein purification and protein assessment.....41
3.2. Characterization of HIV-1 protease structure.....41
3.2.1. Estimation of HIV-1 protease secondary structure.....41
3.2.2. Analysis of HIV-1 protease tertiary structure.....45

3.2.3. Assessing the quaternary structure of HIV-1 protease……45
3.3. Functional characterization of HIV-1 protease ……45
3.3.1. HIV-1 protease enzyme kinetics ……45
3.3.2. Thermodynamics of HIV-1 protease ……45
3.4. The three-dimensional structure of HIV-1 protease ……49
4. Discussion ……55
5. Conclusion ……60
6. References ……61

List of Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral therapy
CA	Capsid
CCR5	C-C chemokine receptor type 5
CD	Circular dichroism
CRFs	Circular recombinant forms
CXCR4	C-X-C chemokine receptor type 4
DI	Dimer interface
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Env	Envelope protein
FDA	Food and Drug Administration
Gag	Group-specific antigen
Gp	Glycoprotein
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
IN	Integrase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
kDa	kilo-Dalton
MA	Matrix
M	Major
N	Non-M, non-O
Nef	Original negative factor
NC	Nucleocapsid
NNRTI	Non-Nucleoside Reverse Transcriptase Inhibitor
NOPs	Naturally occurring polymorphisms
NRTI	Nucleoside Reverse Transcriptase Inhibitor
O	Outlier
OD	Optical Density
PDB	Protein Data Bank
PI	Protease Inhibitor
PMSF	Phenylmethylsulfonylfluoride

Pol	Polyprotein of HIV-1
PR	Protease
p6	Assembly protein
Rev	Regulatory of viral protein expression
RNA	Ribonucleic acid
RT	Reverse Transcriptase
RSV	Rous sarcoma virus
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SIVs	Simian immunodeficiency viruses
Tat	Transactivator of transcription
URFs	Unique recombinant forms
Vif	Viral infectivity factor
Vpu	Viral protein U
Vrp	Viral protein R

List of Figures

Figure 1. A map showing prevalence of group M subtypes around the world 2

Figure 2. HIV genome and virion organization 4

Figure 3. A schematic view showing the replication cycle of HIV 5

Figure 4. Structure of HIV-1 protease and different conformations 8

Figure 5. Processing of HIV-1. 10

Figure 6. Nomenclature of the active site 12

Figure 7. The catalytic mechanism of the protease. 13

Figure 8. Techniques in protein analysis 23

Figure 9. ITC principle 25

Figure 10. Steps involved in X-ray crystallography 26

Figure 11. Expression construct and sequence verification 39

Figure 12. Assessment of gene expression and growth curve for HIV-1 protease 39

Figure 13. Analysis of HIV-1 protease induction trials using tricine SDS-PAGE 40

Figure 14. HIV-1 protease purification profile and tricine SDS-PAGE 42

Figure 15. The UV Absorption spectrum of HIV-1 protease 42

Figure 16. spectrophotometric quantification of HIV-1 protease quantification at 280 nm and 340 nm 43

Figure 17. Concentration analysis of HIV-1 protease using active site titration 43

Figure 18. Analysis of HIV-1 protease secondary structure using far-UV CD 46

Figure 19. Fluorescence spectra of HIV-1 protease 47

Figure 20. Quaternary structure analysis of HIV-1 protease 48

Figure 21. Typical displacement ITC experiment of HIV-1 protease with PIs 51

Figure 22. Thermodynamic parameters of HIV-1 protease interaction with ATV, DRV and LPV 52

Figure 23. HIV-1 protease crystals obtained from using the vapor diffusion method 53

Figure 24. A ribbon representation of the HIV-1 C-SA crystal structure 53

List of Tables

Table 1. HIV-1 protease inhibitors and their properties 15

Table 2. Thermodynamic parameters of HIV-1 protease from ITC 44

Table 3. Enzyme kinetic parameters for HIV-1 protease..... 50

Table 4. Data collection and refinement of the HIV-1 protease structure..... 54

1. Introduction

1.1 A brief view of HIV/AIDS

Human immunodeficiency virus (HIV) is one of the most devastating pandemics in human history (1). This subclass of retroviruses was initially discovered in 1981, and arose due to transmission of simian immunodeficiency viruses (SIVs) from primates to humans (2). The high mortality rates associated with HIV through its progression to AIDS can be attributed to its mode of infection that primarily targets infection-fighting human immune cells (CD4+ T cells) and macrophages. Due to their ability to withstand the structural changes of the virus, the latter maintains and propagates viral infection for longer periods of time (3). While CD4+ T cells are destroyed with a half-life of less than two days as a result of a cell death signal that is induced during HIV replication (4, 5). Despite efforts to curb the surge of HIV it remains a global crisis largely due to the wavering progress in treatment and prevention worldwide. In 2021, the AIDS pandemic claimed numerous lives and new infections were taking place more frequently. This led to 650 000 deaths with new infections amounting to 1.5 million and 38.4 million people were said to be living with the virus worldwide (5). Sub-Saharan Africa has the largest HIV epidemic in the world accounting for ~ 70% of the global disease burden, with ~ 8.3 million adults living with HIV in South Africa (5).

Two genetically distinct types exist for the AIDS precursor, HIV-1 and HIV-2, with HIV-1 being the most prevalent consisting of various strains that have been divided into M (major), N (non-M, non-O), O (outlier) and P (putative) categories (6, 7). The world's infection burden is due to the M category (Figure 1) and because of its diversity has been further classified into several variants which include A-D, F-H, J-K clades with subtypes (A1-A4, F1- F2), circular recombinant forms (CRFs) and unique recombinant forms (URFs) (7). Naturally occurring polymorphisms (NOPs) are what cause the variances between the subtypes (8). Globally, subtypes A, B and C are the most widely distributed viral strains, with subtype C being the primary cause of the pandemic in South Africa (Figure 1) and accounting for 46% of global infections (9). Heterosexual transmission is the primary cause of this pandemic in SA, females account for a high incidence rate (10). Subtype B, though accounting for 12% of infections was the main focus of earlier research into antiretroviral therapy and prevention strategies. This was due to subtype B being prevalent in the United States and Europe (Figure 1) (9), which were the regions that provided a majority of resources and funding towards early research in HIV/AIDS.

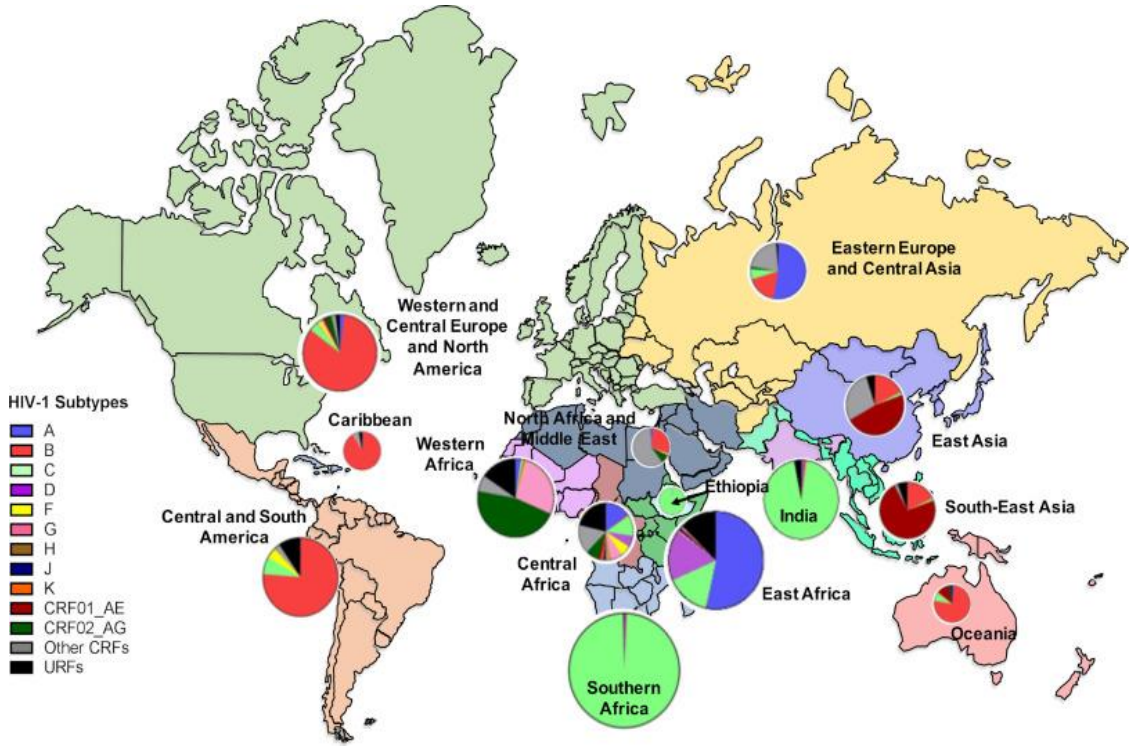


Figure 1. A map showing prevalence of group M subtypes around the world. The percentage of each circulating subtype within a region is represented by the pie graph with the pie size indicating the number of infections and each region is colour coded. (Figure taken from Gartner *et al.*, 2020) (9).

The discovery of HIV as a lentivirus in 1985 (11), shed light on some of the unique features of the virus. The complex genomic structure of lentiviruses which includes numerous genes encoding proteins associated with viral replication, immune evasion and host cell interaction distinguishes them from other viruses (12). This virion ~ 100 nm is characterised by the spherical outermost portion, spicule-studded appearance, and cone-shaped core (Figure 2). The HIV genome encodes an RNA and fifteen crucial structural and non-structural proteins involved in the virus's life cycle (Figure 2). Three open reading frames group specific antigen (gag), pol and envelope (env) are found from the 5'- to 3'-end of long terminal repeats (LTRs). Structural elements include viral matrix (MA), capsid (CA), nucleoprotein (NC) and p6 of the Gag and surface (SU) glycoproteins (gp) gp120 with transmembrane (TM) glycoprotein gp41 of the Env proteins. Enclosed within the virus are enzymes reverse transcriptase (RT), integrase (IN) and protease (PR) encoded by the pol gene and play a crucial role in viral replication. The virus encodes three of the six accessory proteins, viral infectivity factor (Vif), viral protein R (Vpr), and original negative factor (Nef), with transactivator of transcription (Tat) and regulatory of viral protein expression (Rev) serving as regulatory genes. While two RNA molecules are contained within the particles and viral protein U (Vpu) plays a role in virion assembly (13, 14). To date HIV continues to be one of the most thoroughly reviewed lentiviruses, with its research providing insights into the replication mechanism, immune evasion and pathogenesis of the virus (11).

1.2 HIV-1 replication cycle

HIV, like all viruses, cannot exist outside of a living cell because it is reliant on the replication apparatus of the host cell to carry out self-replication (15). HIV transmission occurs when infected blood or bodily fluids enter the host (16). HIV-1 though only encoding 15 mature proteins (Figure 2) can relentlessly attack and impair the individuals innate and adaptive immune systems (17). It mainly targets the hosts CD4+ helper T cells which have a crucial role in regulating both cellular and humoral immune responses (15). Due to the complexity of the viral replication, it is divided into an early and late phase. The virion's attachment to the cell surface and the integration of the proviral DNA into the host genome are the first two stages of the early phase (Figure 3), while the late phase begins with proviral transcription initiation and to conclude the process a fully infectious virion is released (15).

Host cell infection begins when the virus binds to CD4+ receptors via the glycoproteins (gp120 and gp41) (17). Subsequent binding between the host cell's CCR5 or CXCR4 chemokine co-receptors and the virus allows fusion with the cell membrane through the formation of an irreversible conformational shift, permitting the release of the of the viral core into the cytoplasm of the host (18, 19). The genomic viral RNA is reversed transcribed by the virions RT enzyme into DNA upon destruction of the viral core (20). Due to the RT enzyme's lack of proofreading activity

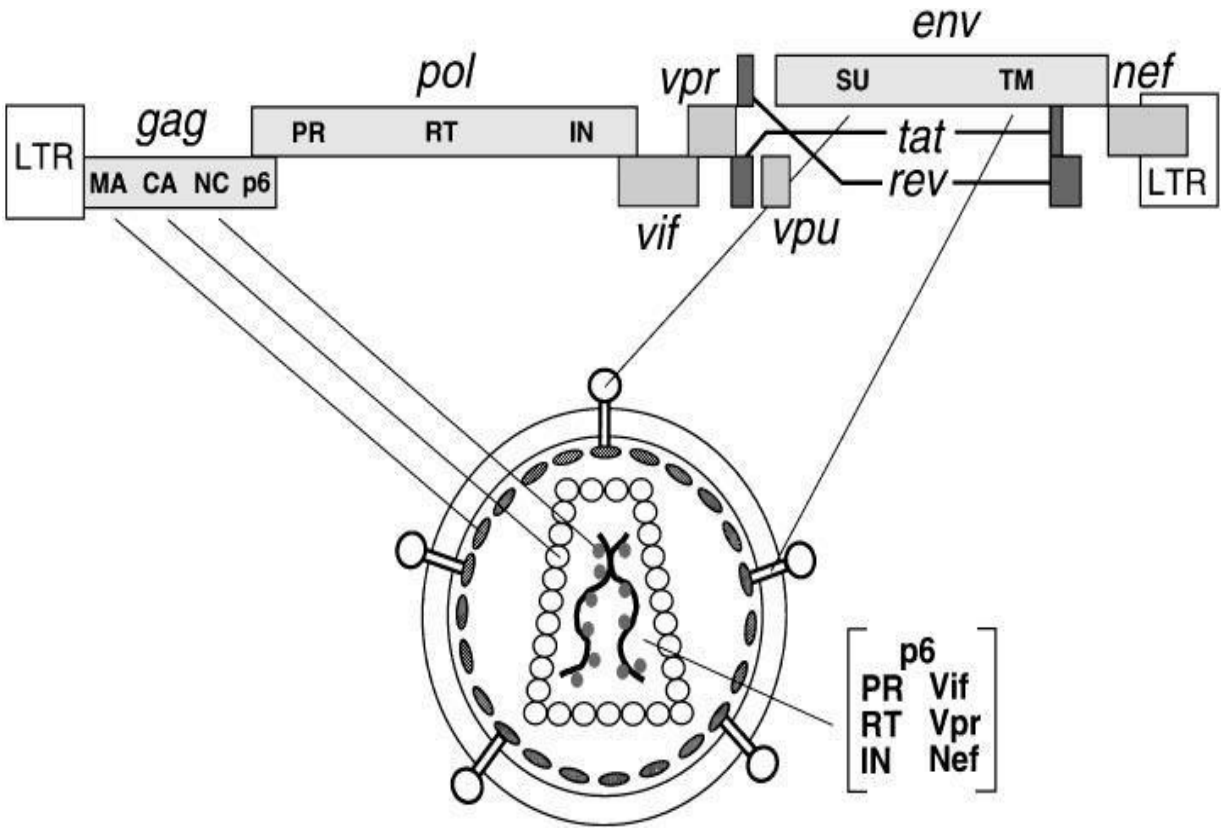


Figure 2. HIV genome and virion organisation. Long terminal repeats (LTR), matrix protein (MA), Capsid (CA), Nucleocapsid (NC), assembly protein (p6), Protease (PR), Reverse Transcriptase (RT), Integrase (IN), Regulatory of viral protein expression (Rev), Surface envelop glycoprotein gp120 (SU), Transmembrane glycoprotein gp41 (TM), Transactivator of transcription (Tat), Viral infectivity protein Vif), Viral protein R (Vpr), Negative regulatory factor (nef). (Figure taken from Frankel and Young 1998) (14).

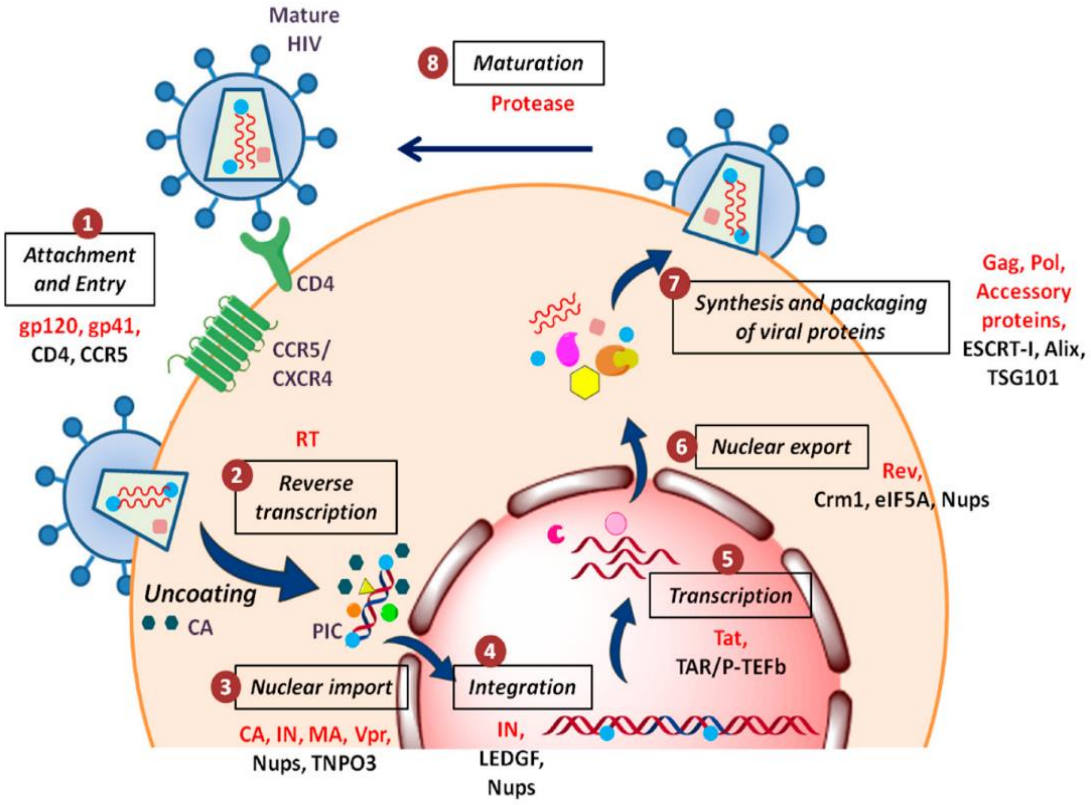


Figure 3. A schematic view showing the replication cycle of HIV (17). (1) The HIV-1 particle enters the host cell cytoplasm after adhering to and fusing to the cell membrane using its surface glycoproteins. (2) viral genomic RNA is reverse transcribed into DNA, which is followed by the viral capsid's removal. (3) the pre-integration complex enters the nucleus with the assistance of several host nuclear proteins. (4) Viral DNA is integrated into the chromatin of the host. (5) Viral transcription occurs during productive infection. (6) Splicing of viral RNA and nuclear exportation takes place. (7) New viral particles that will bud out from the membrane are produced. (8) Virions mature and become infectious through the action of the HIV protease. Letters in red and black indicate viral and host proteins respectively, and their involvement in each replication step. (Figure taken from Shukla *et al*., 2019).

throughout the transcription process, several closely similar viral variants are produced. The viral IN enzyme inserts the viral genome into the host genome after transcription. Many host nuclear proteins facilitate the nuclear entry of pre-integration complex like the presence of lens epithelium-derived growth factor. This integrase host binding factor is found in the host's chromosomal DNA (21). The transcription machinery of the host transcribes the integrated viral and host DNA into many copies of HIV RNA followed by the translation of mRNA outside the nucleus to proteins (22). The assemble of new viral RNA copies and proteins at the cell membrane results in the formation of non-infectious HIV virions which bud out of the cell (15). The lengthy protein chains are disassembled and reassembled into mature infectious virions by the action of viral protease enzyme (15, 17). The newly formed mature virions spread out and infect other CD4+ helper T cells, resuming the whole process of replication again and infecting other cell types such as the macrophages, more resting T cell subsets and immature dendritic cells. It is of significant interest to understand how HIV interacts with host cells during replication as it gives insights on how to improve antiretroviral drugs and design strategies to curb the virus (15).

Various antiretroviral drugs have been designed to target HIV at different stages of the viral replication cycle, by inhibiting the protease role to produce mature virions which will prevent the virus from binding to cellular receptors, preventing the action of RT enzyme and viral genome integration (17).

1.3 Antiretroviral therapy of HIV-1

Antiretroviral drugs have played a critical role in the reduction of HIV incidences, morbidity and mortality associated with HIV, at the end of 2021 there were 28.7 million people globally on antiretroviral therapy (ART). At present more than 30 drugs are being used for HIV treatment targeting various stages of the replication cycle (17). There are currently six classes of ART drugs classified based on their therapeutic mechanism of action. The advised AIDS highly active antiretroviral therapy (HAART) regimen comprises at minimum with three drugs from two kinds of ART: fusion inhibitors, non-nucleoside reverse transcriptase Inhibitor (NNRTIs), Chemokine receptor antagonists, nucleoside reverse transcriptase inhibitor (NRTIs), protease inhibitors and integrase inhibitors (23). The ART initiation for all individuals infected with HIV is still an ongoing struggle due to factors such as lack of funding, effective infrastructure, treatment access, long-term adherence to ART and weak indicators of HIV drug resistance (24). The accelerated development of drug resistance strains creates a barrier to viral suppression and renders ART ineffective, 40-50% of patients who initially experience viral suppression will subsequently succumb to failure of treatment (25). When a frameshift mutation in the pol gene that encodes the protease led to the generation of immature,

non-infectious virions, the HIV protease was initially proposed as a druggable target in the HIV/AIDS treatment resulting to it robust characterisation (26) and will be discussed in detail.

1.4 The structure and function of HIV-1 protease

Very early on during the HIV epidemic it was discovered that the viral genome encoded an aspartyl protease within the pol region (27). The aspartyl protease contained a signature sequence corresponding to a Asp-Thr-Gly sequence close to the N-terminus of one of the coding regions (27). Pepstatin inhibition and the inactivation of the active site of the HIV-1 protease's aspartate (Asp25) were used to confirm the existence of aspartic protease (28). It is known that large aspartic proteases have two Asp-Thr-Gly sequences indicating that the two subunits of the molecule are the result of gene duplication (27). Homology modelling techniques with the Rous sarcoma virus (RSV) PR structure were used for the structural confirmation of the HIV-1 protease homodimeric aspartyl enzyme (28).

1.4.1 The HIV-1 protease structure

The protease (PR) has a homodimeric molecular weight of 22 kDa and was the first HIV-1 protein structure to be described (28, 29). Each monomer of the C2 symmetric homodimer contains 99 amino acids commencing at proline-69 and ending at phenylalanine-167 from the *pol* gene open reading frame (30). Three major domains of the protease include the flexible flaps, dimer interface (DI) and the active site with different elements (31, 32). According to the protease secondary structure (e.g., Figure 4A), each monomer contains one short helix (colors), and antiparallel sheet β strands predominate the structure (33). The interaction of 8 amino and carboxyl terminal residues (residues 1-4 and 96–99) from each chain form the dimer interface. These residues make up the four-stranded β-sheet protease, which has the hydrophilic side chains Gln2, Thr4, Thr96, and Asn98 on the outside and the hydrophobic side chains Pro1, Ile3, Leu97, and Phe99 inside the interface. The 96–99 carboxyl terminus of this periodic structure, four salt bridges, and 34 hydrogen bonds are all responsible for the dimer's stability (34). 75% of the total Gibbs energy stabilizing PR comes from a DI (35). During dimerization, a central active site made up of the triad residues Asp25-Thr26-Gly27 from the two subunits is formed with both Asp25 having catalytic activity (33). Each monomer's Asp25 carboxyl groups are nearly coplanar and many residues contributing to the formation of the active site are Arg8, Leu23, Asp25-Thr-Gly-Ala-Asp-Asp30, Val32, Ile47-Gly-Gly-Ile50, Phe53, Thr80-Pro-Val82, and Ile84. These residues clearly show that the lining of this pocket is extensively distributed with negative and positive charge in addition to hydrophobic structural features (27). The loop in which the active site is positioned within receives its structural stability from a web of hydrogen bonds it forms with another monomer's loop. These interactions of hydrogen called the "fireman's grip" contribute to the rigidity of network (28, 36).

