Impact of L38↑N↑L Insertions on Structure and Function of HIV-1 South African Subtype C Protease

Xolisiwe Maputsoe

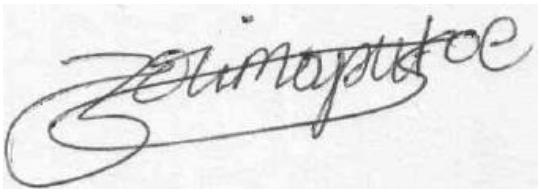
A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science.

Johannesburg, 2012

Declaration

I hereby declare that this dissertation is my own unaided work. It is submitted for the degree of Master of Science, at the University of the Witwatersrand, Johannesburg. It has not been submitted for any other degree or examination at another university.

Xolisiwe Maputsoe



18th Day of May , 2012

Abstract

The Human Immunodeficiency Virus (HIV) subtype C accounts for the majority of infections in Southern Africa. The HIV protease is one of the targets in HIV treatment due to its pivotal role in HIV maturation in the host cell. However, because of polymorphisms in the HIV genome, drug resistance becomes a major problem in HIV treatment. Polymorphisms in the HIV protease gene result in altered substrate cavities, and /or flap hinge modifications leading to unfavourable drug interaction with the enzyme. The most common form of drug resistant mutations is single amino acid substitutions. Although, amino acid insertions have been reported, this form of mutation in the HIV protease is rare. L38↑N↑L insertion is a unique form of HIV protease polymorphism that was isolated from a patient failing drug therapy in South Africa. The objective of this research was to assess the impact of the L38↑N↑L insertions, with accompanying background mutations, on the structure and function of this form of polymorphism in HIV-1 South African subtype C protease. The far-UV circular dichroism (CD) spectra of L38↑N↑L protease shows a trough at 203 nm, suggesting alterations in the secondary structure content of this mutant. Whereas the wild type (WTCSA-HIVPR) displays a trough at 215 nm. However, tertiary structure characterisation using fluorescence spectroscopy did not detect changes within the local tryptophan environment of L38↑N↑L protease in comparison with the wild type due to no significant shift in emission wavelength. The specific activity of L38↑N↑L protease and wild type was 28.0±1.3 $\mu mol.min^{-1}.mg^{-1}$ and 123.45±6.4 $\mu mol.min^{-1}.mg^{-1}$ respectively. The turn-over number for L38↑N↑L protease and wild type was $1.0 \times 10^{-3} \pm 6.0 \times 10^{-5}$ and $7.7 \times 10^{-3} \pm 5.6 \times 10^{-4}$ respectively. As much as the presence of known drug resistance mutations in L38↑N↑L can be attributed to drug resistance, it should also be noted that the insertions may have also caused local structural alterations that may have enhance drug resistance of L38↑N↑L. These changes could have lead to the decreased catalytic activity of the L38↑N↑L protease. Homology modelling studies show that the insertions in L38↑N↑L protease may have resulted in a fold similar to 2HS1 (PDB code), which has a modification on the flap hinge. In addition, the homology modelling studies suggest that L38↑N↑L protease may have a second inhibitor binding site next to one of the flap hinge regions as seen in the 2HS1 model. In conclusion, the L38↑N↑L insertions and accompanying background mutations may have contributed to the local structural modifications that lead to drug resistance in L38↑N↑L protease.

This work is dedicated to those who have always been a source of encouragement:

My mother, Khohlisa, for the sacrifices and endless support.

My brothers, Moponya and Bongani

To Tshele, for believing in me

In memory of my grandparents, James and Julia Dingizwayo

"Our deepest fear is not that we are inadequate. Our deepest fear is that we are powerful beyond measure. It is our light, not our darkness that most frightens us. We ask ourselves, Who am I to be brilliant, gorgeous, talented, fabulous? Actually, who are you not *to be? You are a child of God. Your playing small does not serve the world. There is nothing enlightened about shrinking so that other people won't feel insecure around you. We are all meant to shine, as children do. We were born to make manifest the glory of God that is within us. It's not just in some of us; it's in everyone. And as we let our own light shine, we unconsciously give other people permission to do the same. As we are liberated from our own fear, our presence automatically liberates others."*

Marianne Williamson

Acknowledgements

I acknowledge my supervisor Dr Yasien Sayed for his guidance and patience throughout.

I also acknowledge my co-supervisor Dr Ike Achilonu, for his advice and always going the extra mile to assist.

Prof. Heini Dirr for his support and the pleasure of allowing me to work in his laboratory.

Prof. Lynne Morris from the AIDS research unit at the NICD for providing me with the sequence of the variant protease used in this study.

Members of the Protein Structure-Function Research Unit. Special thanks are due to Nishal Parbhoo.

Special thanks goes to the following people: my cousin Leutla Lephatswane, my friends, Mpho Choene, Thandeka Khoza, Maabo Moralo, Lerato Mpye, Nomxolisi Ngubane, Obakeng Ntshudisane, Palesa Seele.

I acknowledge the National Research Foundation for funding.

TABLE of CONTENTS

List of figures

List of tables

List of abbreviations

Å	Angstrom
AIDS	Acquired Immunodeficiency Virus
ARV	Antiretroviral
CD	Circular Dichroism
C-terminus	Carboxyl terminal
DEAE	Diethylaminoethyl
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FDA	Food and Drug Administration
g	Gravitational force
Gag	Group antigens
HIV	Human Immunodeficiency Virus
IN	Integrase
IPTG	Isopropyl β-D thiogalactoside
k_{cat}	Turnover number
k_{cat}/ K_M	Catalytic efficiency
K_M	Michaelis constant
LB	Lysogeny Broth
LTR	Long Terminal Repeat
NOP	Naturally Occurring Polymorphism
N-terminu	Amino terminal
PCR	Polymerase Chain Reaction
PI	Protease Inhibitor
PMSF	phenylmethylsulfonylfluoride

PR	Protease
RMSD	Root Mean Square Deviation
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Reverse Transcriptase
SD	Standard deviation
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TEMED	Tetramethylethylenediamine
U	Unit
V_{max}	Maximum velocity

The IUPAC-IUBMB one and three letter code for naming amino acids was used throughout.

CHAPTER 1

1. Introduction

1.1. HIV/AIDS epidemic

The Human Immunodeficiency Virus (HIV), the causative agent of acquired immunodeficiency syndrome (AIDS) (Weiss *et al.*, 1985; Gallo and Montagnier, 1988) remains one of the major causes of death worldwide, particularly in sub-Saharan Africa, with an estimated 1.8 million deaths in 2010 (UNAIDS, 2010; WHO, 2011). According to the 2010 UNAIDS global report, sub-Saharan Africa is home to 68% of the world's total population living with HIV (UNAIDS, 2010). According to this report, South Africa is one of the African countries most affected by the epidemic, with a reported 17 – 18% incidence rate (UNAIDS, 2010; SANAC, 2011; WHO, 2011).

HIV has a high genetic diversity. For this reason, it is classified according to types, groups, subtypes and recombinant forms, based on phylogenetic and genetic distance analysis (Charneau *et al.*, 1994; Subbarao and Schochetman, 1996; Simon *et al.*, 1998; Roques *et al.*, 1999; Robertson *et al.*, 2000; Leitner *et al.*, 2005). HIV is divided into two distinct groups, HIV-1 and HIV-2. HIV-1 is mainly responsible for the worldwide epidemic, whereas HIV-2 exists in West Africa (Clavel *et al.*, 1986). It is proposed that HIV-2 is restricted to West Africa because it does not spread as easily as HIV-1, and infected individuals acquire immunodeficiency slower than those infected with HIV-1 (Reeves and Doms, 2002; Parkin and Schapiro, 2004).

HIV-1 is divided into nine subtypes; A-D, F-H, J and K whereas HIV-2 is divided into groups A to G (Gao *et al.*, 1994; Chen *et al.*, 1997; Yamaguchi *et al.*, 2000). In terms of global distribution of HIV-1 subtypes, sub-Saharan Africa shows the greatest diversity, with subtypes A and C predominating. Subtype A is found in the northern part of sub-Saharan Africa and subtype C in southern Africa as shown in Figure 1. Subtype B is the main subtype in developed countries such as North America, Western Europe and Australia (Montano *et al.*, 1997; Essex, 1999). HIV-1 subtypes can be further divided into sub-subtypes; circulating and the unique recombinant forms (Brodine *et al.*, 1995; Fleury *et al.*, 2003; Kantor and Katzenstein, 2004). These variant forms differ in phenotype and genotype and are referred to as naturally occurring polymorphisms (NOPs) (Coman *et al.*, 2008b).

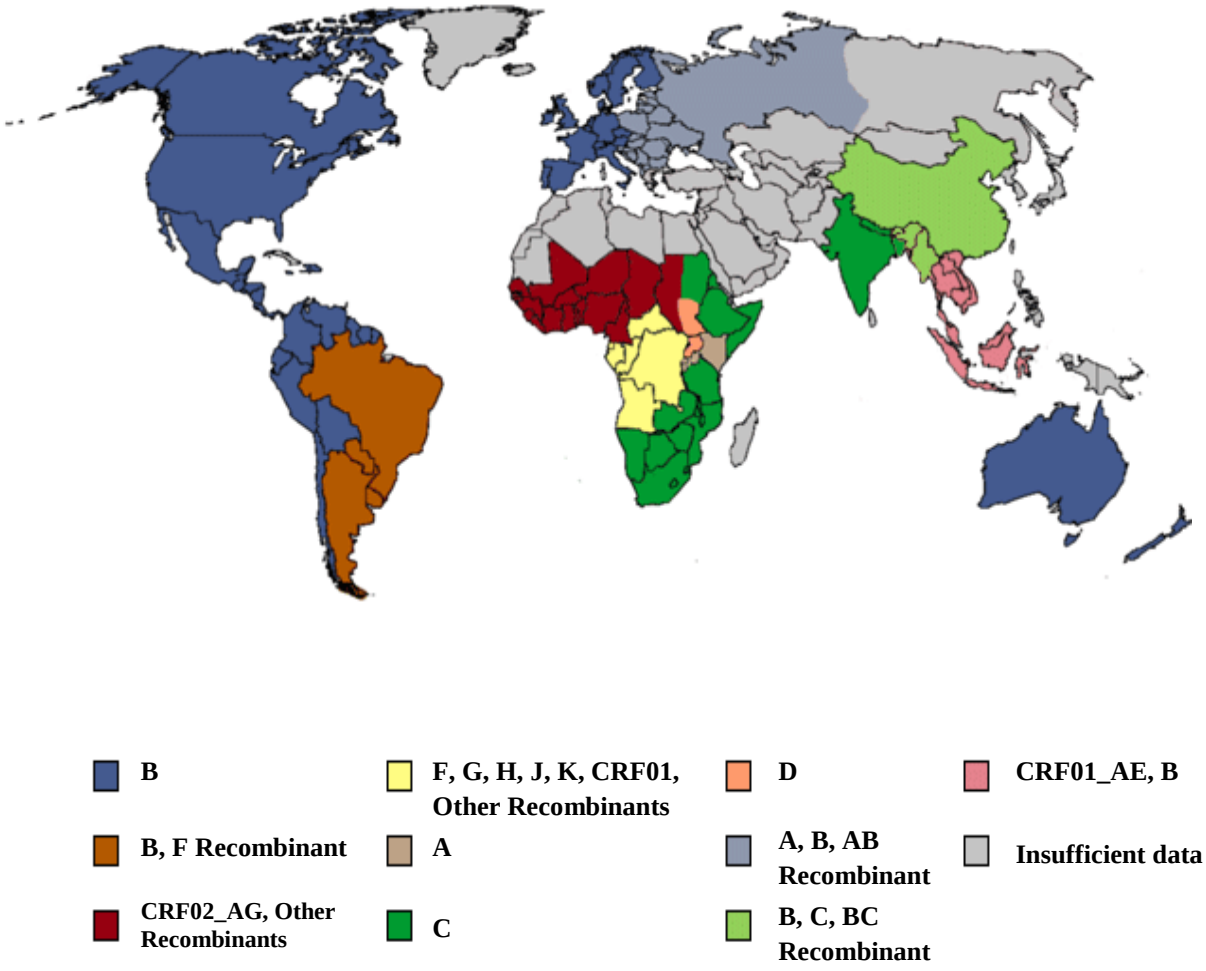


Figure 1 Geographical distribution of HIV-1 subtypes.

Subtype B is found in North America, Western Europe and Australia. Sub-Saharan Africa shows the greatest diversity; with subtype C predominating in the South and East. Subtypes A and D are also present, but with lower frequencies. The CRF02_AG and other recombinant forms exist in West and West central Africa, although at a low frequency. In central Africa there is an array of rare subtypes and recombinant forms with no predominating strain (http://www.pbs.org/wgbh/pages/frontline/aids/atlas/clade.html accessed 20 October 2011).

One of the biggest challenges with HIV-1 has been the radical emergence and spread of HIV -1 C viruses, which are responsible for 50% of global infections (Esparza and Bhamarapravati, 2000; Kantor and Katzenstein, 2004; Hemelaar *et al.*, 2006). It is proposed that the increase in the number of subtype C infections could be a result of the ease with which subtype C viruses are transmitted. The cellular transcription factor, NFkβ, is responsible for viral transcription up-regulation (Baeuerle, 1991). Apparently, the presence of an extra NFkβ binding site in the long terminal repeat (LTR), found in the subtype C may enhance gene expression, which in turn alters transmission and pathogenesis (Tatt *et al.*, 2001). Another reason could be that subtype C viruses have an enhanced catalytic activity compared to other subtypes as shown by the Freire group (Velazquez-Campoy *et al.*, 2001c).

1.2. HIV life-cycle

HIV infects lymphocytes and macrophages due to the specific interactions it makes with the $CD4^+$ receptor, when entering these cells (Maddon *et al.*, 1986; Ashorn *et al.*, 1990). An outline of the HIV life-cycle is depicted in Figure 2. The virus fuses with the host's plasma membrane. Once it is taken up by the host, the virus becomes partially uncoated, allowing reverse transcription to take place (Farnet and Haseltine, 1991), resulting in production of double-stranded DNA from single-stranded RNA. It is suggested that both linear and circular DNA are produced from this process, but only the linear DNA is capable of being integrated (Brown *et al.*, 1987). The linear viral DNA is transported to the nucleus as part of a protein complex involving matrix protein and integrase (Burkrinsky *et al.*, 1992). The integrase enzyme facilitates the incorporation of the viral DNA with the host cell DNA. Integration of the viral DNA is not site-specific (Ratner, 1993).

The viral DNA is transcribed along with the host cell DNA. The resulting viral RNAs are spliced and moved from the nucleus to the cytoplasm, in order for translation to take place. These RNA molecules encode different structural and functional precursor proteins of the virus (Nutt *et al.*, 1988). The Gag and Gag-Pol precursor proteins, along with two RNA molecules are transported to the plasma membrane where they are assembled, thus forming an immature viral particle (Beschiaschvili and Baeuerle, 1991). Viral assembly occurs at the plasma membrane of the host cell. The Gag and Gag-Pol products migrate towards the membrane and are anchored to the cytoplasmic side by a covalent N-terminal-linked myristic acid (Henderson *et al.*, 1983; Mervis *et al.*, 1988).

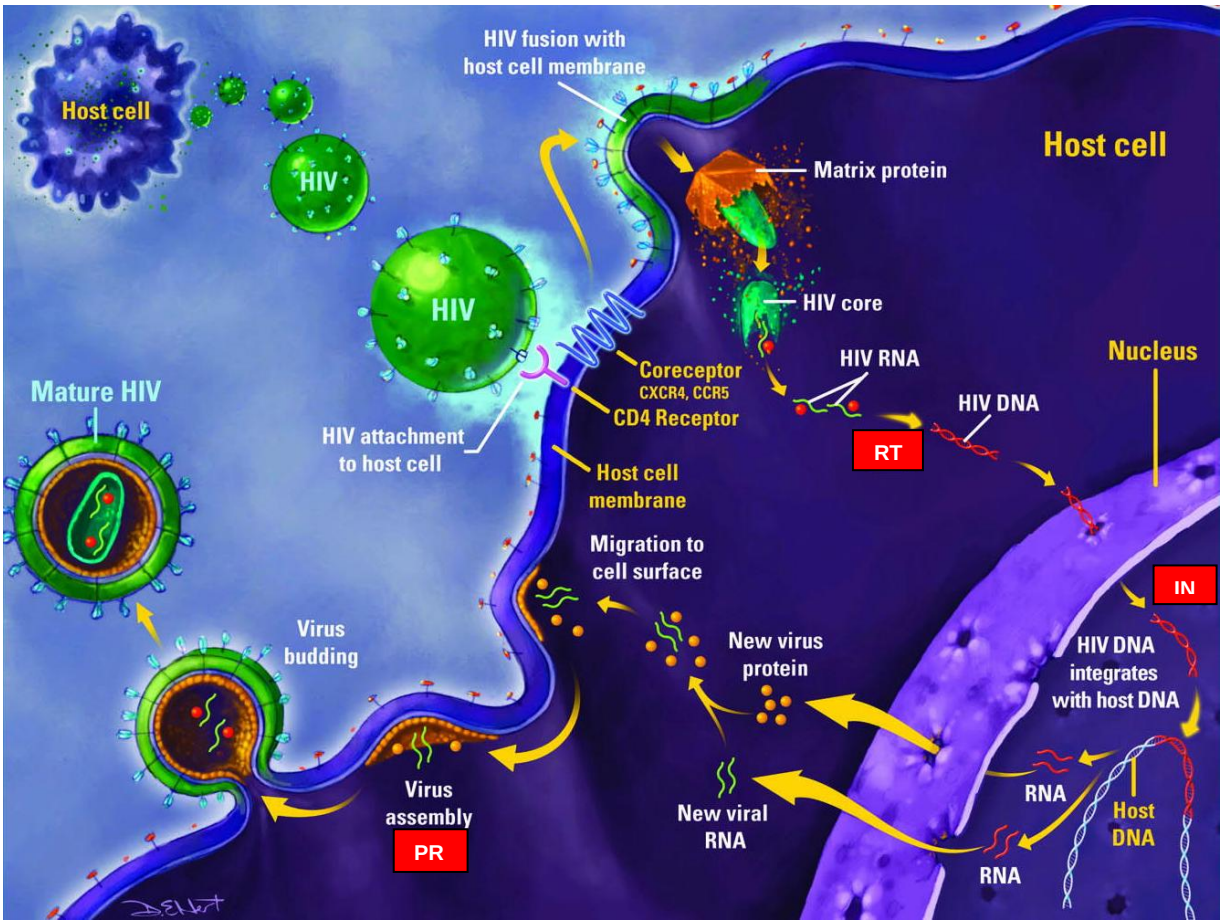


Figure 2 The life cycle of HIV.

The HIV life-cycle begins with the virus attaching to the host cell via the $CD4^+$ receptor. The virus then fuses with the membrane becoming partially uncoated, releasing important structural and functional components. Reverse transcription of viral RNA takes place. New viral DNA is transported to the nucleus where it is integrated with host DNA. After DNA transcription, the new viral RNA is transported out of the nucleus where it is transported to the plasma membrane along with viral protein. Viral assembly takes place at the plasma membrane and an immature virion is formed. The virus then buds of the host cell and is ready to infect new cells. Important functional proteins in the HIV life cycle are shown in the red boxes (taken from www.HIVwebstudy.org accessed 19 July 2011).

Electron micrographs showing the steps involved in viral assembly suggest that the polyproteins aggregate, followed by the membrane bulging outward. The nascent viral particles then bud from the plasma membrane (Gonda *et al.*, 1985). It is postulated that viral particles mature after budding has taken place and HIV protease may have a role in controlling the maturation of the virus. This is based on the fact that dimerisation of the HIV protease monomers takes place before proteolysis of the polyproteins (Navia *et al.*, 1989). Secondly, the protease assumes a soluble form after cleavage from the Gag-Pol polyprotein. Once soluble, the protease diffuses away from the plasma membrane, preventing premature initiation of viral maturation. Due to the major role it plays in the life-cycle of HIV, the protease is one of the major drug targets in anti-HIV therapy (De Clercq, 2007).

1.3. HIV protease

HIV protease is a member of the aspartyl protease family on the basis of the conserved Asp-Thr-Gly active site sequence (Toh *et al.*, 1985). In addition, HIV protease is inhibited by pepstatin, which is a common aspartyl protease inhibitor (Hansen *et al.*, 1988; Seelmeier *et al.*, 1988; Darke *et al.*, 1989). It is the only member of the aspartyl protease family that is dimeric. HIV protease catalyses its release from the viral Gag-Pol precursor through cleavage at the amino and carboxyl - terminus (Debouck *et al.*, 1987; Farmerie *et al.*, 1987; Giam and Boros, 1988). It is also responsible for release of the poly-protein precursor products (Witte and Baltimore, 1978; Crawford and Goff, 1985; Katoh *et al.*, 1985; Farmerie et al., 1987; Darke *et al.*, 1988). HIV protease plays an essential role in viral maturation, as shown by inactivation through mutagenesis (Asp 25 to Asn, Thr or Ala) or chemical inhibition (pepstatin) (Hansen et al., 1988; Kohl *et al.*, 1988; Mous *et al.*, 1988; Seelmeier et al., 1988; Darke et al., 1989; McQuade *et al.*, 1990). For this reason, HIV protease is one of the major targets in anti-HIV therapy (De Clercq, 2007).

There are 25 antiretroviral (ARV) drugs approved for the treatment of HIV (Alfonso and Monzote, 2011). These drugs include reverse transcriptase inhibitors (nucleotide, nucleoside and non-nucleoside inhibitors), integrase inhibitors, HIV cell entry inhibitors (fusion inhibitors, and co-receptor antagonists) and protease inhibitors (De Clercq, 2009). These inhibitors are used in combination and complement each other in HIV drug therapy (Kantor and Katzenstein, 2004; De Clercq, 2009) .

1.3.1. Structure of HIV-1 protease

Protein crystallography is one of the important techniques used in elucidating the structure of proteins and HIV protease. Interactions displayed in the X-ray crystal structures of protease are exploited for the design of antiretroviral drugs, improving specificity, and the potency of protease inhibitors (Kantor and Katzenstein, 2004). HIV protease is an obligate homodimer, with each monomer possessing 99 amino acids. The homodimeric protease structure consists of nine β strands, and two α helices (Figure 3), and these structural characteristics are conserved in all HIV proteases. The two subunits, interact by means of a two-fold axis of symmetry (Navia *et al*., 1989), and are stabilised by anti-parallel interactions between the catalytic residues; Asp 25, Thr 26, Gly 27 in one subunit, and Asp 25´, Thr 26´, Gly 27´ in the other (Navia *et al.*, 1989). The protease structure has distinct regions namely; the flaps, the flap elbow (flap hinge), the active site, and the dimer interface, as shown in Figure 3.

1.3.2. The flap region

The protease flaps are dynamic moieties that result from the folding of two anti-parallel β strands (Miller *et al.*, 1989). The flaps have a high glycine content which allows for flexibility of the protease in this region (Lapatto *et al.*, 1989; Navia *et al*., 1989; Wlodawer *et al.*, 1989). The protease flaps play an essential role in the activity of the protease by controlling entry of substrate or inhibitor to the active site (Navia *et al*., 1989). The flaps are also involved in substrate recognition (Prabu-Jeyabalan *et al.*, 2006). When the active site is free of substrate or inhibitor, the flaps are flexible but still maintain reasonable coverage of the active site in an effort to keep the hydrophobic residues buried. This conformation is referred to as the "semi-open" form (Hornak *et al.*, 2006b). It has also been, suggested that the flaps are rigid and that flexibility is confined to the flap tips, and the flap elbows act as hinges (Freedberg *et al.*, 2000; Hornak *et al*., 2006b). The flap and the flap elbow work as a single entity. Therefore, any movements or changes that occur in the elbow region are conveyed to the flap tip (Clemente *et al.*, 2004; Perryman *et al.*, 2006; Coman *et al*., 2008b).

