

https://dspace.org/

The contribution of F99 to the structure and function of South African HIV-1 subtype C protease

Seele, Palesa Pamela

2013-01-29
http://hdl.handle.net/10539/12317

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

The contribution of F99 to the structure and function of South African HIV-1 subtype C protease

Palesa Pamela Seele

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

Johannesburg, 2012

DECLARATION

I 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.



Palesa Pamela Seele

18 day of June, 2012

ABSTRACT

The HIV/AIDS still remains a global health challenge with 42 million people infected with the virus. An alarming 70% of these people reside in sub-Saharan Africa with HIV-1 subtype C being the most prevalent subtype in this region and globally. HIV-1 protease (PR) is an obligate homodimer which plays a pivotal role in the maturation and hence propagation of the HI virus. Although successful developments on PR active site inhibitors have been achieved, the major limiting factor has been the emergence of HIV drug resistant strains. It has been postulated that disruption/dissociation of the dimer interface may lead to an inactive enzyme. The development of small molecules and peptides has been a major research area with the key target being the N- and C-termini antiparallel β-sheet. This is due to its highly conserved nature and because it consists of a cluster of amino acids that contribute most of the binding energy and stability of the dimer interface. Hence it is referred to as a 'hot-spot'. Therefore, binding of protease inhibitors at this site could cause destabilisation and/or dissociation of the enzyme. The terminal residue, F99, was mutated to an alanine disrupting the presumed lock-and-key motif it forms and in turn creating a cavity at the N- and C-termini antiparallel β-sheet. A second mutant, W42F/F99A, was created for monitoring tertiary structural changes exclusively at the N- and C-termini antiparallel β-sheet. The F99A and W42F/F99A, compared to the wild-type, showed a higher expression yield and also migrated further when separated using tricine SDS-PAGE. Wild-type protease CD spectra showed a minimum at 214 nm and a local maximum at 230 nm, while the mutants exhibited minima at 203 nm and absence of the local maxima. A 50% higher fluorescence intensity and a 2 nm red-shift for the mutants versus the wild-type was observed. According to SE-HPLC data the relative molecular weight of the wild-type, F99A and W42F/F99A are 16.4 kDa, 20.7 kDa and 18.1 kDa, respectively. Although the thermal unfolding of all three proteases was irreversible, the unfolding transition of the wild-type was clearly defined between 55 °C and 63 °C. The F99A and W42F/F99A unfolding curves were linear without clearly defined transition states. The specific activity of the F99A (0.13 µmol/min/mg) amounted to a ten-fold reduction compared to the wild-type (1.5 µmol/min/mg). The substrate binding affinity (K_M) for the F99A was 41% lower than the wild-type when 2 µM of protein was used.

The V_{max} and k_{cat} values were about 30-fold and two-fold, respectively, higher for the wild-type when compared to the F99A. Therefore, the tricine SDS-PAGE analysis, secondary and tertiary structural characterisation and thermal denaturation curve showed that the F99A mutation has altered the structure causing ‘partial’ unfolding of the protein. But, the protein still maintained minute activity. The overlap between the ANS binding spectra of the wild-type and variants suggests that the dimeric form still exists.

DEDICATIONS

This work is dedicated to:

My Family 'Basia batubatsi'

And special dedications to my great-grandfather Nimrod Yedwa Tubane, grandparents Matshidiso Job and Mary Violet 'mummy' Seele

Your legacies are engraved in my being

My late friend Pamela Mamosebetsi Mdhluli

'Genius is patience'
Sir Isaac Newton

ACKNOWLEDGEMENTS

My supervisor, Dr Yasien Sayed and advisor, Prof. Heini Dirr for giving me the opportunity, inspiration and support to do this project. To Dr Ikechukwu Achilonu my co-supervisor thanks for the patience, guidance and for challenging me.

To Dr Salerwe Mosebi for his input at the beginning of this project

National Research Foundation for the funding

Members of the PSFRU for all the help, advice and patience

My family, friends (my Vegas family) thank you for keeping me sane, without them this journey would be pointless

RESEARCH OUTPUT

Oral presentation

Author: Palesa Pamela Seele

Co-authors: Dirr Heini, Achilonu Ikechukwu and Sayed Yasien

Title: Probing interactions at the dimer interface required for HIV-1 subtype C protease conformational stability and function.

Postgraduate Cross Faculty Symposium, 27 October 2010

University of the Witwatersrand, Johannesburg

TABLE OF CONTENTS

DECLARATION ii
ABSTRACT iii
DEDICATIONS v
ACKNOWLEDGEMENTS vi
RESEARCH OUTPUT vii
TABLE OF CONTENTS viii
LIST OF FIGURES xi
LIST OF TABLES xiii
ABBREVIATIONS xiv
CHAPTER 1 INTRODUCTION 1
1.1 HIV/AIDS 1
1.2 HIV life-cycle and drug therapy 3
1.3 The common features and organisation of the dimer interface of proteins 6
1.3.1 The physicochemical properties of the dimer interface 6
1.3.2 Contribution of 'hot-spots' and conserved residues at the dimer interface 7
1.4 The conformation of β-sheets and role at the dimer interface 8
1.5 The structure and catalytic mechanism of HIV-1 protease 9
1.6 The N- and C-termini antiparallel β-sheet of HIV-1 protease: influence on structure, function and stability 13
1. 7 Objective and aims 17
CHAPTER 2 EXPERIMENTAL PROCEDURES 19
2.1 Materials 19
2.2 The HIV-1 expression vector 19
2.3 Engineering of primers and mutants using site-directed mutagenesis 19
2.4 Transformation of *Escherichia coli* T7 Express cells with plasmid DNA 21
2.5 Over-expression and purification of HIV-1 protease 23

2.6 Analysis of the purity of the proteases (wild-type, F99A and W42F/F99A) by tricine-SDS-PAGE 24
2.7 Protein concentration determination 25
2.8 Structural characterisation of wild-type and mutant proteases 25
2.8.1 Secondary structure assessment using far-UV circular dichroism 25
2.8.2 Tertiary structure assessment using fluorescence spectroscopy 26
2.8.3 Quaternary structure assessment using SE-HPLC coupled with DLS 27
2.8.4 ANS binding studies to native and unfolded wild-type versus the F99A and W42F/F99A proteases 28
2.9 Thermal unfolding of the wild-type, F99A and W42F/F99A proteases 28
2.10 Functionality of the proteases 29
Chapter 3 Results 31
3.1 Multiple sequence and secondary structural alignments 31
3.2 Sequence verification of DNA insert 31
3.3 Over-expression and purification of the wild-type, F99A and W42F/F99A 36
3.4 Secondary structure analysis of F99A, W42F/F99A and wild-type 36
3.5 Tertiary structure analysis of F99A, W42F/F99A and wild-type 39
3.6 Extrinsic ANS-binding fluorescence 42
3.7 Analysis of molecular weight and hydrodynamic volume of wild-type, F99A and W42F/F99A 42
3.8 Thermally-induced unfolding of wild-type, F99A and W42F/F99A proteases 45
3.9 Effects of F99A mutation on the catalytic activity of the protease 45
CHAPTER 4 DISCUSSION 53
4.1 Role of F99 on the structure of HIV-1 subtype C protease 56
4.1.1 Secondary structural characterisation 56

4.1.2 Tertiary structural characterisation 60
4.1.3 Role in quaternary structural characterisation 62
4.2 Importance of F99 on the thermal stability of HIV-1 subtype C protease 63
4.3 Effects of the F99A mutation on the catalytic activity of HIV-1 subtype C protease 64
4.4 Conclusions 66
CHAPTER 5: REFERENCES 68

LIST OF FIGURES

Figure 1. The global distribution of different HIV-1 subtypes 2
Figure 2. The HIV-1 life-cycle 4
Figure 3. A ribbon representation of the dimeric structure of HIV-1 subtype C protease 10
Figure 4. Schematic of the scissile peptide bond hydrolysis by the aspartic protease 12
Figure 5. Monomeric HIV-1 protease colour coded according to residues that contribute to the stability of the dimer interface 15
Figure 6. The lock-and-key motif formed at the dimer interface of HIV-1 protease 16
Figure 7. Primer sequences of the F99A and W42F mutations 20
Figure 8. Multiple sequence and structural alignment of different HIV-1 protease subtypes 32
Figure 9. Segments from sequenced HIV-1 subtype C wild-type and F99A insert DNA 33
Figure 10. Sections from sequenced HIV-1 subtype C wild-type and W42F/F99A insert DNA 34
Figure 11. Amino acid sequences of the wild-type, F99A and W42F/F99A cDNA 35
Figure 12. Molecular weight determination using tricine SDS-PAGE of purified proteases 37
Figure 13. Far-UV circular dichroism spectra of native HIV-1 subtype C proteases 38
Figure 14. Fluorescence spectra of HIV-1 subtype C proteases with excitation at 280 nm 40
Figure 15. Tryptophan emission spectra of wild-type, F99A and W42F/F99A proteases 41
Figure 16. Comparisons of ANS binding to wild-type, F99A and W42F/F99A 43
Figure 17. SE-HPLC of the wild-type, F99A and W42F/F99A 44

Figure 18. Thermal unfolding of the wild-type, F99A
and W42F/F99A 47
Figure 19. Specific activity of wild-type and F99A
HIV-1 subtype C proteases 48
Figure 20. Determination of the Michaelis-Menten constants and maximum
velocities of the wild-type and F99A proteases using 2 µM of enzyme 49
Figure 21. Michaelis-Menten constant and maximum velocity
determination of the wild-type and F99A using 0.24 µM enzyme 50
Figure 22. Turn-over determination of the wild-type
and F99A proteases 51
Figure 23 Global view of the F99 position and cavity
formed by F99A mutation in the HIV-1 protease 54
Figure 24. Simulation of the cavity created by substituting
phenylalanine with alanine. 55
Figure 25. Local environment around F99 and F99A 58
Figure 26. Illustration of the W6 and W42 neighbouring residues 61

LIST OF TABLES

Table 1: PCR-based mutagenesis assay preparation 22

Table 2: Catalytic profiles of the wild-type and F99A 52

ABBREVIATIONS

A_{280}	Absorbance at 280 nm
Å	Angström
AIDS	Acquired Immunodeficiency Syndrome
ANS	8-anilino-1-napthalene sulfonate
C-SA	South African subtype C
CD	Circular Dichroism
DLS	Dynamic Light Scattering
DEAE	Diethyl Amino Ethyl
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
DTT	Dithiothreitol
EDTA	Ethylenediaminetetra-acetic Acid
ε	molar extinction coefficient
FDA	Food and Drug Administration
FL	Fluorescence
GST	Glutathione *S*-Transferases
HIV	Human Immunodeficiency Virus
HT	voltage applied to the circular dichroism photomultiplier tube or High tension voltage
IPTG	Isopropyl –β- D-Thiogalactopyranoside
K_a	Dissociation constant
k_{cat}	Catalytic constant
k_{cat}/K_M	Catalytic efficiency

K_M	Michaelis constant
LB	Lysogeny Broth
Mw	Molecular weight
OD_{600nm}	Optical Density at a wavelength of 600 nm
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pI	isoelectric point
pK_a	-log K_a
PR	Protease
RNA	Ribonucleic Acid
RSV	Rous Sarcoma Virus
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SE-HPLC	Size Exclusion High Performance Liquid Chromatography
TEMED	*N*, *N*, *N*, *N*'-tetramethylenediamine
[Θ]	Mean Residue Ellipticity
WHO	World Health Organisation
UV	Ultraviolet
UNAIDS	United Nations AIDS Organisation

CHAPTER 1 INTRODUCTION

1.1 HIV/AIDS

Since its emergence in the 1980s, Human Immunodeficiency Virus has had a devastatingly massive and rapid infectious and fatality rate globally. HIV is a *Lentivirus* that has been shown to cause AIDS, a disease that causes about 1.9 million deaths annually (WHO, 2010; UNAIDS, 2009). About 42 million people in the world are HIV positive (Papathanasopoulos *et al.*, 2003; UNAIDS, 2009) and more than 70% of these infections occur in sub-Saharan Africa (UNAIDS, 2009). The HI virus is classified into different types, groups, subtypes, sub-subtypes and CRFs (Circulating Recombinant Forms) (Chakrabarti *et al.*, 1987; Peeters *et al.*, 1991; Osmanov *et al.*, 2002). The two types, HIV-1 and HIV-2 are believed to originate from the SIVsm from the sooty mangabey monkey (*Cerocebus atys*) and the common chimpanzee (*Pan troglodytes*), respectively (Chakrabarti *et al*., 1987; Peeters *et al*., 1991). Despite that they have emerged about the same time, HIV-1 is the most prevalent (Chakrabarti *et al*., 1987; Peeters *et al*., 1991). HIV-1 is further divided into group M (Major), O (outlier) and N (New or Non-M or Non-O), which are a consequence of cross-species transmission (Peeters *et al*., 1991).

The genetic variability/distinction within group M has been mainly attributed to the HIV envelope glycoproteins giving rise to nine subtypes, A, B, C, D, F, G, H, J and K. The subtype B, which is widely spread in North America and Western Europe (Figure 1), has been the most studied. The manufactured HIV inhibitors are targeted against this subtype. In the sub-Saharan African region, subtype C is the most prevalent (Papathanasopoulos *et al*., 2003; Walker *et al.*, 2005). Recently, studies have focused on subtype C, which accounts for 56% of infections worldwide (Walker *et al*., 2005). Amongst Southern African countries, South Africa is one of the most affected by this disease (Walker *et al*., 2005). A high prevalence of 35% in KwaZulu-Natal, makes HIV one of the leading epidemics in the sub-Saharan Africa (Walker *et al*., 2005).

