Impact of the hinge region L38↑H↑L double insertion mutation on the structure and function of the HIV-1 South African subtype C protease

By

Mpho Setshedi

(1353021)

Dissertation

Submitted in partial 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

Co-supervisor: Dr. R. Pandian

Academic advisor: Dr V. Meyer

July 2022

Declaration

I, **Mpho Setshedi (1353021)**, am a student registered for the degree of Master of Science in Biochemistry at the University of the Witwatersrand in the academic year 2022

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.
- I have not submitted this work before for any other degree or examination at this or any other 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:

Date: 18 July 2022

Abstract

The development of mutations in the human immunodeficiency virus (HIV) genome has become a crucial factor in limiting antiretroviral therapy for HIV/AIDS treatment. By the end of June 2021, approximately 37 million people globally were infected with HIV. South Africa, in particular, bears the brunt of the HIV-1 subtype C epidemic with 7.7 million infections. The HIV protease is a homodimeric protein that naturally contains 99 amino acids per monomeric subunit. The protease is vital in the HIV life cycle because it cleaves Gag and Gag-Pol precursor polyproteins into proteins necessary for viral assembly, maturation, and infection. Herein, we performed a comparative study between the HIV-1 subtype C protease (wild-type) and a protease containing a double amino acid insertion (histidine and leucine at codon 38), L38↑H↑L. Each recombinant protease was overexpressed and isolated from *E. coli* BL21 (DE3) pLysS cells, and purified using ion-exchange chromatography. Far-UV circular dichroism spectra of the WT-CSA and L38↑H↑L protease displayed a minimum at 218 nm which was indicative of a predominantly beta-sheeted protein. Size exclusion chromatography results indicated that the L38↑H↑L protease has a homodimeric molecular weight of approximately 22 kDa which is consistent with characteristics pertaining to the WT-CSA protease. Steady-state enzyme kinetic assays were performed by using a fluorogenic substrate that mimics a cleavage site on the Gag polyprotein. The L38↑H↑L protease displayed a 15-fold reduction in K_M and a 1.25-fold reduction in the catalytic turnover number, k_{cat} relavite to the WT-CSA protease. The L38↑H↑L protease was found to be 11 times more efficient in catalyzing the hydrolysis of the fluorogenic substrate relative to the WT-CSA protease. The L38↑H↑L protease displayed a 2.25-fold and 0.71-fold reduction in V_{Max} and specific activity relative to the WT-CSA protease. The data suggested that amino acids in the hinge region of HIV-1 proteases indirectly affect enzyme catalysis and substrate specificity but has no effect on the structural stability of the L38↑H↑L protease. Enzyme inhibition studies in the presence of Saquinavir, Darunavir and Atazanavir showed that the L38↑H↑L binds the inhibitors more tightly relative to the WT-CSA protease. This was confirmed by the lower IC_{50} concentrations of the respective protease inhibitors relative to the WT-CSA protease. Thermodynamics data obtained from performing displacement isothermal titration calorimetry showed that the L38↑H↑L protease displayed a 2-fold increase (-79.02 ± 6.81 kJ/mol) in ΔH when bound to Atazanavir and a 1-fold reduction in ΔH when bound to

Saquinavir (-9.35 ±3.1 kJ/mol) and Darunavir (-27.61 ± 1.76 kJ/mol) relative to the WT-CSA protease. The binding of Saquinavir and Darunavir to the L38↑H↑L protease was found to be entropically unfavourable (where -TΔS was -32.81 ± 6.40 kJ/mol and -22.74 ± 3.70 kJ/mol, respectively), whereas the binding of the protease to Atazanavir was entropically unfavourable (-TΔS was 37 ± 4.83 kJ/mol). Upon analysis of the K_d, it was noted that L38↑H↑L protease displayed a 2-fold increased drug susceptibility towards Saquinavir (79 ± 7.50 nM), and 2-fold reduced drug susceptibility towards Atazanavir (83.63 ± 7.10 nM), while the drug susceptibility for Darunavir (3.3 ± 1.20 nM) remained unchanged relative to the WT-CSA protease. Due to the important role of the hinge and flap region in accommodating substrates and inhibitors into the active site of the protease, the mutations affected the kinetics of substrate hydrolysis and the thermodynamics of drug binding.

Acknowledgements

To my supervisor and mentor, Professor Yasien Sayed, thank you for your patience, understanding and support towards my research. You continue to inspire me and assist in the expansion of my knowledge and passion for biochemistry.

To my co-supervisor, Dr. Pandian, thank you for providing me with extensive insights into protein crystallography.

To my family and friends, specifically my mom, brother and Aasiya Lakhi, thank you for always listening to my biochemistry dilemmas and supporting me through difficult times. Thank you for your love and support.

To the members of the Protein Structure-Function research Unit (PSFRU) at the University of the Witwatersrand, thank you for the guidance, and support.

To the National Research Foundation (NRF), thank you for the financial assistance.

Table of Contents

List of Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
CCR5	C-C chemokine receptor type 5
C-SA	HIV-1 South African Subtype C
CD	Circular Dichroism
CXCR4	C-X-C chemokine receptor type 4
DEAE	Diethylaminoethyl
DMSO	Dimethly-sulfoxide
DTT	Dithiothreitol
ε	Molar extinction coefficient
EDTA	Ethylenediaminetetraacetic acid
ΔG	Change in Gibbs free energy
ΔH	Change in enthalpy
ΔS	Change in entropy
Far-UV CD	Far-ultraviolet Circular Dichroism
FDA	Food and Drug Administration
HIV	Human Immunodeficiency Virus
IN	Integrase
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
K_a	Binding affinity constant
k_{cat}	Catalytic turnover
k_{cat}/K_M	Catalytic efficiency
kDa	kilo-Dalton
K_i	Inhibition constant in the presence of substrate
K_M	Michaelis-Menten constant
N	Stoichiometry
NOP	Naturally Occurring Polymorphism
OD_{600nm}	Optical Density at a wavelength of 600 nm

PDB	Protein Data Bank
PI	Protease Inhibitor
PMSF	Phenylmethylsulfonyl fluoride
PR	Protease
RT	Reverse Transcriptase
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SE-HPLC	Size Exclusion-High Performance Liquid Chromatography
ssRNA	Single-stranded ribonucleic acid
L38↑H↑L	HIV protease containing two amino acid (Histidine and Leucine) insertion mutations at codon 38
UNAIDS	United Nations AIDS Organisation
V_{max}	Maximum velocity

Amino Acids are named according to the IUPAC three and one-letter codes.

The words WT-CSA protease and wild-type protease are used interchangeably.

The words L38↑H↑L protease and mutant protease are used interchangeably.

List of Figures

List of Tables

1. Introduction

1.1 HIV/AIDS epidemic

Human immunodeficiency virus (HIV) is one of the most pressing health concerns facing the modern world. At the end of June 2021, 37,7 million people globally were estimated to be infected with HIV and of those infected, only 28.2 million people were receiving antiretroviral therapy. Approximately 68% of these infections occurred in sub-Saharan Africa (UNAIDS, 2021). Among sub-Saharan countries, South Africa is the most affected, accounting for 19% of the global number of infections with 8.2 million infections and 126 755 AIDS-related deaths. The number of infections has increased by approximately one million from 2018 to 2019 (Stats SA, 2021). HIV remains a leading cause of death worldwide and the leading cause of death globally among women of reproductive age (UNAIDS, 2021) This alarming statistic shows how HIV/AIDS remains a health challenge in sub-Saharan Africa and that designing effective treatment is of utmost importance.

HIV is a *Lentivirus* that falls within the *Retroviridae* family. HIV was found to be the etiological agent of Acquired Immunodeficiency Syndrome (AIDS), a state in which the immune system is left severely compromised, leaving the person susceptible to opportunistic illnesses such as tuberculosis and Kaposi sarcoma (Weiss *et al.*, 1985 and Powell *et al.*, 2016). HIV can be categorised into two types: HIV-1 and HIV-2. HIV-2 is limited to West Africa and is less virulent relative to HIV-1. HIV-1 is mostly found in sub-Saharan Africa, India, China and Brazil (Brodine *et al.*, 1995). The diversity among the HIV types is a result of the viral reverse transcriptase (RT), which introduces mutations into the genes of the retrovirus as well as the high replication rate of the virus which produces an average of ten billion virions in a day (Roberts *et al.,* 1988).

HIV-1 can be subdivided into four groups (M, N, O, and P). The M (Major) group is the most common and can be further classified into nine genetically distinct subtypes (A to D, F to H, J and K) and circulating recombinant forms (CRFs). CRFs are derived from the recombination of different subtypes in a single host (Osmanov *et al.,* 2002). The differences among the subtypes are due to naturally occurring polymorphisms (NOPs); that is, mutations that arise naturally in the genome of the virus (Velazquez-Campoy *et al.*, 2001). Among HIV-1 subtypes, subtype C is the prevalent subtype in sub-Saharan Africa (Buonaguro *et al.*, 2007). Subtype C accounts for approximately 46.6% of infections in the world (Hemelaar *et al.*, 2019). According to the Protein Data Bank (PDB), there are only four HIV-1 subtype C protease crystal structures available as compared to the vast number of crystallised subtype B protease structures. Only one of the four crystallised subtype C protease structures represent the wild-type South African protease (C-SA) (PDB ID: 3U71). Although HIV-1 subtype C is the most prevalent in the world, antiretroviral therapy (ART) drugs are not designed and tested against this subtype. This poses a question about whether the treatment is as effective in C subtypes as it is in other subtypes.

1.2 The replication cycle of HIV-1

HIV is an enveloped RNA virus of approximately 100 nm in diameter (Greene, 1993). The core of the virus consists of two identical copies of single-stranded RNA molecules which make up the viral genome. The HIV replicative cycle consists of the early phase in which infection and integration occur and the late phase in which the transcription and translation of the viral genome occur (Turner and Summers, 1999). Routes of HIV infection are predominantly through mucosal surfaces such as male and female genital tracts, rectal surfaces and gut surfaces (Simon *et al.,* 2006; Shaw and Hunter, 2012; Showa *et al*., 2019).

Infection of the host cell begins when the viral gp120 and gp41 glycoprotein receptors recognise the CD4 (cluster of differentiation 4) receptor on immune cells such as macrophages, T cells and dendritic cells (Chad and Kim, 1998 and Cunningham *et al.,* 2010). The virus binds and fuses with the cell membrane by either using the C-C chemokine receptor type 5 (CCR5) or C-X-C chemokine receptor type 4 (CXCR4) of the host cell (Kwong, 1998) (Figure 1). The binding of viral gp120 envelope protein to the CD4 receptor causes a structural change in the gp120, exposing the binding site for the co-receptors. Once the co- receptor is bound, further structural rearrangements occur, mostly in gp41 transmembrane protein, which leads to viral entry. A cohort study by Ruiz-Mateos and colleagues (2018) found a relationship between a CCR5 delta 32 mutation and the inability of the virus to enter host cells. CCR5 delta 32 mutation is caused by a 32-base pair deletion in the gene encoding the CCR5 receptor (Marmor *et al.,* 2006). A distinct characteristic of HIV-1 subtype C is the preferential use of viral CCR5 receptors instead of CXCR4 receptors when binding and fusing with the host cell (Zhang *et al.,* 2002). The fusion of the viral envelope and the cell membrane allows for the translocation of the constituents of the viral core into the cytoplasm of the host cell. Translocation is proceeded by reverse transcribing the viral single-stranded RNA (ssRNA) into double-stranded DNA (dsDNA) by the viral reverse transcriptase.

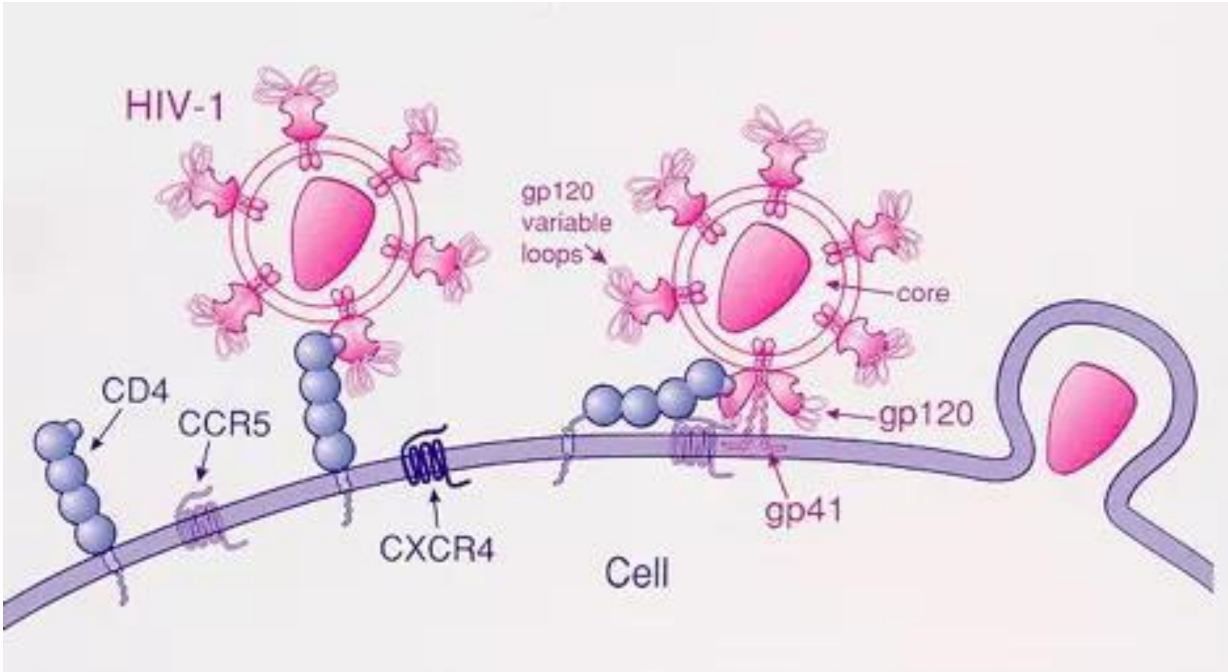


Figure 1. Mechanism of viral entry. Interactions between the viral envelope glycoprotein, gp120 and the cell surface receptor CD4 and CCR5 chemokine receptor are responsible for the fusion and entry of HIV-1 into host cells. A mutation in the CCR5 receptor has been shown to affect the ability of the virus to bind and fuse with the host cell (Marmor *et al*., 2006). The image was taken from:
https://www.nature.com/scitable/blog/viruses101/hiv_resistant_mutation/ [accessed on 17/02/2022].

The dsDNA is then transported into the nucleus where it is integrated into the genome of the host cell by the HIV-1 integrase (Farnet and Haseltine, 1991). The integration of viral DNA was thought to occur at random sites but, in fact, the viral dsDNA is incorporated at sites where actively transcribed genes are found (Schroder *et al.,* 2002). The incorporated viral genome is called a provirus. The provirus is transcribed and translated along with the genes of the host cell (Figure 2).

The genome of HIV-1 encodes for structural and functional proteins required for viral replication, assembly regulatory proteins. HIV-1 contains three open reading frames (ORFs) encoding for structural enzymes; namely, group-specific antigen (*gag*), *pol*, and *env*. The *gag* ORF encodes the Gag polyprotein that ultimately forms the virus capsid, nucleocapsid, and the matrix, the *pol* ORF encodes the Pol polyprotein that forms three retroviral enzymes; namely, reverse transcriptase (RT), protease (PR), and integrase (IN). The *env* gene encodes the viral envelope proteins such as the gp120 and gp41 glycoproteins. The HIV-1 protease is initially synthesised along with retroviral enzymes and other viral proteins as part of the Gag- Pol precursor polyprotein. The Gag, Gag-Pol polyproteins and two copies of the viral RNA genome are transported to the plasma membrane where they assemble and bud off from the cell, taking along with it some elements of the cell membrane (Henderson *et al.,* 1983). The budding mechanism generates non-infectious immature virus particles which only become mature upon processing of the Gag and Gag-Pol polyprotein (Freed, 1998). Maturation occurs when two Gag-Pol polyprotein chains align, allowing two protease monomers to dimerise and undergo autoproteolysis from the polyprotein (Navia *et al.,* 1989). The functional protease homodimer can now cleave the Gag-Pol and Gag polyproteins at eleven sites to generate structural and functional proteins needed to generate an infectious virus (Figure 3). The protease-mediated cleavage of the Gag-Pol polyprotein is a crucial step in the HIV replicative cycle. The binding cleft that runs through the active site of the protease is able to accommodate seven amino acids. This coincides with the specificity of the protease, which is determined by the nature of the four amino acids upstream and three amino acids downstream of the scissile bond in the substrate (Vella *et al*., 2012). For this reason, HIV protease is a drug target when designing antiretroviral therapy regimens.

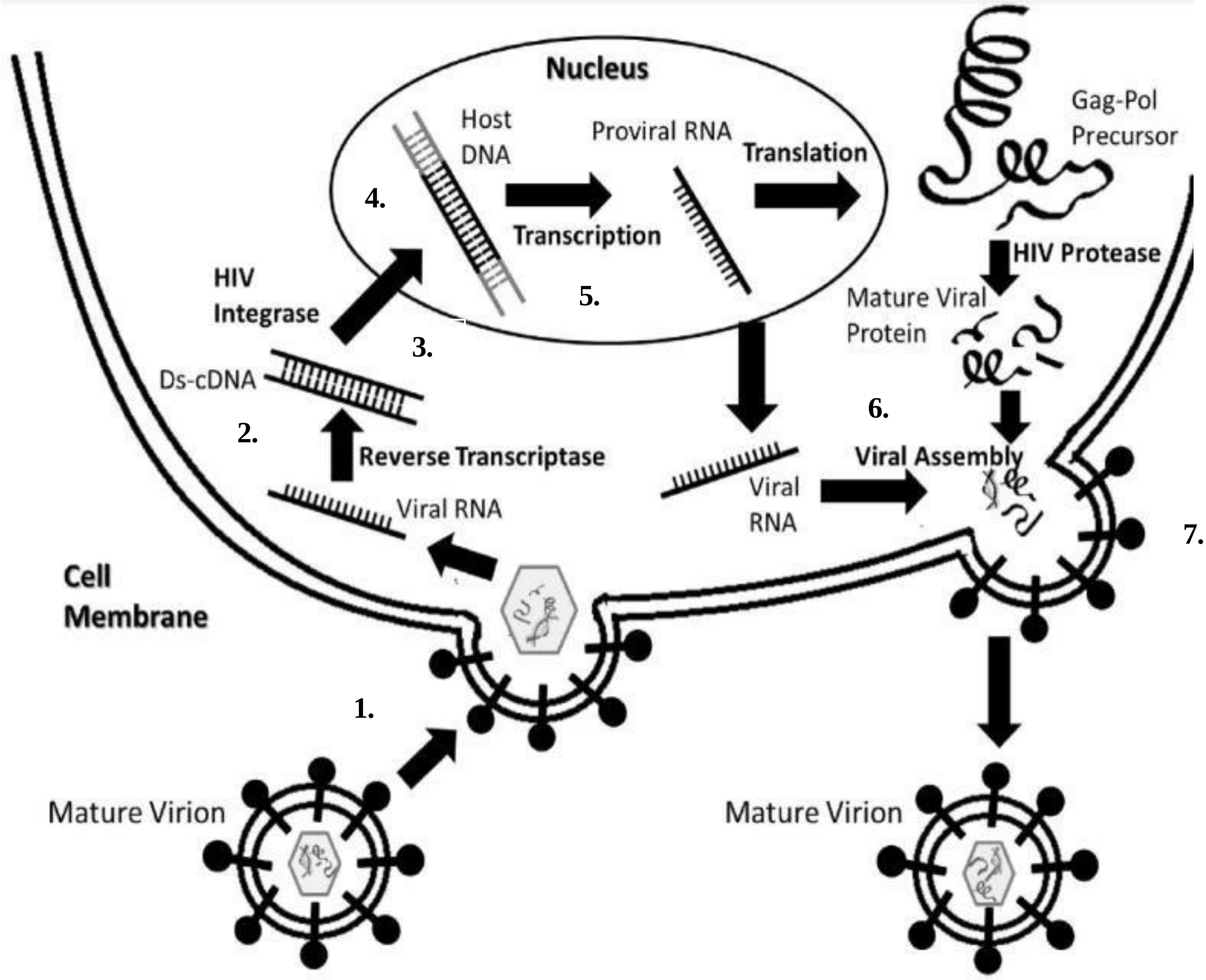


Figure 2. The replication cycle of HIV-1. The eight stages of the HIV life cycle are: 1) binding and fusion of virus, 2) reverse transcription of viral genome into double cDNA, 3) entrance of viral cDNA into nucleus, 4) integration of the viral genome into the genome of the host, 5) transcription and translation of viral genome into functional and structural proteins and polyprotein precursors, 6) provirus assembly and budding, 7) provirus release and virus maturation. The protease plays a critical role in viral assembly and maturation. Image was taken from Yang *et al*., 2012.

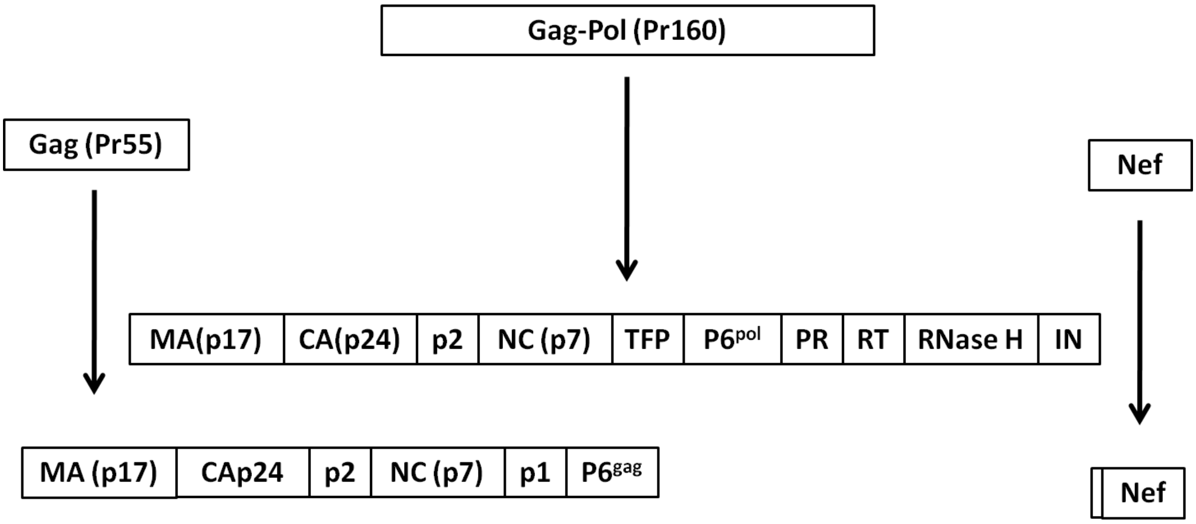


Figure 3. Schematic representation of the Gag and Gag-Pol polyproteins and Nef protein at the sites at which the HIV-1 protease cleaves. Gag and Gag-Pol polyprotein cleavage yields functional viral enzymes and structural proteins required for viral assembly and maturation. The alignment of two Gag-Pol polyproteins results in the dimerisation of protease monomers to generate a catalytically functional protease. HIV-1 protease cleaves the Gag polyprotein at five sites (MA, CA, p2, NC, P1, and P6^{gag}) and the Gag-Pol polyprotein at six sites (NC, TFP, p6^{pol}, PR, RT, RNase H and IN and a single site in Nef (Côté *et al.,* 2001). The image was taken from de Oliveira *et al.,* 2003.

