

https://dspace.org/

# Structural and functional effects of an I36TT insertion in the South African HIV-1 subtype C protease

**Mpye, Keleabetswe Lerato**

**2010-09-07**
http://hdl.handle.net/10539/8617

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

# Structural and Functional Effects of an I36TT Insertion in the South African HIV-1 Subtype C Protease

**Keleabetswe Lerato Mpye**

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, 2010

# DECLARATION

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

_____

Keleabetswe Lerato Mpye

_____ day of _____, 2010.

# ABSTRACT

Human immunodeficiency virus type 1 (HIV-1) is the cause of an estimated 1.9 million deaths in the world annually. The HI virus has three enzymes, notably; reverse transcriptase, integrase and protease, which are crucial for its maturation and continued infection. HIV relies on the catalytic efficacy of the protease enzyme for cleaving precursor polyproteins to yield structural and functional proteins. Inhibition of the viral protease proffers major therapeutic benefits in the battle against HIV. However, the use of protease inhibitors in HIV regimens is limited by the emergence of drug resistant mutations in the protease coding region. Drug resistant mutations are characterized by amino acid substitutions, deletions, and/or insertions in the viral sequence. Thus, this study set out to assess structural and functional characteristics of an HIV-1 South African subtype C (C-SA) protease with a single amino acid substituted and another inserted at codon 36, i.e., I36TT, with respect to the wild-type HIV-1 C-SA protease. An I36TT protease was generated with its background polymorphisms (P39S, D60E, and Q61E), over-expressed, and purified. Secondary and tertiary structural properties of the wild-type and the variant protease were evaluated using far-UV circular dichroism and fluorescence spectroscopy. Both proteases exhibited typical secondary structural features of a predominantly β-sheeted protein, indicated by circular dichroism spectra with minima at 216 nm. Using intrinsic tryptophan as a probe, the tertiary structures of both proteases revealed that the local structural environments of both proteases had not been perturbed. This was indicated by fluorescence emission intensity peak at 355 nm. Proteolytic efficiency of the protease enzymes was evaluated following hydrolysis of a synthetic chromogenic HIV substrate mimicking the conserved protease cleavage site in the *gag-pol* polyprotein precursor. A comparison of the kinetic properties of the enzymes indicated that the I36TT variant protease has a slightly (7%) enhanced catalytic activity relative to the wild-type HIV-1 C-SA protease. Enzymatic parameters of the two proteases in the presence of various protease-inhibitors showed that both proteases' catalytic activity is highly affected by saquinavir with $IC_{50}$ values of 7.6 nM and 6.3 nM for the wild-type and the variant, respectively. The variant protease enzyme, compared to the wild-type, appears to have acquired resistance towards indinavir with $IC_{50}$ value of 16.2 nM and 9.5 nM for the variant and wild-type, respectively. In the presence of

ritonavir, the variant protease ($IC_{50}$ value of 36.4 nM) proved to have retained most of its enzymatic characteristic as compared to the wild-type protease ($IC_{50}$ value of 19.1 nM). The wild-type and variant protease enzymes showed similar susceptibility ($IC_{50}$ value of 17.3 and 16.8 nM for wild-type and variant, respectively) to nelfinavir. Thermodynamic analysis of the protease enzymes indicated that the I36TT mutation does not affect the binding energetics of acetyl pepstatin to the protease. Thus, vitality studies suggest that the I36TT substitution/insertion mutation and background polymorphisms were incorporated in the HIV-1 C-SA protease enzyme to improve viral replication rate.

**This work is dedicated to:**

**My mother Albertinah “Dinah”,**

**without whom none of this would have been possible.**

**To my sisters: Tebogo “Sonia” and Nompendulo,**

**my brothers: Thapelo and Zweli,**

**who always encourage and support me in all I do.**

“Correction does much, but encouragement does more”

Johann Wolfgang von Goethe

# ACKNOWLEDGEMENTS

My supervisors Dr Yasien Sayed, Dr Salerwe Mosebi and Prof. Heini Dirr for their patience, guidance and the frequent discussions

Dr Salerwe Mosebi for purifying the protease inhibitors

Johanna Ledwaba, Prof. Lynn Morris and everyone from the AIDS research unit at the NICD, Johannesburg, for allowing me to work in their lab and giving me the sequence of the variant protease used in this study

Members of the Protein Structure-Function Research Unit, University of the Witwatersrand, Johannesburg, for being there whenever I needed help and the good, loud laughs

National Research Foundation and Carnegie for funding this project

My mother and my “big family” for their constant support and believing in me

My “Vegas crew” for those crazy weekends. They helped me maintain my sanity

My friends and everyone else for their contributions

Last but not least, I thank the good gracious Lord for being with me and keeping me safe from conception to date.

# TABLE OF CONTENTS

**DECLARATION** ..... **ii**

**ABSTRACT** ..... **iii**

**ACKNOWLEDGEMENTS** ..... **vi**

**TABLE OF CONTENTS** ..... **vii**

**LIST OF FIGURES** ..... **x**

**LIST OF TABLES** ..... **xi**

**ABBREVIATIONS** ..... **xii**

**CHAPTER 1** ..... **1**

**1 INTRODUCTION** ..... **1**

**1.1 Human immunodeficiency virus** ..... **1**

**1.2 Structure of the HI virus** ..... **1**

**1.3 HIV-1 protease structure and function** ..... **3**

**1.4 HIV therapy and drug resistant mutations** ..... **8**

**1.5 Insertion mutations** ..... **10**

**1.6 Objectives** ..... **11**

**CHAPTER 2** ..... **14**

**2 MATERIALS AND EXPERIMENTAL PROTOCOLS** ..... **14**

**2.1 HIV-1 expression vector** ..... **14**

**2.2 Reagents** ..... **14**

**2.3 HIV-1 South African subtype C variant protease sequence** ..... **15**

**2.4 Generation of variant plasmid DNA** ..... **15**

2.4.1 Mutagenic primer design ..... 15

2.4.2 Generation of the I36TT substitution/insertion and the background polymorphisms ..... 17

**2.5 Transformation of Escherichia coli cells with plasmid DNA ..... 19**

**2.6 Over-expression, extraction and purification of the C-SA protease ..... 19**

**2.7 Protein concentration determination ..... 20**

**2.8 Sodium dodecyl sulphate polyacrylamide gel electrophoresis ..... 21**

**2.9 Purification of the protease inhibitors ..... 22**

**2.10 Calorimetric assays ..... 22**

2.10.1 Isothermal titration calorimetry ..... 22

**2.11 Protease characterization ..... 26**

2.11.1 Structural assessment ..... 26

2.11.1.1 Secondary structure ..... 26

2.11.1.2 Tertiary structure ..... 27

2.11.2 Spectrophotometric analyses ..... 28

2.11.2.1 Protease enzyme kinetics ..... 28

2.11.2.2 Inhibition studies ..... 30

**2.12 Data analysis ..... 30**

**CHAPTER 3 ..... 31**

**3 RESULTS ..... 31**

**3.1 Plasmid DNA sequence confirmation ..... 31**

**3.2 Protease purity and monomeric molecular weight determination ..... 31**

**3.3 Protein concentration determination ..... 34**

**3.4 Structural characterization ..... 34**

3.4.1 Secondary structural content assessment ..... 34

3.4.2 Tertiary structure determination ..... 37

**3.5 Enzymatic characterization ..... 37**

3.5.1 Catalytic activity studies ..... 37

3.5.2 Inhibition kinetics ..... 46

**3.6 Thermodynamic characteristics ..... 46**

**CHAPTER 4 ..... 54**

**4 DISCUSSION ..... 54**

**4.1 Structural content assessment ..... 54**

4.1.1 Secondary structure ..... 54

4.1.2 Tertiary structure ..... 55

**4.2 Enzymatic characteristics ..... 56**

4.2.1 Catalytic properties ..... 56

4.2.2 Inhibition by ARVs ..... 58

**4.3 Conclusion ..... 59**

**CHAPTER 5 ..... 60**

**5 REFERENCES ..... 60**

# LIST OF FIGURES

**Figure 1: Structural arrangement of the *gag-pol* polyprotein precursor in HIV-1 .... 2**
**Figure 2: HIV life-cycle ..... 4**
**Figure 3: Ribbon representation of a ligand-bound homodimeric HIV-1 protease ... 5**
**Figure 4: Mechanism of peptide bond hydrolysis by aspartic protease ..... 7**
**Figure 5: Chemical structures of the HIV-1 protease inhibitors..... 9**
**Figure 6: Ribbon representation of the crystal structure of a free (unbound) conformation of HIV-1 protease..... 12**
**Figure 7: Schematic representation of ITC cells..... 23**
**Figure 8: HIV-1 binding isotherms ..... 25**
**Figure 9: Variant protease sequence confirmation ..... 32**
**Figure 10: HIV-1 C-SA protease purity and molecular weight determination ..... 33**
**Figure 11: HIV-1 C-SA absorption spectrum ..... 35**
**Figure 12: Far-UV CD spectra of HIV-1 C-SA protease ..... 36**
**Figure 13: Intrinsic tryptophan fluorescence emission spectra of the HIV-1 C-SA protease ..... 38**
**Figure 14: Progress curve for HIV-1 substrate hydrolysis..... 39**
**Figure 15: Determination of specific activity ..... 40**
**Figure 16: Wild-type protease catalytic constant and maximum velocity determination ..... 41**
**Figure 17: Variant protease catalytic constant and maximum velocity determination ..... 42**
**Figure 18: Determination of enzyme turn-over number..... 43**
**Figure 19: Catalytic efficiency determination..... 44**
**Figure 20: Determination of $IC_{50}$ value for saquinavir ..... 47**
**Figure 21: Indinavir $IC_{50}$ value determination ..... 48**
**Figure 22: Ritonavir $IC_{50}$ value determination..... 49**
**Figure 23: Nelfinavir $IC_{50}$ value determination ..... 50**
**Figure 24: Representation of the HIV-1 C-SA protease binding isotherm ..... 53**

# LIST OF TABLES

**Table 1: Primer sequences ..... 16**

**Table 2: PCR cycling parameters..... 18**

**Table 3: Summary of the catalytic parameters..... 45**

**Table 4: Summary of inhibitors $IC_{50}$ values..... 51**

# ABBREVIATIONS

| | |
|---|---|
| $A_{280}$ | absorbance at 280 nm |
| AIDS | acquired immunodeficiency syndrome |
| amp | ampicillin |
| ARV | antiretroviral |
| bp | base pairs |
| °C | degrees Celsius |
| $\Delta G$ | change in Gibbs free energy |
| $\Delta H$ | change in enthalpy |
| $\Delta S$ | change in entropy |
| CA | capsid protein |
| CD | circular dichroism |
| CSIR | Council for Scientific and Industrial Research |
| D60E | Asp60 (wild-type) replaced with a Glu |
| Da | Daltons |
| DEAE | diethylaminoethyl |
| $dH_2O$ | distilled water |
| DMSO | dimethylsulphoxide |
| DNA | deoxyribonucleic acid |
| Dnase | deoxyribonuclease |
| DR | drug resistant |
| $\varepsilon$ | molar extinction coefficient |

| | |
|---|---|
| EDTA | ethylenediaminetetraacetic acid |
| Far-UV CD | far-ultraviolet circular dichroism |
| FDA | food and drug administration |
| Gly-Gly | N-glycylglycine |
| HAART | highly active antiretroviral treatment |
| HIV | human immunodeficiency virus |
| HIV-1 C-SA | human immunodeficiency virus type 1 South African subtype C |
| HPLC | high performance liquid chromatography |
| $IC_{50}$ value | inhibitor concentration needed to decrease the catalytic activity of an enzyme by 50% |
| IN | integrase |
| IPTG | isopropyl-β-D-thiogalactopyranoside |
| ITC | isothermal titration calorimetry |
| I36TT | Ile at position 36 (wild-type) substituted with a Thr and another Thr residue inserted next to it |
| $K_a$ | association constant |
| $K_d$ | dissociation constant |
| $k_{cat}$ | catalytic constant |
| $k_{cat}/K_M$ | catalytic efficiency |
| kcal | kilocalories |
| $K_M$ | Michaelis constant |
| ℓ | litre |
| LB | lysogeny broth |

| | |
|---|---|
| M | molar |
| $\lambda_{max}$ | maximum fluorescence wavelength |
| MA | matrix protein |
| ME | β-mercaptoethanol |
| µcal | microcalories |
| µℓ | microlitre |
| µM | micromolar |
| mdeg | millidegrees |
| mℓ | millilitre |
| mM | millimolar |
| MRE | mean residue ellipticity |
| *N* | stoichiometry |
| NaCl | sodium chloride |
| NaOH | sodium hydroxide |
| NC | nucleocapsid protein |
| NICD | National Institute for Communicable Diseases |
| Nle | norleucine |
| nM | nanomolar |
| nPhe | *p*-nitrophenylalanine |
| OD | optical density |
| ORF | open reading frame |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |

| | |
|---|---|
| PDB | protein data bank |
| PI | protease inhibitor |
| PMSF | phenylmethylsulfonyl fluoride |
| PR | protease |
| P39S | Pro39 (wild-type) replaced with a Ser |
| Q61E | Gln61 (wild-type) replaced with a Glu |
| RNA | ribonucleic acid |
| RT | reverse transcriptase |
| SD | standard deviation |
| SDS | sodium dodecyl sulphate |
| SE-HPLC | size exclusion high performance liquid chromatography |
| TFR | transframe region |
| UV | ultraviolet |
| YT | yeast extract and tryptone powder |

The IUPAC-IUBMB one and three letter notations for the amino acids have been used.

# CHAPTER 1

# 1 INTRODUCTION

## 1.1 Human immunodeficiency virus

Human immunodeficiency virus (HIV), the causative agent of the acquired immunodeficiency syndrome (AIDS), is a retro-virus (an RNA virus) and is the cause of an estimated 1.9 million deaths in the world annually (Weiss *et al*., 1985; Gallo and Montagnier, 1988; WHO, 2008). The HI virus mutates rapidly, producing variable strains of the virus within a single host. Based on genetic similarities, the HI virus has been categorised into types, groups, and subtypes. HIV-1 is the most prevalent of the two types of HIV (HIV-1 and HIV-2) and is further classified into four groups (M, N, O and P). Of the four groups, the Major (M) group of HIV-1 is the most common and it is responsible for more than 90% of HIV infections globally. The "M" group is sub-divided into ten genetically distinct subtypes (A to H, J and K) (Robertson *et al*., 2000; Gonzales *et al*., 2001; Velazquez-Campoy *et al*., 2003). Although the B subtype was the first to be studied, as the HIV pandemic continues, more studies are focusing on subtype C since this subtype accounts for more than 56% of all new infections globally (Esparza and Bhamarapravati, 2000). The higher level of viral fitness and easy transmission of the C-subtype has been proposed to be responsible for its increased spread (Gordon *et al*., 2003; de Oliveira *et al*., 2003).

## 1.2 Structure of the HI virus

The HI virus, like most retroviruses, has three main genes; specifically, *gag*, *pol* and *env* which are expressed as zymogens (inactive polyprotein precursors) (Coffin, 1984; Luban *et al*., 1993). The zymogens are later hydrolyzed at specific sites to yield mature structural and functional proteins crucial for viral assembly and maturation, and continued infection of the virus (Vogt, 1996; Louis *et al*., 1999; Turner and Summers, 1999) (see Figure 1 for protease cleavage sites in the *gag-pol* polyprotein and the products thereof). Proteolysis of the *gag* region gives rise to the structural proteins; N-terminal matrix protein (p17), the capsid protein (p24), the

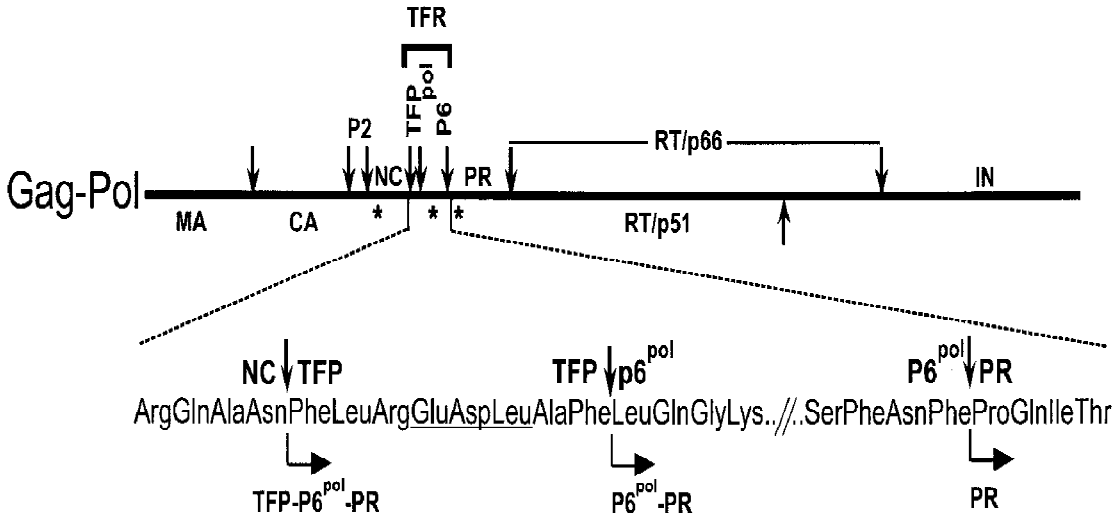



**Figure 1: Structural arrangement of the *gag-pol* polyprotein precursor in HIV-1**

Specific and additional novel viral protease cleavage sites are indicated by straight arrows and asterisks, respectively. The focused region depicts amino acids spanning the NC-TFP, TFP-6pol and p6pol-PR junctions. The transframe region (TFR) consists of the transframe octapeptide TFP, 48 amino acids of p6pol and an inhibitor of the protease (i.e., the underlined Glu-Asp-Leu tripeptide). Bracketed arrows indicate the N-termini of the protease precursor (TFP-p6pol-PR), the intermediate p6pol-PR and the mature protease. Figure adapted from Louis *et al*., 1999.

nucleocapsid (p7) and the C-terminal p6 (Luban *et al*., 1993). The two envelope glycoproteins, gp41 and gpl20, in the golgi apparatus are encoded in the *env* gene and are processed in the final stages of HIV-1 life-cycle to produce infectious virion particles (Debouck *et al*., 1987; Niu, 2008). HIV-1 *pol* gene processing yields three viral enzymes; namely, the protease (p10), reverse transcriptase (p66 and p51) and integrase (p32) crucial for viral maturation and propagation (Figure 2 summarises the contribution of these proteins in the HIV life-cycle). Maturation of the polyproteins is achieved by proteolytic processing of the genes by the protease enzyme, encoded in the *pol* gene of the retrovirus, during late stages of virion formation (Coffin, 1984; Debouck *et al*., 1987).

## 1.3 HIV-1 protease structure and function

Contrary to other cellular proteases, retroviral proteases (HIV-1 protease included) are symmetric homodimers with their substrate binding clefts formed along the interface of the dimer (Katoh *et al*., 1989; Louis *et al*., 1999). HIV-1 protease is composed of 198 amino acids (99 from each subunit) and classified under the aspartic proteinases (Richards *et al*., 1989; Dash and Rao, 2001; Trylska *et al*., 2007). The formation of the protease dimer is stabilized by the conserved intra-monomer contact between Asp29 and Arg87 residues and the active site inter-monomer contact between Asp25 and Thr26 residues (Ishima *et al*., 2007). Interface interactions between the N-terminal (amino acid 1 to 4) and C-terminal (amino acids 96 to 99) residues (see Figure 3) also play a fundamental role in dimer formation (Weber, 1990). Studies have indicated that dimer stability is also influenced by the sequences flanking the N-terminus (native transframe region) of protease domain in the *gag-pol* precursor (Louis *et al*., 1999; Ishima *et al*., 2007). The above observation was indicated by a drastic reduction of the catalytic activity of the protease prior to its maturation at the N-terminus. The active site of the HI viral protease is covered by “flaps” (flexible β-hairpins formed by residues 45 to 55 of each monomer) which control access to the active site (Trylska *et al*., 2007) (Figure 3). Flexibility of the “flaps” is suppressed when the protease is bound to an inhibitor or substrate (Ishima *et al*., 1999; Trylska *et al*., 2007).