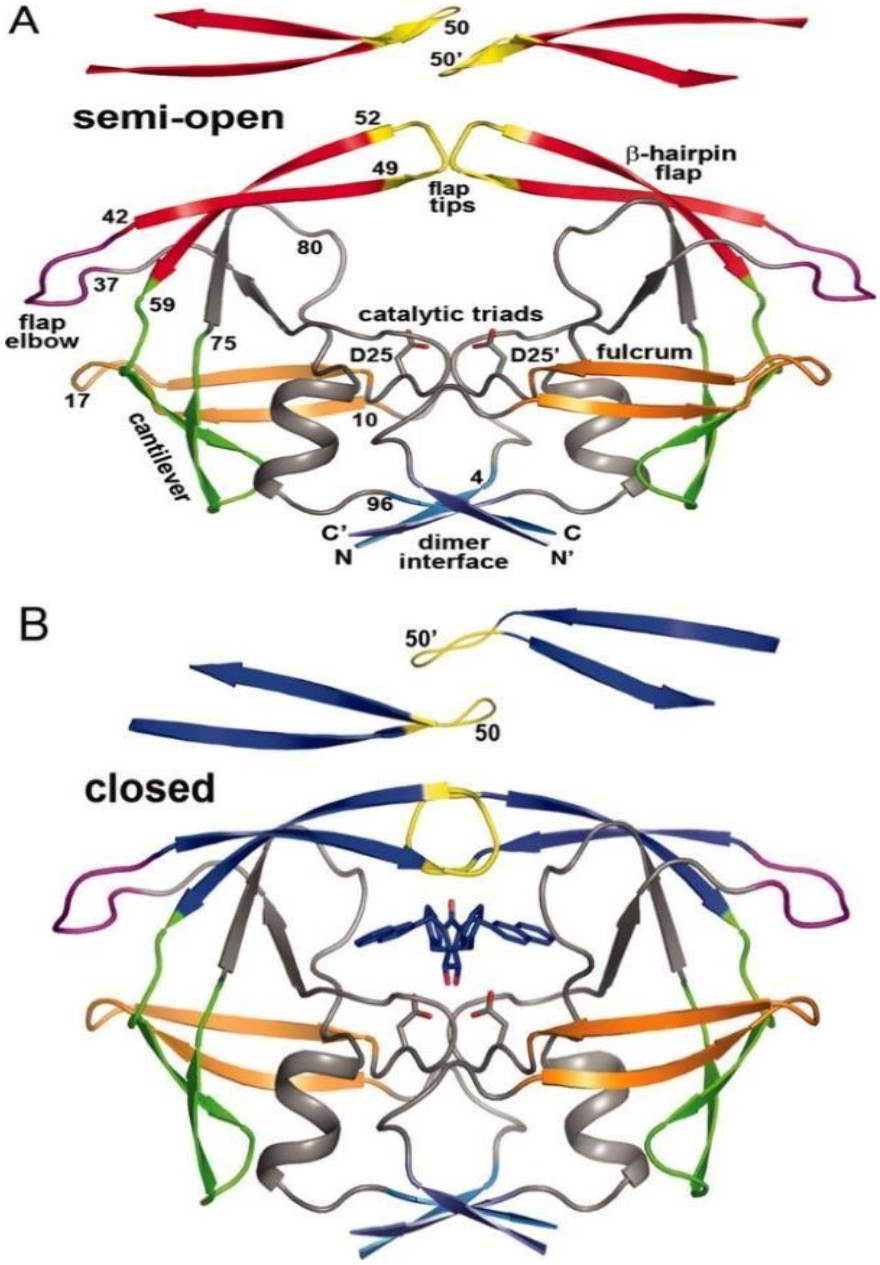


Figure 4. Three-dimensional ribbon representation of the HIV-1 protease structure showing different conformations. Flaps (red/blue), flap tips (yellow), flap elbow/hinge (magenta), cantilever (green), fulcrum (orange), DI (blue/cyan). (A) Upon inhibitor removal the protease returns to a semi-open configuration with flap tips distance of about 4.3 Å (PDB 1HHP). (B) In the closed conformation there is approximately 5.8 Å between the flap tips from each monomer (PDB 1HVR). Understanding the open and closed conformations has an impact on understanding drug resistance mechanisms. (Figure taken from Hornak *et al*., 2006) (37).

Each monomer has flaps that contain two antiparallel β strands connected through a β turn and form a β hair pin loops that are flexible as they are glycine rich (27, 29). They form a protective roof over the protein's active site and their dynamics which are important for enzyme function and dictate the volume and pocket size (38). The flaps have a crucial role in substrate/inhibitor binding and will undergo structural changes depending on the presence or absence of this binding (39). The unbound crystal structure of the protease showed three conformational changes that take place at equilibrium open, semi-open and closed form (40) and the flap tips (Ile-50C^{α}-Ile-50' C^{α}) distance is used for measuring these changes (37). The overlapping of the two flap tips blocks substrate entry to the active site, when there is an overlap with the substrate a 15 Å upward motion occurs in the flaps and once the inhibitor binds allowing for its lockage in the catalytic site creates a closed conformation Figure 4B (33, 41). An open configuration is required for normal substrate binding, molecular simulation studies depict transient openings with a broad flap reshuffling and curling back of the flap tips into protein structures with an orientation of approximately 20 Å (42). Flaps stability and movement is closely linked to residues 35-42 and 35'-42' of the hinge region (43). The shift of the flaps (43-58 and 43' -58') occurs through the cantilever with the compensatory change being experienced by residues 59-79 and 59'-79'. A fulcrum with residues 11-22 and 11'-22' is a β-sheet region and corresponds to the flap and cantilever, the flaps close down when the cantilever shifts up (44). The site-directed mutagenesis was not only instrumental for the protease characterisation but also played a crucial role in demonstrating the protease vital function in viral maturation (36).

1.4.2 The catalytic mechanism and specificity of HIV-1 protease

The proteolytic nature of the HIV-1 protease enables it to liberate itself from the Gag-Pol precursor, it further catalyse the cleavage events of polyproteins to produce structural proteins necessary for viral maturation (45). The Gag is translated as a fusion protein of 55 kDa ($p55^{gag}$) during the viral replication and the Pol as a $p160^{gag.pol}$ precursor (36). Gag and Pol are cleaved at twelve different processing sites, where the asymmetric shape of the substrate is recognised regardless of the amino acid sequences (46, 47). Figure 5 shows that the sequential processing of the Gag polyproteins starts with the cleavage of the C-terminal part of spacer peptide 1 (p2) which separates nucleocapsid and capsid protein (MA-CA-p2↓NC-p1-p6, arrow indicates cleavage). The events that follow occur simultaneously between the capsid and the matrix protein (MA↓CA-p2) and the C-terminal sequence release of the p6 from the C-terminal (NC-p1↓p6) of the spacer peptide 2 (p1). Lastly, the small spacer peptides 1 (p2) and 2 (1), are cleaved from the capsid (AC↓p2) and nucleocapsid (NC↓p1) respectively (46, 48). The Gag maturation and ordered regulation of cleavages is facilitated by the spacer peptide 1 (p2) (48) with spacer peptide 2 (p1) playing a critical role in Gag and Pol incorporation

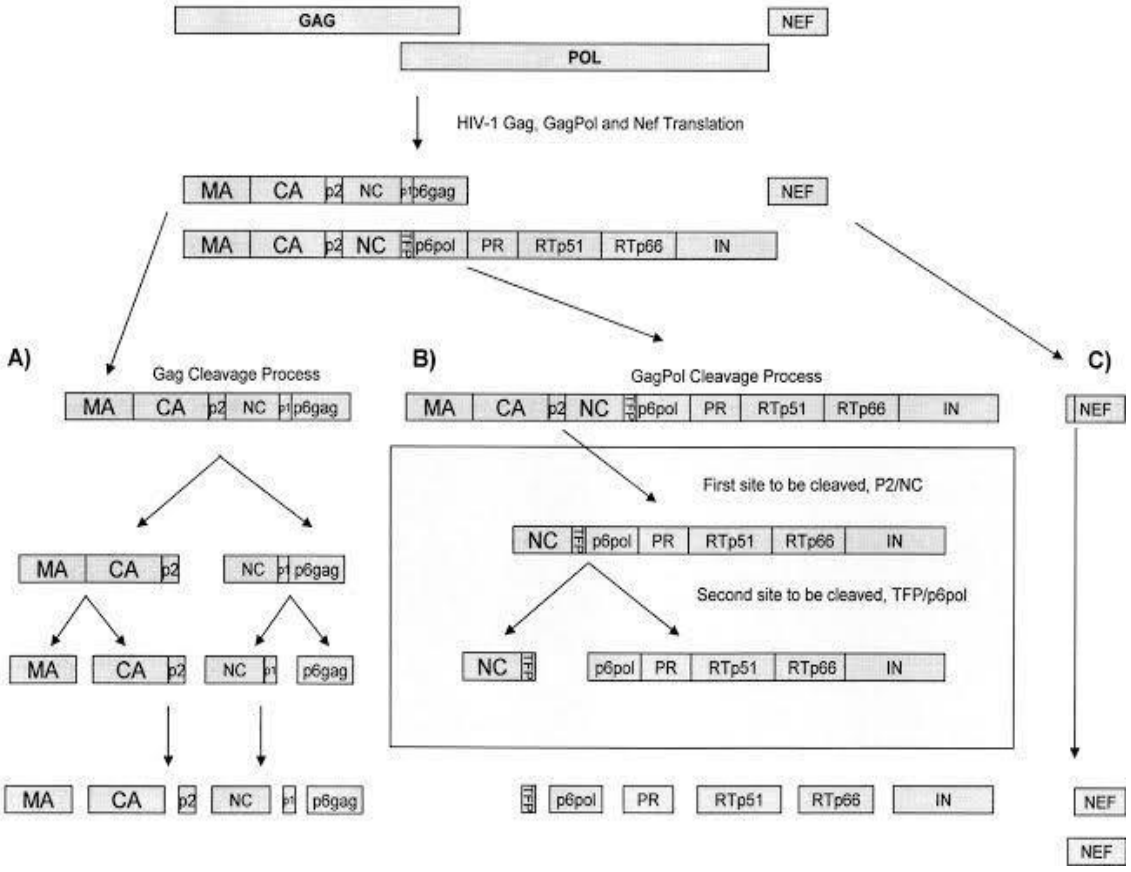


Figure 5. Processing of HIV-1. Cleavage of Gag (A), Gag-Pol (B) and Nef (C) by the protease at different cleavage sites resulting in structural proteins and functional viral enzymes needed for the assembly and maturation of the virus. (Figure taken from De Oliveira *et al*., 2003) (51)**.**

into the virus (49). Primarily, the Pol is cleaved from the Gag-Pol and further processed to protease, p51 (RT), p66 (RT) and integrase (50).One of the interesting features of the protease is its broad substrate specificity. It selects numerous unrelated sequences and effectively cleaves them (27). The peptide's superimposable secondary structure results in a substrate envelope that fits well into the binding site of the protease (46). In proteins, the regions that are protease-susceptible consist of the octapeptides or a seven amino acid sequence (52). The interaction for these amino acids can be from the P4 to P4' in the substrate with the protease S4-S4' corresponding subsites leading to the scissile bond cleavage (33) as shown in Figure 6. The allocation of the substrate can be done in one of the three groups according to the cleavage site flanking sequences with the first-class containing Phe-Pro or Tyr-Pro at P1 and P1', while the second class contains Arg at P4 and Phe-Leu at P1'-P2' and the third class has Glu or Gln at P2' (36). The most common amino acid found at cleavage sites is glutamic acid at the P2' position, and its presence is linked with a decrease in protease discrimination towards P1 and P1' substituents (33). When P2-P2' interacts with distinct protease groups, hydrophobic side chains are observed, while some polar interactions take place between the amide groups of P2-P2' and the polar side chains of the protease (53).The pockets of S2- S2' are hydrophobic, but hydrophilic and hydrophobic residues can both occupy them, with the site being highly specific due to P2 restriction on residue size and type (53).

The protease catalyses the hydrolysis of the peptide bonds using two distinct mechanisms. The first one uses activated water molecules to hydrolyse the scissile bond of the substrate's amide carbonyl bond (Figure 7B). This mechanism involves the interaction of catalytic aspartate residues (Asp25-Asp25') with a water molecule. After this interaction, a general acid-base process takes place within these residues, causing the water molecules to act as the attacking nucleophile and hydrate the carbonyl carbon of the substrate, breaking the peptide bond and restoring the active site to its initial protonation state (33). In the second mechanism (Figure 7A) a nucleophilic atom from amino acids side chain is used to hydrolyse the amide bond (53). The unique specificity of the HIV-1 protease and its critical role in viral replication has made it a key target for the HIV/AIDS drug therapy. Various protease inhibitors have been developed and are currently being used, they will be discussed in detail.

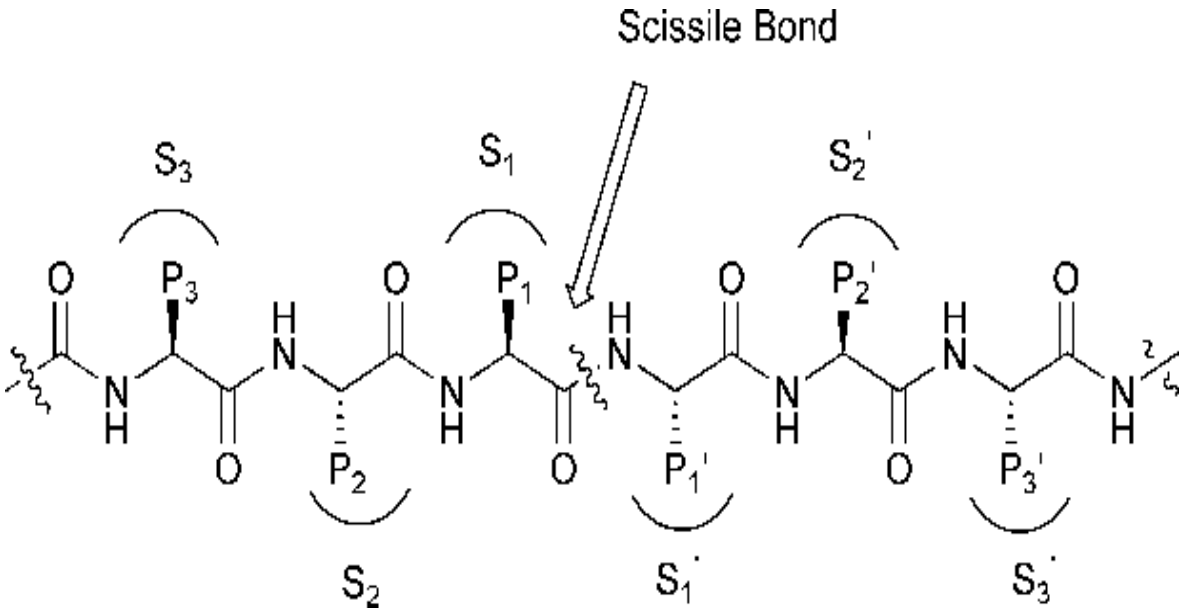
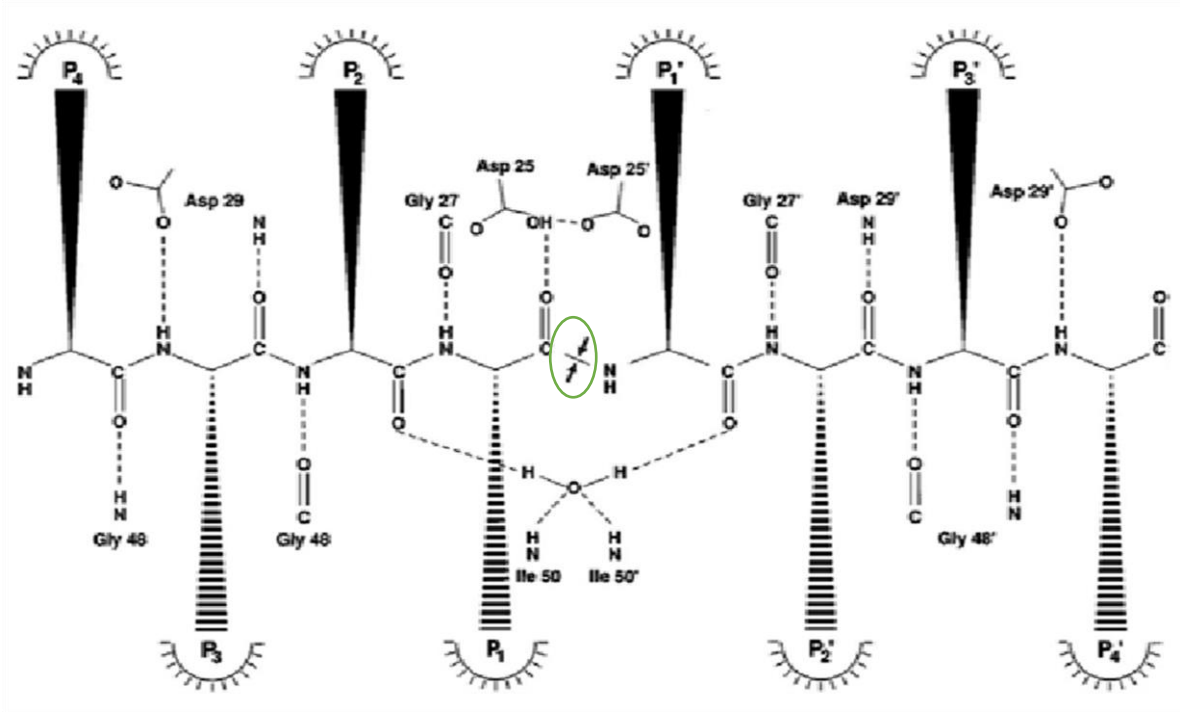


Figure 6. Nomenclature of the active site**.** The protease subsites are depicted as S3-S3' and the corresponding substrate subsites as P3- P3' the subsites are numbered in respect to the scissile bond represented by an arrow it where cleavage occurs. (Figure taken from Brik and Wong, 2003) (54).

A.



B.

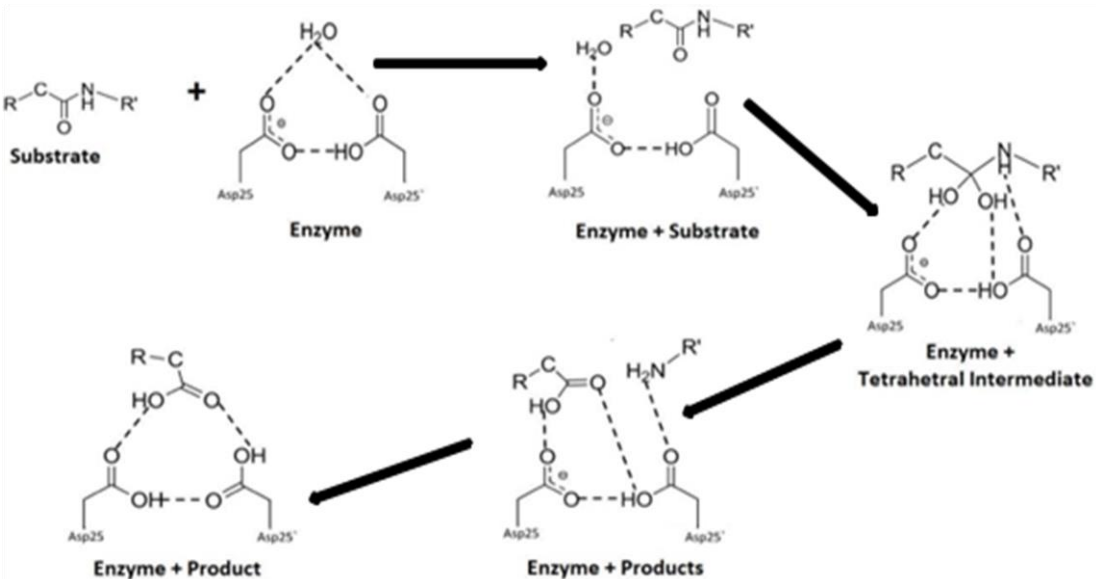


Figure 7. The catalytic mechanism of the protease**.** (A) Shows the reaction mechanism using a nucleophilic atom from amino acid side chains with Asp 25 and Asp 25' being the catalytic residues, while the green circle represents the scissile peptide bond (Figure taken from Wlodawer and Vondrasek, 1998) (53). (B) Shows the hydrolysis of the substrate scissile bond using an activated water molecule resulting in a free enzyme and two products being formed (Figure taken from Shen *et al*., 2012) (55).

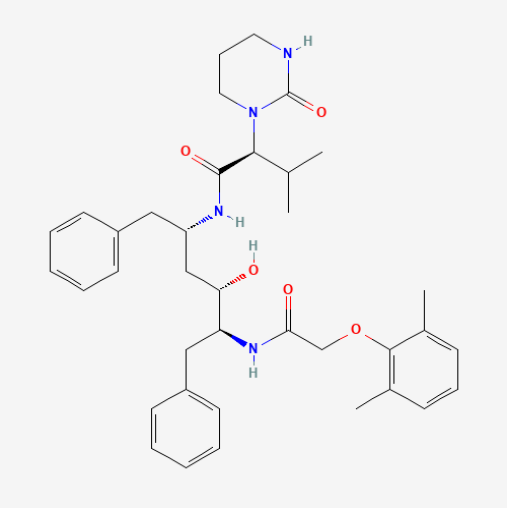
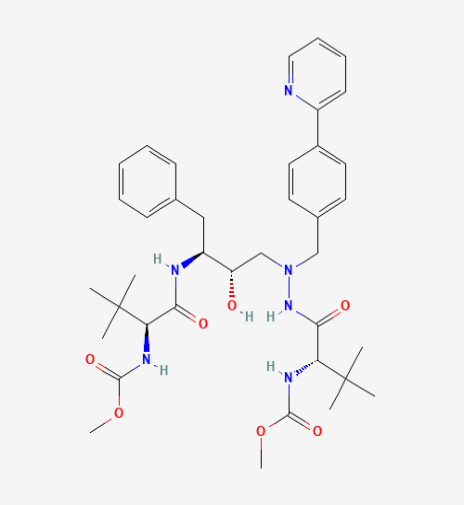
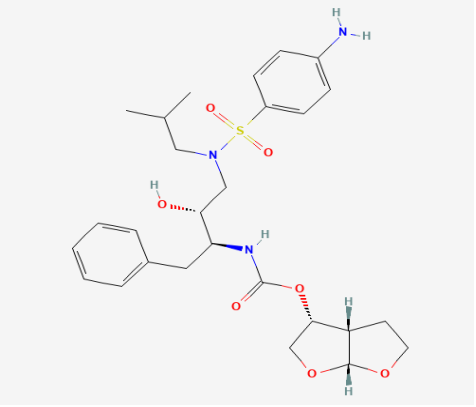
1.5 HIV-1 protease inhibition

The development of HIV protease inhibitors (PIs) was largely derived from information provided by aspartic PIs intended to inhibit the effects of angiotensinogenase (56). Most of these aspartyl protease inhibitors had non-cleavable inserts in place of the scissile-bond amino acid sequence and were peptidomimetic substrate-based compounds (57). There are ten protease inhibitors approved by the Food and Drug Administration (FDA) that are on the market and depending on their structural features are classified as first or second-generation protease inhibitors. First generation protease inhibitors like saquinavir and indinavir were demonstrated to have high binding affinity for the protease through increased hydrophobic interactions (45). The first PI to receive FDA approval was saquinavir, which has an isostere of hydroxyethylamine instead of a scissile bond, ritonavir and indinavir quickly followed (58). Amprenavir, fosamprenavir and nelfinavir are other three first-generation protease inhibitors and nelfinavir was the first PI with the ability to be administered in both adults and children (58).

First-generation PIs made a significant impact, but they also had limitations such as decreased clinical efficacy due to limited absorption, quick excretion, and a high burden of adverse effects (25, 58). The development of second-generation PIs also came about due to high level of drug resistant mutations associated with these PIs (45). The design of the second-generation PIs was mainly to target drug resistant strains, improve bioavailability and reduce side effects (58). These PIs have an improved polar interaction with main chain residues while decreasing the size of hydrophobic groups to minimise the impact of active site mutation on the inhibitor (45). This group comprises of atazanavir, darunavir, lopinavir and tipranavir. PIs darunavir, atazanavir, and lopinavir co-administered along with a ritonavir booster in HAART are the first line of treatment for the HIV infection, for this study the focus will be on these three second generation PIs, and they are briefly described on Table 1 below.