1.3 Current treatment of HIV/AIDS

HIV/AIDS is a lifelong disease and there is currently no cure for it. Antiretroviral therapy is currently the gold standard of HIV/AIDS treatment. Without treatment, the average survival time is approximately nine to eleven years (UNAIDS, 2021). Antiretroviral therapy is a combination of drugs that inhibit some part of the HIV replication cycle. The goals of antiretroviral therapy were to provide maximal suppression of the viral load, restore and preserve immune function, prolong life expectancy and improve quality of life (Simon *et al.*, 2006). The antiretroviral drugs consist of five classes, that is, nucleoside and nucleotide reverse transcriptase inhibitors (NRTIs and NtRTIs, respectively), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), integrase inhibitors also called integrase strand transfer inhibitors (InSTIs) and entry inhibitors. Their mechanisms of action are described in Table 1 below. Among the different classes, protease inhibitors are the most efficient in reducing the viral load in infected cells (Yang *et al*., 2012). The HIV protease was first suggested as a druggable target in the treatment of HIV/AIDS when a frameshift mutation in the *pol* gene encoding the protease resulted in a formation of non-infectious immature virions (Kramer *et al.,* 1986). This mutation prevents the protease from cleaving the Gag and Gag-Pol polyproteins.

1.4 HIV-1 protease

The first scientists to obtain a crystal structure of HIV-1 protease were Navia *et al* in 1989. Four more structures were reported subsequently by Wlodawer *et al* (1989). HIV-1 protease is a member of the aspartyl protease family (E.C 3.4.23.16). The protease is a homodimeric globular protein consisting of two identical two-fold axis symmetrical monomers made up of 99 amino acid residues each (Wlodawer *et al.*, 1989). Aspartic proteases are characterised by conserved catalytic residues; Asp25, Thr26, and Gly27 in one monomer and Asp25', Thr26', Gly27' in the other monomer (Toh *et al*., 1985).

The secondary structure of the protease consists of 48% beta-sheets, 48% random coils and 4% alpha-helix (PDB ID: 3U71) (Naicker *et al.*, 2013). The protease has regions critical in the proteolysis of polyproteins; namely, the flaps region, hinge region, cantilever, fulcrum and dimeric interface and the active site (Figure 4). The active site (residues 23-30), found between the monomers, houses the catalytic triad in each monomer. The flaps (residues 45-57 in each monomer) are highly flexible structures that are found above the active site.

Table 1. Classes of antiretroviral drugs

Class	Abbreviation	Mechanism of action	Specific action
Nucleoside and nucleotide reverse transcriptase inhibitors	NRTIs and NtRTIs	Reverse transcriptase inhibition	Nucleic acid analogues mimic the normal building blocks of DNA, preventing transcription of viral RNA into complimentary DNA
Non-nucleoside reverse transcriptase inhibitors	NNRTIs	Reverse transcriptase inhibition	Alter the conformation of the catalytic site of reverse transcriptase and directly inhibits its action
Protease inhibitors	PIs	Protease inhibition	Inhibit the final stages of viral assembly and maturation, resulting in the formation of inactive viral particles.
Integrase strand transfer inhibitors	InSTIs	Inhibition of viral integration	Prevent the transfer of viral DNA strands into the host genomic DNA
Entry inhibitors	-	Entry inhibition	Bind to viral gp41, gp120, host cell $CD4^+$ or CCR5 receptors.

(Table adapted from Meitjes *et al*., 2017)

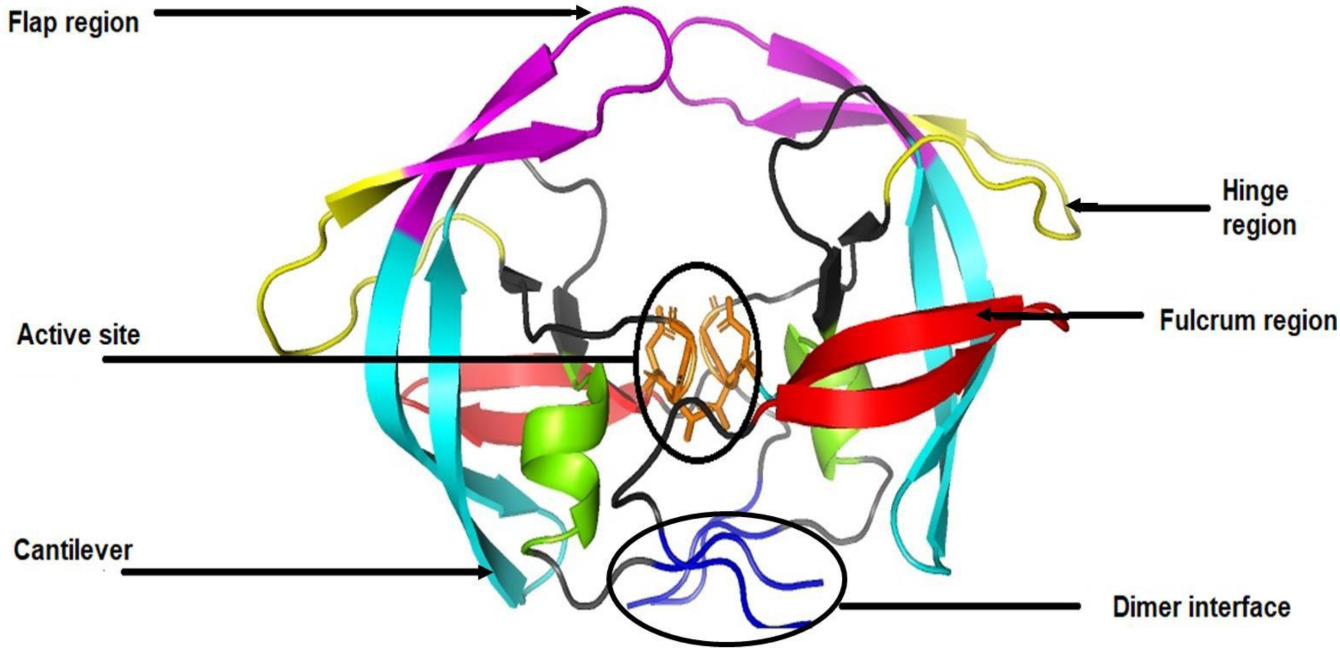


Figure 4. Ribbon representation of the crystal structure of HIV-1 South African subtype C protease in the unbound form. The secondary structures consist mainly of β- strands (shown in light blue, purple and yellow) and two alpha-helices (green). Flap conformational movements are controlled by hinge regions (shown in yellow) consisting of residues 35-42 and 35'-42' and residues 57-61 and 57'-61' which act as a cantilever (light blue) and assists in the conformational change of the flaps. The fulcrum (red) consists of residues 10-23. The active site (shown in orange) houses the catalytic residues: Asp25, Thr26, and Gly27 in one monomer and Asp25', Thr26', Gly27' in the other. The dimer interface (shown in dark blue) lies below the active site, and it consists of four-stranded, antiparallel β-sheets at the C- and N-terminus residues1-4 and 96-99, respectively of each monomer. This figure was generated using PyMOL with PDB ID: 3U71 (Naicker *et al.*, 2013).

The flaps assist in the binding of the polyproteins to the active site (Gustchina and Weber, 1990). Molecular dynamics (MD) simulation studies indicate that the protease flap regions can adopt open, semi-open and closed conformations in solution depending on whether the polyprotein or inhibitor is bound (Nicholson *et al.,* 1995). Spin-labelled pulse EPR spectroscopy measurements, a technique for investigating the structure and local dynamics of proteins using electron spin resonance, indicate that the open and semi-open conformations are seen mostly in the HIV-1 subtype C relative to other subtypes (Babe *et al*., 1992). The hinge regions (residues 35-42 and 57-61) assist in the conformational movements of the flaps. The dimer interface, which lies at the bottom of the active site, stabilises the homodimeric form of the protease. Dimerisation of the two monomers is required for the protease to become catalytically active. It has been indicated that a disruption of the dimer interface results in the collapse of the globular protein into unstable monomer structures (Xu *et al.,* 1997 and Naicker *et al*., 2014).

Stabilisation of the dimer interface is achieved by the interactions between the four β hairpin loops located in the N-terminus (residue 1-4) and C-terminus (residue 96-99). Due to the critical role of the protease in generating mature and infectious human immunodeficiency virus, it has become an attractive target in designing anti-HIV drugs that inhibit the processing of Gag and Gag-Pol polyprotein.

1.5 The catalytic mechanism of HIV-1 protease

HIV-1 protease follows a general acid-base mechanism of catalysis. The widely accepted model of aspartic protease catalytic mechanism is based on the crystal structure of *Rhizoys Chinensis* aspartic protease complexed with a reduced peptide inhibitor (Suguna *et al.,* 1987).

According to the Schechter and Berger convention (Schechter and Berger, 1967), HIV-1 protease cleaves a peptide bond, also known as a scissile bond, that is found between P1 and P1' where P refers to the peptide. The amino acid residues found towards the N-terminus of the polyprotein are termed P1, P2, and P3 whereas the amino acid residues found at the C- terminus are termed P1', P2' and P3'. The corresponding sub-sites in the protease are termed from the central aspartates S1, S2, S3, and S1', S2' and S3' (Figure 5 A).

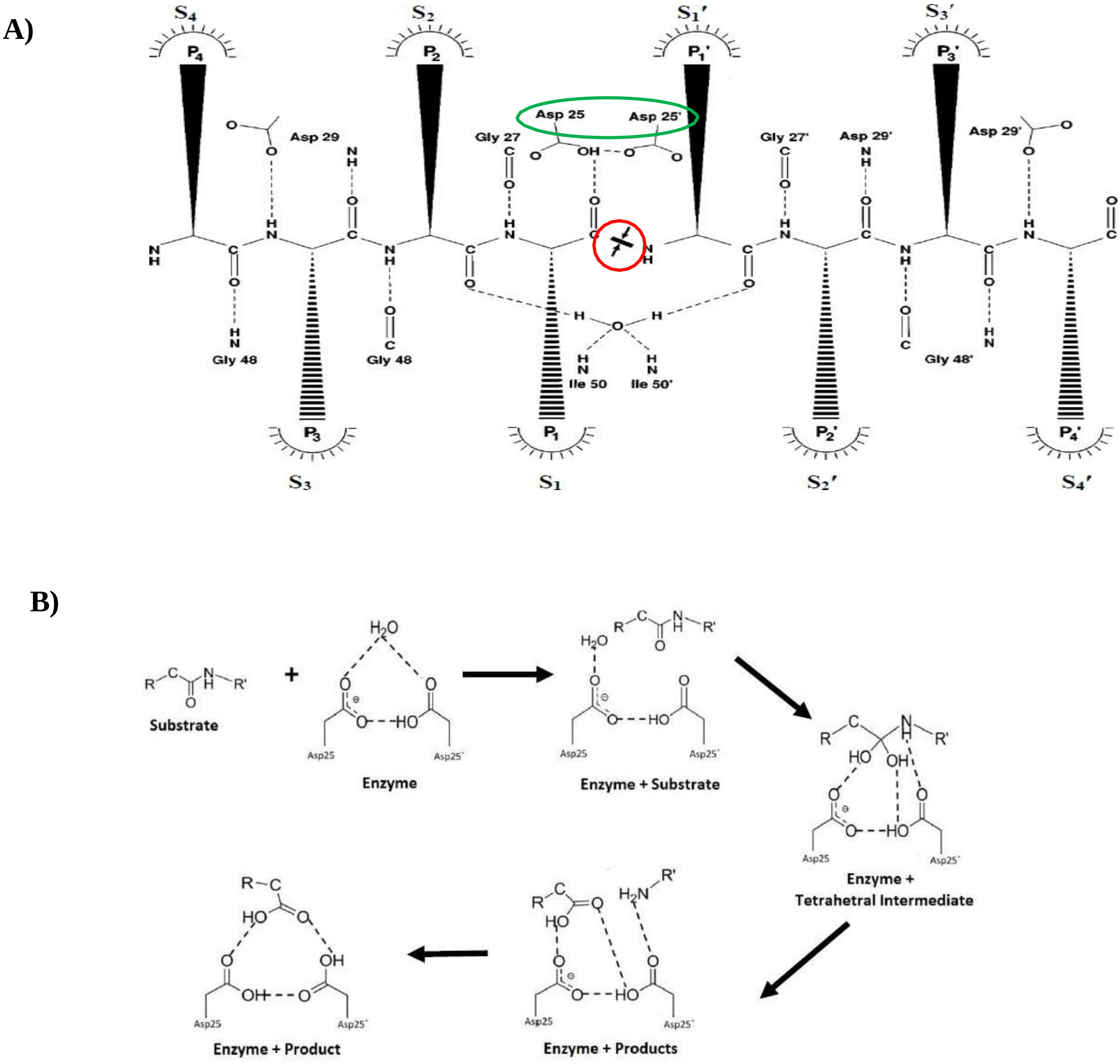


Figure 5. A) A schematic diagram showing the nomenclature used to assign amino acid residues of peptide substrates (P) as well as the corresponding binding sites (S) on the protease enzyme and the location of the scissile bond. The catalytic residues Asp 25 and Asp 25' are highlighted in the green circle (). The highlighted circle in red () of the scissile peptide bond B) Catalytic mechanism of the HIV-1 WT-CSA protease follows a general acid-base mechanism of catalysis. Two products and a free enzyme are generated at the end of the reaction. Image A) was taken from Wlodawer and Vondrasek, 1998. B) was taken from Shen *et al.*, (2012).

The catalytic residues Asp 25 and Asp 25' (Figure 5 B) polarise a nearby water molecule located at the catalytic site. One Asp residue (Asp-COO^-) acts as a general base to polarise the water molecule while the second Asp residue (Asp-COOH) acts as a general acid to protonate the leaving amine product. This renders the oxygen atom on the water molecule electronegative. The water molecule performs a nucleophilic attack on the amide carbonyl carbon on the scissile bond of the substrate. The substrate is then converted into a gem-diol intermediate in the transition state. The unstable tetrahedral intermediate collapses to form two products and a free enzyme molecule (Figure 5 B) (Hyland *et al.,* 1991).

1.6 HIV protease inhibitors

Of all the drugs used in the treatment of HIV/AIDS, protease inhibitors are the most efficient (Wang *et al.*, 2001). As previously noted, the catalytic cleavage of the polyproteins results in the formation of viral enzymes and proteins needed for viral maturation. Protease inhibitors improve the clinical outcome of patients by firstly, preventing the assembly and maturation of the virus, thus preventing viral replication and decreasing the viral load. Secondly, protease inhibitors inhibit the protease-mediated cleavage of host cell proteins which reduces protease-related cytotoxicity, apoptosis and necrosis in an infected host cell (Korant *et al.,* 1998 and Sainski *et al.,* 2011). The first HIV-1 protease crystal structure by Navia and colleagues (1989) has enabled the structure-based design of ten US Food Drug and Administration (FDA)-approved HIV protease inhibitors (PIs); namely: Atazanavir, Amprenavir, Darunavir, Indinavir, Lopinavir, Nelfinavir, Ritonavir, Saquinavir, Tipranavir and Fosamprenavir. All the protease inhibitors, apart from Tipranavir, are peptidomimetic hydroxyethylamine or hydroxyethelene inhibitors (Wang *et al.,* 2001).

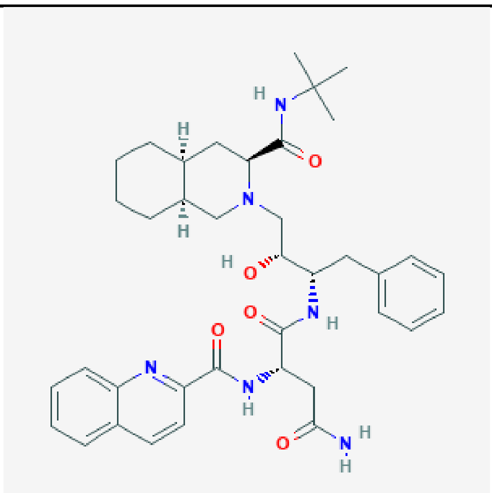
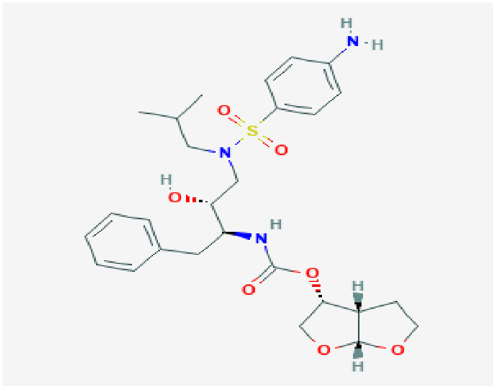
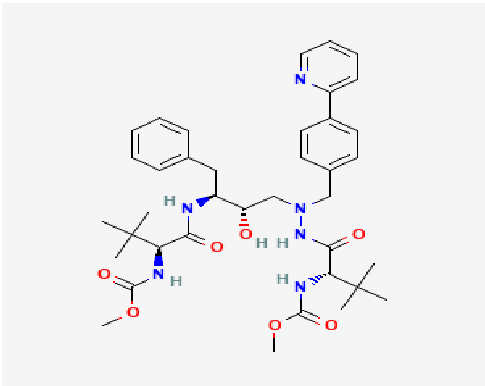
A peptidomimetic inhibitor is a small non-cleavable peptide chain that mimics a section of the Gag polyprotein (Marshall and Ballante, 2017). HIV-1 PIs are transition-state analogues of the polyprotein substrates. The transition-state of the substrate has a significantly larger affinity for the enzyme active site than the polyprotein substrate (Wolfenden, 1999). However, PIs are not 100% efficient at inhibiting the catalytic function of the protease (Titanji *et al.,* 2013). A small amount of viral replication still occurs in the presence of PIs. This replication process, combined with the effects of the error-prone RT, results in the development of mutations such as naturally occurring polymorphisms (NOPs) and mutations that confer resistance to PIs. Despite a wide variety of FDA approved protease inhibitors in use, there has been the development of resistance towards these inhibitors as a result of the development of mutations in the protease gene. This results in a decrease in the potency of the protease inhibitors (Velazquez-Campoy *et al*., 2003). The protease inhibitors that will be used in this study are Saquinavir, Darunavir and Atazanavir (Table 2). Darunavir is exclusively prescribed for salvage therapy after failure on protease inhibitor-based second- line regimen.

1.7 Protease inhibitor resistance

HIV-1 is prone to mutations due to the high replication rate of the virus as well as the error-prone viral reverse transcriptase due to the lack of proof reading abilities (Kozal *et al.*, 1996). Many of the mutations that occur within the protease sequence are NOPs that are commonly found in the viral genome while some mutations result in PI resistance.

PI resistance occurs as a result of continued viral replication in the presence of inhibitors, which allows for the accumulation of mutations in the flaps and hinge regions (Kantor and Katzenstein, 2004 and Titanji *et al.,* 2013). Mutations such as deletions, insertions, and substitutions in the viral genome result in changes in the amino acid sequence of the protease (Velazquez-Campoy *et al.,* 2001). These mutations alter the conformational shape of the substrate-binding site of the protease and/or flexibility of the flap region, resulting in a decreased binding affinity of the inhibitors for the enzyme but do not affect the binding affinity of the substrate (Debouck *et al.*, 1987). HIV-1 protease mutations can be classified into primary and secondary mutations. Primary mutations occur at residues that are directly involved in the catalysis of protease substrates (Chen *et al.*, 1995). Primary mutations usually occur near or at the substrate-binding cleft. Primary mutations alter the interactions between the protease and the protease inhibitor but rarely involve residues involved in active catalysis of the Gag and Gag-Pol substrate (Kozisek *et al.,* 2007). Mutations that occur at regions that are not directly involved in catalysis, also known as secondary mutations, contribute to decreased PI susceptibility (Olsen *et al*., 1999). Like primary mutations, secondary mutations arise as a result of PI drug therapy. Grossman and colleagues (2004) indicated that when subtype C protease was exposed to Nelfinavir, Indinavir and Saquinavir, resistance-related mutations occurred more frequently in subtype C than in subtype B. To date, 45 residues out of 99 residues that compose the HIV-1 subtype C protease monomer have been found to cause protease inhibitor resistance (Turner *et al.,* 2004). Some of the common mutations are mentioned in Table 3 below. All currently available protease inhibitors have specific mutations associated with them (Virgil *et al.*, 2010).