1.3.3. The active site

The protease active site is a symmetrical cavity that is located below the flaps (Navia *et al*., 1989), as viewed in Figure 3. Each protease monomer contributes a conserved amino acid

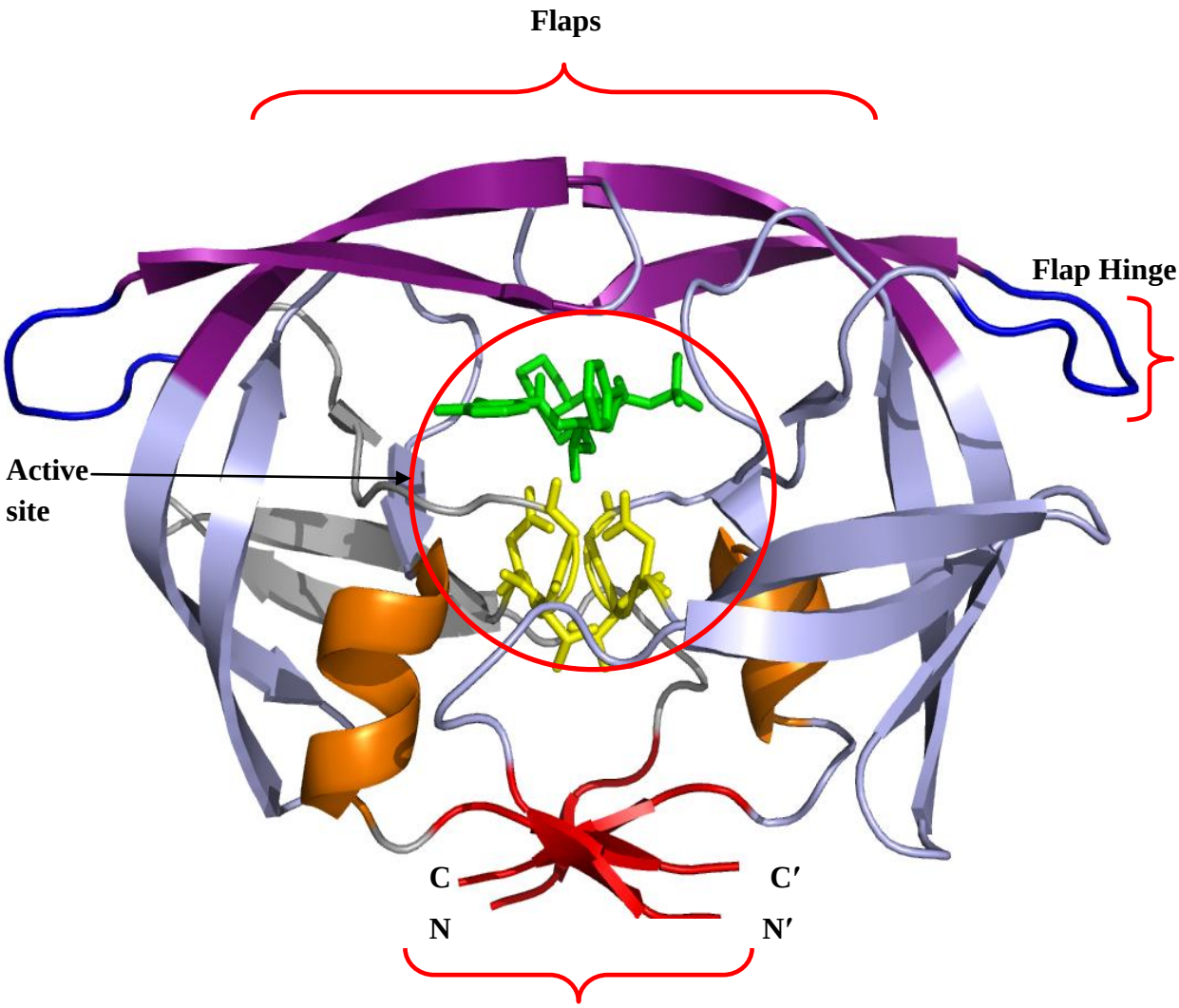


Figure 3 Ribbon representation of HIV-1 subtype C protease with bound Nelfinavir.

The structure of HIV protease consists mainly of β sheets and has two α helices (shown in orange). The flap region is shown in deep purple and consists of residues 44 to 57 from each monomer. Flap tips 49-53. Flap movements are controlled by the flap hinge / elbow (shown in dark blue) made up of residues 37 to 43. The protease inhibitor Nelfinavir (shown in green) is bound to the active site. Each monomer contributes one catalytic Asp 25 residue (shown in yellow). The dimer interface (shown in red) lies below the active site is made up of the C- and N -termini of each monomer. It is made up of residues 1 to 4; and 69 to 99. This Figure was generated using PyMOL with PDB code 2R5Q (DeLano, 2002).

sequence known as the catalytic triad (Asp 25 –Thr 26 – Gly 27) to the active site (Navia *et al*., 1989). The two aspartates are in close proximity, in a coplanar orientation, with their carboxylate oxygens hydrogen bonded to the amide hydrogens of Gly 27 and Gly 27′. At the same time, the two active site threonine residues are hydrogen bonded to main chain amide and carboxyl groups from the other chain to form a hydrogen network known as the 'fireman's grip'(James and Sielecki, 1983).

1.3.4. The dimer interface

Dimerisation of the two protease monomers is essential for catalysis because both substrate binding and the active site require interactions from both monomers (Wlodawer *et al.*, 1989). The majority of forces stabilising the dimer interface are formed by the interactions between four anti-parallel β strands (residues 1 – 4; and 96 – 99). This region consists of both the N- and C-termini of each subunit (Todd *et al.*, 1998). The flap region also has minor interactions between residues 48 and 54, from each monomer. Stability studies have shown that the dimer interface is important for the stabilisation of the global structure of HIV protease. Upon disruption of the interface, the structure collapses into two unstable monomers. This suggests that stabilisation of the monomer is not due to intrinsic forces but rather the interaction with each other to form the dimer (Todd *et al*., 1998).

1.4. The catalytic mechanism of HIV-1 protease

Substrate or inhibitor binding causes the HIV protease flaps to assume a less dynamic conformation. The flaps are pulled in towards the bottom of the active site, close to the catalytic residues (Hornak *et al*., 2006b). Hydrogen bonds between side chains also keep the flaps in close contact. An important result of flap closure is the exclusion of water, providing the protease with a hydrophobic environment essential for catalysis (Velazquez-Campoy *et al.*, 2003a). The flaps coordinate a catalytically important water molecule while in the closed conformation. This process helps position the substrate for catalysis (Baca and Kent, 1993).

The peptide bond that is cleaved by HIV protease is termed the "scissile" bond and it is found between P1 and P1′ according to the Schechter and Berger convention (Schechter and Berger, 1967), where "P" refers to peptide. The flanking amino acids towards the amino - terminus are termed P1, P2, P3, and P4. Similarly, amino acids located towards the carboxyl - terminus are named P1′, P2′, P3′, and P4′. The corresponding sub-sites on the protease are named from

the central aspartates S1, S2, S3, S4, S1′, S2′, S3′, S4′ (Schechter and Berger, 1967). This is depicted in Figure 4.

The two catalytic Asp 25 residues from each monomer have a water molecule between them and they exist in opposite states of protonation (Hyland *et al.*, 1991a; Hyland *et al.*, 1991b; Ido *et al.*, 1991; Das *et al.*, 2010).The pK_a of these aspartates are lower than those existing in aqueous environments. Substrate binding increases the pK_a from approximately 3.3 to 5.1 (Ido *et al*., 1991). The catalytic mechanism of HIV protease, as shown in Figure 5, follows general acid-base catalysis. In the free native enzyme, represented as EH^- in Figure 5, the catalytic aspartates share a proton as well as a water molecule. Since Asp 25 and Asp 25′ are identical, the negative charge is assigned to the water oxygen. Upon binding of substrate, an enzyme-substrate complex (EH^-•S) is formed. The negative charge on the water causes the water molecule to conduct a nucleophilic attack on the substrate carbonyl carbon. At the same time, the carbonyl oxygen interacts with hydrogen 1 (H_1). This leads to the formation of an unstable tetrahedral oxyanion intermediate in the transition-state (TS). Collapse of the oxyanion intermediate is caused by peptide bond breakage through the transfer of the water oxygen, and hydrogen 3 (H_3) to the peptide nitrogen to give an enzyme product complex (E•P), as shown in Figure 5. Release of products leads to the re-formation of EH^- (Ido *et al*., 1991; Tomasselli and Heinrikson, 2000; Dunn, 2002).

1.5. Protease inhibitors (PIs)

The development of HIV PIs is one of the most successful examples of structure-based drug design. The three-dimensional structure of HIV protease, interactions of the protease flaps with the natural substrate, as well as the catalytic mechanism were central in the structure-based drug design of PIs. The primary sequence of the natural substrates was also taken into account (Wlodawer and Erickson, 1993). The main role of PIs is to prevent the protease from cleaving the Gag-Pol polyprotein into essential products. Using an analogy similar to that of the lock and key, inhibitor binding acts as a key, ‘locking’ the enzyme, thus rendering it inactive (Sawant *et al.*, 2008). The resulting immature virions are cleared from the cell by a mechanism that is still not well understood. It is important to note that PIs prevent maturation of viral particles but cannot influence cells that are already infected. In essence, they act in the prevention of the spread of new infections (Robins and Plattner, 1993; Leonard, 1996; Sawant *et al*., 2008).

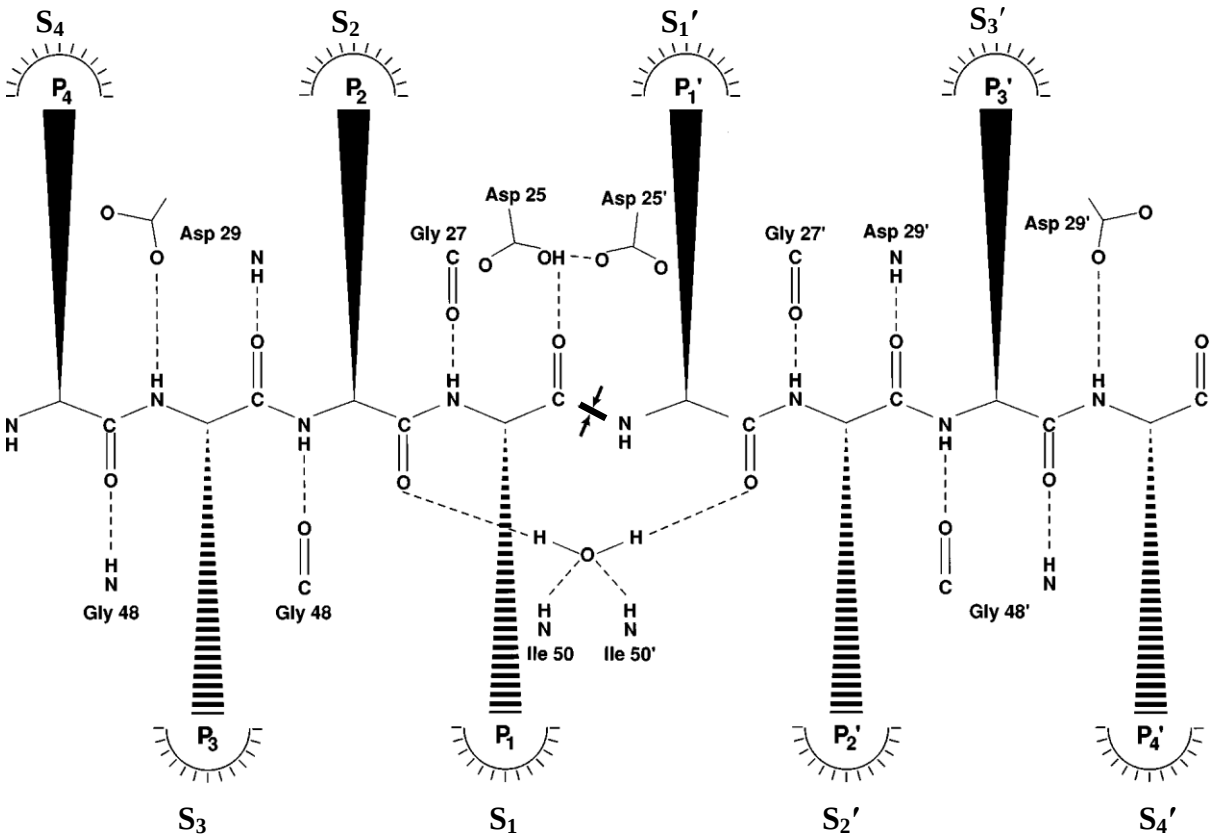


Figure 4 Diagram showing the Schechter and Berger conventional nomenclature.

The amino acids towards the amino - terminus are named P1, P2, P3, and P4. Similarly, substrate amino acids towards the carboxyl - terminus are named P1′, P2′, P3′, and P4′. The corresponding sub-sites on the protein are named from the central aspartates S1, S2, S3, S4, and S1′, S2′, S3′, S4′. (Adapted from Wlodawer and Vondrasek, 1998)

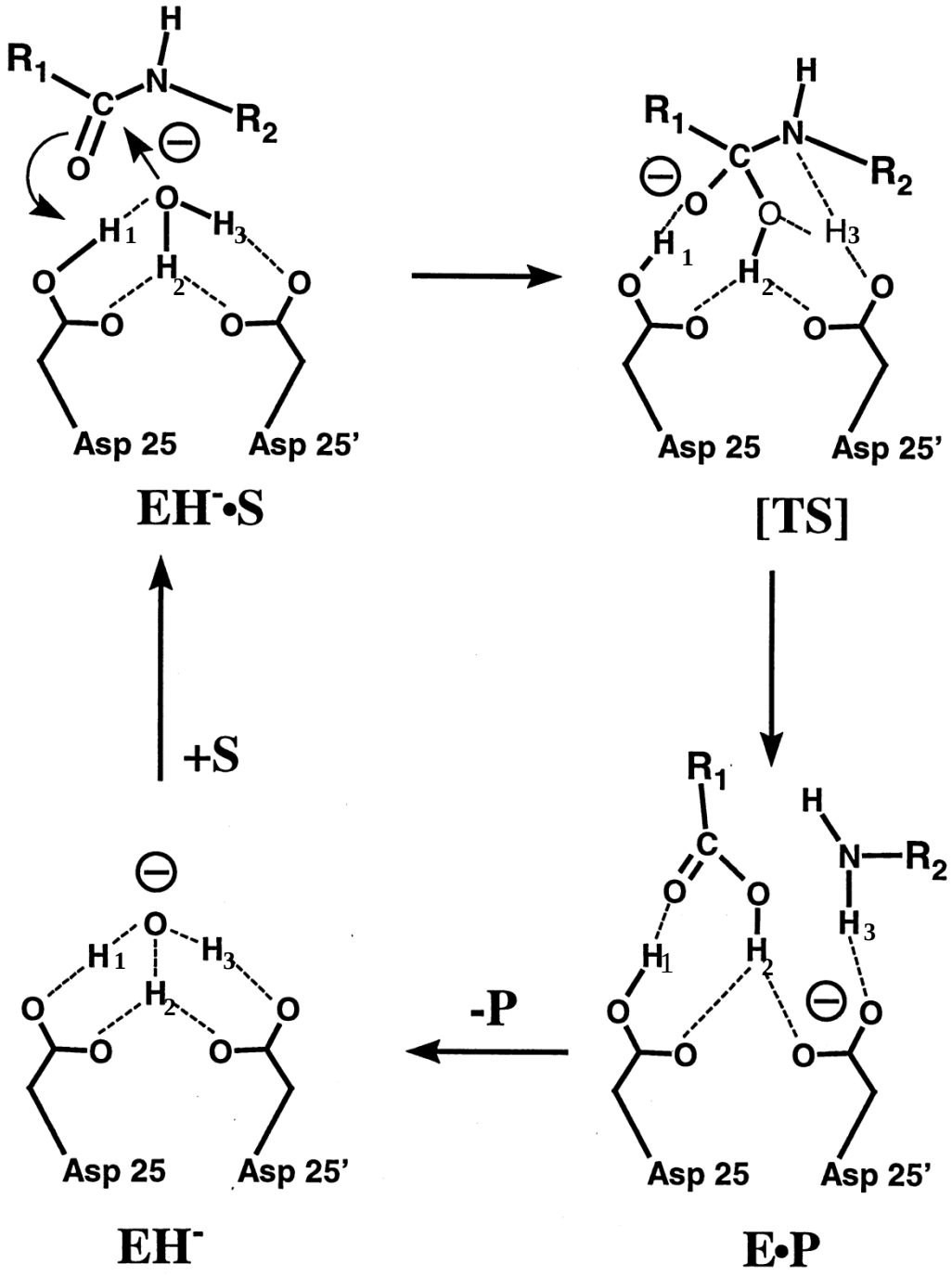


Figure 5 A schematic representation of the catalytic mechanism of the HIV protease.

The Asp residues from each monomer form hydrogen bonds with the conserved water molecule. The conserved water molecule becomes activated in order to carry out a nucleophilic attack on the carbonyl group of the peptide scissile bond producing an unstable tetrahedral substrate that collapses to give hydrolysed products and free enzyme (Taken from Tomasselli and Heinrikson, 2000).

Three generations of PIs have been developed in an effort to improve efficacy, longevity, and reduce side effects of infected individuals (Piacenti, 2006). To date there are ten PIs approved by the USA, Food and Drug Administration (FDA). These drugs are: Saquinavir, Ritonavir, Indinavir, Nelfinavir, Amprenavir, Lopinavir, Fosamprenavir, Atazanavir, Tipranavir and Darunavir (De Clercq, 2009). The first five PIs are classified as first generation inhibitors. Lopinavir, fosamprenavir and atazanavir are second generation inhibitors. Tipranavir and darunavir are third generation drugs. Second and third generation inhibitors, also known as the adaptive inhibitors were developed due to the decrease in efficacy of first generation drugs. Mutations found within the protease cavity cause changes in geometry of the active site, leading to a decrease in the number of hydrogen bonds, van der Waals interactions as well as other favourable interactions. Adaptive inhibitors have flexible moieties that allow better adaptation to mutations (Todd *et al.*, 2000; Velazquez-Campoy *et al.*, 2001a; Vega *et al.*, 2004; Ohtaka and Freire, 2005).

The interaction between HIV protease and PIs is mainly hydrophobic, which is similar to the interaction it has with the natural substrates. Structurally, HIV PIs can be classified as peptidomimetic and non-peptidomimetic. A structural characteristic of peptidomimetics is that they contain a non-hydrolysable transition-state mimic. The natural cleavage site is replaced by a transition-state isostere, such as hydroxyethylene. All the above mentioned PIs are peptidomimetics except for tipranavir, which is based on a coumarin scaffold (Navia *et al*., 1989; De Clercq, 2009).

Dimeric HIV protease is symmetrical, and so the design of most PIs makes use of this characteristic. Symmetrical inhibitors exhibit tighter binding, however, they are more prone to viral resistance resulting from amino acid changes that affect binding. This is because a single amino acid mutation can have a great impact on inhibitor binding (Sawant *et al*., 2008). The chemical structures of three PIs that were available in our laboratory while this study was conducted are shown in Figure 6. Important structural elements for each inhibitor are shown.

1.5.1 Saquinavir

Saquinavir or Invirase® is a PI designed by Hoffman-La Roche (Basel, Switzerland). It was the first PI to undergo clinical trials. It is a pentapeptide analogue. Characteristic features of Saquinavir are 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 (Roberts *et al.*, 1990). These structural elements are shown in Figure 6.

1.5.2 Ritonavir

Ritonavir or Norvir® was developed by Abbott Laboratories (Illinois, USA). It is derived from a C2-symmetric, peptidomimetic inhibitor. It is a hydroxyethyl transition-state isostere. It has two thiazole termini, and hydrophobic Phe and Val substituents, increasing its oral bioavailability relative to that of Saquinavir (Ho *et al.*, 1994). It is described as a potent inhibitor, without killing normal cells, as seen when tested against a variety of laboratory strains and clinical isolates (Kuroda *et al.*, 1995). Ritonavir has been shown to inhibit cytochrome P450–3A. It is, therefore used to inhibit the rapid metabolism of other protease inhibitors such as Saquinavir, delaying the development of resistance (Kempf *et al.*, 1997).

1.5.3 Nelfinavir

Nelfinavir or Viracept® was developed by Agouron Pharmaceuticals (La Jolla, USA). It has the same bulky hydrophobic group and hydroxyethylamine transition-state isostere as Saquinavir, except it has an extended P1 substituent making it a lipophilic inhibitor.

Despite the wide range of clinically used PIs there is still an increase in the number of resistance associated infections. This is attributed to a decrease in PI efficacy as a result of viral strains with specific amino acid mutations that lower their affinity for inhibitors, while maintaining normal function (Velazquez-Campoy *et al.*, 2003a).

1.6. Mutations

The emergence of multi-drug resistant HIV strains is caused by a number of reasons. To begin with, it is because of the high replication rate of HIV, as well as the low fidelity and lack of proof-reading abilities of reverse transcriptase (Roberts *et al.*, 1988; Coffin, 1995). Drug pressure, poor patient compliance to ARV treatment and cross infection between patients also gives rise to new viral strains which encode mutated reverse transcriptase and protease enzymes (Richman *et al.*, 1994; Coffin, 1995; Ho *et al.*, 1995; Robertson *et al.*, 1995a; Robertson *et al.*, 1995b; Wei *et al.*, 1995).