Although HAART has been effective in treating HIV, the emergence of drug resistance to PIs has hampered treatment efficacy, leaving few antiviral choices for patients who contract the resistant strains of the virus (32). To control the HIV infection there is a need for new approaches in the design and development of novel PIs that have a high genetic barrier against resistance.

Table 1. HIV-1 protease inhibitors and their properties

Generic name	Abbreviation	Properties	Structure*
Lopinavir	LPV	Developed by Abbott laboratories and FDA approved it in 2000 it is based on improvement of structural features for RTV (59). The core region of LPV contains a hydroxyethylene dipeptide isostere, the P2 pocket in RTV has it 5-thiazolyl group replaced by a phenoxyacetyl group (60). Also, a six-member cyclic urea replaces the 2-isopropylthiazolyl group of the P2' thus increasing potency (K_i = 1.3 pM) through this bulky P2 and P2' replacement (54)	
Atazanavir	ATV	Approved in 2003 and developed by Bristol-Myers Squibb, has a K_i of 10 pM it has an aza-di-peptide core with 2-hydroxy-1,3-di-aminopropane transition state isostere makes up the ATV structure. The removal of one of the chiral centers is facilitated by the aza-dipeptide connection between the P1-P2 pockets enabling a straightforward and large-scale synthesis (61)	
Darunavir	DRV	Approved in 2006 and developed by Tibotec Inc it is an APV structural homolog with DRV containing a bis-THF moiety which improves the hydrogen bond interaction at the protease active site thus slowing down drug resistance (62)	

*All chemical structures for the PIs were downloaded from PubChem [21/02/2023]

1.6 Resistance to protease inhibitors

Protease inhibitors have a high genetic barrier to drug resistance, despite resistance being associated with all drugs being used in the treatment of HIV. This is mainly because there are numerous advantages associated with HIV protease including its size and dimeric property. Its small size will limit mutation possibilities and the dimeric characteristic will require the numerous mutations prior to being PI resistant (33, 63). Nevertheless, when compared to other ARV classes, PIs have the highest mutation frequencies (32). In individuals infected, it has been estimated that within the HIV protease the daily occurrence of a single possible mutation is 10^4-10^5 times. There is a link between genotypic variations in the PR gene and the emergence of clinical resistance. Frequent mutation selection occurs due to the mutant protease being more fit compared to the wild-type in the presence of the inhibitor (33). The resistant strain development is due to RT lacking the proofreading activity resulting in a mismatch rate of 10^{-3}-10^{-5} nucleotide bases per cycle, the virus's rapid replication cycle which generates 10^{10} virions per day, its capacity for genomic recombination and drug pressure (25). The degree of cross-resistance among different PIs rises as the number of accumulated mutations rises (33). Depending on the mutation location within the PR it can either be a primary (active site) or secondary (non-active site) mutation. Mutations conferring drug resistance occur at 45 residues of the 99 residues and from those only 11 are linked to the active site and others are secondary mutations (64).

1.7 South African subtype C protease

The currently used PI in antiretroviral therapy are designed with improved binding affinity to HIV-1 subtype B protease instead the subtype C protease which is causing rampage worldwide. Drug resistance may develop more quickly in non-B subtypes due to the high occurrence of some resistance mutations in these subtypes, and some non-B subtypes may already be intrinsically resistant to some antiretroviral drugs if not all of them. The South African subtype C protease consist of eight NOPs these are substitution amino acid found in each monomer, enabling distinguishing between the subtype C and subtype B. In the fulcrum region three mutation are found these are T12S, I15V and L19I substitution mutations, M36I and R41K are in the hinge region, H69K and L89M are positioned in the 60s and 80s loops. I93L forms the α helix and because polymorphisms occur distal to the active site, they are said to have an indirect effect on the binding of the inhibitor. Subtype C protease thermodynamic analysis with PIs have depicted a weaker binding affinity in comparison to subtype B proteases (65). The HIV-1 subtype C conformational stability is moderately reduced in contrast to the subtype B indicating that changes in structure may in turn affect PI binding to the subtype C protease (8). Studies on HIV-1 subtype C protease previously conducted in the lab by Naicker and colleagues

(43, 66) also demonstrated that NOPs in the subtype C had an effect on drug binding, as the binding affinity was weak for some PIs and vitality studied indicated a selective polymorphisms advantage for subtype C relative to subtype B.

1.8 Gene cloning

The understanding of the chemical structure and activity of cells at a molecular level has greatly benefited from gene cloning studies (67). It is a fundamental tool in molecular biology that has enabled researchers to study the function of specific genes and their encoded proteins. This method typically involves transferring the desired DNA fragment from one organism to a self-replicating medium, such as a bacterial plasmid. Once inside a suitable host cell the vector proliferates, producing multiple identical copies of not only itself but also the inserted gene of interest to generate a new DNA fragment called recombinant DNA (68, 69). During host cell division, the recombinant DNA is further replicated in progeny cells resulting in the formation of colonies or clones which can be used in sequencing, genotyping, mutagenesis and heterologous protein expression (70, 71). By cloning a gene of interest and expressing it in a heterologous host, researchers can obtain large quantities of the encoded protein for downstream applications, such as protein purification, structural studies and drug discovery.

1.9 Overexpression, purification, and characterisation

The production of recombinant proteins in microbial systems has had a significant impact on molecular biology (71). It allows for the extraction and purification of a large quantity of a desired recombinant protein, which provides a wide range of applications including its use for research, diagnosis and treatment of diseases. To produce recombinant proteins, a gene of interest is cloned downstream of an expression vector promoter, and recombinant proteins are synthesised as a result. The blueprint for proteins is stored in DNA and is decoded through tightly regulated transcriptional processes to produce messenger RNA (mRNA), which is then translated into polypeptide chains and undergoes post-translational modifications to form proteins (72, 73).

After the successful expression of recombinant proteins, it is essential to purify the protein for various applications. Several steps are involved in isolating a specific protein from a complex mixture of proteins during the purification process (74), these steps may take an advantage of protein solubility allowing a part of the mixture to precipitate through alteration of some solvent properties, while others include taking advantage of the adsorption principle as observed in column chromatography (75). Depending on nature, size and the final goal for the purification there are numerous chromatography techniques that can be used (76). These techniques may include the use of Ion

exchange chromatography (IEC), which separates proteins based on their net surface charge by ionic interactions occurring between a protein and a charged stationary phase (77). Others may use hydrophobic interaction chromatography (HIC) (78) which separates proteins according to the difference on their hydrophobicity. Also affinity chromatography (79) can be used and will separate proteins by the interaction between a specific immobilised substrate (80) and a peptide "tag" (like GST and Histidine tags) adhered to the recombinant DNA during expression, and size exclusion chromatography (SEC) (81) which separates proteins through a porous gel based on their hydrodynamic radii. In protein purification of recombinant HIV-1 protease ion-exchange and affinity chromatography are widely used (82).

One of the most important aspects in recombinant protein expression is the identification and the characterisation of the purified protein. Briefly, quality control of purified protein must not be overlooked as the correct interpretation of various experiments on structural and biophysical characterisation depends on the premise that the protein samples are homogeneous and pure, that their concentration has been precisely assessed and that all the protein is in the active native state and completely solubilised (83). Numerous techniques can be used for protein quality control which include the most widely used technique SDS-PAGE which separates molecules according to their subunit molecular weight (84) and spectroscopy. Studying enzyme kinetics, thermodynamics and three-dimensional structure of proteins is necessary for the development and design of effective drugs.

1.10 Enzyme kinetics

One of the most important features of enzymology is catalytic activity, the study of enzyme catalysed chemical reactions is the basis of enzyme kinetics. In which valuable information about the enzyme catalytic and inhibition mechanism can be obtained (85). Enzymes speed up reaction rates by decreasing free energy of activation and have a high specificity for the reaction they catalyse (86). Briggs and Haldane (87) proposed basic principles to model kinetics involved in understanding enzymes, they proposes that at substrate saturating conditions the rate of reaction will not be increased by increasing substrate amount as the concentration of enzyme-substrate complex remains the same. In the reaction:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} E.S \xrightarrow{k_2} E.P \qquad \text{(Equation 1)}$$

E and S are free enzyme and substrate while E.S, and P represent enzyme-substrate complex, and product, The association rate constant of enzyme to substate binding is denoted as k_1, while the dissociation rate constant of the enzyme-substrate complex is denoted as k_{-1} and k_2 represents the rate constant of product and enzyme formation. When the concentration of enzyme-substrate complex remains the same over time, the reaction is said to have achieved steady state making it possible to measure product formation as well as determine enzyme kinetic parameters. Since the enzyme concentration remains constant in the sample, the formation of enzyme-substrate complex is dependent on substrate concentration. Michaelis-Menten equation links reaction rate to substrate concentration by kinetic parameters K_M and V_{max} (88)

$$V_0 = \frac{V_{max}[S]}{K_M + [S]} \qquad \text{(Equation 2)}$$

where V_0 (initial velocity) is the product increase rate with time, while maximum velocity (V_{max}) is when there is complete saturation of enzyme with substrate. Michaelis-Menten constant (K_M) is the substrate concentration [S] needed for half-maximal velocity to be reached and represents binding affinity of enzyme to substrate. During enzyme kinetic experiments fluorometric assays are performed using artificial fluorogenic substrate (89) and will be used for this study.

1.11 Thermodynamics studies

The crucial stage in drug discovery is understanding the mechanisms of protein-drug binding and obtaining precise information on the equilibrium thermodynamics of the interaction (90, 91). The term thermodynamics in drug discovery refers to studies of heat change that takes place upon interaction of biomolecules (92). General thermodynamics govern a non-covalent association of two macromolecules, where K_a the binding affinity is due to the change in free energy upon binding and like with any other spontaneous processes binding only occurs when it is associated with a negative Gibbs free energy of binding (ΔG). This in turn consist of both components of enthalpy (ΔH) and entropy ($-T\Delta S$) given by:

$$\Delta G = \Delta H - T\Delta S = -RT\,\ln K_a \qquad \text{(Equation 3)}$$

where T is the absolute temperature in Kelvin and R is the gas constant (8.314 $J.mol^{-1}K^{-1}$)(93). The use of this thermodynamic profile summarises and reflects on the interplay between various intermolecular interactions involved in drug target complex formation (90). The contribution of

enthalpy to free energy reflects the specificity of the interactions between partners which include halogen, ionic, hydrogen bonds, van der Waals and electrostatic interactions, whilst the entropic contribution is the measure of the overall system dynamics (93). The Simple thumb rules of thermodynamics state that a favorable enthalpy is due to hydrogen bonds with hydrophobic interactions resulting in favorable entropy, and that changes in conformation are associated with unfavorable entropy (92). The *ΔH* which is associated with the first law of thermodynamic measures the flow of heat at constant pressure between a system and its surroundings. A positive value indicates an endothermic reaction while a negative value denotes an exothermic reaction. The *ΔS* associated with the second law of thermodynamics measures the degree of randomness of the system where a positive value corresponds to an increase entropy and a negative value denote a decrease in entropy. Favorable or spontaneous processes are the ones with a negative enthalpy change and a positive entropy change. Since entropic and enthalpic contributions are correlated, enthalpy increase by tighter binding may directly impact entropy by mobility restriction of interaction molecules (94). Regarding this enthalpy-entropy compensation phenomenon, the energy of binding associated with ligand protein interactions varies from the individual sum contributions to free binding energies (93). Isothermal titration calorimetry (ITC) tool is used for the complete understanding of the full thermodynamic picture of binding reactions (95).

1.12 Three-dimensional structure of proteins

The native conformation of proteins is critical for their biological function as it determines how it will interact with other molecules in the cell. The proteins primary amino acid sequence with factors such as disulfide bonds, hydrogen bonding and van der Waals forces dictates the 3D structure of proteins (96, 97). The study of protein structures has led to significant advances in the field of drug discovery. Researchers can use knowledge of a protein's three-dimensional structure to design drugs that target specific binding sites on the protein. Understanding the precise shape of a protein's active site at the atomic level enables researchers to design small molecules that bind and inhibit protein activity, thus blocking its function (98). This is the approach that has been used in the development of a wide range of drugs with X-ray crystallography being the most preferred method for the elucidating the protein’s 3D structure accounting for 87 % of 3D protein structures reported so far (99).

1.13 Techniques in structural and functional characterisation of proteins

Currently several methods are being used to characterise protein structure, determine thermodynamics and enzyme kinetics of proteins. Protein biochemistry research requires sensitive, fast, simple and reliable techniques when performing this kind of research, which is why scientists are

always pushing the envelope when it comes to the development of such techniques. Some of the commonly used techniques include the ones that use light, i.e., spectrophotometry, spectrofluorimetry spectropolarimetry, liquid chromatography, X-ray crystallography and a modern technique which makes use of heat i.e., ITC.

1.13.1 Spectrophotometry

The spectrophotometry technique (100) is used to determine quantity of light absorbed by a chemical substance through light intensity measurement as a light beam passes through the sample. A spectrophotometer instrument enabling work at UV range was initially developed in 1940 by Beckman and colleagues (101) is used for measuring reflected or transmitted visible light, UV light or infrared light and measures intensity as the function of wavelength of the light source. The monochromator permits light of specific wavelength to pass through the sample allowing detection by a photodetector, resulting in an electric signal corresponding to light intensity. During the experiment positioning the cuvette the same way in the light beam is important to prevent variations caused by differences in glass composition. Some of the applications of UV-Vis absorption spectrophotometry (Figure 8B) in protein evaluation include qualitative analysis, quantitative analysis, enzyme assays and physiochemical studies.

1.13.2 Spectrofluorimetry

Fluorimetry is a rapid, highly sensitive and specific tool that is widely used for analysis of dynamics, structure and protein interactions in solution. UV absorption or radiation of visible light by molecules (fluorophores) results in electron transition from a ground state to an excited state, because this state is unstable electrons return to the ground state emitting excess energy as a photon of lower wavelength (102). The released photons can be detected by measuring emission intensity using a fluorometer. The shift in energy level of the excited and emitted energy is termed stokes shift and intrinsic or extrinsic fluorophores can be used. Intrinsic fluorescence (Figure 8A) entails the use aromatic amino acids (tyrosine, tryptophan and phenylalanine) within proteins. The absorption and emission wavelengths of phenylalanine are short. Based on solvent polarity, tyrosine and tryptophan can be excited at wavelengths between 280 nm and 285 nm, with tryptophan having a maximal emission of 350 nm in water while tyrosine has wavelength of 304 nm, and the emission of tyrosine can be avoided by selectively exciting tryptophan at 295 nm. Extrinsic fluorescence uses numerous fluorescent probes that can bind to a molecule (103). Also, the interaction of biomolecules in solution can be studied using fluorescence quenching which can be defined as the physiochemical process that results in the reduction of fluorescence intensity for a given molecule (104).

1.13.3 Far-UV circular dichroism spectroscopy

Far-UV CD spectroscopy (105) can be used for the determination of secondary structural content and monitoring of conformational changes of proteins in aqueous solution (Figure 8C). This non-destructive technique measures absorption difference of left and right circularly polarised light, chiral molecules (e.g., proteins) will absorb circularly polarised light in one direction thus allowing for quantification of absorption difference between the two components. In the far-UV region (240 nm - 190/180 nm), there are two electronic transitions: one from n $\rightarrow \pi*$, which is sensitive to the formation of hydrogen bonds and appears as a negative band at approximately 220 nm, and another from $\pi \rightarrow \pi*$, with a positive absorption observed at around 190 nm and a negative band at approximately 210 nm. The above mentioned provide insight on the composition of protein secondary structure as the peptide bond is the principal absorbing group and distinct spectra are demonstrated by various forms of regular secondary structure obtained in peptides and proteins (103, 106). Proteins that are predominantly α-helical have a positive band at 193 nm and two negative ones at 222 nm and 208 nm while β-dominant proteins exhibit a pronounced positive band at 195 nm and a trough between 215-225 nm. The β structure CD signal comprises of various types of β-sheets arrangements and turns in comparison with the α-helix signal its less intense. Although far-UV CD cannot determine three-dimensional structure of proteins it does have some advantages including its rapidness, possibility of sample recovery and small sample requirements.

1.13.4 Size-exclusion high performance liquid chromatography

SE-HPLC (107) can be used for high order structural characterisation of proteins through sizing as it allows for separation of molecules based on their size. In this technique the exclusion of large molecules from the pores of the column packing material by mobile phase is much faster, compared to small molecules which largely interact with the stationary phase of the column and elute late. Time taken for elution of each analyte is expressed as the retention time. The mode of detection when a sample travels through the column is predominantly UV. The amount of light absorbed is related to the concentration of specific compounds in a sample. For accuracy in analysis a dual wavelength detection system is preferred. This allows for purity profiling as longer wavelengths are sensitive to aromatic amino acids and are used for protein measurements (108). Shorter wavelengths provide great sensitivity for aggregates and other impurities (108, 109). The size of unknown proteins can be estimated from a calibration curve derived from molecular weights of known proteins (108). Even though systems of SE-HPLC and FPLC can accommodate silica-based SEC columns due to their adequate pressure thresholds they may still require to be coupled with light scattering detectors.

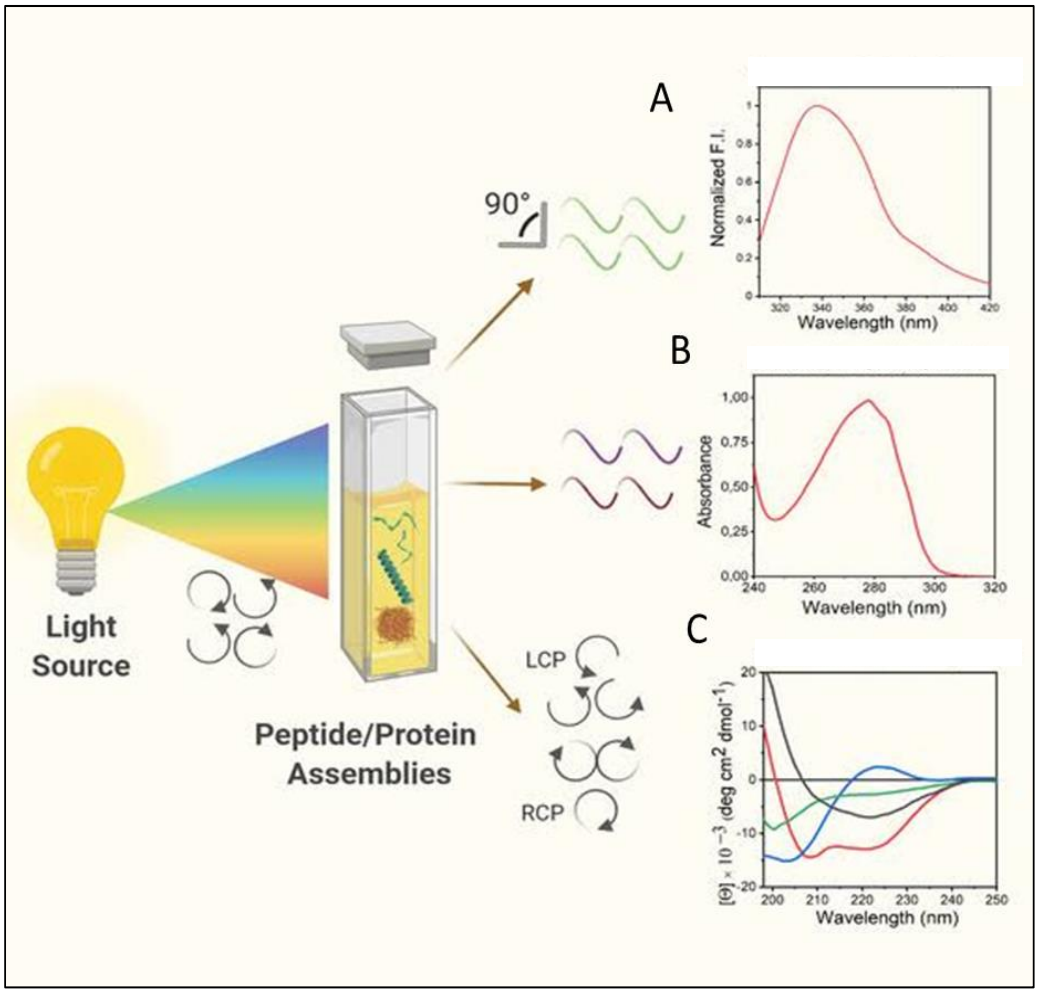


Figure 8: A schematic overview of techniques in protein analysis.(A) Tryptophan fluorescence emission spectrum, (B) UV-Vis absorption spectrum and (C) Far-UV CD spectra. These techniques use light in different ways to study protein properties and function (Figure taken from Pignataro *et al*., 2020) (103).

1.13.5 ITC

Isothermal titration calorimetry determines the amount of heat released or absorbed when two molecules interact and allows for binding characterization between two molecules. This label-free, non-immobilised technique has a unique feature of not only providing the binding affinity (K_a) of the interaction, but also the thermodynamic parameters (Gibbs free energy (ΔG), (enthalpy (ΔH) and entropy (ΔS) and binding stoichiometry (N) in a single experiment (110). Other binding assays do not provide as much thermodynamic information and may require additional techniques. The ITC instrument (Figure 9) has a reference cell and a sample cell which are in thermal equilibrium. During a stepwise titration of ligand into protein solution in the sample cell, heat is either absorbed or released. The heat change is quantified by determining the amount of power required to sustain the same temperature in respect to the reference cell (111, 112). The amount of this power is measured as peak signals and to obtain the required heat the area under individual peaks is determined. The binding of ligand to protein corresponds to heat released or absorbed (113) and upon protein saturation with ligand the reaction heat signal approaches zero and the observed signal is due to heat of dilution. Successive accumulation of reaction heats shows a saturation curve which enables the deduction of K_a, ΔH and N directly from the plot, once these parameters have been determined ΔG and ΔS can be calculated from equation described in section 1.11. The evolution of traditional ITC to a more robust widely used method in protein science has seen the development of the new ITC methods such as reverse titration ITC and the displacement ITC (114).

1.14 X-ray crystallography

X-ray crystallography remains a preferred method for determining the molecular and atomic structure of a protein crystal. The discovery of X-rays in 1985 is attributed to Wilhen Conrad Rontager (99) with the determination of the first three-dimensional of protein in 1958 (115). A pure protein sample is crystallised at high concentration and subjected to X-ray beam upon crystallization (Figure 10). The crystal information can be obtained from the pattern of the diffraction spots with the intensity of spots being used to deduce structure factors which will make it possible to calculate the electron density map. To improve this map quality numerous methods can be used and once a sufficient map quality has been achieved the protein sequence is used to construct a molecular structure that is further refined to precisely fit the map and assume a thermodynamically favoured conformation (116). Even though protein crystallography is well understood to predict conditions of protein crystallization still presents a challenge. Therefore, the approach is based on trial-and-error method with screening of a wide range of chemical conditions and commercially available screening kits simplifies the screening process.