Table 2. Protease inhibitors used in this investigation and their properties

Generic name	Abbreviation	Properties	Structure*
Saquinavir	SQV	The first protease inhibitor approved by the FDA. (Roberts *et al.*, 1990). It is a highly specific inhibitor of HIV-1 proteases. Contains a hydroxylethylamine transition-state analogue replacing the cleavable scissile bond, a bulky decahydroquinoline in place of proline at position P1′ and a quinoline at position P3 (Figure 5 A)	
Darunavir	DRV	The latest protease inhibitor on the market. Binds to the backbone of HIV-1 protease which makes it difficult to acquire drug resistance against darunavir (Ghosh *et al.*, 2007). The hydrogen bonds it forms with the backbone of the HIV-1 protease slows down the development of drug resistance (Lefebvre and Schiffer, 2008)	
Atazanavir	ATV	Structural analog of the substrate transition state during which the bond between the phenylalanine and proline is broken (Lv *et al*., 2015) *Bis*-aryl moiety on the (hydroxethyl)hydrazine group makes it a potent inhibitor against wild-types and mutant forms of the HIV-1 protease (King *et al*., 2004)	

*Protease inhibitor chemical structures were downloaded from PubChem

(http://punchem.ncbi.nlm.nih.gov) [accessed on 17/02/2022]

Table 3. Drug resistance mutations

Position	Mutation	Protease inhibitors affected
L23	I	Nelfinavir
L24	I	Lopinavir, Saquinavir
D30	N	Nelfinavir, Darunavir
V32	I	All protease inhibitors except Saquinavir
L33	F	Lopinavir, Atazanavir, Darunavir
M36	I, F	Atazanavir, Tipranavir, Lopinavir, Nelfinavir, Indinavir, Fosamprenavir
M46	I	Amprenavir, Indinavir, Lopinavir, Nelfinavir, Saquinavir, Atazanavir, Darunavir
M46	L	Amprenavir, Indinavir, Lopinavir, Nelfinavir, Saquinavir, Atazanavir
I47	V	Atazanavir, Tipranavir, Lopinavir
I47	A	Atazanavir
G48	V, M	Lopinavir, Nelfinavir, Saquinavir
I50	L, V	Atazanavir, Darunavir, Lopinavir
F53	L	Atazanavir, Indinavir, Nelfinavir
I54	V, T, A, L, M	Fosamprenavir, Lopinavir, Atazanavir, Nelfinavir, Darunavir
G73	S, T	Atazanavir, Darunavir, Fosamprenavir
L76	V	Saquinavir, Atazanavir, Tipranavir
V82	A, T, F, S, L	Atazanavir, Fosamprenavir, Indinavir
I84	V, A, C	Tipranavir
N88	D, S	Atazanavir, Nelfinavir, Indinavir
L90	M	Atazanavir, Darunavir, Fosamprenavir, Saquinavir

(Table adapted from Mahdi *et al.*, 2015 and Shafer *et al.*, 2007, Stanford HIV Drug Resistance Database [https://hivdb.stanford.edu/cgi-bin/PRPosMutSummary.cgi] accessed on 18/07/2022)

1.8 Insertion mutations

Insertion mutations are associated with HIV-1 protease inhibitor resistance when found in combination with primary mutations and secondary mutations. These mutations can cause minor structural changes to the parts involved in accommodating the protease inhibitor into the substrate-binding cleft (Todd and Freire, 1998). The hinge region (Figure 4) is most prone to insertion mutations. Insertion mutations in the protease are caused by the slippage and stalling of the reverse transcriptase (Kim *et al.*, 2001). Insertion mutations rarely occur. The prevalence of insertion mutations in the WT-CSA protease has been found to be 2.5% (Pereira-Vaz *et al.*, 2009; Ledwaba *et al*., 2019). It has been shown that mutations in the HIV-1 protease may exhibit different inhibitor binding kinetics and energetics (Mosebi *et al*., 2008; Williams *et al.*, 2019; Zondagh *et al*., 2019). These mutations may alter the movement and flexibility of the hinge and flaps region which then affects the binding of substrates and inhibitors (Mittal *et al*., 2013). A molecular dynamics (MD) study by Naicker *et al.*, (2013) indicated that differences in protease inhibitor effectiveness may be explained by key differences in the flexibility of the hinge region and flap region of the protease caused by mutations. The stability of the protease hinge influences the movement of the closely associated flaps which play a role in accommodating the substrate into the binding cleft. Previously investigated insertion mutation N37T↑V protease which as a mutation in the hinge region, was found to have a reduced catalytic ability compared to the WT-CSA protease (Zondagh *et al.*, 2019). The data suggested that the N37T↑V insertion mutation increases viral fitness and decreases protease inhibitor susceptibility (Zondagh *et al*., 2019). In another study, *in vitro* drug testing showed that mutant protease L38↑N↑L displayed a lowered sensitivity to Atazanavir, Darunavir and Lopinavir protease inhibitors (Williams *et al*., 2019). The L38↑N↑L protease displayed a reduction in k_{cat}, K_M and k_{cat}/K_M values as well as a reduction in the amount of protease found in the active conformation when compared to WT- CSA protease. Williams *et al.* (2019) and Zondagh *et al.* (2019) reported different amino acid insertions in the hinge region of the HIV-1 protease in ARV naïve patients which exhibit different substate catalysis kinetics and inhibitor binding thermodynamics. These alarming reports may suggest that acquired protease inhibitor resistance due to insertion mutations in the HIV-1 protease may become more common in the future.

1.9 L38↑H↑L protease

A major limiting factor in the HIV-1 epidemic has been the emergence of viral strains that exhibit resistance to protease inhibitors due to mutations in the protease sequence. The HIV-1 South African subtype C L38↑H↑L mutant protease, with no background mutations, is the first to be discovered in South Africa. The mutant protease was isolated from a drug-naïve infant (Ledwaba *et al*., 2019). The variant protease has a double insertion of Histidine and Leucine in the hinge region at position 38 of the protease amino acid sequence. As a result of the double insertion mutation, each subunit of the has 101 amino acids compared to 99 amino acids per monomer found in the WT-CSA protease The amino acid sequence alignment of the WT-CSA and the L38↑H↑L protease is found in Figure 6. Mutations at or near the hinge region tend to change the flexibility of the hinge region and the extent to which the flap region opens, thereby changing the shape of and access to the substrate-binding cleft. Consequently, reducing the ability of the protease inhibitor to bind to the substrate-binding cleft (Goldfarb *et al.*, 2015). Hinge region mutations may affect kinetic and thermodynamic binding properties of substrates and/or protease inhibitors. The loss of susceptibility to protease inhibitors usually occurs because the mutant protease molecules lower the binding affinity for the inhibitors yet maintain sufficient affinity for the Gag and Gag-Pol substrate molecules. There is a growing need in understanding how hinge region mutations affect the overall efficiency of protease inhibitors. This work will add to the growing knowledge of the impact of hinge region mutations on the structure and function of the WT-CSA protease.

A.

```
WT-CSA      PQITLWKRPL VSIKVGGQIK EALLDTGADD TVLEEINL-- PGKWKPKMIG GIGGFIKVRQ   58
L38↑H↑L     PQITLWKRPL VSIKVGGQIK EALLDTGADD TVLEEINLHL PGKWKPKMIG GIGGFIKVRQ   60
            ********** ********** ********** ********   ********** **********

WT-CSA      YDQILIEICG KKAIGTVLVG PTPVNIIGRN MLTQLGCTLNF   99
L38↑H↑L     YDQILIEICG KKAIGTVLVG PTPVNIIGRN MLTQLGCTLNF   101
            ********** ********** ********** ***********
```

B.

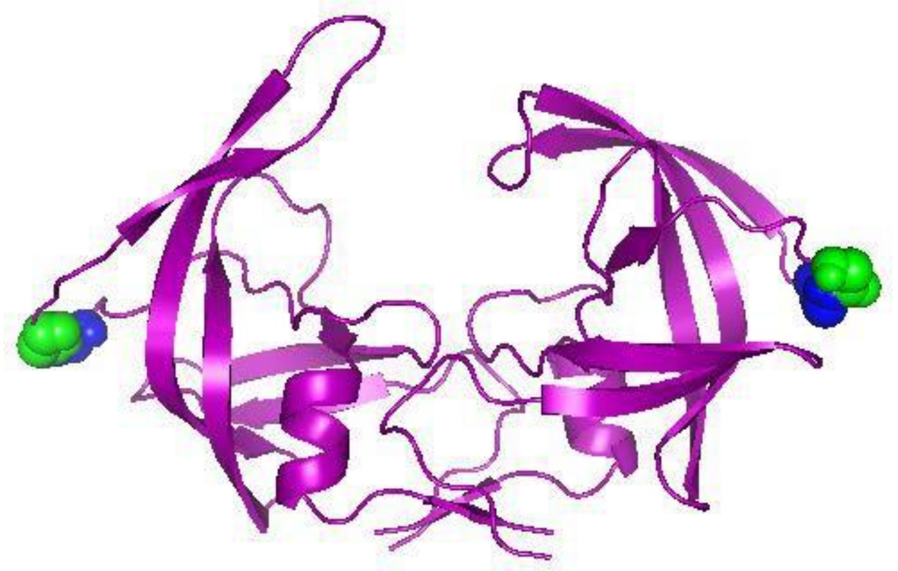


Figure 6. (A) **Amino acid sequence alignment**. The amino acid sequence alignment of L38↑H↑L variant and the wild-type protease. The highlighted box indicates the position of the Histidine and Leucine insertion mutation at position 38 of the protease amino acid sequence. The alignment was performed using Clustal Omega tool (EMBLEBI). **(B) Homology model of the L38↑H↑L protease**. The green (●) and blue (●) spheres represent the position of the histidine and leucine mutation in each subunit. The figure was generated with PyMol version 2.5.2.0 (DeLano, 2002).

1.10 Aim

The aim of this study is to assess the possible effects of the L38↑H↑L double insertion mutation on the structure and function of the HIV-1 South African subtype C protease as well as to assess the potency of Atazanavir, Darunavir and Saquinavir protease inhibitors towards the L38↑H↑L mutant protease.

1.11 Objectives

1.11.1 Recombinantly overexpress the WT-CSA and L38↑H↑L protease in *Escherichia coli* BL21 (DE3) pLysS cells

1.11.2 Purify, solubilise and refold the protease into its native functional form

1.11.3 Quantify the protease by using UV absorbance spectroscopy and active site isothermal titration calorimetry (ITC)

1.11.4 Assess and compare the secondary structure content of both proteases by using far-UV circular dichroism spectropolarimetry

1.11.5 Assess the oligomeric structure of the both proteases by using size exclusion high-performance liquid chromatography (SE-HPLC)

1.11.6 Perform comparative enzyme kinetic assays to determine the specific activity, K_M, k_{cat}, V_{max}, k_{cat}/K_M and IC_{50}

1.11.7 Assess thermodynamic parameters (ΔG, ΔH and ΔS) and K_d of protease-inhibitor binding interactions by using displacement isothermal titration calorimetry

2. Methods and Materials

2.1 Expression vector and expression host

pET-11a vectors were used for the recombinant expression of the WT-CSA and the L38↑H↑L mutant protease. The pET-11a vectors carried ampicillin resistance markers for selection. The pET-11a vectors with the WT-CSA and L38↑H↑L protease complementary deoxyribonucleic acid (cDNA) insert were synthesised by GeneScript (USA).

Escherichia coli BL21 (DE3) pLysS (Novagen®) competent cells were used as the expression host. This host cell was used because of its capability of expressing cytotoxic recombinant proteins. Transcription of transformed cDNA is induced by the T7 RNA polymerase which is under the control of isopropyl-β-D-thiogalactopyranoside (IPTG). pLysS encodes T7 lysozyme which minimises the uninduced expression level of genes under the control of the T7 promoter thus reducing cytotoxic effects by the HIV-1 protease (Studier and Moffatt, 1986). The expression host carries a chloramphenicol resistance marker for selection.

2.2 Transformation of *E. coli* BL21 (DE3) pLysS cells

The pET-11a vectors each encoding the cDNA sequence for the L38↑H↑L and the WT-CSA protease gene were transformed into *E. coli* BL21 (DE3) pLysS cells, respectively, by using the heat shock method. Initially, 10 µg of each of the pET-11a vectors were incubated with of *E. coli* BL21 (DE3) pLysS cells on ice for 30 minutes to ensure that the vector comes in close proximity to the competent cells. This ensures efficient transformation of the competent cells. Thereafter, the cells were heat-shocked at 45 ºC on a heating block for 45 seconds. The heat increases the permeability of the bacteria cell membrane which allows for the vectors to enter the cells and thereafter chilled on ice for 30 minutes. The cells were then incubated in a shaking incubator for 1 hour at 37 ºC, 250 rpm for the WT-CSA and L38↑H↑L protease. The transformed cells were grown overnight on nutrient agar plates (15% (w/v) agar), 1% (w/v) tryptone, 0.5% (w/v) yeast extract (0.5% (w/v) and 0.5% (w/v) sodium chloride) supplemented with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol antibiotics. A single colony from each of the agar plates was selected and used to inoculate 100 mL LB media (1% (w/v) tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) sodium chloride 0.5% supplemented with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol to allow for the growth of successfully transformed cells. The cells were grown overnight at 37 ºC, 200 rpm. The overnight cell culture was used to make glycerol stocks of transformed cells. The glycerol stocks were stored at -80 ºC.

To confirm the successful transformation of the cells, a GeneJET plasmid miniprep kit (Thermo Scientific™, Massachusetts, USA) was used to extract and isolate the vectors from the transformed cells. The vector isolates were sent to Inqaba Biotech® (Pretoria, South Africa) for Sanger sequencing as well as for WT-CSA and L38↑H↑L protease cDNA insert validation. The sequence results were received as a DNA sequence. The DNA sequence was then transcribed and translated into the protein sequence by using the ExPASy tool (https://web.expasy.org). The WT-CSA and L38↑H↑L protein sequence was aligned using the Basic Local Alignment Search Tool (Blastx) (Altschul *et al.*,1997). Upon successful validation of the WT-CSA and L38↑H↑L cDNA sequence, expression trials of the L38↑H↑L protease were performed.

2.3 Growth characteristics

The high yield of recombinant WT-CSA and L38↑H↑L protease production is highly desirable, but due to its ability to be cytotoxic to bacterial cells, achieving a desirable amount of expressed protein is often hindered. Expression trials of the L38↑H↑L protease were performed to determine the optimum conditions required for overexpression of the L38↑H↑Lprotease gene. A growth curve of bacterial cells transformed with the pET-11a vector containing the L38↑H↑L cDNA insert was constructed to determine the OD_{600nm} required for the cells to reach the mid-log phase. The transformed L38↑H↑L protease glycerol stocks were used to inoculate 100 mL of LB media. The cells were grown overnight at 37 ºC, 200 rpm. A 1:20 dilution of the bacterial culture was done into 50 mL LB media. The cells were grown for 16 hours while taking an OD_{600nm} measurement every 5 minutes for the first 30 minutes and thereafter every hour. The OD_{600nm} readings were plotted as a function of time in minutes. The growth curve was used to determine the time it took the cells to reach the mid-log phase. The mid-log phase of a bacterial growth curve is indicative of the exponential growth of the cells, there are therefore more cells available. This is particularly beneficial since we need high recombinant protein overexpression.

2.4 Expression trials

The L38↑H↑L protease glycerol stocks were used to inoculate 100 mL LB media. The cells were allowed to grow overnight at 37 ºC, 200 rpm. A 1:20 dilution of the overnight culture into 5 x100 mL LB media was performed. The cells were grown at 37 ºC, 200 rpm until the mid-log phase (OD_{600nm} of 0.6) was achieved. The IPTG concentrations required for the overexpression of the mutant protease were determined by using 0, 0.2, 0.5, 0.8, and 1.0 mM IPTG. The amount of cells were normalized using the rule of thumb *for E.coli* cells that states that 1 unit of OD_{600nm} = 8 x 10^8 cells/mL (Volkmer and Heinemann, 2011, Couto *et al.*, 2018, Elbing and Brent, 2020). The same concentration of cells were taken at various time points (0, 2, 5, 8, and 16 hours) after IPTG induction. The cells were harvested by centrifugation at 5000 xg for 3 minutes and the supernatant was discarded. The cell pellet was resuspended in TE buffer (10 mM Tris-HCl, 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0) and sonicated by using the Microson XL-2000® sonicator for 15 seconds at 50% power. The cells were further centrifuged at 5000 xg for 5 minutes to isolate the soluble and insoluble fractions of the cell lysate. Thereafter, the samples were resuspended in reducing sample buffer (12% (w/v) SDS, 30% (v/v) glycerol, 0.05% (w/v) Coomassie Blue G-250, 10% (v/v) β-mercaptoethanol and 150 mM Tris-HCl, pH 7.0) in a 2:1 ratio.

2.5 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), a technique developed by Laemmli (1970), was used to assess the homogeneity of the protease molecules and to determine their monomeric molecular size. Sodium dodecyl is an anionic detergent that binds proteins in proportion to their mass and confers a negative charge onto the denatured polyproteins. A tricine SDS-PAGE (Schagger and von Jagow, 1987), which provides a better resolution of separating proteins lower than 30 kDa, was used to assess the size and homogeneity of the cell lysate samples. The difference between the Laemmli method and the tricine SDS-PAGE method is the trailing ions used, in particular, their pK_a values. Glycine which is used in the Laemmli method has a pK_a value of 9.6 whereas the tricine used in the tricine SDS-PAGE method has a pK_a value of 8.15. The pK_a value affects the mobility of the trailing ions with respect to the mobility of proteins in the gel. Tricine moves faster relative to the proteins in the stacking gel and decreases the stacking limit of the gel. This allows for a more effective separation of proteins of low molecular mass (Schagger and von Jagow, 1987).

Expression of the target protein was analysed on a tricine SDS-PAGE gel containing 4% (w/v) stacking and 16% (w/v) separating gel. The soluble and insoluble fractions were boiled for five minutes prior to loading onto the gel. The gels were subjected to electrophoresis by using a cathode buffer (0.1 M Tris-HCl, 0.017% (w/v) Tricine, 0.001% (w/v) SDS, the pH was not altered) and anode buffer (0.1 M Tris-HCl pH 8.9). The electrophoresis of the gels was performed at 140 V for 1 hour using the Bio-Rad system. The PAGE Ruler Pre-stained Protein Ladder (Thermo Scientific®) was used as the molecular weight marker. The gels were stained using a solution containing 50% (v/v) methanol, 10% (v/v) acetic acid and 0,025% (w/v) Coomassie Blue G-250. Thereafter, the gels were destained in a solution containing 50% (v/v) methanol and 10% (v/v) acetic acid.

The relative migration distance (R_f) of the molecular weight marker and the protein samples from the induction trials were used to validate the overexpression of the gene of interest. The R_f value was calculated as the distance migrated by a polyprotein band over the distance migrated by the dye front. A calibration curve of log molecular weight versus R_f was used to determine the size of the protease monomer.

Upon analysis of the gels, the optimum conditions required for protein overexpression were determined. Expression trials were only performed for the L38↑H↑L mutant protease. Conditions for the overexpression of the WT-CSA were taken from Naicker *et al*., 2013.

2.6 Overexpression and purification

For the overexpression of the WT-CSA and L38↑H↑L protease, Erlenmeyer flasks containing 200 mL LB media supplemented with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol were inoculated with the respective glycerol stocks. The cells were grown overnight at 37 ºC, 200 rpm. The overnight bacterial cultures were used to inoculate 14 Duran® Erlenmeyer flasks (Sigma Aldrich, Missouri, USA) containing 500 mL LB media in a 1:20 ratio. The cells were allowed to grow at 37 ºC, 200 rpm until OD_{600nm} of 0.6 was achieved. Overexpression of the WT-CSA and the L38↑H↑L protease gene was induced by the addition of 1 mM (WT-CSA) and 0.6 mM (L38↑H↑L) IPTG, respectively. The cells were grown for a further 4 hours (WT-CSA) and 4 hours (L38↑H↑L) after IPTG induction at 37 ºC, 200 rpm. The bacterial cells were harvested by centrifugation at 5000 xg for 10 minutes at 4 ºC. The resulting cell pellets were resuspended in buffer A (10 mM Tris-HCl, 2 mM (EDTA), and 1 mM phenylmethylsulfonyl fluoride (PMSF), pH 8.0) and stored at -80 ºC.

The cells were thawed at room temperature to aid in cell lysis. The thawed cells were resuspended in an additional 20 mL of buffer A with 1 mg/mL bovine pancreatic deoxyribonuclease I (DNase) dissolved in 20 mM Tris-HCl, pH 7.5, and 2 M Magnesium Chloride ($MgCl_2$). The cells were homogenised by using a Bio-Gen series PRO200 homoginiser (Pro Scientific Inc., Connecticut, USA) for one minute. The viscosity of the cell suspension was left to decrease for 30 minutes at room temperature. Further cell lysis with eight cycles of Sonication with 30 seconds intervals at 70% power were performed by using the Microson XL-2000® sonicator. The bacterial cell suspension was centrifuged at 18 000 xg to separate the soluble and insoluble portions of the cell lysate. The protease was expressed in the insoluble fraction (inclusion bodies) and was washed by resuspending the pellet in ice-cold Buffer A^+ (10 mM Tris-HCl, 2 mM EDTA, 2% Triton X-100 and 1 mM PMSF, pH 8.0) and centrifuged at 18000 xg for 30 minutes at 4 °C. The pellet wash step with buffer A^+ was performed twice. The solubilisation of the inclusion bodies was performed by resuspending the pellet in Buffer B (10 mM Tris-HCl, 8 M urea, 5 mM Dithiothreitol (DTT), pH 8.0) for 1 hour at room temperature. The unfolded protein solution was filtered using a 0.45 µM cellulose acetate syringe filter and loaded onto a 5 mL HiTrap FF DEAE sepharose column attached to a 5 mL HiTrap FF CM column (Sigma Aldrich, Missouri, USA). The columns were pre-equilibrated with five column volumes of Buffer B at a flow rate of 3 mL/min, 0.3 kPa. The purification experiment was performed on the ÄKTA Start™ liquid chromatography purification system (GE Healthcare Life Sciences, Chicago, USA). The difference between the theoretical isoelectric point of 9.32 for both the WT-CSA and L38↑H↑L as predicted by the ExPasy ProtPARAM tool (Gasteiger *et al.*, 2005) and the pH of the buffer enables the positively charged protease molecules to bind to the negatively charged resin of the HiTrap FF CM column. Non-specific, negatively charged proteins bound to the HiTrap FF DEAE sepharose column. The unfolded positively charged proteases (WT-CSA and L38↑H↑L) bound the CM cation exchange column and were eluted with Buffer B containing 1 M sodium chloride. The protein samples were refolded and dialysed using SnakeSkin™ dialysis tubing with a MWCO of 3.5 kDa (Thermo Fischer Scientific™, Massachusetts, USA) against Buffer C (10 mM formic acid, 10% glycerol, and 0.02% sodium azide) for 4 hours at 4 ºC. The refolded protein samples were further dialysed into Buffer D^+ (10 mM sodium acetate, 2 mM DTT, 0.02% sodium azide, pH 5.0) overnight at 4 ºC.

Thereafter, the dialysed protein samples were centrifuged at 18000 xg, 4 ºC to get rid of aggregated proteins and debris. The resulting supernatant containing the protein of interest was further purified by using a HiTrap FF CM column connected to an ÄKTA Start™ liquid chromatography purification system (GE Healthcare Life Sciences, Chicago, USA) to remove any non-specific protein contaminants. The CM column was equilibrated using five column volumes of buffer D (10 mM sodium acetate, 0.02% sodium azide, pH 5.0). The positively charged protease molecules bound to the column were eluted using a 0-1 M sodium chloride gradient (300 mL at a flow rate of 3 mL/min flow rate). The eluted proteins were dialysed using SnakeSkin™ MWCO 3.5 kDa dialysis tubing (Thermo Fischer Scientific™, Massachusetts, USA) against buffer D (10 mM sodium acetate, 0.02% sodium azide, pH 5.0) overnight at 4 ºC. The collected protein samples were concentrated using a 50 mL Amicon® stirred concentrator MWCO 3.5 kDa (MilliporeSigma, Massachusetts, USA) to 0.4–0.5 mg/mL and stored at -80ºC until required.

The collected protein samples were evaluated for purity by using tricine SDS-PAGE as described in section 2.5. A calibration curve of log molecular weight versus R_f was used to determine the size of the protease monomer as well as to assess the purity.