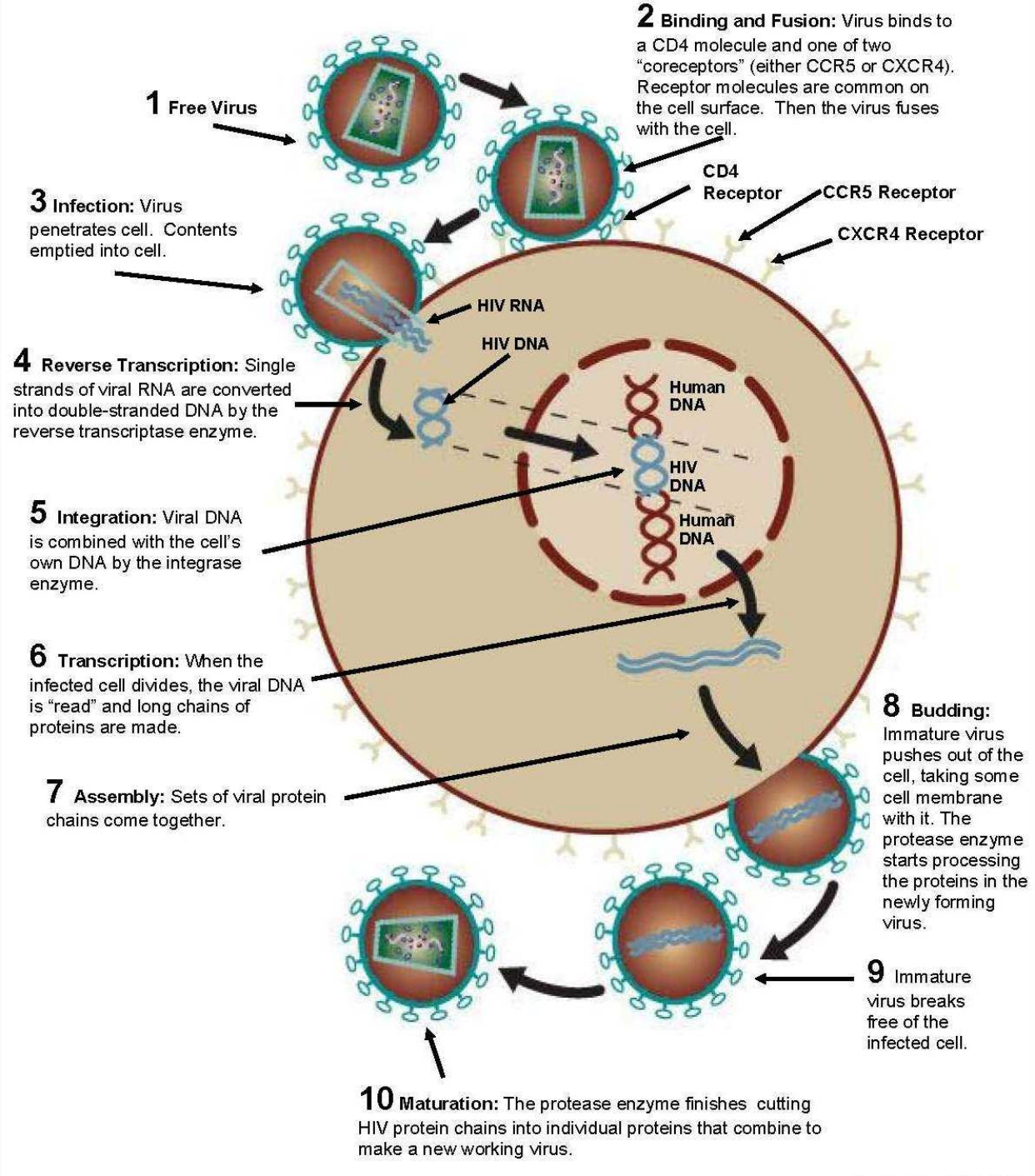



**Figure 2: HIV life-cycle**

Production of mature and infectious HI viral particles is achieved through synergistic contribution of the *gag-pol* gene products as depicted above. Picture taken from AIDSinfo.org.

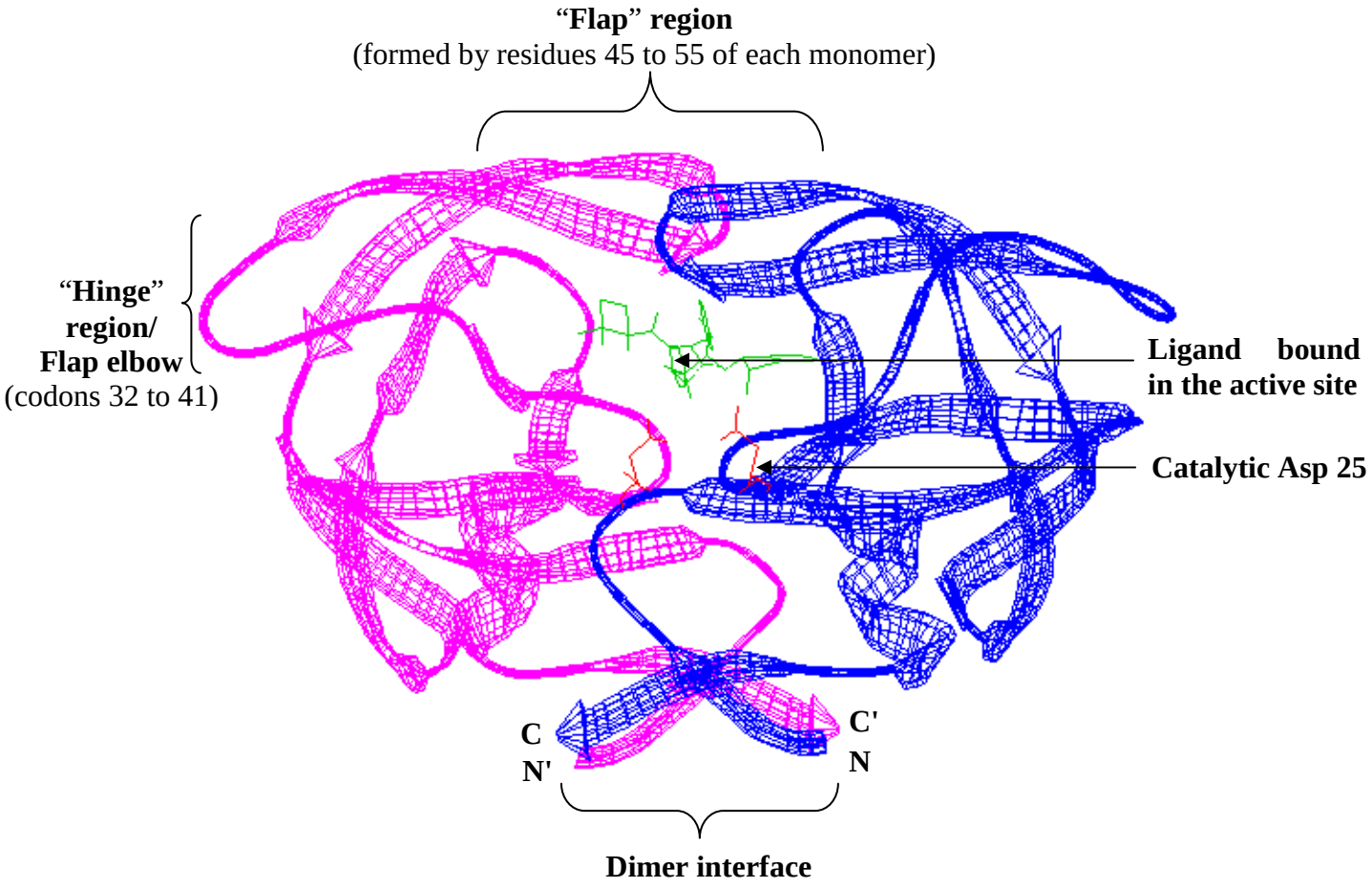



**Figure 3: Ribbon representation of a ligand-bound homodimeric HIV-1 protease**

The active site of the enzyme is formed from subunit A (in blue) and subunit B (in pink) with the two catalytic aspartic acid residues 25 and 25', one from each subunit, shown in red stick-format at the bottom of the cavity. Dimer formation is stabilised by interface interactions between the N-terminal (amino acid 1 to 4) and C-terminal (amino acids 96 to 99) residues. The above structure was generated using Swiss-pdb viewer (Guex *et al*., 1995). The PDB code used to generate this structure is 2R8N.

Based on the mechanisms they use to catalyse peptide hydrolysis, protease enzymes are divided into two classes. The first class uses an activated nucleophilic (can either be a hydroxy group or thiol) atom of an amino acid side chain to initiate hydrolysis of the amide bond carbonyl carbon of the substrate's scissile bond. The second class uses an activated water molecule to attack the carbonyl carbon of the scissile amide bond. In the former class, the activated nucleophilic atom, activated by another amino acid side chain, attacks the amide bond carbonyl of the substrate's scissile bond to yield an acyl enzyme intermediate (an ester or a thioester). The intermediate is then hydrolysed by a water molecule to yield final products. In the latter class, the water molecule is activated by either a zinc cation (the zinc metallo-proteases) or by an aspartic acid residue, Asp-$COO^-$(Figure 3) (the aspartyl proteases) at the active site. Aspartyl proteases have a conserved Asp-Thr-Gly motif within the active site/substrate binding cleft, located at positions 25 to 27 within the HIV-1 protease, which is implicated in the activity of the HIV protease (Miller *et al*., 1989; Dash and Rao, 2001). Figure 4 summarizes the catalytic mechanism of aspartyl proteases, HIV-1 protease included.

Okimoto *et al*. (2000) demonstrated that HIV-1 protease hydrolysis mechanism consists of three crucial reactions; namely, the reaction of the formation of the amide hydrate intermediate, the reaction of the protonation of the proline nitrogen of the substrate and the reaction of the C-N bond cleavage of the substrate. The second reaction being the rate-limiting step because of the presence of proline in the amide hydrate intermediate, meaning that proline would cause the streric hindrance for the protonation reaction. Studies have demonstrated that replacing the catalytic Asp25 residue with an asparagine residue leads to inability of the HIV-1 protease to cleave its substrate (Debouck *et al*., 1987; Seelmeier *et al*., 1988) and reduced stability of the protein during folding (Ishima *et al*., 2007).

Biochemical and genetic characterization of the HIV genome revealed that viral genomes with malfunctioning protease yield immature and noninfectious virion particles, rendering the virus harmless (Debouck *et al*., 1987; Kohl *et al*., 1988). For this crucial role played by the protease in the HIV-1 life-cycle, extensive studies are continuously conducted to design drugs that will bind to the active site of the protease enzyme and inhibit it.

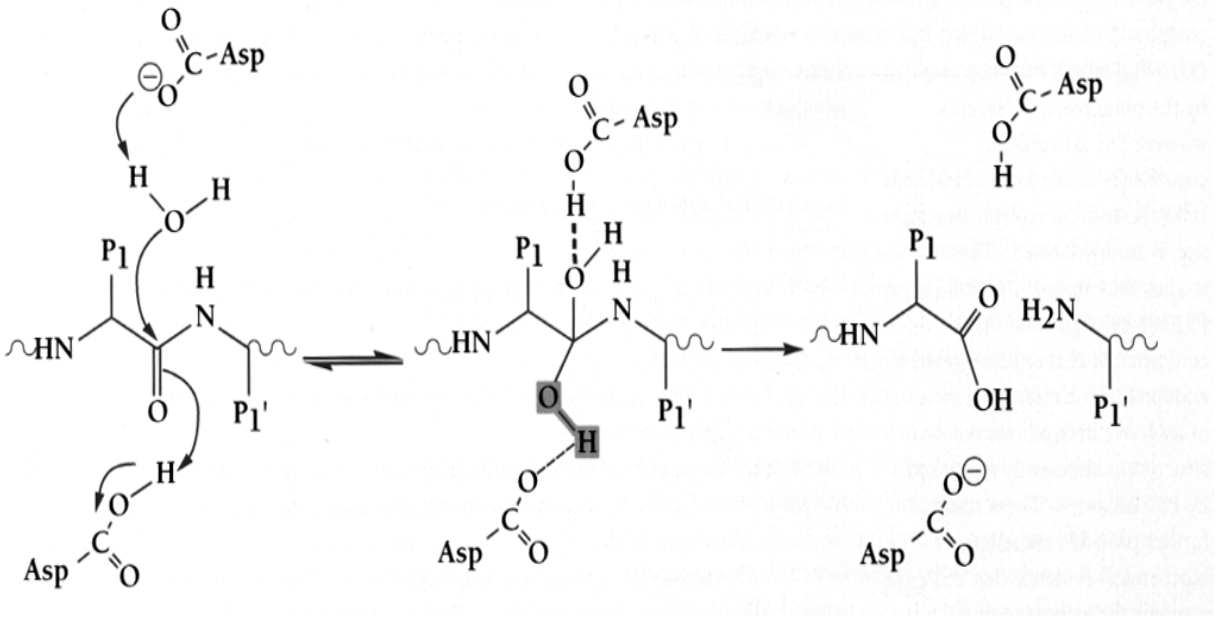



**Figure 4: Mechanism of peptide bond hydrolysis by aspartic protease**

Two aspartic acid residues are located on opposite sides of the peptide bond to be catalyzed in the protease active site. One aspartic acid residue (Asp-$COO^-$) acts as general base to activate the water molecule (H-OH) which will in turn attack the carbonyl carbon of the scissile amide bond of the substrate. The second aspartic acid residue (Asp-COOH) acts as general acid to protonate the leaving amine product. Adapted from Walsh, 2003.

## 1.4 HIV therapy and drug resistant mutations

Studying and understanding the structure and life-cycle of the HI virus led to development of potential drugs that target specific viral proteins. HIV treatment employs drugs that inhibit the catalytic properties of the major enzymes (protease, integrase, and reverse transcriptase) which the HI virus uses for its continued infection. The regimen is termed HAART and uses a combination of drugs (e.g., protease inhibitors, integrase inhibitors and reverse transcriptase inhibitors) which suppress viral multiplication and ultimately, improve the host's immune response. Studies have revealed that inhibition of the viral protease offers major therapeutic benefits in the battle against HIV (Debouck and Metcalf, 1990) since the enzyme plays a crucial role in the viral life-cycle. A large number of currently used antiviral drugs target and inhibit the catalytic activities of this viral enzyme. However, effective treatment of HIV-1 infected patients is often hindered by the rapid emergence of drug-resistant mutants with reduced susceptibility to inhibition by antiretroviral drugs used in HIV-1 treatment (Condra *et al*., 1995; Schinazi *et al*., 1997; Lorenzi *et al*., 1999; Rodriguez-Barrios and Gago, 2004).

Drug resistant mutations arise as a result of high viral replication rate, lack of post-transcriptional modification (i.e., proof-reading) of the RT, and poor patient compliance to ARV treatment, amongst others (Kantor *et al*., 2004, Wensing *et al*., 2005; Cozzi-Lepri *et al*., 2007). In some cases, drug resistant mutations have also demonstrated to be partly accounted for by transmission of the infection from drug-experienced patients to uninfected individuals (Wensing *et al*., 2005). Additionally, Gulnik *et al*. (1995) also showed the possibilities of emergence of cross resistance to other protease inhibitors from exposure to just one drug. The structure-based drug design combines computational biochemistry and structural biochemical methods to develop new lead compounds. The lead compounds are designed in a way that they have improved potency and specificity towards the targeted protein (Wlodawer and Vondrasek, 1998). Thus, second generation inhibitors (Figure 5) of the HIV-1 protease were designed (by optimizing favorable polar interactions with main chain atoms and to conserved residues, and by reducing the size of the inhibitor's hydrophobic groups) to overcome resistance (Kovalevsky *et al*., 2006).

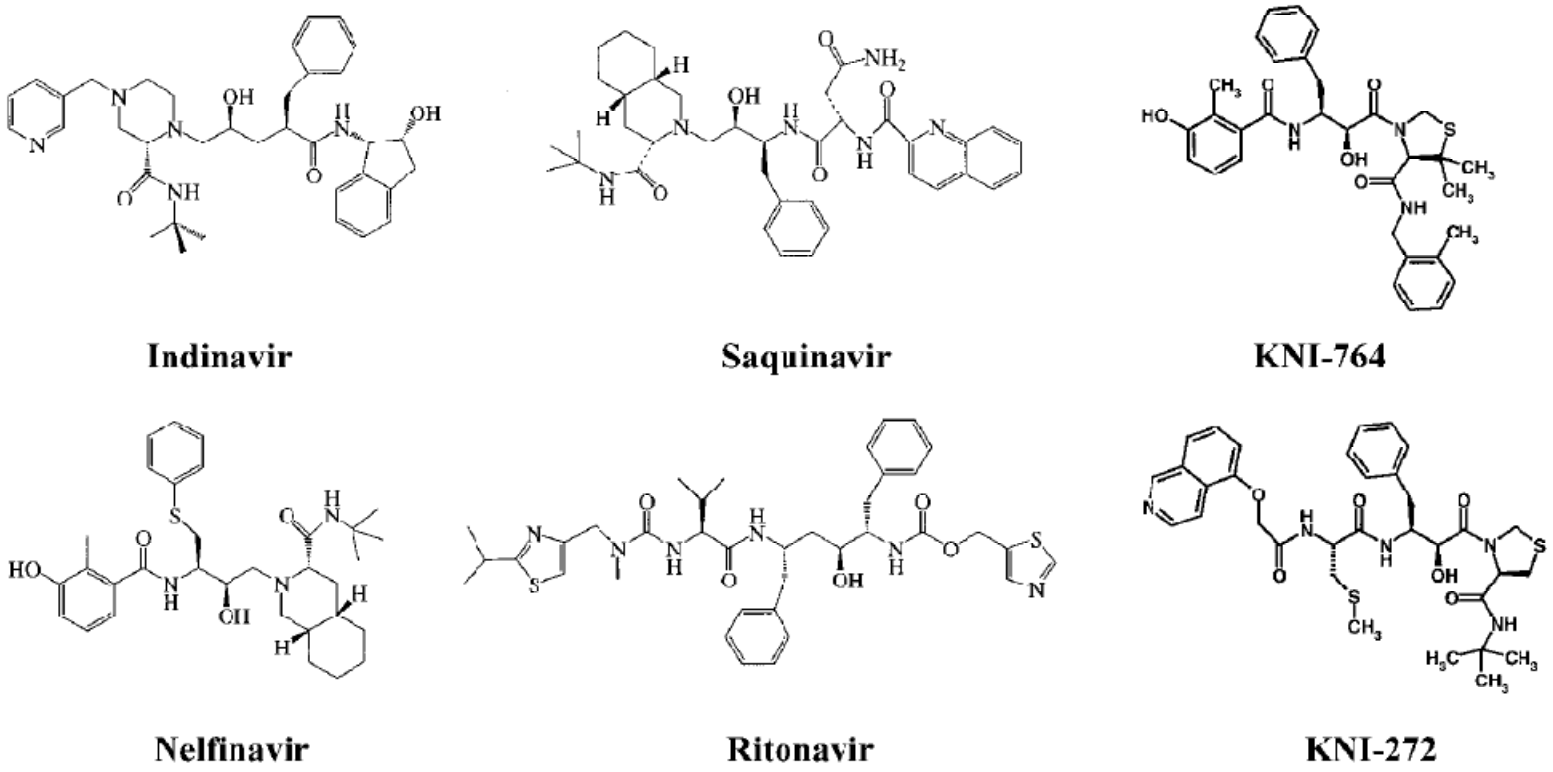



**Figure 5: Chemical structures of the HIV-1 protease inhibitors**

The first generation inhibitors (Indinavir, Saquinavir, Nelfinavir and Ritonavir) of HIV PR were amongst the first drugs developed using structure-based drug design. The second generation inhibitors (KNI-764 and KNI-272) were designed to overcome resistance due to DR mutations arising from exposure to the first generation inhibitors. The second generation inhibitors bind to the active site with improved specificity. The PIs bind at the active site of the PR. Picture taken from Velazquez-Campoy *et al.*, 2001b.

DR mutations are characterized by amino acid substitutions, deletions, and/or insertions in the viral sequence. These mutations frequently emerge in the reverse transcriptase and protease enzyme of the HI virus. Studies have revealed that such mutations may alter the viral structure and can affect the functional kinetics of the enzyme (Debouck *et al*., 1987; Kim *et al*., 2001; Ishima *et al*., 2007).

## 1.5 Insertion mutations

Amino acid insertions are a common feature of the reverse transcriptase enzyme of HIV and are implicated in the development of drug resistance in HIV positive patients on RT inhibitors (Kim *et al*., 2001; Masquilier *et al*., 2001; Kozisek *et al*., 2008). However, in rare cases, insertion mutations have also been detected in the protease gene of HIV. In 2001, Kim *et al*. estimated the prevalence of insertions in the protease to be 0.09% from a database of more than 24 000 genotyped patients' sera. Most of the insertions in the protease are usually duplications of neighboring sequences and are likely generated by RT slippage during reverse transcription process (Winters and Merigan, 2001; Kozisek *et al*., 2008). Part of the "hinge" region (i.e., codons 32 to 41, see Figure 3), which leads to the "flaps", in the protease, is most prone to insertions. Mutations in this region cause conformational changes in the active site of the enzyme and lead to compromised flap movement (Todd and Freire, 1999). This confers the virus an effective way to block access of protease inhibitors to the active site. Consequently, the protease develops resistance towards PIs and enhances viral fitness (Kim *et al*., 2001; Kozisek *et al*., 2008).

Following Grant's report on the possibility of transmitting insert-containing viral strains between patients (Grant *et al*., 2001), Sturmer *et al.* (2003) reported on a patient with amino acid inserts near codon 36 (M36TNL) of the protease prior to protease inhibitor therapy. This suggested that viral strains with inserts in the protease enzyme, with acquired PI resistant mutations, may become more common in the future.

## 1.6 Objectives

Insertion mutations have been detected at various positions in the protease coding region of HIV, but codons 32 to 41 of the protease enzyme appear to be more prone to amino acid insertions (Winters *et al*., 2005). Grant *et al*. (2001) reported on the possibility of transmission of the insert-containing HI virus amongst patients, suggesting that viral strains with insertion mutations in the protease enzyme, that have acquired protease inhibitor resistance, might be more common in the future. Later on, amino acid insertions (M36TNL) in the protease sequence of HIV-1 were detected in protease inhibitor-naïve patients (Sturmer *et al*., 2003). The M36TNL variant was discovered in the subtype B HIV-1 protease and did not show major effects on the biochemical characteristics of the protease enzyme.