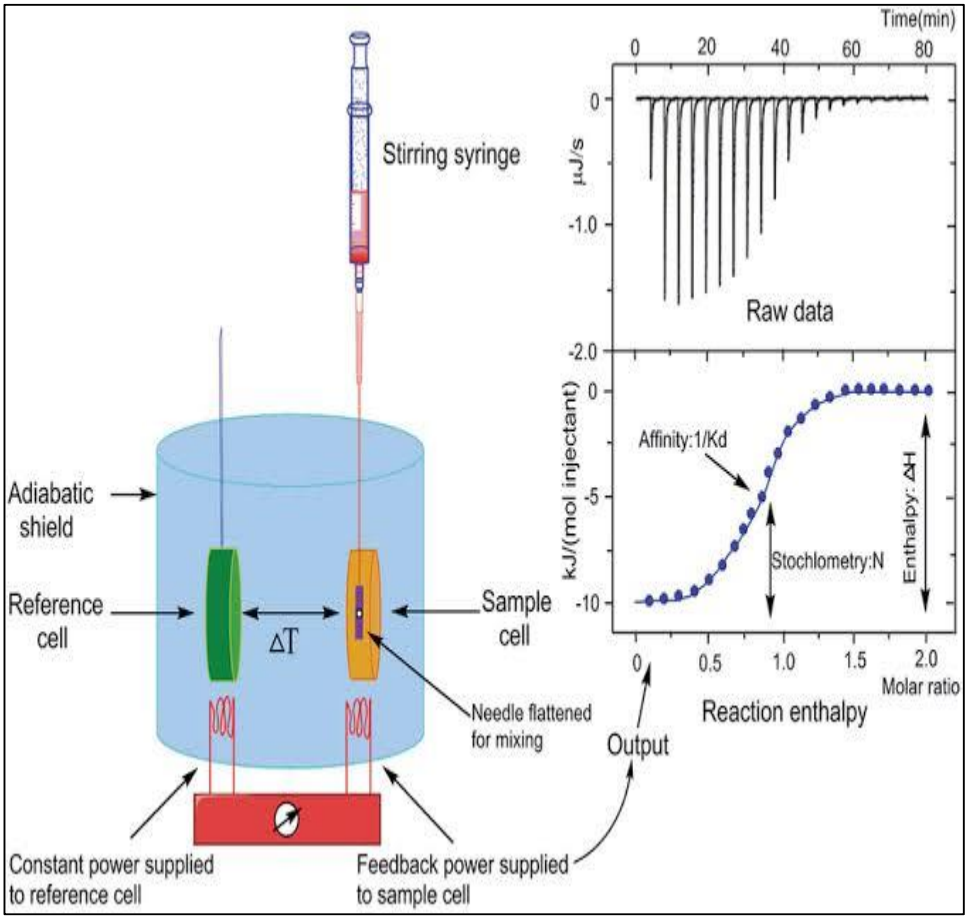


Figure 9: A schematic ITC presentation. Instrument components (left), raw ITC data from the titration experiment (top-right) and fitted ITC data showing deduction of parameters directly from the plot (bottom-right) Top right image is the titration thermogram indicating heat generated after each injection of the ligand into the protein (in black). The bottom right picture is the relationship between the heat released during each injection and the ratio of the total ligand concentration to the total protein concentration present. The circles represent experimental data points. (Figure taken from Song *et al.*, 2015) (117).

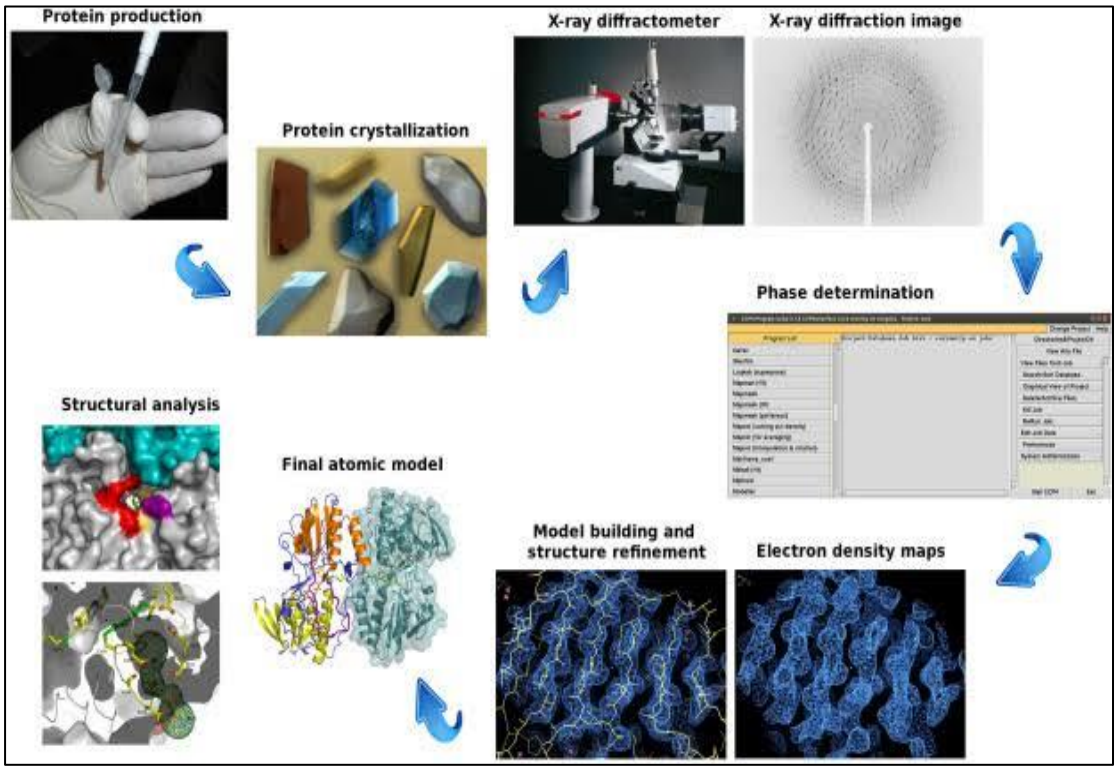


Figure 10: Steps involved in X-ray crystallography. elucidation of the three-dimensional structure of a protein. The steps include the production of pure target protein, which is crystalised and subjected to diffraction. The collected diffraction data are processed, and phases of the diffracted X-rays are determined. The electron density map of the protein molecule within the crystal is calculated from the phases. This is followed by building the protein model into the electron density map and refining the model. The refined model is rigorously validated to ensure its accuracy and reliability. Numerous validation tools are used to assess the structure and structural analysis is performed. Figure taken from (Brito and Archer, 2020) (118).

1.15 Aim

The aim of this study is to perform structural and functional characterisation on HIV-1 C-SA protease and assess the potency of atazanavir, darunavir and lopinavir on the protease while elucidating the three-dimensional structure of the protease.

1.16 Objectives

1.16.1 To overexpress and purify the recombinant HIV-1 C-SA protease in *E. coli* BL21 (DE3) pLysS cells

1.16.2 To perform qualitative and quantitative assessment of the protease using UV absorbance spectroscopy and ITC

1.16.3 To use far-UV circular dichroism spectropolarimetry to estimate the secondary structure of the protease

1.16.4 To assess the tertiary structure of the protease using intrinsic tryptophan fluorescence

1.16.5 To assess the quaternary structure of the protease using SE-HPLC

1.16.6 To determine specific activity, K_M, k_{cat}, V_{max}, k_{cat}/K_M from the enzyme kinetic assays

1.16.7 To determine thermodynamic parameters (ΔG, ΔH, ΔS) and K_d of the protease PI (ATV, DRV and LPV) complex using displacement ITC

1.16.8 To elucidate the three-dimensional structure of the HIV-1 C-SA protease using X-ray crystallography

2. Materials and Methods

2.1 HIV-1 protease expression construct and strain

HIV-1 subtype C (HIV-1CSA) protease sequence (with Q7K to reduce autocatalysis) previously generated in our laboratory was recombinantly constructed into a pET-11a vector. This T7 promoter-based expression system conferring ampicillin resistance was synthesised by GenScript (USA). The origin site within the vector enables gene replication in the host cell. *E. coli* BL21 (DE3) pLysS cells from Novagen (USA) suitable for expression of toxic genes and conferring chloramphenicol resistance were used as the host strain. The strain consists of both the DE3 lysogen which upon IPTG induction expresses T7 RNA polymerase and the pLySs plasmid encoding T7 lysozyme which inhibits non-target protein expression. The regulation of T7 promoter system allows for controlled recombinant protein expression in BL21 (DE3) pLysS cells, wherein the IPTG addition induces expression of target gene.

2.2 Expression construct transformed into host strain

Heat shock method was used for the transformation of pET-11a-HIV-1CSA proteas vector into *E. coli* BL21 (DE3) pLysS competent cells. This entailed thawing competent cells on ice and directly adding the vector, gently flicking ensured a proper mixture. After ice incubation for 30 minutes the mixture was heat shocked for 45 seconds at 42 ^{o}C followed by a 5 minute ice incubation. As a recovery step after transformation super optimal broth with catabolite repression (SOC) medium (at 37 ^{o}C) was added and the mixture was incubated at 37 ^{o}C with shaking (250 rpm) for 1.5 hours. Transformed cells were grown on LB-agar plates (15% (w/v) agar, 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) NaCl) with appropriate antibiotics (100 µg. mL^{-1} ampicillin and chloramphenicol 30 µg. mL^{-1}) overnight at 37 ^{o}C. Confirmation of successful transformation was confirmed by plasmid DNA sequencing. A single colony was inoculated into 100 mL LB media (1% (w/v) tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) NaCl) supplemented with ampicillin (100 µg. mL^{-1}) and chloramphenicol (30 µg. mL^{-1}) the culture was allowed to grow overnight in the shaker at 37 ^{o}C, 200 rpm. Glycerol stocks were prepared from the overnight culture and stored at -80 ^{o}C.

Isolation of recombinant plasmid from the overnight culture was carried out using a GeneJET plasmid miniprep kit (Thermo Scientific, Massachusetts, USA) and sent to Inqaba Biotech (Pretoria, South Africa) for DNA sequencing. The received sequence was translated using ExPASy translate tool and Basic Local Alignment Search Tool (Blastx)(119) was used to perform sequence alignment with the sequence of PDB ID: 3U71 (43).The overnight culture was also used to assess the expression of gene of interest and its solubility using tricine SDS-PAGE under reducing conditions.

2.3 HIV-1 protease growth curve

The role of HIV-1 protease also involves cleavage of cellular proteins and contribution to HIV-induced cytotoxicity (120). It is therefore crucial to monitor its behavior within the cell and deduce an appropriate induction time for over-expression. The glycerol stock was used to inoculate 100 mL LB media supplemented with appropriate antibiotics and allowed to grow overnight at 200 rpm, 37 ^{o}C. A 100-fold dilution of the overnight culture in LB was done and cells allowed to grow for 6 hours with OD_{600nm} measurements being taken at zero time point and every 20 minutes for 6 hours. Readings at OD_{600nm} were plotted against time in minutes and from the growth curve mid-log phase was deduced. This is an appropriate time for induction as cells are in their exponential growth and will lead to the over-expression of gene of interest.

2.4 HIV-1 protease induction trials

The insolubility and low expression of HIV-1 protease has hindered its large-scale production and in turn made downstream applications a strenuous process (121). Induction trials were performed with different IPTG concentrations to decide on the concentration to use for HIV-1 C-SA protease over-expression. LB media (100 mL) was inoculated with a glycerol stock and allowed to grow overnight at 37 ^{o}C, 200 rpm. The overnight culture was diluted 100-fold in 6 x100 mL LB media and allowed to grow (37 ^{o}C, 200 rpm) until log phase wherein different IPTG concentrations (0.2, 0.4, 0.6, 0.8 and 1 mM) were used to induce overexpression (OD_{600nm} ~ 0.6). At each time point (0,1,2,3,4,5,6, and 24 hours) 1 mL of cell culture was harvested at 10 000 xg, 4 ^{o}C for 5 minutes with supernatant being discarded. TE buffer (2 mM EDTA, 10 Mm Tris-HCl) was used to resuspend the pellet and sonication was done using Microson XL-2000 sonicator at 15 amps for 3 seconds in 3 cycles. Insoluble and soluble fractions were separated by centrifugation at 12 000 xg, 4 ^{o}C for 10 minutes followed by a 2:1 ratio resuspension in reducing sample buffer (30% (v/v) glycerol, 0.05% (w/v) Coomassie Brilliant blue G-250, 12% (w/v) SDS, 6% (w/v) BME and 150 mM Tris-HCl, pH 6.8).

2.5 Tricine SDS-PAGE

Tricine SDS-PAGE which is an adaptation of the Laemmli method (122) was used to assess protein purity and its molecular weight. The technique allows for high resolution separation of proteins with low molecular mass in a single polyacrylamide gel. The tricine in the stacking gel allows for stacking of proteins with lower molecular weights making the method more efficient (122). The resuspended pellet and supernatant were diluted in reducing buffer in a 1:1 ratio and boiled (5 min, 100 ^{o}C) prior to loading 20 µl on the gel. Tricine SDS-PAGE gel had 4% stacking and 16% separating gel and a cathode

[10x (1 M Tris-HCl, 1M Tricine and 1% (w/v) SDS pH remained unaltered as 8.25)]-anode [10x (1 M Tris-HCl pH 8.9)] buffer system was used for electrophoresis this was done at an initial voltage of 30 V for proper stacking and a final voltage of 140 V. Gel visualization was carried out using Coomassie staining (0,025% (w/v) Coomassie Blue G-250, 10% (v/v) acetic acid and 50% (v/v) methanol). The gel was then distained in 10% (v/v) acetic acid and 50% (v/v) methanol, known protein standards from the Pre-stained Protein Ladder (Thermo Scientific, USA) was used as the molecular size marker and a plot of log molecular weight of standards against relative migrated distance (R_f) was used to estimate the monomeric size of the HIV-1 C-SA protease.

2.6 Overexpression of HIV-1 protease

Upon successful determination of overexpression conditions, the overnight culture prepared by inoculating glycerol stock into 100 mL LB media supplemented with ampicillin (100 µg. mL^{-1}) and chloramphenicol (30 µg. mL^{-1}) and incubated overnight at 37 ^{o}C, 200 rpm was diluted a 100-fold in 8x650 mL LB media supplemented with appropriate antibiotics. Cells were grown (37 ^{o}C, 200 rpm) to mid log phase (OD_{600nm} ~ 0.6) for induction with 1 mM IPTG and gene overexpression was allowed to occur for 4 hours at 37 ^{o}C, 200 rpm. This was followed by harvesting the cells through centrifugation at 5000 xg, 4 ^{o}C for 10 minutes, the resulting supernatant containing media was discarded and 20 mL ice-cold buffer A (10 mM Tris-HCl, 1 mM PMSF and 2 mM EDTA, pH 8.0) was used to resuspend the cell pellet which was then short term stored at -20 ^{o}C. The cell pellet was thawed to aid lysis followed by further resuspension in ice-cold 30 mL buffer A with 10 mM $MgCl_2$. A Bio-Gen series PRO200 homogenizer (Pro Scientific Inc., Connecticut, USA) was used to homogenise resuspended cells and DNase (1 mg/mL) was added to increase protein extraction efficiency by removing DNA allowing reduction in viscosity this was done on ice while stirring the cell suspension (30 min). The cell suspension was then subjected to 30 sec sonication cycles (x10) at 70 µm amplitude whilst on ice using Microson XL 2000 sonicator. To separate soluble and insoluble fractions the cell lysate was centrifuged (4 ^{o}C, 18000 xg, 30 min) and target protein was expressed as inclusion bodies. The resulting pellet was washed by re-suspension in ice- cold buffer A^{+} (buffer A with 1% Triton X-100) and centrifuged (4 ^{o}C, 18000 xg, 30 min) the pellet was washed twice.

2.7 Inclusion body isolation and purification

The purification procedure used followed the one previously established in lab by Naicker *et al,.(43)* briefly the target protein was recovered from inclusion body pellet by unfolding at room temperature for 1 hour using buffer B (8 M urea, 10 mM Tris-HCl, 5 mM DTT, pH 8.0), followed by centrifugation (18000 xg, 30 min, 20 ^{o}C). The supernatant was subjected to acidification through dialysis at 4 ^{o}C using

a 3.5 kDa Snakeskin dialysis tubing (Thermo Fischer Scientific, Massachusetts, USA) in buffer C (10 mM Formic Acid, 10 % Glycerol and 0.01% (w/v) sodium azide) for 4 hours and refolded through dialysis at 4 ^{o}C in buffer D^{+} (10 mM sodium acetate, 2 mM DTT, 0.01% (w/v) sodium azide, pH 5.0) overnight. The refolded protein was centrifuged (18000 xg, 4 ^{o}C, 30 min) to remove aggregated proteins and filtered using cellulose acetate syringe filter (0.45 µM). The AKTA start protein purification system (GE Healthcare Life Sciences, Chicago, USA) was used for purification. The estimated pI (9.32) of HIV-1 protease (ExPASy ProtParam tool) and buffer pH allowed binding of loaded protein (2 mL/min) on HiTrap CM Sepharose FF (5 mL) cation exchange column that was initially equilibrated with 5 column volumes of buffer D^{+} (4 mL/min). Elution of bound protease was done using a 0-1 M NaCl gradient at a 3 mL/min flow rate and dialysed into storage buffer D (buffer D^{+} minus 2 mM DTT) overnight at 4 ^{o}C. After protein dialysis a 3.5 kDa stirrer Amicon concentrator (MilliporeSigma, Massachusetts, USA) was used to concentrate the protein sample and stored at -80 ^{o}C for further analysis and downstream applications.

2.8 HIV-1 protease characteristics

2.8.1 UV-Vis spectroscopy for protein quality and quantity

Jasco V-630 UV-Vis spectrophotometer was used to assess protein quality and quantity, HIV-1 protease has four cysteine and four tryptophane residues allowing for maximum light absorption at 280 nm. Assessment of protein purity was done by diluting (10-fold) concentrated protein with the dialysate in a 1 mL reaction volume and monitoring absorbance (230 nm to 360 nm) at 20 ^{o}C using a quartz cuvette (micro-QS). This was done in triplicate, an average of the resulting spectra was corrected by subtracting for buffer components (baseline) and the absorbance data was plotted as a function of wavelength of incident light. Protein concentration was determined spectrophotometrically by performing a serial dilution and taking measurements at 280 nm and 340 nm. The average reading at 280 nm was corrected for scattering in the sample (from aggregates, dust and particles) by the subtraction of absorbance reading at 340 nm. A plot of corrected absorbance readings against dilution was constructed and a linear regression in conjunction with Beer-Lambert law was used with the protease extinction coefficient (25230 $M^{-1}\,cm^{-1}$) for the estimation of protein concentration. Since the purification process of HIV-1 protease entails the unfolding (inactive) and refolding (active) of the protein, results from UV-Vis absorption spectroscopy are rendered in adequate as the technique cannot distinguish between different forms of proteins .

2.8.2 ITC for active site titration

HIV-1 protease is active and functional in its native conformation (homodimer), its therefore paramount to determine the percentage of properly refolded protein and correct the concentration obtained from UV-Vis absorption spectroscopy. Nano ITC (TA instruments, New Castle, USA) was used to perform an active site titration of HIV-1 protease against acetyl-pepstatin (Merck, South Africa) as a ligand. This statin containing compound has been demonstrated to be an excellent inhibitor for dimeric HIV-1 protease (123) The experiment was conducted in a 1:10 ratio with a dilution of acetyl-pepstatin stock (8 mM) to 200 µM using storage buffer (buffer D) to avoid buffer mismatch and loaded on an ITC syringe (250 µL). HIV-1 protease (~ 20 µM) in storage buffer was loaded into the sample cell while reference cell contained milli-Q water. Samples were degassed (20 ^{o}C, 0.5 atm, 10 min) prior to loading and the loading of the syringe and both cells (precisely filled) was done with caution in order avoid bubbles as they interfere with the titration profile during the experiment. The run was performed at 20 ^{o}C by injecting acetyl-pepstatin into the sample cell with a stirring rate of 200 rpm. The initial injection volume was 2 µL and is regarded as the void volume since its purpose is to provide interface equilibration between the injection syringe tip and the solution surrounding the sample cell (110). The rest of the titration was continued till saturation was reached using 5 µL injection volumes and a 300 sec spacing in between them allowing the system with sufficient time to return to equilibrium.

The fitting and analysis of the calorimetric data was executed using NanoAnalyze software (TA Instruments) and the parameters were described for the independent model. The generated peak signals are automatically integrated by the software and the stoichiometry value (*N*) corresponds to the percentage of refolded protein. The binding ratio of Acetyl-pepstatin to the protease is 1:1 thus an N value of 1 indicates 100% of refolded protein. This percentage is used to correct for the active HIV-1 protease concentration and the new working concentration is used for further experiments. A titration of ligand against buffer D serves as a control measurement and is used for subtracting heat change due to the dilution of ligand into buffer D.

2.8.3 Protein secondary structure via far-UV CD

Jasco J-810 spectropolarimeter was used to determine the secondary structure of HIV-1 C-SA protease. Previous work on HIV-1 protease (43) demonstrated that the protein is predominantly β-sheeted and the CD spectrum is expected to have a strong positive band between 190 and 200 with a negative band between 215 and 225. HIV-1 C-SA protease concentration of 5 µM was used to perform the experiment (20 ^{o}C, data pitch of 0.2 nm, 0.5 nm bandwidth, 100 nm/min scanning speed, 8

accumulations) with a quartz cuvette (2 mm) over a wavelength range of 180 nm to 250 nm. The experiments were performed triplicates of the experiments were performed, the resulting ellipticity (ϑ) data was corrected for buffer components and normalised to mean residue ellipticity *(MRE*) by the equation:

$$MRE = (\vartheta \times 100)/cnl \qquad \text{(Equation 4)}$$

Where *c* is concentration of protein (mM), *n* represent the number of protein residues, *l* quartz cuvette pathlength (cm) (124).

2.8.4 Tertiary structure using Fluorescence spectroscopy

Intrinsic tryptophan fluorescence was performed on HIV-1 protease using Jasco FP-6300 spectrofluorometer. The tryptophan has high sensitivity to the emission spectra of its environment and therefore allows for its characterization relative to the local environment. When solvent exposure is decreased, or tryptophan is buried the emission spectrum is blue shifted (low wavelength) and red-shifted when the exposure is increased (123), the emission wavelength indicates if tryptophan is buried or solvent exposed. The excitation of aromatic amino acid is at 280 nm and 295 nm can be used for selective excitation of tryptophan which has a maximal emission of 350 nm in water and 304 nm for tyrosine (123). HIV-1 protease as a dimer has four tryptophan residues two in each monomer and two tyrosine residues one in each monomer (Trp 6, Trp 6', Trp 42, Trp 42' Tyr 59 and Tyr 59'). 2 µM of HIV-1 protease in buffer D was used to perform the experiment at 20 ^{o}C using excitation wavelengths of 280 nm and 295 nm with scanning speed of 500 nm/min between 270 nm and 450 nm. The scans were performed in triplicates, averaged and corrected for buffer components. The plot of fluorescence intensity in relation to wavelength was generated.

2.8.5 Determination of quaternary structure using SE-HPLC

The protein quaternary structure was verified by measuring the proteins retention time using UFLC (SE-HPLC) Shimadzu instrument (Tokyo, Japan). A 30 cm x 4.6 mm TSKgel SuperSW2000 column (Tosoh Bioscience, Tokyo, Japan) with a stationary phase composed of silica base material (4 µm particle size) connected to a TSKgel guard column was equilibrated with filtered and degassed buffer D comprising of 500 mM NaCl for 1 hour at 20 ^{o}C in a 0.2 mL/min flow rate. Protein sample (20 µM) of 20 uL was loaded into the column and a run was performed at 0.2 mL/min using a dual wavelength of 254 nm

(nucleic acids) and 280 nm (protein) for 30 min. A chromatogram of protein elution was detected at 280 nm showing absorbance readings against retention time (min). After column re-equilibration 20 µL of proteins standard mix (Sigma-Aldrich, USA) was loaded and analysed using the same conditions. The standard mix consisting of para-aminobenzoic acid, albumin, bovine thyroglobulin, γ-globulin and ribonuclease A was used to generate a calibration curve. The data was plotted using log molecular weight (Mr) against retention time to estimate the protease size. The linear regression function was used to fit the curve and the protease size was calculated using the retention time.

2.9 HIV-1 protease functionality

2.9.1 Enzyme kinetics via fluorescence spectroscopy

To perform enzyme kinetic studies of HIV-1 protease a fluorogenic substrate (Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala- Nle-NH_2) was used. The substrate mimics the KARVL/AEAM cleavage site between the capsid/p2 in the Gag-polyprotein precursor (8). The closeness of Abz (the fluorogenic 2-aminobenzoyl) and the quenching amino acid group Phe(NO_2) reduces the substrate's ability to fluoresce. The HIV-1 protease recognise and hydrolyse the substrate between Nle (norleucine) and Phe(NO_2) (nitro-phenylalanine) (125). The peptide bond cleavage between Nle-Phe(NO_2) results in increased fluorescence as the quenching effect is diminished (125). Enzyme kinetic assays were all performed by exciting the Abz group at 337 nm and measuring emission at 425 nm. Substrate stock solution (500 µM) prepared in 50% acetonitrile and reaction buffer (10 mM sodium acetate, 2 M NaCl, pH 5.0) were used to prepare a reaction mixture (800 µL) which also contained the protease and milli-Q. The experiments were done in triplicate using Jasco V-630 spectrophotometer at 20 ^{o}C, with fluorescence intensity being measured for 30 seconds at 1 sec data pitch, 0.5 sec of response time, excitation bandwidth (2.5 nm) and emission bandwidth of (5 nm). The conversion of fluorescence intensity to activity was done by measuring the intensity value that result upon complete cleavage of 1 nmol of substrate. All the obtained data was analysed using Microsoft excel software (Microsoft corporation, USA).