2.7 Quantification of protein concentration

2.7.1 Absorbance spectroscopy

The concentration of the WT-CSA and L38↑H↑L protease samples was determined spectrophotometrically by taking absorbance readings of the samples at 280 nm and 340 nm using a Jasco V-630 UV-Vis Spectrophotometer (Jasco, Pfungstadt, Germany). A serial dilution consisting of 1:10, 1:12, 1:14, 1:16, 1:18, and 1:20 protein: buffer samples were used to determine the protein concentration. Aromatic amino acid residues such as Tyrosine and Tryptophan absorb maximally at 280 nm (Lakowicz, 1983). Light scatter from protein aggregates and dust particles were corrected by subtracting the 340 nm absorbance reading from the absorbance reading at 280 nm. The corrected absorbance readings were plotted as a function of dilution fractions. The plots were fit with a linear regression function and the resulting straight-line equation coupled with the Beer-Lambert Law was used to determine the protein concentration. The Beer-Lambert law is as follows:

$$A_\lambda = \varepsilon_\lambda c l \quad [1]$$

where A is the corrected absorbance at 280 nm, ε_λ is the molecular extinction coefficient of the protein at 280 nm, c is the molar concentration and l is the path length (one centimeter) of the light through the medium in a quartz cuvette. The theoretical molecular extinction coefficient of both the dimeric WT-CSA and the L38↑H↑L protease is 25 230 $M^{-1}cm^{-1}$, (ExPASy ProtPARAM (Gasteiger *et al*., 2005)). The concentration of purified protein samples were determined for every experiment.

Due to the autocatalytic nature of the HIV-1 protease as well as the isolation, unfolding and refolding methods involved in the purification process, absorbance spectroscopic techniques are deemed insufficient. Absorbance spectroscopy measurements do not distinguish between fully folded and functional as well as cleaved proteins.

2.7.2 Active site isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a powerful and sensitive technique used to characterise the binding thermodynamics between biomolecules such as peptides, nucleic acids, antibodies, and proteins. An ITC machine consists of a sample cell and reference cell located in an adiabatic jacket (Núñez *et al*., 2012). A titrant, such as a ligand, is injected into the sample cell containing the protein solution using a spinning syringe. Titrations of the ligand into the sample cell are performed until saturation of the protein binding sites has occurred. Saturation results in fewer interactions between the titrant and the protein and less heat transfer occurs. The enthalpic change resulting from heat absorption (endothermic) or release (exothermic) in proportion to the amount of binding occurring is measured by a microcalorimeter located inside the machine (Sheehan, 2009). Measuring heat transfer during the binding process enables the determination of binding affinity (K_a), reaction stoichiometry (N), and enthalpy (ΔH) (Holdgate and Ward, 2005), From these parameters, Gibbs free energy (ΔG) and entropy (ΔS) can be calculated using the equation:

$$\Delta G = -\mathrm{RT}\ln(K_\mathrm{a}) = \Delta H - \mathrm{T}\Delta S \qquad [2]$$

where R is the gas constant (8.314 $J.mol^{-1}K^{-1}$) and T is the absolute temperature in kelvin.

Since the protein concentration determined from performing serial dilutions does not distinguish between native and cleaved protease molecules, the N value was used to determine the percentage concentration of functional protease in the sample. It is essential to determine the proportion of protease molecules that have refolded correctly and have formed disulfide bonds between the protease monomers. It is essential to perform active site titrations since the HIV-1 protease is functional in its homodimeric form (Naicker *et al*., 2013).

Active site titration experiments were performed using the Nano ITC standard volume instrument (TA® instruments, New Castle, USA). Acetyl-pepstatin (Merck (Pty) Ltd, South Africa), a well-documented aspartyl protease inhibitor was used as a ligand in this experiment (Seelmeier *et al*., 1988). The acetyl-pepstatin stock solutions of 8 mM were prepared by dissolving 5.15 mg/mL of the lyophilised powder in 10 mM sodium hydroxide. Prior to use, buffer D was used to dilute the stock solutions to 200 µM acetyl-pepstatin. The ligand used in performing ITC experiments should be dissolved in the same buffer as the protein sample in order to avoid buffer mismatch. The diluted ligand was loaded into a 250 µL ITC syringe. Excess volumes of the ligand are required to ensure the syringe is filled to maximal capacity while ensuring that there is no bubble formation. Bubbles produce high heat signals that mask the heat signals produced by the protein-ligand binding reaction. Both the ligand and the protein sample were degassed under a volume of 0.5 atm for 10 minutes at 20 ºC. An initial injection of 2 µL, followed by 29 subsequent 5 µL of 200 µM acetyl-pepstatin, were injected into the sample cell containing 1.5 mL of 20 µM protease while stirring at 310 rpm. The reference cell contained milli-Q® water. Measurements from performing the active titrations were plotted as the power (µcal/sec) needed to maintain the sample and reference cell at an equal temperature of 20 °C (Wiseman *et al*., 1989). Acetyl-pepstatin binds to the protease in a 1:1 ratio. A stoichiometry value of 1 indicates that the concentration of protease molecules is equal to that of the inhibitor and that 100% of the protease molecules in the sample are in their fully folded and functional form and no autolysis has occurred. The integrated heat per injection data was fitted and analysed with the NanoAnalyze™ software (TA Instruments) using the independent model. The heat of dilution (HOD) of the ligand into buffer was subtracted from the data and an independent model was used to fit the data. The HOD is often obtained by titrating the ligand into the buffer. Active site ITC experimental conditions were used to perform HOD titrations.

2.8 Structural characterisation

2.8.1 Far-UV circular dichroism to estimate secondary structure content

Far-UV circular dichroism (CD) spectropolarimetry is a technique used to measure the secondary structure content of optically active molecules such as proteins and DNA. Circular dichroism in the far-UV region (240 nm-180 nm) is observed when optically active molecules absorb left- and right-handed circularly polarised light. The CD spectrum results from the interaction of optically active molecules and an optically asymmetrical environment.

Optically active parts of a protein include the peptide backbone, aromatic amino acids, and disulfide bonds (Adler *et al.,* 1973). The CD spectrum of a protein is influenced by phi (Φ) and psi (Ψ) angles of the peptide backbone (Kelly *et al*., 2005). The peptide backbone angles reflect the formation of secondary structure elements such as α-helices, β-sheets, and random coils. These secondary structure elements absorb circularly polarised light differently, resulting in different shapes and magnitudes of the CD spectrum (Whitmore and Wallace, 2007). Predominantly α-helical proteins have troughs at 222 nm and 208 nm and a peak at 193 nm. Whereas predominantly β-sheeted proteins have a trough at 218 nm. However, far- UV CD cannot indicate where the detected secondary structure elements of a protein are located within the protein or the secondary structure of specific amino acid residues in contrast to X-ray crystallographic and NMR structural determinations. The use of the far-UV CD to probe proteins' secondary structural content can allow for the analyses of the structural changes that may be brought about by mutations in the protein structure. We can, therefore, determine whether the L38↑H↑L mutation causes apparent changes in the secondary structural content of the L38↑H↑L protease relative to the WT-CSA. According to the crystal structure of the WT-CSA protease (PDB ID: 3U71) (Naicker *et al.,* 2013), the protein is predominantly β-sheeted.

Far-UV CD spectra of the homodimeric WT-CSA and L38↑H↑L protease were measured by using a Jasco J-810 spectropolarimeter connected to a Peltier temperature controller (Jasco, Pfungstadt, Germany). The ellipticity (θ) was blanked with Buffer D prior to taking protein measurements. Protein concentrations of 5 µM for the WT-CSA and L38↑H↑L were used to perform the experiments. The ellipticity of the protease samples was measured over a wavelength range of 210-250 nm at 20 ºC using a data pitch of 0.2 nm and a bandwidth of 0.5 nm. The spectra were an average of eight accumulations. A quartz cuvette with a path length of 2 mm was used to perform the experiments. The experiments were performed in triplicates. The ellipticity of the WT-CSA and L38↑H↑L protein samples was converted to the mean residue ellipticity (MRE) using the equation:

$$\mathrm{MRE} = \frac{\theta \text{ x } 100}{\mathrm{cn}l} \qquad [3]$$

where θ is the CD signal (mdeg), c is the protein concentration (mM), n is the number of amino acids in the protein (198 residues for the dimeric WT-CSA and 202 residues for the dimeric mutant protease) and *l* is the pathlength of the quartz cuvette (cm) (Woody,1995).

2.8.2 SE-HPLC to confirm quaternary structure and size

Size exclusion-high performance liquid chromatography (SE-HPLC) is a technique used to separate molecules according to their molecular size in a non-denaturing manner. This technique was first described by Ruthven *et al.,* 1954. The molecules are separated by porous beads of a particular size. Molecules with smaller hydrodynamic volumes are retained much longer in the porous beads compared to larger molecules with large hydrodynamic volumes. This allows for the larger molecules to elute first (Barth *et al.,* 1994). The use of liquid under pressure with a constant flow rate aids with the separation of the molecules. This technique can also be used to assess protein purity. A contaminated protein sample results in non- specific absorbance peaks on the chromatogram. According to the crystal structure obtained by Naicker *et al*., 2013 (PDB ID: 3U71), the WT-CSA is a homodimer, consisting of two symmetrical subunits. SE-HPLC was used to determine the effects that may be caused by the L38↑H↑L mutation on the quaternary structure of the protease and to assess whether the quaternary size is maintained.

Approximation of the molecular size and purity was determined by using a TSKgel® SuperSW2000 30 cm x 4.6 mm, 4 µm particle size column (Tosoh Corporation©, Tokyo, Japan). The column was connected to a TSKgel® Size exclusion (SW-Type) HPLC guard column (Tosoh Corporation©, Tokyo, Japan). The experiments were performed on a Shimadzu ultrafiltration liquid chromatography (UFLC) SPD-20A SE-HPLC machine (Tokyo, Japan). The WT-CSA and L38↑H↑L protease samples were concentrated to 40 µM and 60 µM, respectively, prior to use.

The column was equilibrated with degassed and filtered buffer D containing 200 mM NaCl at a flow rate of 0.3 mL/min for the WT-CSA and 0.2 mL/min for the L38↑H↑L mutant protease. After sufficient column equilibration (five column volumes), 20 µL of 40 µM (WT- CSA) and 60 µM (L38↑H↑L) the protein samples were injected onto the column. A chromatogram showing absorbance readings at 280 as a function of time (minutes) required for the protein sample to pass through the column was generated. The column was allowed to re-equilibrate for 30 minutes after subsequent runs. Thereafter, 20 µL of a protein standard mixture (69385 Protein standard mix, Sigma-Aldrich, USA) consisting of thyroglobulin bovine, γ-globulin, albumin, ribonuclease A and Para-aminobenzoic acid was injected into the column. The standard mixture was allowed to pass through the column at a flow rate of 0.2 mL/min and 0.3 mL/min in order to match the conditions used to run the WT-CSA and L38↑H↑L protease samples. A standard curve of known molecular weight (Log molecular weight) as a function retention time (min) was plotted to estimate the molecular size of the WT-CSA and L38↑H↑L protease. A linear regression function was used to fit the standard curve and the subsequent straight-line equation was used to calculate the molecular size of the proteases based on their retention times.

2.9 Functional characterisation

2.9.1 Enzyme kinetics

Enzyme kinetics is the study of the kinetics and reaction rates involved in enzyme catalysed reactions. Kinetic studies provide a better understanding of how enzymes work. Some enzyme assay conditions such as protein/substrate/ligand concentrations, pH, buffers, and temperatures need to be considered when assessing enzyme kinetics. Conditions that favour an enzyme's optimal activity are the most desirable but often physiological conditions under which the enzyme works cannot be replicated *in vitro*.

The kinetics involved in the understanding of enzymes have traditionally been modelled by following the basic principles proposed by Briggs and Haldane (1952). According to Briggs and Haldane (1952), at saturating conditions of the substrate, the concentration of the enzyme-substrate complex remains relatively the same. Increasing the amount of substrate will not increase the rate of the reaction. Consider the following reaction:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E.S \xrightarrow{k_2} E + P \qquad [4]$$

where E, S, E.S, and P represent the free enzyme, substrate, enzyme-substrate complex, and product, respectively, and k_1, k_{-1}, and k_2 are reaction rate constants. The reaction rate constant k_1 is the association rate constant for the binding reaction of the free enzyme to the substrate to form the enzyme-substrate complex, k_{-1} is the dissociation rate constant of the enzyme-substrate complex into free enzyme and substrate, k_2 is the rate constant for the formation of product and free enzyme from the enzyme-substrate constant. The reaction achieves a steady state when the concentration of the enzyme-substrate complex remains constant over time. The formation of a product from the enzyme-substrate complex is the rate-limiting step. The rate at which the enzyme-substrate complex is formed is dependent on the amount of substrate in the sample since the concentration of the enzyme in the sample remains constant.

If the concentration of the enzyme remains constant with increasing increments of the substrate, a hyperbolic Michaelis-Menten plot can be generated. A conventional approach used to understand steady state enzyme kinetics is based on the Michaelis-Menten equation. The Michaelis-Menten plot follows the equation:

$$\mathrm{V}_0 = \frac{V_{\mathrm{max}}\,[S]}{K_{\mathrm{M}}+[S]} \qquad [5]$$

where V_0 is the initial velocity, V_{max} is the maximum velocity under saturating levels of the substrate and K_{M} is the Michaelis-Menten constant. Product inhibition is assumed to be negligible (Segel, 1988). The Michaelis-Menten constant is the substrate concentration required to reach half-maximal velocity. It also indicates the affinity the enzyme has for the substrate. The lower the K_{M} value, the higher the affinity. The maximal velocity is the reaction rate achieved when all the enzyme active sites are saturated with the substrate.

Under the assumption that E + S and E.S are in equilibrium and that the enzyme-catalysed reaction occurs under steady-state conditions, the Michaelis-Menten constant (Michaelis and Menten, 1913) can be written as follows:

$$K_M = \frac{(k_{-1} + k_2)}{k_1} \qquad [6]$$

Enzyme kinetics coupled with fluorometric assays provided a powerful tool that can be used to study how enzymes function.

2.9.2 Fluorometric enzyme assays

Fluorometric assays are the most widely used technique for assessing enzyme kinetics. This technique is based on the ability of an enzyme to convert a substrate into a product that absorbs and releases light at specific wavelength/s. The fluorescence phenomenon is observed when a fluorophore absorbs light at a specified wavelength from a source. This then results in the excitation of an electron in the fluorophore from the singlet ground state to an excited state of a higher energy level. After the electron spends a fluorescence lifetime in the excited state, the electron falls back down to the lowest available energy state of a longer wavelength, releasing its energy as an emitted photon (Kogure *et al*., 2008). The fluorescence lifetime is the amount of time an electron spends in the excited state. The electron in the excited state undergoes energy loss through thermal or vibrational energy. The energy level shift between the excited and emission energy levels is known as the Stokes shift (Abbyad *et al*., 2007). The photons released by a fluorophore can be detected and measured by using a fluorometer. Analytical measurements provided by a fluorometer are used to detect changes in fluorescence intensity which allows for the determination of the activity of the enzyme as well as other enzyme kinetics parameters. Enzyme kinetic studies coupled with fluorometric assays were performed to determine the possible effects the L38↑H↑L double insertion mutation had on the activity of the enzyme relative to the WT-CSA.

Enzyme kinetic studies of the WT-CSA and L38↑H↑L protease were determined by following the hydrolysis of a fluorogenic substrate, Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala- Nle-NH_2, that mimics the capsid/p2 cleavage site in the Gag polyprotein (Velazquez-Campoy *et al*., 2001). The substrate is recognised and hydrolysed between Nle (norleucine) and Phe(NO_2) (nitro-phenylalanine) by the protease (Carmel and Yaron, 1978). The close proximity of the fluorogenic 2-aminobenzoyl (Abz) and the quenching Phe(NO_2) amino acid group reduces the ability of the substrate to fluoresce. The cleavage of the Nle-Phe(NO_2) peptide bond diminishes the quenching effect and results in an increase in fluorescence (Carmel and Yaron, 1978; Polgar *et al*., 1994). The time-dependent cleavage of the Nle- Phe(NO_2) peptide bond results in an increase in fluorescence. For all enzyme kinetics assays performed, the 2-aminobenzoyl group was excited at a wavelength of 337 nm and emission was measured at a wavelength of 425 nm. Stock solutions of 500 µM substrate were prepared by dissolving the lyophilised powder in 50% acetonitrile. The reaction mixture (total volume of 800 µL) consisted of protease, substrate, milli-Q® and a reaction buffer (2 M NaCl, 100 mM sodium acetate, pH 5.0). Glacial acetic acid was used to pH the reaction buffer. The fluorescence intensity measurements for the kinetic assays were measured for 30 seconds using a data pitch of 1 sec, a response time of 0.5 seconds, and an excitation bandwidth and emission bandwidth of 2.5 nm and 5 nm, respectively. The intensity value due to the complete cleavage of 1 nmol of the substrate was measured and used to convert the fluorescence intensity to activity. The experiments were performed in triplicate at 20 ºC using a Jasco V-630 spectrophotometer. All kinetic assay data were plotted and analysed on GraphPad Prism version 9 (GraphPad Software, San Diego, California, USA)

2.9.3 Michaelis-Menten enzyme kinetics

Enzyme kinetic parameters such as the K_M and V_{max} of the WT-CSA and L38↑H↑L protease were determined by following the hydrolysis of the fluorogenic substrate. The K_M and V_{max} kinetic parameters were determined under constant enzyme concentration of 50 nM and varying concentrations of substrate (0- 200 µM). A hyperbolic curve of enzyme velocity as a function substrate concentration was plotted to determine the kinetics parameters.

2.9.4 Specific activity and catalytic turnover number

In order to measure the specific activity and catalytic turnover number of the WT-CSA and L38↑H↑L mutant protease, substrate concentrations greater than the K_M should be used. Specific activity enzyme assays were performed with increasing protease concentrations of 0-50 nM while the substrate concentration was kept constant at 50 µM. A linear regression function was used to fit the plot of enzyme activity (µmol/min) as a function of enzyme amount (mg). The slope from the linear regression plot was used to determine the specific activity of the enzymes.

The catalytic turnover number, a first-order rate constant, is a measure of how fast the product is formed from the enzyme-substrate complex. The catalytic turnover number is represented as k_2 in equation 4 (section 2.9.1). It represents the number of substrate molecules transformed into a product using a single catalytic site of the enzyme per unit time. The k_{cat} number was determined from the specific activity assays. A linear regression function was used to fit the plot of enzyme activity (µmol/sec) per µmol of the enzyme. The slope of the straight-line generated was used to determine the catalytic turnover number (sec^{-1}).

2.9.5 Catalytic efficiency

Catalytic efficiency (k_{cat}/K_M), a second-order rate constant, was used as a measure to compare the effectiveness of proteins in catalysing the same reaction (Eisenthal *et al.*, 2007). Catalytic efficiency can only be measured using substrate concentrations much lower than the K_M values obtained. Constant enzyme concentration of 50 nM was used while the substrate concentration ranged from 0-10 µM. A linear regression function was used to fit the plot of k_{cat} as a function of substrate concentration (nM). The resulting slope of the straight-line equation generated was used to determine the catalytic efficiency (µM^{-1}.sec^{-1}).

2.9.6 IC50 determination

The concentration of an inhibitor that is required to reduce the catalytic activity of the protease by 50% is termed the half-maximal inhibitory concentration (IC_{50}). IC_{50} values are often used to compare the relative effectiveness of inhibitors. The values of IC_{50} are obtained at varying ranges of inhibitor concentrations while keeping the concentration of the substrate constant.

The competitive inhibition between the fluorogenic substrate used and the protease inhibitor can allow for the calculation of IC_{50} (Netsu *et al.*, 2001; Ohtaka *et al.*, 2002; Naicker *et al.*, 2013). The IC_{50} values for competitive inhibition can be calculated from the Cheng-Prusoff (1973) equation:

$$IC_{50} = K_i \left(1 + \frac{[S]}{K_M} \right) \qquad [7]$$

where K_i is the inhibition constant, [S] is the substrate concentration and K_M is the Michaelis-Menten constant. The IC_{50} values were determined from dose-dependent plots of percentage enzyme activity as a function of inhibitor concentration. Percentage enzyme activity can be calculated using the following formula:

$$\text{Activity} = \frac{V \text{ of inhibited reaction}}{V \text{ of uninhibited reaction}} \qquad [8]$$

where V is the rate of the protease catalysed reaction.

Lower IC_{50} values obtained from different inhibitors can suggest tighter inhibition. For example, in an inhibition assay, inhibitor A with an IC_{50} value of 100 nM requires a thousand times less inhibitor to reduce the enzyme activity by half than inhibitor B with an IC50 value of 100 µM.

Standard inhibition assays containing a reaction mixture of 50 nM active protease and 50 µM fluorogenic substrate in the presence of increasing concentrations of protease inhibitor (0-200 nM) were performed. Inhibitor stocks of 10 mM of each protease inhibitor were prepared by dissolving the inhibitor lyophilised powder in 100% DMSO. Initially, 80 µL of the inhibitor stocks (10 mM) were added to 20 µL protease (2 µM stock) and 220 µL of 72.73 mM sodium acetate buffer, pH 5.0. This reaction mixture (reaction mixture 1) was incubated at 4 ºC for 15 hours to allow for sufficient time for equilibrium to be maintained. A second reaction mixture (reaction mixture 2) consisting of 80 µL fluorogenic substrate (500 µM stock) and 400 µL of 60 mM sodium acetate, 2 M sodium chloride, 2.4% DMSO, pH 5.0 was made. Glacial acetic acid was used to pH both buffers. Initially, 320 µL of reaction mixture 1 was added to 480 µL reaction mixture 2 to initiate the reaction. Thereafter, 400 µL of this mixture was dispensed into the cuvette. The reaction was measured after 30 seconds for 30 seconds. The fluorescence intensity was measured by using a data pitch of 1 sec, a response time of 0.5 seconds, and an excitation bandwidth and emission bandwidth of 2.5 nm and 5 nm, respectively. The intensity value due to the complete cleavage of 1 nmol of the substrate was measured and used to convert the fluorescence intensity to activity. The experiments were performed in triplicates using a Jasco V-630 spectrophotometer.

2. 10 Displacement isothermal titration calorimetry

In a displacement ITC experiment, also known as a competitive binding assay, the dissociation constant and protein-ligand binding thermodynamics of high-affinity inhibitors can be measured by displacing a prebound weaker competitive inhibitor, such as acetyl- pepstatin (K_a = $10^6 – 10^7$ M^{-1}), using an inhibitor with a higher binding affinity ($K_a > 10^8$ M^{-1}) (Velazquez-Campoy and Freire, 2006). Displacement ITC has successfully been used to determine the thermodynamic parameters of HIV-1 protease inhibitors and can provide information on how mutated proteases either lower or increase their susceptibility to inhibitors. Measuring the thermodynamic parameters such as a change in Gibbs free energy (ΔG), dissociation constant (K_d), change in enthalpy (ΔH), and change in entropy (ΔS) can demonstrate valuable information. A spontaneous binding process must have a negative ΔG. The ΔG should become increasingly negative as the binding of the inhibitor to the protease becomes tighter (Velazquez-Campoy and Freire, 2018). The K_d denotes the binding affinity of the protein for the ligand, the lower the K_d the tighter the binding. The ΔH which is the objective of the first law of thermodynamic, the law of energy conservation denotes the total internal energy or heat of a biologic system and the product of the volume and the constant pressure of the system. If the ΔH is positive, the binding reaction is endothermic in nature whereas a negative entropy denotes an exothermic reaction. The ΔS is the objective of the second law of thermodynamics. It denotes the change in the state of disorder of a system.