Thus, in this study, an HIV-1 South African subtype C protease with an amino acid substituted and inserted at codon 36 (I36TT) was assessed (see Figure 6 for the location of the insertion and background amino acids polymorphisms). The protease variant was selected from 240 genotyped patients' samples. The PR sequence was obtained from Prof. Lynn Morris (NICD, Johannesburg, South Africa). The variant has distinct background polymorphisms (P39S, D60E, and Q61E) which accompany the insertion. The protease variant in the current study (I36TT) was detected in drug-naïve patients.

The current study was aimed at assessing the contribution of the I36TT substitution/insertion mutation to the overall structure and enzymatic properties of the South African HIV-1 subtype C protease. Position 36 of the protease is located in the hinge region (Figure 6) which controls access to the active site of the enzyme. Mutations in this region may alter protease flexibility and activity (Kozisek *et al*., 2008). Secondary and tertiary structural content of the I36TT variant protease were probed using far-UV circular dichroism spectroscopy and intrinsic tryptophan fluorescence spectroscopy, respectively. Isothermal titration calorimetric experiments with the aspartyl protease inhibitor, acetyl pepstatin, were carried out to evaluate binding energetics of the inhibitor to the HIV-1 C-SA protease. Kinetic parameters of the enzyme were evaluated using a chromogenic HIV-1 substrate which resembles the protease cleavage site in the *gag-pol*

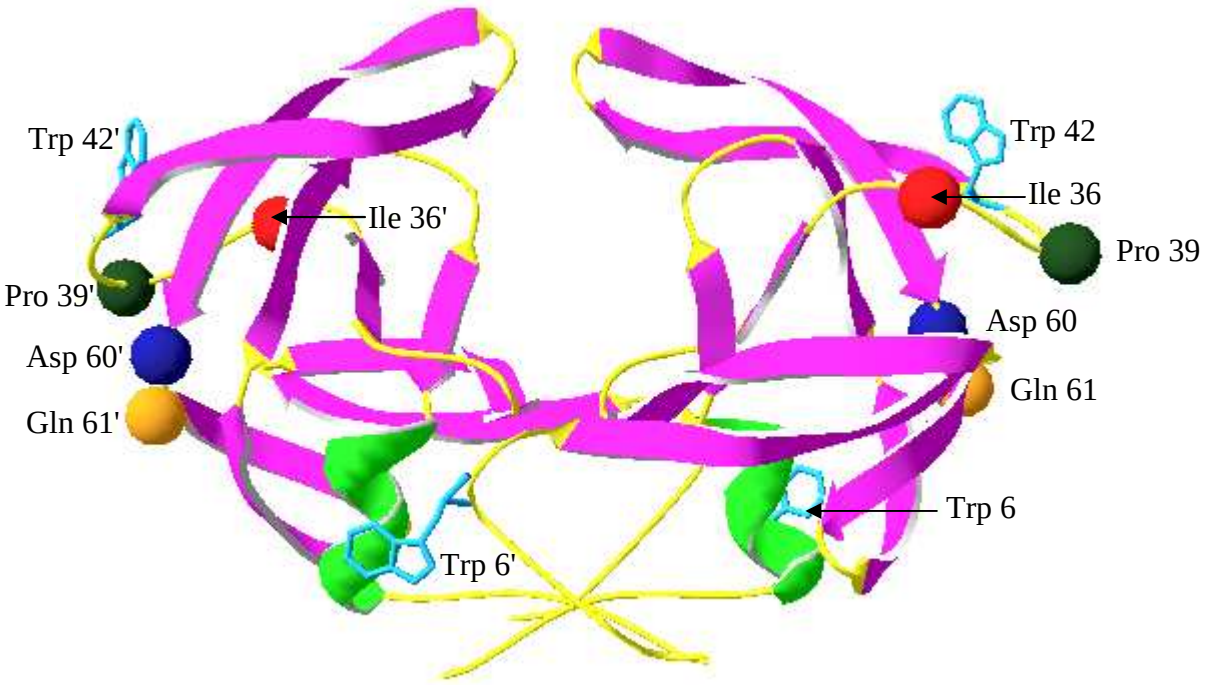



**Figure 6: Ribbon representation of the crystal structure of a free (unbound) conformation of HIV-1 protease**

Locations of amino acid residues which are mutated (background polymorphisms) in the I36TT variant used in this study are highlighted with spheres of various colors. Shown in green, pink and yellow are the α-helices, β-strands and aperiodical structures, respectively. The backbone and side chains of the tryptophan residues from each subunit are highlighted in blue stick-format. The above structure was generated using Swiss-pdb Viewer (Guex *et al.*, 1995) (PDB code: 1HHP).

polyprotein precursor. Furthermore, inhibition assays were performed using four protease inhibitors which are available in our laboratory; notably, saquinavir, indinavir, ritonavir and nelfinavir. These drugs are currently employed in HIV therapy. The variant HIV-1 C-SA protease was compared to the wild-type protease.

# CHAPTER 2

# 2 MATERIALS AND EXPERIMENTAL PROTOCOLS

## 2.1 HIV-1 expression vector

The pET-11b plasmid carrying a gene encoding subtype B protease was a gift from Dr. Jordan Tang from the University of Oklahoma Health Sciences Center, Oklahoma city. The plasmid DNA was mutated to code for the wild-type HIV-1 South African subtype C protease by Dr. Salerwe Mosebi at the University of the Witwatersrand, Johannesburg, South Africa. Autocatalysis of the protease was minimized by introducing the Q7K substitution which reduces the most autocatalytic site (Mildner *et al*., 1994; Velazquez-Campoy *et al*., 2001a). The plasmid is 5.7 kb and is under control of a T7 promoter. Plasmid identity was confirmed by DNA digestion using specific restriction enzymes, i.e., *BamH*I and *Nde*I, and then analyzed on a 1% agarose gel. The cDNA sequence was verified by sequencing (at Inqaba Biotechnical Industries) using the universal T7 primer.

## 2.2 Reagents

Chromogenic HIV substrate was purchased from California Peptide Research, Inc. California, USA. FlexiPrep kit for DNA purification was purchased from Amersham Biosciences, New Jersey, USA. Oligonucleotide primers, DNA and protein ladders, and restriction enzymes (*BamH*I and *Nde*I) were purchased from Inqaba Biotechnical Industries, Pretoria, RSA. The QuikChange$^{TM}$ site-directed mutagenesis kit was purchased from Stratagene, California, USA. All other chemicals were of analytical grade.

## 2.3 HIV-1 South African subtype C variant protease sequence

Sequence data of a patient infected with South African subtype C (C-SA) HIV-1 was obtained from Prof. Lynn Morris's laboratory at the National Institute for Communicable Diseases, Johannesburg, South Africa. Viral RNA was purified from the patient's serum using QIAmp MinElute® Spin Kit (QIAGEN) according to the manufacturer instructions. The extracted viral RNA was reverse transcribed to synthesize a cDNA which was then amplified with a nested PCR. Amplification of cDNA was verified on an agarose gel (1%) containing ethidium bromide. PCR product was purified using Roche PCR purification (Roche) according to the manufacturer instructions and purity was confirmed using 1% agarose gel with ethidium bromide. DNA was further quantitated using a nanodrop spectrophotometer (Thermo Scientific). The pure cDNA was sequenced using BigDye Terminator v3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems) on a PRISM 3100 Genetic Analyzer (Applied Biosystems). Analysis of the resultant sequence was performed using Sequencher software program. The protease coding sequence contained a single amino acid substitution and insertion at codon 36, i.e., I36TT, and three background mutations; notably, P39S, D60E, and Q61E, as compared to the wild-type HIV-1 C-SA protease. The above-mentioned mutations were incorporated into the wild-type C-SA protease using site-directed mutagenesis to generate the variant protease.

## 2.4 Generation of variant plasmid DNA

### 2.4.1 Mutagenic primer design

Mutagenic primers were designed to incorporate specific oligonucleotides at defined positions (see Figure 6) on the wild-type protease sequence to produce the variant protease. The primers were designed based on the HIV-1 C-SA protease sequence by Mosebi *et al*., (2008) with the aid of Gene Runner software (v3.01). The oligonucleotide primers used to generate the HIV-1 C-SA variant protease incorporating an ATC to ACA ACA (I36TT) substitution/insertion plus the background (CCG to TCA (P39S), GAT to GAG (D60E), CAG to GAA (Q61E)) mutations had the following sequences (Table 1):

**Table 1: Primer sequences**

Sequences of primers used to incorporate the I36TT substitution/insertion mutation including background polymorphisms in the HIV-1 C-SA protease to generate the I36TT variant protease

| Mutation | Primer | Sequence |
|---|---|---|
| D60E/Q61E | Forward | 5' GTTCGTCAGTATGAGGAAATCCTGATCGAAATCTGCGG 3' |
| | Reverse | 5' CCGCAGATTTCGATCAGGATTTCCTCATACTGACGAAC 3' |
| I36TT/P39S | Forward | 5' GTTCTGGAAGAAACACAACAAATCTGTCAGGTAAATGGAAGCC 3' |
| | reverse | 5' GGCTTCCATTTACCTGACAGATTTGTTGTTTCTTCCAGAAC 3' |

*Inserted and substituted bases are shown in blue and red, respectively.*

### 2.4.2 Generation of the I36TT substitution/insertion and the background polymorphisms

A cDNA, incorporated in pET-11b expression vector, encoding the wild-type HIV-1 South African subtype C specific viral mutations was used as a template to generate the variant protease. The I36TT substitution/insertion mutation plus the background mutations were introduced at specific positions of the plasmid using mutagenic oligonucleotides and a QuikChange$^{TM}$ site-directed mutagenesis kit from Stratagene (Papworth *et al*., 1996). PCR-based site-directed mutagenesis was performed according to the manufacturer's instructions. Briefly, PCR reactions contained 5 μℓ of 10 X reaction buffer, 50 ng of double-stranded DNA template, 125 ng of each (forward and reverse) oligonucleotide primers, 10 mM dNTP mix, 1 μℓ *PfuTurbo* DNA polymerase and distilled water added to a total volume of 50 μℓ. Mutagenesis was achieved through eighteen amplification cycles at three different temperatures. Amplification was performed at 95 °C for 30 seconds to denature the DNA, at 55 °C for 30 seconds to anneal the primers to the template DNA strand and at 68 °C for 120 seconds to polymerize the DNA. The PCR product was treated with 1 μℓ of the *Dpn*I restriction enzyme to digest the methylated and hemimethylated DNA template. The *Dpn* I-treated DNA was then used to transform *Escherichia coli* XL1-Blue supercompetent cells supplied with the kit (Chung *et al.,* 1989). See Table 2 for a summary of the PCR reaction parameters.

Cells were plated onto lysogeny broth (LB) agar plates supplemented with 100 μg/mℓ of ampicillin. Transformants were selected by plating the cells on LB plates supplemented with 100 μg/mℓ of ampicillin and incubated at 37 °C for 18 hours. Subsequently, the mutant plasmid DNA encoded in the pET-11b vector was extracted from the transformed *Escherichia coli* XL1-Blue cells and purified using the FlexiPrep kit as per manufacturer's instructions. To make sure that the correct base substitutions and the insertion have been incorporated, and that no other alterations were made during PCR amplification, the purified plasmid DNA was sent to Inqaba Biotechnical Industries (Pretoria, RSA) for sequencing.

**Table 2: PCR cycling parameters**

The I36TT variant was generated using the aforementioned mutagenic oligonucleotides and Stratagene QuikChange™ site-directed mutagenesis kit using PCR under the following conditions:

| Segment | Cycles | Temperature (°C) | Time (sec) |
|---|---|---|---|
| 1 | 1 | 95 | 30 |
| 2 | 18 | 95 | 30 |
| | | 55 | 60 |
| | | 68 | 120 |

## 2.5 Transformation of *Escherichia coli* cells with plasmid DNA

*Escherichia coli* BL21 (DE3) pLysS cells were transformed with the plasmid DNA encoding the wild-type HIV-1 C-SA protease. Transformants were selected by plating the cells on LB agar plates supplemented with 100 μg/mℓ of ampicillin and 35 μg/mℓ of chloramphenicol and incubated at 37 °C for 18 hours. Cells harboring the plasmid were selected and grown at 37 °C in a fresh LB medium supplemented with 100 μg/mℓ of ampicillin and 35 μg/mℓ of chloramphenicol. The same was done for the variant (containing the I36TT, P39S, D60E, and Q61E mutations) plasmid DNA. The wild-type and the I36TT variant C-SA proteases were over-expressed, purified from washed inclusion bodies and re-folded as described below.

## 2.6 Over-expression, extraction and purification of the C-SA protease

Plasmid encoding either variant or wild-type HIV-1 SA subtype C protease (containing the mutation Q7K designed to reduce the hypersensitive autolytic site) was expressed as inclusion bodies (Ido *et al*., 1991; Todd *et al*., 1998) in *Escherichia coli* BL21 (DE3) pLysS cells. Briefly, *Escherichia coli* cells harboring the plasmid DNA were grown at 37 °C in LB medium supplemented with 100 μg/mℓ of ampicillin and 35 μg/mℓ of chloramphenicol. The overnight culture was diluted 100-fold into fresh 2 × yeast extract and tryptone powder (YT) medium supplemented with ampicillin (100 μg/mℓ) and chloramphenicol (35 μg/mℓ) and grown at 37 °C. When the optical density ($OD_{600}$) of the culture reached 0.4 to 0.5 (after two and half hours), over-expression of the HIV-1 C-SA protease was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG). Upon binding of IPTG to the lac repressor, which regulates transcription, the inhibitor is removed from the operator and transcription proceeds (Chung *et al*., 1989). IPTG was added to final concentrations of 0.4 mM and 0.2 mM for the wild-type and the variant, respectively. Over-expression of both proteases was allowed to continue for two hours.

The cells were pelleted after growth and resuspended in ice-cold extraction buffer (10 mM Tris, 1mM EDTA, and 1 mM PMSF, pH 8). Following the addition of $MgCl_2$ and DNase I to final concentrations of 10 mM and 10 U/μℓ, respectively, the culture medium was stirred on ice until the viscosity of the mixture decreased. The cells were then ruptured by sonication and centrifuged at 15 000 × g for 30 minutes at 4 °C. The pellet was resuspended in ice-cold extraction buffer containing 1% (v/v) of Triton X-100. Cell debris and protease-containing inclusion bodies were pelleted by centrifugation at 15 000 × g for 30 minutes at 4 °C. The pellet was then resuspended in a freshly prepared solubilisation buffer containing 10 mM Tris, 2 mM DTT, 8 M urea, pH 8.0, at room temperature, and centrifuged at 15 000 × g for 30 minutes at 20 °C.

The protease, in the supernatant, was purified using anion exchange (DEAE) column chromatography previously equilibrated with solubilisation buffer. HIV-1 PR has a pI of 9.3 and will not bind the DEAE column. Upon elution from the column, the protease was acidified by adding formic acid to a final concentration of 25 mM. Precipitation of significant amount of contaminating proteins occurred upon acidification. Following an overnight incubation, the precipitated contaminants were removed by centrifugation at 15 000 × g for 30 minutes at 4 °C. The protease was concentrated to a final volume of ~20 mℓ and stored at 4 °C.

HIV-1 protease was refolded by dialysis into 10 mM formic acid at 4 °C. Subsequently, the protease was dialyzed into a buffer containing 10 mM sodium acetate, 1 mM NaCl and 1 mM DTT, pH 5.0. The folded protease was then desalted into 10 mM sodium acetate buffer, pH 5.0 using a PD-10 gel filtration column (Amersham biosciences. NJ, USA) and stored at -20 °C.

## 2.7 Protein concentration determination

The concentration of the variant and the wild-type HIV-1 C-SA protease was determined spectrophotometrically (using a JASCO V-630 UV-VIS spectrophotometer) using the Beer-Lambert law:

$$\mathbf{A} = \boldsymbol{\varepsilon_{\lambda} c l} \qquad \textbf{Equation 1}$$

where A is the absorbance, ε is the molar extinction coefficient at a given wavelength (λ), c is the concentration of the absorbing medium, and $l$ is the pathlength of the light through the medium in a cuvette. The concentration of pure HIV-1 C-SA protease was subsequently determined spectrophotometrically using an extinction coefficient, $E_{1\%}$, of 11.8 at 280 nm (Polgár *et al*., 1994). A quartz cuvette of 1 cm path length was used for all spectrophotometer readings. The purity and the monomeric molecular weight of the proteases were evaluated as described in section 2.8 using 15% SDS-PAGE (Laemmli, 1970).

## 2.8 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), the discontinuous buffer system described by Laemmli (1970), was used to determine the purity and subunit molecular weight of the protein. Briefly, the technique uses SDS, an anionic detergent, to denature proteins. Upon binding to and denaturation by SDS, the proteins acquire a rod-like shape with an overall negative charge. The charged particles migrate towards the electrode of opposite charge (i.e., positive terminus) in an electric field. ß-mercaptoethanol is used in SDS-PAGE to reduce any inter- and intramolecular disulphide bonds. Thus, proteins are separated on the basis of size.

The protein samples were diluted five-fold in 5 × loading buffer (0.5 mM Tris-HCl, pH 6.8, 20% (v/v) glycerol, 10% (w/v) SDS, 100 mM ß-mercaptoethanol, 0.05% (w/v) bromophenol blue) and boiled for five minutes to denature the proteins before loading onto the gel. Electrophoresis was performed at 200 volts for 45 minutes using a Bio-Rad system. Electrode buffer, pH 8.3, contained 0.192 M glycine, 0.124 M Tris base and 0.5% (w/v) SDS. The sizes of the proteins were determined in relation to a set of standard proteins, of known molecular weight, electrophoresed under the same denaturing conditions. The protein molecular weight markers (Fermentas #SM0431) used were: β-galactosidase (116.0 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45.0 kDa), lactate dehydrogenase (35.0 kDa), REase *Bsp981* (25.0 kDa), β-lactoglobulin (18.4 kDa) and lysozyme (14.4 kDa). Immediately after electrophoresis, the gel was stained in a solution containing 0.25% (w/v) Coomasie Brilliant Blue R250 dissolved in

50% (v/v) methanol and 10% (v/v) acetic acid for two hours to visualize the separated protein bands. The gel was then destained in a solution containing 15% (v/v) acetic acid and 10% (v/v) ethanol.

## 2.9 Purification of the protease inhibitors

Commercially available protease inhibitors (PI) were purified by Dr Salerwe Mosebi at CSIR, South Africa. Four of the HIV protease inhibitors employed in clinical settings were used; notably, indinavir, nelfinavir, saquinavir and ritonavir. The active ingredients were extracted from the tablets and purified using a semi-preparative C-18 reversed-phase HPLC column (Velazquez-Campoy *et al*., 2001a). The inhibitors were lyophilized and stored as powder (indinavir and nelfinavir) and as suspensions in 100% dimethylsulphoxide (DMSO) (saquinavir and ritonavir) at -20 °C prior to use (Mosebi *et al.*, 2008).

## 2.10 Calorimetric assays

### 2.10.1 Isothermal titration calorimetry

An isothermal titration calorimeter (ITC) directly measures the heat evolved or absorbed in aqueous solutions as a result of injecting precise amounts of reactants into the sample cell. The heat signal is a universal phenomenon of binding reactions in the sample cell. Thus, it is applicable to most ligand-macromolecule and macromolecule-macromolecule interactions. The isothermal power compensation calorimeter is composed of a sample and a reference cell (see Figure 7). Both cells are maintained at a constant temperature throughout the experiment (Wiseman *et al.*, 1989). The reference cell is usually filled with buffer alone while the sample cell has either the protein or the ligand of interest dissolved in the same buffer as the syringe contents. Buffer identity is crucial in order to eliminate the large heats of dilution which may result and, consequently, mask the desired observations.

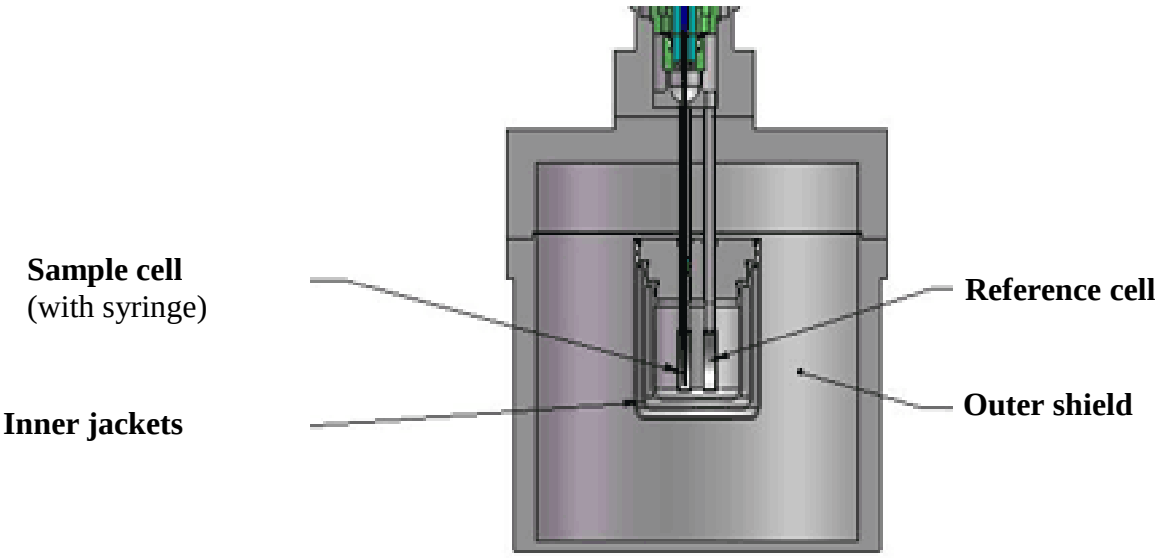



**Figure 7: Schematic representation of ITC cells**

ITC systems use a cell feedback network to differentially measure and compensate for heat produced or absorbed between the sample and reference cell. The two cells (i.e., sample cell and reference cell) are mounted in a cylindrical adiabatic environment. During ITC experiment, the syringe, in the sample cell, rotates to provide continuous mixing of the reactants. Picture taken from MicroCal Inc.