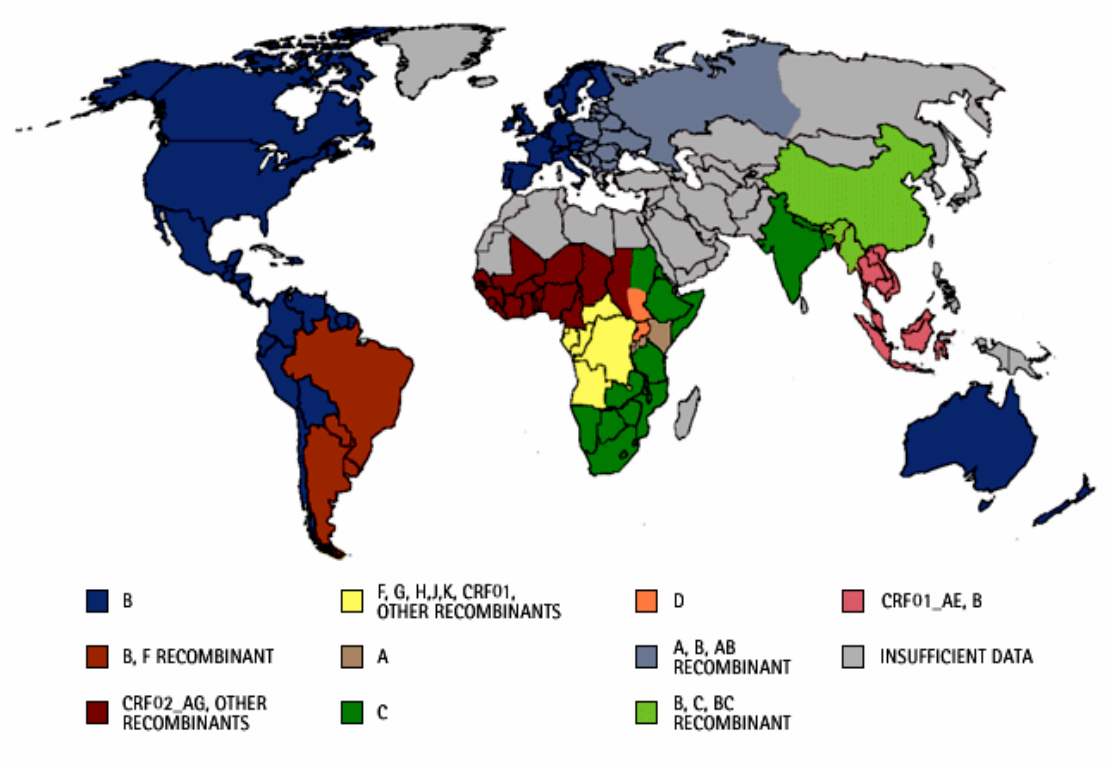


Figure 1. The global distribution of different HIV-1 subtypes
The HIV-1 subtypes vary in their prevalence across the globe and within the African continent. The subtype B is mostly spread in North America and Europe, while the subtype C is more prevalent in sub-Saharan Africa. Picture was taken from http://www.pbs.org/wgbh/pages/frontline/aids/atlas/clade.htm.

1.2 HIV life-cycle and drug therapy

The HI virus is approximately 100 to 120 nm in diameter and has a high replication rate, producing about 10^{10} virions daily (Greene, 1993; Coffin, 1995; Tavassoli, 2011). HIV infection (Figure 2) commences when the gp120 of the viral envelope binds to the host CD4 receptor of the T-lymphocytes cells (Freed, 2001; Plantier *et al.*, 2009). This is followed by the fusion step, which is mediated by the ternary complex comprising of the gp120, CD4 receptor and the co-receptor CXCR4. Once viral contents are released into the host cytosol, the single-stranded viral RNA is reverse-transcribed into double-stranded DNA by viral reverse transcriptase. This viral DNA enters the nucleus of the host-cell where it is ultimately integrated into the host genome by the viral enzyme, integrase. The recombinant DNA is then transcribed to mRNA which in turn is translated into viral Gag and Gag-Pol polyproteins (Freed, 2001; Mugnaini *et al.*, 2005; Plantier *et al.*, 2009). When the viral RNA and proteins have assembled, immature virions bud off. The last, but very crucial step (Kramer *et al.*, 1986; Lambert *et al.*, 1992), is maturation of the virus where the protease processes the Gag and Gag-Pol polyproteins producing a mature and infectious virus.

The viral enzymes (reverse transcriptase, integrase and protease) are key targets for the development of HIV therapeutic drugs. Highly-active antiretroviral therapy (HAART), which is thus far, a well-established therapy, relies on a combination of three or more of these drugs (Broder, 2010). The solving of the first crystal structures of the HIV-1 Protease (Navia *et al.*, 1989; Wlodawer *et al.*, 1989; Weber, 1990) has enabled development of successful protease inhibitors, which target the protease active site. The FDA-approved protease inhibitors include Amprenavir (APV), Atazanavir (ATV), Darunavir (TMC-114), Fosamprenavir (FOS-APV) Indinavir (IDV), Lopinavir (LPV), Nelfinavir (NFV), Ritonavir (RTV), Saquinavir (SQV) and Tipranivir (TPV) (Fitzgerald and Springer, 1991; Wlodawer and Vondrasek, 1998; Broder, 2010).

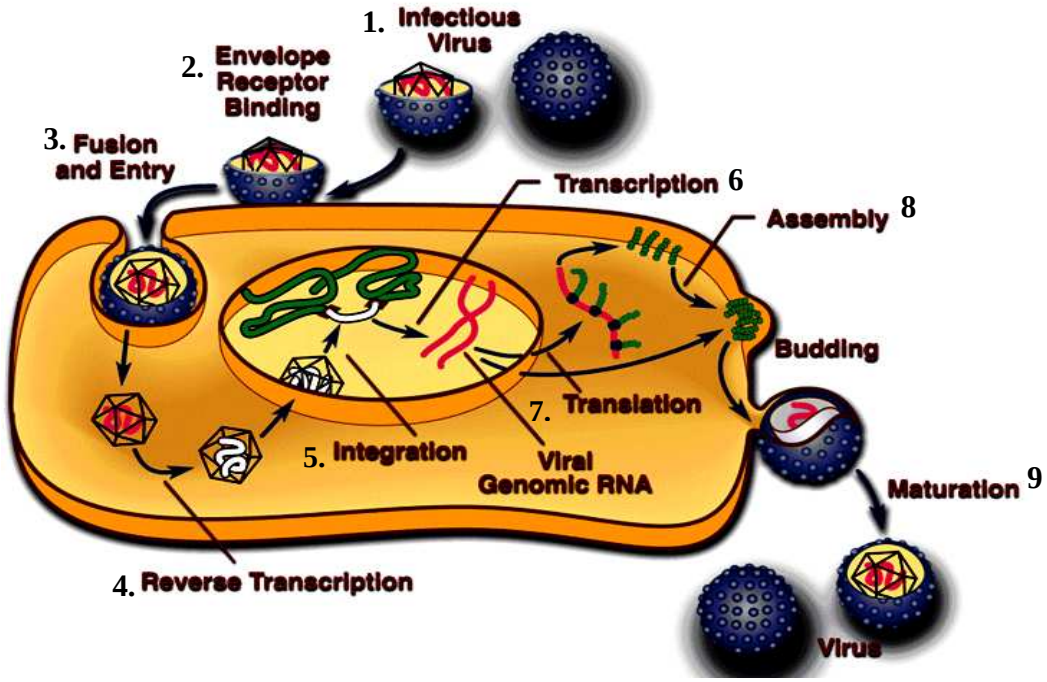


Figure 2. The HIV-1 life-cycle

The HI virus gains entry into the host cell by binding to the host CD4 cell receptors via the virus gp120 (1, 2 and 3). The viral reverse transcriptase and integrase reverse transcribe the viral RNA into double-stranded DNA and integrate it into the host DNA, respectively (4 and 5). DNA transcription (6), followed by translation (7), occur prior to assembly (8) and maturation (9) steps. The 8^{th} and 9^{th} steps are critical and are carried out by the HIV protease. This figure was adapted from http://home.ncifcrf.gov/hivdrp/RCAS/replication.html.

These inhibitors impede the catalytic function of the enzyme, by preventing maturation and propagation of HIV. The NRTIs (nucleoside reverse transcriptase inhibitors) and NNRTIs (non-nucleoside reverse transcriptase inhibitors) target the reverse transcriptase, but with reduced efficiency in chronically infected cells (Lambert *et al*., 1992; Balzarini *et al*., 2005). This is because chronically infected cells are characterised by viral particles, which contain integrated proviral genome (Lambert *et al*., 1992; Balzarini *et al*., 2005). Therefore the reverse transcriptase is not needed in chronically-infected cells for production of new viral particles, hence targeting the RT is not as significant as targeting the PR for anti-viral drug chemotherapy. Protease inhibitors are efficient in both acutely and chronically infected cells, hence the protease is a more preferred target for rational drug design over the reverse transcriptase which is only effective in acutely infected cells (Kramer *et al*., 1986; Lambert *et al*., 1992).

The emergence of HIV drug resistant strains and the genetic diversity of the virus have posed a major challenge in development of a durable anti-retroviral regimen and eradication of the virus from the host (Hoetelmans *et al*., 1997; Papathanasopoulos *et al*., 2003; Bannwarth *et al*., 2009). The lack of a proof-reading mechanism by the reverse transcriptase, drug-pressure and other factors promote mutations and selection of these drug resistant strains (Bebenek *et al*., 1989; Papathanasopoulos *et al*., 2003). One of the proposed attributes to resistance is the susceptibility of the protease to mutations. These mutations may alter catalysis by direct interaction with the inhibitor (flap mutations) or inducing conformational changes at the active site (hinge mutations) (Galiano *et al*., 2009). The altered conformation of the active site allows efficient substrate catalysis but still lowering the affinity of the protease to inhibitor (Galiano *et al*., 2009). Analysis of one of the first crystal structures of HIV-1 protease, led to the postulation that disruption/dissociation of the dimer interface may result in an inactive enzyme (Wlodawer *et al*., 1989). The dimer interface region; in particular the antiparallel β-sheet which is formed by the N- and C-termini has been identified as a potential drug-target. This owes to its highly conserved nature (Wlodawer *et al*., 1989; Broglia *et al*., 2008; Bannwarth *et al*., 2009). Recently, the major focus has been on the development of novel peptides and small molecules that target this region so as to block dimerisation of the protease (Koh *et al*., 2007; Yan *et al*., 2008; Bannwarth *et al*., 2009).

1.3 The common features and organisation of the dimer interface of proteins

Protein dimer interfaces include the surface area buried upon dimerisation and residues that form inter-subunit interactions. Oligomerisation is an important phenomenon adopted by nature to reduce the protein surface area that is exposed to solvent (Jones and Thornton, 1995; Jones and Thornton, 1996; Tsai *et al.*, 1997; Xu *et al.*, 1997). Understanding of this phenomenon originates from early studies of haemoglobin and cytochrome c′ just to mention a few (Perutz *et al.*, 1960; Finzel *et al.*, 1985). This enables the protein to adopt a more spherical/globular form (Pace *et al.*, 1996; Tsai *et al.*, 1997; Chothia and Janin, 1975). This is important in the cell as it has an influence on molecular crowding. Like the HIV-1 protease, some proteins elicit functionality only when they have oligomerised. Most of these proteins rely on dimer interface interactions for complete stability and optimum functionality (Lancaster *et al.*, 2004; Alves *et al.*, 2006; Walton *et al.*, 2009). For subunits to dimerise, they require specific recognition complementarity, which is dependent upon their physical shape and charge (Jones and Thornton, 1995).

1.3.1 The physicochemical properties of the dimer interface

It has been established that the recognition sites of proteins are predominantly hydrophobic, globular, and planar (Chothia and Janin, 1975; Tsai *et al.*, 1997; Jones and Thornton, 1995). They also fold into shape and carry electrostatic groups that influence the complementarity of binding (Chothia and Janin, 1975; Tsai *et al.*, 1997; Jones and Thornton, 1996). Generally, interactions at the dimer interface include salt-bridges, hydrogen bonds, hydrophobic and van der Waals interactions (Chothia and Janin, 1975; Dill, 1990; Xu *et al.*, 1997). The combination and magnitude of these interactions is unique for different proteins. The same principles that govern interactions for protein-protein interactions also apply to dimer interfaces (Janin and Chothia, 1990). A study reviewing the nature of protein-protein interactions (Janin and Chothia, 1990) showed that on average, the accessible surface area of a protein is rich with non-polar residues which populate about 55% to 75% of the total protein surface area (Janin and Chothia, 1990; Han *et al.*, 2006).

The polar and charged residues form 25% and 20% of the accessible surface area, respectively (Janin and Chothia, 1990). Salt-bridges and hydrogen bonds can significantly enhance stability across the dimer interface but are mainly implicated in binding specificity (Janin and Chothia, 1990; Xu *et al*., 1997).