2.9.1.1 Michaelis-Menten kinetics

The hydrolysis of the fluorogenic substrate was monitored to determine the K_M and V_{max} parameters of the HIV-1 protease. This was done by varying substrate concentration (0-200 µM) under constant enzyme concentration (50 nM). A plot of the hyperbolic curve of enzyme velocity as a function of substrate concentration was used to determine kinetic parameters.

2.9.1.2 Specific activity and catalytic turnover number

Specific activity for HIV-1 protease and its catalytic turnover number were measured using substrate concentrations higher than K_M value. To perform the assay an increasing protease concentration (0-50 nM) was used with a constant substrate concentration (50 µM). The data was plotted for enzyme activity (µmol/min) against enzyme amount (mg) and fitted using linear regression to deduce the enzymes specific activity.

The measure of how rapidly the enzyme-substrate complex form product is termed the catalytic turnover number and denoted as k_2 in equation 1 (section 1.12). The k_{cat} value was obtained from the specific activity data and represents a maximum number of substrate molecules converted to product by a single enzyme per unit time. A plot of enzyme activity per µmol of enzyme was used for the determination of catalytic turnover number (sec^{-1}) using the slope of linear regression.

2.9.1.3 Catalytic efficiency

To compare how effective the protein is in catalysing the same reaction catalytic efficiency which is k_{cat}/K_M was measured (126). This measurement can only be done by using concentrations of substrate that are lesser than obtained K_M values. For the measurement the substrate concentrations were varied (0-10 µM) and enzyme concentration remained constant (50 nM). The resulting data was used to generate a plot of k_{cat} as a substrate concentration function, a function of linear regression was used for the fitting and from the resulting slope the determination of catalytic efficiency (μM^{-1} .sec^{-1}) was possible.

2.9.2 Thermodynamics by displacement ITC

The main goal in drug development is designing drugs that have high binding affinities for their target molecules. To measure this drug binding affinity ITC can be used, but it presents challenges once the binding affinity approaches and exceeds the nanomolar level (127). An experimental mode called displacement titration provides a solution to this limitation as it allows for a complete determination of thermodynamic binding parameters of a ligand down to a picomolar range (127). Two titrations are required by this approach, the first one is the titration of a low -affinity inhibitor to the target molecule followed by a displacement titration wherein a high-affinity inhibitor is titrated into an enzyme solution complexed with a low-affinity inhibitor (128). This is also called a competitive binding assay as both inhibitors compete for the same binding site within the enzyme (128). In some cases, reversed titrations are performed in ITC this is conducted by placing a macromolecule in syringe and the ligand in the sample cell (114).

Thermodynamic parameters of HIV-1 protease inhibitors were determined using displacement titration. A weak binding inhibitor (acetyl-pepstatin) complexed with HIV-1 protease is titrated with a strong binding inhibitor which will displace acetyl-pepstatin allowing for the determination of thermodynamic parameters of the strong binding inhibitors) (127). Three of the FDA-approved protease inhibitors (ATV, DRV and LPV) were prepared to 10 mM in 100% DMSO, during the experiment a reaction mixture (2 mL) of HIV-1 protease (20 µM), acetyl-pepstatin (200 µM) in 2% DMSO that was initially incubated at room temperature for 30 minutes was loaded into the sample cell and the PI of 200 µM in 2% DMSO that was prepared from the initial stock using buffer D was loaded in a 250 µL syringe and titrated into the sample cell. Displacement ITC was performed for ATV and DRV, whilst reversed displacement ITC was performed LPV. The resulting raw data was analysed using NanoAnalyze software (TA instruments) and data fitting was done using the independent model. Prior to data fitting the heats of dilution were subtracted, it is important to avoid buffer mismatch during the run by maintaining the same buffer conditions between the mixture in the sample cell and the syringe.

2.10 Three-dimensional structure using X-ray crystallography

2.10.1 Crystallisation and collection of data

To determine the three-dimensional structure of HIV-1 protease X-ray crystallography was used. The conditions for crystal growth were screened using the sitting drop vapor diffusion method in a 96-well microplate. A commercially available screening kit, Hampton research index HR2-144 (Hampton Research, USA) was used. The kit allowed for the rapid screening of crystals using 96 different reagents. Crystals were grown at 293 K and within two days crystals were observed. Various conditions from the screening kit resulted in crystal formation as sitting drop involves small volumes, the plate quickly dried up and obtained crystals were small; therefore, few of those conditions were prepared and used for the hanging drop vapor diffusion method in a purpose of growing diffraction worthy crystals. The conditions were prepared to 20 mL (2 mL salt, 2 mL buffer, 10 mL precipitant and 6 mL water) and crystals were grown using the hanging drop method at 293 K. Three conditions resulted in the formation of good quality crystal and the crystal obtained from using a reservoir buffer containing 0.1 M Bis-tris pH 6.5 and 3.0 NaCl was successfully diffracted. The protein stock concentration used was 5 mg/mL in buffer D. The hanging drop (6 µL) had a 1:1 ratio of protein stock and mother liquid and plate was thoroughly covered. The diffraction quality crystal was placed on a cryoloop and briefly immersed in a cryoprotectant solution containing 50% paratone in the mother liquor. The chosen crystal was mounted on the goniometer under cryo stream at 100K stabilised while liquid nitrogen was sprayed continuously on the sample. The X-ray diffraction data of HIV-1 C-SA protease was done

using a home source Bruker D8 Venture Bio PHOTON III 28-pixel array area detector (208 ×128 mm2) diffractometer with a Cu Kα IμS DIAMOND source with the wavelength 1.5418Å. X-ray diffraction intensity data of HIV-1 C-SA protease diffracted up to 2.40 Å resolution were collected at a crystal-to-detector distance of 106 mm. The diffraction images were collected with an oscillation angle of 0.4˚ with an exposure time of 5 seconds. The unit cell and full data set were collected using PROTEUM4: SAINT was used to integrate the data, and SADABS was used to make empirical absorption corrections. A POINTLESS program in built in PROTEUM4 software package was performed for scaling, determination of space group and generating intensity statistics for the data (129, 130).

2.10.2. Model building and refinement of structure

The protease structure phases were solved by a molecular replacement method using PHASER (131). The C-SA PR was used as a search probe for molecular replacement (PDB ID: 3U71; (43). Coot (132) was used for model building followed by using PHENIX (133) for cycles of global reciprocal space refinement. Coot was used again but for local real space refinement and stereochemical validation of the HIV-1 protease target model was performed using MolProbity (134) and PROCHECK (135). Images of the structure were generated using the PyMOL Molecular Graphics System (Schrödinger LLC., Portland, USA).

3. Results

3.1 HIV-1 protease overexpression and purification

3.1.1 Expression construct and sequence verification

Sequence verification of HIV-1 protease was done through Sanger sequencing for the isolated vector (Inqaba biotech). The analysis of the obtained sequence indicated that the sequence in the pET-11a vector was that of HIV-1 protease (ExPASy translate tool). The results from BLASTp tool demonstrated that the sequence was accurate with no lack or additional amino acids as it had 99 amino acids which is the number of amino acids found in each monomer of HIV-1 protease and that the amino acids in the sequence were 100% correct (Figure 11).

3.1.2. Target expression and growth curve

HIV-1 protease expression by *E. coli* BL21 (DE3) pLysS cells was verified by assessing the expression of target protein using tricine SDS-PAGE under reducing conditions. The protease was expressed by the cells as the gel shows faint bands (Figure 12A) around the expected theoretical molecular weight of HIV-1 protease (~11 kDa). The indefinite bands are also indicative of low expression associated with HIV-1 protease. To facilitate overexpression of HIV-1 protease an appropriate time for induction was determined, this was done by constructing a growth curve of transformed cells and deducing the OD_{600nm} for mid log phase. The growth curve (Figure 12B) of *E. coli* BL21 (DE3) pLysS cells with HIV-1 protease indicated mid log phase at OD_{600nm} ~ 0.6 and was reached approximately 2.5 hours after inoculation.

3.1.3. Induction trials

Information obtained from the growth curve was used to performed induction trials in order deduce optimum conditions for overexpression of HIV-1 protease. The overnight culture was diluted a 100-fold and induced at an $OD_{600\,nm}$ ~ 0.6 and allowed to grow at 37 ^{o}C, 200 rpm as detailed in section 2.4. After analysis of induction trials using tricine SDS-PAGE (Figure 12) it was determined that the optimum conditions for overexpression of HIV-1 protease in *E. coli* BL21 (DE3) pLysS cells was 1 mM IPTG for 4 hours at 37 ^{o}C, 200 rpm this was depicted by a thick prominent band from the gel (Figure 13E).

A.

B.

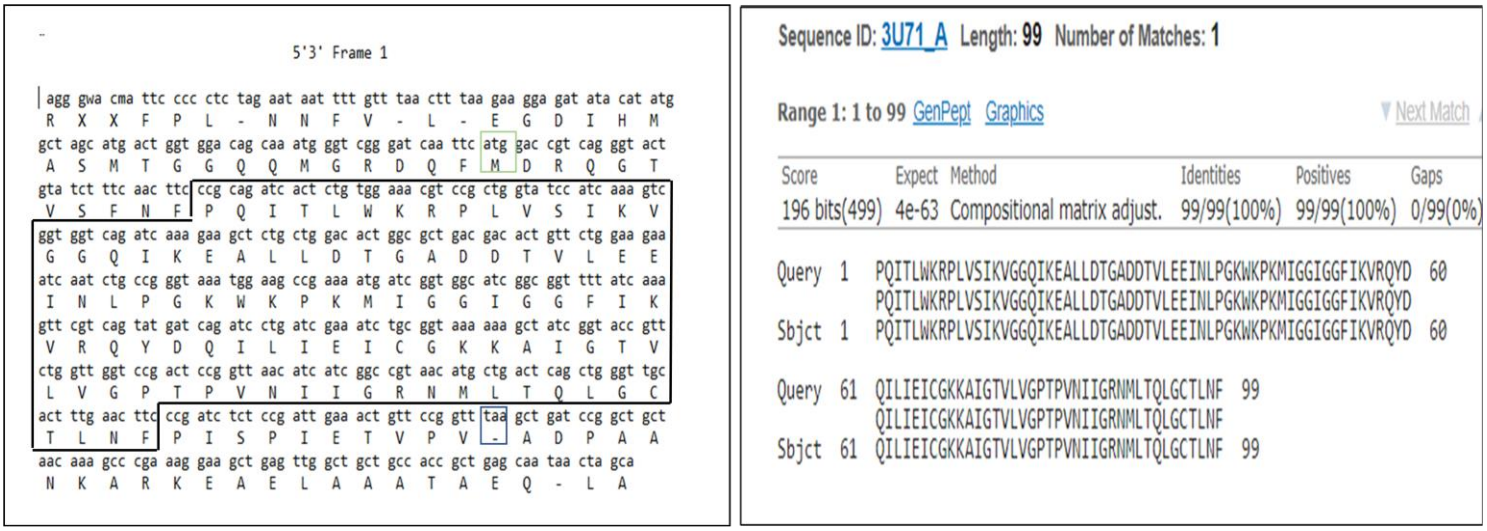


Figure 11: Expression construct and sequence verification. (A) Insertion of HIV-1 protease into pET-11a vector was verified using ExPASy translate tool, the selected sequence represents the HIV-1 sequence in the open reading frame with atg as a start codon and taa as a stop codon. (B) HIV-1 protease sequence (query) sequence alignment with the WT-CSA sequence in protein data bank (PDB ID: 3U71) as the subject using BLASTp tool indicated a 100% sequence match. The amino acid sequences both upstream (N-terminal) and downstream (C-terminal) of the protease do not have implications in protein folding, solubility, and functionality. Studies dealing with protease structure, function and stability are performed in the absence of sequences upstream and downstream of the protease sequence.

A.

B.

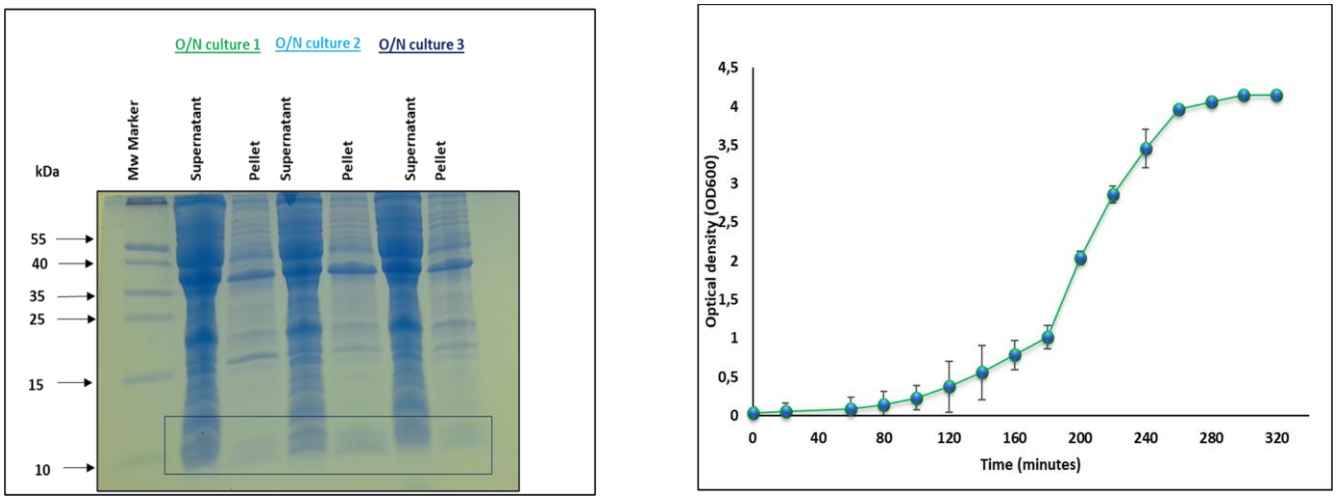


Figure 12: Assessment of gene expression and growth curve for HIV-1 protease. (A) Tricine SDS-PAGE showing expression of HIV-1 protease in *E. coli* BL21 (DE3) pLysS cells. The gel shows the supernatant and cell pellet of the overnight (O/N) culture. It was determined that HIV-1 protease was expressed by the cells as indicated by faint bands. (B) The growth curve generated as described in section 2.3, indicated that mid log phase OD_{600nm} of 0.6 was reached ~ 2.5 hours after inoculation.

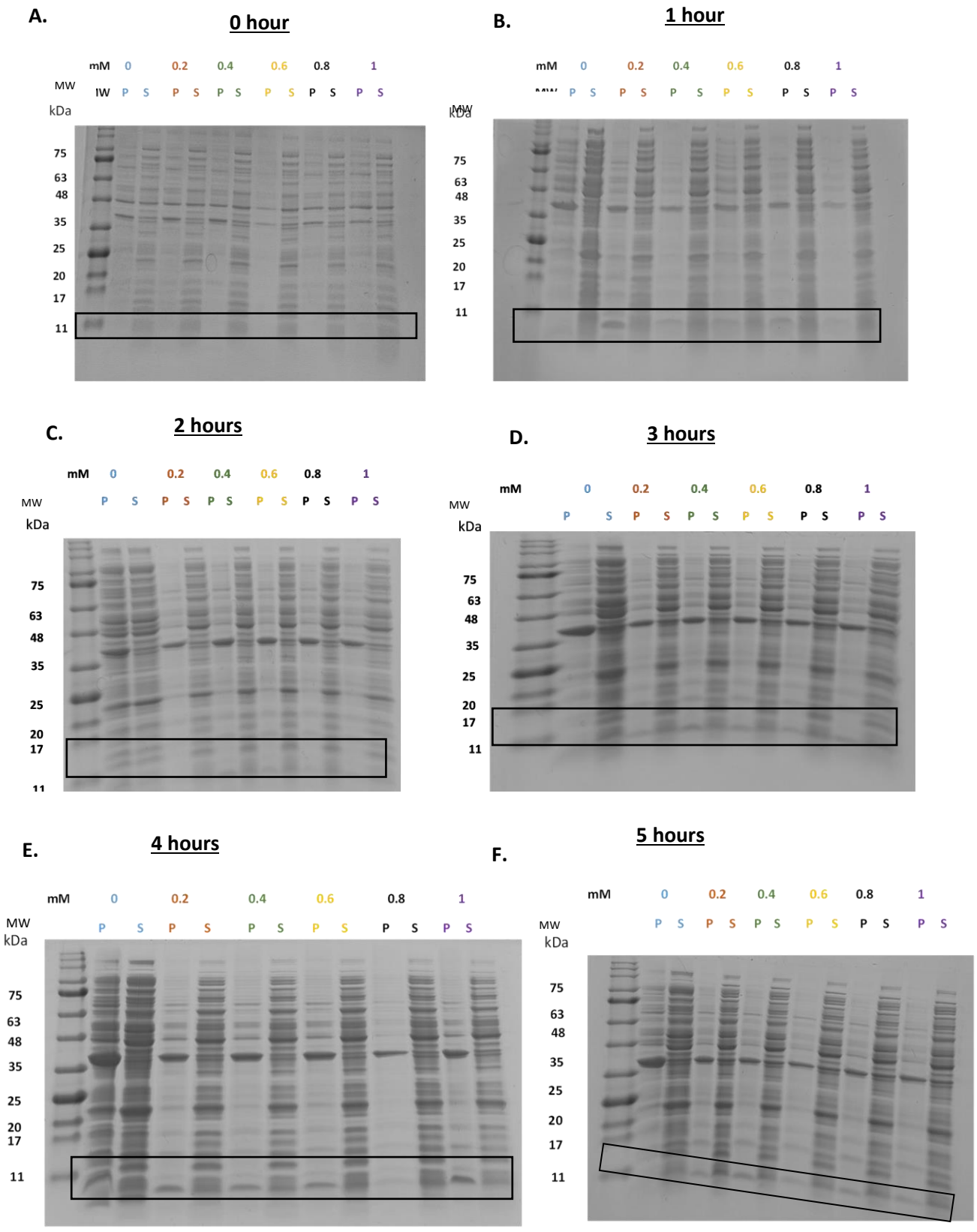


Figure 13: Analysis of HIV-1 protease induction trials using tricine SDS-PAGE. The induction trials were performed in different IPTG concentrations (Mm) for 0 (A),1 (B),2 (C),3 (D),4 (E),5 (F). Supernatant (S) and pellet (P) cell samples were analysed using the tricine SDS-PAGE gel and the optimal conditions for over-expression were determined to be 1 mM IPTG for 4 hours at 37 °C, 200 rpm.

3.1.4. Protein purification and protein assessment

HIV-1 protease was overexpressed and purified as detailed in section 2.6 and 2.7 respectively. Purification of HIV-1 protease from inclusion bodies using ion-exchange chromatography proved successfully. The last peak in the purification profile (Figure 14A) demonstrated the elution of the recombinant protease and the eluted protein was highly pure with tricine SDS-PAGE gel showing only the band of HIV-1 protease (Figure 14B). Furthermore, the quality of HIV-1 C-SA protease was also evaluated using the UV absorption spectrum by monitoring absorbance from 230 nm to 360 nm as described in section 2.8.1, the spectrum showed a flat baseline at 330 nm with a trough being observed at 250 nm (Figure 15). Quantitative assessment of HIV-1 protease was performed by estimating the monomeric size of HIV-1 protease from the calibration curve of log molecular weight versus R_f values. From the curve, the monomeric size of HIV-1 protease was estimated to be ~11 kDa (Figure 16A). The protein concentration was spectrophotometrically determined as described in section 2.8.1. It was determined that from 5.2 liters of overexpression, the average protein yield was around 2132 mg (Figure 16B) in a volume of approximately 10 mL.

Results from UV-Vis absorbance spectroscopy alone are not sufficient for protein quantification as the technique cannot discriminate between folded and improperly folded protein. To overcome this challenge active site titration of HIV-1 protease was performed to determine the percentage of properly refolded protein. This was done as detailed in section 2.8.2 by titrating acetyl-pepstatin against HIV-1 protease and fitting the data using NanoAnalyze (Figure 17). The experiment showed the stoichiometry value (*N*) to be 0.63 indicating that 63% of HIV-1 protease (Table 2) was properly refolded this was then used to correct the concentration obtained using UV-Vis absorption spectroscopy. Thermodynamic parameters (Table 2) obtained from the experiment. The reaction was endothermic (ΔH = 53.42 kJ/mol), spontaneous (ΔG = -35.44 kJ/mol), and entropically driven (ΔS = 303.1 J/mol. K) with the binding affinity of acetyl-pepstatin in Nanomolar range (K_d = 83.7 nM).

3.2. Characterisation of HIV-1 protease structure

3.2.1. Estimation of HIV-1 protease secondary structure

Far-UV CD experiments were performed using 5 µM of HIV-1 protease in buffer D over 8 accumulations as detailed in section 2.8.3. The protein was found to be predominantly β-sheeted exhibiting a pronounced positive band at 195 nm and a trough between 215-225 nm (Figure 18). The results were shown to be highly reproducible as indicated by the spectra (A, B and C).

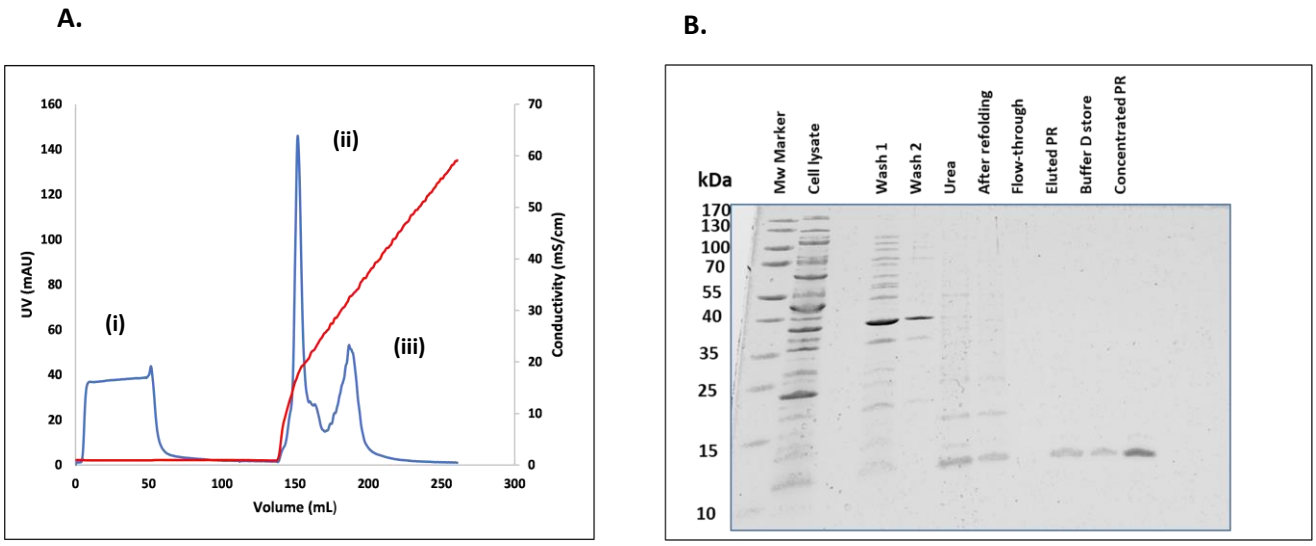


Figure 14: HIV-1 protease purification profile and tricine SDS-PAGE. (A) Chromatogram of HIV-1 protease showing UV absorbance (blue) and conductivity (red) as a function of NaCl . (i) Flow through during column protein loading. (ii) Elution of contaminants. (iii). Eluted HIV-1 protease. (B) Tricine SDS-PAGE gel showing steps involved in HIV-1 protease purification. The single band of HIV-1 protease is indicative of high protein purity.