During an ITC experiment, the reference cell is supplied with a user-defined amount of power and the temperature difference between the sample and the reference cell is measured. The variable level of thermal power is supplied to the sample cell in order to drive this temperature difference to zero. The raw signal from the ITC experiment represents the thermal power supplied to the sample cell. The measured change in instrumental thermal power supplied to the sample cell, after addition of a ligand or enzyme, is used to determine the reaction rates (Gomez and Freire, 1995; Todd and Gomez, 2001). As the syringe contents are injected into the sample cell and the two interact, depending on the nature of the reaction (exothermic or endothermic) the ITC supplies power to maintain constant temperature between the sample and the reference cell. For a heat consuming interaction (endothermic), a positive deflection is observed. Conversely, if the reaction generates heat (exothermic), a negative deflection in thermal power is observed (see Figure 8).

Resolvability of the binding parameters from the ITC data depends entirely on the concentration of the cell reactant relative to the affinity of the interaction. Hence, the initial concentration of the protein and the ligand needs to be taken into consideration and both reactants must be of highest purity. Too low or too high cell reactant concentration will result in either a data which does not allow for determination of interaction parameters or insufficient information in the transition region to resolve the affinity constant ($K_a$), respectively (Todd and Gomez, 2001). Precise measurement of the energy required to maintain temperature of the cell during the experiment enables simultaneous determination of all binding parameters i.e., Gibbs free energy, change in enthalpy, change in entropy, binding affinity and stoichiometry of binding from one ITC experiment. For a strong interaction, $\Delta G$ must be large and negative, $\Delta H$ must be large and negative and $\Delta S$ must be large and positive. Determination of HIV-1 C-SA protease active site concentration was performed using a high-precision VP-ITC titration calorimeter from MicroCal Inc., USA. The enzyme (protease) solution at 20 μM was loaded into the ITC sample cell and titrated with the ligand (300 μM) dissolved in the same buffer (10 mM sodium acetate, pH 5.0). Acetyl pepstatin, an aspartic protease inhibitor, was used a ligand when determining the concentration of the active site of the HIV-1 protease. The heat evolved after each inhibitor injection was obtained from the integration of the area under the peak of the heat signal. The heat due to the binding reaction between the protease enzyme and the inhibitor was obtained from the

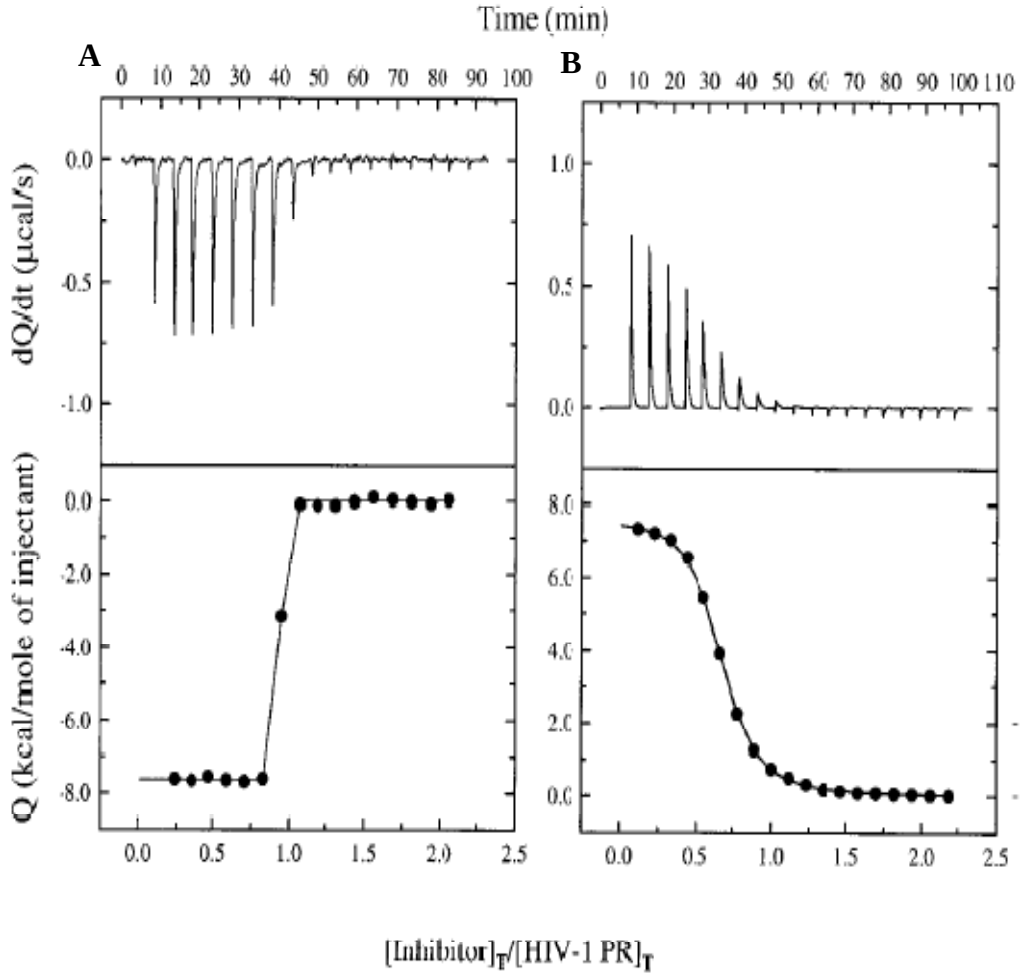



**Figure 8: HIV-1 binding isotherms**

Exothermic (**A**) and endothermic (**B**) ITC profiles of the wild-type HIV-1 subtype B protease titrated with KNI-764 (in **A**) and acetyl pepstatin (in **B**) inhibitors. Figure adapted from Velazquez-Campoy *et al*., 2001b.

difference between the heat of reaction and the corresponding heat of dilution (dilution of inhibitor in a buffer). At the end of the experiment, saturation of the macromolecule is reached and it is possible to measure the heats associated with binding events; particularly, $\Delta H$, $\Delta S$, $\Delta G$, $K_a$ and $N$ (Velazquez-Campoy *et al*., 2001b). The independent variables; $K_a$, $N$ and $\Delta H$ were directly determined from the integrated heat signal. The $\Delta G$ of binding was calculated by substituting the value of $K_a$ in the equation:

$$\Delta \boldsymbol{G} = -\mathbf{RT}\boldsymbol{lnK_a} \qquad \textbf{Equation 2}$$

where R is the universal gas constant (1.98 cal.mol$^{-1}$.K$^{-1}$) and T is the temperature in Kelvin. $\Delta S$ was in turn calculated using the equation:

$$\Delta \boldsymbol{G} = \Delta \boldsymbol{H} - \mathbf{T}\Delta \boldsymbol{S} \qquad \textbf{Equation 3}$$

The stoichiometry obtained from the titration data is then used to estimate the concentration of active site in the protease molecule. A stoichiometry of 1 corresponds to a 100% active site concentration.

## 2.11 Protease characterization

### 2.11.1 Structural assessment

#### 2.11.1.1 Secondary structure

Circular dichroism (CD) is observed when optically active matter absorbs left and right handed circularly polarized light differently. To generate a CD signal, the optically active molecule (chromophore), must either be in an optically asymmetric environment or be intrinsically chiral (Pain, 2004). The chromophores in proteins are the aromatic acid residues, peptide backbone and disulphide bonds (Woody, 1995). CD signal is generated during the transition of an electron from a filled ground state orbital (n or $\pi$) to an empty excited state orbital ($\pi$*) with higher energy. The raw CD signal is expressed in terms of ellipticity ($\theta$) given as:

$$\theta = 2.303 \times \frac{A_L - A_R}{4} \qquad \textbf{Equation 4}$$

where $A_L$ and $A_R$ are the absorbances of the left- and right-handed circularly polarised light (Modi *et al.*, 1993). The raw CD signal ($\theta$) was buffer-corrected then converted to mean residue ellipticity (deg.cm$^2$.dmol$^{-1}$) using the equation:

$$\mathbf{MRE} = \frac{\theta \times 100}{cnl} \qquad \textbf{Equation 5}$$

where $\theta$ is the raw signal (mdeg), c is the concentration of the protein (mM), n is the number of residues in the protein and $l$ is the path length of the cuvette (cm) (Woody, 1995).

The technique is used for assessing the structures of proteins and peptides in solution and was employed in this study to determine structural properties of both the wild-type and the variant HIV-1 C-SA protease. Secondary structures were determined using far-UV (250 to 190 nm) CD spectra due to light absorption by the peptide backbone (Adler *et al.*, 1973). The data were taken on a Jasco J-810 spectropolarimeter, constantly flushed with nitrogen to remove oxygen, using a 2 mm path length quartz cuvette. The spectropolarimeter was set at 20 °C and the data were collected under the following conditions: 0.1 nm data pitch, 0.1 nm bandwidth, and 50 nm.min$^{-1}$ scan speed.

#### 2.11.1.2 Tertiary structure

Fluorescence is an emission phenomenon which involves energy transition from higher to lower state. Briefly, molecules in their lower energy state are excited by the absorption of light of a specific wavelength to a higher energy state. The excited molecules return to ground state at longer wavelength than the excitation radiation and the energy emitted is manifested as fluorescence (van Holde *et al.*, 1998). Energy loss is rapid (as heat) and occurs by collision degradation. This loss of energy between absorbed and emitted light is known as “Stokes’ shift”. (Lakowicz, 1999). The aromatic compounds (with delocalised $\pi$-electrons) are most easily excited to higher energy states, and thus are prone to fluoresce. Proteins exhibit characteristic fluorescence spectra according to the environment within which the main fluorescing species are

packed. The greater the exposure of Trp to the polar aqueous environment, the longer its wavelength of maximum emission will be, since the polar solvent molecules lower the energy of the excited state (Royer, 1995). Tyr fluorescence is almost totally quenched if it is ionised or near an amino group, or carboxylic group or Trp.

Fluorescence spectroscopic assays were used to investigate the localized conformational changes induced in the tertiary structure of the protease. Changes occurring in the tertiary structure of the protease were monitored using the intrinsic fluorescence of the phenyl and indole rings of tyrosine and tryptophan, respectively, as probes. Tyr and Trp were excited at 280 nm, and tryptophan was selectively excited at 295 nm (Shirley 1995). C-SA protease has two tryptophan residues (located at positions 6 and 42, see Figure 6) per subunit. These tryptophans are located where they will not detect major/global structural changes and are quite exposed to the solvent; therefore, they were used to monitor local structural changes. Fluorescence assays were performed using a PerkinElmer LS 50 B Luminescence fluorimeter. Fluorescence emission was monitored between 290 and 550 nm. Excitation and emission slits were set at 4.0 nm and 3.5 nm, respectively. The scan speed was set at 350 $nm.min^{-1}$ and ten accumulated scans were averaged for each spectrum. The assays were conducted using 5 μM protein in 10 mM sodium acetate buffer, pH 5.0.

## 2.11.2 Spectrophotometric analyses

### 2.11.2.1 Protease enzyme kinetics

Enzymatic activity of the HIV-1 C-SA protease was measured by following the hydrolysis of the HIV chromogenic substrate, Lys-Ala-Arg-Val-Nle-nPhe-Glu-Ala-Nle-$NH_2$. The substrate resembles the conserved protease cleavage site, KARVL/AEAM, between the capsid protein and the nucleocapsid p2 in the Gag-polyprotein precursor (Velazquez-Campoy *et al*., 2001a). Hydrolysis of the HIV chromogenic substrate was characterised by the decrease in absorbance at 300 nm. Catalytic properties such as the $K_M$, $V_{max}$, $k_{cat}$, and $k_{cat}/K_M$ of the proteases were determined (Velazquez-Campoy *et al*., 2001a and Mosebi *et al*., 2008). All catalytic activity assays were performed using a Jasco V-630 spectrophotometer.

The $K_M$ for the substrate was calculated based on different catalytic activities measured using constant concentration of the protease and varied substrate concentrations. Initial reaction velocities were plotted against substrate concentrations in order to generate Michaelis-Menten plots. Thus, $K_M$ and $V_{max}$ values were determined by fitting the data in the following equation:

$$\boldsymbol{V_0 = \frac{V_{max}\,[S_0]}{[S_0] + K_M}} \qquad \textbf{Equation 6}$$

where $V_0$ is the initial velocity of the reaction, $V_{max}$ is maximum velocity attained by the reaction, and $[S_0]$ is the substrate concentration in the reaction (when substrate concentration is much greater than enzyme concentration).

Protease was added to a micro-cuvette containing substrate at 20 °C. Final concentrations for the $K_M$ and $V_{max}$ determination assays were: 300 nM protease, 100 μM substrate, and 10 mM sodium acetate buffer, 0.1 M NaCl, pH 5.0.

It is known that when the substrate concentration is very high, all the enzyme is present as the enzyme-substrate complex, and the limiting initial velocity, $V_{max}$, is reached. Under such conditions, the turnover number was determined from the slope of the linear plot of catalytic activity versus enzyme concentration using the equation:

$$\boldsymbol{V_{max} = k_{cat}[E]_t} \qquad \textbf{Equation 7}$$

where $[E]_t$ is the total protease concentration in the assay.

When the substrate concentration is very low, there is a linear relationship between $V_0$ and $[S_0]$, since $[S_0] \ll K_M$, then $([S_0] + K_M) \approx K_M$. Hence, the catalytic efficiency of the protease was calculated as the slope of the linear plot of catalytic activity against substrate concentration using the following equation:

$$\boldsymbol{V_0 = \left(\frac{k_{cat}}{K_M}\right)[E]_t} \qquad \textbf{Equation 8}$$

The measured slopes of the plots were then divided by the respective enzyme concentrations to give values for $k_{cat}/K_M$.

#### 2.11.2.2 Inhibition studies

The concentration of the individual protease inhibitors needed to reduce the catalytic activity of the HIV-1 C-SA protease enzyme by 50% ($IC_{50}$ value) was determined at 25 °C by measuring the rate of chromogenic substrate hydrolysis in the presence of various protease inhibitors. Standard assays containing 300 nM protein and saturating concentration (i.e., 250 μM) of chromogenic substrate were performed in the presence of increasing concentrations of inhibitor. Separate assays were performed for each of the following protease inhibitors currently in clinical use; i.e., saquinavir, ritonavir, indinavir, and nelfinavir.

## 2.12 Data analysis

Protease concentrations used are for dimeric protein unless otherwise stated. The terms mutant and variant are used interchangeably in this dissertation. Data fitting was performed using Sigma Plot v11.0 software (Systat softwares) unless stated otherwise. ITC data was fitted using ITCone from ORIGIN v5.0 software (MicroCal. Inc.). All HIV-1 protease structures were generated using Swiss PDB viewer software (Guex *et al*., 1995).

# CHAPTER 3

# 3 RESULTS

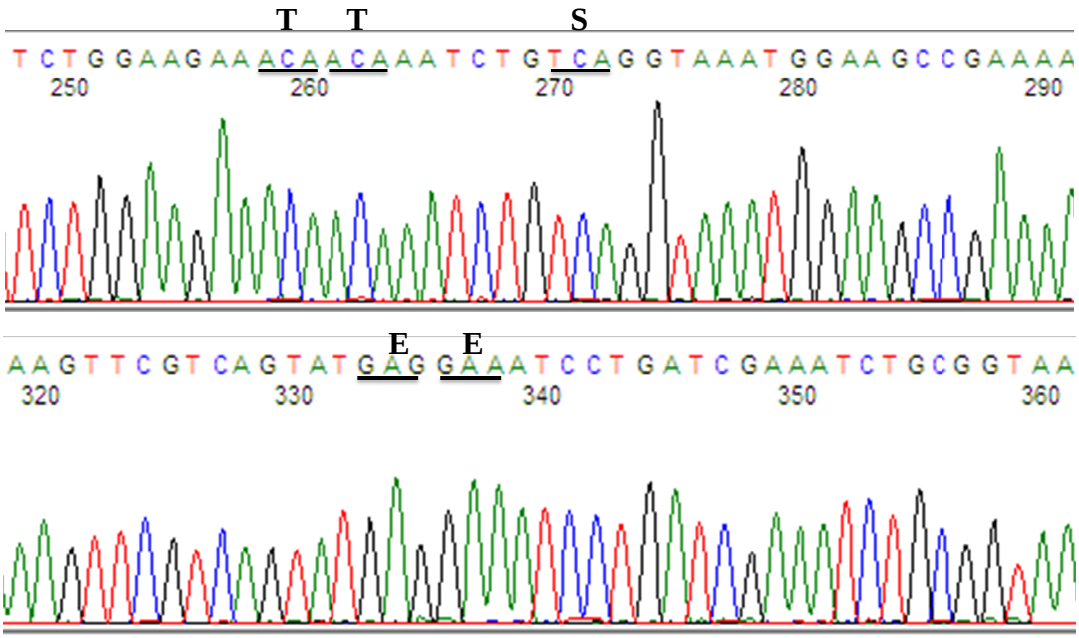
## 3.1 Plasmid DNA sequence confirmation

cDNA encoding wild-type HIV-1 South African subtype C protease was used as a template during PCR-based site-directed mutagenesis to generate the variant cDNA. The cDNA encoded in the pET-11b plasmid was purified and sequenced using the Universal T7 primer. DNA sequencing was performed (at Inqaba Biotechnical Industries) in order to confirm that only the mutations observed in the I36TT variant had been incorporated in the protease sequence, and no other alterations had occurred during amplification. Sequencing (Figure 9A) and sequence alignment results (Figure 9B) revealed that only the I36TT (ATC to ACA ACA) substitution/insertion and its accompanying polymorphisms (CCG to TCA (P39S), GAT to GAG (D60E), CAG to GAA (Q61E)) were incorporated. The variant protease sequence was aligned with the wild-type HIV-1 C-SA protease to highlight differences between the two proteases (Figure 9B) and it showed that the variant protease has 100 amino acids in its sequence in contrast to the wild-type with only 99 amino acids.

## 3.2 Protease purity and monomeric molecular weight determination

Following over expression of the HIV-1 C-SA protease and its subsequent extraction and purification from inclusion bodies, protease purity and subunit molecular weight were determined using SDS-PAGE (Figure 10A). Electrophoresis results revealed that the wild-type and the mutant protease were pure. This was indicated by a single band observed in the lane in which the protein was loaded in the gel. Using the electrophoresis gel (Figure 10A), the calibration curve of logarithm of the molecular weight of the protein molecular weight markers versus the distance migrated by individual proteins was plotted. The monomeric molecular

**A**




**B**

```
Wild-type      PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEEI-NLPGKWKPKMIGG 49
I36TT Variant  PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEETTNLSGKWKPKMIGG 50
               ************************************  **.**********

Wild-type      IGGFIKVRQYDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF 99
I36TT Variant  IGGFIKVRQYEEILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF 100
               **********::************************************
```

**Figure 9: Variant protease sequence confirmation**

**(A)** Sections of the chromatograms highlighting four codons substituted and one more inserted at codon number 36 (I36TT) of the wild-type HIV-1 C-SA protease to generate the variant protease. The background mutations accompanying the insertion/mutation are highlighted at codons 39 (P to S), 60 (D to E), and 61 (Q to E) of the HIV-1 C-SA protease sequence. **(B)** Alignment of the wild-type and the variant HIV-1 South African subtype C protease sequences indicating differences and similarities between the two protease sequences. The Q7K substitution (reduces autocatalysis of the protease enzyme) was also incorporated and is highlighted in pink.