Although hydrophobic interactions are mainly thought to drive dimerisation and also to stabilise the protein, they have been implicated in binding recognition (Keskin *et al.*, 2005; Dill, 1990; Janin and Chothia, 1990). Upon interaction of subunits, packing of the respective side chains culminates in optimisation of the van der Waals interactions. This is very important in determining the surface shape complementarity of subunits (Janin and Chothia, 1990; Dey *et al.*, 2010).

The distribution of α-helices on protein surfaces and interfaces is constant. In contrast, the residues making up the β-strands are more likely to be found at interfaces versus the protein surfaces (Tsai *et al*., 1997). On the other hand, Jones and Thornton (1995) concluded that although all the structural elements are found at dimer interfaces, the helices predominate. The dimer interface residues when compared to residues in other regions of the protein make fewer intra-molecular contacts, making the dimer interface more flexible (Jones and Thornton, 1995).

In contrast to the rest of the protein, the dimer interface exhibits a unique amino acid composition (Keskin *et al*., 2005). The aromatic groups are particularly favoured at the dimer interface because they provide good 'glue' for binding of subunits (Jones and Thornton, 1995). These residues tend to form, what is referred to as, a lock-and-key motif where the aromatic side chain protrudes from one subunit and inserts/buries into the hydrophobic pocket of the other subunit (Jones and Thornton, 1995). This phenomenon has been well-documented for the glutathione *S*-transferases (GSTs) superfamily of proteins (Sayed *et al.*, 2000; Alves *et al*., 2006; Parbhoo *et al*., 2011).

1.3.2 Contribution of 'hot-spots' and conserved residues at the dimer interface

Previously, Xu *et al.* (1997) suggested that the polar residues and hydrogen bonding groups are uniformly distributed throughout the dimer interface, implying equal contribution of the amino acid residues to the energy of stabilisation (Xu *et al*., 1997).

Despite that, a general amino acid pattern has not been established at dimer interfaces, hence the binding energy and stability contributed by amino acids is not uniform. That is, there maybe a single residue or a cluster of residues contributing more to the binding energy and stability of a protein. The amino acids that confer the most to binding energy and stability of the dimer interface are referred to as 'hot-spots' (Burgoyne and Jackson, 2006). These may occur as clusters that are surrounded by residues that contribute far less to the stability (Bogan and Thorn, 1998). Thus, residues that possess this feature make their identity, size and charge very important. Therefore, it only makes evolutionary sense that these 'hot-spots' and their neighbouring residues are conserved (Keskin *et al.*, 2005). The tendency of conserved hot-spot residues to cluster, leads to the presumption that they could be useful in predicting their interacting partners (Caffrey *et al.*, 2004; Burgoyne and Jackson, 2006). But, it is also known that dimer interface residues are not any more or less conserved than other regions of the protein (Caffrey *et al.*, 2004; Burgoyne and Jackson, 2006).

1.4 The conformation of β-sheets and role at the dimer interface

The amino acid side chains of the potential pairing β-strands are thought to play a role in recognition, proper folding conformation and hence stability of the resulting β-sheet (Mandel-Gutfreund *et al.*, 2001). The specificity with which β-strands recognise their pairing partners is still not clearly understood. This is because conservation of complementary strands is not drastically distinct from other non-binding β-strands. Also, the identity of pairing side-chains plays a minute role in binding recognition (Mandel-Gutfreund *et al.*, 2001). However, it is understood that the identity of side-chains may play a more important role when protein-protein binding is dependent on pairing of β-strands (Mandel-Gutfreund *et al.*, 2001). Non-polar residues, similar to the protein core, are also more conserved in β-sheets (Mandel-Gutfreund *et al.*, 2001). More specifically, the aromatic and β-branched amino acid pairs are more conserved, hence, they are more constrained to maintain hydrophobic interactions (Merkel *et al.*, 1999; Mandel-Gutfreund *et al.*, 2001).

These, in turn, confer the most stability in β-sheets than other residues (Mandel-Gutfreund *et al.*, 2001). In particular, phenylalanine has high propensity for β-sheets and, in most part, is highly conserved when present in β-sheets. This aromatic residue has been implicated in the so called glycine 'rescue'. This phenomenon involves the phenyl ring shielding the glycine amide and carbonyl groups from forming hydrogen bonds with the competing polar solvent (Mandel-Gutfreund *et al.*, 2001; Merkel *et al.*, 1999). This, in turn, counteracts the destabilising effects of glycine (Merkel *et al.*, 1999).

1.5 The structure and catalytic mechanism of HIV-1 protease

The HIV-1 protease is described as an obligate homodimer because its functionality is dependent on the dimerisation of its symmetrical monomers (Hansen *et al.*, 1988). These monomers, in turn, contribute two symmetrical aspartate residues at the single active site (Wlodawer and Gustchina, 2000). It, thus, belongs to the aspartic protease family; hence, it is also inhibited by pepstatin (Hansen *et al.*, 1988; Wlodawer and Gustchina, 2000). The active site structural organisation is unique for retroviral proteases, and makes it different from the non-viral proteases. Each monomer of the HIV-1 protease is composed of 99 amino acids forming nine β-strands and one α-helix (Figure 3). The prominent regions of this enzyme are the dimer interface, the flaps and hinge regions. The dimer interface is comprised of four regions: residues 1 – 4 and 94 – 99, which form a four-stranded N- and C-termini antiparallel β-sheet; residues 24 – 29, with residues 25 – 27 (Asp25 - Thr26 - Gly27) forming the catalytic triad; residues 87, 90, 91, 92 and 93 which make up the α-helices and residues 48 – 54 forming the flap tips (Weber, 1990; Coman *et al.*, 2008b).

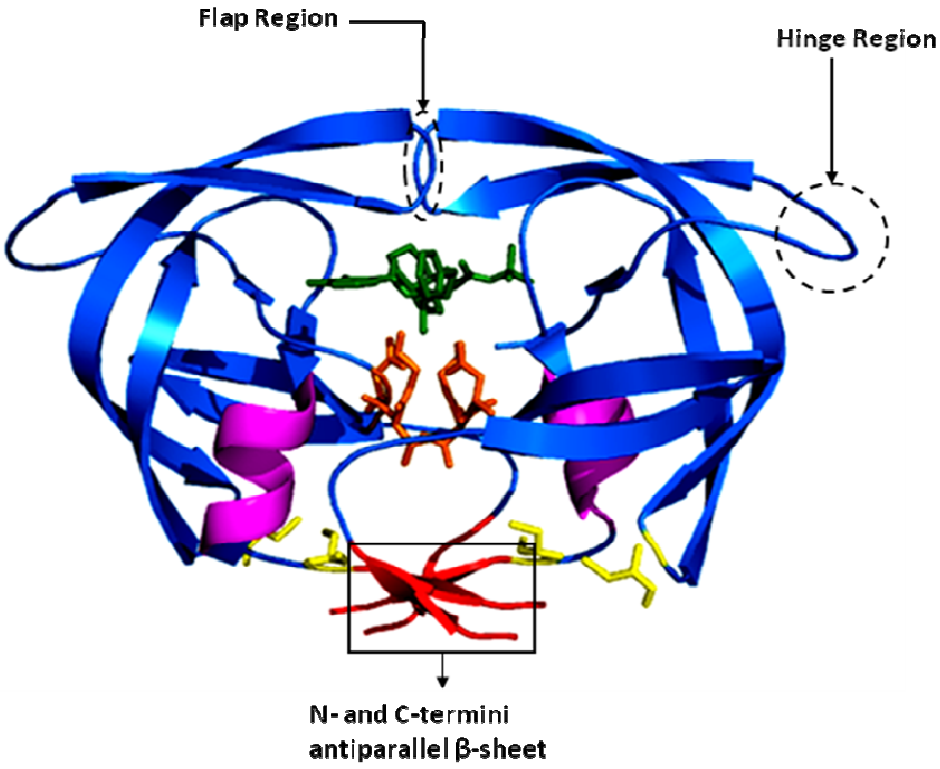


Figure 3. A ribbon representation of the dimeric structure of HIV-1 subtype C protease

The α-helices of each monomer are in purple with the N- and C-termini antiparallel β-sheet shown in red. The rest of the β-strands are represented in blue. The flap and hinge regions are also indicated. The catalytic triad residues (D25 – T26 – G27), from each monomer are coloured in orange. The C67 and C95 are yellow. The protease is complexed with Nelfinavir which is coloured green. Figure generated using PyMOL v0.99 (DeLano Scientific, 2006). PDB code 2R5Q (Coman *et al.*, 2008b).

β-strand two to eight form a jelly-roll β-barrel topology within each subunit (Noel *et al.*, 2009), whereas each α-helix anchors the N- and C-termini antiparallel β-sheet to the rest of the protease moiety (Louis *et al.*, 2003). The active site, as previously mentioned, is positioned at the dimer interface and above it (Figure 3) the flaps are arranged such that they seem to enclose it from the environment. Consequently, for the substrate (inhibitor) to access the active site region, the flaps (residues 43 to 58) need to open up, and also to possibly exclude water molecules (Liu *et al.*, 2008; Kozisek *et al.*, 2008). The flaps are highly flexible and their flexibility is thought to be controlled by the hinges (Kozisek *et al.*, 2008; Mosebi *et al.*, 2008). Although the protease is symmetrical, its natural substrates are asymmetrical (Fitzgerald and Springer, 1991; Prabu-Jeyabalan *et al.*, 2000). Binding of the substrate induces conformational changes which cause the monomers to assume a more asymmetrical arrangement (Prabu-Jeyabalan *et al.*, 2000). Once the substrate has gained access to the active site, the D25/D25′ side-chain activates the conserved active site water molecule (Figure 4) which launches a nucleophilic attack on the carbonyl group of the polyprotein scissile bond. This is followed by protonation of the leaving amine group by the D25/D25′ residue resulting in hydrolysis of the polyprotein (Hansen *et al.*, 1988).

Protease binding subsites are designated S1, S2, S3 and S1′, S2′, S3′ starting from the D25 residue. The S1 and S1′ subsites (similarly S2 and S2′ etc.) are contributed by each monomer at symmetric positions. The respective side-chains of the substrate are termed P1, P2, P3 and P1′, P2′, P3′ going toward the N-terminus and C-terminus, respectively (Wlodawer and Vondrasek, 1998; Prabu-Jeyabalan *et al.*, 2000). The protease binding pockets comprise residues that form direct and indirect interactions with the substrate. These residues originate from different regions of the protease and either make direct and/or water-mediated hydrogen bonding, van der Waals and/or hydrophobic interactions with the substrate (Prabu-Jeyabalan *et al.*, 2000). Therefore, substitution of these residues may directly alter the catalytic activity of the HIV-1 protease by altering the binding affinity of substrate or inhibitor.

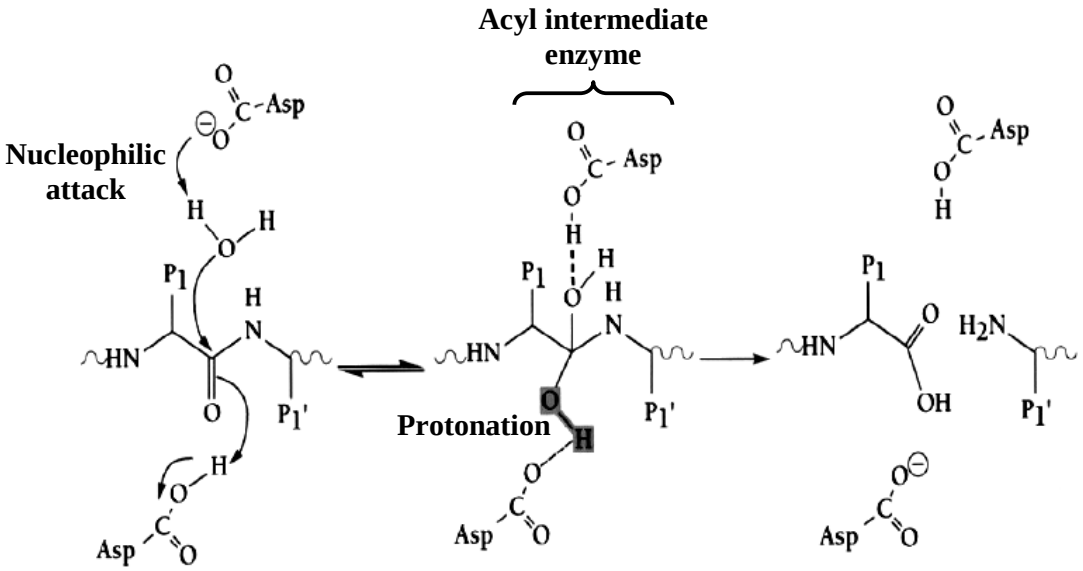


Figure 4. Schematic of scissile peptide bond hydrolysis by the aspartic protease
One of the catalytic aspartate residues activates the conserved active site water molecule. The activated water molecule launches a nucleophilic attack on the carbonyl scissile peptide bond. The second aspartic acid residue then protonates the leaving amine group forming the acyl intermediate enzyme. The hydrolysis of the acyl enzyme is carried out by a water molecule thus completing the process of catalysis. The substrate side-chains are assigned by either P1 or P1′. Picture was adapted from http://www.jiaowu.buct.edu.cn/Courseware/Harvard/BCMP201/pdf/.