Primary stocks of Saquinavir, Darunavir, and Atazanavir, which are FDA-approved tight binding HIV-1 protease inhibitors, were dissolved in 100% DMSO to a final concentration and volume of 10 mM and 500 µL, respectively. During the displacement titration experiment, 2 mL of 20 µM WT-CSA and L38↑H↑L protease (corrected concentration from section 2.7.2) was preincubated with 200 µM acetyl-pepstatin in 2% DMSO for 30 minutes at room temperature Thereafter, the protease-inhibitor mixture (i.e., protease-acetyl-pepstatin) was loaded into the sample cell. In a separate experiment, the tight-binding protease inhibitors were used to displace acetyl-pepstatin from both the WT-CSA and L38↑H↑L protease active sites. The secondary protease inhibitor stocks of 200 µM in 2% DMSO were made by diluting the primary inhibitor stocks in Buffer D (same buffer in which the protein is dissolved in). A 250 µL ITC syringe was used to inject the inhibitors into the sample cell. It is essential that both the sample cell and the syringe contents are dissolved in 2% DMSO to avoid buffer mismatch. The integrated heat per injection data were analysed by using the NanoAnalyze™ software (TA instruments). The HOD of the ligand into buffer was subtracted from the data and an independent model was used to fit the data.

3. Results and Discussion

The emergence of mutations/natural polymorphisms and circular recombinant forms between the different sub-groups and sub-types drives the need to understand how mutations affect the protease overall structure and function. Protease inhibitors form the main component of second-line ARV treatment in South Africa. While the effects of primary mutations, which are directly involved in the binding of substrates and protease inhibitors, are constantly investigated, the mutations that occur in the hinge region of the protease are rarely investigated. How hinge region mutations affect enzyme kinetics and drug binding interactions are less clear. This study examines the effects of the L38↑H↑L double insertion mutation on the structure and function of the HIV-1 South African subtype C protease.

3.1. Protein overexpression and purification

3.1.1 Protease cDNA insert validation

The WT-CSA and L38↑H↑L protease cDNA sequences were validated by sending the extracted vectors to Inqaba biotech (Pretoria, South Africa) for Sanger sequencing. Upon sequence result analyses, it was confirmed that the pET-11a vectors contained the WT-CSA and L38↑H↑L protease gene, respectively (Figure 7). Basic Local Alignment Search Tool (Blastx) (Altschul *et al.*,1997) confirmed that the L38↑H↑L protease gene contains a double amino acid insertion of leucine and histidine at position 38 of the amino acid sequence relative to the WT-CSA sequence (Figure 8). The sequence alignment confirmed that the WT-CSA has 99 amino acids per monomer whereas the L38↑H↑L protease has 101 amino acids per monomer. No additional mutations were found in the L38↑H↑L protease cDNA relative to the WT-CSA protease cDNA.

3.1.2 Expression trials

The high yield expression of the HIV-1 protease is particularly difficult due to its cytotoxic and autocatalytic nature (Gustchina and Weber, 1990, Naicker *et al*., 2013). High yield recombinant expression of proteins in *E. coli* is particularly desirable when studying the structure and function of the proteins. Expression trials were performed to determine the optimal conditions required to produce high yields of the novel L38↑H↑L protease.

A.

```
Gwc ccc tcc att tct taa aat ttt gtt caa ctt taa gaa gga aaa ata cat atg cct caa
 X   p   S   I   S   -   N   F   V   Q   L   -   E   G   K   I   H   M   P   Q
atc act ctt tgg aaa cga ccc ctt gtc tca ata aaa gta ggg ggc cag ata aag gag gct
 I   T   L   W   K   R   P   L   V   S   I   K   V   G   G   Q   I   K   E   A
ctc tta gat aca gga gca gat gat aca gta tta gaa gaa ata aat ttg  cca gga aaa
 L   L   D   T   G   A   D   D   T   V   L   E   E   I   N   L    P   G   K
tgg aaa cca aaa atg ata gga gga att gga ggt ttt atc aaa gta aga cag tat gat
 W   K   P   K   M   I   G   G   I   G   G   F   I   K   V   R   Q   Y   D
caa ata ctt ata gaa att tgt gga aaa aag gct ata ggt aca gta tta gta ggg cct aca
 Q   I   L   I   E   I   C   G   K   K   A   I   G   T   V   L   V   G   P   T
cct gtc aac ata att gga aga aac atg ttg act cag ctt gga tgc aca cta aat ttt taa
 P   V   N   I   I   G   R   N   M   L   T   Q   L   G   C   T   L   N   F   -
tcc tcc ggc taa gga caa agc ccg aaa gga agc tga ata
 s   S   G   -   G   Q   S   P   K   G   S   -   I
```

B.

```
gwc aat tcc ctc tag aat aat ttt gtt taa ctt taa gaa gga gat ata cat atg cct caa
 X   N   S   L   -   N   N   F   V   -   L   -   E   G   D   I   H   M   P   Q
atc act ctt tgg aaa cga ccc ctt gtc tca ata aaa gta ggg ggc cag ata aag gag gct
 I   T   L   W   K   R   P   L   V   S   I   K   V   G   G   Q   I   K   E   A
ctc tta gat aca gga gca gat gat aca gta tta gaa gaa ata aat ttg cac ttg cca gga
 L   L   D   T   G   A   D   D   T   V   L   E   E   I   N   L   H   L   P   G
aaa tgg aaa cca aaa atg ata gga gga att gga ggt ttt atc aaa gta aga cag tat gat
 K   W   K   P   K   M   I   G   G   I   G   G   F   I   K   V   R   Q   Y   D
caa ata ctt ata gaa att tgt gga aaa aag gct ata ggt aca gta tta gta ggg cct aca
 Q   I   L   I   E   I   C   G   K   K   A   I   G   T   V   L   V   G   P   T
cct gtc aac ata att gga aga aac atg ttg act cag ctt gga tgc aca cta aat ttt taa
 P   V   N   I   I   G   R   N   M   L   T   Q   L   G   C   T   L   N   F   -
gga tcc ggc tgc taa caa agc ccg aaa gga agc tga gtt ggc tgc tgc cac cgc tga gca
 G   S   G   C   -   Q   S   P   K   G   S   -   V   G   C   C   H   R   -   A
taa cta gca
 -   L   A
```

Figure 7. Verification of the WT-CSA and L38↑H↑L protease gene insert in pET-11a vectors. (A) ExPASy TRANSLATE results of the HIV-1 WT-CSA and (B) L38↑H↑L protease gene insert into the pET-11a vector. The highlighted part (■) of the WT-CSA and (■) for the L38↑H↑L protease sequence represents the open reading frames. The atg start codon (■) and the taa stop codon (■) flanks the open reading frames of both sequences.

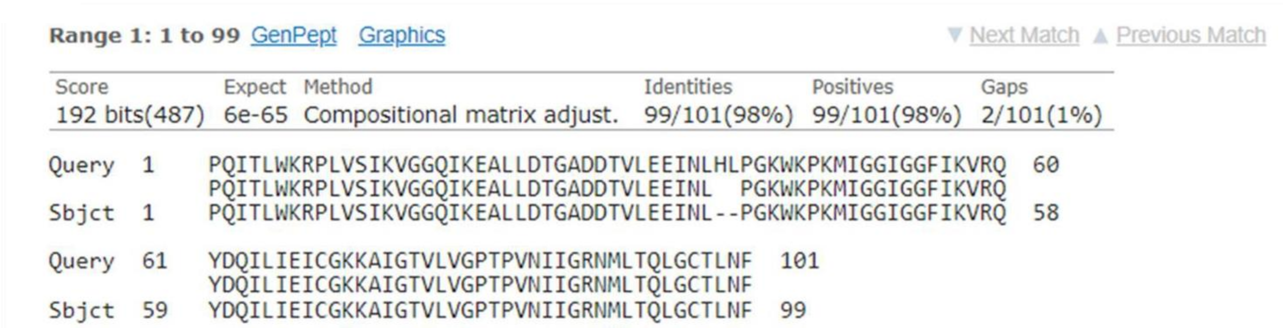


Figure 8. Amino acid sequence alignment of the WT-CSA and L38↑H↑L protease. Blast Basic Local Alighnment Tool for proteins (BLASTp) sequence alignment of the WT-CSA (PDB ID: 3U71) (subject) and L38↑H↑L (query) protease gene. There is a 98% sequence identity between the 2 amino acid sequences. The WT-CSA and the L38↑H↑L protease gene has 99 and 101 amino acids per monomer, respectively.

Expression conditions of the WT-CSA have been previously described (Naicker *et al*., 2013). The L38↑H↑L growth characteristics (section 2.3) showed that the transformed cells reached the mid-log phase (OD_{600nm} of 0.6) after approximately 2 hours. Upon performing expression trials (section 2.4), it was determined that the L38↑H↑L protease expresses optimally in *E. coli* BL21 (DE3) pLysS cells using 0.6 mM IPTG at 37 ºC, 200 rpm. The L38↑H↑L protease was expressed optimally 4 hours after IPTG induction (Figure 9) using normalised cell numbers.

Overexpression of the L38↑H↑L protease was confirmed by an apparent thick and dark band on the tricine SDS-PAGE gel (Figure 9). The WT-CSA and L38↑H↑L protease molecules were expressed as inclusion bodies found in the insoluble portion of the cell lysate. The high yield expression of the protein as inclusion bodies is extremely beneficial since the structure of the protein hinders its ability to autocatalyse. The overexpressed protein, under denaturing conditions, was determined to be approximately 11 kDa from the construction of an R_f value versus log molecular weight plot. The L38↑H↑L calculated monomeric size was close to that predicted by ExPAsy ProtPARAM (Gasteiger *et al*., 2005).

3.1.3 Protein purification and purity assessment

Overexpression and purification of the WT-CSA and L38↑H↑L proteases were performed by using the conditions as described in section 2.6. The WT-CSA and L38↑H↑L protease samples were successfully purified by using ion-exchange chromatography (Figure 10). Both the WT-CSA and the L38↑H↑L protease samples were purified from the insoluble portion of the whole-cell lysate. It is to be noted that a significant proportion of the L38↑H↑L protein sample, collected after a round of protein refolding and dialysis, was aggregated. An attempt in removing the aggregated proteins was performed by centrifuging the samples at 18000 xg, 4 ºC for 30 minutes. The unfolding and refolding process of the L38↑H↑L protease could be the source of the aggregation. Protein refolding by dialysis relies on the slow removal of urea from the protein sample. However, if this process occurs too fast, it may not allow sufficient time for the proteins to refold correctly which then resulted in protein aggregation. This provides a need for an improved purification protocol. Note that *in vivo*, the protease does not undergo protein unfolding and refolding and it is, therefore, recovered from the soluble fraction of the whole-cell lysate. However, this purification protocol requires the recovery of the protease from inclusion bodies and is different from the protease recovery process *in vivo*. This may also contribute in the development of protease aggregation

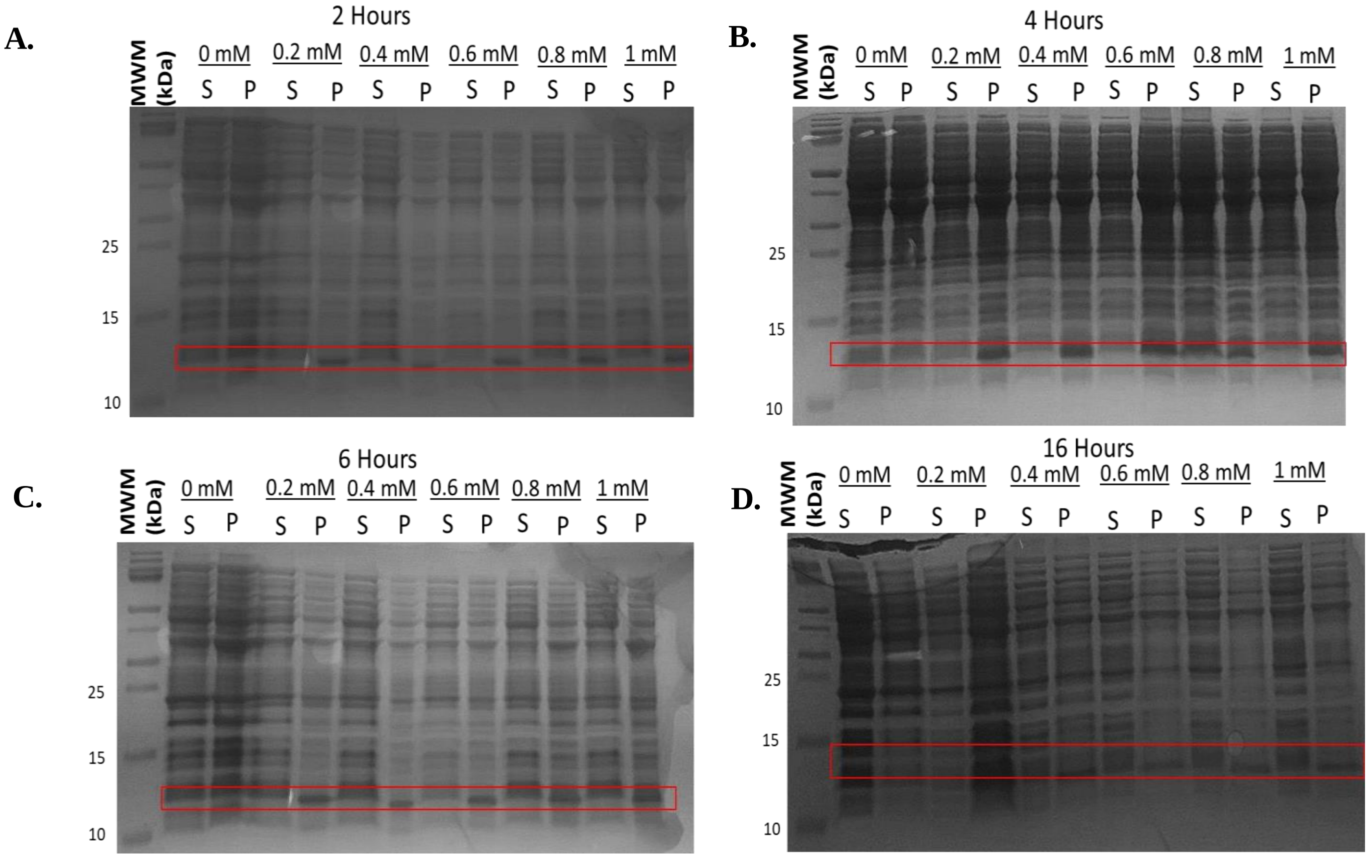


Figure 9. Tricine SDS-PAGE gels showing L38↑H↑L protease expression trials. Expression trials for the L38↑H↑L protease were performed in *E. coli* BL21 (DE3) pLysS cells as described in section 2.4. Tricine SDS-PAGE gels show the soluble and insoluble cell extracts collected at 2 (A), 4 (B), 6 (C), and 16 hours (D) after IPTG induction. S and P denotes supernatant and pellet, respectively. It was determined that the protease expressed optimally by using 0.6 mM IPTG for 4 hours, at 37 ºC and 200 rpm. A large portion of the overexpressed protein was found in the pellet (insoluble) portion of the whole-cell lysate, as seen by the thick band on the SDS-PAGE gels.

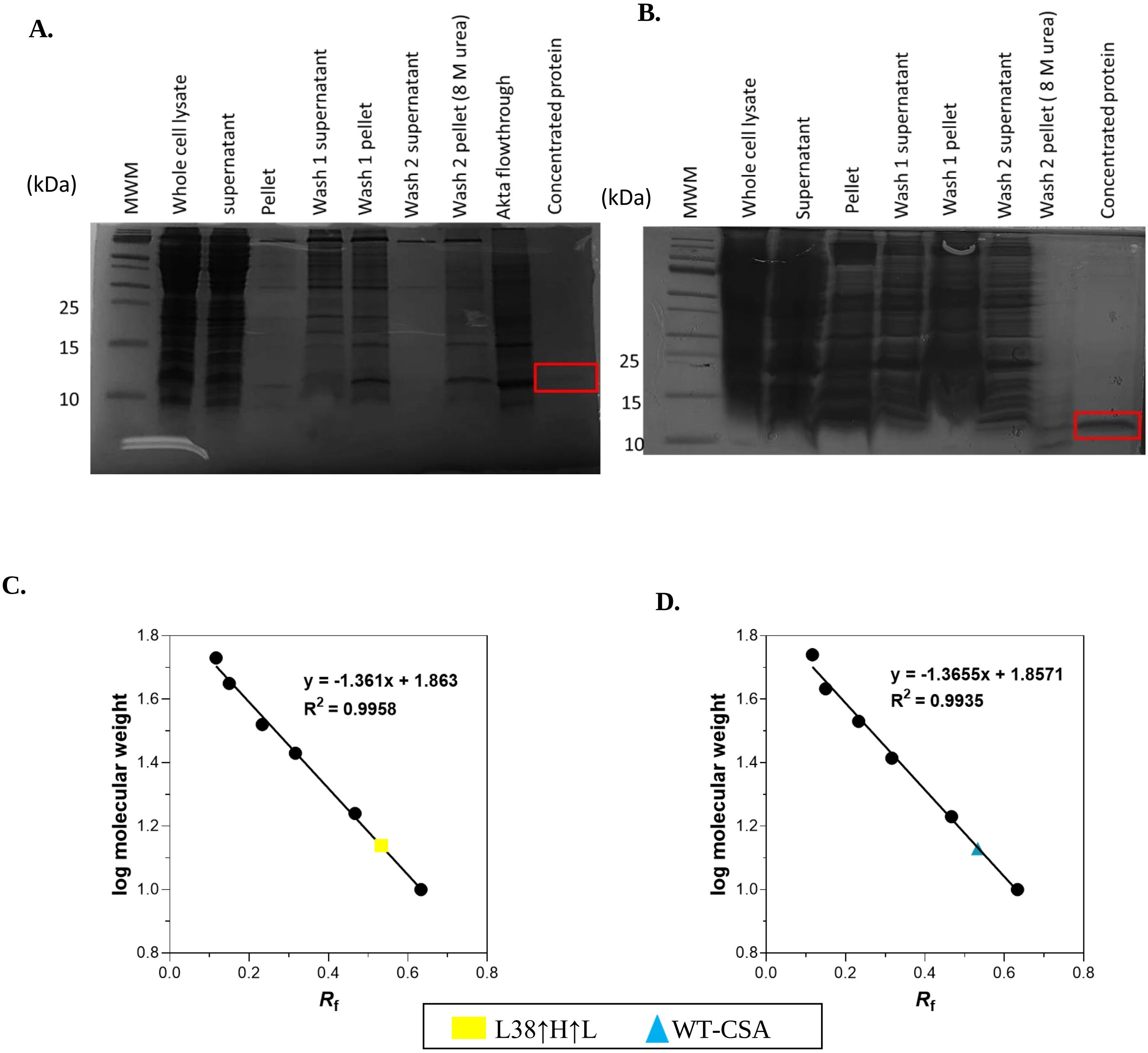


Figure 10. Purification profile of the L38↑H↑L and WT-CSA protease. Tricine SDS-PAGE gels showing the purification profiles of the L38↑H↑L protease (A) and WT-CSA protease (B). The protein samples were prepared under reducing conditions. From the R_f value calibration curves of the L38↑H↑L (C) and the WT-CSA protease (D), it was determined that the monomeric sizes of the L38↑H↑L and WT-CSA protease were ~10.61 kDa and ~11 kDa, respectively.

It was also noted that most of the refolded L38↑H↑L protease from the second round of ion-exchange chromatography did not bind to the cation-exchange column (Figure 10, AKTA flowthrough). This could be because the aggregated proteins remained in the protein sample even after centrifugation. Protein aggregation may impair the ability of the proteins to bind to the column resin. In contrast, no protein aggregation was observed after the refolding and dialysis of the WT-CSA protease and no WT-CSA protein was detected within the AKTA flowthrough. This result entails that a majority, if not all, the protein was able to bind to the column prior to elution.

The purity of the WT-CSA and L38↑H↑L protease was analysed on a tricine SDS-PAGE gel as described in section 2.5. Analyses from the construction of log molecular weight versus R_f values calibration curve (Figure 10 C and D) revealed that the WT-CSA and L38↑H↑L protease monomeric size was approximately 11 kDa and that the purified proteins were relatively pure.

3.2. Protein concentration determination

3.2.1 Protein concentration determination through absorbance spectroscopy

The concentration of the purified WT-CSA and L38↑H↑L protease was determined spectrophotometrically by performing serial dilutions as described in section 2.7.1. A linear regression plot coupled with the Beer-Lambert law was used to calculate the concentration of the protein from the slope of the plot. The average protein yield per 14 liters of LB media ranged from 0.56-0.61 mg/mL with an average volume of 5 mL for the WT-CSA and 0.38-0.42 mg/mL with an average volume of 5 mL for the L38↑H↑L protease.

3.2.2 Corrected protein concentration determination through ITC

Active site ITC experiments were performed to determine the concentration of protease in the sample that is functional and folded correctly. Active site experiments were performed as described in section 2.7.2. Due to the purification protocol that includes isolation of the protease molecules from inclusion bodies, and unfolding and refolding steps, it is important to determine whether the protease has folded correctly. It is also important to assess whether disulfide bonds have formed sufficiently between the protease monomers since the protein is functional as a homodimer.

Acetyl-pepstatin only binds to dimerised protease molecules. Acetyl-pepstatin binds to the homodimeric HIV-1 protease in a 1:1 ratio; therefore, it is possible to determine the concentration of the functional protease in a sample. This can then be used to correct the concentration determined from spectroscopic experiments. Upon analysis of the ITC thermograms and isotherms, the stoichiometry value (percentage) was approximately 58% (0.58) and 71% (0.71) for the WT-CSA and L38↑H↑L protease, respectively (Figure 11 A). This indicates that only 58% and 71% of the WT-CSA and L38↑H↑L protease concentration determined from section 3.2.1 represents 'useful' protein.