There are specific mutations (single or group point mutations, which are essentially amino

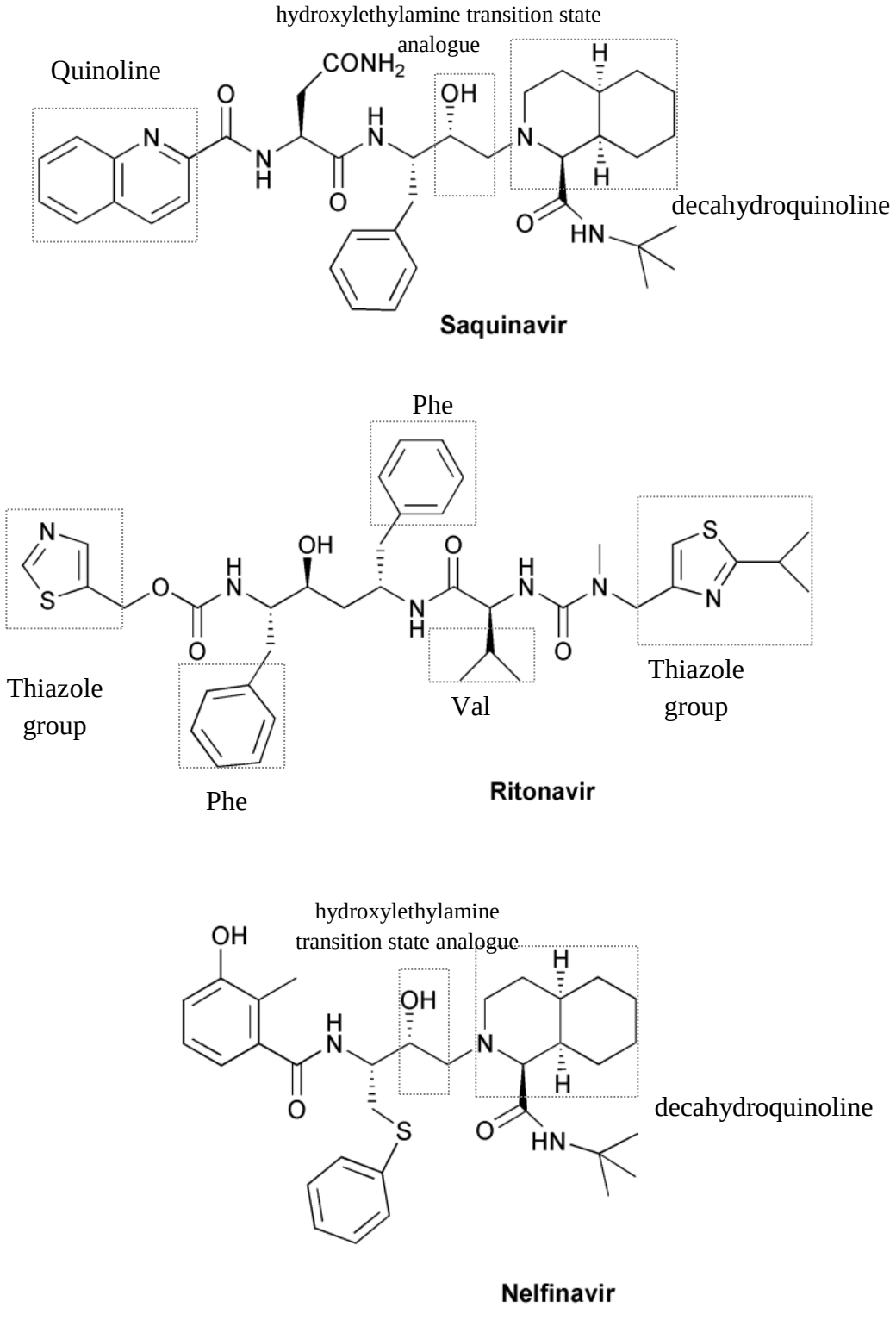


Figure 6 The chemical structures of three FDA approved protease inhibitors.

PIs are generally bulky and highly hydrophobic. The PIs shown above are first generation inhibitors. A common characteristic is that they are peptidomimetics.

acid substitutions) that occur in response to specific drugs. For example, Saquinavir resistance is due to G48V and I84V; Indinavir and Ritonavir V82A/T/F/S and I84V; Nelfinavir by D30N and I84V (Velazquez-Campoy *et al*., 2003a). Drug resistance mutations are classified as either primary or secondary mutations. Most primary mutations are found within the substrate binding pocket. They are conservative mutations (charge and polarity are conserved) but the geometry is altered. Primary mutations directly reduce drug binding, and in some cases even the binding of the natural substrate, thus affecting viral replication. Secondary mutations rectify this problem by playing a compensatory role. They improve proteolytic efficiency and increase viral replication (Boom *et al.*, 1990; Nijhuis *et al.*, 1999; Shafer, 2002). In addition to primary and secondary mutations within the protease, amino acid changes are also introduced in the cleavage sites of viral Gag and Gag-Pol polyproteins so that the mutated protease is able to bind and catalyse these polyproteins (Doyon *et al.*, 1996; Zhang *et al.*, 1997; Mammano *et al.*, 1998)

The type and location of an amino acid polymorphism is confined by the need for the enzyme to maintain catalytic function. Mutations are normally found in the flap hinge region, in the loop connecting the β strands and the α-helix and opposite β strand (Velazquez-Campoy *et al.*, 2003b). The hinge region and the α-helix play an important role in flap dynamics and subunit rotation. Mutations in these regions affects inhibitor binding (Velazquez-Campoy *et al*., 2001c).

Another aspect of decreased drug efficacy involves NOPs. Some amino acids that occur naturally in non-subtype B are associated with drug resistance in subtype B protease, questioning the efficiency of PIs in non-subtype B proteases. For example the M36I mutation in subtype B exists as a drug resistant mutation, but in subtype C, Ile 36 occurs naturally (Becker-Pergola *et al.*, 2000). NOPs, in the absence of drug pressure, are found in areas that do not lead to loss of structural stability or .loss of catalytic activity (Velazquez-Campoy *et al*., 2003b).

1.6.1. Insertions

During ARV therapy insertions, instead of mutations are sometimes selected. In isolates from patients failing drug therapy, insertions in both reverse transcriptase and protease are sometimes found (Masquelier *et al.*, 2001; Winters and Merigan, 2005). It is suggested that

most insertions are duplicates of neighbouring DNA sequences. This may be a result of primer or template slippage during the reverse transcription process (Kozisek *et al.*, 2008).

Insertions in the HIV protease gene are fairly rare with a prevalence rate of 0.1% (Kim *et al.*, 2001). According to a mutation survey conducted by the Konvalinka group (Kozisek *et al*., 2008), most insertions occur between residues 32 and 41. This includes the substrate cleft and the flap hinge. Insertions occur in regions of the protease that are next to externally exposed loops. The reason could be that these regions may be able to accommodate the extra amino acids by extending outwards, thus, preventing tremendous changes in the overall structure of the protein (Winters and Merigan, 2005).

The susceptibility of PIs to insertions is dependent on whether they occur in the presence of major protease resistance mutations. It was found that insertions that occurred in the absence of major drug resistant mutations were completely susceptible to PIs. Insertions that occurred in addition to drug resistant mutations showed a reduced susceptibility to PIs (Kim *et al*., 2001; Sturmer *et al.*, 2003). Although the prevalence rate of insertions is low, there seems to be a slow increase since 1999 (Kozisek *et al*., 2008). Insertions are transmittable between individuals and could lead to the spread of new forms of drug resistant strains (Winters and Merigan, 2005).

The first insertion to be reported was M36T↑N↑L, found in a drug naïve patient infected with subtype B (Sturmer et al., 2003). It is a six base-pair insertion at position 36. The M36T↑N↑L insertion did not show any drug susceptibility because it does not occur in the presence of drug resistant mutations. Position 36 in HIV protease is prone to alterations either as point mutations or insertions; M36I is a drug resistant mutation found in subtype B (Sturmer *et al*., 2003). Other examples of insertions include L33↑L and E35↑E. The E35↑E mutant contributes towards PI resistance by increasing protease vitality in the presence of inhibitors such as Saquinavir, Ritonavir and Lopinavir. Vitality in the presence of Nelfinavir is, however, decreased (Kozisek *et al*., 2008). The L33↑L insertion increases protease vitality in the presence of all the above mentioned inhibitors, including Nelfinavir. HIV protease binding affinity for Nelfinavir is decreased by a factor of ten as a result of the L33↑L insertion (Kozisek *et al*., 2008).

1.6.2. L38↑N↑L protease

Sequencing of the blood samples of a drug-treated patient infected with HIV-1 South African subtype C revealed a unique protease, L38↑N↑L, with two insertions. This patient was reportedly failing drug therapy. The protein sequence of L38↑N↑L is shown in Figure 7 A, which shows the position of Leucine 38, and the amino acid mutations (E35D, I36G, N37S, M46L, D60E) that accompany the insertions. According to the Stanford database (http://hivdb.stanford.edu/cgi-bin/), E35D, N37S, M46L and D60E have all been identified as drug resistance mutations. Figure 7 B is the sequence alignment of L38↑N↑L protease with South African subtype C protease. The Q7K mutation in South African subtype C as shown in the alignment was introduced to minimise auto-proteolysis (Mildner *et al.*, 1994). In the context of this study

1.7. Objective

The L38↑N↑L HIV subtype C protease sequenced from a patient failing drug therapy selected two amino acid insertions, along with drug resistance mutations. The specifics as to which drug regimen the patient was on, was not of importance in the context of this study. It was however, important to study this unique protease that has a hundred and one amino acids compared to the expected ninety nine amino acids. The main objective of this study was therefore to assess the impact these insertions have on the structure and function of South African subtype C protease.

To achieve this objective, this study aims to:

1. Construct L38↑N↑L protease and associated mutations (E35D, I36G, N37S, M46L and D60E) using polymerase chain reaction (PCR) and mutagenesis.

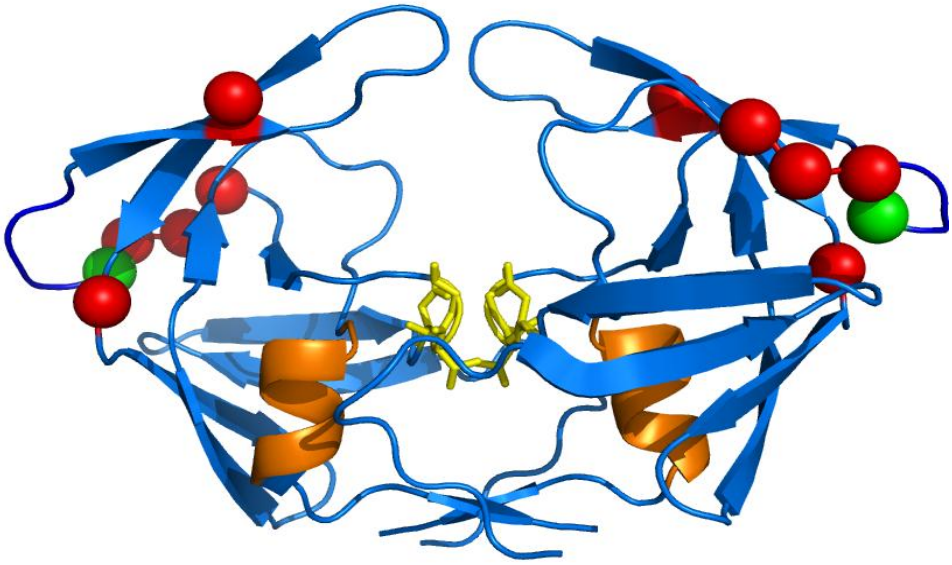
2. Optimise the expression of the L38↑N↑L protease. This is achieved by performing induction studies.

3. Purify recombinant wild type (WTCSA-HIVPR) and L38↑N↑L protease using anion exchange chromatography.

4. Analyse and compare the secondary and tertiary structure of both wild type (WTCSA-HIVPR) and L38↑N↑L protease.

5. Obtain protease enzyme kinetic parameters in order to assess the enzyme activity of the L38↑N↑L protease in comparison to the WTCSA-HIVPR.

6. Use homology modelling to predict the structure of L38↑N↑L protease.



```
                10        20        30          40        50
C-SA    PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEEINL--PGKWKPKMIGGIGGFIKVRQ
        ::::::.:::::::::::::::::::::::::...:  :::::::.::::::::::::
L38     PQITLWQRPLVSIKVGGQIKEALLDTGADDTVLEDGSLNLPGKWKPKLIGGIGGFIKVRQ
              10        20        30        40        50        60

         60        70        80        90     100
C-SA    YDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF
        :.:::::::::::::::::::::::::::::::::::::::
L38     YEQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF
              70        80        90         101
```

Figure 7 A) HIV-1 South African subtype C protease B) Sequence alignment

A) ribbon representation of HIV-1 South African subtype C protease. The location Leu 38 is represented with the green sphere. The position of the accompanying point mutations are represented with the red spheres. This Figure was generated using PyMOL (PDB code 3U71) (DeLano, 2002). B) A sequence alignment of South African subtype C WTCSA-HIVPR (Mosebi *et al*., 2008) (A) with L38↑N↑L protease (B). The amino acid residues highlighted in red are the amino acid point mutations, and the ones highlighted in blue are the insertions.

CHAPTER 2

2. Materials and methods

2.1. Materials

The pET-HIVPR expression vector contains a gene encoding HIV protease subtype B. The vector was a kind gift from Dr J. Tang, (Health Science Centre, University of Oklahoma, Oklahoma city, United States of America) (Ido *et al*., 1991). Site directed mutagenesis was used to convert subtype B protease to South African subtype C protease (WTCSA-HIVPR) by Dr S. Mosebi, (Protein Structure-Function Research Unit, University of the Witwatersrand, Johannesburg, South Africa) (Mosebi *et al.*, 2008). A Q7K point mutation was introduced in order to minimise autocatalysis (Mildner *et al*., 1994). Molecular grade urea used was purchased from Merck (Darmstadt, Germany). A GeneJet^{TM} plasmid and purification kit purchased from Fermentas (Vilnius, Lethuania). Stratagene QuikChange® Lightning Site-Directed Mutagenesis Kit was purchased from Stratagene (Illinois, USA.), and the mutagenesis oligonucleotide primers from Inqaba Biotec^{TM} (Pretoria, South Africa).The chromogenic substrate was a kind gift from Dr. T. Govender, (School of Pharmacology, University of KwaZulu-Natal). Protease inhibitors Saquinavir, Nelfinavir and Ritonavir were a kind gift from the National Institute of Allergy and Infectious Diseases (NIH AIDS Research and Reference reagents program, Germantown, United States of America). Other reagents were of analytical grade.

2.2. Source of HIV protease sequence

HIV-1 protease sequence data from a South African patient infected with South African subtype C was provided by Prof. Lynn Morris from the National Institute for Communicable Diseases, South-Africa (NICD, Johannesburg, South Africa). This information was used to design oligonucleotide primers required for the generation of the L38↑N↑L protease, including the following with background mutations: E35D, I36G, N37S, M46L and D60E.

2.3. Primer design and site-directed mutagenesis

Oligonucleotide primers were designed based on the sequence obtained from the NICD. Firstly, when designing the primers, it was important to ensure that the codon which was

selected coded for a particular amino acid, is found in both *Escherichia coli* (*E. coli*) and Homo sapiens. This was done using graphic codon usage analyser (http://gcua.schoedl.de/). After designing the primers, the oligonucleotide sequences were uploaded into Oligocalc™ (http://www.basic.northwestern.edu/biotools/oligocalc.html), an online software tool that calculates the melting temperatures of the primers. The ideal melting temperature is ≥ 78 °C. A reverse complement of the sequences was generated using an online software tool that generates the reverse complement of the oligonucleotide primers (http://www.bioinformatics.org/sms/rev_comp.html). Table 1 shows sequences of all the primers that were used to generate the L38↑N↑L protease. The Stratagene QuikChange® Lightning Site-Directed Mutagenesis Kit was used to generate the L38↑N↑L protease along with the following background mutations: E35D, I36G, N37S, M46L and D60E. For mutagenesis and the polymerase chain reaction (PCR) 20 ng of double stranded DNA and 125 ng plasmid DNA were used. Other solutions were added according to the manufacturer's instructions. The PCR cycle parameters that were used are summarised in Table 2. After completion of PCR, 1 µL of *Dpn*I restriction enzyme was to digest the parent plasmid according to the manufacturer's instructions.

After PCR it was important to propagate the mutated DNA, this was done by transforming E. cloni® 10G competent cells (Lucigen, Middleton, USA) with 10 ng of pET-HIVPRL38 from the PCR reaction mix and incubated on ice for 30 minutes. The plasmid DNA reaction mix was heat shocked at 42 °C for 45 seconds and immediately cooled on ice for 2 minutes. This was followed by the addition of 900 mL of SOC medium (super optimal broth supplemented with glucose) (2% tryptone (w/v), 0.5% yeast extract (w/v), 10 mM sodium chloride, 2.5 mM potassium chloride and distilled water). The mixture was incubated at 37 °C for 1 hour after which 200 µL was plated onto Lysogeny Broth (LB) plates (1 g tryptone, 0.5 g of yeast and sodium chloride, 1.5 g agar in 100 mL of distilled water) with 100 µg/mL of ampicillin and grown overnight. A single colony was selected and grown in 10 mL of LB media (1g tryptone, 0.5 g of yeast and sodium chloride, 1.5 g agar in 100 mL of distilled water). A GeneJet™ plasmid and purification kit (Fermentas) was used to isolate the plasmid from the E. cloni® cells. The pET-HIVPRL38 sample containing 10 µg of plasmid was sent to Inqaba Biotec™, for sequencing. Inqaba Biotec™ sent results in the form of a chromatogram that was viewed using Finch TV.

Table 1: Primer sequences used to generate L38↑N↑L protease and the background mutations (E35D, I36G, N37S, M46L and D60E).

The nucleotides representing the substitution mutations are in red, and the insertion mutations are in blue.

Mutation	Primer	Sequence
E35D, N37S	Forward	5′- GAC GAC ACT GTT CTG GAA GAT ATC AGC CTG CCG GGT AAA TGG AAG -3′
	Reverse	5′- CTT CCA TTT ACC CGG CAG GCT GAT ATC TTC CAG AAC AGT GTC GTC -3′
I36G	Forward	5′- GAC GAC ACT GTT CTG GAA GAT GGT AGC CTG CCG GGT AAA TGG AAG -3′
	Reverse	5′- CTT CCA TTT ACC CGG CAG GCT ACC ATC TTC CAG AAC AGT GTC GTC -3′
N insert	Forward	5'- GTT CTG GAA GAT GGT AGC CTG AAC CCG GGT AAA TGG AAG CCG AAA CTG ATC GGT G -3'
	Reverse	5′- CAC CGA TCA GTT TCG GCT TCC ATT TAC CCG GGT TCA GGC TAC CAT CTT CCA GAA C - 3'
N and L insert	Forward	5'- GTT CTG GAA GAT GGT AGC CTG AAC CTG CCG GGT AAA TGG AAG CCG AAA CTG ATC G -3'
	Reverse	5'- CGA TCA GTT TCG GCT TCC ATT TAC CCG GCA GGT TCA GGC TAC CAT CTT CCA GAA C -3'
M46L	Forward	5′- GGG TAA ATG GAA GCC GAA ACT GAT CGG TGG CAT CGG CGG -3′
	Reverse	5′- CCG CCG ATG CCA CCG ATC AGT TTC GGC TTC CAT TTA CCC -3′
D60E	Forward	5′- ATC AAA GTT CGT CAG TAT GAA CAG ATC CTG ATC GAA ATC TG -3′
	Reverse	5′- CAG ATT TCG ATC AGG ATC TGT TCA TAC TGA CGA ACT TTG AT -3′

Table 2: PCR cycling parameters

Segment	Cycles	Temperature (°C)	Time
1	1	95	2 minutes
2	18	95	20 seconds
		60	10 seconds
		68	3 minutes
3	1	68	5 minutes

2.4. Recombinant expression of WTCSA-HIVPR and L38↑N↑L protease in *E. coli*

HIV-1 protease WTCSA-HIVPR and L38↑N↑L protease were expressed as inclusion bodies (Ido *et al*., 1991) in BL21 (DE3) pLysS cells. These competent cells were transformed with the pET-HIVPRC for WTCSA-HIVPR and pET-HIVPRL38 plasmid for L38↑N↑L protease Expression conditions for WTCSA-HIVPR required 0.4 mM of isopropyl β-D thiogalactoside (IPTG) with 4 hours of induction. Induction studies were carried out in order to determine optimal conditions for *E. coli* expression of the L38↑N↑L protease. The concentration of IPTG and times of over-expression were varied. There were six flasks in total (0 mM, 0.2 mM, 0.4 mM, 0.6 mM, 0.8 mM and 1 mM IPTG). An overnight culture was prepared by growing the *E. coli* BL21 (DE3) pLysS cells harbouring the pET-HIVPRL38 plasmid in fresh LB medium at 37 °C in the presence of 100 μg/mL ampicillin and 35 μg/mL chloramphenicol. Fresh LB medium (100 mL) was inoculated with the overnight culture making a 100-fold dilution and the cells were allowed to grow at 37 °C at 180 rpm until the OD_{600} had reached 0.4 - 0.5. A sample which represented 0 M IPTG, after zero hours of induction, was collected. The other five flasks were induced with 0.2 mM, 0.4 mM, 0.6 mM, 0.8 mM and 1 mM IPTG. The cells in the six flasks were induced for a total of six hours and 1.5 mL of samples was collected from each flask every hour. The collected samples were centrifuged, (16 000 *g*, 20 minutes, 10 °C) and the pellet resolved on tricine sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) to determine optimal conditions for over-expression. The reason the pellet and not the supernatant was analysed is because the protease is expressed as inclusion bodies.

The optimal conditions for WTCSA-HIVPR and L38↑N↑L protease expression were 0.4 mM IPTG (at OD_{600} of approximately 0.4), 4 hours induction at 37 °C in 5 L Erlenmeyer flask (shaking at 250 rpm). After four hours of induction, the cells were harvested by centrifugation (2700 *g*, 12 minutes, 10 °C). The cell pellet was re-suspended with ice-cold 50 mL of Buffer A (10 mM Tris, 2 mM EDTA and 1 mM PMSF pH 8.0) and stored at - 20 °C.

2.5. Purification and refolding of HIV-1 protease

Purification of HIV-1 protease was performed using modifications of the method previously described by (Velazquez-Campoy *et al*., 2001c). The cells (stored at - 20 °C) were thawed at 4 °C and additional buffer A was added to a final volume of 100 mL. Magnesium chloride ($MgCl_2$) and DNAse I were added to a final concentration of 10 mM and 10 U/μL, respectively. Once the viscosity of the mixture had decreased, the suspension was poured into

sterile tubes and sonicated. The sonicated suspension was centrifuged (25000 × *g*, 30 minutes, 8 °C). The pellet was re-suspended in Buffer A containing 1% (v/v) Triton X-100 and centrifuged (25000 × *g*, 30 minutes, 8 °C). The pellet was then re-suspended in Buffer B (10 mM Tris, 10 mM dithiothrietol and 8 M urea pH 8). The homogenate was centrifuged (32500 × *g*, 30 minutes, 20 °C). The supernatant was loaded onto a DEAE anion exchange column pre-equilibrated with buffer B. Bradford's assay reagent (Bio-Rad, California, USA) was used to determine which of the eluted fractions contained protein. The protein-containing fractions were pooled and 25 mM formic acid was added, reducing the pH to approximately 3.0. This results in precipitation of any non-aspartyl protein contaminants.

The protein was refolded by extensive dialysis against 10 mM formic acid at 4 °C, and then against (10 mM sodium acetate buffer, pH 5.0, containing 2 mM NaCl and 2 mM DTT). The dialysis buffers were hundred times the volume of the sample. The refolded protein was concentrated using ultra-filtration under nitrogen gas. Purity of the L38↑N↑L protease and WTCSA-HIVPR protease was assessed using tricine SDS-PAGE.

2.6. Assessment of protein purity using SDS-PAGE

SDS-PAGE is a technique that separates proteins according to molecular weight. Based on how far protein samples travel and how many bands appear, the size and purity can be analysed (Laemmli, 1970). SDS is a negatively charged detergent. SDS binds to proteins by the same amount and causes proteins to acquire a rod-like shape, with an overall negative charge (Reynolds and Tanford, 1970).

The Laemmli SDS-PAGE method is a discontinuous gel system consisting of a stacking and a separating gel. Each gel has a different percentage of acrylamide (Laemmli, 1970). The stacking gel stacks the protein sample into a thin band. This is an important step that ensures efficient separation of protein samples that have entered the separating gel. If a situation arises whereby the proteins are not separated from bulk SDS, streaking of the separated protein bands occurs (Schagger and von Jagow, 1987; Schagger, 2006).