The flap region, with respect to its structural conformation, harbours significant differences amongst the retroviral proteases especially between the HIV-1 subtype A, B, C and F (Wlodawer and Gustchina, 2000). This region elicits drug resistance via residues at positions 46, 47, 48, 50, 53 and 54 (Liu *et al*., 2008; Galiano *et al*., 2009).

HIV-1 subtypes harbour differences at the genetic and protein level and are called naturally occurring polymorphisms (NOPs) (Coman *et al.*, 2008a). The NOPs that occur in the dimer interface are N88D, L89M and I93L, with L89M occurring in more than 95% of the subtype C strains. No major drug resistant mutations have been reported for the antiparallel β-sheet termini except for the conservative substitution, L97V. In other retroviral proteases, L97 is replaced by an isoleucine (Weber, 1990). This emphasises the importance of maintaining hydrophobic packing in the C-terminus (Louis *et al*., 2003; Bowman and Chmielewski, 2002).

1.6 The N- and C-termini antiparallel β-sheet of HIV-1 protease: influence on structure, function and stability

The N- and C-termini antiparallel β-sheet, often referred to as the dimer interface, is formed by four β-strands in retroviruses and by six β-strands in non-viral proteases (Wlodawer and Gustchina, 2000). Previous studies showed that the overall topology between rous sarcoma virus and HIV-1 proteases are similar (Wlodawer *et al*., 1989; Weber, 1990). Proteases from both viruses show significant intersubunit interactions in the “fireman’s grip” at the active site, between residues 8, 29 and 87 for HIV-1 and the N- and C-termini antiparallel β-sheet. The structural comparison between the two proteases also depicts the N- and C-termini β-strands arranged in a similar fashion (Wlodawer *et al*., 1989). The evolutionary conservation in the structural arrangement of this region suggests that it is crucial for maintaining the stability and function of the protease (Lesk and Chothia, 1980; Wlodawer *et al*., 1989; Mirny and Shakhnovich, 1999). Upon dimerisation, the HIV-1 protease subunits form a network of 34 hydrogen bonds and four ionic interactions between them. Two of these hydrogen bonds occur at the flaps, five at the active site residues and eight occurring between residues 6 – 8, 29 and 87. A substantial 19 hydrogen bonds and two ion pairs are formed at the N- and C-termini antiparallel β-sheet (Weber, 1990).

The C-terminal (residues 96 to 99) contributes 50% of subunit-subunit ionic and 56% hydrogen bond interactions and buries 45% of the surface area following dimerisation (Weber, 1990). Therefore, it is obvious that structure-based thermodynamic analysis showed that the N- and C-termini antiparallel β-sheet contributes about 75% to the total Gibbs energy (Todd *et al.*, 1998). This region is, thus, a 'hot-spot'. It was suggested that most of the energy of stabilisation at the N- and C-termini antiparallel β-sheet is mainly accounted for by interactions formed by residues F99, L97, N98, T96, C95, P1, I3 and L5 (Figure 5) (Todd *et al*., 1998). Notably, most of these residues are hydrophobic. The F99 residue contributes the most to the energy of stabilisation of the protease (Todd *et al*., 1998). This residue is oriented toward the interior of the dimer interface and is located in the presumed hydrophobic cleft which includes W6 (Hostomsky *et al.*, 1989). The F99 locks into the hydrophobic pocket seemingly, with its phenyl ring facing and locking with P1 of the other subunit (Figure 6). P1 forms one inter-monomeric hydrogen bond with F99 (Coman *et al*., 2008b).

The F99 is conserved in HIV-1 and SIV (Simian Immunodeficiency Virus) proteases and a leucine is located at the position in HIV-2 and EIAV (equine infectious anaemia virus) proteases (Wlodawer and Gustchina, 2000). This implies that, not only do hydrophobic interactions drive subunit-subunit interactions but packing of these hydrophobic residues contributes significantly to the stability of the dimer (Louis *et al*., 2003; Bowman and Chmielewski, 2002). The N- and C-termini regions do not contribute directly to substrate binding nor does it partake directly in catalysis. The catalytic function of HIV-1 protease has been shown to be rather more sensitive to substitutions at the C-terminal residues than the N-terminal residues (Masso and Vaisman, 2003). But, both the N- and C-termini have been shown to be crucial for dimer formation and stability, especially the C-terminal (Louis *et al*., 2003; Ishima *et al.*, 2003). Since HIV-1 protease is only active as a homodimer, the N- and C-termini therefore play an indirect but essential role in the function of the protease (Oroszlan and Luftig, 1990; Louis *et al.*, 2003; Ishima *et al.*, 2003). Previous studies have reported that the F99A mutation causes complete inhibition of the subtype B HIV-1 protease (Pettit *et al.*, 2003).

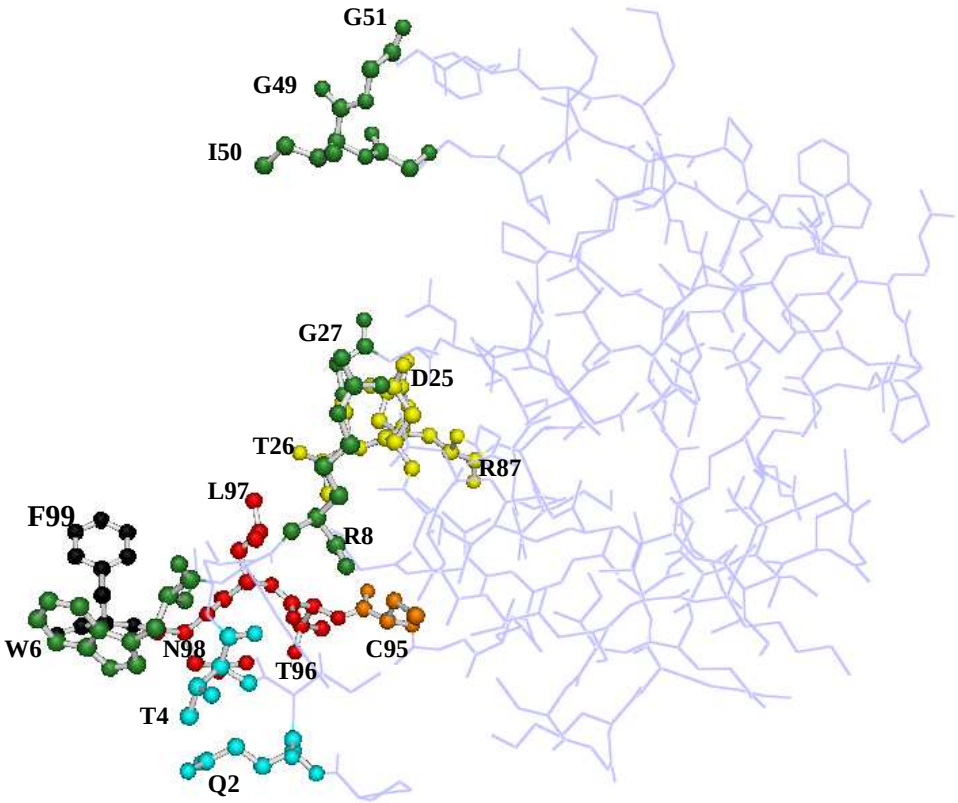


Figure 5. Monomeric HIV-1 protease colour coded according to residues that contribute to the stability of the dimer interface

The F99 (black) contributes the most to Gibbs energy of stabilisation, followed by residues coloured red, orange, yellow, green and cyan, respectively. Those in blue do not contribute to the dimer interface. The picture was adapted from a study by Todd *et al*., 1998 and generated using PyMOL v0.99 (DeLano Scientific, 2006). PDB code 2R5Q (Coman *et al*., 2008b).

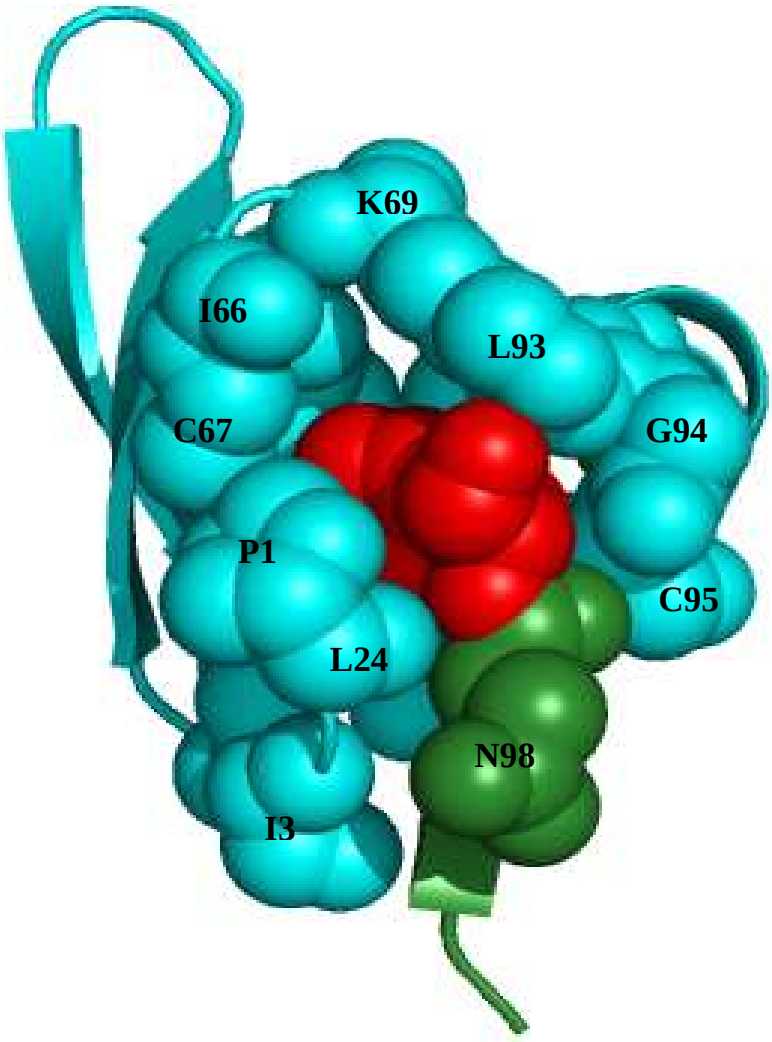


Figure 6. The lock-and-key motif formed at the dimer interface of HIV-1 protease

The 'key' residue F99 (red) is inserted into the hydrophobic pocket ('lock') which is formed by the respective residues (cyan) from the partnering subunit. These residues are within 4 Å of F99. The phenyl ring of F99 and proline (Pro1) side-chain lock. The N98 (green) is from the same subunit as F99. The prime indicates residues from the partnering subunit. Figure generated using PyMOL v0.99 (DeLano Scientific, 2006). PDB code 2R5Q (Coman *et al.*, 2008b).

Urea and thermal equilibrium unfolding/refolding studies done on subtype B HIV-1 protease showed that the unfolding transitions followed a two-state mechanism given by the equation below (Todd *et al*., 1998; Noel *et al*., 2009). In contrast, the presence of a fully folded monomeric intermediate was detected by kinetic folding studies (Noel *et al*., 2009). The transient nature of the intermediate is due to the large hydrophobicity of the dimer interface, which makes it unstable when exposed to the solvent. Hence it is reported that the dimerisation step contributes the most Gibbs energy of stabilisation (Todd *et al*., 1998; Noel *et al*., 2009). The stability of the protease is both pH- and concentration-dependent, the latter being due to its dimeric nature (Todd *et al*., 1998; Noel *et al*., 2009). The protease is more stable at higher pH values, and at pH 5 the dissociation constant is in the micromolar range (Todd *et al*., 1998; Noel *et al*., 2009). There is a knowledge gap on the unfolding mechanism for the South African subtype C HIV-1 protease. So far, the mechanism by which the protease dimerises within the Gag-Pol precursor *in vivo*, and the exact sequence of events that follow with respect to its cleavage and activation has not been elucidated. But it has been suggested that *in vivo* dimerisation and activation of the protease occurs whilst embedded within the Gag-Pol (Pettit *et al.*, 2003).

$$N_2 \Leftrightarrow 2N \qquad \textbf{Equation 1}$$

Therefore, introducing mutations at highly conserved hydrophobic regions may result in destabilisation of the dimer and, hence, induce conformational changes at the active site, which may alter the function of HIV-1 protease homodimer (Broglia *et al*., 2008; Louis *et al*., 2003). The disruption of the N- and C-termini antiparallel β-sheet has also been a subject of rational drug design in other proteins, such as the amyloid fibrils responsible for Alzheimers disease and the adhesins from the bacterial pili responsible for uropathogenic infections (Li *et al*., 2006; Pinkner *et al*., 2006).

1. 7 Objective and aims

The implication of regions near the active site in the rise of drug resistant mutations has led to development of novel ideas for inhibiting the HI virus. Highly conserved hydrophobic residues play an important role in the structure, function and stability of proteins (Lesk and Chothia, 1980a; Mirny and Shakhnovich, 1999). Previous studies have indicated that the N- and C-termini antiparallel β-sheet which forms part of the dimer interface is indispensable for dimerisation and stability of HIV-1 protease (Ishima *et al.*, 2003; Ishima *et al*., 2007). Thus, the objective of this study is to detect residues at the dimer interface of HIV-1 subtype C protease that are required for HIV-1 subtype C protease homodimeric structure, function and stability.