The thermodynamics parameters obtained from performing active site ITC provided fundamental information on the binding interactions of acetyl-pepstatin to the WT-CSA and L38↑H↑L protease, which include the nature of the binding reaction as well as the intermolecular forces driving the binding reactions. The division of the Gibbs free energy ΔG into its constituents, the change in enthalpy ΔH and change in entropy ΔS, reflect the thermodynamic means of interaction between the enzymes and the ligand. The ΔG, ΔH and -TΔS results were similar between the WT-CSA and L38↑H↑L protease (Figure 11). The binding reaction of acetyl-pepstatin to the WT-CSA and L38↑H↑L protease was found to be spontaneous due to the negative ΔG (-35.21 ± 4.49 kJ/mol and -34.98 ± 1.27 kJ/mol, respectively) (Figure 11 B). The binding reaction of acetyl-pepstatin to the WT-CSA and the L38↑H↑L protease was found to be endothermic (27.31 ± 6.27 kJ/mol and 31.42 ± 7.65 kJ/mol, respectively). Active site ITC experiments of HIV-1 proteases with acetyl-pepstatin by Todd and Freire (1999) was found to be endothermic as well. The ΔH upon ligand binding denotes the net formation and disruption of noncovalent bonds upon enzyme-ligand binding. Noncovalent bonds, such as van der Waals forces, hydrogen bonds and electrostatic interactions, results from interactions of the enzyme-solvent, ligand-solvent, as well as the enzyme-ligand complex (Chen *et al.*, 2016; Du *et al.*, 2016). The endothermic nature of the binding reaction of acetyl-pepstatin to the proteases is due to the hydrophobic nature of the ligand and the reduced ability to form strong noncovalent bonds with the enzymes active site upon binding (Figure 11 C). The positive ΔH indicated that the binding reaction is enthalpically unfavourable and that the favourable binding energy comes from the favourable entropy (-TΔS was -55.02 ± 11.76 kJ/mol and -65.57 ± 9.2 kJ/mol for the WT-CSA and L38↑H↑L protease, respectively). These results revealed that the binding reaction of acetyl- pepstatin to the WT-CSA and the L38↑H↑L protease is entropically driven i.e., $\left| T\Delta S \right| > \left| \Delta H \right|$. When a highly hydrophobic ligand binds to the active site of the enzyme,

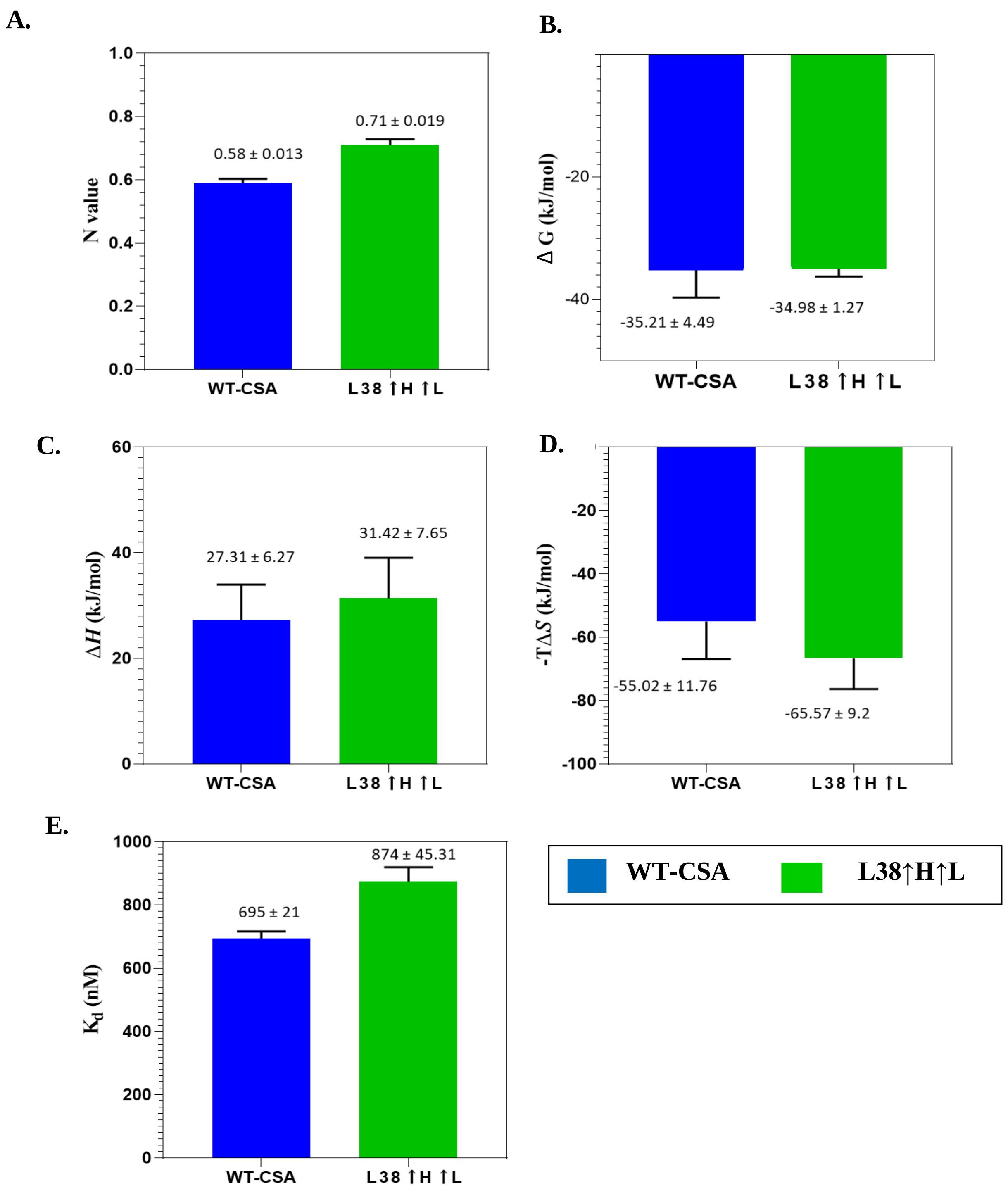


Figure 11. Thermodynamic parameters obtained from performing active site titration experiments. (A) stoichiometry (N value) plot, (B) Gibbs free energy (ΔG) plot, (C) Enthalpy (ΔH) plot, (D) Entropy (ΔS) plot (E) dissociation constant (K_d) plot. Active site ITC experiments were performed as described in section 2.7.2. The reaction of acetyl-pepstatin with the WT-CSA and L38↑H↑L protease occurred with a stoichiometry value of ~0.58 ± 0.013 and 0.71 ± 0.019, respectively. The data were fit using the GraphPad Prism version 9 software (GraphPad Software, San Diego, California, USA).

the solvent/water molecules occupying the binding cleft are displaced to the solvent bulk, which result in an unfavourable change in enthalpy. This is because the solvent molecules that were previously occupying the binding cleft and forming noncovalent bonds now partake in weaker hydrogen bonding in the solvent bulk (Fox *et al*., 2018). Whereas when ligands with polar groups become desolvated upon binding to the enzyme, they form relatively strong noncovalent bonds with the enzyme (Du *et al*., 2016). These strong noncovalent interactions contribute favourably to the binding enthalpy. It can be determined that the change in enthalpy due to the disruption of energetically favourable noncovalent bonds (endothermic) outweighs the formation of energetically noncovalent interactions involved in the binding reactions of acetyl-pepstatin to the WT-CSA and L38↑H↑L protease.

The major thermodynamic parameter difference that can be noted from the binding of acetyl-pepstatin to the WT-CSA and L38↑H↑L protease is the dissociation constant (K_d). Upon analysis of the K_d, it was noted that acetyl-pepstatin binds slightly less tight to the L38↑H↑L protease (874 ± 45.31 nM) relative to the WT-CSA (695 ± 21 nM). The lower the K_d value, the tighter the binding of a ligand to the protein. The K_d also represents the affinity the protein has for the ligand. The lower WT-CSA K_d value indicated that the protein has a higher affinity for the ligand relative to L38↑H↑L protease. The thermodynamic parameter findings for the WT-CSA are consistent with the findings in the literature (Mosebi *et al.,* 2008, Naicker *et al*., 2013).

3.3. Structural characterisation

3.3.1 Secondary structure content determination

The secondary structural content of the WT-CSA and L38↑H↑L protease was estimated by performing far-UV CD experiments as described in section 2.8.1. The WT-CSA and L38↑H↑L protease spectra both have a trough at 218 nm (Figure 12). The spectra exhibited a profile typical of a predominantly β-sheeted protein. The far-UV CD spectrum findings of the WT-CSA are consistent with crystallisation results of corresponding with a WT-CSA crystal structure obtained by Naicker *et al.*, (2013) (PDB ID: 3U71). The spectrum implies that the L38↑H↑L mutation does not alter the secondary structural content. However, far-UV CD

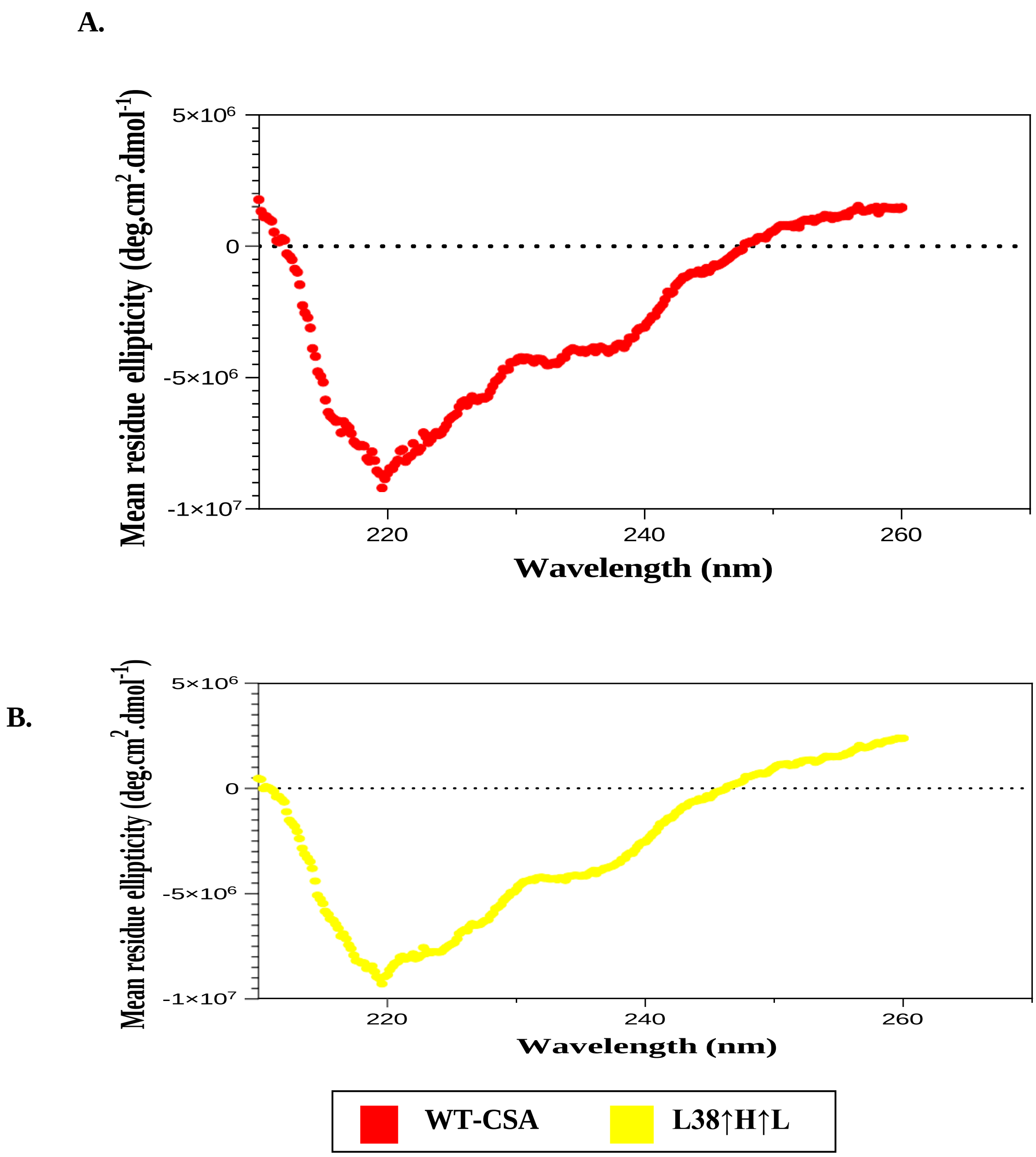


Figure 12. Assessment of the secondary structure content by using far-UV CD. The spectra the WT-CSA (A) and L38↑H↑L (B) both exhibit a minimum at 218 nm which is typical for a predominantly β-sheeted protein. The experiments were performed as described in section 2.8.1. the experiment was performed in Buffer D. The protein concentrations used were 5 µM for both the WT-CSA and L38↑H↑L protease. The spectra were generated with an average of 8 accumulations. The experiment was performed in replicates. The data were fit using the GraphPad Prism version 9 software (GraphPad Software, San Diego, California, USA).

spectra generated do not provide information on where the secondary structures are located within a protein.

3.3.2 Quaternary structure determination

The chromatograms generated from performing SE-HPLC experiments were used to determine the oligomeric size of the WT-CSA and L38↑H↑L protease. SE-HPLC experiments were performed as described in section 2.8.2. It was observed that the WT-CSA and L38↑H↑L protease exist in both monomeric and dimeric forms (Figure 13 and 14). The molecular sizes of the WT-CSA protease were estimated to be ~21233 Da (dimer) and ~10665 Da (monomer). Whereas the molecular sizes of the L38↑H↑L protease were estimated to be ~22284 Da (dimer) and-MSC ~10256 Da (monomer). These findings are not significantly different from predicted molecular sizes. According to ExPasy ProtPARAM tool (Gasteiger *et al*., 2005), the dimeric and monomeric sizes of the WT-CSA were 22010 Da and 11005 Da, respectively. Whereas the predicted molecular weight for the dimeric and monomeric L38↑H↑L protease were 22010 Da and 11005 Da, respectively. The difference in molecular sizes determined from the calibration curves can be attributed to behavior of the protease when it passes through the size exclusion column and the low sensitivity of the HPLC detector in detecting the elution of proteins from the column. This can cause inaccuracies in the retention times obtained. The preferred conformation of the protease molecules seems to be in lower concentrations relative to monomeric forms since the monomeric absorbance reading is higher than that dimeric form. Findings from performing active site ITC experiments, as previously mentioned in section 3.2.2, also confirm that not all of the WT-CSA and L38↑H↑L purified protease molecules are in their dimeric and functional form.

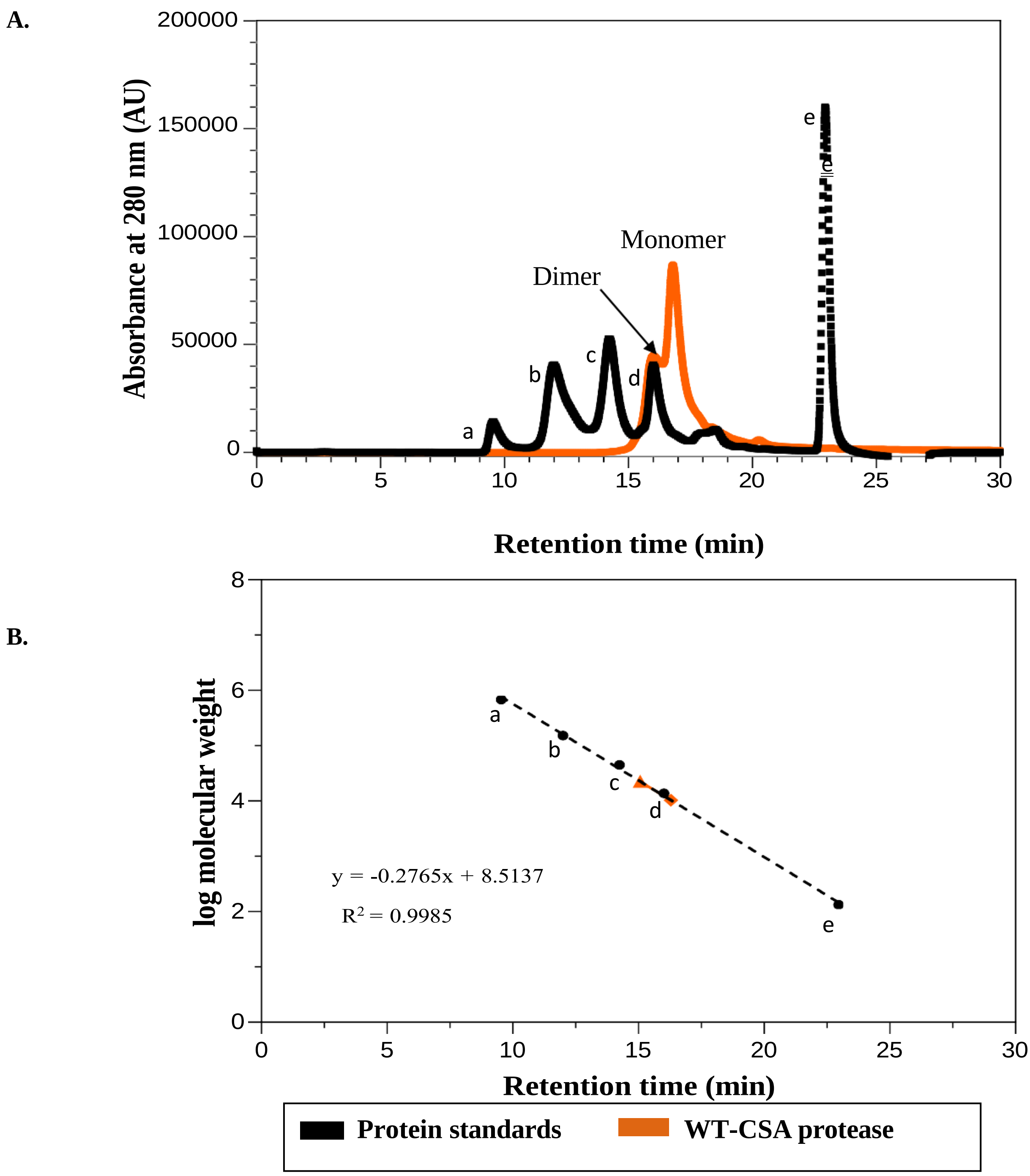


Figure 13. Quaternary structure assessment of the WT-CSA protease through SE- HPLC. SE-HPLC chromatogram (A) and the molecular weight calibration curve (B) of the WT-CSA. Both plots show that the protease existed in monomeric and dimeric forms when passing through the column. The monomeric and dimeric size of the protease was determined to be ~10665 Da (■) and ~21233 Da (▲), respectfully. The protein standard mixture consisted of (a) thyroglobulin bovine (670 000 kDa), (b) γ-globulins (150 000 kDa), (c) Albumin (44 300 kDa), (d) ribonuclease A (13 700 kDa) and (e) Para-aminobenzoic acid (137.14 Da).

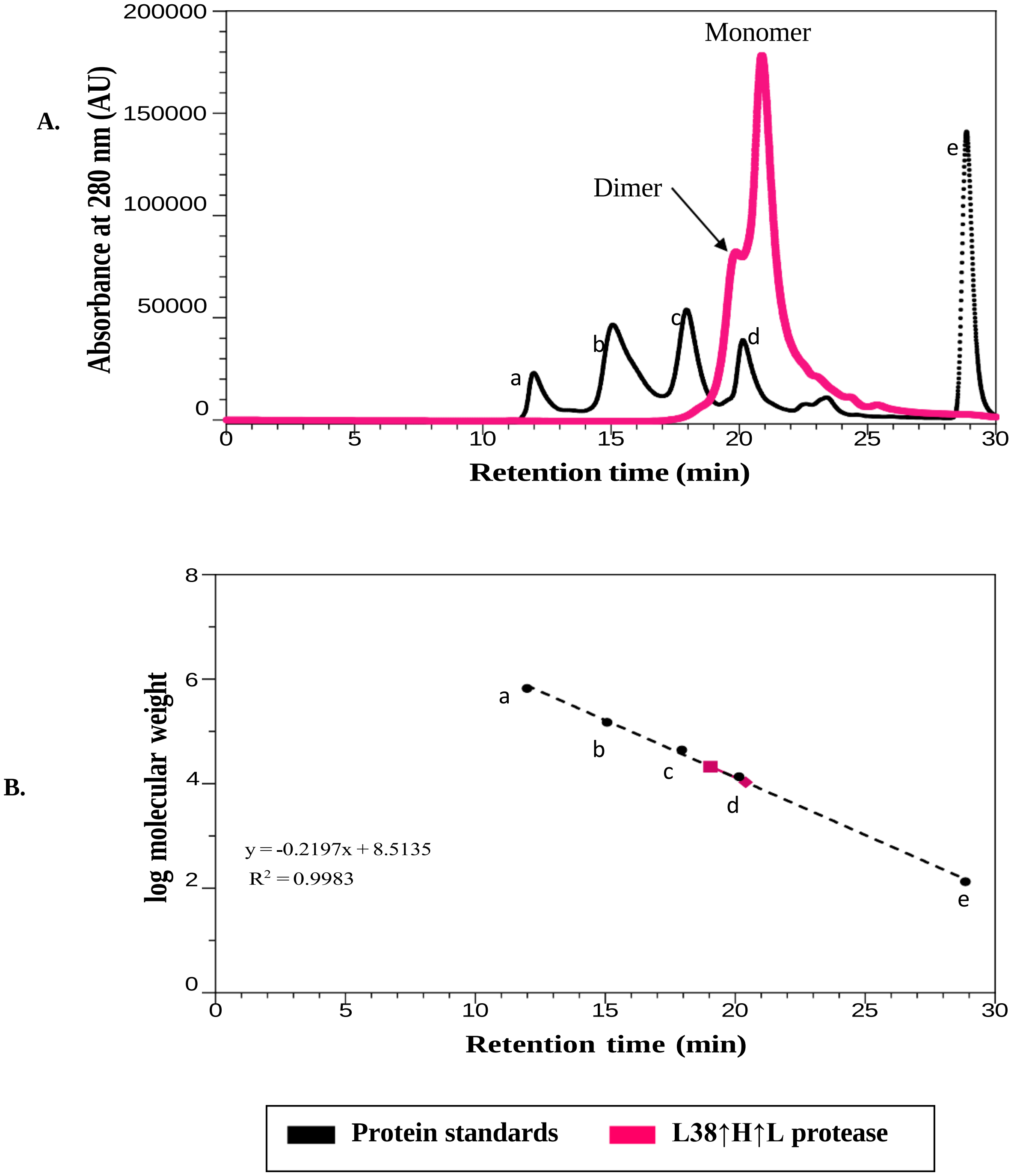


Figure 14. Quaternary structure assessment of the L38↑H↑L protease through SE-HPLC. SE-HPLC chromatogram (A) and the molecular weight calibration curve of the L38↑H↑L protease. Both plots show that the protease exists in monomeric and dimeric forms when passing through the column. The monomeric and dimeric sizes of the protease were determined to be were ~11005 Da (◆) and ~22010 Da (■) respectfully. The protein standard mixture consisted of (a) thyroglobulin bovine (670 000 kDa), (b) γ-globulins (150 000 kDa), (c) Albumin (44 300 kDa), (d) ribonuclease A (13 700 kDa) and (e) Para-aminobenzoic acid (137.14 Da).