The Laemmli method provides good resolution for proteins above 100 kDa, but ineffectively resolves proteins below 30 kDa, and these protein bands appear as smears near the gel front (Schagger, 2006). The main reason for this is because SDS-protein complexes of proteins below 30 kDa have approximately the same size and charge and therefore cannot be

distinguished from each other (Fish *et al.*, 1970). It is, therefore, difficult to separate these proteins from bulk SDS. This separation of protein from bulk SDS is important for efficient separation to take place within the separating gel (Schagger and von Jagow, 1987). Tricine SDS-PAGE was used in this study because it provides better resolution of proteins below 30 kDa.

A difference in the degree of separation between the Laemmli and Tricine SDS method relies on differences with the trailing ions used, in particular their pK_a values. The conventional Laemmli method uses glycine as a trailing ion, which has a pK_a value of 9.6. Tricine SDS-PAGE uses tricine, with a pK_a value of 8.15. The pK_a value affects the electrophoretic mobility of the trailing ions with respect to the electrophoretic mobility of proteins (Schagger and von Jagow, 1987; Schagger, 2006). Glycine is able to stack large proteins within the stacking gel because it migrates slowly within the acidic medium of a stacking gel, therefore, increasing the stacking limit. Tricine on the other hand, lowers the stacking limit by moving much faster within the stacking gel, relative to the electrophorectic mobility of the protein. This is because tricine exists in its anionic form, allowing for faster migration. By decreasing the stacking limit, proteins of lower molecular mass are separated efficiently (Schagger and von Jagow, 1987; Schagger, 2006).

Expression and purity of WTCSA-HIVPR and L38↑N↑L protease protein were assessed using Tricine SDS-PAGE. Buffers and solutions were prepared as shown in Table 3 (Schagger, 2006). Figure 8 summaries preparation procedures for Tricine SDS-PAGE samples. The protein samples were diluted one two times with sample buffer (10% (w/v) SDS, 10% beta mercaptoethanol, 10% (v/v) glycerol and bromophenol blue). The samples were incubated at 37 °C for 15 minutes. These samples, along with the protein molecular weight marker (Fermentas SM0431) were loaded onto a tricine gel (4% stacking and 16% separating). In order to ensure that accuracy of SDS-PAGE results, the same amount of sample was loaded on to the tricine gel. The initial voltage used was 30 V until the samples reached the separating gel. The voltage was then increased to 130 V.

2.7. Determination of protein concentration

The concentrations of WTCSA-HIVPR and L38↑N↑L protease were determined spectrophotometrically at 280 nm, using the Beer-Lambert law:

Table 3: Preparation of tricine gels.

	4% sample gel	16% gel
AB-3* (μL)	1	10
Gel buffer** (μL)	3	10
Glycerol (g)	-	3
Add water to a final volume (μL)	12	30
APS (10%) (μL)	90	100
TEMED (μL)	9	10

*AB3 [48 g acrylamide, 1.5 g bisacrylamide, 100 mL distilled water]

** Gel buffer (3 M Tris, 1 M HCl, 0.3% (w/v) SDS, pH 8.45)

Anode buffer ×10 stock solution [1 M Tris, 0.225 M HCl, pH 8.9]

Cathode buffer ×10 stock solution [1 M Tris, 1 M Tricine, 1% (w/v) SDS, pH 8.25]

Fixing solution [50% methanol (v/v), 10% acetic acid (v/v), 100 mM ammonium acetate]

Staining solution [0.025% Coomassie dye (w/v) in 10% acetic acid (v/v)]

Destaining solution [10% acetic acid (v/v) solution]

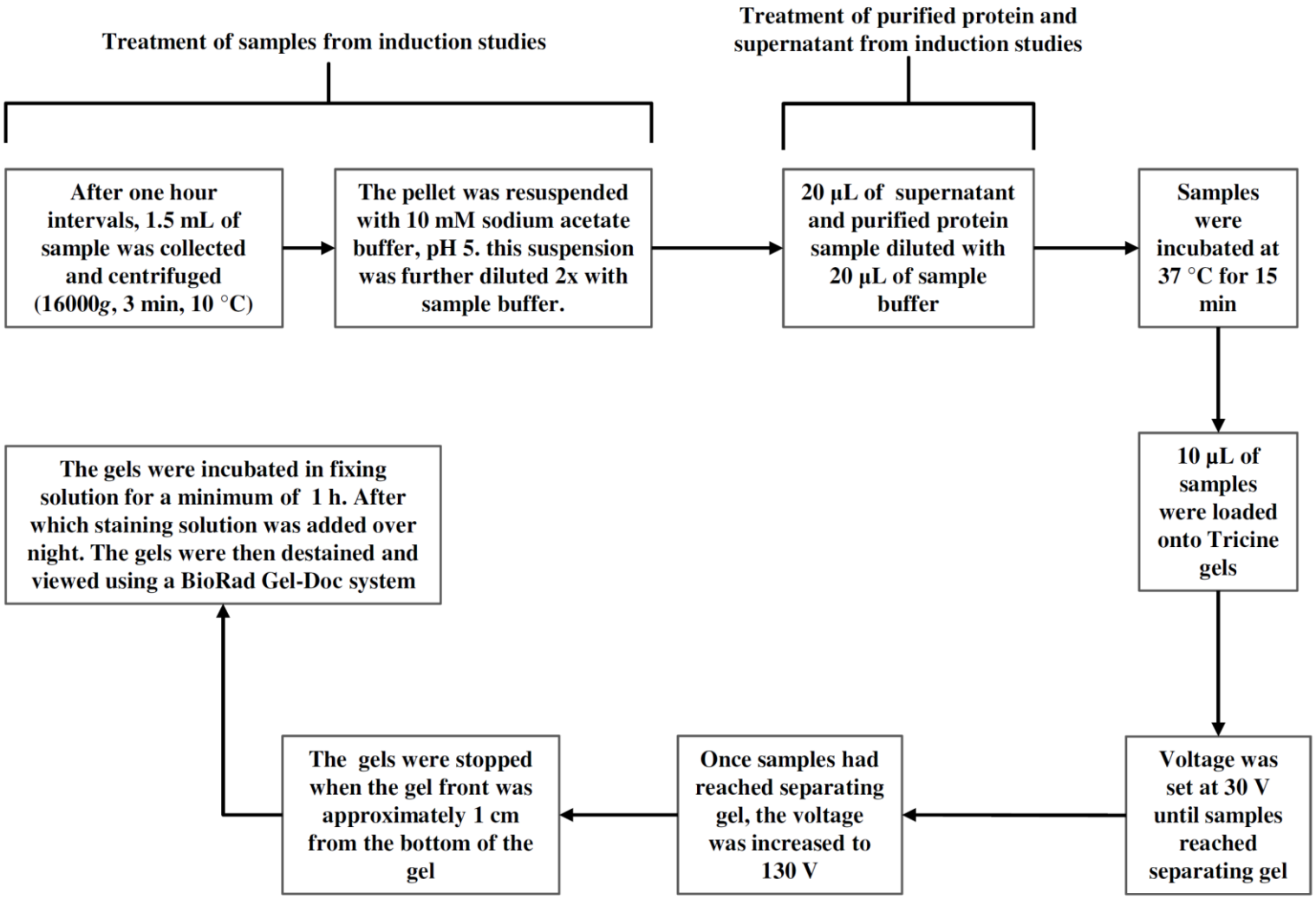


Figure 8 Flow diagram showing the steps involved when resolving protein samples on tricine SDS-PAGE.

$$A = \varepsilon cl \qquad \text{Equation 1}$$

where A is the absorbance of the sample at wavelength λ_{280} (nm), ε is the molar extinction coefficient (M^{-1} cm^{-1}) at wavelength λ_{280}, c is the molar concentration of the absorbing solution and *l* (cm) is the path length of light passing through the solution. The value of the extinction coefficient used was 11800 M^{-1} cm^{-1} (Polgár *et al.*, 1994).

The protein concentration was determined by diluting the protein sample with 10 mM sodium acetate buffer, to make a four times dilution. The absorbance reading was taken at λ_{280} and λ_{340}. The protein was diluted progressively until a ten time's dilution was obtained, and the absorbance read after each dilution. The 10 mM sodium acetate buffer was used as the blank. The λ_{340} for each dilution was subtracted from the 280 nm absorbance for each dilution. This value was then corrected for the buffer. A plot of protein fraction against absorbance was plotted. The slope was divided by the extinction coefficient to obtain the protein concentration.

2.8. Protein structure characterisation

2.8.1. Far-UV Circular Dichroism

A circular dichroism (CD) spectrum results from the interaction of a chromophore in an optically asymmetrical environment with circularly polarised light. In proteins, the backbone amide bonds and aromatic amino acids are the main chromophores. In far-UV (250-190 nm) CD, the main chromophore is the peptide backbone (Adler *et al.*, 1973; Greenfield, 1996). When the peptide backbone has ordered and regular turns, as is the case in alpha helices, beta turns and beta sheets, it interacts with circularly polarised light in a specific way to give a characteristic CD spectrum (Adler *et al.*, 1972). A single CD spectrum is the sum of the spectra of all the conformational elements of that protein. It therefore provides an estimate of the secondary structure (Greenfield, 1999). Proteins with a high alpha helical content exhibit two negative bands at 222 nm and 210-208 nm, and a positive band at 190 nm (Venyaminov and Yang, 1991; Greenfield, 1996). Alternatively, if a protein is composed predominantly of β sheets, a single negative band is obtained between 216 and 218 nm and a positive band at 190 nm (Venyaminov and Yang, 1991; Greenfield, 1996). Disordered proteins have very low ellipticity above 210 nm and a negative band near 195 nm (Venyaminov *et al.*, 1993). HIV protease has a high content of β-sheets and has two α helices. The expected CD spectrum will, therefore, have one negative band between 216 and 218 nm and a positive band at 190 nm.

Far-UV CD spectra for WTCSA-HIVPR and L38↑N↑L protease were obtained using a Jasco J-810 spectropolarimeter. The spectropolarimeter parameters were set at a data pitch of 0.1 nm and 0.5 nm bandwidth. A 2 mm path length cuvette was used. Since far-UV CD is sensitive to chloride ions, protein stocks used for CD studies were diluted to a final concentration of 20 μM using sodium acetate buffer pH 5.0 without sodium chloride. Studies were carried out at 20 °C.

The raw CD data was converted to mean residue ellipticity using the following equation:

$$\mathrm{MRE} = \frac{\theta \times 100}{\mathrm{cn}l} \qquad \text{Equation 2}$$

Where θ is the raw signal (mdeg), c the protein concentration (mM) used, n the number of amino acid residues (WTCSA-HIVPR has a total of 198 and L38↑N↑L protease 202 residues) and *l* path length (cm) of light that passes through the cuvette (Woody, 1995).

2.8.2. Fluorescence spectroscopy

Fluorescence is the emission of light that results from the electronically excited states to the ground state (Lakowicz, 2006). The main fluorophores are the aromatic amino acids, tyrosine and tryptophan, with little contribution from phenylalanine. These aromatic residues have conjugated double bonds which can be excited with light of a specific wavelength to a higher excited electronic state. Once excited, the molecules undergo rapid energy loss through internal conversions. The electrons return to the lowest vibrational state of the excited electronic state and then the ground state emitting a photon of lower energy. Energy absorbed by phenylalanine and tyrosine is often transferred to tryptophan within the same protein. Conversely, tryptophan can be quenched by nearby residues within the same protein (Lakowicz, 2006). Intrinsic fluorescence is highly dependent on the indole ring of the tryptophan residue. The emission spectrum of the indole ring is highly sensitive to solvent polarity. The spectrum may be blue-shifted (lower wavelength) when the indole ring is buried, and red-shifted (higher energy) when it is unfolded or solvent exposed. The indole ring absorbs light energy at 280 nm and emits at a wavelength of approximately 340 nm. Therefore, the emission wavelength is indicative of whether tryptophan is buried, or solvent

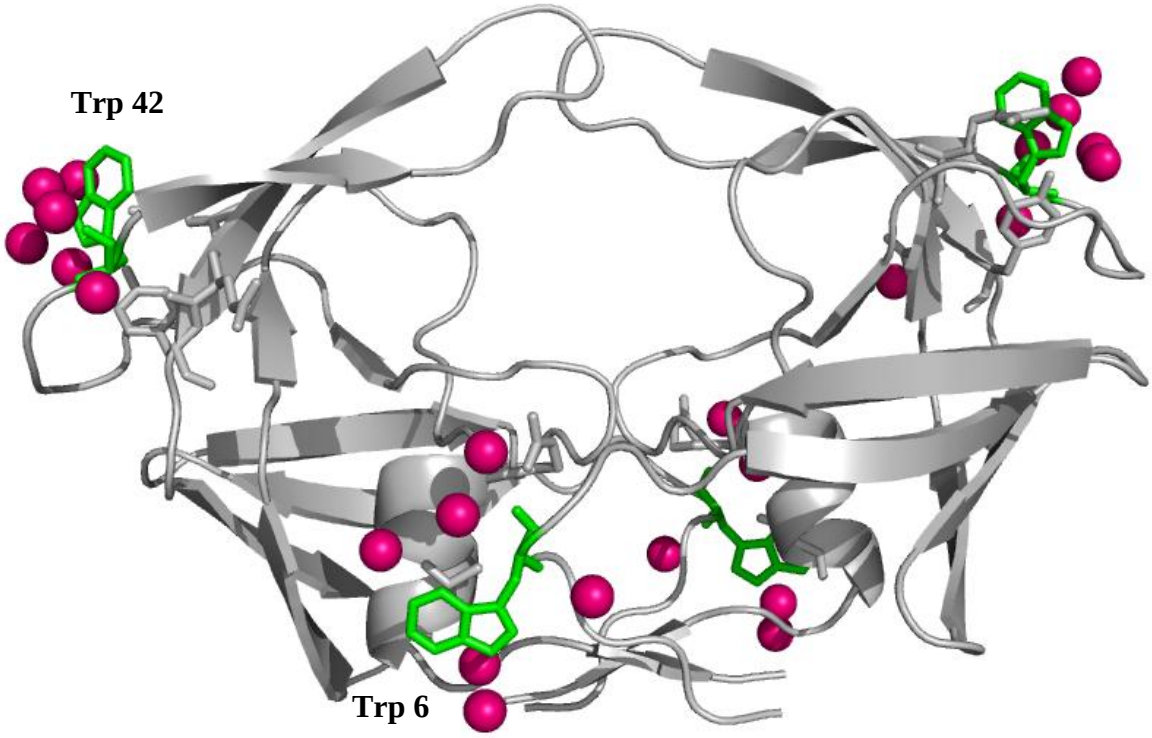


Figure 9 Three dimensional structure of HIV protease showing the position of tryptophan and tyrosine residues

The positions of the fluorescent probes found in HIV protease are shown. Trp 42 is found in the flap elbow, whereas Trp 6 is found close to the dimer interface. The magenta spheres represent water molecules that are within 4 Å of each tryptophan. Trp 42 is partially buried, whereas Trp 6 is highly solvent exposed. This figure was generated using PyMOL (PDB code 2R8N) (DeLano, 2002).

exposed (Lakowicz, 2006). For tryptophan residues that are solvent exposed, an emission maximum is expected around 350 nm. The emission maximum occurs at a higher wavelength (relative to buried tryptophan residues) because when tryptophan residues are solvent exposed, the tryptophan indole group involved in hydrogen bonding, decreasing the energy (Lakowicz, 2006). Fluorecence spectroscopy was used to characterise the environment of the four tryptophan residues in HIV protease. The fluorescent probes in dimeric HIV protease are: Trp 6, Trp 6′, Trp 42, Trp 42′ Tyr 59 and Tyr 59′. Figure 9 shows the interaction of the four tryptophan residues with solvent.

Fluorescence spectroscopy is a highly sensitive technique which is why a low protein concentration is used. The fluorescence spectra for WTCSA-HIVPR and L38↑N↑L protease were obtained using a Jasco FP-6300 spectrofluorometer. Fluorescence emission was observed between 280 and 500 nm. The excitation and emission slit widths were set at 0.5 nm. A total of three accumulated scans were averaged to obtain a single spectrum. Fluorescence assays were performed using 3 μM of protein in sodium acetate buffer pH 5.0. Studies were carried out at 20 °C.

2.9. Enzyme kinetics

Initial rates are important in understanding the chemical mechanism of an enzyme. When performing enzyme kinetic studies it is important to ensure that reactions occur at initial velocity. This is because at initial velocity the concentration of reactants is known from the amounts added. Secondly, at time zero, there are no products present so there will be no back reactions taking place. This ensures that the formation of product is limited by available enzyme, and not because of back reactions. In addition, initial velocity studies guarantee that no errors occur due to loss of enzyme activity over time (Palmer, 1995b; Michaelis *et al.*, 2011). At significantly high substrate concentrations, the enzyme becomes saturated, forcing the enzyme and substrate to form an enzyme-substrate complex. This ensures that there is no unbound enzyme; therefore, the total amount of enzyme in the reaction will be the concentration of the enzyme-substrate complex. This follows zero order kinetics, since the rate of reaction cannot be increased by increasing the amount of substrate(Michaelis *et al.*, 2011).

For a single substrate reaction, the substrate binds the enzyme at a specific binding site, forming a relatively stable enzyme-substrate complex. Stability is ensured by placing the

reacting groups in close proximity to each other, and the catalytic site of the enzyme. Catalysis may involve a second transition-state that quickly collapses to give products. Products may still be bound, producing another unstable enzyme-product complex. The rate-limiting step is the conversion of the enzyme-substrate complex to enzyme-product complex. The reaction rate is through determination of maximum reaction velocity (V_{max}). This can be determined using equation 3 (Palmer, 1995b):

$$V_{max} = k_2 [E]_0 \qquad \text{Equation 3}$$

Where $[E]_0$ is the total amount of enzyme present, free enzyme and the enzyme-substrate complex.

According to Michaelis and Menten, equilibrium is established between enzyme, substrate and enzyme-substrate complex. The breakdown of the enzyme-substrate complex is too slow to disturb the equilibrium, so it is excluded. They depicted the relationship between initial velocity and substrate concentration to be the following(Michaelis *et al*., 2011):

$$v_0 = \frac{V_{\max}[S_0]}{[S_0] + K_M} \qquad \text{Equation 4}$$

Where V_{max} is the maximum velocity, $[S]_0$ is the initial velocity and K_M is the substrate concentration half V_{max}. It is further assumed that the substrate is usually present in much higher concentrations than the enzyme. In situations where the substrate concentration is assumed to be 'low', in order to give first order kinetics with a small degree of saturation, the substrate concentration may still be a thousand times higher than concentration of enzyme. So, if the initial substrate concentration $[S]_0$ is much greater than the initial enzyme concentration $[E]_0$, then the formation of the enzyme-substrate complex will result in an insignificant change in free substrate concentration. Therefore, $[S]_0$ can be substituted for [S] (Michaelis *et al*., 2011).

The Michaelis-Menten equation cannot, however, be applicable to all enzyme catalysed systems as some reactions proceed so fast that such an equilibrium is not established. The Briggs-Haldan modification states that the steady-state assumption is more valid. The Briggs-

Haldan claim states that since the concentration of enzyme relative to that of substrate is negligible, it also means that the rate of change for the enzyme-substrate concentration will also be negligible, compared to that of product [P] over time (Briggs and Haldane, 1925).

2.9.1. Specific activity

Enzyme activity can be described in terms of the enzyme unit (U), which is the amount of enzyme that will catalyse the formation of product from substrate (Wharton and Eisenthal, 1981). Specific activity is an effective way of expressing enzyme purity and it is referred to as the activity of an enzyme per unit weight (U/mg) (Dixon and Webb, 1958; Wharton and Eisenthal, 1981). The chromogenic peptide substrate, Lys-Ala-Val-Nle-*p*-nitro-Phe-Glu-Ala-Nle-NH_2, which mimics the conserved KARVL/AEAM cleavage site between the capsid protein and nucleocapsid (CA-p2) within the Gag polyprotein precursor, was used. The hydrolysis of this substrate was monitored at 300 nm, using a Jasco V-630 spectrophotometer. The extinction coefficient used was $1800 M^{-1} cm^{-1}$ (Velazquez-Campoy et al., 2001c). The reaction mixture contained 50 μM substrate, protein (within range of 100 - 200 nM) and 50 mM sodium acetate buffer, (pH 5, with 0.1 M sodium chloride), at a final volume of 120 μL.

2.9.2. Kinetic parameters

Catalytic parameters are useful in elucidating the catalytic mechanism. The catalytic parameters that were determined for WTCSA-HIVPR and L38↑N↑L protease are the Michaelis constant (K_M), the turnover number (k_{cat}), the catalytic efficiency (k_{cat}/K_M) and the maximum velocity (V_{max}). The K_M is defined in molarity (M). A low K_M value indicates high substrate binding affinity (Wharton and Eisenthal, 1981). The enzyme turn-over describes the number of substrate molecules transformed into product for every enzyme molecule, per unit time (s^{-1}). The catalytic efficiency ($M^{-1} s^{-1}$) describes the ability of an enzyme to increase the reaction rate at low substrate concentrations. The maximum velocity (V_{max}) occurs under saturating substrate concentrations (Wharton and Eisenthal, 1981; Palmer, 1995a).

The catalytic parameters for both WTCSA-HIVPR and L38↑N↑L protease were determined by monitoring the cleavage of the chromogenic substrate Lys-Ala-Val-Nle-*p*-nitro-Phe-Glu-Ala-Nle-NH_2. The hydrolysis of HIV protease chromogenic substrate was monitored at 300 nm. All experiments were carried out at 20 °C in 50 mM sodium acetate buffer (pH 5.0

buffer, with 0.1 M sodium chloride) with a final reaction volume of 120 µL. Experiments were carried out in triplicate and the data is reported as ± SD. For determination of K_M, the protein concentration used was 200 nM. The substrate concentration was gradually increased until saturation had occurred. Determination of the k_{cat} values, saturating substrate concentrations were used (for WTCSA-HIVPR saturation was obtained at 280 µM of substrate and 340 µM for L38↑N↑L protease). The enzyme concentration was varied (within range of 100 – 500 nM). The k_{cat}/K_M values were determined by using non-saturation substrate concentrations (10 – 50µM) and constant protein concentration of 200 nM.

2.9.3. Inhibition studies

The IC_{50} value is defined as the concentration of inhibitor that is required to inhibit the activity of an enzyme by 50%. The IC_{50} values were determined using three PIs (Saquinavir, Ritonavir and Nelfinavir) suspended in 3% (v/v) DMSO. The study was carried out for both WTCSA-HIVPR and L38↑N↑L protease. A reaction mixture consisted of 50 mM sodium acetate buffer (pH 5.0 buffer, with 0.1 M sodium chloride), 200 nM protein, 280 µM substrate concentration for WTCSA-HIVPR and 340 µM substrate concentration for L38↑N↑L protease and different concentrations of inhibitor (0.01 – 0.06 µM). The final reaction was 120 µL. Each reaction was carried out in triplicate. Saturating substrate concentrations were used to ensure that enzyme-substrate complexes are formed within the reaction mixture. As a control, all the above mentioned components without either WTCSA-HIVPR or L38↑N↑L protease were used to make up the reaction mixture.