The aims of this study are to:

- Use sequence and secondary structural alignment bioinformatics tool to detect highly conserved residues at the dimer interface of South African HIV-1 subtype C protease.
- Employ PCR-based site-directed mutagenesis to generate the F99A mutation.
- Compare the secondary, tertiary and quaternary structure of the mutant to the wild-type using far-UV circular dichroism, intrinsic and extrinsic fluorescence spectroscopy and SE-HPLC.
- Investigate thermal stability by using far-UV circular dichroism.
- Compare the activity of F99A and wild-type proteases by following the hydrolysis of a chromogenic substrate spectrophotometrically.

CHAPTER 2 EXPERIMENTAL PROCEDURES

2.1 Materials

The QuickChange® Lightning Site-Directed Mutagenesis kit was purchased from Stratagene (La Jolla, CA, USA). The synthesis of oligonucleotide primers and DNA sequencing was done by Inqaba Biotec (Pretoria, SA). Chromogenic substrate was a kind gift from Dr T. Govender from University of KwaZulu-Natal. The chromogenic substrate is a peptide with the sequence Lys-Ala-Arg-Val-Nle-*p*-nitro-Phe-Glu-Ala-Nle-NH_2. All the other chemicals were of analytical grade.

2.2 The HIV-1 expression vector

The pET-HIVPR plasmid encoding wild-type HIV-1 subtype B protease was a kind gift from Dr J. Tang, University of Oklahoma Health Science Centre, Oklahoma City (Ido *et al.*, 1991). The South African subtype C HIV-1 protease, when compared to the B subtype protease, encodes the following polymorphisms; T12S, I15V, L19I, M36I, R41K, H69K, L89M and I93L. These polymorphisms, including the Q7K that minimises the autoproteolysis of the PR, were introduced into the pET-HIVPR expression vector by Dr. S. Mosebi (University of the Witwatersrand, Johannesburg). The expression vector will be referred to as pET-WTCSA encoding, the 'wild-type' HIV-1 subtype C SA protease.

2.3 Engineering of primers and mutants using site-directed mutagenesis

To construct the F99A and W42F/F99A mutants, specific oligonucleotide primers were designed based on the pET-WTCSA sequence (Dr Y. Sayed). The primers were designed based on the pET-WTCSA sequence using the Primer-X software (http://bioinformatics.org/primex). The primer sequences (Figure 7) were analysed using the Gene Runner software programme v3.01 (Hastings Software Inc., NY.USA) to check for possible formation of secondary structures.

F99A Fwd : 5' CAGCTGGGTTGCACTTTGAAC**GCG**TAGTAATGTCCGATTGAAAT 3'

F99A Rev : 5' CATTTCAATCGGAGATTACTAC**CGC**GTTCAAGTGCAACCAGCTG 3'

W42F Fwd: 5' GAAATCAATCTGCCTGGTAAG**TTC**AAGCCTAAAATGATCGGTGGC 3'

W42F Rev: 5' GCCACCGATCATTTTAGGCTT**GAA**CTTACCAGGCAGATTGATTTC 3'

Figure 7. Primer sequences of the F99A and W42F mutations
Primer sequences for the construction of the F99A and W42F mutations are indicated. The codons for alanine and phenylalanine are in bold.

To generate both the F99A and W42F/F99A mutants, the QuickChange® Lightning Site-Directed Mutagenesis kit from Stratagene (Braman *et al.*, 1996) was used. Briefly, the reactions were carried out at 95 °C to denature the double-stranded DNA for 30 seconds, followed by primer annealing at 55 °C for 60 seconds and then DNA elongation at 68 °C for 5 minutes. The PCR products were then digested with *Dpn*I (10 U/µg DNA) restriction enzyme at 37 °C for an hour and then at 20 °C for another hour. The *Dpn*I digests methylated DNA template (Nelson and McClelland, 1992; Chung *et al.*, 1989). The digested products were then used to transform XL 10-Gold Ultra Competent *E. coli* cells (Stratagene). The transformation was carried out by the one-step method (Chung *et al*., 1989). Cells were then plated on Lysogeny Broth (LB)-agar plates (0.5% (w/v) yeast, 1% (w/v) tryptone, 1% (w/v) NaCl), 1.5% (w/v) agar) supplemented with 100 µg/ml of ampicillin and incubated for 16 hours at 37 °C. The plasmid DNA was extracted from cells using the Strataprep miniprep kit by following the manufacturer's instructions (Stratagene, La Jolla, USA). Quantification and analysis of the plasmid DNA was performed by using the nanodrop spectrophotometer (Thermo Scientific). The plasmids encoding the F99A and W42F/F99A are referred to as pET-F99A-CSA and pET-W42F/F99A-CSA, respectively. See Table 1 for a summary of the PCR assays.

2.4 Transformation of *Escherichia coli* T7 Express cells with plasmid DNA

E. coli T7 Express I^q competent cells (New England Biolabs Inc., Ipswich, MA, USA) were transformed with either pET-WTCSA (10 ng), pET-F99A-CSA (10 ngl) or pET-W42F/F99A-CSA (10 ng) plasmid. The cells were then plated on Lysogeny Broth (LB) agar plates supplemented with 100 µg/ml ampicillin and 35 µg/ml chloramphenicol to specifically select for transformed cells. The plates were incubated for 16 hours at 37 °C. Cell colonies which were successfully transformed were grown in LB medium (0.5% (w/v) yeast, 1% (w/v) tryptone, 1% (w/v) NaCl) supplemented with 100 µg/ml ampicillin and 35 µg/ml chloramphenicol. These cells were grown at 37 °C, shaking at 250 rpm for 20 hours.

Table 1: PCR-based mutagenesis assay preparation
Mutagenesis was performed as described in section 2.3 using the reagents below:

Reagent	Volume (µl)	Final concentration (ng/µl)
10 × Reaction buffer	5	-
Forward primer	1	125
Reverse primer	1	125
Double-stranded DNA template	4.8	15
QuickChange® dNTP mix	1.5	-
milliQ water	36.7	-
QuickChange® Lightning enzyme	1	-

2.5 Over-expression and purification of HIV-1 protease

HIV-1 PR was over-expressed as inclusion bodies (Ido *et al.*, 1991) in *Escherichia coli* T7 Express cells. Cells were grown overnight in LB containing 100 μg/ml ampicillin and 35 μg/ml chloramphenicol at 37 °C. The culture was then diluted 100-fold into fresh LB media and grown until an optical density ($OD_{600\ nm}$) of 0.4 – 0.5. IPTG was then added to a final concentration of 0.4 mM to induce expression for 4 hours at 37 °C. Cells were harvested by centrifugation at 5000 × *g*, and the pellet was resuspended in buffer A (10 mM Tris, 2 mM EDTA and 1 mM PMSF, pH 8) and then stored at -20 °C overnight. The thawed cells were sonicated following the addition of $MgCl_2$ and DNase I to final concentrations of 10 mM and 10 U/μl respectively, and then centrifuged at 15 000 × *g* for 30 minutes. The resulting pellet was resuspended in ice-cold buffer A containing 1% (v/v) Triton X-100 and again centrifuged at 15 000 × *g* for 30 minutes. The inclusion bodies containing the recombinant HIV-1 protease and cell debris were collected by centrifuging at 15 000 × *g* for 30 minutes at 4 °C. Finally, the pellet was resuspended in buffer B (10 mM Tris, 8 M urea and 2 mM DTT, pH 8) to solubilise the pellet. The solubilised pellet was passed through a column packed with diethylaminoethyl (DEAE) anion sepharose matrix, pre-equilibrated with buffer B. The Bradford assay reagent (Bradford, 1976) was used to identify protein-containing fractions, which were then pooled and acidified with formic acid to a final concentration of 25 mM. Acidification caused precipitation of non-aspartic protease contaminants (Ido *et al*., 1991). The protease was refolded by dialysis against 10 mM formic acid at 4 °C. The refolded protease was dialysed against protease storage buffer C (10 mM sodium acetate, 1 mM NaCl and 2 mM DTT, pH 3.5).

A further purification step was required to remove contaminating proteins. The carboxymethyl (CM) cation exchange column was equilibrated using buffer C. The refolded protease which was dialysed against buffer C was loaded onto the CM column which was washed using the same buffer to remove proteins which are bound non-specifically. The protease, bound to the column, was eluted using a salt gradient from 0 to 1 M NaCl. The pooled fractions were desalted and dialysed against buffer C.

2.6 Analysis of the purity of the proteases (wild-type, F99A and W42F/F99A) by tricine-SDS-PAGE

Tricine-SDS-PAGE (Schagger and von Jagow, 1987; Schagger, 2006) was used to confirm the size and purity of the wild-type and mutants. The principle of this method is similar to glycine-SDS-PAGE (Laemmli, 1970), but in this case tricine is the trailing ion. This system is ideal for resolving proteins below 20 kDa because of its ability to separate these proteins from bulk SDS before reaching higher acrylamide concentrations (Schagger and von Jagow, 1987; Schagger, 2006). This is due to the higher pK_a of tricine. This technique relies on the dodecyl group of the SDS to bind to hydrophobic regions of the protein while the sulfonate part gives the protein an overall negative charge. The respective protein denatures and adopts a rod-like shape with an overall negative charge. Therefore, the compactness and hence the abundance of hydrophobic amino acid influences SDS binding and migration of the SDS-protein complex on the acrylamide gel.

Stacking and separating gels were prepared to a final concentration of 4% (w/v) and 16% (w/v) acrylamide/bisacrylamide, respectively, in gel buffer (3 M Tris and 0.3% (w/v) SDS, pH 8.45). The separating gel also contained 0.1% glycerol. The protein samples were diluted two-fold in sample buffer (12% (w/v) SDS, 6% (v/v) β-mercaptoethanol, 30% (w/v) glycerol, 0.05% (w/v) Coomasie Blue and 150 mM Tris, pH 7) and then incubated at 37 °C for 30 minutes. The cathode (1 M Tris, 1 M Tricine and 1% (w/v) SDS; pH 8.25) and anode (1 M Tris; pH 8.9) buffer were added to the upper and bottom compartments, respectively, before the protein samples were loaded. Initially, 30 V was applied until the samples reached the separating gel, followed by 130 V for separation. Standard molecular weight markers were used to assess the migration and purity of the wild-type and mutant proteases.

2.7 Protein concentration determination

To determine the concentrations of the proteins, the Beer-Lambert law was applied:

$$A = \varepsilon c l \qquad \textbf{Equation 2}$$

where A is the absorbance, ε ($M^{-1}.cm^{-1}$) is the molar extinction coefficient, c (M) is the concentration of the sample and *l* (cm) is the path length of light taken to travel through the cuvette.

Since the aromatic residues absorb maximally at 280 nm (Lakowicz, 1983), absorbance readings at 280 nm were recorded using a Jasco V-630 UV-Vis spectrophotometer. The readings were recorded for 10 serial dilutions of the respective protein. The molar absorption coefficient ($\varepsilon_{280\ nm}$) of 11800 $M^{-1}.cm^{-1}$ was used for the wild-type and F99A and 6250 $M^{-1}.cm^{-1}$ for the W42F/F99A (Perkins, 1986).

2.8 Structural characterisation of wild-type and mutant proteases

It was important to examine the secondary, tertiary and quaternary structural modifications, if any, induced by the F99A and/or W42F/F99A.

2.8.1 Secondary structure assessment using far-UV circular dichroism

The inherent ability of proteins to differentially absorb circularly polarised light aids in the characterisation of their secondary structure (Woody, 1995). Optical activity is observed in amino acids with aromatic and sulfide groups. The amino acid backbone also exhibits optical activity, and it is the different organisation of the backbone that gives distinct CD spectra for various proteins (Woody, 1995). Absorption range for far-UV is between 190 nm and 250 nm. The secondary structural modifications induced by the F99A and W42F/F99A were investigated exploiting this technique.

The spectral measurements were obtained using a Jasco J-810 spectropolarimeter running Spectra Manager with the v1.5.00 software over the wavelength range of 190 nm to 250 nm. The data were collected at 20 °C using a data pitch of 0.5 nm, bandwidth of 0.1 nm and scan speed of 200 nm.min^{-1}. The spectra were an average of 10 accumulations. All the proteins were prepared to a final concentration of 15 μM in a buffer containing 10 mM sodium acetate, 2 mM DTT at pH 5. To avoid discrepancies due to protein concentration, the mean residue ellipticity [Θ] (deg.cm^2.dmol^{-1}.residue^{-1}) was calculated using the equation

$$[\Theta] = \frac{100 \times \theta}{cnl} \qquad \textbf{Equation 3}$$

where c is the concentration of the protein in mM; *l* is the path length in cm, n is the number of residues in the protein chain and θ is the ellipticity (mdeg).

2.8.2 Tertiary structure assessment using fluorescence spectroscopy

The underlying principles of fluorescence spectroscopy is that a molecule absorbs light at a certain wavelength and emits it at a higher wavelength where the energy lost between the absorbed and emitted light is referred to as Stokes shift (Lakowicz, 1983) This loss in energy is due to interactions between the fluorophore and its immediate environment (Lakowicz, 1983). In proteins, aromatic rings of residues such as phenylalanine, tyrosine and tryptophan are used as fluorescent probes, but tryptophan is the dominant fluorophore. These probes can absorb light at 280 nm; however, the indole ring of tryptophan can be exclusively excited at 295 nm. The indole ring of the tryptophan makes it very sensitive to the polarity of its environment and was, therefore, used as an indicator of any local tertiary structural changes of the mutants with respect to the wild-type.