3.4 Functional characterisation

3.4.1 Kinetics of substrate hydrolysis

The kinetic analysis of the WT-CSA and L38↑H↑L protease was done by performing enzyme kinetic studies as described in section 2.9.1. The buffer used to perform the enzyme assays (pH 5.0) mimics the environment near the plasma membrane of the CD4 cells where the cleavage of the Gag-Pol and Gag polyproteins occur (Henderson *et al*., 1983; Mervis *et al*., 1988). Specific activity assays were performed to ensure that both proteases are catalytically active. It was important to ensure activity since the WT-CSA and L38↑H↑L protease were recovered from inclusion bodies during the purification protocol. The specific activity of the L38↑H↑L protease was 0.71-fold lower relative to the WT-CSA protease (Figure 15 A). The V_{Max} of the L38↑H↑L protease was 2.25-fold lower than the WT-CSA protease. This demonstrated that the L38↑H↑L protease requires less substrate to achieve maximal velocity. The K_M of the L38↑H↑L protease displayed a 15-fold reduction relative to the WT-CSA protease (Figure 15 C). The lower K_M value indicated that the mutant has an increased affinity for the substrate relative to the WT-CSA protease. The L38↑H↑L protease displayed a reduced rate of substrate processing (k_{cat} value was 6.89 ± 1.76 sec^{-1}) relative to the WT- CSA protease, which displayed a k_{cat} value of 9.83 ± 0.89 sec^{-1} (Figure 15 D). The k_{cat} kinetic parameter was influenced by substrate binding which corresponded well with the K_M findings. It was determined that the substrate was held more tightly in the active site of the protein which then delays the release of the product from the active site. If the substrate binds too tightly to the protease, the binding reaction intermediate becomes too stable for the product to be released. This then decreases the number of product molecules released from the active site per second, hence the relatively lower k_{cat} value (Albery and Knowles, 1976). The WT-CSA protease experimental results did not vary from published literature (Naicker *et al*., 2013; Williams *et al*., 2019). The L38↑H↑L protease displayed a higher k_{cat}/K_M value (3.02 ± 0.08 sec^{-1}.µM^{-1}) relative to the WT-CSA protease (0.0265 ± 0.14 sec^{-1}.µM^{-1}) (Figure 15 E). This demonstrated that the L38↑H↑L protease was more catalytically efficient in catalysing the hydrolysis of the fluorogenic substrate relative to the WT-CSA protease.

The fluorogenic substrate used was specifically designed to mimic the capsid/p2 cleavage sites on the WT-CSA Gag polyprotein. Due to the close link between the Gag polyprotein and the subsequent protease product, mutations in the Gag-polyprotein cleavage

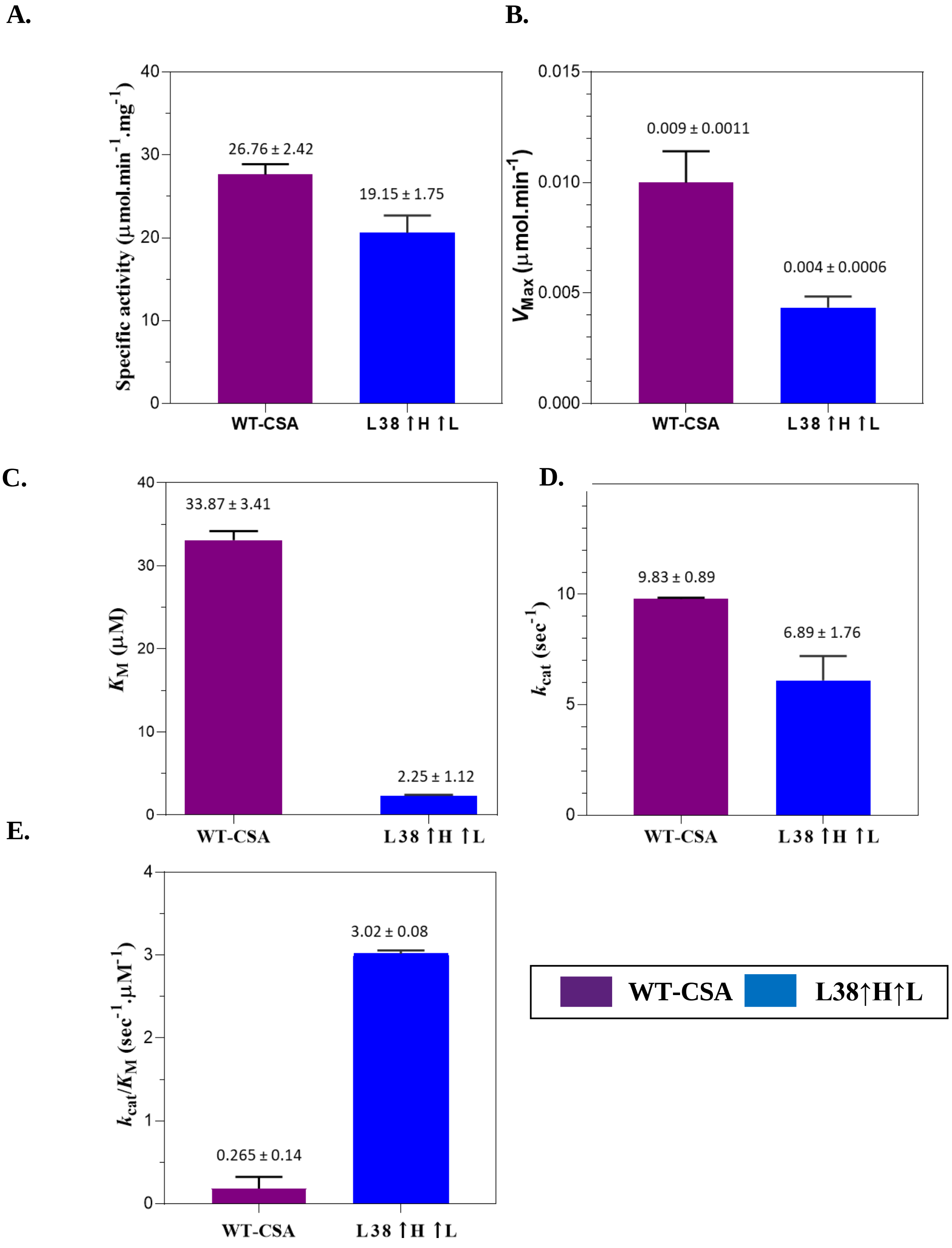


Figure 15. Kinetics parameters obtained from performing enzyme kinetics assays. (A) Specific activity (B) Maximum velocity (C) Michaelis-Menten (D) Catalytic turnover (E) Catalytic efficiency plots collectively show that the L38↑H↑L protease displays enhanced catalytic efficiency in the hydrolysis of the fluorogenic substrate. The error bars represent ± SD of each triplicate experiment. The data were fit and analysed using the GraphPad Prism version 9 software (GraphPad Software, San Diego, California, USA).

and non-cleavage sites have been observed (Doyon *et al*., 1996; Clavel and Mammano., 2010; Su *et al*., 2019). Some of the mutations found in the Gag polyprotein result in mutated protease products (Williams *et al*., 2019). An important factor in determining rates of cleavage is the amino acid sequence of the polyprotein. Therefore, we need to perform genotypic analysis of the L38↑H↑L associated Gag polyprotein in order to determine whether there are mutations present in the polyprotein. The mechanisms through which Gag cleavage site mutations can affect protease inhibitor binding are yet to be understood (Williams *et al*., 2019; Zondagh *et al.*, 2019). Cleavage and non-cleavage site mutations in the Gag polyprotein may improve the binding interactions of the polyprotein to the protease. Improved or reduced cleavage of the Gag-polyprotein may affect enzyme kinetics and even drug binding (Clavel and Mammano, 2010).

3.4.2 Competitive binding of the substrate and inhibitor

Inhibition studies were performed to determine the concentration of protease inhibitor required to reduce the activity of the protease by 50% (Burlingham and Widlanski, 2003). The IC_{50} experiments were performed by initially preincubating the enzyme with each of the respective inhibitors then adding the fluorogenic substrate to initiate the hydrolysis reaction. As a result, the activity of the enzyme decreases as the inhibitor concentration increase. This is due to the competitive binding of the substrate and inhibitor for the active site of the enzyme (Velazquez-Campoy *et al*., 2001; Naicker *et al*., 2013). The IC_{50} for Saquinavir, Darunavir and Atazanavir were determined for the WT-CSA and L38↑H↑L protease as described in section 2.9.5 A final DMSO concentration of 2% was used to ensure inhibitor solubility. IC_{50} values were estimated from dose-dependent curves, and the data were fit and analysed using the GraphPad Prism version 9 software (GraphPad Software, San Diego, California, USA). According to figure 16, the IC_{50} values between the WT-CSA and L38↑H↑L protease varied remarkably, with the L38↑H↑L protease exhibiting lower IC_{50} values when compared to the WT-CSA protease. The L38↑H↑L protease displayed a 5.5-fold, 2.7-fold and 2-fold reduction in IC_{50} when bond to Saquinavir (9.84 ± 1.73 nM), Darunavir (12.40 ± 1.34 nM) and Atazanavir (16.98 ± 2.65 nM) respectively, relative to the WT-CSA protease which displayed IC_{50} values of 26.43 ± 5.10 nM, 15.07 ± 3.80 nM and 20.02 ± 2.1 nM for Saquinavir, Darunavir and Atazanavir, respectively. This suggested that the double insertion mutation influenced the binding ability of protease inhibitors The molecular basis for the difference in Saquinavir binding to the WT-CSA and L38↑H↑L protease is unknown at this stage. However, detailed molecular docking studies or x-ray crystallography may shed light on the differences observed. It can be concluded that the L38↑H↑L protease requires far less inhibitor concentrations in order to reduce the activity of the protease by 50% relative to the WT-CSA protease. The IC_{50} values also provided information of the binding affinity of the inhibitors to the proteases. It was determined that the L38↑H↑L protease displayed increased affinities for Saquinavir, Darunavir and Atazanavir relative to the WT-CSA protease. This can be confirmed by the lower IC_{50} values obtained from performing the enzyme inhibition kinetic assays. The IC_{50} values obtained for the WT-CSA are comparable with those in literature (Naicker *et al*., 2013).

3.4.3. Thermodynamics of enzyme-inhibitor binding

To date, the only technique that can measure protein-ligand binding thermodynamics at constant temperatures comes from performing ITC experiments. The affinity measurements, such as association/dissociation constants, for tight binding inhibitors cannot be accurately determined from active site titrations, therefore displacement titration experiments are employed to solve the problem. Displacement titration ITC allows for the determination of the binding thermodynamics and association constants of tight binding inhibitors (K_a > 10^9 M^{-1}) such as FDA-approved protease inhibitors (Ohtaka *et al*., 2002). The experimental conditions between the WT-CSA and L38↑H↑L protease titration experiments were kept identical. In order to evaluate the thermodynamic parameters underlying protease- inhibitor binding, Saquinavir, Darunavir and Atazanavir protease inhibitors were chosen to displace prebound acetyl-pepstatin molecules from the active site of the WT-CSA and L38↑H↑L protease. Acetyl-pepstatin was chosen as the inhibitor to be displaced because it binds to HIV-1 proteases relatively weaker than FDA-approved (Ohtaka *et al*., 2002; King *et al*., 2004), it also binds to HIV-1 proteases in an endothermic and unfavourable manner (positive ΔH) as seen in section 3.2.2. A comprehensive understanding of protein-ligand interactions requires a complete description of several thermodynamic parameters. A quantitative description of the forces driving protein-ligand interactions are the ΔG, -TΔS, ΔH and K_d. Displacement titration experiments were performed as described in section 2.10. Upon analysis of the thermograms and isotherms obtained from performing the experiments, the apparent differences between the WT-CSA and L38↑H↑L protease-inhibitor binding reactions came from the ΔH, -TΔS and K_d (Figure 17).

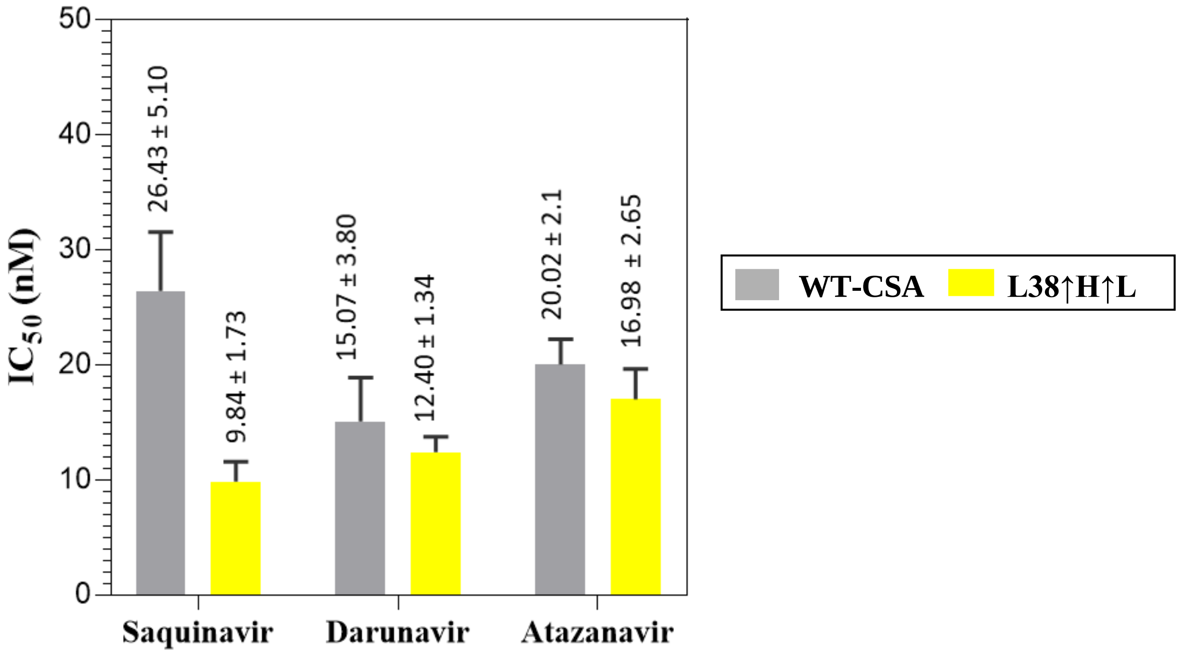


Figure 16. IC_{50} determinations of the WT-CSA and L38↑H↑L protease with the respective inhibitors. It was determined that the L38↑H↑L protease requires less inhibitor to reduce the activity of the protease relative to the WT-CSA protease. The IC_{50} values were determined from dope-response curves. The experiments were performed as described in section 2.9.6. The error bars represent ± SD of each triplicate experiment. The data were fit and analysed using the GraphPad Prism version 9 software (GraphPad Software, San Diego, California, USA)

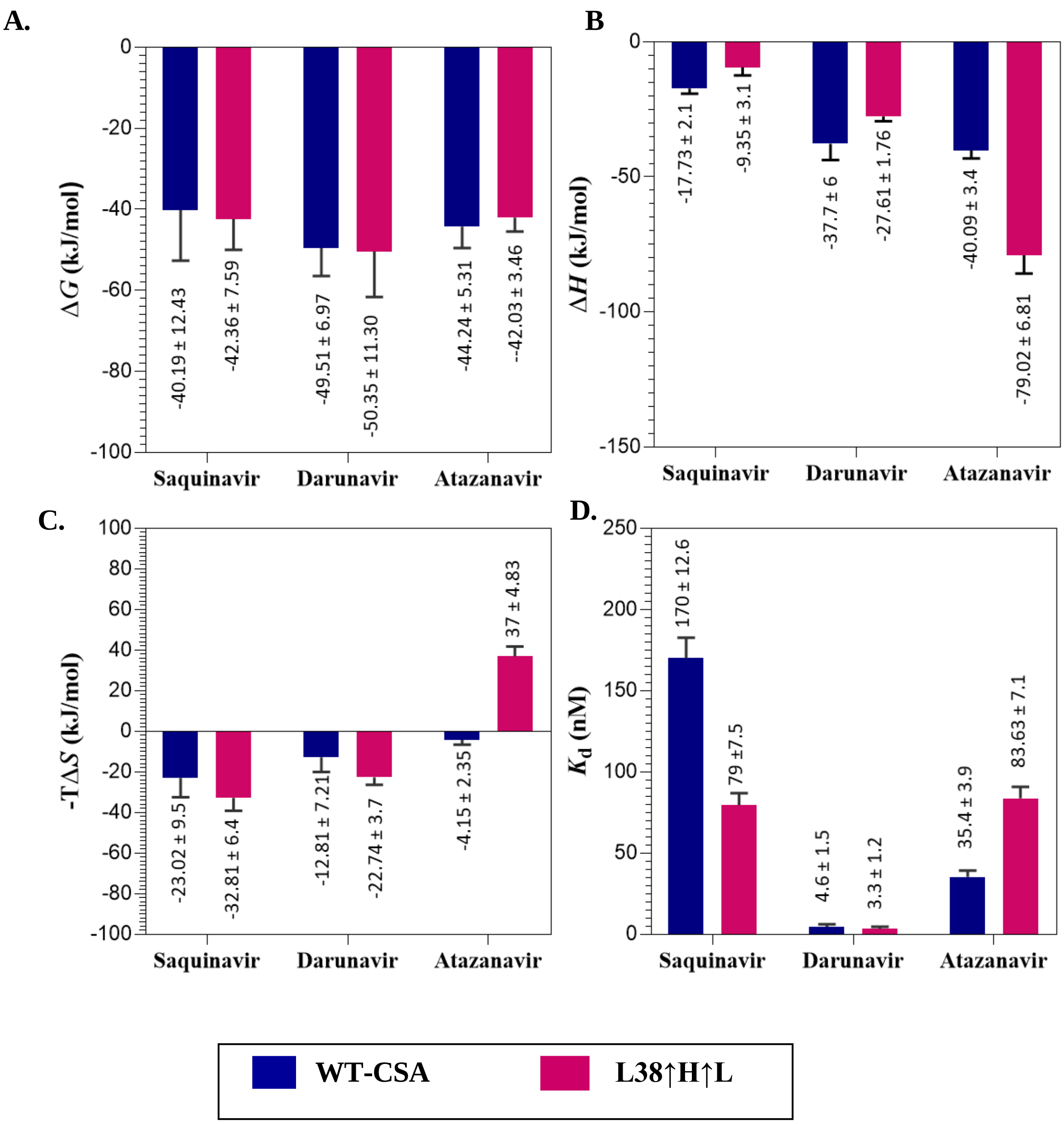


Figure 17. Thermodynamic parameters obtained from protease-inhibitor binding interactions. (A) Gibbs free energy plot (B) Enthalpy plot (C) Entropy plot (D) dissociation constant plot. Darunavir binds the tightest to the L38↑H↑L protease, followed by Atazanavir and Saquinavir. The binding reaction of both proteases with the respective inhibitors was found to be spontaneous and enthalpically favourable. The binding of the L38↑H↑L protease to Atazanavir is entropically unfavourable as opposed to Darunavir and Saquinavir. The error bars represent ± SD of each duplicate experiment. The data were fit using the GraphPad Prism version 9 software (GraphPad Software, San Diego, California, USA)

The change in Gibbs free energy (Δ*G*) remained relatively unchanged for both the WT-CSA and L38↑H↑L protease when bound to the respective protease inhibitors (Figure 17 A). The binding of the respective protease inhibitors to the WT-CSA and L38↑H↑L protease were found to be spontaneous, given the negative Δ*G* values obtained. In protein-ligand binding reactions, the Δ*G* also provides information on the stability of the ligand-bound protein. Protein-ligand complexes which have relatively larger and more negative Δ*G* values are considered to be more stable. It was noted that the L38↑H↑L protease displayed increased stability when bound to Saquinavir (-42.36 ± 7.59 kJ/mol) and Darunavir (-50.35 ± 11.30 kJ/mol) relative to the WT-CSA protease which displayed Δ*G* values of -40.19 ± 12.43 kJ/mol and -49.51 ± 6.97 kJ/mol, respectively. However, when bound to Atazanavir, the L38↑H↑L protease displayed a reduced stability (-42.03 ± 3.46 kJ/mol) relative to the WT-CSA protease (-44.24 ± 5.31 kJ/mol).

In a protein-ligand binding reaction, the net enthalpy change (Δ*H*) is a result of the disruption of noncovalent bonds between the protein-solvent and ligand-solvent interactions, the dispersion of solvent molecules from the binding interface and formation of noncovalent bonds between the protein-ligand (Perozzo *et al*., 2004). The binding of the respective protease inhibitors to the WT-CSA and L38↑H↑L protease were found to be enthalpically favourable and exothermic (negative Δ*H*) (Figure 17 B). A favourable Δ*H* often indicates the formation of noncovalent bonds such as van der Waals forces, hydrogen bonds and electrostatic interactions at the binding interface of the protein (Bronowska, 2011). The magnitude of the Δ*H* varied considerably between the WT-CSA and the L38↑H↑L protease. The binding reaction of the L38↑H↑L protease with Atazanavir (-79.02 ± 6.81 kJ/mol) resulted in a ~2-fold increase in the magnitude of Δ*H* relative to the WT-CSA protease (-40.09 ± 3.40 kJ/mol). Whereas the binding reaction of the L38↑H↑L protease with Saquinavir (-9.35 ± 3.10 kJ/mol) and Darunavir (-27.61 ± 1.76 kJ/mol) resulted in ~1 fold decrease in Δ*H* magnitude relative to the WT-CSA protease, which exhibited Δ*H* values of -17.73 ± 2.10 kJ/mol and -37.7 ± 60 kJ/mol, respectively. The magnitude of a favourable Δ*H* provided insights into ligand/inhibitor specificity for a protein (Chen *et al*., 2016; Du *et al*., 2016). From figure 17 B, the L38↑H↑L protease displayed a greater specificity for Atazanavir (more negative Δ*H*), relative to the WT-CSA protease. Atazanavir may form more noncovalent bonds with the L38↑H↑L protease relative to the WT-CSA protease. This would result in a larger and more negative Δ*H* for Atazanavir binding to the L38↑H↑L protease. On the other hand, the L38↑H↑L protease displayed a reduced specificity for Saquinavir and Darunavir relative to the WT-CSA protease.

This implies that the L38↑H↑L protease forms less noncovalent bonds with Saquinavir and Darunavir when compared to the WT-CSA protease. However, the Δ*H* thermodynamic parameter alone cannot provide all the necessary information on the thermodynamic interactions involved in ligand binding. Entropy is another contributing factor to protein-ligand binding interactions.