2. 10 Homology modelling of L38↑N↑L protease

Homology modelling is a method used to accurately ascertain the three dimension model of a target protein based on its amino acid sequence. The three dimensional structure of proteins is more conserved amongst evolutionarily related proteins as compared to their amino acids sequences (Kovalevsky *et al.,* 2006). The target protein is required to at least 50% amino acid sequence similarity with an experimentally solved three dimensional structure (Marti-Renom *et al.*, 2000; Tramontano *et al.*, 2001; Schwede *et al.*, 2003). Since a crystal structure of the L38↑N↑L protease was not determined, homology modelling was to predict its three dimensional structure. HHpred™ (Soding *et al.*, 2005) was used to model the L38↑N↑L protease structure. HHpred™ is an online homology detection and structure prediction tool

(http://toolkit.tuebingen.mpg.de/hhpred). The three dimensional modelling was done using the crystal structure of HIV protease in complex with Darunavir (PDB id 2HS1, 0.84 Å and 1.22 Å resolution, resolved by SER-CAT, Advanced Photon Source) (Kovalevsky *et al.*, 2006). The flow chart in Figure 10 shows the steps taken in the homology modelling. In order to validate the success of energy minimisation, refinement was perfomed on both L38↑N↑L protease and 2HS1. Ramachandran plots were generated by Molprobity™ (Chen *et al.*, 2010)

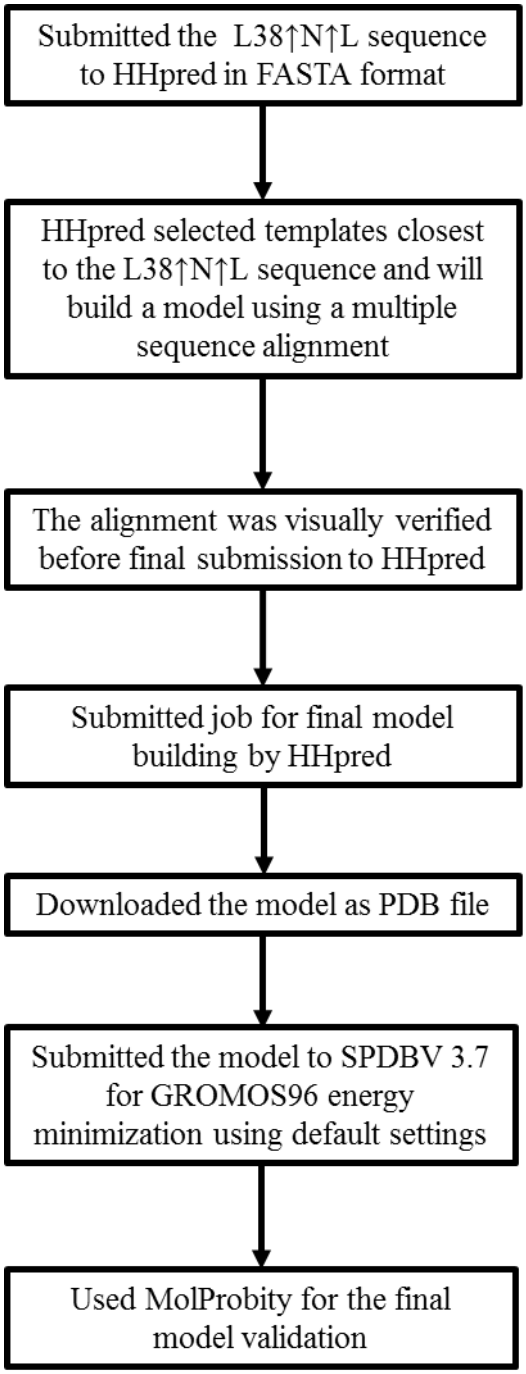


Figure 10 Flow chart showing the steps used for the prediction of the three dimensional structure of L38N↑L↑ mutant.

CHAPTER 3

3. Results

3.1. Verification of pET-HIVPRL38 sequence

A Stratagene QuikChange® Lightning Site-Directed Mutagenesis Kit was used to generate the L38↑N↑L protease within the pET-HIVPRC plasmid DNA. Background mutations that accompanied amino acid insertions between point 38 and 39 are E35D, I36G, N37S, M46L, and D60E. The success of mutagenesis was confirmed by DNA sequencing at Inqaba Biotec®. The chromatograms in Figure 11 A show the positions of the generated mutations. Figure 11 B is a sequence alignment of WTCSA-HIVPR, sequence obtained from NICD and L38↑N↑L protease.

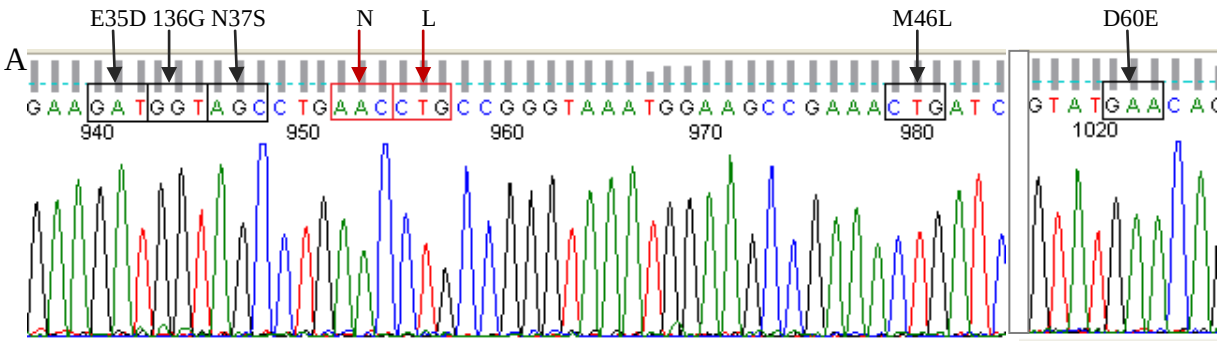
3.2. Induction studies

In order to optimise the over-expression of the L38↑N↑L protease, induction studies were carried out. This was done using varying IPTG concentrations (0 mM, 0.2 mM, 0.4 mM, 0.6 mM, 0.8 mM, and 1 mM). Samples were collected at different time intervals; 0, 2, 4 and 6 hours. Collected samples were resolved on a 16% Tricine gel in order to check for expression. Figure 12 shows a representative tricine gel of samples collected after four hours of induction. The gel shows seven lanes labelled A to G. Lanes A to F represent pellet samples with 0 mM, 0.2 mM, 0.4 mM, 0.6 mM, 0.8 mM and 1 mM IPTG, respectively. Lane G shows the molecular weight marker. Lanes A to F, indicate that HIV protease is expressed Although there was expression with different concentrations, 0.4 mM IPTG is the optimal IPTG concentration. WTCSA-HIVPR and the L38↑N↑L protease were expressed in *E. coli* BL21 (DE3) pLysS cells for 4 hours with 0.4 mM IPTG.

3.3. Assessment of protein purity

The protein was purified using DEAE anion exchange column. Purity of WTCSA-HIVPR and L38↑N↑L protease was assessed using Tricine SDS-PAGE as shown in Figure 13. Lane A is WTCSA-HIVPR protease, lane B is L38↑N↑L protease, and lane C is the molecular weight marker (Fermentas SM0661). A space between the protein samples and molecular

A



B

Figure 11 Verification of mutagenesis success

A) Chromatogram showing regions in the pET-HIVPRL38 nucleotide sequence where the mutations were incorporated. Positions of E35D **(GAA to GAT)**, I36G **(ATC to GGT)**, N37S **(AAT to AGC)**, inserts N **(AAC)** and L **(CTG)**, M46L **(ATG to CTG)** as well as D60E **(GAT to GAA)** respectively. B) Multiple sequence alignment of WTCSA-HIVPR, sequence obtained from NICD and L38↑N↑L protease.

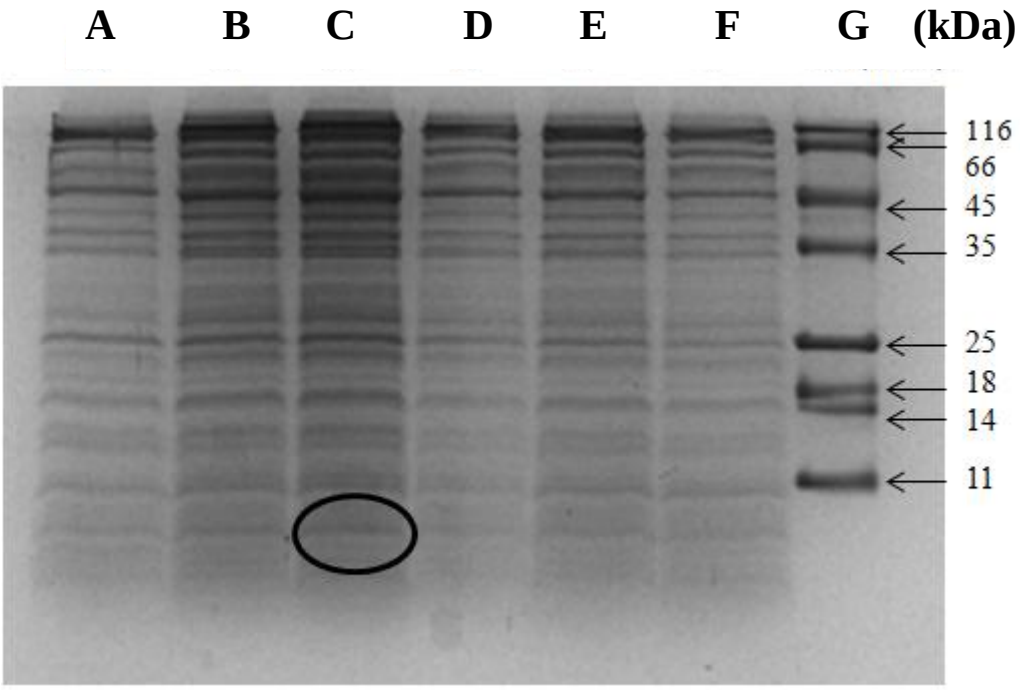


Figure 12 Optimisation of L38↑N↑L protease

Induction studies were performed on *E. coli* BL21 (DE3) pLysS cells transformed with pET-HIVPRL38 plasmid. The gel above shows the samples after four hours of induction Lanes A to F represents samples resulting from different IPTG concentrations that were used (0, 0.2, 0.4, 0.6, 0.8, 1mM). Lane G is the molecular weight marker. Lane C shows higher levels of over-expression relative to other Lanes.

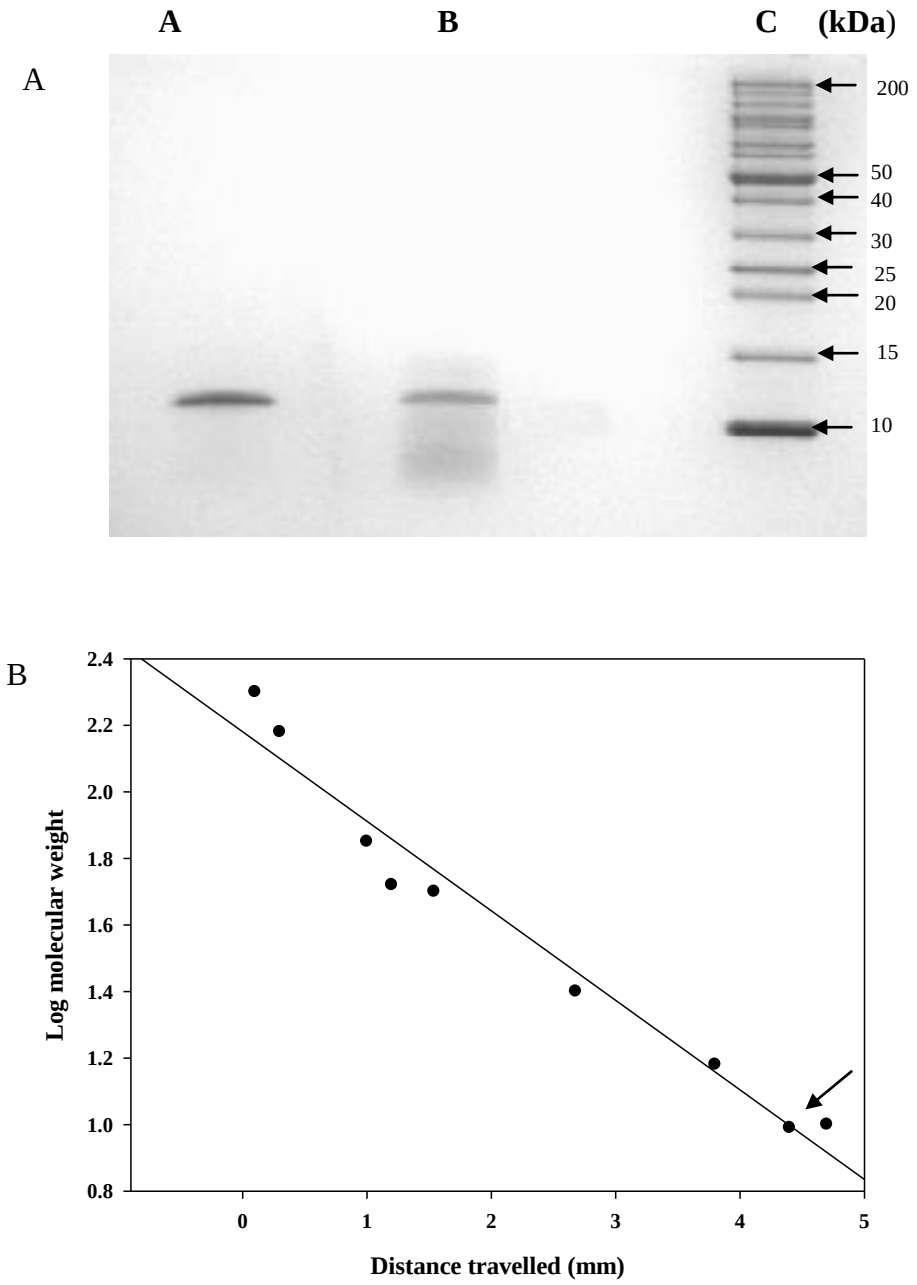


Figure 13 A) Tricine SDS-PAGE showing WTCSA-HIVPR and L38↑N↑L protease B) Calibration curve for WTCSA-HIVPR and the L38↑N↑L protease

WTCSA-HIVPR protease is shown in Lane A, Lane B is L38↑N↑L protease and the molecular weight marker is shown in Lane C. Figure B is a calibration curve for determination of the molecular weights of WTCSA-HIVPR and L38↑N↑L protease. The arrow shows the position of the two proteases. The two proteases travelled 4.4 mm along the tricine gel and the approximate size of the proteins is 10 kDa. The correlation coefficient is 0.95.

weight marker was left to avoid cross contamination between the lanes. According to Figure 13 A WTCSA-HIVPR protease appears to be pure, as no other bands visible. In lane B however, the band appears as a smear. This could be due to protease degradation products. Figure 13 B is the calibration curve used in the verification of protein size. According to the graph the size of both WTCSA-HIVPR and L38↑N↑L is 10 kDa.

3.4. Secondary structural analysis

Far-UV CD was used as a probe to determine the secondary structural integrity of WTCSA-HIVPR and L38↑N↑L protease. Figure 14 is the CD spectra of WTCSA-HIVPR and L38↑N↑L protease. WTCSA-HIVPR CD spectrum is shown in red and L38↑N↑L in black. The spectra of both proteins start at the origin. WTCSA-HIVPR shows a trough at 216 nm and L38↑N↑L protease shows a trough at 203 nm.

3.5. Tertiary structural characterization using fluorescence spectroscopy

Fluorescence spectroscopy was used to establish the local environment surrounding the four tryptophan residues in HIV protease. Excitation at 280 nm results in excitation of tryptophan and tyrosine residues from both monomers (Trp 6, Trp 6′, Trp 42, Trp 42′, Tyr 59 and Tyr 59′), whereas excitation at 295 nm is the excitation of just the tryptophan residues. Figure 15 A shows the fluorescence spectra of WTCSA-HIVPR and L38↑N↑L protease protein at 280 nm excitation. There is no wavelength shift in the fluorescence spectra of WTCSA-HIVPR and L38↑N↑L protease. An emission maximum is obtained at 349 nm for both proteins. However, the fluorescence intensity of L38↑N↑L is decreased. A similar pattern is observed with excitation at 295 nm, as shown in Figure 15 B. Emission maximum is also obtained at 349 nm and there is a decrease in fluorescence intensity of the L38↑N↑L protease.

3.6. Determination of catalytic parameters

The catalytic parameters for both WTCSA-HIVPR and L38↑N↑L were determined by monitoring the cleavage of the chromogenic substrate, Lys-Ala-Val-Nle-*p*-nitro-Phe-Glu-Ala-Nle-NH_2.

Figure 16 shows the graphs from which the specific activity of WTCSA-HIVPR and L38↑N↑L protease were obtained. The specific activity was obtained from the slopes of these graphs. The specific activity value for WTCSA-HIVPR is higher than that obtained for

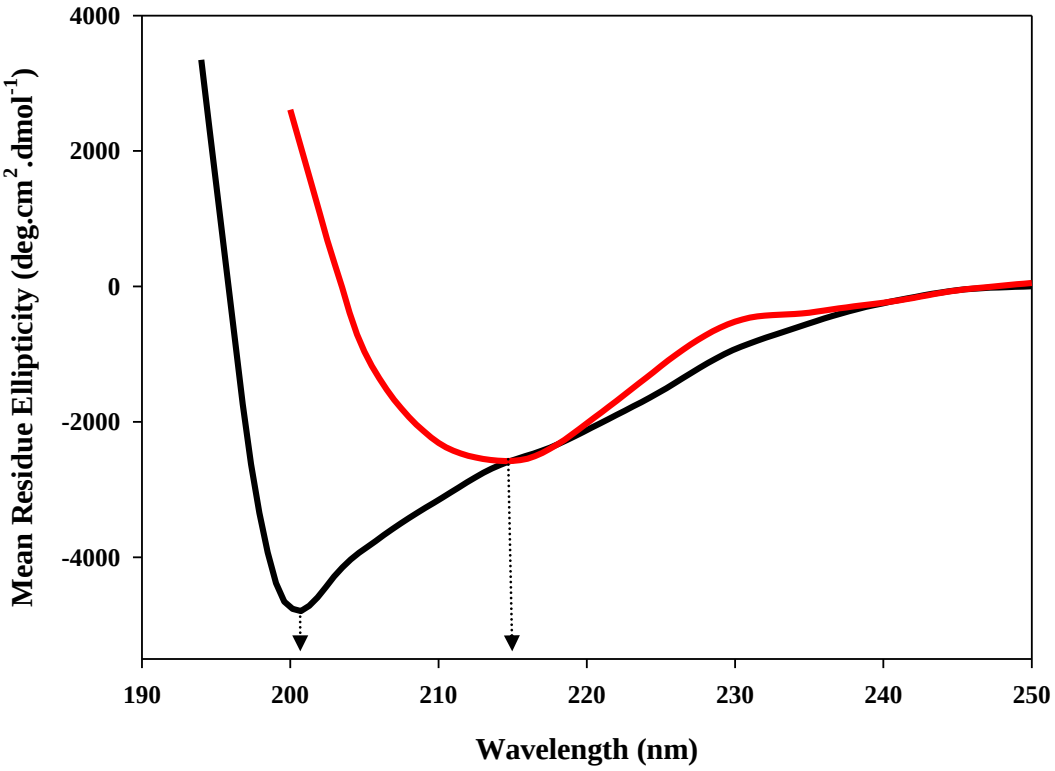


Figure 14 Far-UV CD spectra for WTCSA-HIVPR and L38↑N↑L protease

The far-UV CD spectra for both WTCSA-HIVPR (—) and L38↑N↑L protease (—) show troughs at 215 nm and 203 nm respectively.

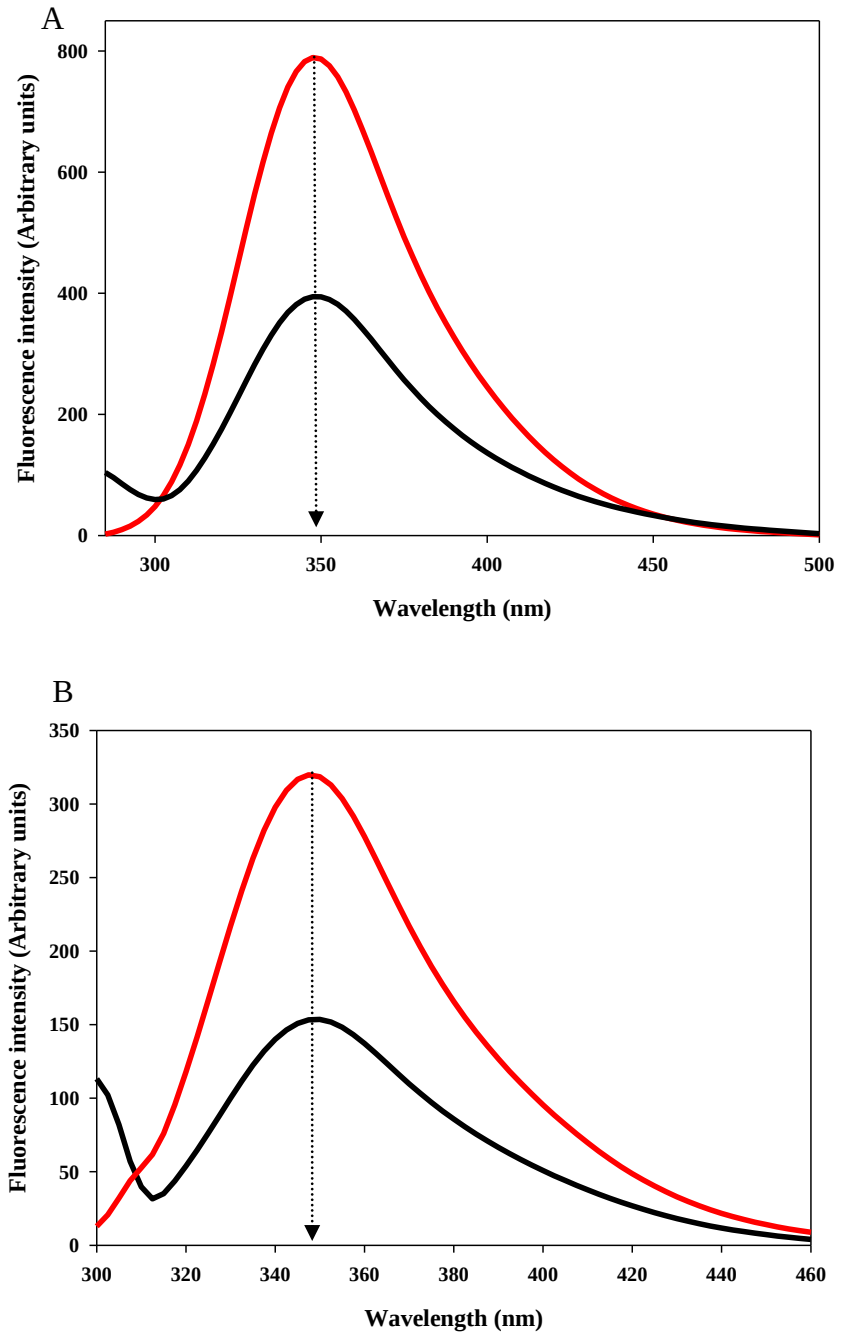


Figure 15 Fluorescence spectra of WTCSA-HIVPR and L38↑N↑L protease

Fluorescence spectra of both WTCSA-HIVPR (—) and L38↑N↑L protease (—) at (A) 280 nm excitation with emission maximum at 349 nm and (B) 295 nm excitation with an emission maximum at 349 nm.