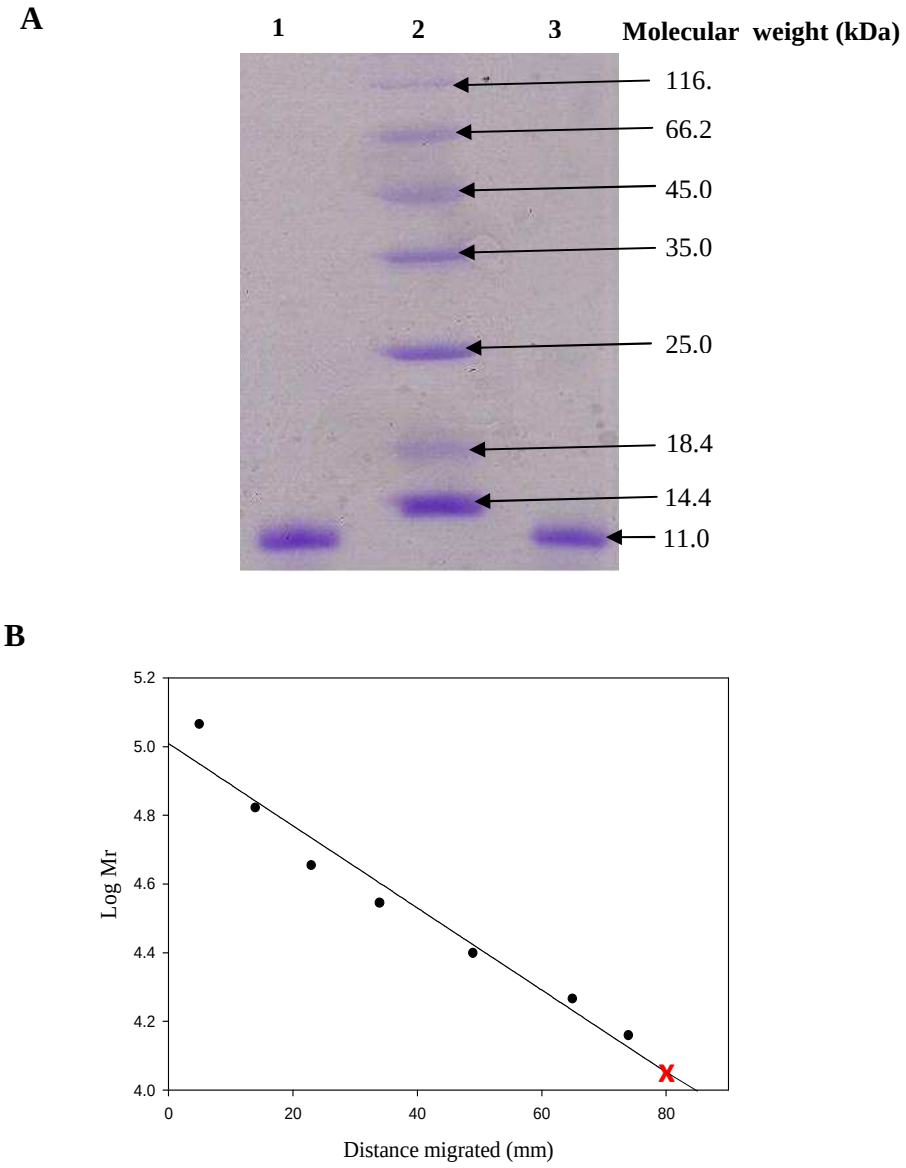



**Figure 10: HIV-1 C-SA protease purity and molecular weight determination**

**(A)** 15% SDS-PAGE gel and **(B)** calibration curve for determination of the molecular weight of the monomeric HIV-1 C-SA proteases. Lanes 1, 2 and 3 represent the purified wild-type protease, protein molecular weight marker, and purified variant protease, respectively. The protein molecular weight markers used were: β-galactosidase (116.0 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45.0 kDa), lactate dehydrogenase (35.0 kDa), REase *Bsp981* (25.0 kDa), β-lactoglobulin (18.4 kDa), and lysozyme (14.4 kDa). Wild-type and variant HIV-1 C-SA protease both migrated a distance corresponding to a size of 11.2 kDa. Position of the protease is marked (**X**). The correlation coefficient is 0.96.

weight of the wild-type and the variant HIV-1 C-SA proteases were determined from the calibration curve, Figure 10B, by fitting data from the plot to the linear equation (y = mx + c). The monomeric molecular weight of the C-SA protease was found to be 11.2 kDa, corresponding to the value of the HIV-1 protease previously reported by Ido *et al*. (1991).

## 3.3 Protein concentration determination

The concentration of the purified South African HIV-1 subtype C protease was determined by measuring the absorbance of light in the ultra-violet region of the spectrum. Optimum absorbance intensity was observed at 280 nm (Figure 11) and was due to the chromophores (i.e., aromatic amino acid residues) in the HIV-1 protease. Using the protein absorption spectrum, Figure 11, the protease concentration was calculated and it was found to be 20 μM.

## 3.4 Structural characterization

### 3.4.1 Secondary structural content assessment

Circular dichroism was used in the far-UV range (250 to 190 nm) to probe changes in the secondary structure of both the wild-type and the variant protease. Both proteases exhibited far-UV CD spectra distinctive of a predominantly β-sheeted protein, characterized by a negative peak at 216 nm (Figure 12). This was in agreement with the secondary structural content of HIV-1 protease as previously reported (Navia *et al.*, 1989; Dash and Rao, 2001; Foulkes *et al.*, 2006; Mosebi *et al.*, 2008) (Figure 6).

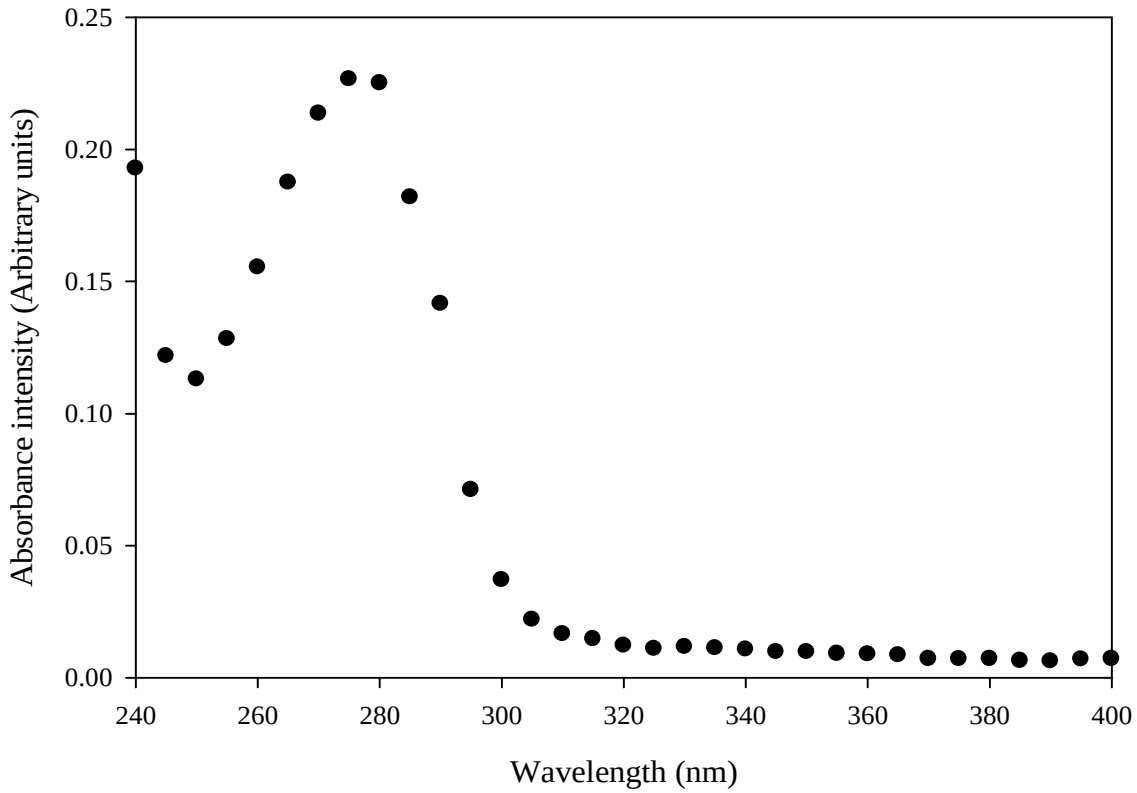



**Figure 11: HIV-1 C-SA absorption spectrum**

Absorption spectrum of 20 μM HIV-1 South African subtype C protease. The experiment was carried out at 20 °C in 10 mM sodium acetate buffer, 0.1 M sodium chloride, pH 5.0. The spectrum illustrates the absorption of ultra-violet light by the aromatic amino acids' side chains of the protein at 280 nm wavelength.

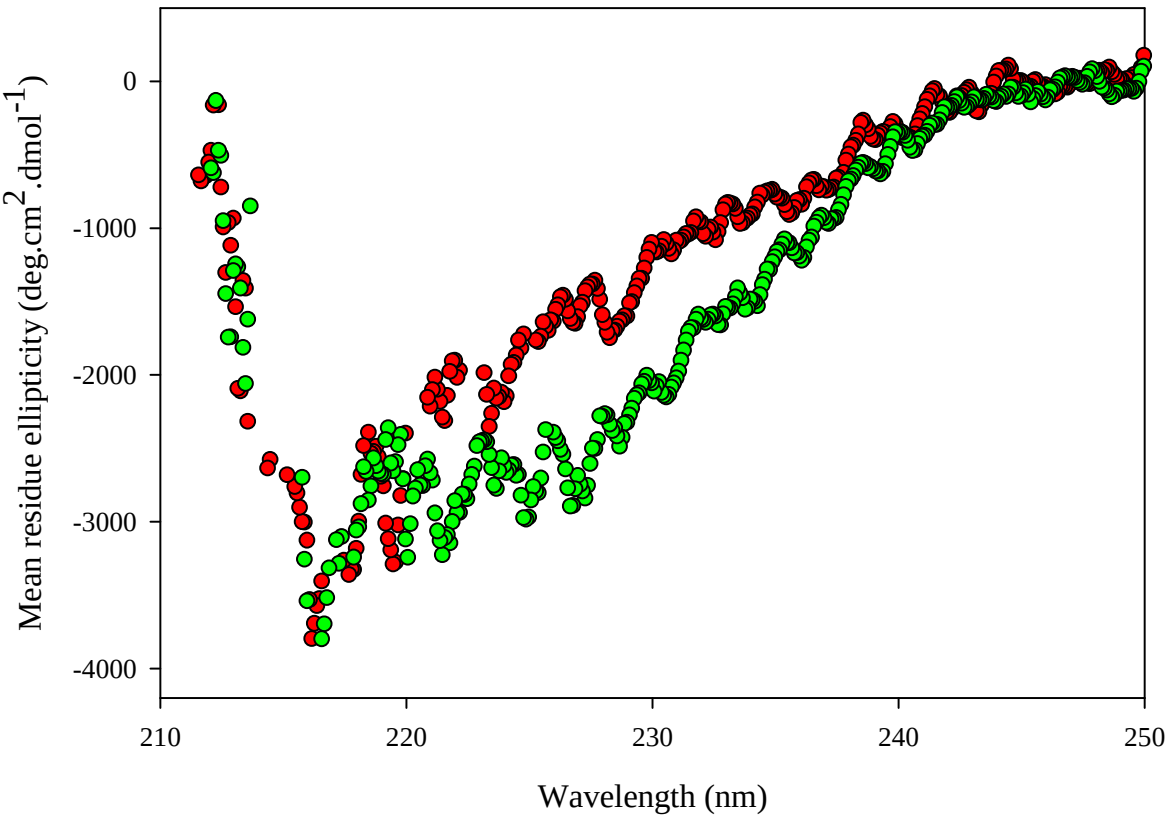



**Figure 12: Far-UV CD spectra of HIV-1 C-SA protease**

Representation of the far-UV circular dichroism spectra of the wild-type (in green) and the variant (in red) HIV-1 C-SA protease in 1 mM sodium acetate buffer, pH 5.0. The spectra were used to assess the secondary structural content of the protease. The data were collected from a Jasco J-810 spectropolarimeter using 25 μM protein in 2 mm path length cuvette. The experiments were conducted at 20 °C under the following conditions: 0.1 nm data pitch, 0.1 nm bandwidth, 250 to 190 nm, and 50 nm.min$^{-1}$ scan speed. Each spectrum is an average of ten scans.

### 3.4.2 Tertiary structure determination

Changes in the tertiary structure of the protease were monitored using intrinsic fluorescence spectroscopy and the four solvent-exposed tryptophan (W6, W6', W42 and W42') residues in the protease sequence as probes. Fluorescence spectra (Figure 13) from both excitation wavelengths (280 and 295 nm) revealed that the local environment of the tryptophan residues in the variant protease had not been perturbed when compared to the wild-type. Both the wild-type and the variant proteases exhibited fluorescence emission spectra typical of free tryptophan in solvent, with an emission intensity maximum at 355 nm (Lakowicz, 1999).

## 3.5 Enzymatic characterization

### 3.5.1 Catalytic activity studies

Kinetic parameters of the HIV-1 South African subtype C protease were assessed following hydrolysis of chromogenic HIV-1 substrate, mimicking the protease cleavage site in the *gag-pol* protein precursor. The reaction was monitored at 300 nm using an absorbance spectrophotometer (Figure 14). Hydrolysis of the substrate was monitored for three minutes for each kinetic assay and only the first fifty seconds of the enzyme reaction was considered when measuring catalytic activity of the enzyme. This was considered due to the linear relationship between the reaction rate and the initial velocity. Overall enzymatic activity (depicted in Figures 15 to 19 and further summarized in Table 3) of the C-SA variant protease is slightly enhanced in the presence of the I36TT insertion/mutation and its accompanying background polymorphisms.

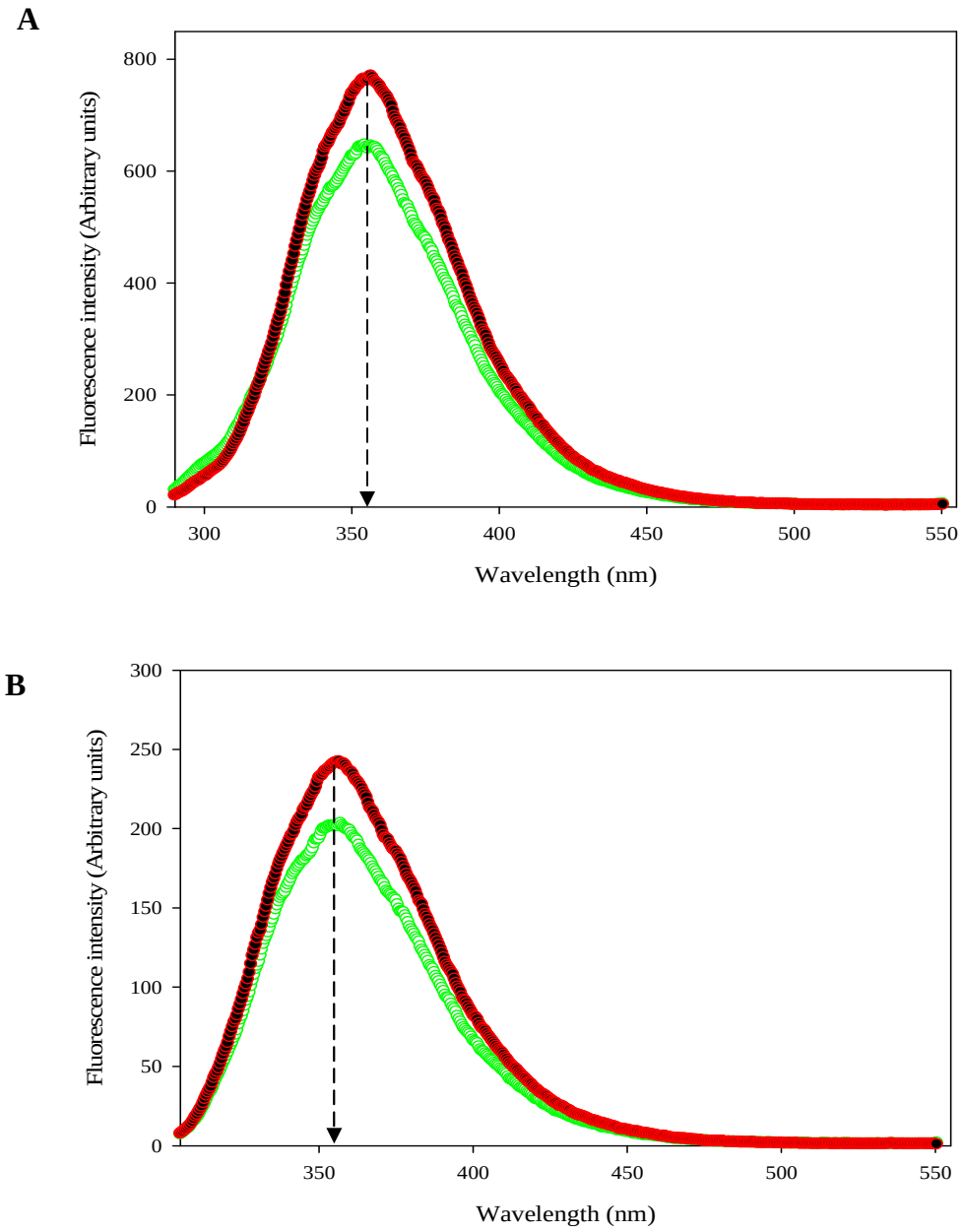



**Figure 13: Intrinsic tryptophan fluorescence emission spectra of the HIV-1 C-SA protease** **(A)** The emission spectrum of tryptophan and tyrosine excited at 280 nm. **(B)** The emission spectrum of tryptophan selectively excited at 295 nm. Both proteins fluoresce maximally at 355 nm at both excitation wavelengths. Fluorescence emission of the wild-type (in green) and the variant (in red) HIV-1 C-SA protease in 10 mM sodium acetate buffer, pH 5.0. Each spectrum is an average of ten accumulations. The dashed vertical lines highlight the wavelength at which maximum fluorescence emission intensities were attained**.**

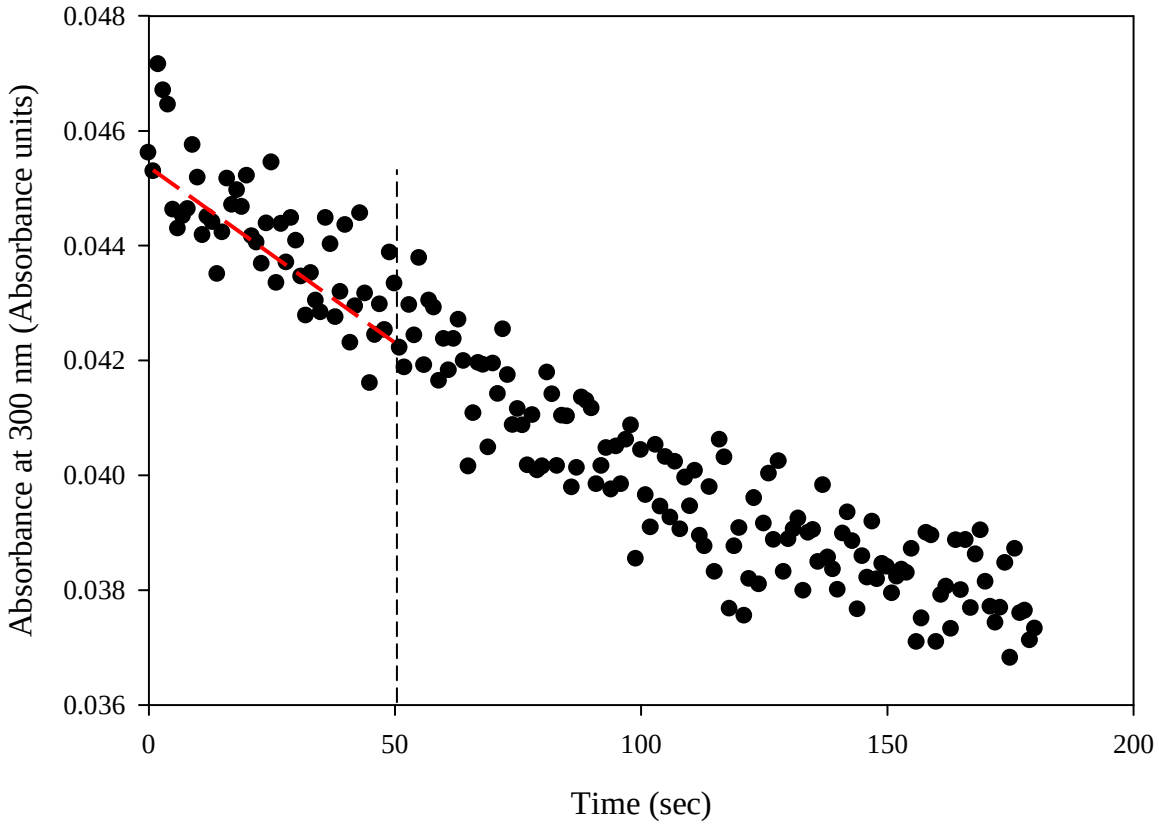



**Figure 14: Progress curve for HIV-1 substrate hydrolysis**

Progress curve of the HIV-1 South African subtype C protease catalyzing the hydrolysis of the synthetic HIV-1 chromogenic substrate. The assay was performed at 20 ° C using 300 nM enzyme concentration and 50 μM substrate in 10 mM sodium acetate buffer, 0.1 M sodium chloride, pH 5.0. The initial velocity (first 50 seconds of the reaction) considered when determining the reaction rate is indicated (**---**). The assay was monitored at 300 nm.