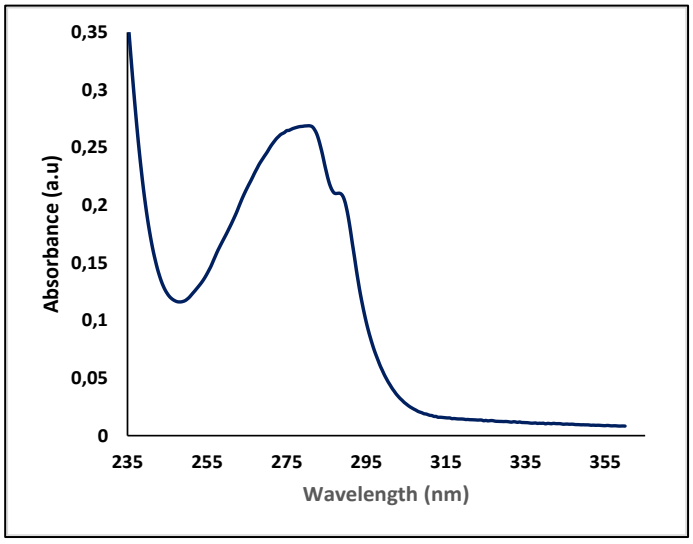


Figure 15: The UV absorption spectrum of HIV-1 protease. The spectrum was obtained by monitoring absorbance from 230 nm to 360 nm using HIV-1 protease in buffer D. A minimum is observed at 250 nm and a flat baseline at 330 this indicates that there is no DNA contamination or aggregation in the purified protein.

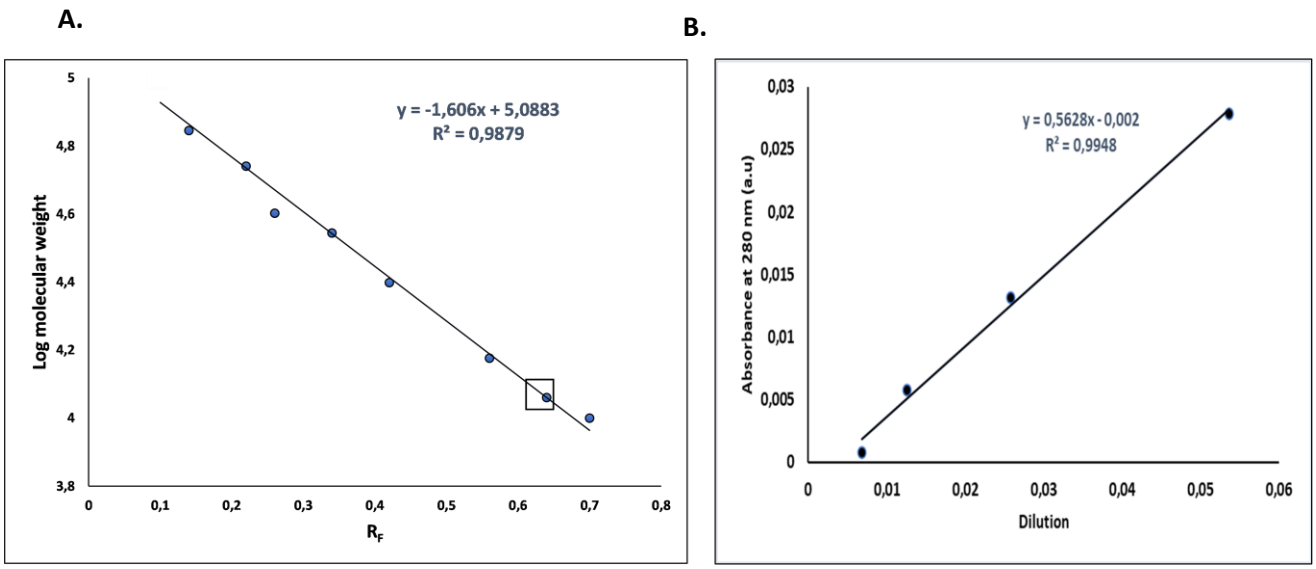


Figure 16: spectrophotometric quantification of HIV-1 protease quantification at 280 nm and 340 nm. (A) Calibration curve used to determine the protein monomeric size; the estimated size of HIV-1 protease (point in the square box) was ~11 kDa. (B) HIV-1 protease concentration was estimated from the serial dilution plot used in conjunction with Beer-Lambert law , it was estimated that the concentration of HIV-1 protease was 18.6 µM.

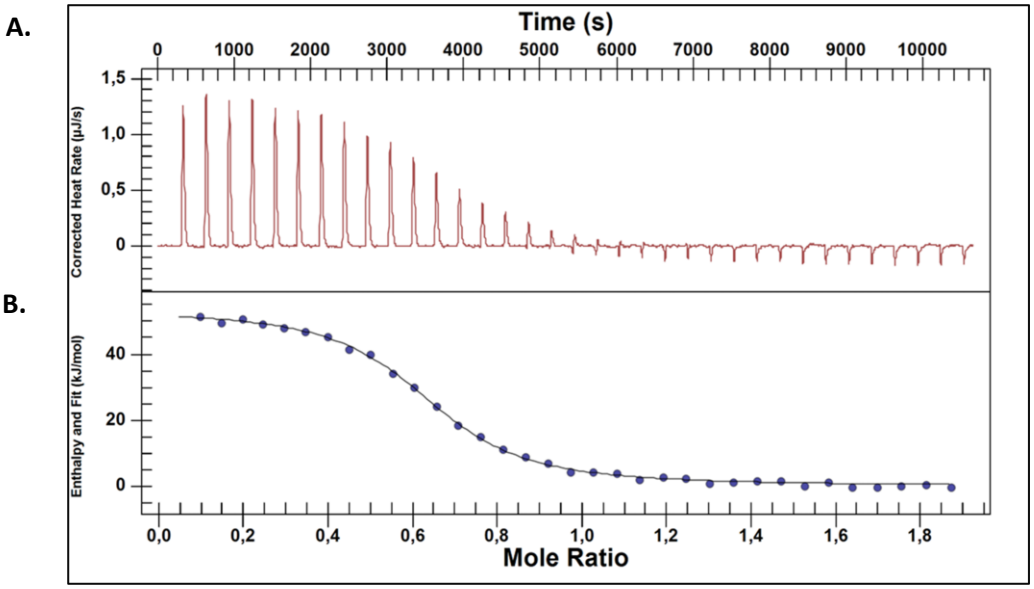


Figure 17: Analysis of HIV-1 protease active site titration. (A) Thermogram of ITC showing raw data from the titration reaction (B) An ITC isotherm showing fitted data of active site titration using independent model, titration of acetyl-pepstatin into HIV-1 protease was done until saturation was reached this was demonstrated by the sigmoidal shape of the isotherm.

Table 2: Active site titration thermodynamic parameters of HIV-1 protease with acetyl-pepstatin.

ΔH (kJ/mol)	ΔG (kJ/mol)	ΔS (J/mol.K)	N	K_d (nM)
53.42	-35.44	303.1	0.63	83.7

3.2.2. Analysis of HIV-1 protease tertiary structure

HIV-1 protease 2 µM in buffer D was used to perform intrinsic tryptophan fluorescence using fluorescence spectroscopy at excitation wavelengths of 280 nm and 295nm. At 280 nm both the tryptophan and tyrosine residues will be excited while only tryptophan is excited at 295 nm. The fluorescence spectra of HIV-1 protease at 280 nm overlayed with the 295 nm spectra demonstrates the local environment of the HIV-1 protease residues. The maximal emission of HIV-1 C-SA protease is observed at 347 nm (Figure 19).

3.2.3. Assessing the quaternary structure of HIV-1 protease

SE-HPLC was performed to determine the quaternary structure of HIV-1 C-SA protease the run was 30 minutes long using 20 µM of protein as described in section 2.8.5, resulting data from the generated chromatogram was used to estimate protein size it was observed that the protein came out as two peaks which can be classified as being a dimer and a monomer (Figure 20A). The peaks appeared as though they were delayed as shown in the calibration curve (Figure 20B) with size 21456 Da for the dimer and 10212 Da for the monomer which is close to what is expected from the HIV-1 protease size estimated using ExPasy ProtPARAM tool.

3.3. Functional characterisation of HIV-1 protease

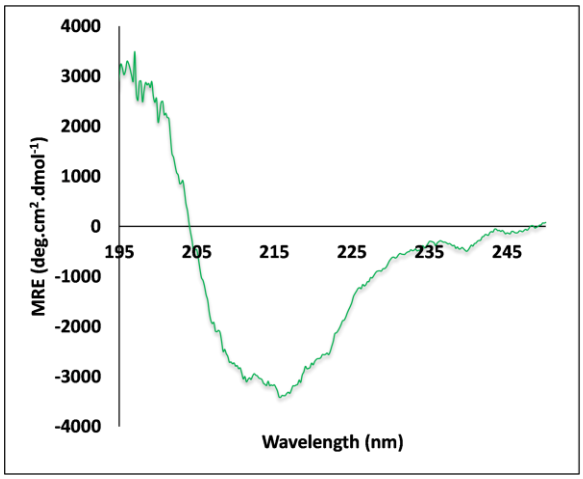
3.3.1. HIV-1 protease enzyme kinetics

Kinetics of HIV-1 protease in buffer D were performed by monitoring the hydrolysis of the fluorogenic substrate (Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala- Nle-NH_2) as described (section 2.9.1). Obtained data was analysed and tabulated as shown in Table 3. To ensure that catalytic activity of HIV-1 protease was retained after purification specific activity was determined (24.22 ± 1.72 µmol.min^{-1}mg^{-1}). Michaelis-Menten equation was used to assess the effect of substrate concentration on enzyme activity giving V_{max} and K_M at 0.036 ± 0.003 µmol. min^{-1} and 79.546 ± 6.491 µM respectively. The active site was determined to undergo 8.88 ± 0.62 sec^{-1} catalytic cycles. The conversion of substrate to product by the enzyme was determined to be 0.102 ± 0.01 sec^{-1}µM^{-1}.

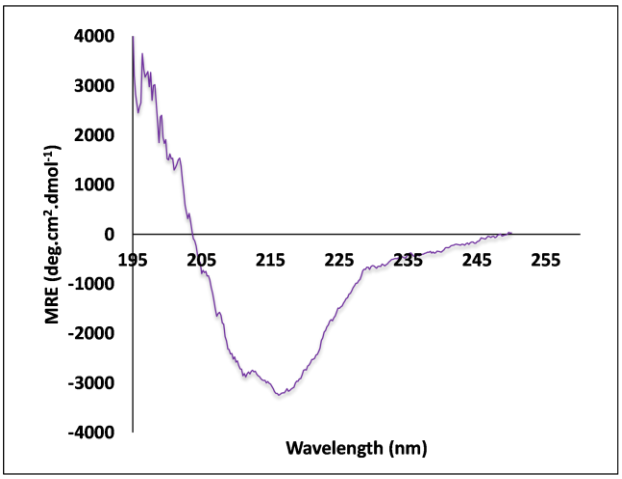
3.3.2. Thermodynamics of HIV-1 protease

Displacement ITC was performed to study thermodynamics of HIV-1 protease using three drugs (ATV,DRV and LPV). The titration of the drugs into the protease-acetyl-pepstatin complex displaces the weakly bound acetyl-pepstatin from the active site allowing for drug binding.

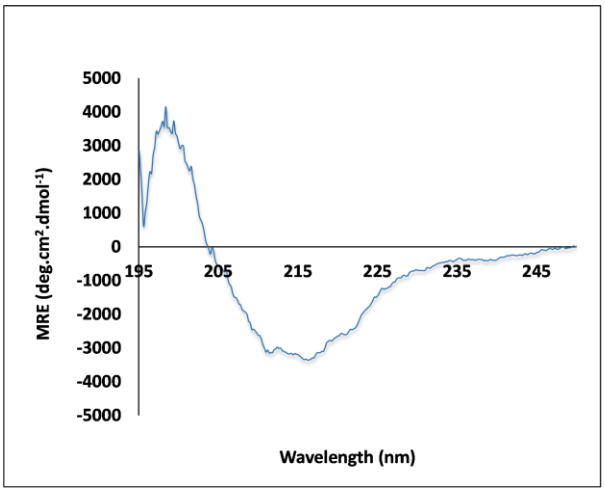
A.



B.



C.



D.

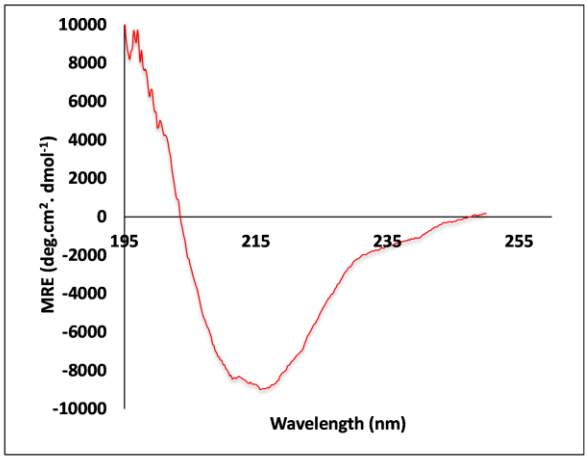


Figure 18: Analysis of HIV-1 protease secondary structure using far-UV CD. Experiments were performed in triplicate over 8 accumulations each using 5 µM of HIV-1 protease in buffer D as described in section 2.8.3. Data was fitted using Microsoft Excel software (Microsoft Corporation, USA). The resulting spectra indicate that the protein is β-dominant exhibiting a pronounced positive band at 195 nm and a trough between 215-225 nm. The obtained data is highly reproducible as shown by the spectra (A, B and C) with D showing the average spectrum of the triplicates. The shift in the spectrum observed in 18 C is likely due to slight changes in secondary structural contents as the molecules are conformationally flexible.

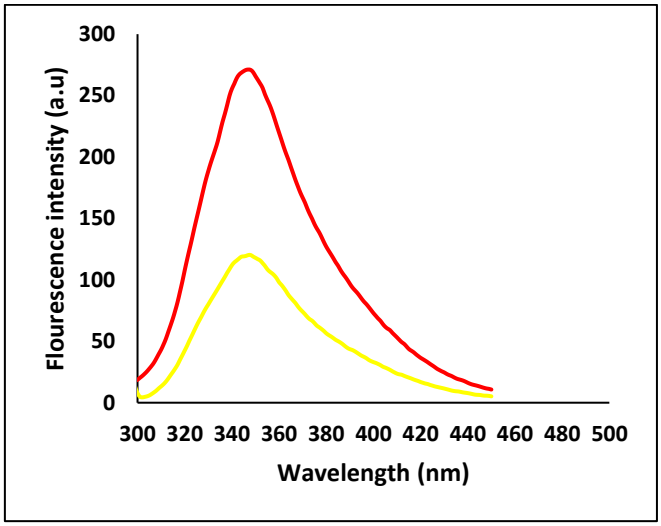
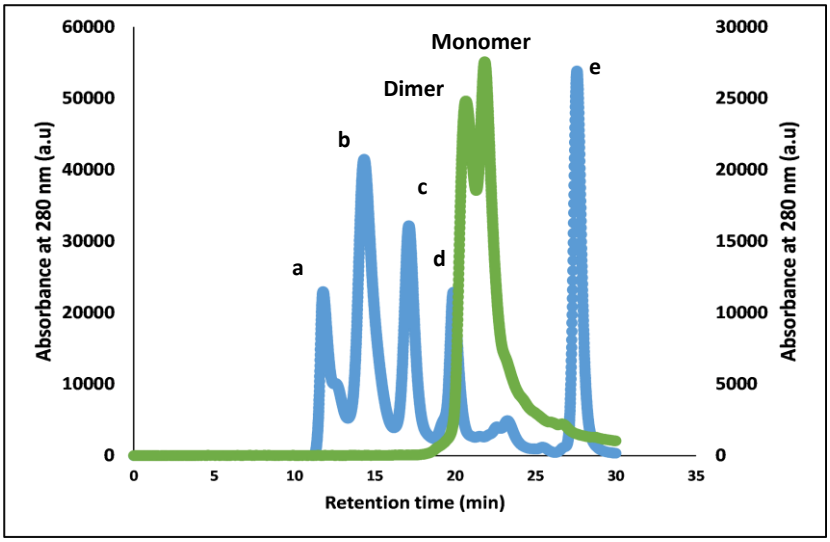


Figure 19: Fluorescence spectra of HIV-1 protease. Intrinsic fluorescence was performed using 2 µM HIV-1 protease in buffer D at excitation wavelengths of 280 (red) and 295 (yellow). The resulting data indicated a maximum emission at 347 nm for both excitation wavelengths.

A.



B.

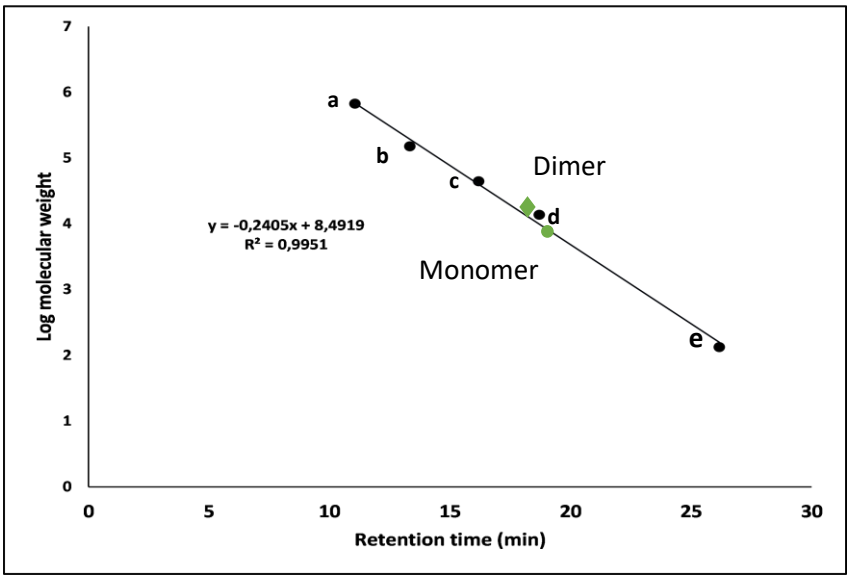


Figure 20: Quaternary structure analysis of HIV-1 protease. (A) The chromatogram from SE-HPLC indicating the nature of proteins as they pass through the column. (B) A calibration curve of log molecular weight that was used to estimate the protein size. The protein is shown to exist in two forms as it passes through the column. The standard mixture of proteins contains (a) thyroglobulin bovine (670 000 Da), (b) γ-globulins (150 000 Da), (c) Albumin (44 300 Da), (d) ribonuclease A (13 700 Da) and (e) Para-aminobenzoic acid (137.14 Da). The experiment was performed in 10 mM sodium acetate with 500 mM NaCl pH 5.0 at 20 °C.

The experiments were performed, data analysed as detailed in section 2.9.2 The reactions occurred at constant temperatures and reached saturation depicted as a sigmoidal shape as shown by the thermograms and isotherms in Figure 21, data analysis revealed significant information about binding interactions that occurred between the drugs and the HIV-1 protease. It was observed (Figure 22) that LPV had a high binding affinity with K_d = 1 nM followed by ATV which had K_d = 18.57 nM while DRV had a value of K_d = 42.26 nM. All the reactions were spontaneous (negative ΔG), entropically unfavoured (negative ΔS) but enthalpically driven (negative ΔH).

3.4. The three-dimensional structure of HIV-1 protease

The three-dimensional structure of HIV-1 protease was determined using X-ray crystallography. The trial-and-error experiments to determine the optimum conditions for crystal growth were performed, using sitting drop vapor diffusion method using the Hampton research index HR2-144 screening kit. The screening of different crystal conditions allows for optimal conditions that will result in crystals that can be diffracted. As part of our trial-and-error experiments different concentrations, buffers, temperatures and precipitant conditions were explored. The condition resulting in crystal growth Figure 23A was used for the growth of diffraction worthy crystals using hanging drop vapor diffusion method (Figure 23B) using 0.1M Bis-Tris pH 3.0M NaCl. The structure (Figure 24) was resolved at a resolution of 2.4 Å diffraction, structure modelling and refinement was done as described in section 2.10 and Table 4 represents a summary data of the obtained crystal structure.

Table 3 : Enzyme kinetic parameters for HIV-1 protease using a fluorogenic substrate e (Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle- NH2) (n = 3)

Parameters	values
V_{max} (µmol. min^{-1})	0.04 ± 0.003
K_M (µM)	79.5 ± 6.5
Specific activity (µmol. $min^{-1}.mg^{-1}$)	24.2 ± 1.7
k_{cat} (sec^{-1})	8.9 ± 0.6
k_{cat}/K_M (sec^{-1}. μM^{-1})	0.10 ± 0.01

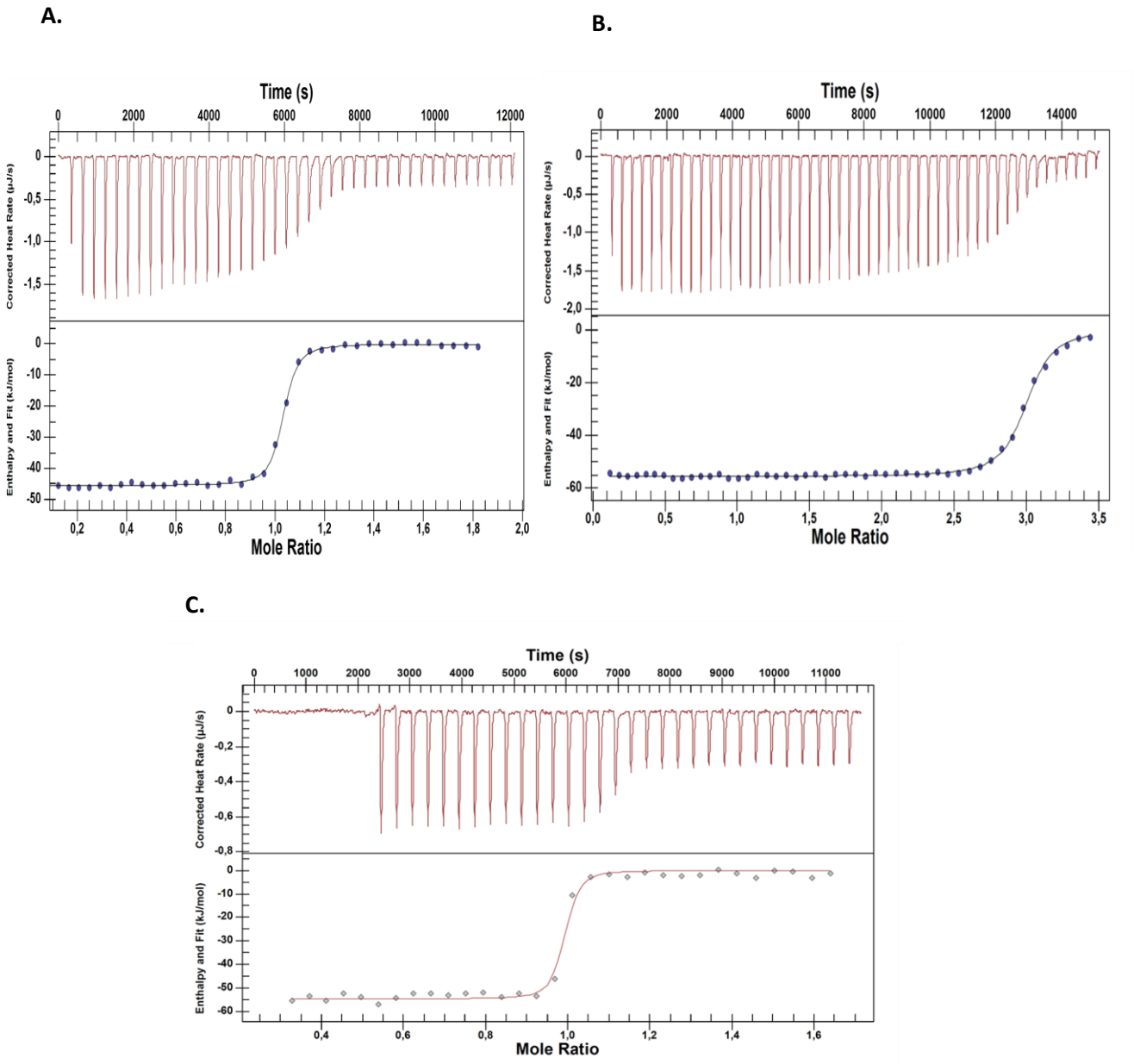


Figure 21: Typical displacement ITC experiment of HIV-1 protease with PIs. Thermogram of ITC showing raw data from the titration of inhibitor to protein reaction (upper plot) with ITC isotherm (bottom plot) showing fitted data using NanoAnalyze. Three drugs were used for the thermodynamic experiments, ATV (A), DRV (B) and LPV (C). Drug titration was done to fully saturate HIV-1 protease as demonstrated by the sigmoidal shape of the isotherm. HOD were subtracted and baseline correction made where was necessary prior and analysis was done using NanoAnalyze. The experiments were performed at 20 ^{o}C, 10 mM sodium acetate, 2% DMSO, pH 5.0.