L38↑N↑L protease. These values are shown in Table 4. Figure 17 A and B are the graphs from which the K_M of both WTCSA-HIVPR and L38↑N↑L protease were determined. Sigma plot® was used to fit the two curves, using the equation for a hyperbolic relationship. The K_M and V_{max} values were obtained for WTCSA-HIVPR. However, in the case of the L38↑N↑L protease, the data did not fit to graph described for a hyperbola. The experiments were performed several times using different protein concentrations and a similar pattern was obtained.

The slopes of Figure 18 were used to determine the k_{cat} values of WTCSA-HIVPR and L38↑N↑L protease. The k_{cat} value of the WTCSA-HIVPR is much higher than of L38↑N↑L protease. Figure 19 shows the linear progress curves that were used to determine the k_{cat}/ K_M values. A final k_{cat}/ K_M value for WTCSA-HIVPR and L38↑N↑L protease was obtained by averaging values obtained from the slope of each graph. The k_{cat}/ K_M of WTCSA-HIVPR is three times higher than that of L38↑N↑L protease. These values are summerised in Table 4.

Figure 20 shows the inhibition profiles for WTCSA-HIVPR and L38↑N↑L protease with Saquinavir, Ritonavir and Nelfinavir, respectively. Figure 20 A, B and C are the inhibition profile for WTCSA-HIVPR and L38↑N↑L protease with Saquinavir, Ritonavir. and Nelfinavir. Comparison between WTCSA-HIVPR and L38↑N↑L protease for each curve are not too different from each other. The IC_{50} values for WTCSA-HIVPR and L38↑N↑L protease with each PI are summarised in Table 5.

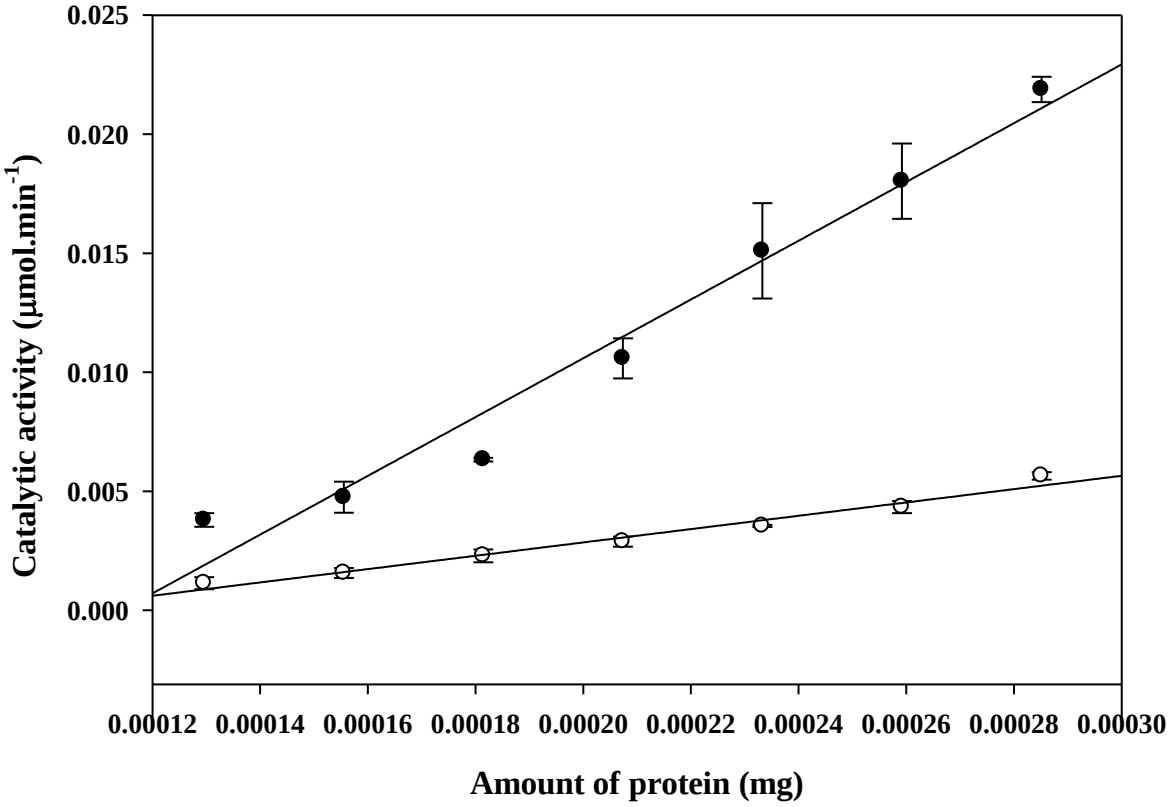


Figure 16: Specific activity of WTCSA-HIVPR and L38↑N↑L protease

The specific activity of WTCSA-HIVPR (●) and L38↑N↑L protease (○) were determined by monitoring the hydrolysis of HIV protease chromogenic substrate at 300 nm. The specific activity was obtained from the slopes of the graphs. The correlation coefficients for the graphs are 0.95 and 0.96 for WTCSA-HIVPR and L38↑N↑L protease respectively.

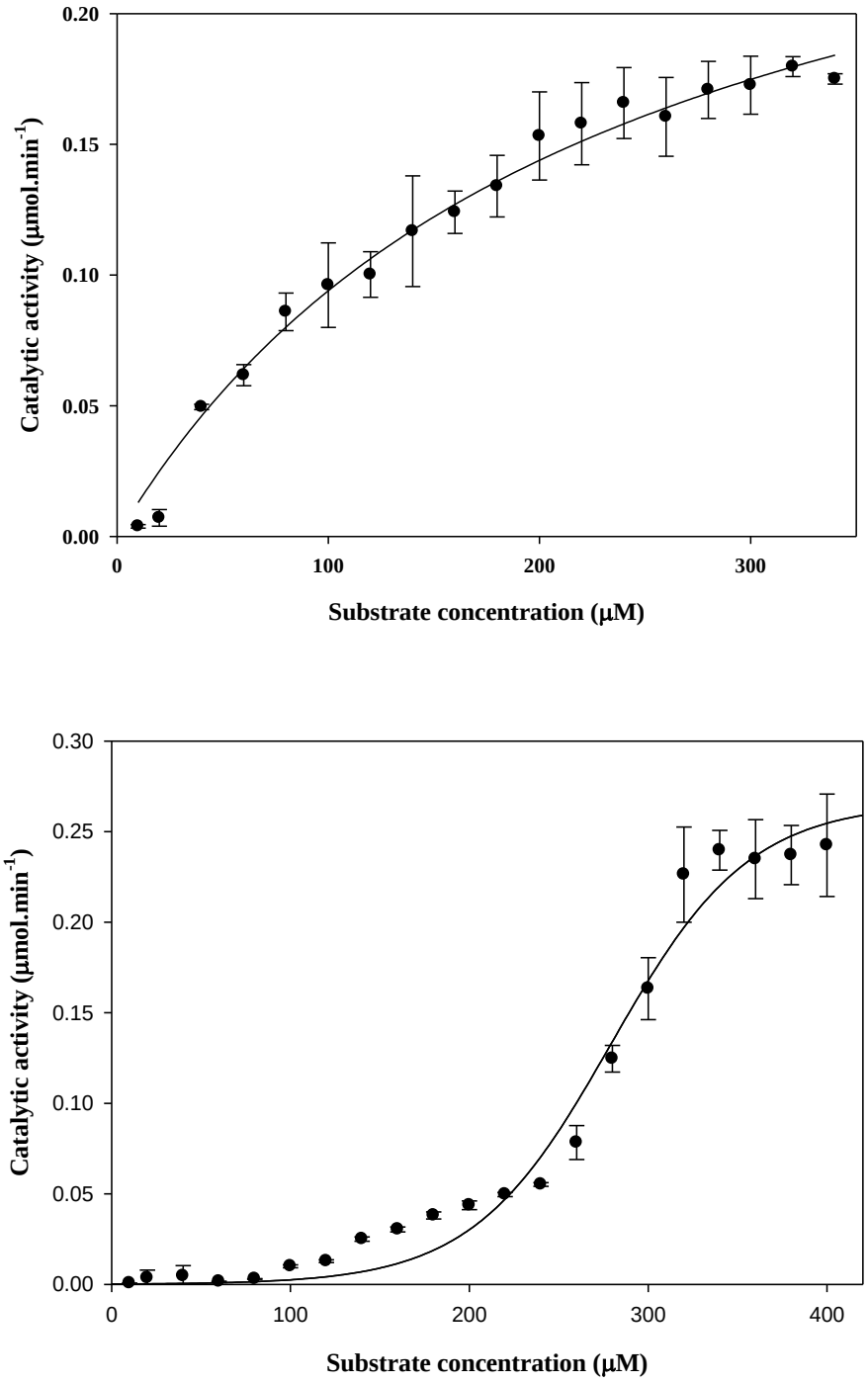


Figure 17: Michaelis-Menten plot for determination of the K_M values for A) WTCSA-HIVPR and B) L38↑N↑L protease

The WTCSA-HIVPR data fit to a graph described for a hyperbola, the L38↑N↑L protease data fit to a sigmoidal graph.

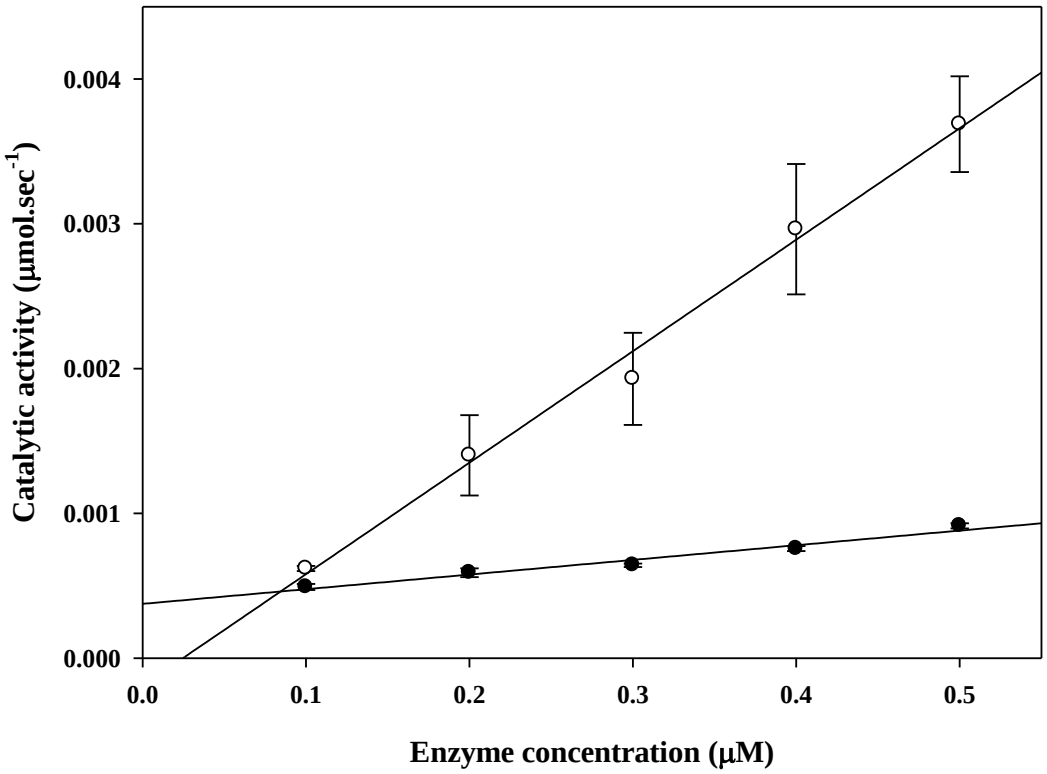


Figure 18 Determination of enzyme turn-over (k_{cat})

The turn-over values for WTCSA-HIVPR (●) and L38↑N↑L protease (○) were determined from the slope of each graph. The correlation coefficient for the graph is 0.94 for WTCSA-HIVPR and 0.96 for L38↑N↑L protease.

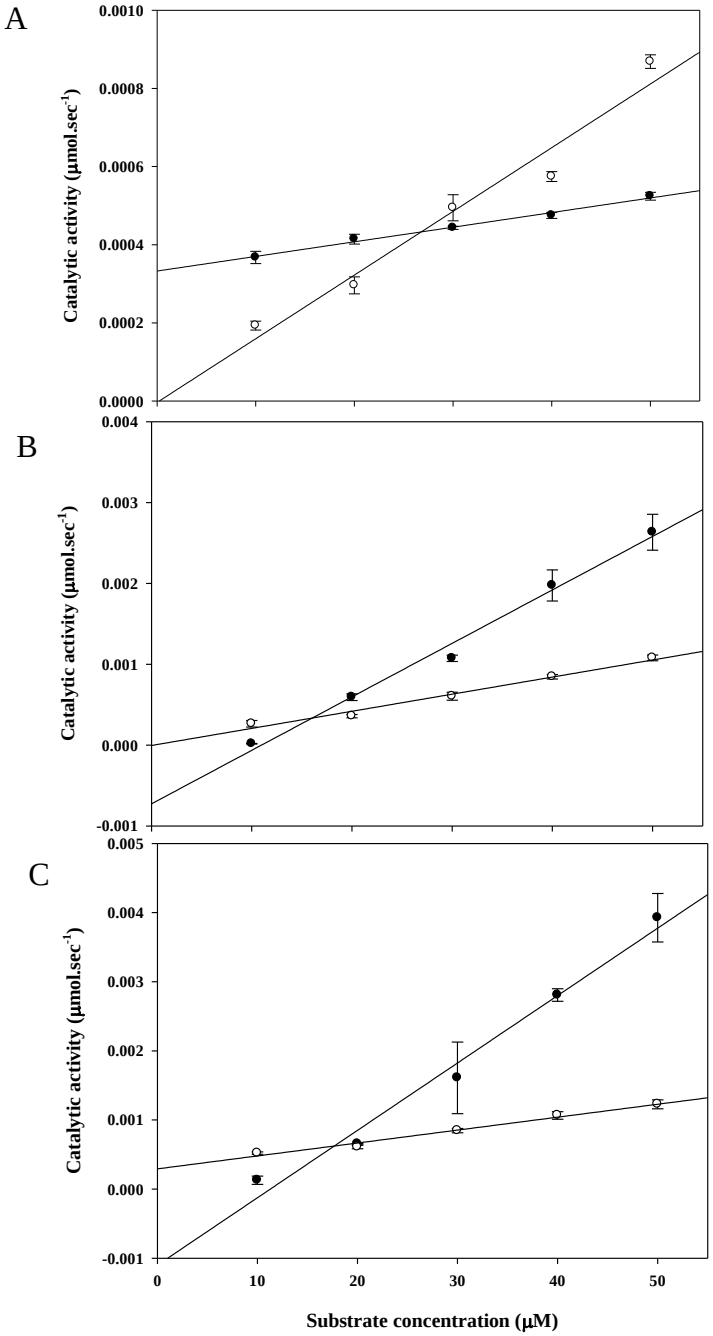


Figure 19 Determination of the catalytic efficiency (k_{cat}/K_M)

The catalytic efficiencies were determined from the slopes of the graphs. Figure A, B and C show the k_{cat}/K_M values obtained using 100 nM, 200, and 300 nM of protein, respectively. This was done for WTCSA-HIVPR (●) and L38↑N↑L protease (○).The correlation coefficients for graphs in Figure A are 0.96 for WTCSA-HIVPR and 0.95 for L38↑N↑L protease. Figure B, 0.97 for both graphs and in Figure C the correlation coefficients for the graphs are 0.96 for WTCSA-HIVPR and 0.97 for L38↑N↑L protease.

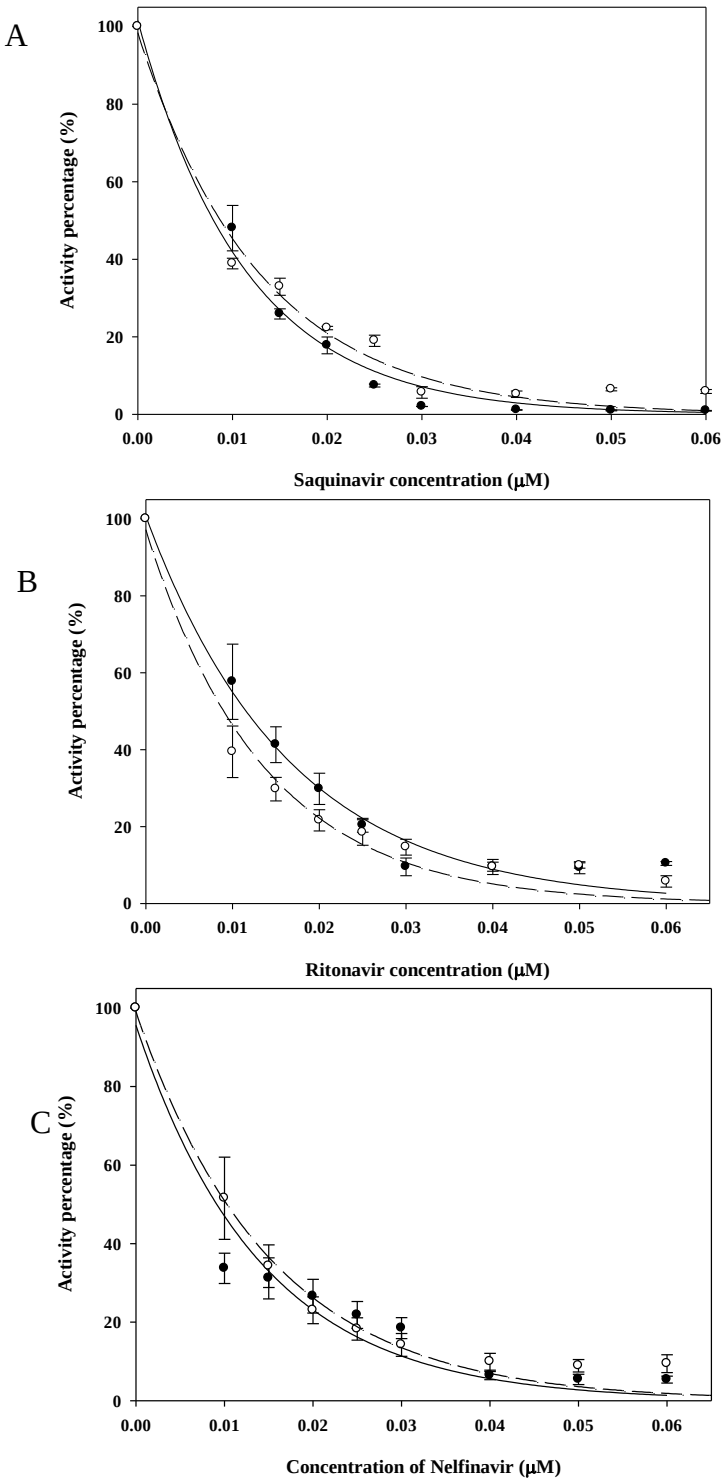


Figure 20 Determination of WTCSA-HIVPR and L38↑N↑L protease IC_{50} values

The IC_{50} values were determined for A) Saquinavir, B) Ritonavir and C) Nelfinavir.

Table 4: Catalytic parameters for WTCSA-HIVPR and L38↑N↑L protease

	Specific activity ($\mu mol.min^{-1}.mg^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu M^{-1}.s^{-1}$)
WTCSA-HIVPR C-SA	123.45 ± 6.4	226.43 ± 27.84	$7.7 \times 10^{-3} \pm 5.0 \times 10^{-4}$	$5.6 \times 10^{-5} \pm 3.0 \times 10^{-6}$
L38↑N↑L	28.00 ± 1.3	-	$1.0 \times 10^{-3} \pm 6.0 \times 10^{-5}$	$1.9 \times 10^{-5} \pm 9.7 \times 10^{-7}$

Table 5: IC_{50} values of WTCSA-HIVPR and L38↑N↑L with three protease inhibitors

	IC_{50} Saquinavir (nM)	IC_{50} Ritonavir (nM)	IC_{50} Nelfinavir (nM)
WTCSA-HIVPR C-SA	8 ± 3.6	11 ± 5.3	9 ± 6.4
L38↑N↑L	8 ± 4.1	9 ± 5.4	10 ± 5.2

3.7. Homology modelling

The protein data bank (Bernstein *et al.*, 1977), has 588 protease structures. However HHpred picked one sequence that closely resembles the sequence of L38↑N↑L protease. HHpred selected 2HS1 (Kovalevsky et al., 2006) as the best template for L38↑N↑L protease. The sequence alignment is shown in Figure 21. The total energy before minimisation was - 3857 $kJ.mol^{-1}$ for L38↑N↑L protease and - 624.31 $kJ.mol^{-1}$ for 2HS1. Table 6 below summarises the results obtained from MolProbity, before and after energy minimisation. This summary describes the percentage of residues with poor rotamers, Ramachandran outliers, and Ramachandran favoured conformations (Chen) The Ramachandran analyses show that 94% for L38↑N↑L protease and 97% for 2HS1 of non glycine and non-proline residues have conformational angles (φ and ψ) in the most favoured regions of the Ramachandran plot, Figure 22 and Figure 23 show the structural alignment of HIV proteases between subtype A (PDB id 3IXO) Robbins 2010), subtype B (PDB id 2HS1) (Kovalevsky , 2006), subtype C (PDB id 2R5Q) (Coman), South African subtype C protease (L38), subtype F (PDB id 2P3C) (Sanches 2007), and CRF_01 A/E (PDB id 2AQU) (Clemente, 2006). The Root Mean Square Deviation (RMSD) between L38↑N↑L protease and 2HS1 is 0.38 Å, with a structural alignment of 396 atoms. The RMSD between L38↑N↑L protease and 2R5Q subtype C protease is 0.87 Å, with 368 atoms.

```
L38↑N↑L.            PQITLWQRPLVSIKVGGQIKEALLDTGADDTVLEDGSLNLPGKWKPKLIGGIGGFIKVRQ 60
2HS1.               PQITLWKRPLVTIKIGGQLKEALLDTGADDTIIEEMSL--PGRWKPKMIGGIGGFIKVRQ 58
                    ********************************::*::**  *******:************

L38↑N↑L            YEQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF 101
2HS1.              YDQIIIEIAGHKAIGTVLVGPTPVNIIGRNLLTQIGATLNF 99
                    *:**:***:*:********************:***:*:****
```

Figure 21 Sequence alignment of L38↑N↑L with 2HS1.

The sequence identity between the two proteins is 81%.

Table 6 Summary of the results obtained from MolProbity™ before and after energy minimisation in order to verify the stereochemistry of the model.