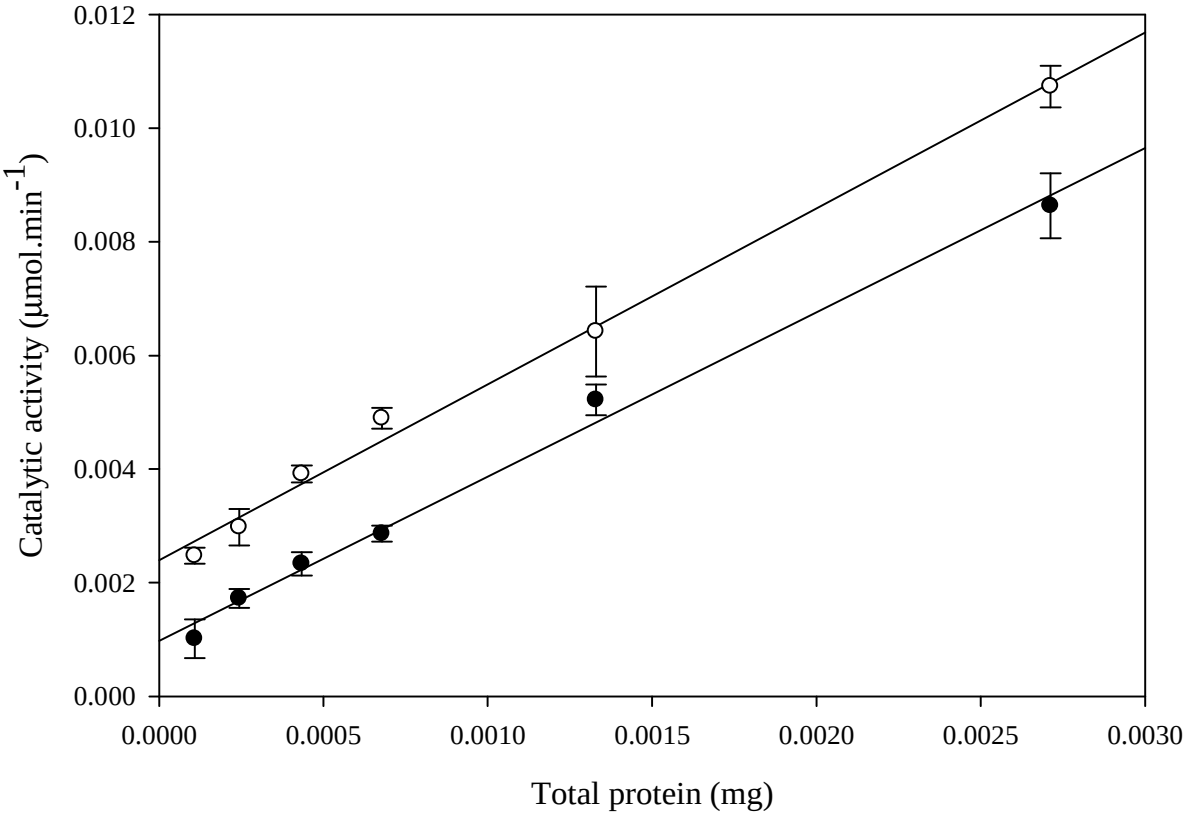



**Figure 15: Determination of specific activity**

Specific activity of the wild-type (●) and the variant (○) HIV-1 C-SA proteases. The catalytic activities of both enzymes were calculated by monitoring the hydrolysis of the chromogenic HIV-1 substrate which was indicated by the decrease in absorbance at 300 nm. The experiment was carried out at 20 °C in 10 mM sodium acetate buffer, 0.1 M sodium chloride, pH 5.0, and saturating substrate concentration (i.e., 250 µM). The experiments were conducted in triplicates and the data is reported as the mean ± SD. Catalytic activity of the protease enzyme was calculated by dividing the spectrophotometrically measured absorbance per minute of the hydrolyzed substrate by the enzyme's molar extinction coefficient. The specific activity of the protease was calculated from the slope of the plot and was found to be 2.9±0.095 µmol/min/mg and 3.1±0.11 µmol/min/mg for wild-type and variant protease enzyme, respectively. The solid lines represent linear fits to the experimental data. The correlation coefficient is 0.98 for both the wild-type and the mutant.

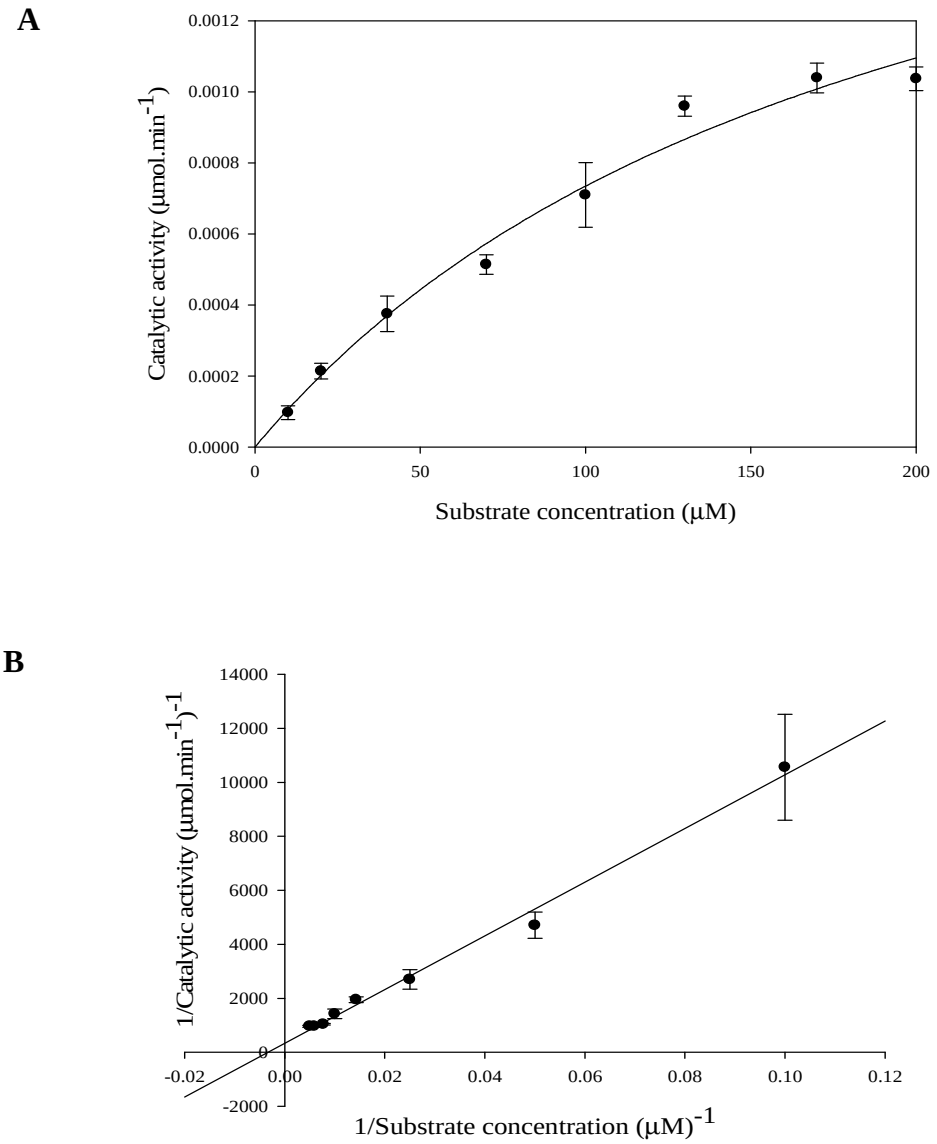



**Figure 16: Wild-type protease catalytic constant and maximum velocity determination**

**(A)** Michaelis-Menten plot and **(B)** Lineweaver-Burke plot for determination of $K_M$ and $V_{max}$ of the wild-type HIV-1 C-SA protease towards the chromogenic HIV substrate. $K_M$ and $V_{max}$ values were found to be 193.3±35.69 µM and 0.0021±2.34×10$^{-4}$ µmol.min$^{-1}$, respectively. The assays were carried out at 20 °C in 10 mM sodium acetate buffer, 0.1 M sodium chloride, pH 5.0, and 300 nM protease enzyme. The solid lines through the data points are hyperbolic and linear fits, respectively, to the experimental data. The correlation coefficient for the linear fit is 0.95. The experiments were conducted in triplicates and the data is reported as the mean ± SD.

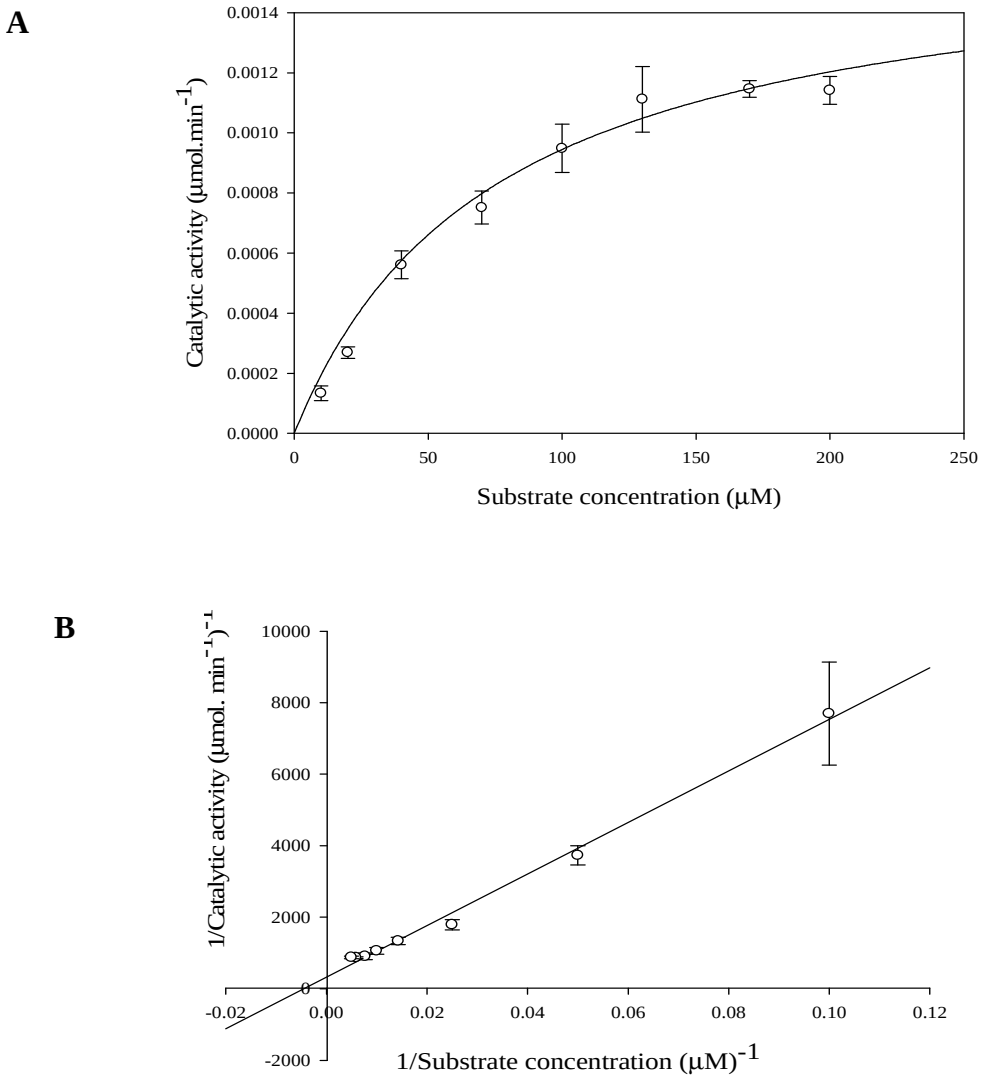



**Figure 17: Variant protease catalytic constant and maximum velocity determination**

**(A)** Michaelis-Menten plot and **(B)** Lineweaver-Burke plot for the determination of $K_M$ and $V_{max}$ of the I36TT variant HIV-1 C-SA protease towards its substrate. $K_M$ and $V_{max}$ values were found to be 89.14±12.12 µM and $0.0017 \pm 1.03 \times 10^{-4}$ µmol.min$^{-1}$, respectively. The experiments were conducted in triplicates and the data is reported as the mean ± SD. The assays were conducted at 20 °C in 10 mM sodium acetate buffer, 0.1 M sodium chloride, pH 5.0, and 300 nM enzyme. The solid lines represent hyperbolic and linear fits, respectively, to the experimental data. The correlation coefficient for the linear fit is 0.96.

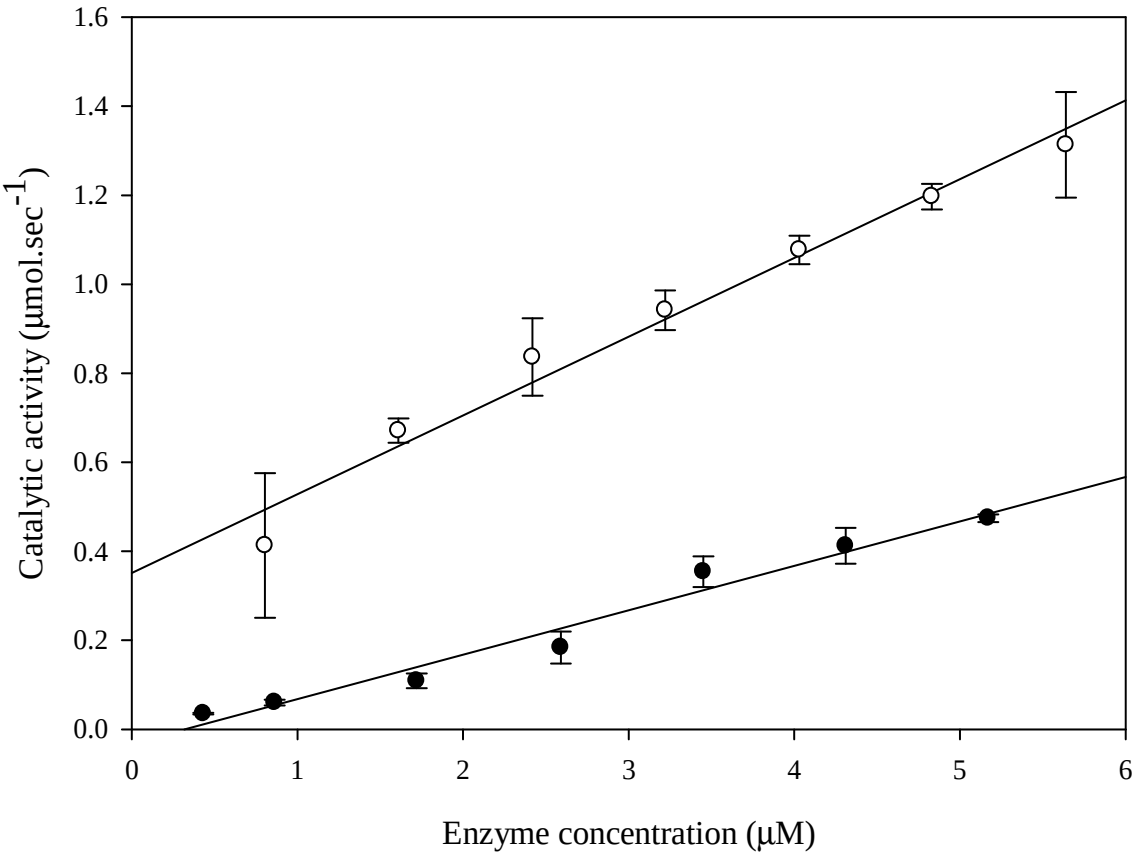



**Figure 18: Determination of enzyme turn-over number**

Linear curves for determining the turn-over number ($k_{cat}$) of the wild-type (●) and the variant (○) HIV-1 C-SA proteases. Turn-over number was determined from the slopes of the plots and was found to be $0.10\pm0.49\times10^{-2}$ $s^{-1}$ and $0.18\pm1.18\times10^{-2}$ $s^{-1}$ for wild-type and variant protease, respectively. The solid lines through the data points are linear fits to the experimental data. The correlation coefficient for the linear fit is 0.96 and 0.92 for the wild-type and mutant, respectively. The experiments were performed at 20 °C in 10 mM sodium acetate buffer, 0.1 M sodium chloride, pH 5.0, and saturating substrate concentration (i.e., 250 µM). The experiments were conducted in triplicates and the data is reported as the mean ± SD.

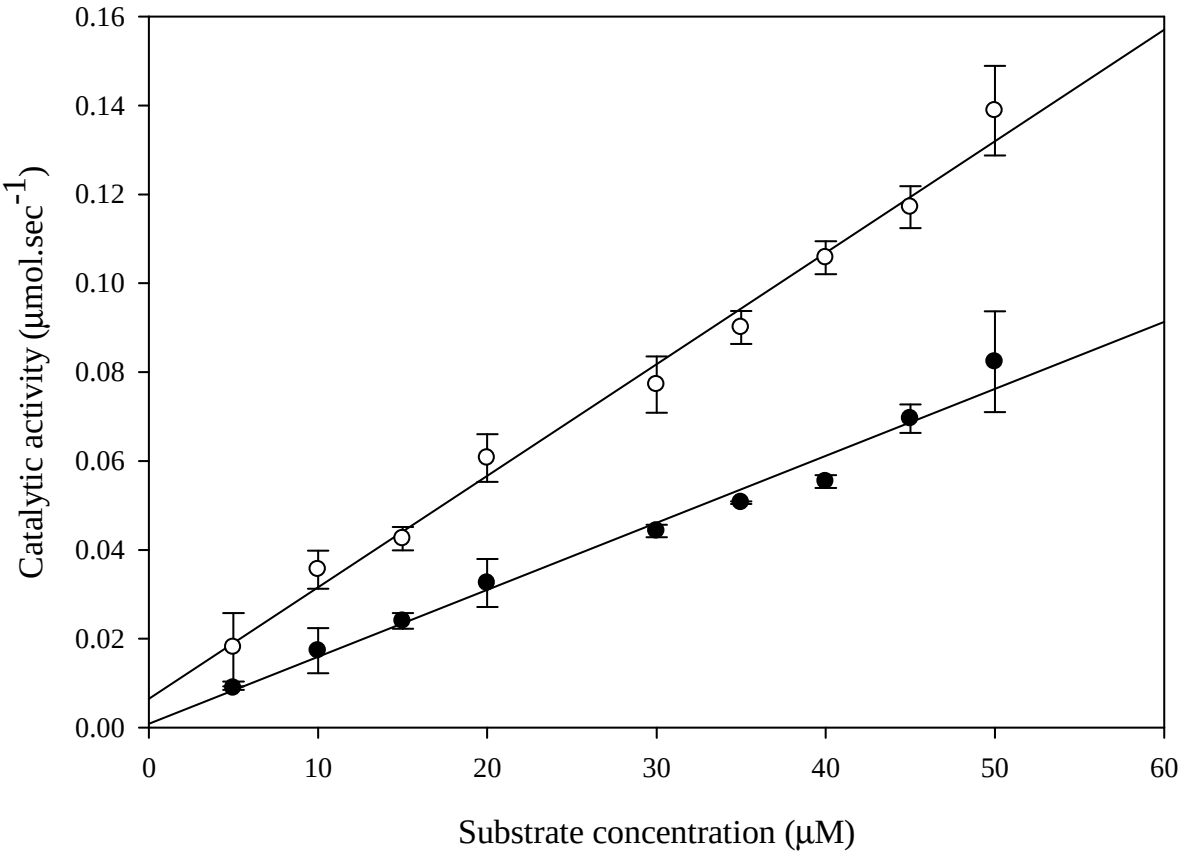



**Figure 19: Catalytic efficiency determination**

Linear curves for determining the catalytic efficiencies ($k_{cat}/K_M$) of the wild-type (●) and the variant (○) HIV-1 C-SA proteases. Catalytic efficiencies were determined by calculating the slopes of the plots and multiplying by the concentration of the enzyme in the assay. $k_{cat}/K_M$ values were found to be $0.007 \pm 6.56 \times 10^{-5}$ $\mu M^{-1}.s^{-1}$ and $0.011 \pm 8.03 \times 10^{-5}$ $\mu M^{-1}.s^{-1}$ for wild-type and variant protease enzyme, respectively. The assays were conducted at 20 °C in 10 mM sodium acetate buffer, 0.1 M sodium chloride, pH 5.0, and 300 nM enzyme. The experiments were conducted in triplicates and the data is reported as the mean ± SD. The solid lines through the data points are linear fits to the experimental data. The correlation coefficient for the linear fit is 0.95 and 0.98 for the wild-type and the mutant, respectively.

**Table 3: Summary of the catalytic parameters**

Enzymatic characteristics of the wild-type and the variant HIV-1 C-SA proteases

| **Parameter** | **Wild-type** | **Variant** |
|---|---|---|
| Specific activity (μmol/min/mg) | 2.9±0.095 | 3.1±0.11 |
| $V_{max}$ (μmol.min$^{-1}$) | 0.0021±2.34×10$^{-4}$ | 0.0017±1.03×10$^{-4}$ |
| $K_M$ (μM) | 193.3±35.69 | 89.14±12.12 |
| $k_{cat}$ (s$^{-1}$) | 0.10±0.49×10$^{-2}$ | 0.18±1.18×10$^{-2}$ |
| $k_{cat}/K_M$ (μM$^{-1}$.s$^{-1}$) | 0.007±6.56×10$^{-5}$ | 0.011±8.03×10$^{-5}$ |

### 3.5.2 Inhibition kinetics

$IC_{50}$ values of the four clinical protease inhibitors; notably, saquinavir, indinavir, ritonavir and nelfinavir, were determined following catalysis of the same chromogenic substrate. The four drugs used were the only ones available in our laboratory and are amongst some of those prescribed during HIV therapy. Wild-type and variant enzymes did not possess significant differences for saquinavir and nelfinavir (Figures 20 and 23). However, variant $IC_{50}$ values for indinavir and ritonavir were increased two-fold relative to the wild-type (Figures 21 and 22). Table 4 summarizes figures for $IC_{50}$ values of the four protease inhibitors for the variant and wild-type protease enzymes. $IC_{50}$ values for all four inhibitors were determined from the inhibition percentage versus inhibitor concentration plots. Activity percentages were calculated using the following equation:

$$\textbf{Activity } \% = \frac{\boldsymbol{V_0}\textbf{of inhibited reaction}}{\boldsymbol{V_0}\textbf{of uninhibited reaction}} \times \mathbf{100} \qquad \textbf{Equation 9}$$

where $V_0$ is the enzyme catalyzed reaction rate.