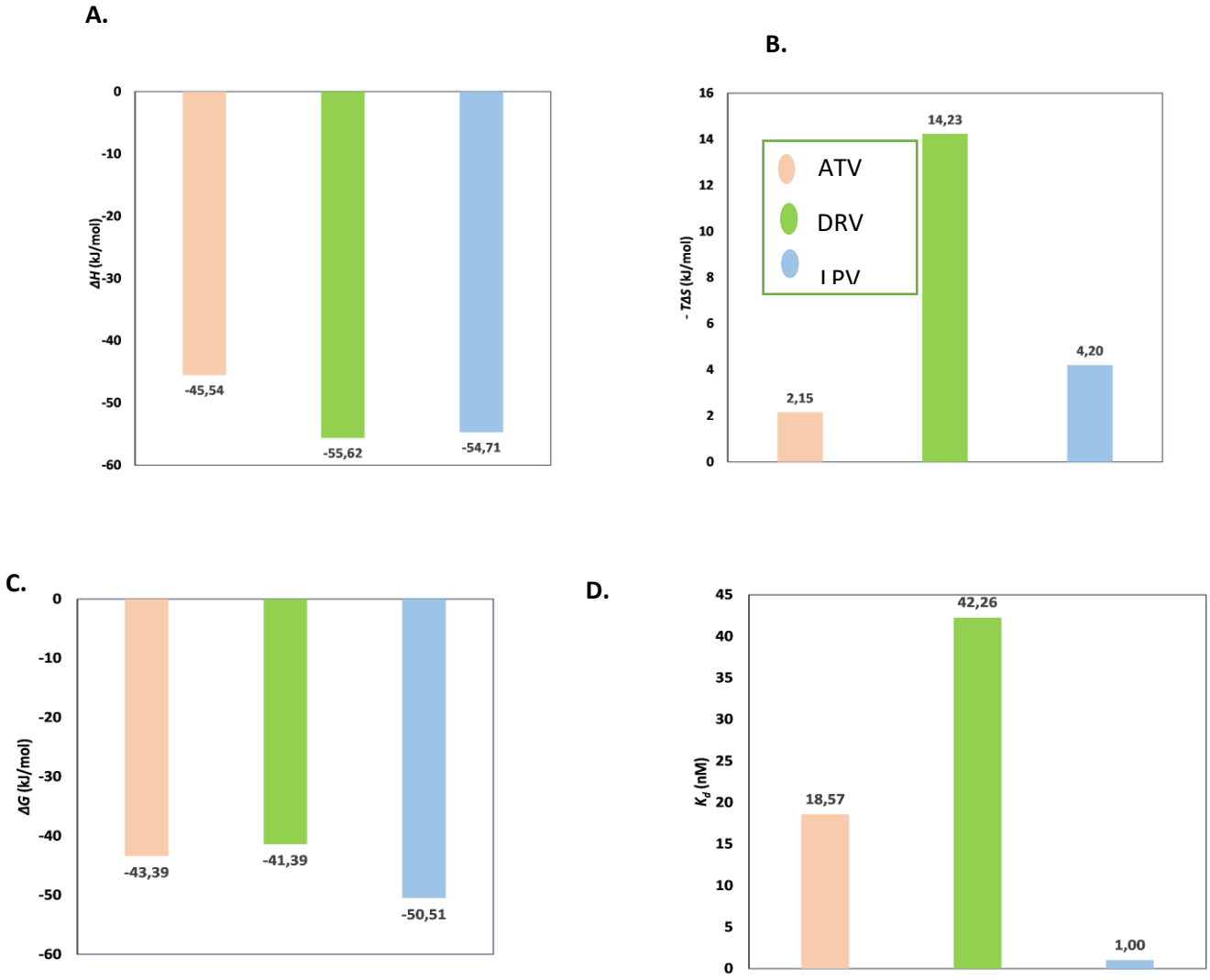


Figure 22: Thermodynamic parameters of HIV-1 protease interaction with ATV, DRV and LPV. (A) Showing the *ΔH* of the three drugs, (B) *ΔS* for the three drugs (C) Indicates *ΔG* (D) Binding affinity for the three drugs (K_d). From the data it was depicted that LPV has a higher binding affinity for HIV-1 protease followed by ATV. All the reactions involving the drugs showed that they were exothermic, spontaneous, and entropically unfavored. All the data was fitted using Microsoft excel (Microsoft corporation, USA).

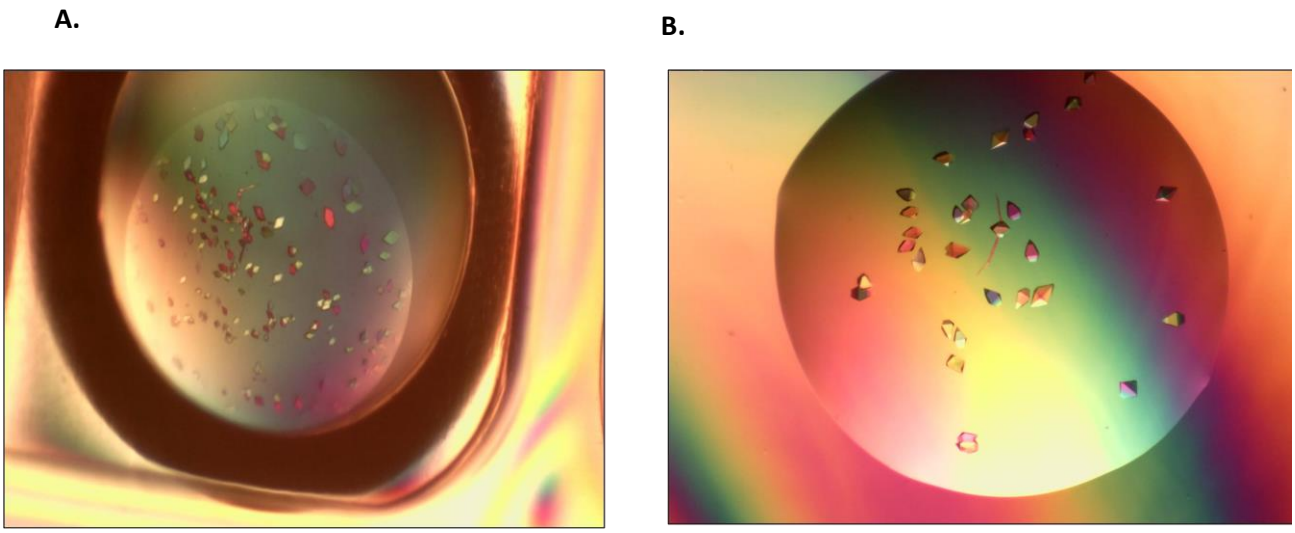


Figure 23: HIV-1 protease crystals obtained from using the vapor diffusion method (A) Screening kit results showing crystals for HIV-1 protease using sitting drop vapor diffusion method (B) HIV-1 protease crystals obtained from the hanging drop vapor diffusion method. Stock protein of 5 mg/mL in 10 mM sodium acetate pH 5.0 was used and crystals were grown at 293 K.

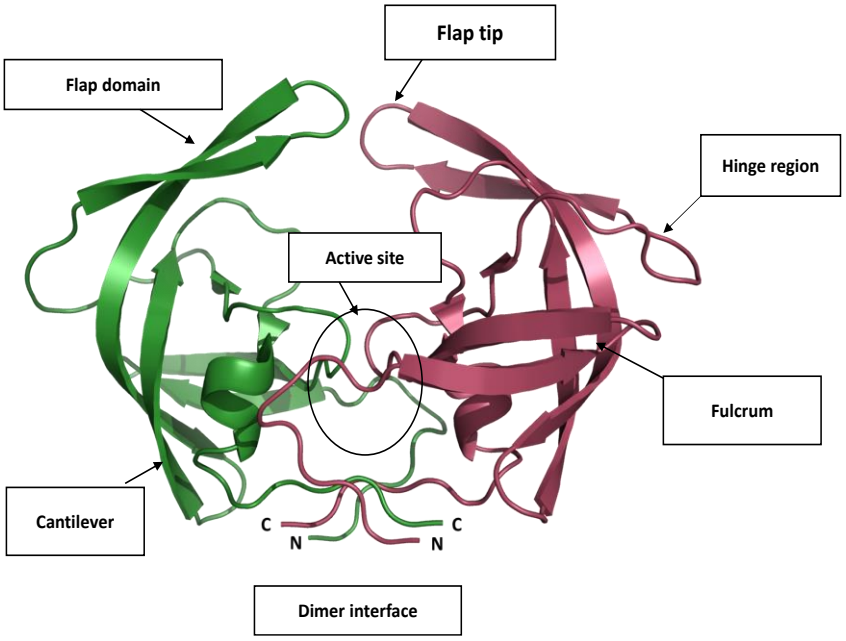


Figure 24: A ribbon representation of the HIV-1 C-SA crystal structure solved at 2.4 Å PDB ID: 8CI7. The Figure was generated using PyMOL.

Table 4: Data collection and refinement statistics for the HIV-1 protease structure

Data collection and processing statistics	
Wavelength (Å)	1.54184
X-ray source	CuKα
Space group	P 4_1 2_1 2
Unit-cell parameters	
a, b, c (Å)	46.60, 46.60, 102.14
α, β, γ (°)	90.00, 90.00, 90.00
Resolution range (Å)	22.39 – 2.40 (2.49-2.40)
No. of observed reflections	21173 (1915)
No. of unique reflections	4659 (433)
Completeness (%)	97.2 (88.7)
I/σ(I)	9.7 (0.8)
Multiplicity	4.5 (4.4)
R_{merge}	0.138 (1.654)
Matthews coefficient (VM) (Å3/Dalton)	2.37
Solvent content (%)	48.1
Asymmetric unit content	Monomer
Refinement statistics	
Resolution limit	2.40 - 22.39
R /R_{FREE}(%)	20.4 /22.1
COMPLETENESS(%)	96.4
No. of protein	830
No. of ligand atoms	0
Average B value (Å^2)	34.0
RMSD from standard geometry	
RMSD in bond length (Å)	0.003
RMSD in bond angles (°)	0.563
Wilson B-factor (Å^2)	36.6
Ramachandran statistics	
Residues in outlier regions (%)	0
Residues in favored regions (%)	98
Residues in allowed regions (%)	2
PDB ID	8CI7

Data in brackets represents high resolution shell

4. Discussion

HIV-1 protease as an important target for antiretroviral therapy has seen the development of various PIs, sadly these PIs are mostly developed for the HIV-1 subtype B protease and not the subtype C which is predominant in the sub-Saharan Africa, Brazil, India and China (136). The development of protease inhibitors that specifically target the subtype C protease is a necessity. The current mode of treatment using ART is associated with treatment failure which has been extensively studied, and these studies have recognised NOPs in subtype C as the main driver of resistance in the use of PIs (137). The pivotal aspects of drug discovery include the understanding of enzyme kinetics, thermodynamics and structure especially at a molecular level so drugs with high genetic barrier against resistance, high specificity or binding affinity for the target molecule can be designed and developed effectively.

In order to characterise the structure and function of HIV-1 protease, the *E. coli* BL21(DE3) pLysS cells were successfully transformed with a pET-11a vector containing a 99 amino acid sequence corresponding to HIV-1 protease, as demonstrated in Figure. 11. The 99 amino acid sequence aligned correctly to the amino acid sequence of the previously solved protease structure (PDB ID: 3U71)(43). The expression of recombinant protease (Figure 12A) was observed as faint bands around 10 kDa of the molecular weight marker, which is approximately the size of the monomeric HIV-1 protease (~11 kDa) that has been reported in various studies (43). These results also highlighted the challenge that many researchers face when working with HIV-1 protease which is its low yield and insolubility. To study HIV-1 protease its large-scale production is paramount, and scientist are constantly optimising conditions for recombinant protease production (138). Since the role of HIV-1 protease extends to the cleavage of an array of host cellular proteins resulting in induced cytotoxicity (120) it is necessary to monitor how it behaves within the cell and deduce appropriate time for induction, a typical *E. coli* growth curve (Figure 12B) was obtained for the recombinant protease and the time to induce was determined to be ~ 2.5 after inoculation (OD_{600nm} ~0.6). Upon performing induction trials for large scale production of recombinant HIV-1 protease, it was deduced that 1 mM of IPTG with growth of 4 hours after induction was required for optimal overexpression at 37 ^{o}C, 200 rpm (Figure 13E). This was demonstrated by a dark pronounced thick band around the expected size of the monomeric protease (~11 kDa), the results also showed that a large quantity of the recombinant protein was in the inclusion bodies which is what is expected for HIV-1 protease (138, 139).

The recombinant protease was effectively overexpressed and purified, it eluted as the second of the two peaks consistent with literature (140, 141) upon gradient elution with increasing NaCl concentration (Figure 14A). This suggests the slightly stronger ionic interaction of the recombinant protease with the column, thus requiring a higher salt concentration for its elution. The second peak protein (eluted PR) migrated as a single sharp band on the tricine SDS-PAGE (Figure 14B) indicating its high purity and the band size was estimated to be ~11 kDa (Figure 14A) Approximately all the refolded protein was bound to the column as no protein band was detected in the flow-through lane (Figure 14B). The quality of the recombinant protease was also demonstrated through the UV absorption spectrum (Figure 15), the generated spectrum showing a prominent shoulder around 295 nm due to the four tryptophan residues found in the HIV-1 protease dimer was the same as the one previously reported by Ramlucken *et al,* (142) on a similar study. It was deduced that there was no DNA contamination as the spectrum had a minimum or a trough around 250 nm in essence no bumps were spotted at 260 nm. The protein was also free from aggregates as the symmetry of the absorbance peak at 280 nm remained unaltered and the flat baseline at 330 nm was observed (83). The concentration of the recombinant protease was estimated from the calibration curve (Figure 16B) to be 2132 mg (10 mL) from 5.2 liters of overexpression which is consistent with low yields associated with HIV-1 protease.

The percentage of properly refolded protein was accurately determined from the ITC experiment as acetyl-pepstatin did bind to the protease, and no undesired peaks were observed during the experiment (Figure 17A). The saturation curve obtained was in a sigmoidal shape (Figure 17B) all the obtained data was consistent with the results obtained by (43, 127). The binding ratio of acetyl-pepstatin to HIV-1 protease is 1:1 this allowed for concentration determination from the experiment in terms of the stoichiometry value which was found to be 0.63 (Table 2), meaning only 63% of recombinant protease was active from the initially determined concentration. This percentage was used to correct the concentration for subsequent experiments thus ensuring all the experiments were conducted using accurate concentrations. Thermodynamic parameters (Table 2) gave valuable insights on the binding interactions of acetyl-pepstatin to HIV-1 protease, the ligand was demonstrated to weakly bind to the recombinant protease (K_d = 83.7 nM) this is line with the knowledge that acetyl-pepstatin has a low affinity and is used in displacement ITC for high affinity inhibitors (127) and the value does not vary significantly from 400 nM reported by Mosebi *et al*, (143). The reaction was endothermic (ΔH = 53.42 kJ/mol) but spontaneous (ΔG = -35.44 kJ/mol) this is said to be due to the enthalpy-entropy compensation effect which is largely observed in the thermodynamics of protein-ligand binding (113) making the reaction entropically driven (ΔS = 303.1 J/mol. K).

The secondary structure of HIV-1 protease depicted in Figure 18 indicated that the HIV-1 protease was predominantly β-sheeted exhibiting a pronounced positive band at 195 nm and a trough between 215-225 nm. The structure correlate with literature (144) wherein it was demonstrated that each subunit consists of nine β-strands and one α-helix with an antiparallel β-sheet making up the dimer interface. The technique is limited in that it does not provide structure details on a molecular level as X-ray crystallography. Intrinsic tryptophan fluorescence of HIV-1 protease had an emission maximum at 347 nm for both the excitation at 280 nm and 295 nm (Figure 19). The results suggested that the tryptophan residues within the protease are solvent exposed as the maximal emission of tryptophan in water is 350 nm (103). Furthermore, this was consistent with the crystal structure for HIV-1 protease which indicates that the protease has four tryptophan residues two in each monomer (Trp 6, Trp 6', Trp 42, Trp 42') and two tyrosine residues Tyr 59 and Tyr 59'. Proteins containing tryptophan show a fluorescence emission that is a linear combination of individual tryptophan residues present in its sequence(103). The results are also similar to recently reported data on the protease (145). The analysis of quaternary structure of the protease showed presence of the dimer and a monomer (Figure 20A), with their respective sizes in Figure 20B the dimer was estimated to be 21456 Da and 10212 Da for the monomer which was close to what is expected for the HIV-1 protease size estimated using ExPasy ProtPARAM tool and the widely reported 22 kDa (dimer) in literature.

The summary of kinetic parameters for HIV-1 protease are depicted in Table 3, showing that HIV-1 protease was catalytic active with the specific activity of 24.2$\pm$1.7 µmol. $min^{-1}.mg^{-1}$ which is not largely different from 21.0±1.1 µmol. $min^{-1}.mg^{-1}$ reported by Williams *et al* (146). The V_{max} was determined to be 0.04$\pm$0.003 µmol.min^{-1} which is low and indicates that the enzyme is not rapidly catalysing the reaction and high enzyme concentration is required to increase the V_{max}. The K_M was determined to be 79.6$\pm$6.5 µM meaning that the substrate was weakly bound to the enzyme, and that V_{max} will be reached at low substrate concentration. The K_M value obtained does differ from the 14±1.7 µM reported by Williams *et al* but correlate to 76 µM reported by Mosebi *et al*, (143). The K_M values which reflects affinity of an enzyme for its substrate might be different between studies due to different factors. These factors may include experimental conditions such as slight changes in pH or fluctuations in enzyme purity, which will in turn affect enzyme activity and therefore the K_M value. The catalytic turnover number was determined to be 8.9$\pm$0.6 sec^{-1} which is almost the same as 7.7±0.4 sec^{-1} reported by Williams *et al*,(146). The catalytic efficiency was deduced to be 0.10 $\pm$ 0.01 sec^{-1}. μM^{-1} it can be concluded that the protease has a low capacity to convert the substrate to product.

Displacement ITC thermograms and isotherms for the selected PIs are depicted in Figure 21. The PIs (ATV, DRV and LPV) have a high binding affinity ($K_i \leq 1$ nM) making it impossible to carry out traditional ITC experiments as they are beyond the determination limit hence the application of displacement

ITC. All the experiments were successfully carried out as there was binding of PIs to HIV-1 protease depicted by the sigmoidal shapes in each experiment indicating saturation as all available binding sites are being fully occupied. Thermodynamic parameters obtained from each of these experiments demonstrated in Figure 22D indicated that LPV (1 nM) had the highest binding affinity for HIV-1 protease while DRV had the lowest binding affinity (K_d = 42.26 nM) lower than ATV (K_d = 18.57 nM). The K_d values indicate that the binding affinity for these drugs for HIV-1 protease is LPV > ATV > DRV and this follows the same pattern as with the established K_i values for these drugs (LPV: 1.3 pM, ATV: 10 pM and DRV: 15 pM) as described in section 1.5 The binding energetics of the selected PIs indicated that the change in free energy of binding (*ΔG*) is favourable as characterised by negative values (Figure 22C) like for any spontaneous processes the spontaneity of these reactions is consistent with literature (66). The *ΔG* in protein-ligand interactions also give insight on the stability of the interaction with protein-ligand complexes having more negative *ΔG* values being said to be more stable the *ΔG* values for the three PIs (ATV = -43.39 kJ/mol, DRV = -41.39 kJ/mol and LPV = -50.51 kJ/mol) indicated that HIV-1 protease complexed with LPV is most stable, this does correlate with the obtained K_d values as the more stable the ligand-protein interaction is the higher the energy of the ligand binding affinity.

The change in free energy of binding has the enthalpic and entropic components meaning the event of recognition is linked with changes in the structure and dynamics for each counterpart (93). The enthalpy changes (Figure 22A) from the protein-ligand binding of the protease and the PIs indicated that all the interactions were exothermic (negative *ΔH*) with ATV = -45.54 kJ/mol, DRV = -55.62 kJ/mol and LPV = -54.71 kJ/mol. These values did vary in magnitude from the ones reported in literature when similar studies were performed by Naicker *et al*, (66) demonstrating ATV and DRV *ΔH* values of -20.08 kJ/mol and -79.91 kJ/mol, respectively. The change in enthalpy during the binding process is a result of formation and disruption of various noncovalent interactions. These include the formation of van der Waals interactions and loss of hydrogen bonds between ligand and solvent, as well as protein and solvent. A favourable *ΔH* suggests the formation of noncovalent interactions like van der Waals forces, hydrogen bonds, and electrostatic interactions, within the binding interface of the protein (93). It can be concluded that DRV interaction with HIV-protease is more exothermic than that of ATV and LPV interactions. Since the *ΔH* magnitude provides information about the specificity of inhibitor to protein (147) the results also indicated that DRV is more specific for the HIV-1 protease. Similar information can be deduced from the study performed by Naicker *et al*, (66). The dynamics of the overall system for individual experiments was shown to be unfavourable as shown by positive values of *-TΔS* in Figure 22B, because the contributions of enthalpy and entropy are related there is an enthalpy-entropy compensation for spontaneous reactions (93). The entropy may be affected through mobility restrictions of interacting molecules when there is an increase in enthalpy, making the reaction

enthalpically driven. The interaction of DRV with HIV-1 protease is the most entropically unfavourable interaction with the value of -TΔS DRV = 14.23 kJ/mol followed by LPV: 4.2 kJ/mol and ATV: 2.15 kJ/mol. A previous study has shown that the interaction of DRV with HIV-1 protease is entropically unfavourable (18.83 kJ/mol) (65). This is also in correlation with the observed high *ΔH* values which in turn restricted the mobility of the interacting molecules and had a large contribution to the spontaneity of the experiments.

It can be said that high specificity does not translate to high binding affinity as demonstrated through DRV, though it has high specificity for HIV-1 protease compared to the other two inhibitors it does not have a high binding affinity for HIV-1 protease, rather it is LPV that was demonstrated to strongly bind to the protease. The high specificity of DRV can be attributed to it being the most recently approved FDA PI amongst the three and its low binding affinity can be attributed to the fact that all these PIs were initially designed for the subtype B proteases not the subtype C and can, therefore, behave differently between the subtypes (65). This presents the need for the development of PIs that are specifically designed for HIV-1 subtype C protease and in making this a possibility it is important to have ongoing studies on the structure of HIV-1 subtype C protease so that it can be well documented.

X-ray crystallography is still a preferred tool in elucidating the three-dimensional structure of proteins. Figure 23 shows the tetrahedral shaped HIV-1 crystals using 0.1M Bis-Tris pH 3.0M NaCl with Figure 23A representing showers of crystals that resulted from screening using the sitting drop method and Figure 23B showing more well-defined crystals obtained from the hanging drop method which were worthy for diffraction the crystal structure was solved at 2.4 Å and the coordinate for the structure were deposited to Protein Data Bank (PDB ID: 8CI7), Figure 24 represents the ribbon of the crystal structure of the HIV-1 protease with all the important regions indicated. Table 4 shows the collected data and refinement statistics. The space group of structure is the same as the one previously reported by Naicker *et al*, (43) but the unit cell parameters (*a, b, c*) vary slightly from the PDB ID: 3U71. The solved structure had no rotamers or disallowed bond angles and has a higher resolution than the HIV-1 C-SA protease structure previously solved in our lab. The secondary structural content for both proteases is the same, though the number of protein atoms in 8CI7 is 830 while for 3U71 is 755. The increased number of protein atoms might have implications in functional characterisation of the protein. The solved crystal structure (8CI7) demonstrated low root mean square deviation (RMSD) values compared to the protease previously solved in our lab (3U71).

The improved high resolution three-dimensional structure will provide a better understanding of the HIV-1 protease, as it will provide precise information about the arrangement of atoms within the protease molecule. This detailed molecular architecture will aid in the design and development of

protease inhibitors that will target the subtype C protease. This can be done by using computational tools to visualise interactions between the subtype C protease and potential inhibitors.