The monomeric HIV-1 protease has two tryptophan residues, W6 and W42. The W6 is positioned near the dimer interface N-terminal β-strand and W42 is positioned at the hinge region of the flaps fully exposed to the solvent, and is within 4 Å of Y59. This residue was also mutated to create the W42F/F99A mutant. This was done so that changes that might have only occurred at the N- and C-termini antiparallel β-

sheet accompanying the F99A mutation can be monitored via W6. The W42F/F99A mutant has a single tryptophan per monomer, that is, W6 to monitor changes occurring only at the dimer interface near the N- and C-termini antiparallel β-sheet.

Fluorescence measurements were performed using Perkin Elmer LS 50 B Luminescence fluorimeter. The spectra were recorded as an average of 5 accumulations at a wavelength range of 280 nm to 450 nm at 20 °C. The scan speed of 200 $nm.min^{-1}$ was used. The same protein samples prepared for far-UV CD spectroscopy (section 2.8.1) were used for this experiment.

2.8.3 Quaternary structure assessment using SE-HPLC coupled with DLS

To assess the molecular weight and hydrodynamic volume of the purified protease size-exclusion high performance liquid chromatography (SE-HPLC) coupled with dynamic light scattering was used. In this technique, proteins with smaller hydrodynamic volume get trapped within the column beads and are retained longer.

The HPLC has a Bio-Select SEC matrix with a resolution of 5 to 250 kDa (BIO-RAD, South Africa). The experimental procedures were carried out at a constant flow-rate of 1 $ml.min^{-1}$ with an isocratic pressure of 65 bar. The temperature was kept constant at 20 °C. The protease was first dialysed against a buffer containing 10 mM sodium acetate, 1 mM NaCl and 2 mM DTT, pH 5 (protease storage buffer). The protease was then filtered before loading to the HPLC column. The buffer used for protease dialysis was used to equilibrate the HPLC matrix. The buffers were de-gassed and filtered before use.

2.8.4 ANS binding studies to native and unfolded wild-type versus the F99A and W42F/F99A proteases

Further assessment of the quaternary structure was carried out using an extrinsic fluorescent probe, ANS (8-anilino-1-naphthalene sulfonate). ANS is an anionic dye that can be used to probe hydrophobic patches on a protein.

In its unbound form, when ANS is excited at 390 nm, it emits light at 540 nm (Semisotnov and Gilmanshin, 1991). When this fluorescent dye binds to accessible hydrophobic patches it undergoes a hypsochromic shift to a wavelength of 460 nm with a simultaneous hyperchromic shift.

The ANS was prepared to a final concentration of 2 mM in buffer containing 20 mM sodium phosphate, 0.01% sodium azide and 2 mM DTT. The concentration of ANS was confirmed by using the extinction coefficient $\varepsilon_{350\ nm}$ = 4 950 $M^{-1}.cm^{-1}$. The protein assays were prepared to a final concentration of 15 µM and incubated for 2 hours with ANS which was added to a final concentration of 200 µM. For the unfolded protease assays, the proteins were prepared to a final concentration of 15 µM in 8 M urea. The proteins were incubated in the urea overnight at 20 °C. The spectral measurements were recorded using a Perkin Elmer LS 50 B Luminescence fluorimeter.

2.9 Thermal unfolding of the wild-type, F99A and W42F/F99A proteases

The thermal stability of a protein is defined by the Gibbs free energy difference between the folded and unfolded protein where

$$\Delta G = G^{o}_{U} - G^{o}_{F} \qquad \textbf{Equation 4}$$

Thermal unfolding can be monitored via differential scanning calorimetry or, in this case, by circular dichroism spectroscopy. Conformational changes are observed simultaneously with an increase in temperature applied to the protein sample.

The experimental parameters that can be extracted are the melting temperature (T_m), change enthalpy ($\Delta H(T_m)$) which are used to determine the stability (ΔG(T)). But, this can only be true if the unfolding transition is thermodynamically reversible (Pace, 1986). The reversibility is frequently impeded by aggregation of the heat-unfolded polypeptide. The aggregation can be monitored or is represented by changes in the voltage (Benjwal *et al.*, 2006).

The thermal unfolding of the HIV-1 protease was monitored at a wavelength of 230 nm. A local maximum was observed for the CD spectra of the native wild-type whereas the mutants had a negative ellipticity at 230 nm. The wild-type, F99A and W42F/F99A samples were prepared to a final concentration of 30 µM in the protease storage buffer. Unfolding was monitored at a temperature range from 20 °C and 100 °C. The rate of the temperature increase was set at 1 °C/min and to control the temperature the Jasco PTC-423S Peltier-type temperature control system was used.

2.10 Functionality of the proteases

A chromogenic substrate, which is an analogue of the conserved cleavage site between the capsid protein (CA) and nucleocapsid (p2) in the Gag polyprotein precursor, was used for the activity studies. The substrate is a synthetic peptide with the amino acid sequence Lys-Ala-Arg-Val-Nle-*p*-nitro-Phe-Glu-Ala-Nle-NH_2. The aromatic -*p*-nitro-Phe moiety is largely responsible for absorption at 300 nm hence cleavage/removal of the *p*-nitro- from the phenyl ring causes decrease in absorbance at this wavelength.

The specific activity, which gives an indication of the amount of pure/functional enzyme, was calculated by plotting the initial velocity versus the amount of protein. The buffer contained 50 mM sodium acetate, 0.1 M sodium chloride, pH 5 for all the experiments. The substrate concentration was kept constant at 50 µM while the enzyme concentration was varied between the range 0.1 µM to 0.26 µM (wild-type)

and 1.8 µM to 3 µM (F99A). In order to compute the maximum velocity (V_{max}) and the substrate binding affinity (K_M), the Michaelis-Menten equation was used:

$$V_0 = \frac{V_{\max}[S]_0}{[S]_0 + K_m} \qquad \textbf{Equation 5}$$

where the V_0 is the initial velocity and $[S_0]$ is the initial substrate concentration.

The wild-type and F99A assays included varied substrate concentrations between 40 µM and 300 µM and the enzyme concentration was kept constant either at 0.240 µM or 2 µM. The V_0 was recorded over 60 seconds of the initial rate of the reaction. When the enzyme concentration is much higher than the K_m:

$$V_{max} = k_{cat}[E]_t \qquad \textbf{Equation 6}$$

where the $[E]_t$ is the total concentration of enzyme and k_{cat} is the turn-over number. The turn-over number is the amount of substrate converted into product per active site per second and is determined from the slope of the initial velocities versus the enzyme concentration.

All assays were performed at 20 °C and data were collected using a Jasco V-630 spectrophotometer at a wavelength of 300 nm (Velazquez-Campoy *et al.*, 2001b). In order to convert the absorbance to reaction rates, the substrate extinction coefficient of 1800 $M^{-1}.cm^{-1}$ was used. The effects of the introduced mutation on the functionality of the protease were analysed and compared to the wild-type. The data was fitted using Sigma Plot v.11 software (Systat software).

Chapter 3 Results

3.1 Multiple sequence and secondary structural alignments

To probe conserved residues at the dimer interface of the protease, multiple sequence alignments were performed using Clustalw2 (Larkin *et al.*, 2007). The conserved residues were identified from a set of aligned amino acid sequences which represented the respective HIV-1 subtypes (Figure 8). Ultimately, all the residues comprising the N- and C-termini antiparallel β-sheet were conserved. This region was also found to be structurally conserved across subtype A, C-SA and even in the drug resistant subtype B protease (Figure 8). To investigate if the N- and C-termini antiparallel β-sheet was also structurally conserved, the selected structures were superimposed using the computational tool, PyMOL (DeLano Scientific, 2006).

3.2 Sequence verification of DNA insert

The pET-WTCSA, pET-F99A-CSA and pET-W42F/F99A-CSA encode HIV-1 subtype C protease wild-type, F99A and W42F/F99A mutants, respectively. These inserts were sequenced using the universal T7 primers at Inqaba Biotec (South Africa) in order to verify the nucleotide sequence of the DNA. The desired mutations were incorporated (Figure 9 and 10) in the insert DNA. The wild-type and variant DNA sequences retrieved from the sequencing results were further translated into their amino acid sequence using ExPasy translation tool (University of Geneva, Switzerland) to verify and compare the wild-type and mutants. A short upstream sequence (Figure 11), which coincides with the autoprocessing site also present in or mimicking the p6-protease cleavage site (Wan *et al.*, 1996), was present in the wild-type and mutant constructs. It is, therefore, essential that during heterologous expression of the protease that auto-cleavage occurs for complete activity of the enzyme.

A

```
                    N-                                  -C
(Subtype A)         1 PQITLWQRPL-----LTQLGCTLNF 99
(Subtype B)         1 PQITLWQRPL-----LTQLGCTLNF 99
(Subtype C,Indian)  1 PQITLWKRPL-----LTQLGCTLNF 99
(Subtype C, SA)     1 PQITLWQRPL-----LPQIGCTLNF 99
(Subtype D)         1 PQITLWQRPL-----LTQIGCTLNF 99
(Subtype F)         1 PQITLWKRPL-----LTQIGCTLNF 99
Drug resistant      1 PQITLWQRPI-----MTQIGCTLNF 99
                      ******:**:-----:.*:******
```

B

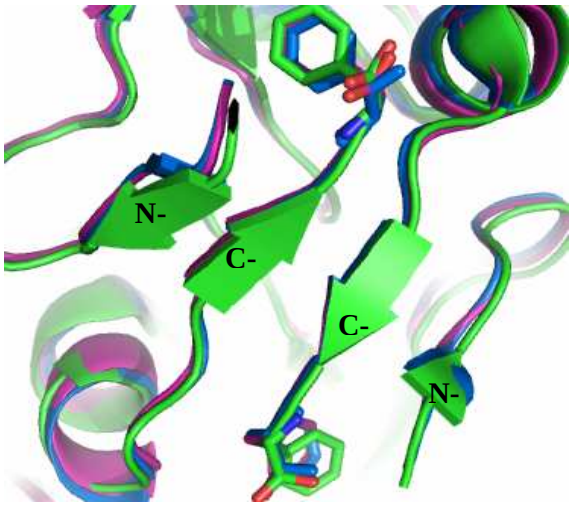


Figure 8. Multiple sequence and structural alignment of different HIV-1 protease subtypes

(A) The segment representing the N- and C-termini residues (red) are indicated. The asterisks represent conserved residues. Double and single dots represent residues which show conserved and semi-conserved substitutions, respectively. The omitted residues are represented by dashed lines. The image was generated using Clustalw2 (Larkin *et al*., 2007). **(B)** Structural alignment of HIV-1 subtype A (green), subtype C SA (pink) and subtype B drug resistant (blue) proteases with PDB codes 3IX0 (Robbins *et al.*, 2010), 3U71 (Naicker, P. *et al.*, unpublished data) and 1RPI (Logsdon *et al.*, 2004), respectively.

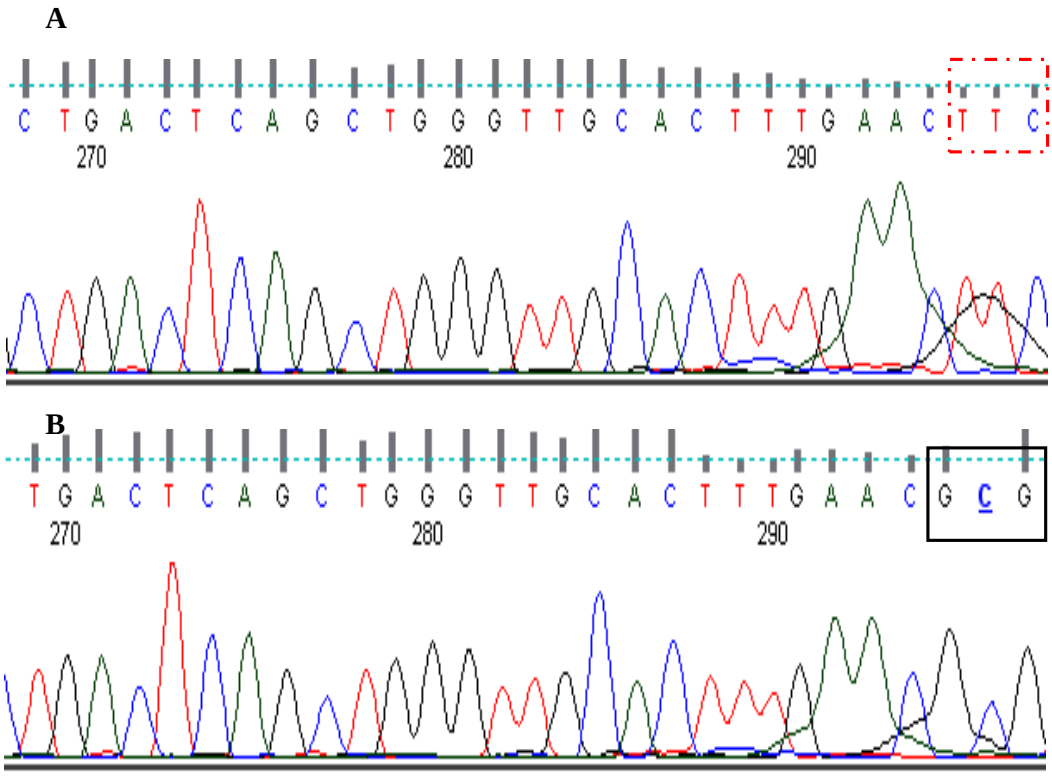


Figure 9. Segments from sequenced HIV-1 subtype C wild-type and F99A insert DNA

The pET-WT-CSA (**A**) and pET-F99A-CSA (**B**) chromatographs are represented. The phenylalanine codon (dotted box) in the wild-type was replaced by an alanine (solid box) yielding the F99A mutant. Images were generated using Finch T.V version 1.4.0. The sequencing data was received from Inqaba Biotec (South Africa).