## 3.6 Thermodynamic characteristics

Thermodynamic parameters of the wild-type HIV-1 C-SA protease and the I36TT variant protease were determined using calorimetric assays. All binding parameters i.e., Gibbs free energy, change in enthalpy, change in entropy and binding affinity were obtained (Table 5) and did not show any significant differences in thermodynamic parameters between the wild-type and the I36TT variant protease. Figure 24 represents an ITC profile obtained from 20 μM HIV-1 C-SA protease binding to acetyl pepstatin (300 μM).

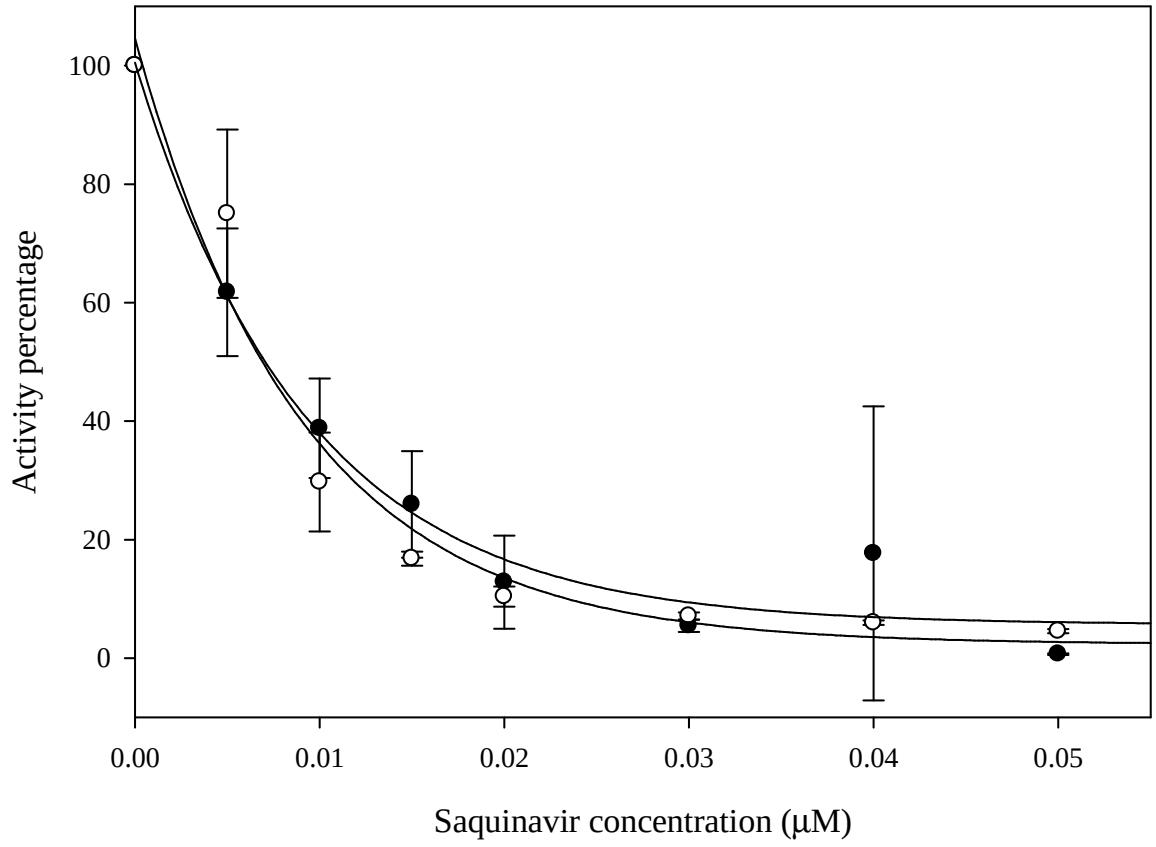



**Figure 20: Determination of $IC_{50}$ value for saquinavir**

Graph used to determine the $IC_{50}$ value (inhibitor concentration needed to decrease the catalytic activity of the enzymes by 50%) of the wild-type (●) and the variant (○) HIV-1 C-SA protease. $IC_{50}$ values for wild-type and mutant were 7.6 nM and 6.3 nM, respectively. The solid lines represent exponential decay fits to the experimental data. The experiments were performed at 20 °C using enzyme concentration of 300 nM, saturating substrate concentration (i.e., 250 μM) and increasing saquinavir concentrations in 10 mM sodium acetate buffer, pH 5.0. Inhibition percentages were calculated using Equation 9. All experiments were conducted in triplicates and the data is reported as the mean ± SD.

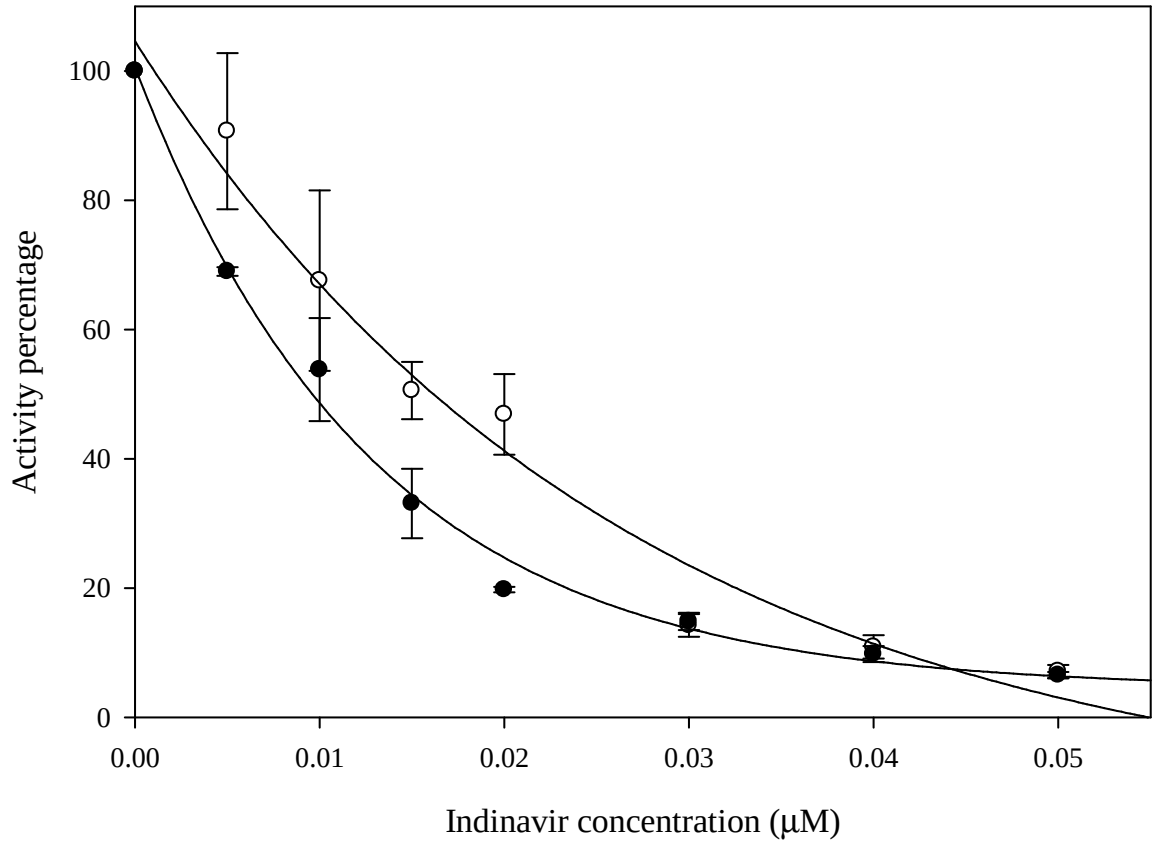



**Figure 21: Indinavir $IC_{50}$ value determination**

Plots used to determine the $IC_{50}$ values of indinavir with the wild-type (●) and the mutant (○) HIV-1 C-SA protease enzyme. $IC_{50}$ values were 9.5 nM and 16.2 nM for wild-type and mutant protease enzymes, respectively. The final concentrations in the assays were: 300 nM enzyme, saturating (250 μM) substrate concentration and increasing indinavir concentrations in 10 mM sodium acetate buffer, pH 5.0 at 20 °C. All experiments were conducted in triplicates and the data is reported as the mean ± SD. The solid lines through the data points are exponential decay fits to the experimental data.

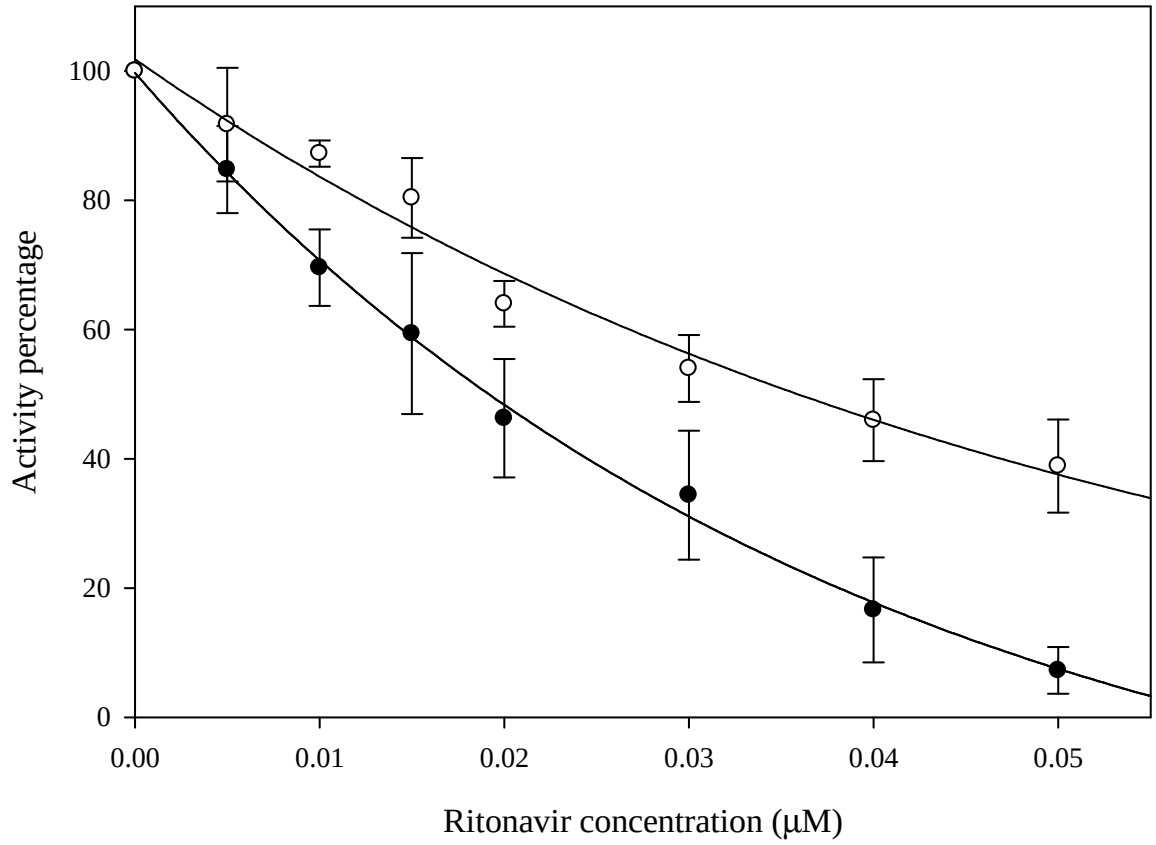



**Figure 22: Ritonavir $IC_{50}$ value determination**

$IC_{50}$ values of ritonavir were determined in the presence of the wild-type (●) and the variant (○) HIV-1 C-SA protease enzymes using these plots. $IC_{50}$ values for wild-type and mutant were 19.1 nM and 36.4 nM, respectively. The solid lines represent exponential decay fits to the experimental data. The assays were performed at 20 °C using 300 nM enzyme concentration, saturating substrate concentration (i.e., 250 μM) and increasing ritonavir concentrations in 10 mM sodium acetate buffer, pH 5.0. All experiments were conducted in triplicates and the data is reported as the mean ± SD.

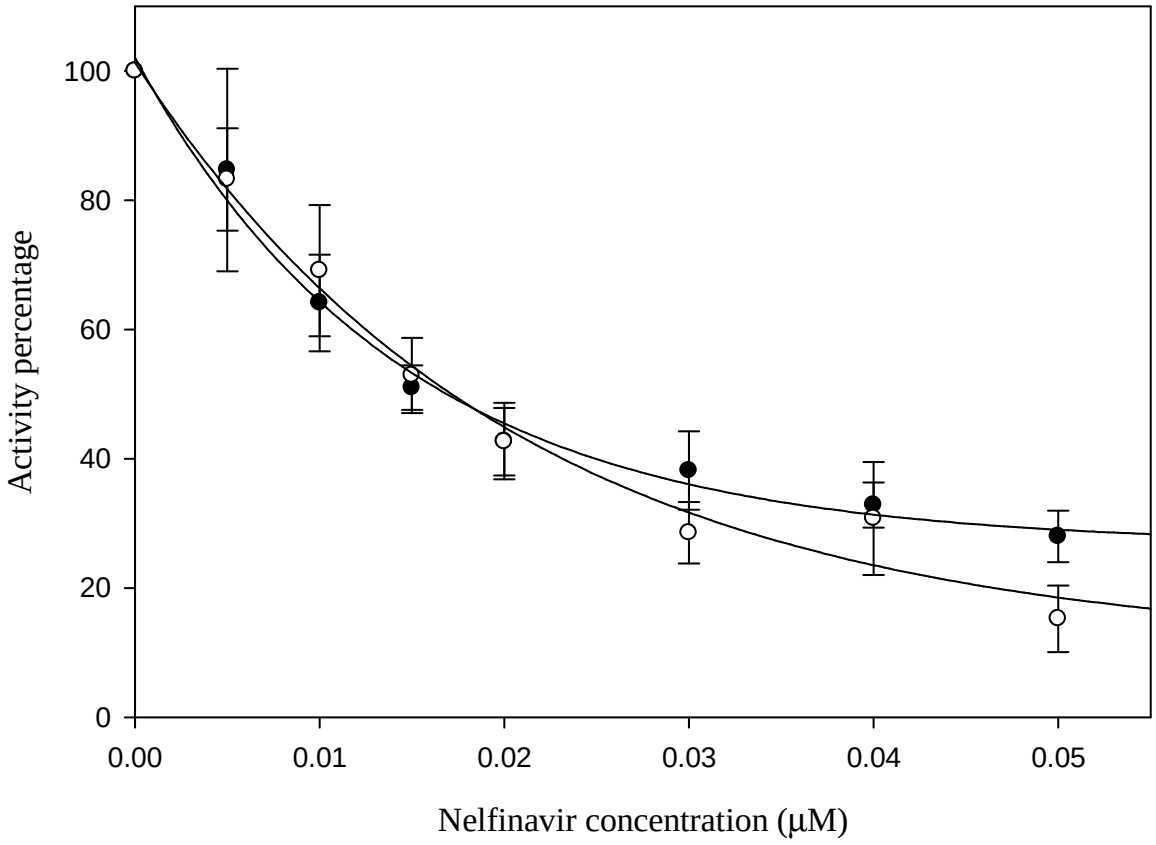



**Figure 23: Nelfinavir $IC_{50}$ value determination**

Graph used to determine the inhibitor concentration needed to decrease the catalytic activity of the wild-type (●) and the variant (○) HIV-1 C-SA protease enzymes by 50%. ($IC_{50}$ values). $IC_{50}$ values were 17.3 nM and 16.8 nM for wild-type and mutant protease, respectively. The experiments were carried-out using enzyme concentration of 300 nM, saturating (250 μM) substrate concentration and increasing nelfinavir concentrations in 10 mM sodium acetate buffer, pH 5.0 at 20 ° C. All experiments were conducted in triplicates and the data is reported as the mean ± SD. The solid lines through the data points represent exponential decay fits to the experimental data.

**Table 4: Summary of inhibitors $IC_{50}$ values**

Inhibition characteristics ($IC_{50}$ values) of the wild-type and the variant HIV-1 C-SA proteases

| | **$IC_{50}$ value (nM)** | | | |
|---|---|---|---|---|
| | Saquinavir | Indinavir | Ritonavir | Nelfinavir |
| **Wild-type** | 7.6 | 9.5 | 19.1 | 17.3 |
| **Variant** | 6.3 | 16.2 | 36.4 | 16.8 |

**Table 5: Thermodynamic parameters of binding**

Wild-type and variant HIV-1 C-SA PR binding parameters were determined using calorimetric assays and the inhibitor acetyl pepstatin

| Parameter | Wild-type | Variant |
|---|---|---|
| $\Delta G$ (kcal/mol) | -8.0 | -8.1 |
| $\Delta H$ (kcal/mol) | 6.2 | 6.5 |
| $-T\Delta S$ (kcal/mol) | -14.0 | -14.6 |
| $K_d$ (nM) | 380 | 395 |

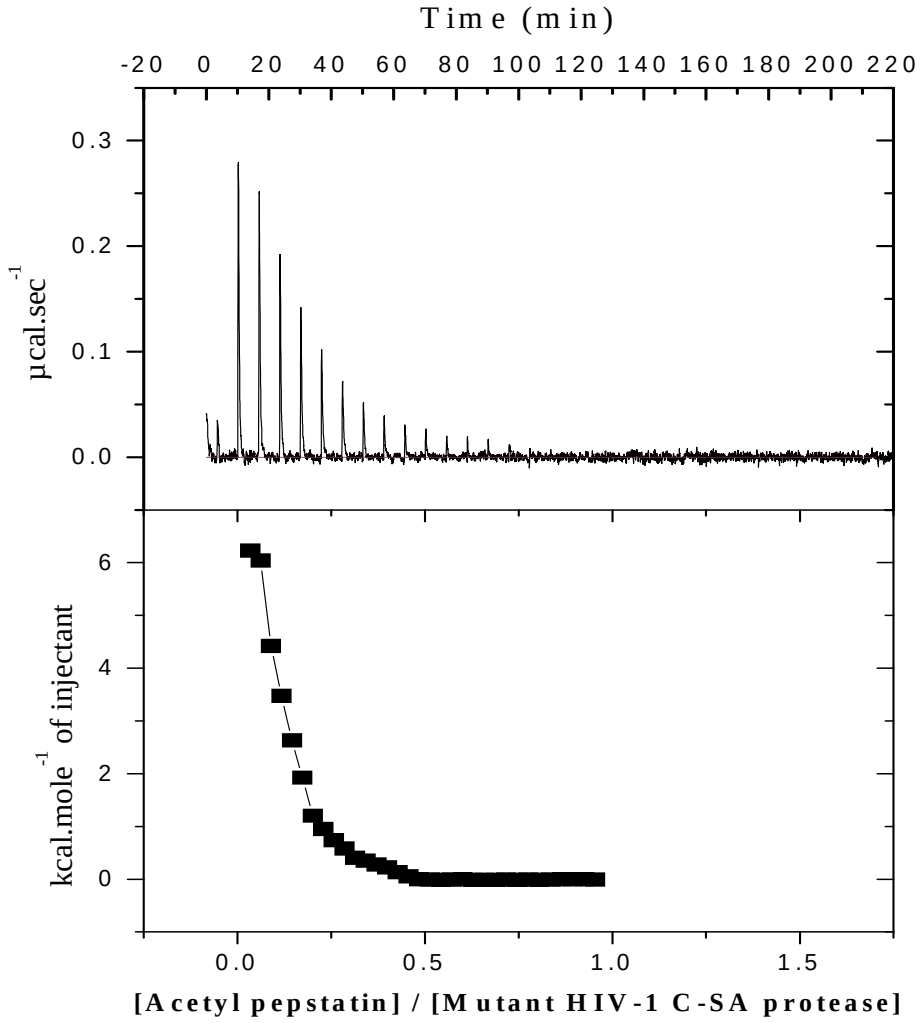



**Figure 24: Representation of the HIV-1 C-SA protease binding isotherm**

Isothermal titration calorimetric plot of the I36TT mutant HIV-1 C-SA protease titrated with 10 µl injections of 300 µM acetyl-pepstatin. Initial enzyme concentration in the ITC sample cell was 20 µM in 10 mM sodium acetate buffer, pH 5.0. Acetyl-pepstatin in the ITC syringe was also dissolved in the same buffer. The top panel shows the heat trace of the thermostat over the time of the experiment with individual injections of the inhibitor as peaks. The lower panel shows the integrated area of the peaks plotted against molar ratio of the inhibitor to protease. Data fitted using ITCone from ORIGIN v5.0 (MicroCal Inc.).