Parameter	Expected value	L38N↑L↑ protease		2HS1 template	
		Before[a]	After[a]	Before[a]	After[a]
Poor rotamers (%)	< 1	2.38	1.39	4.94	3.08
Ramachandran outliers (%)	< 0.2	1.01	1.03	1.03	1.03
Ramachandran favoured (%)	> 98	94.95	93.81	96.91	96.91
Cβ deviations > 0.25 Å	0	0	0	0	0
Residues with bad bonds (%)	0	0	0	0	0
Residues with bad angles (%)	< 0.1	2.97	0	1.01	0

[a] Energy minimization by GROMOS96 implemented on SPDBV v. 3.7

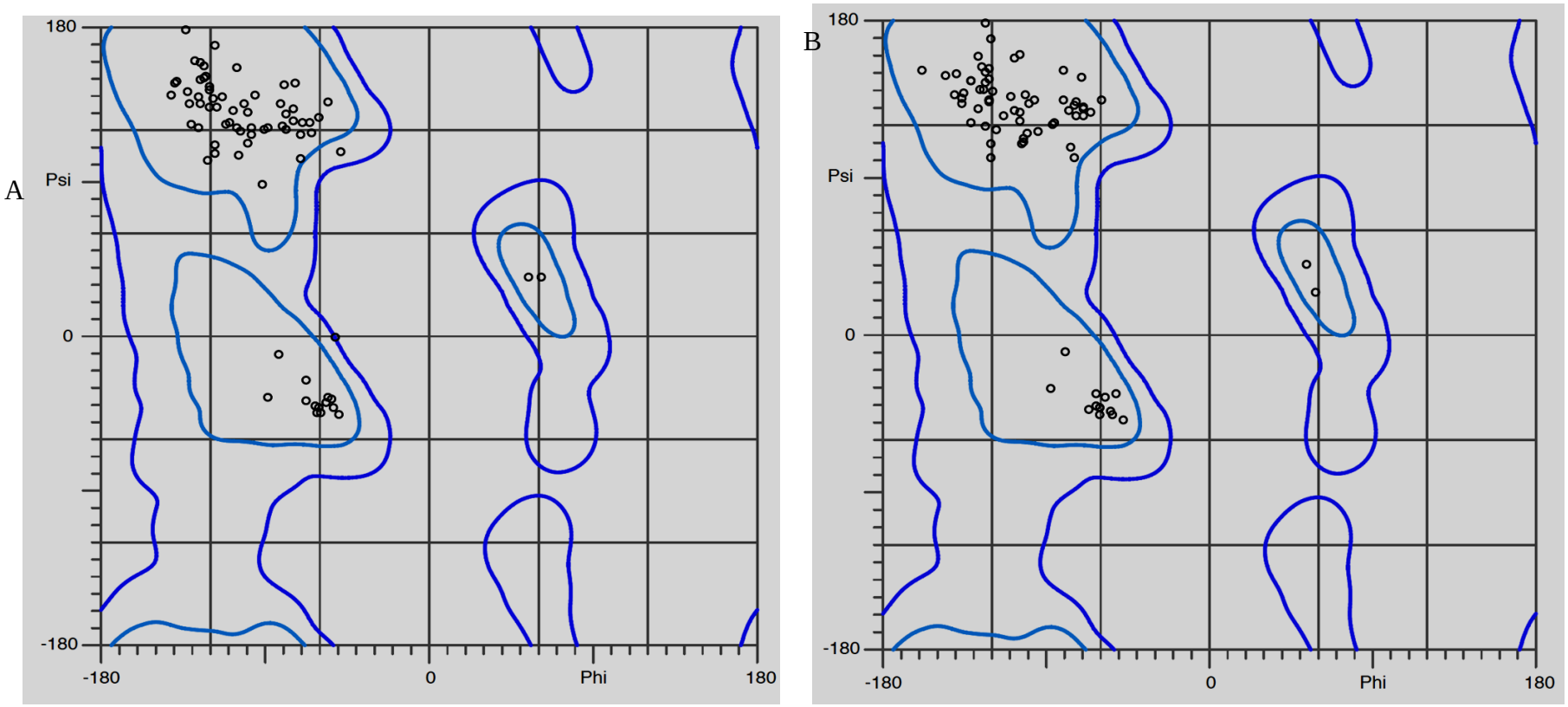


Figure 22 Ramachandran analyses of the homology modelling.

A) L38↑N↑L protease and B) 2HS1 HIV protease by MolProbityTM. These Ramachnadram plots were derived after energy minimisation using GROMOS96. The regions with the cyan line indicate the most allowed region, and the region surrounded by the purple line shows the additionally allowed regions. The white background shows the disallowed regions.

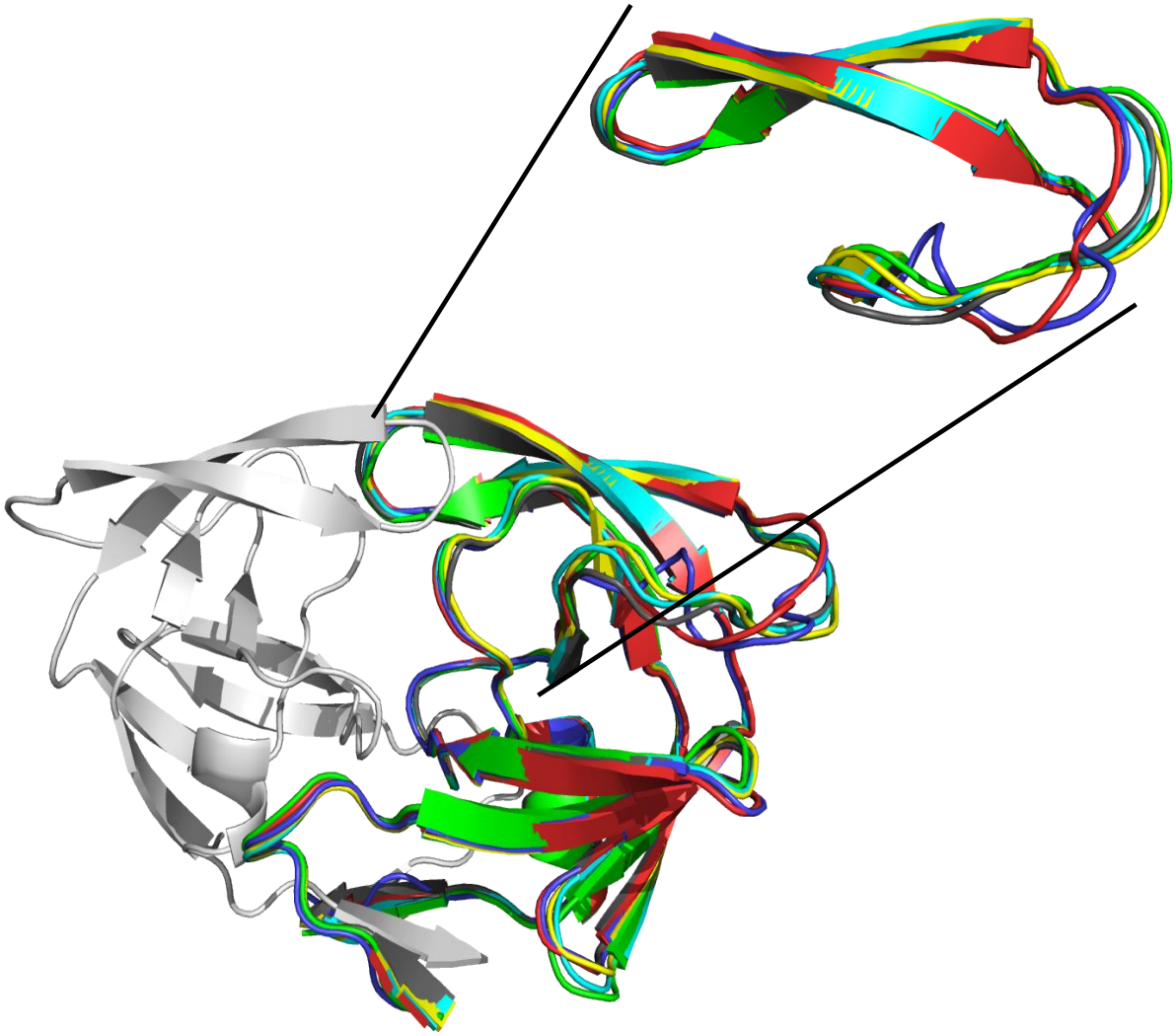


Figure 23 Structural alignment of HIV protease subtypes.

The structures of subtype A (cyan), subtype B (red) subtype C (green), South African subtype C (blue) subtype F (grey) and recombinant form CRF_01 A/E (yellow) were aligned. Most of the structures superimpose well, except for the flap hinge as shown in the model for L38↑N↑L protease.

CHAPTER 4

4. Discussion

HIV PIs were designed on the basis of the subtype B protease structure. How effective these drugs are on non B subtype proteases has been an area of debate. A few reports have suggested that genetic differences between subtypes have no effect on response to PIs (Frater *et al.*, 2001; Frater *et al.*, 2002). Other reports propose that it is the presence of certain polymorphisms in non B subtypes, particularly subtype B drug resistant mutations, that leads to inadequate response to PIs (Descamps *et al.*, 1998; Abecasis *et al.*, 2006). Based on this assumption, it was proposed that studies should be focused on subtype B drug resistant mutations, rather than non B subtype PI susceptibility (Kantor *et al.*, 2005). However, studies conducted by the Freire group (Velazquez-Campoy *et al*., 2001c) suggest an increased catalytic efficiency and vitality for subtype C protease in the presence of inhibitors when compared to subtype B protease. In another study by the Freire group (Velazquez-Campoy *et al*., 2003b) there was evidence of decreased binding affinity in Subtype A, subtype C, South African subtype C and subtype G with PIs. This demonstrated that PIs were developed and optimised for subtype B proteases. In this study it was concluded that decrease in drug susceptibility, as a result of drug resistance mutations, is enhanced by the existence of naturally occurring polymorphisms in non B subtypes (Velazquez-Campoy *et al*., 2003b).

A study done on subtype B proteases L33↑L and E35↑E showed that the insertions had led to an enlarged substrate binding pocket and local changes within the flap region. These changes caused insufficient inhibitor binding (Kozisek *et al*., 2008). In order to understand the effect of the L38↑N↑L insertions on South African subtype C protease a comparative structural and functional analysis between WTCSA-HIVPR and the L38↑N↑L protease was carried out.

South African subtype C L38↑N↑L protease was isolated from a South African patient infected with HIV-1 subtype C. This patient was reportedly failing drug therapy. This is the first time such a polymorphism in has been reported or studied. The insertions are found in the flap hinge. With the emergence of L38↑N↑L protease, it was important to ask the following question: Has the HIV subtype C virus selected mutations that give rise to structural modifications leading to changes in function? Three important aspects will be assessed in order to answer this question. These aspects are; the structure and function of

L38↑N↑L protease, as well as suggestions from the homology model. The expectation is that if significant structural changes were to take place within the flap hinge, then, flap dynamics and therefore drug binding might be affected (Velazquez-Campoy *et al.*, 2001b).

Recombinant L38↑N↑L protease, together with its associated background mutations (E35D, I36G, N37S, M46L and D60E), was used in these studies. The L38↑N↑L protease was successfully purified, and the molecular weight verified (Figure 13 b). Specific activity assays were performed on the protein to ensure that it was catalytically active. This was important because the protease was recovered and refolded from inclusion bodies.

The far-UV CD data shown in Figure 14 implies that the insertions have brought about structural alterations. The WTCSA-HIVPR far-UV CD spectrum shows a trough at 215 nm, as expected for proteins with a high β-sheet content (Adler *et al*., 1972). The far-UV CD spectrum of L38↑N↑L protease however, shows a trough at 203 nm. A trough in this region is found in the spectrum of proteins that are random coils (Greenfield, 1996). The CD spectrum of L38↑N↑L protease was obtained several times to ensure that the spectrum is representative of the mutant's secondary structure. However, the far-UV CD data is inconclusive because if L38↑N↑L protease was random coiled, it would not have been selected for by the virus, since the expectation is that there would be no catalytic activity. In addition, a crystal structure of HIV protease (unpublished data, PDB code 3U71) purified using the same protocol (as stated in section 2.5) was obtained. This structure contained all the three dimensional elements expected for HIV protease. CD studies were performed using protein in its apo form, i.e. unliganded form. Unliganded HIV protease is mainly in the semi-open conformation (Hornak *et al.*, 2006a; Hornak *et al*., 2006b). A possible consequence could be L38↑N↑L protease has increased flexibility because of the additional amino acids in the flap hinge. Inhibitor binding has been shown to stabilise the dimeric structure of protease (Todd and Freire, 1999). It would therefore be useful to obtain far-UV CD studies, for acetylpepstatin – HIV protease complex in addition to the apo HIV protease. A study by L. Mpye 2010 for her MSc dissertation (Protein Structure-Function Research Unit, University of the Witwatersrand, Johannesburg, South Africa) was done on I36T↑T. In the study, the insertion caused no global structural changes, according to the far-UV CD data. The far-UV CD data for both wild type and I36TT mutant showed a trough at 216 nm, as expected for protein predominant in β-sheets (Mpye, 2010). This suggests that the structure of L38↑N↑L proteases is slightly different from I36T↑T. From the CD data, it may be concluded that the L38↑N↑L protease resulted with changes in the secondary structure of protease. The extent to which these

changes have occurred and the specific regions where these alterations are found is still not known. A crystal structure of L38↑N↑L protease would be useful. The structure of the L38↑N↑L protease was further analysed using fluorescence. The information provided by the fluorescence data indicates the changes that have taken place within the vicinity of the tryptophan residues.

The fluorescence data does not suggest any significant changes in the local environment of the four tryptophan residues of protease. According to Figure 15 the emission maximum for WTCSA-HIVPR and L38↑N↑L is 349 nm for both, however, there is a decrease in the fluorescence intensity of L38↑N↑L protease. Differences in the fluorescence intensity between WTCSA-HIVPR and L38↑N↑L protease may possibly be an effect of fluorescence quenching caused by nearby charged amino acids (Lakowicz, 2006). A possible reason for the lack of a wavelength shift in L38↑N↑L protease could be that the local environment of the tryptophan residues is still the same. However, this does not imply that the orientation of the tryptophan residues did not change. The decrease in fluorescence intensity may be due to fluorescence quenching of tryptophan. Tryptophan can be quenched by charged, neighbouring amino acid residues (Lakowicz, 2006). The insertions are located in the flap hinge of the mutant; this is within the vicinity of tryptophan 42. Within a 7 Å sphere of Trp 42 there are five basic residues Lys 41, Lys 43, Lys 45, Lys 55 and Arg 57; three hydrophobic residues, Leu 38, Val 56 and Tyr 59; and four conformationally important residues; Pro 39, Pro44, Gly 40 and Pro 39. There is a polar uncharged group, Gln 58 and a negatively charged residue, Asp 60. Amongst these amino acid residues, those that can act as potential quenchers are: the amide group of glutamine 58, the carboxyl group of aspartic acid 60, the Ɛ-amino group of the lysine residues and the phenol group in tyrosine 59. Quenching by the α-amino group of glutamine, asparagine and lysine also contributes to quenching (Chen and Barkley, 1998). Fluorescence quenching is affected by proximity, specific geometry or local polarity (Chen and Barkley, 1998; Ullrich *et al.*, 2000). Based on this information it can be inferred that in L38↑N↑L protease, there is greater interaction between Trp 42 and the quenching amino acids, possibly due to an extended flap, as compared to the WTCSA-HIVPR. According to the structural information from far-UV CD and fluorescence, it can be assumed that the insertions in L38↑N↑L protease have caused structural changes that require further analysis.

The protease flaps are important during substrate binding. Apart from the obvious role of opening and closing, they also form important interactions: A crystal structure analysis of

HIV protease complexed with acetylpepstatin showed that the inhibitor forms hydrogen bonds with main chain residues in the flap and residues in the active site loops (Fitzgerald *et al.*, 1990). The flap hinge and core regions play an important role in flap dynamics (Rose *et al.*, 1998; Todd and Freire, 1999; Scott and Schiffer, 2000). Comparison of liganded and unliganded protease crystal structures shows that considerable structural alterations take place as a result of substrate binding. Upon substrate binding, the only regions that remain unchanged are the dimer interface and residues of the active site (Fitzgerald *et al*., 1990). This could be a consequence of substrate recognition. It is proposed that substrate recognition in HIV protease is based on structure rather than sequence. The two protease monomers undergo structural alterations in order to accommodate asymmetrical substrates (Prabu-Jeyabalan *et al.*, 2000; Prabu-Jeyabalan *et al.*, 2002).

Sodium acetate buffer pH 5, which was used in the enzymatic assays, is not present under physiological conditions. However, this buffer mimics the environment which HIV protease processes Gag and Gag-Pol polyproteins within the cell. Proteolysis during viral maturation occurs near the plasma membrane (Henderson *et al*., 1983; Mervis *et al*., 1988), which is an acidic pH environment (Honig *et al.*, 1986). The catalytic activity of WTCSA-HIVPR and L38↑N↑L protease were determined using a substrate that mimics a Gag-Pol polyprotein cleavage site. The L38↑N↑L protease has a lower specific activity value compared to the WTCSA-HIVPR, as shown in Table 4. The enzyme turn-over number can be affected by substrate binding. If binding is too loose, the required interactions do not take place, and there is a large energy barrier towards the transition-state. If binding is too tight, the intermediate becomes too stable for product to be released (Albery and Knowles, 1976). The turn-over value for L38↑N↑L is seven times less than that of the WTCSA-HIVPR (Table 4). This suggests that substrate binding may be too loose. If the turn-over number is low, it makes sense why the specific activity is lower in L38↑N↑L protease. The K_M value obtained for WTCSA-HIVPR was 226.43 ± 27.84 μM, as seen in Table 4. The K_M value for L38↑N↑L protease, on the other hand, could not be obtained. The data, as shown in Figure17, does not fit a hyperbolic Michaelis-Menten graph (Michaelis *et al*., 2011), instead it has a sigmoidal shape. A sigmoidal shape for velocity versus substrate graph is often associated with allosteric enzymes. However, HIV protease has one active site (McKeever *et al.*, 1989; Navia *et al*., 1989; Wlodawer *et al*., 1989).

The IC_{50} is the inhibitor concentration that reduces the activity by 50% (Burlingham and Widlanski, 2003). According to Table 5 the IC_{50} values for WTCSA-HIVPR and L38↑N↑L protease were not significantly different from each other. This suggests that the L38↑N↑L protease insertions do not affect inhibitor binding. In a previous study, IC_{50} values were obtained for wild type protease with saquinavir (7.6 nM), nelfinavir (17.3 nM) and ritonavir (19.1 nM) (Lerato Mpye, 2010). The IC_{50} value of 7.6 nM obtained for saquinavir is comparable to the values obtained for both wild type and L38↑N↑L protease as shown in Table 5. The values obtained for nelfinavir and ritonavir are notably different, however, this data is not sufficient enough to draw conclusions with regards to the behaviour of L38↑N↑L protease with the above mentioned PIs. This is because IC_{50} is considered a relative value and is highly reliant on the concentration of substrate used in the assays. Therefore determination of IC_{50} values should be used along with the inhibition constant K_i so that a complete analysis of the behaviour of HIV protease is with inhibitors (Burlingham and Widlanski, 2003).

It should be emphasised that the kinetic studies are not a reflection of what is taking place *in vivo*. When evading drug pressure, HIV selects mutations within the Gag-Pol cleavage sites, along with mutations within the protease. This way it is able to maintain function even in the presence of PIs (Doyon *et al*., 1996; Zhang *et al*., 1997; Mammano *et al.*, 1998). Based on this, it is possible that the substrate used in these studies does not contain the cleavage site mutations that may have been selected for by the virus for cleavage by the L38↑N↑L protease. This could explain why the specific activity of L38↑N↑L protease is so low.

A crystal structure of L38↑N↑L with bound inhibitor could possibly provide an understanding of the interactions involved in this mutant. However, due to time constraints, a crystal structure of L38↑N↑L protease could not be obtained. Homology modelling was used to predict the structure of L38↑N↑L protease. Although this method is important in helping explain the changes that have taken place in the flap region, it is theoretical, and does not override obtaining a crystal structure of L38↑N↑L protease. An online modelling software, HHpred™ was used. In order to avoid any bias, the modeller was allowed to choose structures, from the Protein Data Bank (PDB) that are most likely to resemble L38↑N↑L protease. There are about 588 structures associated with HIV protease in the PDB (Bernstein *et al*., 1977). Out of this array of structures, the modeller chose HIV protease 2HS1 (Kovalevsky *et al*., 2006). This is a subtype B protease in complex with Darunavir. This was against expectation because it did not select the subtype C proteases such as 2R5Q (subtype C protease in complex with nelfinavir), or the 2R8N, the apo enzymes (subtype C protease)

(Coman *et al.*, 2008a). The L38↑N↑L protease has 81% sequence identity with 2HS1. This is much lower than with 2R5Q, which gave a sequence identity of 90%, and WTCSA-HIVPR, 93%.

The refinement of the model using GROMOS96 appeared to be successful on the basis of the improvement of the stereochemical parameters done with the MolProbity^{TM} algorithm. The 2HS1 template was used as a positive control to ensure the success of the refinement. As shown in Table 6, the percentage of poor rotamers decreased by approximately 1% after energy minimisation. In addition, after energy minimisation, the number of rotamers with bad angles was zero. The percentage of Ramachandran outliers and favoured remained the same. For the model, there was an approximate 1% decrease in the percentage of poor rotamers. There was also a decrease in the percentage of residues with poor angles after energy minimisation. The stereochemistry analysis was important in validating the structure of the model. The flap hinge of the L38↑N↑L protease falls within the allowed regions of the Ramachndran plot. This validates the structure of the model. The structural alignment in Figure 23 shows that the global structure of HIV protease is conserved amongst the subtypes. The differences lie in the flap hinge of L38↑N↑L protease and the template 2HS1. 2HS1 has ninety-nine amino acids, and L38↑N↑L protease has 101. If 2HS1 has the same fold as L38↑N↑L protease in the flap hinge, it suggests that the insertions in L38↑N↑L protease are not the cause of this fold in L38↑N↑L protease. The hinge region folding as seen in L38↑N↑L protease could be a polymorphism induced by the background mutations found in L38↑N↑L protease.

Inferences from the 2HS1 M46L crystal structure can be used to explain the structure of L38↑N↑L protease because the same mutation is found in the L38↑N↑L protease. The M46L mutation is selected in seven PIs and is a major Indinavir drug resistant mutation (Johnson *et al.*, 2005). The M46L mutation is found in the flap region and has no direct contact with bound inhibitor. However, Met46 forms hydrogen bonds with substrate analogues (Tie *et al.*, 2005). It is therefore assumed that this mutation affects inhibitor binding by reducing hydrophobic interactions or by actually increasing its interaction with the substrate (Kovalevsky *et al.*, 2006). The M46L mutation within L38↑N↑L protease could be behaving in a similar way. However, the effect of polymorphisms cannot be studied in isolation, the background mutations that exist alongside it, have to be taken into account (Velazquez-Campoy *et al.*, 2003b).