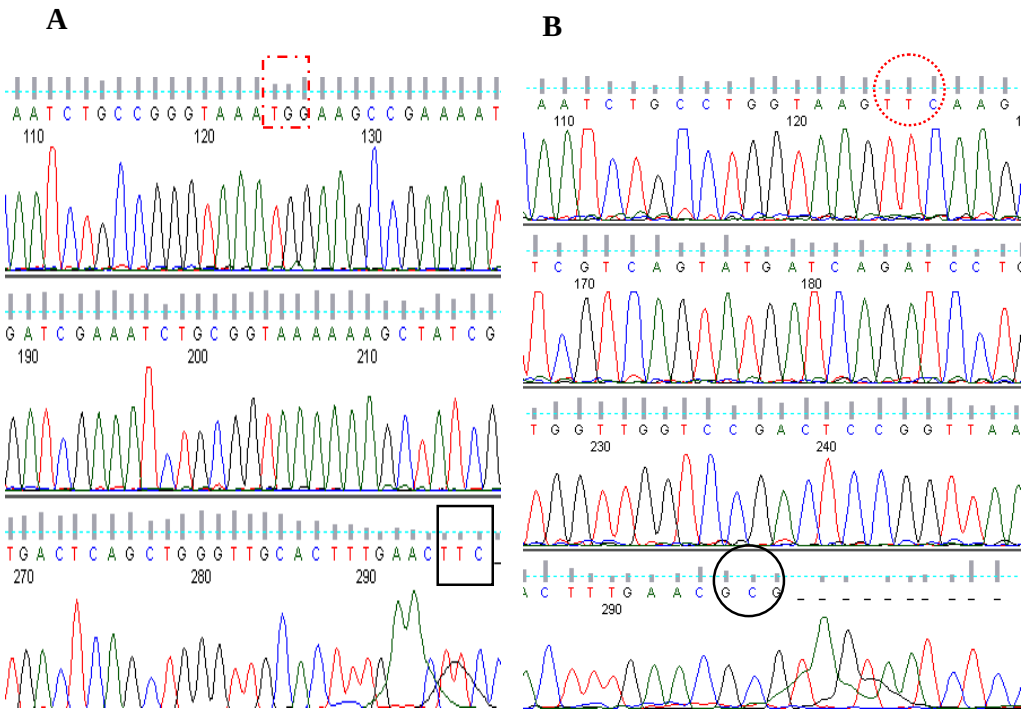


Figure 10. Sections from sequenced HIV-1 subtype C wild-type and W42F/F99A insert DNA

The mutated codons are highlighted in the pET-WT-CSA (A) and pET-W42F/F99A-CSA (B). The W42 (dotted box) and F99 (solid box) codons in the wild-type were replaced by phenylalanine (dotted circle) and alanine (solid circle) codons yielding the W42F/F99A mutant. The images were generated using Finch T.V version 1.4.0. The sequencing data was received from Inqaba Biotec (South Africa).

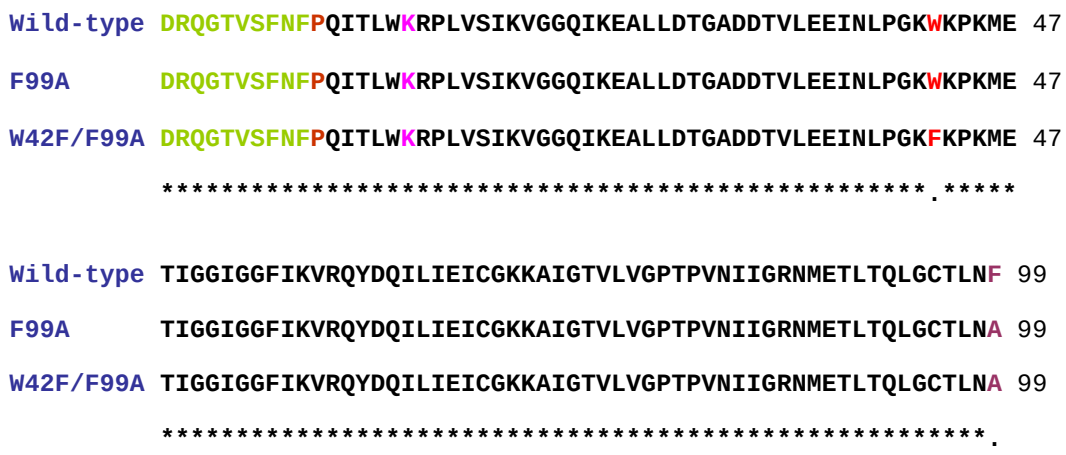


Figure 11. Amino acid sequences of the wild-type, F99A and W42F/F99A cDNA
A decapeptide (green) upstream of the first residue (brown) making up the 99 amino acid residues of the mature protease were identified in all the proteins. The Q7K (pink) mutation was also confirmed in the proteases. The desired mutations at the W42 (red) and F99 (purple) positions were confirmed by comparison to the wild-type sequence. Conserved residues are indicated by asterisks, substitutions are shown by a single dot. The ExPasy (University of Geneva, Switzerland) translation tool was used. The Clustalw2 (Larkin *et al.*, 2007) was used for the alignment.

3.3 Over-expression and purification of the wild-type, F99A and W42F/F99A

The wild-type, F99A and W42F/F99A proteases were over-expressed in *Escherichia coli* T7 Express cells as inclusion bodies and purified under the same set of conditions. The wild-type had a lower expression yield of about 1.25 mg, while the proteases encoding F99A and W42F/F99A, yielded about 16 mg per litre of culture. The differences in concentration were determined spectrophotometrically at A_{280} using the Beer-Lambert law. Tricine SDS-PAGE evaluation indicated that all the proteins were purified to homogeneity as depicted by the single bands corresponding to each protein (Figure 12). The wild-type shows a sharper band while the mutants' migration resulted in smearing. The wild-type and both the F99A and W42F/F99A bands correspond to relative molecular weights of 11 kDa and 7 kDa, respectively. These results were observed with every purification analysis via tricine SDS-PAGE. The wild-type corresponds to the expected HIV-1 protease monomeric molecular weight (Ido *et al*., 1991). The estimated molecular weight for the variants is not reliable since it falls outside the range of the standard molecular weight markers. The wild-type and the mutants have the same pI value of 9.32 (ExPasy tools). It is known that the pI of a protein can interfere with SDS binding hence causing anomalous migration on the SDS-PAGE (Pitt-Rivers and Impiombato, 1968). Therefore the anomalous migration of the mutants may not be attributed to changes in the pI. Further investigations could be done in future to assess the cause of the differences in migration of the wild-type and variants.

3.4 Secondary structure analysis of F99A, W42F/F99A and wild-type

The secondary structural properties of the wild-type, F99A and W42F/F99A were analysed by far-UV circular dichroism spectroscopy. The wild-type spectrum showed a minimum at 214 nm and a local maximum at 230 nm. The minimum is typical of predominantly β-sheeted proteins and the local maximum has also been observed in the literature (Woody, 1995; Noel *et al*., 2009). This, however, is not the case with F99A and W42F/F99A which exhibit minima at 203 nm, and do not portray the local maximum observed for the wild-type at 230 nm (Figure 13).

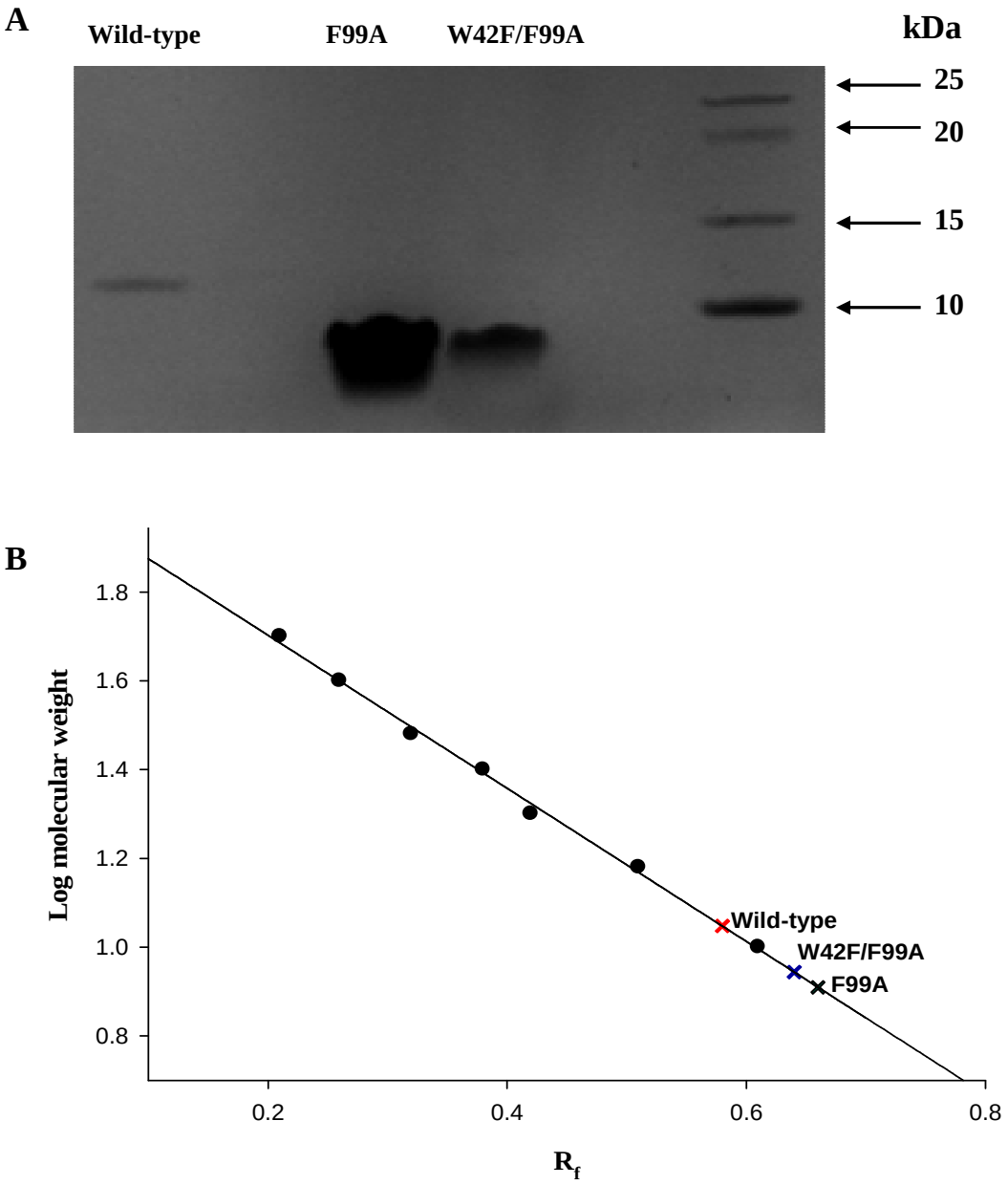


Figure 12. Molecular weight determination using tricine SDS-PAGE of purified proteases

The wild-type, F99A, W42F/F99A and a standard molecular weight marker are indicated in their respective lanes. The proteins were all prepared under reducing conditions. Standard protein molecular weight marker was used to construct the calibration curve. The distance that the proteins migrated and the corresponding molecular weight are indicated on the calibration curve. The curve's regression equation is y = - 1.724 x + 2.048 with the correlation coefficient of 0.99.