# CHAPTER 4

# 4 DISCUSSION

The current study assessed the effects of an amino acid substitution and insertion plus background polymorphisms in the overall structure and catalytic ability of HIV-1 C-SA PR. The variant plasmid encoding HIV-1 C-SA PR, containing the I36TT amino acid substitution/insertion plus background polymorphisms, was generated using site-directed mutagenesis and the plasmid encoding the wild-type as a template. DNA sequencing results revealed that only the desired mutations were incorporated in the wild-type plasmid to yield variant plasmid. The wild-type and variant HIV-1 C-SA PRs were expressed as inclusion bodies in *Escherichia coli* BL21 (DE3) pLysS cells. The variant PR was expressed in higher yields (6 mg/mℓ) as compared to the wild-type PR (4 mg/mℓ). Both PRs were purified using a DEAE anion exchange column containing an equilibration buffer at pH 8.0. Both PRs were isolated to absolute purity and homogeneity (Figure 10). The above results are in agreement with previously determined findings of HIV-1 PR (Ido *et al*., 1991; Mosebi *et al*., 2008).

## 4.1 Structural content assessment

### 4.1.1 Secondary structure

Circular dichroism spectroscopy provides information about differential absorption between left- and right-handed circularly polarized light in optically-active media (Lowry, 1935). The optically active elements in polypeptides and proteins are the peptide backbone, amide interactions (α-helices and β-sheets), and the side chains of aromatic amino acids (Woody, 1995). Manning *et al*. (1988) demonstrated that even small conformational distortions in the backbone of a polypeptide or protein can result in strong alterations of the circular dichroism signal. Thus, highly sensitive circular dichroism was used in this study to monitor conformational changes in the protein backbone. Possible changes in the secondary structural content of the variant and the wild-type South African HIV-1 subtype C protease were probed using circular dichroism spectroscopy in the far-UV (250 to 190 nm) range. Although equal concentrations of the wild-type and the mutant proteins were used, far-UV CD spectra of the

variant protein showed a CD signal with 10% reduction in intensity relative to the wild-type (Figure 12). The reduced CD signal is correlated to the minor structural (peptide back-bone) alterations induced by the mutations in the variant protease. Nonetheless, both the I36TT variant and wild-type protease exhibited far-UV CD spectra with minima at 216 nm. The observed negative peak at 216 nm is typical of a predominantly β-sheeted polypeptide/protein. This is in agreement with previous studies (Navia *et al.,* 1989; Dash and Rao, 2001; Foulkes *et al.,* 2006; Mosebi *et al*., 2008) which characterized the secondary structure of the HIV-1 protease using far-UV circular dichroism as a probe and reported that the protease is 48% β-sheet, 4% α-helix, and 48% aperiodic conformation. This can also be confirmed by studying the crystal structure of the HIV-1 protease (Figure 6). The negative peak near 220 nm is due to the $n$-$\pi^*$ transition of the peptide backbone and it is very sensitive to conformational alterations occurring in the polypeptide chain (Hennessey and Johnson, 1981; Modi *et al*., 1993). Hence, the secondary structural elements of the HIV-1 C-SA protease were not significantly perturbed by the presence of I36TT mutation and its background polymorphisms.

### 4.1.2 Tertiary structure

Fluorescence spectroscopic assays are sensitive to changes in local environments; hence, intrinsic tryptophan fluorescence is employed to investigate the localized conformational changes induced in the tertiary structure of the protein. Conformational changes occurring in the tertiary structure of the protein are monitored using intrinsic fluorescence of the phenyl and indole ring of tyrosine and tryptophan, respectively, as probes (Shirley, 1995). HIV-1 C-SA protease has two tryptophan residues, at positions 6 and 42, in each subunit. Thus, possible changes induced by the I36TT mutation and its background mutations in the tertiary structure of the HIV-1 C-SA protease were probed using intrinsic tryptophan fluorescence. These tryptophan residues are located where they will not detect major/global structural changes (see Figure 6 for location of the tryptophan residues in the HIV-1 protease) and are quite exposed to the solvent. Therefore, they will be used to monitor only local structural changes.

The fluorescence emission spectra of both Trp and Tyr residues were monitored after excitation at 280 nm, and Trp fluorescence was selectively excited at 295 nm (Figures 13 A and B).

Although excitation of the variant protein exhibited about 20% enhanced fluorescence emission intensity as compared to the wild-type, there were no Stokes' (red and/or blue) shifts observed for both proteins at both excitation wavelengths. Intrinsic tryptophan emission fluorescence spectra of the variant and wild-type protease, at both excitation wavelengths (i.e., 280 and 295 nm) were similar, with maximum intensity at 355 nm. This shows that the tryptophan residues in the fully folded native HIV-1 C-SA are highly exposed to the solvent (Figure 6) since free tryptophan fluoresces maximally between 350 and 355 nm in water (Lakowicz, 1999). Even though both proteins exhibited maximum fluorescence intensities at the same wavelengths, the wild-type protein displayed about 20% reduction in fluorescence intensity relative to the mutant. The enhanced intensities may be due to the fact that some of the tryptophan energies, which were quenched in the wild-type protein are no longer quenched in the variant protein. This suggests that there might be some minor alterations induced in the structure of the HIV-1 C-SA protease by the I36TT mutation and its background polymorphisms. These findings are in agreement with a study (Kozisek *et al*., 2008) which compared the structures of mutant HIV-1 subtype B protease containing a single amino acid insertion each (E35EE and L33LL) to the wild-type, and reported that insertion does not significantly change the overall inhibitor binding to the active site for both mutant proteins. For this reason, homology modeling and crystal structures of the wild-type and the I36TT variant protein can report on the degree of conformational alterations induced in the overall structural content of the HIV-1 C-SA protease by the I36TT mutation and its background polymorphisms.

## 4.2 Enzymatic characteristics

### 4.2.1 Catalytic properties

Biochemical evidence has shown that HIV-1 protease catalyzes the proteolysis of *gag* and *pol* gene during formation of virion particle in the life-cycle of HIV (Luban *et al*., 1993; Debouck *et al*., 1987 and Katoh *et al*., 1985) and is highly specific to substrate recognition (Luban *et al*., 1993). Herein, substrate binding and catalytic activity of I36TT variant protease was assessed *in vitro* using a chromogenic HIV-1 substrate which resembles HIV-1 protease cleavage site in the *gag-pol* polyprotein precursor. Comparing the substrate binding and catalytic

properties of the I36TT variant protease to that of the wild-type, it was discovered that the variant enzyme has a greater affinity for the substrate, with $K_M$ values of 193 μM and 89 μM for the wild-type and variant enzyme, respectively (Figures 16 and 17). See Table 3 for a summary of all the determined kinetic parameters. However, maximum velocities attained by the reactions catalyzed by these two enzymes were similar, with $V_{max}$ values of 0.0021 μmol.min$^{-1}$ and 0.0017 μmol.min$^{-1}$ (Figures 16 and 17) for the wild-type and the variant enzyme, respectively.

Catalytic activities of the wild-type and the variant protease enzyme were also assessed by determining the catalytic efficiency, $k_{cat}/K_M$, and turnover number, $k_{cat}$. The latter is used to calculate how much of the substrate compound is converted to product per enzyme active site while the former measures how much substrate is converted to product per time. The I36TT variant enzyme proved to be more catalytically active as it converted more substrate compounds into product comparative to the wild-type. Turnover number and catalytic efficiency of the variant enzyme was increased almost two-fold as compared to that of the wild-type enzyme. This was demonstrated by $k_{cat}$ values of 0.10 s$^{-1}$ and 0.18 s$^{-1}$ (Figure 18) and $k_{cat}/K_M$ values of 0.007 μM$^{-1}$.s$^{-1}$ and 0.011 μM$^{-1}$.s$^{-1}$ (Figure 19) for the wild-type and the variant protease enzyme, respectively. Improved substrate binding, indicated by reduced $K_M$, by the I36TT is implicated in the increased enzyme activity. However, these results are in contrast with those reported by Kozisek *et al*. (2008) which showed over five-fold decrease in the enzyme activity of an insert-containing mutant protease (E35EE) relative to the wild-type HIV-1 subtype B.

Taking into consideration the concentrations of the substrate and enzyme used in the reaction and the time taken by the reaction, specific activity of the enzymes was determined. Specific activity results showed a 7% difference in the catalysis of substrate hydrolysis by the I36TT variant protease relative to the wild-type. This was indicated by specific activity value of 2.9 μmol.min$^{-1}$.mg$^{-1}$ for the wild-type enzyme and 3.1 μmol.min$^{-1}$.mg$^{-1}$ for the variant. Catalytic characteristics determined in this study for the HIV-1 C-SA protease enzyme revealed that the I36TT mutation and background does not reduce the catalytic efficacy of the C-SA protease, but rather improves it slightly (7%).

### 4.2.2 Inhibition by ARVs

Enzymatic characteristics of the wild-type and the I36TT variant protease were also assessed in the presence of four protease inhibitors employed in clinical settings during HIV-1 therapy. Nelfinavir, saquinavir, indinavir and ritonavir were individually included in the assays to assess the efficiency of the HIV-1 C-SA protease enzyme to catalyze hydrolysis of its substrate in the presence of inhibitors. $IC_{50}$ values (amount of inhibitor needed to reduce the catalytic activity of the enzyme by 50%) for individual protease inhibitors were determined for the wild-type and the I36TT variant enzymes.

Mutations at position 36 of the protease are associated with resistance to the protease inhibitor tipranavir in the subtype B HIV- 1 (de Mendoza *et al*., 2007). Insertions at this position lead to a decrease in susceptibility of the protease to the tipranavir and consequently improve viral fitness in the presence of this inhibitor. However, it is more likely that additional factors are also involved in the PR resistance to inhibitor caused by insertion mutations at this position. Comparison of insertion mutant structure to the subtype B wild-type has suggested that the insertion may influence the dynamics of flap movements during inhibitor binding and, thus, impaired flap movement may be implicated in weaker inhibitor binding in insert-containing mutants. However, this is not a general phenomenon for all protease inhibitors. Other insertion mutations detected in the protease region showed different response to various PIs. The M36TNL showed no resistance for all PIs tested and it was, therefore, proposed that this insertion does not contribute to resistance towards most of the FAD-approved protease inhibitors (Sturmer *et al*., 2003).

Herein, like the wild-type, the variant enzyme did not show much resistance to the inhibitors. Catalytic activity of the C-SA proteases was significantly hindered by saquinavir with $IC_{50}$ values of 7.6 nM and 6.3 nM for the wild-type and the variant enzyme, respectively (Figure 20). However, the enzymes proved to be slightly more resistant to nelfinavir with $IC_{50}$ values of 17.3 nM for the wild-type enzyme and 16.8 nM for the variant (Figure 23). Additionally, the I36TT variant retained most of its enzymatic characteristics, with $IC_{50}$ value of 36.4 nM, as compared to the wild-type protease ($IC_{50}$ value of 19.1 nM) in the presence of ritonavir (Figure 22). The same effects ($IC_{50}$ value increased by 58%) were observed for the variant enzyme in the

presence of indinavir (with $IC_{50}$ value of 16.2 nM) relative to the wild-type with $IC_{50}$ value of 9.5 nM (Figure 21). See Table 4 for a summary of all the determined $IC_{50}$ values.

Comparing these results to others reported for other insert-containing protease mutants (Kim *et al*., 2001; Sturmer *et al*., 2003 and Kozisek *et al*., 2008), it may be proposed that amino acid insertion mutations confer PI resistance. Previous studies (Kozisek *et al*., 2008) assessed van der Waals distances between inhibitor (lopinavir) and residues within individual substrate binding subsites and they revealed a slight enlargement of the binding pockets due to rearrangement of neighbouring residues caused by insertions. This enlargement results specifically in the loss of van der Waals contacts between the amino acid side chain and the benzyl ring of lopinavir, and hence, weaker binding of the PI to the active site. Thus, this study reports that the I36TT insertion/mutation and its background polymorphisms do not show large differences in susceptibility to protease inhibitors but proffers protease considerable resistance towards ritonavir and indinavir. Further experimental work would be necessary to describe the role of background polymorphisms accompanying amino acid insertions in the protease coding region of HIV.

## 4.3 Conclusion

This study reports that the I36TT substitution/insertion and its background polymorphisms (P39S, D60E and Q61E), in the protease coding region of the South African HIV-1 subtype C, does not induce gross conformational alterations in the secondary and tertiary structure of the protein. The I36TT substitution/insertion mutant showed more than two-fold increase in affinity for the substrate and about two-fold increase in substrate turnover. The capability of the protease enzyme to cleave its substrate in the presence of various protease inhibitor drugs revealed that the I36TT variant protease is still catalytically active at ritonavir concentrations over 30 nM. These secondary mutations (mutations occurring outside the substrate binding cleft of the protease enzyme) have little effect on overall structure and inhibitor binding but they improved the proteolytic efficiency of the HIV-1 C-SA protease enzyme and hence improved viral maturation and replication rate.

# CHAPTER 5

# 5 REFERENCES


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

AIDSinfo. (2009) http://www.aidseducator.org/factsheets/106-HIV-Life-Cycle.html. Date accessed: 10 December 2009.

Chung, C. T., Niemela, S. L. and Miller, R. H. (1989) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2172-2175.

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

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

Cozzi-Lepri, A., Phillips, A. N., Ruiz, L., Clotet, B., Loveday, C., Kjaer, J., Mens, H., Clumeck, N., Viksna, L., Antunes, F., Machala, L. and Lundgren, J. D. (2007) Evolution of drug resistance in HIV infected patients remaining on a virologically failing combination antiretroviral therapy regimen. *AIDS.* **21**, 721-732.

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

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

Debouck, C. and Metcalf, B. W. (1990) Human immunodeficiency virus protease: a target for AIDS therapy. *Drug. Dev. Res.* **21**, 1-17.

de Mendoza, C., Morello, J., Garcia-Gasco, P., Rodriguez-Novoa, S. and. Soriano, V. (2007) Tipranavir: a new protease inhibitor for the treatment of antiretroviral-experienced HIV-infected patients. *Expert Opin. Pharmacother*. **8**, 839-850.

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

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

Foulkes, J. E., Prabu-Jeyabalan, M., Cooper, D., Henderson, G. J., Harris, J., Swanstrom, R. and Schiffer, C. A. (2006) Role of invariant Thr80 in human immunodeficiency virus type 1 protease structure, function, and viral infectivity. *J. Virol.* **80**, 6906-6916.

Gallo, R. C. and Montagnier, L. (1988) AIDS in 1988. *Sci. Am.* 25941-25948.

Gomez, J. and Freire, E. (1995) Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. *J. Mol. Biol.* **252**, 337-350.

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

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

Grant, R. M., Kahn, J. O., Wrin, T., Drews, B., Javier, J., Webb, M., Petropolous, C. J. and Hecht, F. M. (2001) HIV-1 with an insertion in protease is drug susceptible, replication-competent and transmissible. *Antivir. Ther*. **6** (Suppl. 1), 44.

Guex, N., Peitsch, M., Schwede, T. and Diemand, A. (1995) Swiss-Model and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis.* **18,** 2714-2723.

Gulnik, S. V., Suvorov, L. I., Liu, B., Yu, B., Anderson, B., Mitsuya, H. and Erickson, J. W. (1995) Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry.* **34**, 9282-9287.

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

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**, 24359-24366.

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

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

Kantor, R., Shafer, R. W., Follansbee, S., Taylor, J., Shilane, D., Hurley, L., Nguyen, D., Katzenstein, D. and Fessel, W. J. (2004) Evolution of resistance to drugs in HIV-1 infected patients failing antiretroviral therapy. *AIDS.* **18**, 1503-1511.

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. *J. Virol*. **145**, 280-292.

Katoh, I., Ikawa, Y. and Yoshinaka, Y. (1989) Retrovirus protease characterized as a dimeric aspartic proteinase. *J. Virol*. **63**, 2226-2232.

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

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. USA*. **85**, 4686-4690.

Kovalevsky, A. Y., Tie, Y., Liu, F., Boross, P. I., Wang, Y. F., Leshchenko, S., Ghosh, A. K., Harrison, R. W. and Weber, I. T. (2006) Effectiveness of nonpeptide clinical inhibitor TMC-114 on HIV-1 protease with highly drug resistant mutations D30N, I50V, and L90M. *J Med Chem*. **49**, 1379-1387.

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**, 5869-5878.

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

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

Lorenzi, P. M., Opravil, B., Hirschel, J. P., Chave, H. J., Furrer, H., Sax, T. V., Perneger, L., Perrin, L. and Kaiser, S. (1999) Impact of drug resistance mutations on virologic response to salvage therapy. Swiss HIV Cohort Study. *AIDS.* **13**, 17-21.

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

Lowry, T. M. (1935) Optical rotatory power. London: Longmans Green and Co. LTD

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

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

Masquilier, B., Race, E., Tamalet, C., Descamps, D., Izopet, J., Buffet-Janvresse, C., Ruffault, A., Mohammed, A. S., Cottalorda, J., Schmuck, A., Calvez, V., Dam, E., Fleury, H., Brun-Vezinet, F. and the ANRS AC11 Resistance study group. (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**, 1836-1842.

MicroCal Inc. (2009) http://www.microcalorimetry.com/technology/itc.asp. Date accessed: 1 February 2010.

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. C., Smith, C. W., Heinrikson, R. L., and Tomasselli, A. G. (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,** 9405-9413.

Miller, M., Jaskolski, M., Rao, J. K., Leis, J. and Wlodawer, A. (1989) Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature.* **337**, 576-579.

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

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

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

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

Okimoto, N., Hata, M., Hoshino, T. and Tsuda, M. (2000) Protein hydrolysis mechanism of HIV-1 protease: Investigation by the ab initio MO calculations. *RIKEN Rev.* **29**, 100-102.

Pain, R. (2004) Determining the CD spectrum of a protein. In: Current protocols in Protein Science. Vol. 1, Pp 7.6.1-7.7.20. John Wiley and Sons, Inc., New York, USA.

Papworth, C., Bauer, J. C., Braman, J. and Wright, D. A. (1996) Site-directed mutagenesis using double-stranded plasmid DNA templates. *Strategies.* **9**, 3-4.

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

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

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn. G. H., Leimer, T., McCutchan, F., Osmanov, S., Peeters, M., Pieiazek, D., Salminen, M., Sharp, P. M., Wolinsky, S. and Korber, B. (2000) HIV-1 nomenclature proposal. *J. Sci.* **288**, 55-56.

Rodriguez-Barrios, F. and Gago, F. (2004) Chemometrical identification of mutations in HIV-1 reverse transcriptase conferring resistance or enhanced sensitivity to arylsulfonylbenzonitriles. *J Am Chem*. **126**, 2718-2719.

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

Schinazi, R. F., Larder, B. A. and Mellors, J. W. (1997) Mutations in retroviral genes associated with drug resistance. *Int. Antiviral. news*. **5**, 129-142.

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. USA.* **85**, 6612-6616.

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

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-naïve patient is not associated with indinavir treatment failure. *AIDS Res. Hum. Retr*. **19**, 967-968.

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

Todd, M. J. 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. J. and Gomez, J. (2001) Enzyme kinetics determined using calorimetry: a general assay for enzyme activity. *Anal Biochem*. **15**, 179-187.

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

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

van Holde, K. E., Johnson, W. C and Ho, P. S. (1998) Principles of physical biochemistry. Pp. 452-456. Prentice Hall, Inc., New Jersey, USA.

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

Velazquez-Campoy, A., Kiso, Y. and Freire, E. (2001b) 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., 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.

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

Walsh, C. T. (2003) Proteases: Hydrolysis of peptide bonds: Specificity and mechanism. *Science*. **133**, 1-17.

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

Weiss, A., Hollander, H. and Stobo, J. (1985) Acquired immunofefeciency syndrome: epidemiology, virology, and immunology. *Annu. Rev. Med.* **36**, 545-562.

Wensing, A. M., van der Vijver, D. A., Angarano, G., Asjo, B., Balotta, C., Boeri, E., Camacho, R., Chaix, M. L., Costagliola, D., De Luca, A., Derdelinckx, I., Grossman, Z., Hamouda, Z., Hatzakis, A., Hemmer, R., Hoepelman, A., Horban, A., Korn, K., Kucherer, C., Leitner, T., Loveday, C., MacRae, E., Maljkovic, I., de Mendoza, C., Meyer, L., Nielsen, C., Op de Coul, E. L., Ormaasen, V., Paraskevis, D., Perrin, L., Puchhammer-Stockl, E., Ruiz, L., Salminen, M., Schmit, J. C., Schneider, F., Schuurman, R., Soriano, V., Stanczak, G., Stanojevic, M., van Damme, A. M., Van Laethem, K., Violin, M., Willbe, K., Yerly, S., Zazzi, M. and Boucher, C. A. (2005) Prevalence of drug resistant HIV-1 variants in untreated individuals in Europe: Implications for clinical management. *J. Infect. Dis.* **192**, 958-966.

Winters, M. A. and Merigan, T. C. (2001) Insertions in the human immunodeficiency virus type 1 protease and reverse transcriptase genes: Clinical impact and molecular mechanisms. *Antimicrob. Agents Chemother.* **49**, 2575-2582.

Winters, M. A., Kagan, R. M., Kovari, L., Heseltine, P. N. and Merigan, T. C. (2005) Rare one and two amino acid inserts adjacent to codon 103 of the HIV-1 reverse transcriptase (RT) affects susceptibility to non-nucleoside RT inhibitors. *Antivir. Ther*. **10**, 363-366.

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

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

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

World Health Organization. (2008) Global HIV/AIDS statistics. http://www.who.org. Date accessed: 12 July 2009.