5. Conclusion

In conclusion, HIV-1 CSA protease is important for the development of new antivirals mainly those with high genetic barrier as NOPs will hinder their efficacy. The currently use antiretroviral therapy constitute drugs mainly designed to target the subtype B. Here, we have successfully cloned, overexpressed and characterised HIV-1 CSA protease using a range of biophysical techniques and furthermore elucidated the three-dimensional structure of the protease. enzyme kinetic data indicated that the protein was catalytic active while thermodynamics studies demonstrated that the interaction of PIs with HIV-1 protease may varies from what maybe generally expected due to the NOPs, indicating the need for PIs designed for their specific subtypes as in the long run the efficacy of these PIs diminishes and results in counter effects as more virulent strains emerge.

6. References

1. Freed EO. HIV-1 assembly, release and maturation. Nature Reviews Microbiology. 2015;13(8):484-96.
2. Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH. The origins of acquired immune deficiency syndrome viruses: where and when? Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2001;356(1410):867-76.
3. Carter CA, Ehrlich LS. Cell biology of HIV-1 infection of macrophages. Annu Rev Microbiol. 2008;62:425-43.
4. Cooper A, García M, Petrovas C, Yamamoto T, Koup RA, Nabel GJ. HIV-1 causes CD4 cell death through DNA-dependent protein kinase during viral integration. Nature. 2013;498(7454):376-9.
5. Ugwu CLJ, Ncayiyana JR. Spatial disparities of HIV prevalence in South Africa. Do sociodemographic, behavioral, and biological factors explain this spatial variability? Frontiers in Public Health. 2022;10.
6. Menéndez-Arias L, Tözsér J. HIV-1 protease inhibitors: effects on HIV-2 replication and resistance. Trends in pharmacological sciences. 2008;29(1):42-9.
7. Lessells R, Katzenstein D, De Oliveira T. Are subtype differences important in HIV drug resistance? Current opinion in virology. 2012;2(5):636-43.
8. Velazquez-Campoy A, Todd MJ, Vega S, Freire E. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. Proceedings of the National Academy of Sciences. 2001;98(11):6062-7.
9. Gartner MJ, Roche M, Churchill MJ, Gorry PR, Flynn JK. Understanding the mechanisms driving the spread of subtype C HIV-1. EBioMedicine. 2020;53:102682.
10. Delva W, Abdool Karim Q. The HIV epidemic in Southern Africa–is an AIDS-free generation possible? Current Hiv/Aids Reports. 2014;11:99-108.
11. Greene WC, Peterlin BM. Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. Nature medicine. 2002;8(7):673-80.
12. Narayan O, Clements JE. Biology and Pathogenesis of Lentiviruses. Journal of General Virology. 1989;70(7):1617-39.
13. Freed EO. HIV-1 replication. Somatic cell and molecular genetics. 2001;26:13-33.
14. Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. Annual review of biochemistry. 1998;67(1):1-25.
15. Kirchhoff F. HIV Life Cycle: Overview. 2013. p. 1-9.
16. Showa SP, Nyabadza F, Hove-Musekwa SD. On the efficiency of HIV transmission: Insights through discrete time HIV models. PLOS ONE. 2019;14(9):e0222574.
17. Shukla E, Chauhan R. Host-HIV-1 Interactome: A Quest for Novel Therapeutic Intervention. Cells. 2019;8(10):1155.
18. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Science. 1996;273(5283):1856-62.
19. Nielsen MH, Pedersen FS, Kjems J. Molecular strategies to inhibit HIV-1 replication. Retrovirology. 2005;2:1-20.
20. Saxena SK, Chitti SV. Molecular Biology and Pathogenesis of Retroviruses. Advances in Molecular Retrovirology: IntechOpen; 2016.
21. Mitchell RS, Beitzel BF, Schroder ARW, Shinn P, Chen H, Berry CC, et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS biology. 2004;2(8):e234.
22. German Advisory Committee Blood SAoPTbB. Human immunodeficiency virus (HIV). Transfus Med Hemother. 2016;43(3):203-22.
23. Klimas N, Koneru AOB, Fletcher MA. Overview of HIV. Psychosomatic medicine. 2008;70(5):523-30.

24. Bigna JJR, Plottel CS, Koulla-Shiro S. Challenges in initiating antiretroviral therapy for all HIV-infected people regardless of CD4 cell count. Infectious diseases of poverty. 2016;5:1-6.
25. Ghosh AK, Anderson DD, Weber IT, Mitsuya H. Enhancing protein backbone binding—a fruitful concept for combating drug-resistant HIV. Angewandte Chemie International Edition. 2012;51(8):1778-802.
26. Kramer R, Schaber M, Skalka A, Ganguly K, Wong-Staal F, Reddy E. HTLV-III gag protein is processed in yeast cells by the virus pol-protease. Science. 1986;231(4745):1580-4.
27. Tomasselli AG, Heinrikson RL. Targeting the HIV-protease in AIDS therapy: a current clinical perspective. Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology. 2000;1477(1-2):189-214.
28. Wlodawer A, Erickson JW. Structure-based inhibitors of HIV-1 protease. Annual review of biochemistry. 1993;62(1):543-85.
29. Turner BG, Summers MF. Structural biology of HIV. Journal of molecular biology. 1999;285(1):1-32.
30. DEBOUCK C. The HIV-1 protease as a therapeutic target for AIDS. AIDS research and human retroviruses. 1992;8(2):153-64.
31. Pietrucci F, Vargiu AV, Kranjc A. HIV-1 protease dimerization dynamics reveals a transient druggable binding pocket at the interface. Scientific reports. 2015;5(1):18555.
32. Weber IT, Agniswamy J. HIV-1 protease: structural perspectives on drug resistance. Viruses. 2009;1(3):1110-36.
33. Gulnik S, Erickson JW, Xie D. HIV protease: enzyme function and drug resistance. 2000.
34. Zhang S, Kaplan AH, Tropsha A. HIV-1 protease function and structure studies with the simplicial neighborhood analysis of protein packing method. Proteins: Structure, Function, and Bioinformatics. 2008;73(3):742-53.
35. Todd MJ, Semo N, Freire E. The structural stability of the HIV-1 protease. Journal of molecular biology. 1998;283(2):475-88.
36. Huff JR. HIV protease: a novel chemotherapeutic target for AIDS. Journal of medicinal chemistry. 1991;34(8):2305-14.
37. Hornak V, Okur A, Rizzo RC, Simmerling C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. Proceedings of the National Academy of Sciences. 2006;103(4):915-20.
38. Soares RO, Torres PH, da Silva ML, Pascutti PG. Unraveling HIV protease flaps dynamics by Constant pH Molecular Dynamics simulations. Journal of Structural Biology. 2016;195(2):216-26.
39. Liu Z, Huang X, Hu L, Pham L, Poole KM, Tang Y, et al. Effects of hinge-region natural polymorphisms on human immunodeficiency virus-type 1 protease structure, dynamics, and drug pressure evolution. Journal of Biological Chemistry. 2016;291(43):22741-56.
40. Wlodawer A, Miller M, Jaskólski M, Sathyanarayana BK, Baldwin E, Weber IT, et al. Conserved folding in retroviral proteases: crystal structure of synthetic HIV-1 protease. Science. 1989;245(4918):616-21.
41. Gustchina A, Weber IT. Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap. FEBS letters. 1990;269(1):269-72.
42. Hornak V, Simmerling C. Targeting structural flexibility in HIV-1 protease inhibitor binding. Drug discovery today. 2007;12(3-4):132-8.
43. Naicker P, Achilonu I, Fanucchi S, Fernandes M, Ibrahim MA, Dirr HW, et al. Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance. Journal of Biomolecular Structure and Dynamics. 2013;31(12):1370-80.
44. Harte Jr W, Swaminathan S, Mansuri M, Martin J, Rosenberg I, Beveridge D. Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. Proceedings of the National Academy of Sciences. 1990;87(22):8864-8.
45. Louis JM, Ishima R, Torchia DA, Weber IT. HIV-1 protease: structure, dynamics, and inhibition. Advances in pharmacology. 2007;55:261-98.

46. Wensing AM, van Maarseveen NM, Nijhuis M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. Antiviral research. 2010;85(1):59-74.
47. Lee S-K, Potempa M, Swanstrom R. The choreography of HIV-1 proteolytic processing and virion assembly. Journal of Biological Chemistry. 2012;287(49):40867-74.
48. Clavel F, Mammano F. Role of Gag in HIV resistance to protease inhibitors. Viruses. 2010;2(7):1411-26.
49. Bell NM, Lever AM. HIV Gag polyprotein: processing and early viral particle assembly. Trends in microbiology. 2013;21(3):136-44.
50. Lv Z, Chu Y, Wang Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. HIV/AIDS-Research and palliative care. 2015:95-104.
51. De Oliveira T, Engelbrecht S, Janse van Rensburg E, Gordon M, Bishop K, Zur Megede J, et al. Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? Journal of virology. 2003;77(17):9422-30.
52. Chou K-C. A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins. Journal of Biological Chemistry. 1993;268(23):16938-48.
53. Wlodawer A, Vondrasek J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annual review of biophysics and biomolecular structure. 1998;27(1):249-84.
54. Brik A, Wong C-H. HIV-1 protease: mechanism and drug discovery. Organic & biomolecular chemistry. 2003;1(1):5-14.
55. Shen C-H, Tie Y, Yu X, Wang Y-F, Kovalevsky AY, Harrison RW, et al. Capturing the Reaction Pathway in Near-Atomic-Resolution Crystal Structures of HIV-1 Protease. Biochemistry. 2012;51(39):7726-32.
56. Tang J, Wong RN. Evolution in the structure and function of aspartic proteases. Journal of cellular biochemistry. 1987;33(1):53-63.
57. Andersson HO, Fridborg K, Löwgren S, Alterman M, Mühlman A, Björsne M, et al. Optimization of P1–P3 groups in symmetric and asymmetric HIV-1 protease inhibitors. European journal of biochemistry. 2003;270(8):1746-58.
58. Pokorná J, Machala L, Řezáčová P, Konvalinka J. Current and novel inhibitors of HIV protease. Viruses. 2009;1(3):1209-39.
59. Sham HL, Kempf DJ, Molla A, Marsh KC, Kumar GN, Chen C-M, et al. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. Antimicrobial agents and chemotherapy. 1998;42(12):3218-24.
60. Zeldin RK, Petruschke RA. Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients. Journal of antimicrobial chemotherapy. 2004;53(1):4-9.
61. Gong Y-F, Robinson BS, Rose RE, Deminie C, Spicer TP, Stock D, et al. In vitro resistance profile of the human immunodeficiency virus type 1 protease inhibitor BMS-232632. Antimicrobial agents and chemotherapy. 2000;44(9):2319-26.
62. Ghosh AK, Dawson ZL, Mitsuya H. Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drug-resistant HIV. Bioorganic & medicinal chemistry. 2007;15(24):7576-80.
63. Wu TD, Schiffer CA, Gonzales MJ, Taylor J, Kantor R, Chou S, et al. Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. Journal of virology. 2003;77(8):4836-47.
64. Ghosh AK, Osswald HL, Prato G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. Journal of medicinal chemistry. 2016;59(11):5172-208.
65. Velazquez-Campoy A, Muzammil S, Ohtaka H, Schon A, Vega S, Freire E. Structural and thermodynamic basis of resistance to HIV-1 protease inhibition: implications for inhibitor design. Current Drug Targets-Infectious Disorders. 2003;3(4):311-28.
66. Naicker P, Stoychev S, Dirr HW, Sayed Y. Amide hydrogen exchange in HIV–1 subtype B and C proteases–insights into reduced drug susceptibility and dimer stability. Wiley Online Library; 2014. p. 5395-410.

67. Watson JD, Myers RM, Caudy AA, Witkowski JA. Recombinant DNA: genes and genomes: a short course: Macmillan; 2007.
68. Miles JS, Wolf CR. Principles of DNA cloning. BMJ: British Medical Journal. 1989;299(6706):1019.
69. Glick BR, Patten CL. Molecular biotechnology: principles and applications of recombinant DNA: John Wiley & Sons; 2022.
70. Glover DM, Glover DM. The principles of cloning DNA. Gene Cloning: The Mechanics of DNA Manipulation. 1984:1-20.
71. Srinivasan S, Gunasekaran P, Rajendhran J. Fundamentals of Molecular Biology. Current Developments in Biotechnology and Bioengineering: Elsevier; 2017. p. 59-80.
72. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Frontiers in microbiology. 2014;5:172.
73. Prelich G. Gene overexpression: uses, mechanisms, and interpretation. Genetics. 2012;190(3):841-54.
74. El-Gayar K. Principles of recombinant protein production, extraction and purification from bacterial strains. International Journal of Microbiology and Allied Sciences (IJOMAS). 2015;2(2):18-33.
75. Hedhammar M, Karlström AE, Hober S. Chromatographic methods for protein purification. Stockholm: Royal Institute of Technology. 2006:1-31.
76. Ekinci D. Chemical Biology: BoD–Books on Demand; 2012.
77. Yamamoto S, Nakanishi K, Matsuno R. Ion-exchange chromatography of proteins: CRC Press; 1988.
78. Queiroz J, Tomaz C, Cabral J. Hydrophobic interaction chromatography of proteins. Journal of biotechnology. 2001;87(2):143-59.
79. Cuatrecasas P, Wilchek M, Anfinsen CB. Selective enzyme purification by affinity chromatography. Proceedings of the National Academy of Sciences. 1968;61(2):636-43.
80. Uhlén M. Affinity as a tool in life science. Biotechniques. 2008;44(5):649-54.
81. Corbett RJ, Roche RS. Use of high-speed size-exclusion chromatography for the study of protein folding and stability. Biochemistry. 1984;23(8):1888-94.
82. Hansen J, Billich S, Schulze T, Sukrow S, Moelling K. Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. The EMBO Journal. 1988;7(6):1785-91.
83. Raynal B, Lenormand P, Baron B, Hoos S, England P. Quality assessment and optimization of purified protein samples: why and how? Microbial cell factories. 2014;13(1):1-10.
84. Schägger H. Tricine–sds-page. Nature protocols. 2006;1(1):16-22.
85. Adrio JL, Demain AL. Microbial Enzymes: Tools for Biotechnological Processes. Biomolecules. 2014;4(1):117-39.
86. Jencks WP. Catalysis in chemistry and enzymology: Courier Corporation; 1987.
87. Briggs GE, Haldane JBS. A note on the kinetics of enzyme action. Biochemical journal. 1925;19(2):338.
88. Johnson KA, Goody RS. The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. Biochemistry. 2011;50(39):8264-9.
89. Bisswanger H. Enzyme kinetics: principles and methods: John Wiley & Sons; 2017.
90. Claveria-Gimeno R, Vega S, Abian O, Velazquez-Campoy A. A look at ligand binding thermodynamics in drug discovery. Expert Opinion on Drug Discovery. 2017;12(4):363-77.
91. Amaral M, Kokh D, Bomke J, Wegener A, Buchstaller H, Eggenweiler H, et al. Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. Nature communications. 2017;8(1):2276.
92. O'Brien R, Markova N, Holdgate GA. Thermodynamics in drug discovery. Applied biophysics for drug discovery. 2017:7-28.
93. Bronowska AK. Thermodynamics of ligand-protein interactions: implications for molecular design. Thermodynamics-Interaction Studies-Solids, Liquids and Gases: IntechOpen; 2011.

94. Dunitz JD. Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. Chemistry & biology. 1995;2(11):709-12.
95. Duff Jr MR, Grubbs J, Howell EE. Isothermal titration calorimetry for measuring macromolecule-ligand affinity. JoVE (Journal of Visualized Experiments). 2011(55):e2796.
96. Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, et al. Molecular biology of the cell 6th ed (new york, ny: Garland science). 2014.
97. Kauzmann W. The three dimensional structures of proteins. Biophysical Journal. 1964;4(1):43-54.
98. Krauss IR, Merlino A, Vergara A, Sica F. An overview of biological macromolecule crystallization. International journal of molecular sciences. 2013;14(6):11643-91.
99. Su X-D, Zhang H, Terwilliger TC, Liljas A, Xiao J, Dong Y. Protein crystallography from the perspective of technology developments. Crystallography reviews. 2015;21(1-2):122-53.
100. Shibata K. SPECTROPHOTOMETRY OF INTACT BIOLOGICAL MATERIALS ABSOLUTE AND RELATIVE MEASUREMENTS OF THEIR TRANSMISSION, REFLECTION AND ABSORPTION SPECTRA. The Journal of Biochemistry. 1958;45(8):599-623.
101. Renjini A, Dileep D. Spectrophotometry and Spectrometry-Concept and Applications. Ijariie; 2017.
102. Bose A, Thomas I, Abraham E. Fluorescence spectroscopy and its applications: A Review. International journal of advances in pharmaceutical analysis. 2018;8(1):1-8.
103. Pignataro MF, Herrera MG, Dodero VI. Evaluation of peptide/protein self-assembly and aggregation by spectroscopic methods. Molecules. 2020;25(20):4854.
104. Zhuang X, Ha T, Kim HD, Centner T, Labeit S, Chu S. Fluorescence quenching: A tool for single-molecule protein-folding study. Proceedings of the National Academy of Sciences. 2000;97(26):14241-4.
105. Johnson Jr WC. Secondary structure of proteins through circular dichroism spectroscopy. Annual review of biophysics and biophysical chemistry. 1988;17(1):145-66.
106. Kelly SM, Price NC. The use of circular dichroism in the investigation of protein structure and function. Current protein and peptide science. 2000;1(4):349-84.
107. Lathe G, Ruthven C. The separation of substances on the basis of their molecular weights, using columns of starch and water. The Biochemical Journal. 1955;60(4):xxxiv.
108. Hong P, Koza S, Bouvier ES. A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. Journal of liquid chromatography & related technologies. 2012;35(20):2923-50.
109. Malviya R, Bansal V, Pal OP, Sharma PK. High performance liquid chromatography: a short review. Journal of global pharma technology. 2010;2(5):22-6.
110. Saponaro A. Isothermal titration calorimetry: a biophysical method to characterize the interaction between label-free biomolecules in solution. Bio-protocol. 2018;8(15):e2957-e.
111. Harris G, Scott DJ. Isothermal titration calorimetry: feeling the heat. The Biochemist. 2019;41(2):4-7.
112. Ghai R, Falconer RJ, Collins BM. Applications of isothermal titration calorimetry in pure and applied research—survey of the literature from 2010. Journal of Molecular Recognition. 2012;25(1):32-52.
113. Leavitt S, Freire E. Direct measurement of protein binding energetics by isothermal titration calorimetry. Current opinion in structural biology. 2001;11(5):560-6.
114. Liang Y. Applications of isothermal titration calorimetry in protein science. Acta biochimica et biophysica Sinica. 2008;40(7):565-76.
115. Garman EF. Developments in x-ray crystallographic structure determination of biological macromolecules. Science. 2014;343(6175):1102-8.
116. Smyth M, Martin J. x Ray crystallography. Molecular Pathology. 2000;53(1):8.
117. Song C, Zhang S, Huang H. Choosing a suitable method for the identification of replication origins in microbial genomes. Frontiers in MICROBIOLOGY. 2015;6.

118. Brito JA, Archer M. Structural biology techniques: X-ray crystallography, cryo-electron microscopy, and small-angle X-ray scattering. Practical approaches to biological inorganic chemistry: Elsevier; 2020. p. 375-416.
119. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997;25(17):3389-402.
120. Yang H, Nkeze J, Zhao RY. Effects of HIV-1 protease on cellular functions and their potential applications in antiretroviral therapy. Cell & bioscience. 2012;2:1-8.
121. Nguyen H-LT, Nguyen TT, Vu QT, Le HT, Pham Y, Le Trinh P, et al. An efficient procedure for the expression and purification of HIV-1 protease from inclusion bodies. Protein expression and purification. 2015;116:59-65.
122. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. nature. 1970;227(5259):680-5.
123. Boulanger Y, Sené, cal L, Sauvé G. Solution structure of the HIV protease inhibitor acetyl-pepstatin as determined by NMR and molecular modeling. Journal of Biomolecular Structure and Dynamics. 1997;14(4):421-8.
124. Woody RW. [4] Circular dichroism. Methods in enzymology. 1995;246:34-71.
125. CARMEL A, YARON A. An intramolecularly quenched fluorescent tripeptide as a fluorogenic substrate of angiotensin-i-converting enzyme and of bacterial dipeptidyl carboxypeptidase. European Journal of Biochemistry. 1978;87(2):265-73.
126. Eisenthal R, Danson MJ, Hough DW. Catalytic efficiency and kcat/KM: a useful comparator? Trends in biotechnology. 2007;25(6):247-9.
127. Velazquez-Campoy A, Freire E. Isothermal titration calorimetry to determine association constants for high-affinity ligands. Nature protocols. 2006;1(1):186-91.
128. Krainer G, Keller S. Single-experiment displacement assay for quantifying high-affinity binding by isothermal titration calorimetry. Methods. 2015;76:116-23.
129. Evans P. Scaling and assessment of data quality. Acta Crystallographica Section D: Biological Crystallography. 2006;62(1):72-82.
130. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallographica Section D: Biological Crystallography. 2011;67(4):282-92.
131. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Randy J. Read. 2007."Phaser Crystallographic Software.". Journal of Applied Crystallography.
132. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta crystallographica section D: biological crystallography. 2004;60(12):2126-32.
133. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallographica Section D: Structural Biology. 2019;75(10):861-77.
134. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica Section D: Biological Crystallography. 2010;66(1):12-21.
135. Laskowski R, MacArthur M, Moss D, Thornton J. IUCr. PROCHECK: a program to check the stereochemical quality of protein structures urn: issn. 1993:0021-8898.
136. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global trends in molecular epidemiology of HIV-1 during 2000–2007. AIDS (London, England). 2011;25(5):679.
137. Wallis CL, Mellors JW, Venter WD, Sanne I, Stevens W. Protease inhibitor resistance is uncommon in HIV-1 subtype C infected patients on failing second-line lopinavir/r-containing antiretroviral therapy in South Africa. AIDS research and treatment. 2011;2011.
138. Maseko SB, Govender D, Govender T, Naicker T, Lin J, Maguire GE, et al. Optimized procedure for recovering HIV-1 protease (C-SA) from inclusion bodies. The protein journal. 2019;38:30-6.
139. Volontè F, Piubelli L, Pollegioni L. Optimizing HIV-1 protease production in Escherichia coli as fusion protein. Microbial cell factories. 2011;10(1):1-10.

140. Chattopadhyay D, Evans D, Deibel Jr M, Vosters A, Eckenrode F, Einspahr H, et al. Purification and characterization of heterodimeric human immunodeficiency virus type 1 (HIV-1) reverse transcriptase produced by in vitro processing of p66 with recombinant HIV-1 protease. Journal of Biological Chemistry. 1992;267(20):14227-32.
141. Hui JO, Tomasselli AG, Reardon IM, Lull JM, Brunner DP, Tomich C-SC, et al. Large scale purification and refolding of HIV-1 protease from Escherichia coli inclusion bodies. Journal of protein chemistry. 1993;12:323-7.
142. Ramlucken U, Babu Naidu KS, Govender P. Improved Production of HIV-1 Subtype C Protease from Transgenic. The Open Microbiology Journal. 2021;15(1).
143. Mosebi S, Morris L, Dirr HW, Sayed Y. Active-site mutations in the South African human immunodeficiency virus type 1 subtype C protease have a significant impact on clinical inhibitor binding: kinetic and thermodynamic study. Journal of virology. 2008;82(22):11476-9.
144. Noel AF, Bilsel O, Kundu A, Wu Y, Zitzewitz JA, Matthews CR. The folding free-energy surface of HIV-1 protease: insights into the thermodynamic basis for resistance to inhibitors. Journal of molecular biology. 2009;387(4):1002-16.
145. Eche S, Kumar A, Sonela N, Gordon ML. Acquired HIV-1 Protease Conformational Flexibility Associated with Lopinavir Failure May Shape the Outcome of Darunavir Therapy after Antiretroviral Therapy Switch. Biomolecules. 2021;11(4):489.
146. Williams A, Basson A, Achilonu I, Dirr HW, Morris L, Sayed Y. Double trouble? Gag in conjunction with double insert in HIV protease contributes to reduced DRV susceptibility. Biochemical Journal. 2019;476(2):375-84.
147. Du X, Li Y, Xia Y-L, Ai S-M, Liang J, Sang P, et al. Insights into protein–ligand interactions: mechanisms, models, and methods. International journal of molecular sciences. 2016;17(2):144.