Further analysis of the 2HS1 M46L crystal structure reveals that this protease has an allosteric site, as shown in Figure 24. The second binding site is found on one side of the flap surface. Apparently, the Darunavir binds the active site and flap surface with different configurations (Kovalevsky *et al*., 2006). Although the structure of L38↑N↑L protease

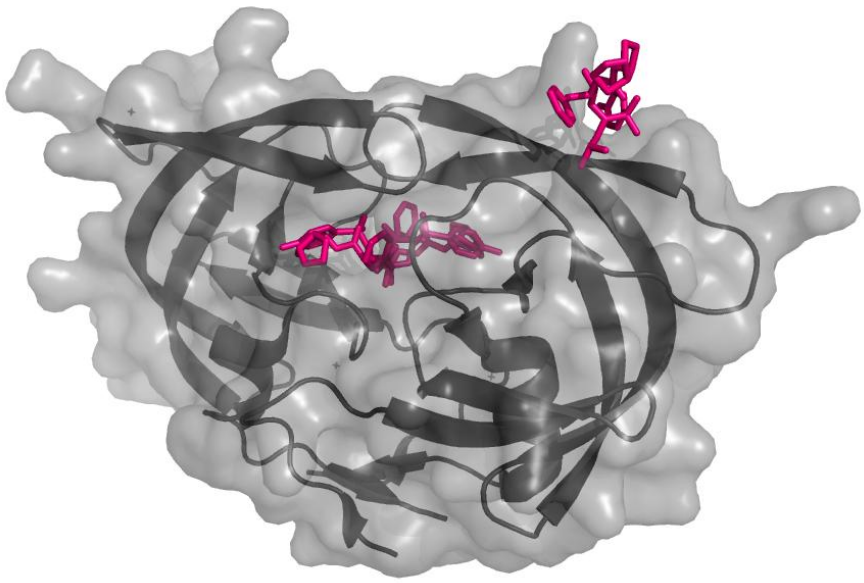


Figure 24 Ribbon representation of HIV protease showing the second inhibitor binding site

The two Darunavir molecular are shown in pink. The second Darunavir binds in a cleft on the surface of one of the flaps. This figure was generated using PyMOL (PDB code 2HS1) (DeLano, 2006).

obtained from the model is theoretical, it re-emphasises the possibility of L38↑N↑L protease having an allosteric site, as this related to the data for K_M determination. The binding of the second inhibitor on the surface induces a conformational change in the protease (Kovalevsky *et al*., 2006). If L38↑N↑L protease does have a second binding site this could have major implications on drug binding. In the 2HS1 study however, there was no mention on how the K_M was obtained and what shape the graph had. This could have helped confirm the shape of the curve obtained for L38↑N↑L protease.

Conclusion

To answer the question previously stated HIV subtype C virus has indeed selected mutations that gave rise to local structural modifications leading to changes in function. The extent of these structural modifications is not known. The homology model of L38↑N↑L protease shows a flap hinge that is slightly altered. In addition, it suggests there is a second inhibitor binding site. A crystal structure of L38↑N↑L protease would therefore be useful for understanding this unique protease with 101 amino acids.

CHAPTER 5

5. References

Abecasis, A. B., Deforche, K., Bacheler, L. T., McKenna, P., Carvalho, A. P., Gomes, P., Vandamme, A. M. and Camacho, R. J. (2006) Investigation of baseline susceptibility to protease inhibitors in HIV-1 subtypes C, F, G and CRF02_AG. *Antivir. Ther.* **11**, 581-589.

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.

Adler, A. J., Greenfield, N. J. and Frasman, G. D. (1972) 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**, 5631-40.

Alfonso, Y. and Monzote, L. (2011) HIV protease inhibitors: Effect on opportunistic protozoan parasites. *Open. Med. Chem. J.* **5**, 40-50.

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

Baca, M. and Kent, S. B. (1993) Catalytic contribution of flap-substrate hydrogen bonds in "HIV-1 protease" explored by chemical synthesis. *Proc. Natl. Acad. Sci. U S A* **90**, 11638-42.

Baeuerle, P. A. (1991) The inducible transcription activator NF-kappa B: regulation by distinct protein subunits. *Biochim. Biophys. Acta.* **1072**, 63-80.

Becker-Pergola, G., Kataaha, P., Johnston-Dow, L., Fung, S., Jackson, J. B. and Eshleman, S. H. (2000) Analysis of HIV-1 protease and reverse transcriptase in antiretroviral drug naive Ugandan adults. *AIDS. res. Hum. Retroviruses.* **16**, 807-813.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* **112**, 535-42.

Beschiaschvili, G. and Baeuerle, H. D. (1991) Effective charge of melittin upon interaction with POPC vesicles. *Biochim Biophys Acta* **1068**, 195-200.

Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M. and van der Noordaa, J. (1990) Rapid and simple method for purifi- cation of nucleic acids. *J. Clin. Microbiol.* **28**, 495–503.

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

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.

Brown, P. O., Bowerman, B., Varmus, H. E. and Bishop, J. M. (1987) Correct integration of retroviral DNA in vitro. *Cell* **49**, 347-356.

Burkrinsky, M. I., Sharova, N., Dempsey, M. P., Stanwick, T. L., Burkrinskaya, A. G., Haggerty, S. and Stevenson, M. (1992) Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc. Natl. Acad. Sci. U S A* **89**, 6580-6584.

Burlingham, B. T. and Widlanski, T. S. (2003) An Intuitive Look at the Relationship of Ki and IC50: A More General Use for the Dixon Plot. *J. Chem. Educ.* **80**, 214-218.

Charneau, P., Borman, A. M., Quillent, C., Guetard, D., Chamaret, S., Cohen, J., Remy, G., Montagnier, L. and Clavel, F. (1994) Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group. *Virology* **205**, 247-53.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

Chen, Y. and Barkley, M. D. (1998) Toward understanding tryptophan fluorescence in proteins. *Biochemistry* **37**, 9976-9982.

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

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

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

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

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

Coman, R. M., Robbins, A. H., Goodenow, M. M., Dunn, B. M. and McKenna, R. (2008b) High resolution structure of unbound human immunodeficiency virus subtype C protease: implications of flap dynamics and drug resistance. *Acta. Cryst.* **D64**, 754-763.

Crawford, S. and Goff, S. P. (1985) A deletion mutation in the 5' part of the pol gene of Moloney murine leukemia virus blocks proteolytic processing of the gag and pol polyproteins. *J. Virol.* **53**, 899-907.

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

Darke, P. L., Nutt, R. F., Brady, S. F., Garsky, V. M., Ciccarone, T. M., Leu, C. T., Lumma, P. K., Freidinger, R. M., Veber, D. F. and Sigal, I. S. (1988) HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins. *Biochem. Biophys. Res. Commun.* **156**, 297-303.

Das, A., Mahale, S., Prashar, V., Bihani, S., Ferrer, J.-L. and Hosur, M. V. (2010) X-ray snapshot of HIV-1 protease in action: Observation of tetrahedral intermediate and short ionic hydrogen bond SIHB with catalytic aspartate. *J. Am. Chem. Soc.* **132**, 6366-6373.

De Clercq, E. (2007) Anti-HIV drugs. *Verh K Acad Geneeskd Belg* **69**, 81-104.

De Clercq, E. (2009) Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int. J. Antimicrob. Agents.* **33**, 307-320.

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. U S A* **84**, 8903-6.

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

San Carlos, CA.

Descamps, D., Apetrei, C., Collin, G., Damond, F., Simon, F. and Brun-Vezinet, F. (1998) Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. *AIDS* **12**, 1109–1111.

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

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

Dunn, B. M. (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.* **102**, 4431-4458.

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

Essex, M. (1999) Human immunodeficiency viruses in the developing world. *Adv. Virus. Res.* **53**, 71-88.

Farmerie, W. G., Loeb, D. D., Casavant, N. C., Hutchison, C. A., 3rd, Edgell, M. H. and Swanstrom, R. (1987) Expression and processing of the AIDS virus reverse transcriptase in Escherichia coli. *Science* **236**, 305-8.

Farnet, C. M. and Haseltine, W. A. (1991) Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J. Virol.* **65**, 1910-1515.

Fish, W. W., Reynolds, J. A. and Tanford, C. (1970) Gel chromatography of proteins in denaturing solvents. Comparison between sodium dodecyl sulfate and guanidine hydrochloride as denaturants. *J. Biol. Chem.* **245**, 5166-8.

Fitzgerald, P. M., McKeever, B. M., VanMiddlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C. T., Herber, W. K., Dixon, R. A. and Darke, P. L. (1990) Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-A resolution. *J. Biol. Chem.* **265**, 14209-19.

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

Frater, A. J., Beardall, A., Ariyoshi, K., Churchill, D., Galpin, S., Clarke, J. R., Weber, J. N. and McClure, M. O. (2001) Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients. *AIDS* **15**, 1493-502.

Frater, A. J., Dunn, D. T., Beardall, A. J., Ariyoshi, K., Clarke, J. R., McClure, M. O. and Weber, J. N. (2002) Comparative response of African HIV-1-infected individuals to highly active antiretroviral therapy. *AIDS* **16**, 1139-46.

Freedberg, D. I., Ishima, R., Jacob, J., Wang, Y. X., Kustanovich, I., Louis, J. M. and Torchia, D. A. (2000) Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci.* **11**, 221-232.

Gallo, R. C. and Montagnier, L. (1988) AIDS in 1988. *Sci. Am.* **259**, 41-8.

Gao, F., Yue, L., Robertson, D. L., Hill, S. C., Hui, H., Biggar, R. J., Neequaye, A. E., Whelan, T. M., Ho, D. D., Shaw, G. M. et al. (1994) Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J. Virol.* **68**, 7433-47.

Giam, C. Z. and Boros, I. (1988) In vivo and in vitro autoprocessing of human immunodeficiency virus protease expressed in Escherichia coli. *J. Biol. Chem.* **263**, 14617-20.

Gonda, M. A., Wong-Staal, F., Gallo, R. C., Clements, J. E., Narayan, O. and Gilden, R. V. (1985) Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus. *Science* **227**, 173-7.

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

Greenfield, N. J. (1999) Applications of circular dichroism in protein and peptide analysis. *Trends Anal. Chem.* **18**, 236-244.

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

Hemelaar, J., Gouws, E., Ghys, P. D. and Osmanov, S. (2006) Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* **20**, W13-23.

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.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. and Markowitz, M. (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123-6.

Ho, D. D., Toyoshima, T., Mo, H., Kempf, D. J., Norbeck, D., Chen, C. M., Wideburg, N. E., Burt, S. K., Erickson, J. W. and Singh, M. K. (1994) Characterisation of human immunodeficiency virus type 1 variants with increased resistance to a C2-symmetric protease inhibitor. *J. Virol.* **68**, 2016-2020.

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

Hornak, V., Okur, A., Rizzo, R. C. and Simmerling, C. (2006a) HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. *J. Am. Chem. Soc.* **128**, 2812-3.

Hornak, V., Okur, A., Rizzo, R. C. and Simmerling, C. (2006b) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. U S A* **103**, 915-20.

Hyland, L. J., Tomaszek, T. A., Jr. and Meek, T. D. (1991a) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry* **30**, 8454-63.

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

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

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

Johnson, V. A., Brun-Vezinet, F., Clotet, B., Conway, B., Kuritzkes, D. R., Pillay, D., Schapiro, J., Telenti, A. and Richman, D. (2005) Update of the Drug Resistance Mutations in HIV-1: 2005. *Topic. HIV Med.* **13**, 51-7.

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

Kantor, R., Katzenstein, D. A., Efron, B., Carvalho, A. P., Wynhoven, B., Cane, P., Clarke, J., Sirivichayakul, S., Soares, M. A., Snoeck, J. et al. (2005) Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. *PLoS. Med.* **2**, e112.

Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Odaka, T. and Oroszlan, S. (1985) Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity. *Virology* **145**, 280-92.

Kempf, D. J., Marsh, K. C., Kumar, G., Rodrigues, A. D., Denissen, J. F., McDonald, E., Kukulka, M. J., Hsu, A., Granneman, G. R., Baroldi, P. A. et al. (1997) Pharcokinetic enhancement of inhibitors of the human immunodeficiency virus by coadministration with ritonavir. *Antimicrob. Agents. Chemother.* **41**, 654-660.

Kim, E. Y., Winters, M. A., Kagan, R. M. and Merigan, T. C. (2001) Functional correlates of insertion mutations in the protease gene of human immunodeficiency virus type 1 isolates from patients. *J. Virol.* **75**, 11227-33.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M. and Sigal, I. S. (1988) Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U S A* **85**, 4686-90.

Kovalevsky, A. Y., Liu, F., Leshchenko, S., Ghosh, A. K., Louis, J. M., Harrison, R. W. and Weber, I. T. (2006) Ultra-high resolution crystal structure of HIV-1 protease mutant reveals two binding sites for clinical inhibitor TMC114. *J. Mol. Biol.* **363**, 161-73.

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

Kuroda, M. J., El-Farrash, M. A. and Harada, S. (1995) Impaired infectivity of HIV-1 after a single point mutation in the pol gene to escape the effect of a protease inhibitor in vitro. *Virology* **210**, 212-216.

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

Lakowicz, J. R. (2006) *Principles of fluorescence spectroscopy.*, New York, USA: Springer.

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

Leitner, T., Korber, B., Daniels, M., Calef, C. and Foley, B. (2005) HIV-1 subtype and circulating recombinant form (CRF) reference sequences 2005. *HIV sequence compendium*

Leonard, J. M. (1996) Perspectives in HIV protease inhibitors. *Antiviral. Chem. Chemother.* **4**, 319-325.

Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A. and Axel, R. (1986) The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**, 333-348.

Mammano, F., Petit, C. and Clavel, F. (1998) Resistance associated loss of viral fitness in human immunodeficiency virus type 1: Phenotypic analysis of protease and gag coevolution in protease inhibitor treated patients. *J. Virol.* **72**, 7632-7637.

Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291-325.

Masquelier, B., Race, E., Tamalet, C., Descamps, D., Izopet, J., Buffet-Janvresse, C., Ruffault, A., Mohammed, A. S., Cottalorda, J., Schmuck, A. et al. (2001) Genotypic and phenotypic resistance patterns of human immunodeficiency virus type 1 variants with insertions or deletions in the reverse transcriptase (RT): multicenter study of patients treated with RT inhibitors. *Antimicrob. Agents. Chemother.* **45**, 1836-42.

McKeever, B. M., Navia, M. A., Fitzgerald, P. M., Springer, J. P., Leu, C. T., Heimbach, J. C., Herbert, W. K., Sigal, I. S. and Darke, P. L. (1989) Crystallization of the aspartylprotease from the human immunodeficiency virus, HIV-1. *J. Biol. Chem.* **264**, 1919-21.

McQuade, T. J., Tomasselli, A. G., Liu, L., Karacostas, V., Moss, B., Sawyer, T. K., Heinrikson, R. L. and Tarpley, W. G. (1990) A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science* **247**, 454-6.

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

Michaelis, L., Menten, M. L., Johnson, K. A. and Goody, R. S. (2011) The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* **50**, 8264-9.

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., Smith, C. W. et al. (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.

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

Montano, M. A., Novitsky, V. A., Blackard, J. T., Cho, N. L., Katzenstein, D. A. and Essex, M. (1997) Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. *J. Virol.* **71**, 8657-65.

Mous, J., Heimer, E. P. and Le Grice, S. F. (1988) Processing protease and reverse transcriptase from human immunodeficiency virus type I polyprotein in Escherichia coli. *J. Virol.* **62**, 1433-6.

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

Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L. and Springer, J. P. (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **337**, 615-20.

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

Nutt, R. F., Brady, S. F., Darke, P. L., Ciccarone, T. M., Colton, C. D., Nutt, E. M., Rodkey, J. A., Bennett, C. D., Waxman, L. H. and Sigal, I. S. (1988) Chemical synthesis and enzymatic activity of a 99-residue peptide with a sequence proposed for the human immunodeficiency virus protease. *Proc. Natl. Acad. Sci. U S A* **85**, 7129-7133.

Ohtaka, H. and Freire, E. (2005) Adaptive inhibitors of the HIV-1 protease. . *Prog. Biophys. and Mol. Biol.* **88**, 193-208.

Palmer, T. (1995a) *Understanding enzymes*, Hertfordshire, Britain: Prentice Hall/ Ellis Horwood.

Palmer, T. (1995b) *Understanding enzymes.*, London: Prentice Hall.

Parkin, N. T. and Schapiro, J. M. (2004) Antiretroviral drug resistance in non-subtype B HIV-1, HIV-2 and SIV. *Antivir. Ther.* **9**, 3-12.

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

Piacenti, F. J. (2006) An update and review of anitretroviral therapy. *Pharmacotherapy* **26**, 1111-1133.

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

Prabu-Jeyabalan, M., Nalivaika, E. and Schiffer, C. A. (2000) How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* **301**, 1207-20.

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

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. *J. virol.* **80**, 3607-16.

Ratner, L. (1993) HIV life cycle and genetic approaches. *Perspect. Drug. Discov.* **1**, 3-22.

Reeves, J. D. and Doms, R. W. (2002) Human immunodeficiency virus type 2. *J. Gen. Virol.* **83**, 1253-65.

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.

Richman, D. D., Havlir, D., Corbeil, J., Looney, D., Ignacio, C., Spector, S. A., Sullivan, J., Cheeseman, S., Barringer, K., Pauletti, D. et al. (1994) Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy. *J. Virol.* **68**, 1660-6.

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

Roberts, N. A., Martin, J. A., Kinchington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B., Galpin, S. A., Handa, B. K., Kay, J., Krohn, A. et al. (1990) Rational design of peptide-based HIV proteinase inhibitors. . *Science* **248**, 358–361.

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C. et al. (2000) HIV-1 nomenclature proposal. *Science* **288**, 55-6.

Robertson, D. L., Hahn, B. H. and Sharp, P. M. (1995a) Recombination in AIDS viruses. *J. Mol. Evol.* **40**, 249-59.

Robertson, D. L., Sharp, P. M., McCutchan, F. E. and Hahn, B. H. (1995b) Recombination in HIV-1. *Nature* **374**, 124-6.

Robins, T. and Plattner, J. (1993) HIV protease inhibitors: their anti-HIV activity and potential role in treatment. *J. Acquir. Immune. Defic. Syndr.* **6**, 162-170.

Roques, P., Menu, E., Narwa, R., Scarlatti, G., Tresoldi, E., Damond, F., Mauclere, P., Dormont, D., Chaouat, G., Simon, F. et al. (1999) An unusual HIV type 1 env sequence embedded in a mosaic virus from Cameroon: identification of a new env clade. European Network on the study of in utero transmission of HIV-1. *AIDS. Res. Hum. Retroviruses.* **15**, 1585-9.

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.

SANAC (2011) National strategic plan on HIV, STIs and TB (2012-2016 summary). *South African National AIDS Council*

Sawant, R. L., Bhatia, M. S., Sawant, M. R. and Wadekar, J. B. (2008) Retroviral proteases: A potential target for development of antiviral agents. *Curr. Trends. Biotechnol. Pharm.* **2**, 1333-141.

Schagger, H. (2006) Tricine-SDS-PAGE. *Nat. Protoc.* **1**, 16-22.

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.

Schechter, I. and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-62.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M. C. (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic. Acids. Res.* **31**, 3381-5.

Scott, W. R. and Schiffer, C. A. (2000) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure* **8**, 1259-65.

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. *Proc. Natl. Acad. Sci. U S A* **85**, 6612-6.

Shafer, R. W. (2002) Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clin. Microbiol. Rev.* **15**, 247-277.

Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F. and Brun-Vezinet, F. (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* **4**, 1032-7.

Soding, J., Biegert, A. and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic. Acids. Res.* **33**, W244-8.

Sturmer, M., Staszewski, S., Doerr, H. W. and Hertogs, K. (2003) A 6-base pair insertion in the protease gene of HIV type 1 detected in a protease inhibitor-naive patient is not associated with indinavir treatment failure. *AIDS. Res. Hum. Retroviruses.* **19**, 967-8.

Subbarao, S. and Schochetman, G. (1996) Genetic variability of HIV-1. *AIDS* **10 Suppl A**, S13-23.

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.

Tie, Y., Boross, P. I., Wang, Y. F., Gaddis, L., Liu, F., Chen, X., Tozser, J., Harrison, R. W. and Weber, I. T. (2005) Molecular basis for substrate recognition and drug resistance from 1.1 to 1.6 angstroms resolution crystal structures of HIV-1 protease mutants with substrate analogs. *FEBS J* **272**, 5265-77.

Todd, M. and Freire, E. (1999) The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins* **36**, 147-156.

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

Todd, M. J., Luque, I., Velazquez-Campoy, A. and Freire, E. (2000) The thermodynamic basis of resistance to HIV-1 protease inhibition. Calorimetric analysis of the V82F/I84V active site resistant mutant. *Biochemistry* **39**, 11876-11883.

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

Tomasselli, A. G. and Heinrikson, R. L. (2000) the HIV-protease in AIDS therapy: a current clinical perspective. *Biochim. Biophys. Acta.* **1477**, 189-214.

Tramontano, A., Leplae, R. and Morea, V. (2001) Analysis and assessment of comparative modeling predictions in CASP4. *Proteins* **Suppl 5**, 22-38.

Ullrich, B., Laberge, M., Tölgyesi, F., Szeltner, Z., Polgár, L. and Fidy, J. (2000) Trp42 rotamers report reduced flexibility when the inhibitor acetyl-pepstatin is bound to HIV-1 protease. . *Protein Sci.* **9**, 2232-2245.

UNAIDS (2010) Global report: UNAIDS report on the global epidemic 2010. *Joint United Nations Programme on HIV/AIDS (UNAIDS)*

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

Velazquez-Campoy, A., Kiso, Y. and Freire, E. (2001a) The binding energetics of first and second generation HIV-1 protease inhibitors: implications for drug design. *Arch. Biochem. Biophys.* **390**, 169-175.

Velazquez-Campoy, A., Muzammil, S., Ohtaka, H., Shon, A., Vega, S. and Freire, E. (2003a) 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. (2001b) Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc Natl Acad Sci U S A* **98**, 6062-7.

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

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

Venyaminov, S. Y., Baikalov, I. A., Shen, Z. M., Wu, C. S. and Yang, J. T. (1993) Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set. *Anal. Biochem.* **214**, 17-24.

Venyaminov, S. Y. and Yang, J. T. (1991) Some problems of CD analysis of protein conformation. *Anal. Biochem.* **191**, 250-255.

Wei, X., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifson, J. D., Bonhoeffer, S., Nowak, M. A., Hahn, B. H. et al. (1995) Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**, 117-22.

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

Wharton, C. W. and Eisenthal, R. (1981) *Molecular enzymology.* , Glasgow and London: Blackie.

WHO (2011) World health statistics. *World Health Organisation*

Winters, M. A. and Merigan, T. C. (2005) Insertions in the human immunodeficiency virus type 1 protease and reverse transcriptase genes: clinical impact and molecular mechanisms. *Antimicrob. Agents. Chemother.* **49**, 2575-82.

Witte, O. N. and Baltimore, D. (1978) Relationship of retrovirus polyprotein cleavages to virion maturation studied with temperature-sensitive murine leukemia virus mutants. *J. Virol.* **26**, 750-61.

Wlodawer, A. and Erickson, J. W. (1993) Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Med.* **62**, 543-585.

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. Enzymol.* **246**, 34-71.

Yamaguchi, J., Devare, S. G. and Brennan, C. A. (2000) Identification of a new HIV-2 subtype based on phylogenetic analysis of full-length genomic sequence. *AIDS. Res. Hum. Retroviruses.* **16**, 925-30.

Zhang, Y. M., Imamichi, H., Imamichi, T., Lane, H. C., Falloon, J., Vasudevachari, M. B. and Salzman, N. P. (1997) Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J. Virol.* **71**, 6662-6670.

Conference output

Cross faculty symposium, 26 October 2010, University of the Witwatersrand, Johannesburg (poster presentation).

Authors: Maputsoe, X., Achilonu, I., Dirr, H. and Sayed, Y.

Title: Impact of L38_NL insertion mutation on structure, function and stability of HIV-1 South African subtype C protease.