The change in entropy (Δ*S*) represents the total order/disorder or randomness of molecules in a system. The Δ*S* associated with protein-ligand binding consist of the following terms:

$$\Delta S = \Delta S_{\mathrm{solv}} + \Delta S_{\mathrm{conf}} + \Delta S_{\mathrm{r/t}} \qquad [9]$$

Where ΔS_{solv} is the entropy change associated with the release of solvent molecules from the binding interface, ΔS_{conf} is the entropy change associated with the conformational freedom of the ligand and protein upon binding, $\Delta S_{\mathrm{r/t}}$ is the entropy change associated with the loss of rotational and translational freedom of the protein and ligand upon binding (MacRaild *et* al., 2007; Du *et al*., 2016). The ΔS_{solv} makes a favourable contribution to the Δ*S* due to its large positive value. The ΔS_{conf} can either make a favourable or unfavourable contribution the Δ*S*. The degree of conformational freedom of the protein-ligand complex can increase or decrease relative to the free protein and ligand in the solution. The $\Delta S_{\mathrm{r/t}}$ contributes unfavourably to the Δ*S* due to the reduced rotational and translational freedom of the protein-ligand complex and reduces the number of free molecules relative to the free protein and ligand in the solution.

It was observed that the binding of Atazanavir to the L38↑H↑L protease was entropically unfavourable (-TΔ*S* was 37 ±4.83 kJ/mol). In contrast, the binding of Atazanavir to the WT-CSA protease was entropically favourable (-TΔ*S* was -4.15 ± 2.35 kJ/mol) (Figure 17 C). The binding of Saquinavir and Darunavir to both the WT-CSA and L38↑H↑L protease was found to be entropically favourable ($-T\Delta S < 0$). The L38↑H↑L protease exhibited a ~1 fold increase in the magnitude of the -TΔ*S* when bound to Saquinavir (-32.81 ± 6.40 kJ/mol) and Darunavir (-22.74 ± 3.70 kJ/mol) relative to the WT-CSA protease, which displayed -TΔ*S* values of -23.02 ± 9.50 kJ/mol and -12.81 ± 7.21 kJ/mol, respectively. The reduced entropic ($-T\Delta S > 0$) contribution of the binding interaction between the L38↑H↑L protease and Atazanavir was compensated by large, negative and favourable enthalpic change with no significant change in Δ*G* (Figure 17 A).

For a spontaneous binding reaction to occur (negative ΔG), enthalpy-entropy compensation phenomenon occurs (Du *et al.*, 2016). Enthalpy-entropy compensation is influenced by the solvation and desolvation of the binding interface and the formation of noncovalent forces (Dunitz, 1995). From the binding interactions of Atazanavir and the L38↑H↑L protease, it can be determined that the unfavourable ΔS_{conf} and $\Delta S_{r/t}$ contributed more to the overall binding ΔS than the favourable ΔS_{solv} (equation 9). This result in an overall negative binding ΔS ($-T\Delta S > 0$). Whereas the binding reaction of Atazanavir to the WT-CSA protease demonstrated an unfavourable binding ΔS. This result suggested that the unfavourable ΔS_{solv} dominates over the favourable ΔS_{conf} and $\Delta S_{r/t}$. The binding ΔS of Saquinavir and Darunavir to both proteases were dominated by unfavourable ΔS_{solv}. The ΔH and ΔS thermodynamic parameters obtained from the binding interaction of L38↑H↑L protease with Atazanavir may be attributed to the increased order of the protease-inhibitor-solvent complex formed upon inhibitor binding (Ladbury, 1996). It can be determined that the binding reaction of Saquinavir to both proteases was entropically driven ($|\Delta H| < |T\Delta S|$), whereas the binding reaction of Darunavir and Atazanavir to both proteases were enthalpically driven ($|\Delta H| > |T\Delta S|$). Another parameter that was used to compare the binding interactions of the respective inhibitors with the WT-CSA and L38↑H↑L protease was the dissociation constant (K_d). The K_d relates to the binding affinity of the inhibitors for the proteases. Enzymes displaying relatively lower K_d values are considered to have a larger affinity and susceptibility for the ligand/inhibitor.

The dissociation constants of the respective protease inhibitors to the WT-CSA and L38↑H↑L protease were determined from performing displacement ITC (Figure 17 D). Upon analysis, it was noted that L38↑H↑L protease displayed a 2-fold increased drug susceptibility towards Saquinavir (79 ± 7.5 nM), and 2-fold reduced drug susceptibility towards Atazanavir (83.63 ± 7.1 nM) relative to the WT-CSA protease which displayed K_d values of 170 ± 12.6 nM and 35.4 ± 3.9 nM, respectively. The drug susceptibility of the L38↑H↑L protease for Darunavir (3.3 ± 1.2 nM) remained unchanged relative to the WT-CSA protease (4.6 ± 1.5 nM). Inhibitors that bind to enzymes with reduced K_d values bind the enzyme with a better affinity relative to proteins with higher K_d values (King *et al.*, 2004; Mittal *et al.*, 2013).

It was observed that a larger binding specificity does not necessarily mean greater binding affinity. Atazanavir displayed the greatest specificity when bound to the L38↑H↑L protease (Figure 17 B) relative to Saquinavir and Darunavir. However, Atazanavir also displayed a reduced affinity i.e., an increased K_d value, relative to the WT-CSA protease. Ligand binding affinity pertains to the strength and quality of the noncovalent bonds, particularly hydrogen bonds, formed between the ligand and the protein, whereas specificity pertains to the number of noncovalent bonds, particularly polar/hydrogen bonds formed between the protein and the enzyme. It was determined that Darunavir exhibited relatively low binding specificity and this is significant considering that Darunavir is the most recent FDA-approved HIV-1 protease inhibitor designed for high potency against WT proteases, mutated proteases as well as multidrug-resistant proteases (King *et al*., 2004). Darunavir also remained potent against the L38↑H↑L protease. The binding affinity of Darunavir for the WT-CSA as well as the mutated L38↑H↑L protease was maintained. The strong interactions of the *bis*-tetrahydrofuranyl urethane group only found in Darunavir and the Asp29 and Asp30 amino acids on the protease subunits contributed to the high binding affinity of the inhibitor (King *et al*., 2004; Lefebvre and Schiffer, 2008).

Computational molecular dynamics simulations as well as protein crystallographic studies of the WT-CSA and L38↑H↑L protease in the apo- and drug complex form can provide additional information on the thermodynamic binding interactions of the WT-CSA and L38↑H↑L protease with the respective inhibitors.

4. Conclusion

A comparative study between the WT-CSA and L38↑H↑L protease was carried out in order to understand the possible structural and functional effects the double insertion causes relative to the WT-CSA protease. The data suggested that amino acids in the hinge region of HIV-1 proteases indirectly affect enzyme catalysis as well as substrate and inhibitor specificity but has no apparent effect on the global structure of the L38↑H↑L protease. Due to the important role of the hinge and flap region in accommodating substrates and inhibitors into the protease active site, mutations in the hinge region appear to affect the stability and movement of this region. This, in turn, affects the kinetics of substrate hydrolysis and thermodynamics of drug binding.

5. References

Abbyad, P., Childs, W., Shi, X., and Boxer, S.G. (2007) Dynamic stokes shift in green fluorescent protein variants. *The national academy of sciences of the USA.* **104**(51):20189-20194

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

Albery, W.J. and Knowles, J.R. (1976) Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry*. **15**(27):5631-5640

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-PLAST: a new generation of protein database search programs. *Nucleic Acids Research.* **25**(17):3389-3402

Babe, L.M., Rose, J., and Craik, C.S. (1992) Synthetic interface peptides alter dimeric assembly of the HIV-1 and HIV-2 proteases, *Protein Sci.* **1**:1244-1253

Barth, H., Boyes, B., and Jackson, C. (1994) Size exclusion chromatography. *Anal. Biochem.* **66**:595-620

Benson, D. A., Boguski, M. S., Lipman, D. J, Ostell, J., and Ouellette, B. F. (1998) GenBank. *Nucleic Acids Res.* **26**:1-7

Bessong, P.O. (2008) Polymorphisms in HIV-1 subtype c proteases and the potential impact on protease inhibitors. *Tropical Medicine and International Health.* **13**(2):144-151

Briggs, G. E., and Haldane, J.B.S. (1925) A note on the kinetics of enzyme action. *Biochemical Journal*. **19:**338.

Brodine, S.K., Mascola, J.R., Weiss, P.J., Ito, S.I., Porter, K.R., Artenstein, A.W., Garland, F.C., McCutchan, F.E., and Burke, D.S. (1995) Detection of diverse HIV-1 genetic subtypes in the USA. *Lancet.* **346**:1198-9

Buonaguro, L., Tornesello, M. L., and Buonaguro, F. M. (2007) Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: Pathogenetic and therapeutic implications. *Journal of Virology*. **8**: 10209–10219

Côté, H.C.F., Brumme, Z.L., and Harrigan, P.R. (2001) Human Immunodeficiency Virus Type 1 Protease Cleavage Site Mutations Associated with Protease Inhibitor Cross-

Resistance Selected by Indinavir, Ritonavir, and/or Saquinavir. *Journal of Virology*. **75**:589–594

Chad, D.C., and Kim, P.S. (1998) HIV entry and its inhibition. *Cell.* **93**(5): 681-684

Chen, Z., Li, Y. and Schock, H.B. (1995) Three-dimensional structure of a mutant HIV-1 protease displaying cross-resistance to all protease inhibitors in clinical trials. *J. Biol. and Chem.* **270**:21433-21436

Cho, Y.S., and Lim, H.S. (2018) Comparison of various estimation methods for the parameters of Michaelis-Menten equation based on *in vitro* elimination kinetic simulation data. *Transl Clin Pharmacol.* **26:**39

Clavel, F. and Mammano, F. (2010) Roles of Gag in HIV resistance to protease inhibitors. *Viruses*. **2**:1411-1426

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

Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E. and Shivaprakash, M. (1995) In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature.* **374**:569-571

Couto, R.A.S., Chen, L., Kuss, S., and Compton, R.G. (2018) Detection of *Escherichia coli* bacteria by impact electrochemistry. *Analyst.***143**, 4840-4843

Cunningham, A.L., Donaghy, H., Harman, A.N., Kim, M. and Turville, S.G. (2010) Manipulation of dendritic cell function by viruses. *Current Opinion in Microbiology.* **13**(4):524-529

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

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

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

Du, X., Li, Y., Xia, YL., Ai, S.M., Liang, J., Sang, P., Ji, X.L., and Liu, S.Q. (2016) Insights into protein-ligand interactions: mechanisms, models, and methods. *Int J Mol Sci.* **17**(2):144

Dunitz, J.D. (1995) Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chemistry and Biology.* **2**:709-712

Eisenthal, R., Danson, M. J., and Hough, D.W. (2007) Catalytic efficiency and k cat/K M: a useful comparator?. *Trends in Biotechnology*. **25**:247-249.

Elbing, K., and Brent, R. (2020) Growth of *E.coli* in liquid medium. *Current protocols in molecular biology*.**125**(1):e81

Farnet, C.M. and Haseltine, W.A. (1991) Determination of viral proteins present in the human immunodeficiency virus tunnel preintegration complex. *Journal of Virology.* **68**(4):1910-1915

Fox, J.M., Zhao, M., Fink, M.J., Kang, K., and Whitesides, G.M. (2018) The molecular origin of enthalpy/entropy compensation in biomolecular recognition. *Annu. Rev. Biophys.* **47**:223-250

Freed, E.O. (1998) HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology.* **251**:1-15

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

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

Grossman, Z., Paxina, E.E., Averbuch, D., Maayan, S., Parkin, T.N., Engelhard, D., Lorber, M., and Schapiro, J.M. (2004) Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. *Antimicrob Agents Chemother.* **48**:2159-2165

Gustchina, A., and Weber, I.T. (1990) Comparison of inhibitor binding in HIV-1 protease and in nonviral aspartic proteases: the role of the flap. *FEBS Lett*. **269**:269-272

Hemelaar, J., Wlangovan, R., Dickson-Tetteh, L., Yun, J., Fleminger, I., Kirtley, S., Williams, B., Gouws-Williams, E., and Ghys, P. (2019) Global and regional molecular epidemiology of HIV-1, 1990-2015: a systemic review, global survey, and trend analysis. *The Lancet: infectious* *diseases.* **19**(2):143-155

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J.Mol. Biol.* **166**(4):557-580

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

Holdgate, G. A., and Ward, W. H. (2005) Measurements of binding thermodynamics in drug discovery. *Drug Discovery Today*. **10**:1543-1550

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

Kantor, R., and Katzenstein, D. (2004) Drug resistance in non-subtype B HIV-1. *J. Clin. Virol.* **29**:152-159.

Kelly, S.M., Thomas, J.J. and Prince, N.C. (2005) How to study protein by circular dichroism. *Biochim. Biophys. Acta.* **1751**:119-139

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

Kogure, T., Kawano, H., Abe, Y., and Miyawaki,A. (2008) Fluorescence imaging using a fluorescent protein with a large stokes shift. *Methods*. **45**(3):223-226

Kramer, R. A., Schaber, M. D., Skalka, A. M., Ganguly, K., Wong-Staal, F., And Reddy, E. P. (1986) HTLV-III *gag* protein is processed in yeast cells by the virus *pol* protein. *Science* **231**:1580-1585

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J. and Hendrickson, W.A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature.* **393**(6686):648–659.

Ledwaba, J., Sayed, Y., Pillay, V., Morris, L. and Hunt, G. (2019) Low frequency of protease inhibitor resistance in HIV-1 subtype C protease inhibitor naïve sequences. *AIDS Research and Retroviruses.* **35**:7

Ladbury, J.E. (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chemistry and Biology.* **3**:973-980

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

Lakowicz, J.R. (1999) *Principles of Fluorescence Spectroscopy 2nd Edition,* (Kluwer Academic ed.), pg 39, Plenum Publishers, New York

Lefebvre, E. and Schiffer, C.A. (2008) Resilience to resistance of HIV-1 protease inhibitors: profile of darunavir. *AIDS Reviews*. **10**(3):131-142

Lv,Z., Chu, Y., and Wang, Y. (2015) HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV/AIDS- research and palliative care*. **7**:95-104

MacRaild, C.A., Daranas, A.H., Bronowska, A., and Homans, S.W. (2007) Global changes in local protein dynamics reduce the entropic cost of carbohydrate binding in the arabinose-binding protein. *J. Mol. Biol.* **368**:822-832

Marasovi, M., Marasovi, T., and Miloš, M. (2007) Robust Nonlinear Regression in Enzyme Kinetic Parameters Estimation. *Journal of Chemistry.* **13**

Marshall, G.R., and Ballonte, F. (2017) Limiting assumptions in the design of peptidomimetics. *Drug Development Research.* **78**:245-267

Meitjes, G., Moorhouse, M.A., Carmona, S., Davies, N., van Vuuren, C., Manzini, T., Mathe, M., Moosa, Y., Nash, J., Nel, J., Pakade, Y., Woods, J., van Zyl, G., Conradie, F., and Venter, F. (2017) Adult antiretroviral therapy guidelines. *Southern African Journal of HIV medicine.* **18**(1):a776

Michaelis, L. and Menten, M.L. (1913) Die Kinetik der Invertinwerkung. *Biochemische Zeitschrift* .**49**:333

Mildner, A. M., Rothrock, D. J., Leone, J. W., Bannow, C. A., Lull, J. M., Reardon, I. M., Sarcich, J. L., Howe, W. J., Tomich, C. S., and Smith, C. W. (1994) The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. *Biochemistry.* **33**:9405-13

Mittal, S., Bandaranayaka, R.M., King, N.M., Prabu-Jeyabalan, M., Nalam, M.N., Nallivaika, E.A., Yilmaz, N.K., and Schiffer, C.A. (2013) Structural and thermodynamic basics of amprenavir/darunavir and atazanavir resistance in HIV-1 protease with mutations at residue 50. *Journal of virology*. **87**(8):4176

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

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

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

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

Nicholson, L. K., Yamazaki, T., Torchia D.A., Grzesiek, S., Bax A., Stahl, S.J., Kaufman, J.D., Wingfield, P.T., Lam, P.Y. and Jadhav, P.K. (1995) Flexibility and function in HIV-1 protease. *Nat. Struct. Biol*. **2**:274-280

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

Ohtaka, H., Velazquez-Campoy, A., and Xie, D. (2002) Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. *Protein Science*. **11**:1908-1916

Olsen, D.B., Stahlhut, M.W., Rutkowski, C.A., Schock, H.B., van Olden, A.L., and Kuo, L.C. (1999) Non-active site changes elicit broad-based cross-resistance of the HIV-1 protease to inhibitors. *J Biol Chem.* **274**:23699-23701

Osmanov, S., Pattou, C., Walker, N., Schwardlander, B., and Esparza J. (2002) Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *J Acquir Immune Defic Syndr.* **29**:184-190

Perozzo, R., Folkers, G., and Scapozza, L. (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *Journal of receptors and signal transduction.* **24**:1-2

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

Powell, M.K., Benkova, K., Selinger, P., Dogosi, M., Kinkorova Lunackova, I., Spurkova, P., Laclova, L., Eis, V., Sach, J. and Heneberg, P. (2016) Opportunistic infections in HIV- infected patients differ strongly in frequencies and spectra between patients with low $CD4^+$ cell counts examined post-mortem and compensated patients examined antemortem irrespective of the HAART era. *PLoS ONE*. **11**(9):e0162704

Prabu-Jeyabalan, M., Nalivaika, E. A., Romano, K., and Schiffer, C. A. (2006) Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *Journal of Virology.* **80**:3607-16

Reynolds, J.A., and Tanford, C. (1970) Binding of Dodecyl Sulfate to Proteins at High Binding Ratios. Possible Implications for the State of Proteins in Biological Membranes. *Proc. Natl. Acad. Sci. U S A.* **66**:1002-1007

Ritchie, R.J., and Prvan, T. (1996) Current statistical methods for estimating the Km and Vmax of Michaelis-Menten kinetics. *Biochemical Education,* **24**:196–206

Roberts, J.D., Bebenek, K., and Kunkel, T.A. (1988) The accuracy of reverse transcriptase from HIV-1. *Science*. **242**:1171-1173

Rose, R. B., Craik, C. S. and Stroud, R. M. (1998) Domain flexibility in retroviral proteases: structural implications for drug resistant mutations. *Biochemistry.* **37**:2607-21

Ruthven, C.R.J., Bernhard, T., Memorial, B. and Charlotte, Q. (1954) The separation of substances and estimation of their relative molecular sizes by the use of columns of starch in water. *Biochemical Journal.* **62**: 665-674

Seelmeier, S., Schmidt, H., Turk, V. and Von Der Helm, K. (1988) Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proceedings of the National Academy of Sciences.* **85**:6612-6616.

Segel, L.A. (1988) On the validity of the steady state assumption of enzyme kinetics. *Bulletin of Mathematical Biology*. **50:**579-593.

Schagger, H. and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate polyacrylamide gel electrophoresis for the separation of proteins in the range from 1-100 kDa. *Anal. Biochem.* **166**:368-379

Schroder, A.R., Shinn, P., Chon, H., Berry, C., Erkel, J.K., and Bushamn, F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell.* **110**:521-529

Sheehan, D. (2009) *Physical biochemistry: Principles and applications.* England (UK): John Wiley and Sons Ltd.

Shaw, G.M., and Hunter, E. (2012) HIV transmission. *Cold Spring Harb Prespect Med.* A006965

Showa, S.P., Nyabadza, F., and Hove-Musekwa, S.D. (2019) On the efficiency of HIV transmission: insights through discrete time HIV models. *PLoS ONE.* **14**(9):e0222574

Simon, V., Ho, D.D., and Karim, Q.A. (2006) HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. *Lancet*. **368**(9534):489-504

Statistics South Africa (Stats SA) (2021) STATISTICAL RELEASE P0302. Mid-year population estimates. http://www.statssa.gov.za/publications/P03 [accessed on 17/02/2022]

Studier, F.W., and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of clone genes. *J. Mol. Biol.* **189**:113-130

Su, C.T., Koh, D.W., and Gan, S.K. (2019) Reviewing HIV-1 Gag mutations in protease inhibitors resistance: insights for possible novel Gag inhibitor designs. *Molecules.* **24**:3243

Tatt, I. D., Barlow, K. L., Nicool, A. and Clewley, J. P. (2001) The public health significance of HIV-1 subtypes. *AIDS.* **15**:S59-S17

The Joint United Nations Program on HIV/AIDS (UNIAIDS) (2012) Global HIV & AIDS statistics: UNAIDS Fact Sheet November 2021 (Geneva, Switzerland: UNAIDS)

Titanji, B.K., Aasa-Chapman, M., Pillay, D., and Jolly, C. (2013) Protease inhibitors effectively block cell-to-cell spread of HIV-1 between T cells. *Retrovirology*. **10**:161

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

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

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

Velazquez-Campoy, A., Muzammil, S., Ohtaka, H., Shon, A., Vega, S., and Freire, E. (2003) Structural and thermodynamic basis of resistance to HIV-1 protease inhibition: implications for inhibitor design. *Curr. Drug Targets Infect. Disord.* **3**:311-328

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

Vella, S., Schwartländer, B., Sow, S.P., Eholie, S.P., and Murphy, R. L. (2012) The history of antiretroviral therapy and of its implementation in resource-limited areas of the world. *AIDS*. **26**:1231-1241

Volkmer, B., and Heinemann, M. (2011) Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS One*. **6**(7):e23126

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

Walker, P.R., Pybus, O.G., Rambaut, A., and Holmes, E.C. (2005) Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. *Infect. Genet. Evol.* **5**:199-208

Wang, Y., Liu, Z., Brunzelle, J.S., Kovari, I.A., Dewdney, T.G., Reiter, S.J., and Kovari, L.C. (2001) The higher barrier of Darunavir and Tipranavir reistance for HIV-1 protease. *Biochem Biophys. Res. Commun.* **412**(4):737-742

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

Whitmore, L. and Wallace, B.A. (2007) Protein secondary structure analyses from circular dichroism spectroscopy : methods and reference databases. *Biopolymers*. **89**:392-400

Williams, A., Basson, A., Achilonu, I., Dirr, H.W. and Sayed, Y. (2019) Double trouble? Gag in conjunction with double insert in HIV protease contributes to reduced DRV susceptibility. *Biochemical Journal.* **476**:375-384

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

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science.* **245**:616-21

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

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

Yang, H., Nkeze, J., and Zhao, R.Y. (2012) Effects of HIV-1 protease on cellular functions and their potential applications in antiretroviral therapy. *Cell and Bioscience.* **2**:32

Zhang, H., Orti, G., and Du, G. (2002) Phylogenetic and phenotypic analysis of HIV type 1 env, gp120 in cases of subtype c mother-to-child transmission. *AIDS Red. Hum. Retrovirus.* **18**:1415-1423

Zhengtong, L.V., Yuan, C. and Wang, Y. (2015) HIV protease inhibitors: a review of molecular selectivity and toxicity. *Dovepress.* **7**:95-104