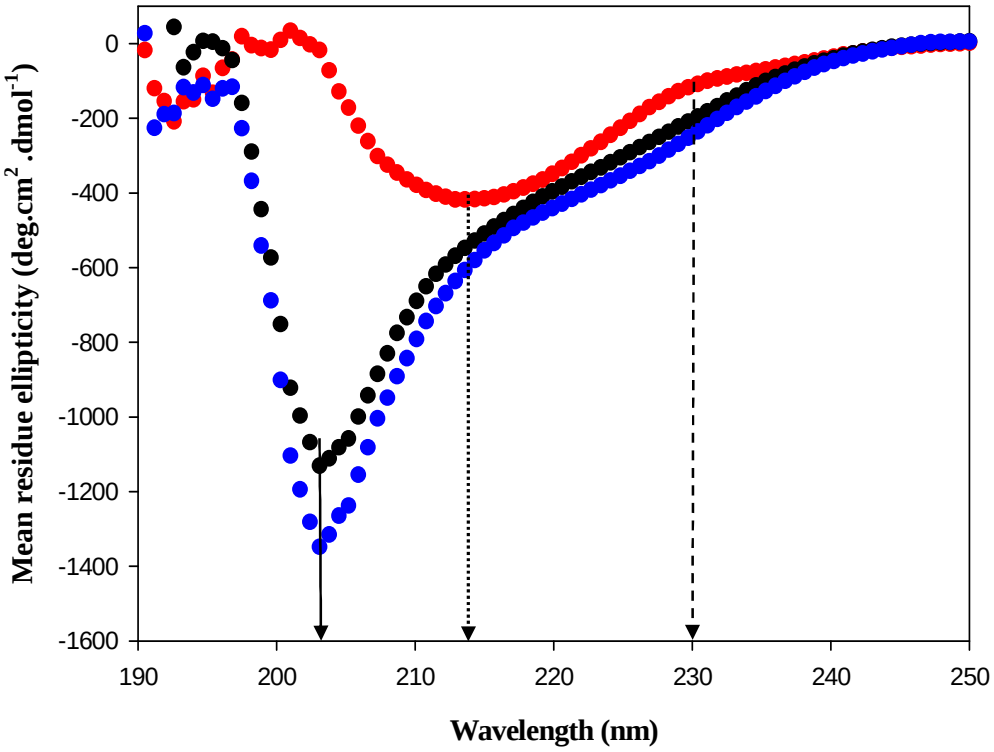


Figure 13. Far-UV circular dichroism spectra of native HIV-1 subtype C proteases

The spectra are a representation of secondary structural elements of native wild-type (●), F99A (●) and W42F/F99A (●). The arrows indicate wavelengths of 203 nm (solid), 214 nm (dotted) and 230 nm (dashed). The proteins were prepared to a final concentration of 15 µM in 10 mM sodium acetate buffer, pH 5. The blanks were subtracted from the raw data.

3.5 Tertiary structure analysis of F99A, W42F/F99A and wild-type

In order to probe tertiary structural changes, if any, intrinsic fluorescence spectroscopy was used. This technique allows the use of phenylalanine, tyrosine and tryptophan residues, but relies mainly on the fluorescence properties of the tryptophan indole with respect to its local environment. The wild-type HIV-1 protease has a single tyrosine (Y59) and two tryptophan residues (W6 and W42) per monomer. Excitation at a wavelength of 280 nm (tryptophan and tyrosine excitation) showed the emission of the wild-type at 355 nm and of both the mutants at 357 nm (Figure 14). The 2 nm red-shift was consistent, occurring after every round of purification. The fluorescence intensity of the mutants was about 50% higher than that of wild-type. The emission wavelength following excitation at 295 nm (tryptophan excitation) was 355, 356 and 357 nm for wild-type, F99A and W42F/F99A, respectively (Figure 15). The 2 nm red-shift is also observed with excitation of the lone tryptophan residue (W6) positioned at the N- and C-termini antiparallel β-sheet. Although a 2 nm shift may be small, but the consistency of this result may indicate that there are tertiary structural changes occurring at this region.

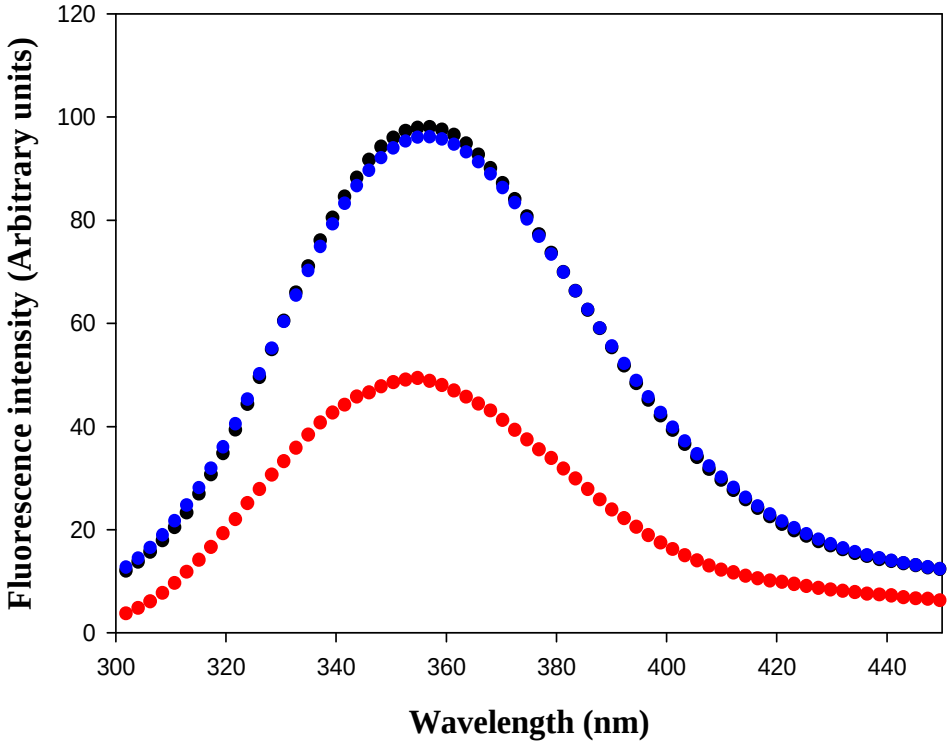


Figure 14. Fluorescence spectra of HIV-1 subtype C proteases with excitation at 280 nm

The emission spectra of the wild-type (●), F99A (●) and W42F/F99A (●) in their native forms are indicated above. Tyrosine and tryptophan residues were excited at 280 nm to give an indication of tertiary structural differences of the proteins. The proteins were prepared to a final concentration of 15 μM in 10 mM sodium acetate buffer, pH 5. The blanks were subtracted from the raw data.

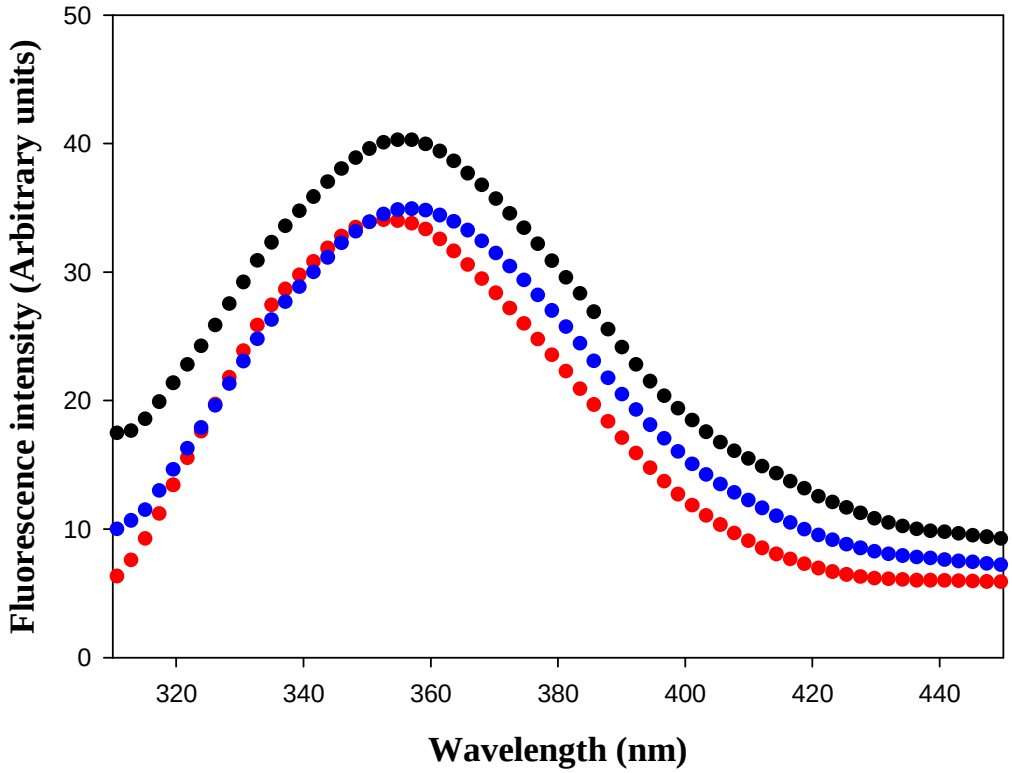


Figure 15. Tryptophan emission spectra of wild-type, F99A and W42F/F99A proteases

The tryptophan emission spectra of native wild-type (●), F99A (●) and W42F/F99A (●).The tryptophan residues were exclusively excited at a wavelength of 295 nm. The proteins were prepared to a final concentration of 15 µM in 10 mM sodium acetate buffer, pH 5. The blanks were subtracted from the raw data.

3.6 Extrinsic ANS-binding fluorescence

The fluorescent dye, ANS (8-anilino-1-naphthalene sulfonate) binds to exposed hydrophobic patches of a protein, and consequently, the emission wavelength is blue shifted with an increase in the quantum yield (Semisotnov and Gilmanshin, 1991). In the presence of the native forms of the wild-type, F99A and W42F/F99A the quantum yield increased from 40 to 46% and there is a blue shift to 480 nm compared to free ANS, which emitted light at a wavelength of 510 nm. The unfolded protein and free ANS emission spectra are red shifted with the quantum yield of the free ANS slightly higher (Figure 16).

3.7 Analysis of molecular weight and hydrodynamic volume of wild-type, F99A and W42F/F99A

Size-exclusion HPLC was employed to compare the hydrodynamic volume and molecular weight of the mutants to that of the wild-type (Figure 17). There are two clear peaks for the wild-type and W42F/F99A. The F99A had a single peak with 'shoulders'. Distinct peaks in Figure 17 suggest different oligomeric states, thus indicating that both dimeric and monomeric species exist in the protein sample. The retention times of the dominant peaks for the wild-type, F99A and W42F/F99A corresponds to 16.4 kDa, 20.7 kDa and 18.1 kDa, respectively. Differences in the elution profiles and molecular weight of the wild-type versus the mutants imply that the F99A mutation could have caused a shift in monomer-dimer equilibrium. The expected relative molecular weight, for the wild-type is 22 kDa. This suggests that the protein behaves anomalously on this HPLC matrix. It is possible that the proteins may be interacting with the matrix causing retardation in the movement of the proteins through the matrix and thus increasing their retention time. Nonetheless there is an observed slight increase in the hydrodynamic volume of the mutants.

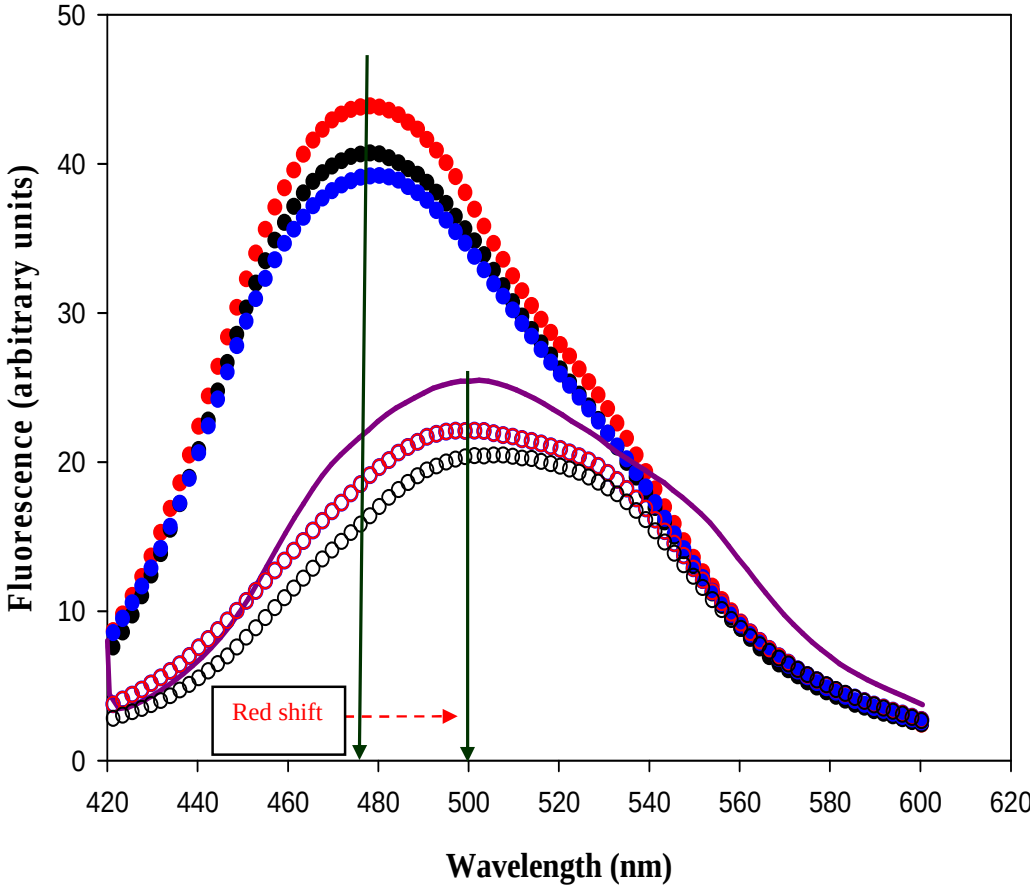


Figure 16. Comparisons of ANS binding to wild-type, F99A and W42F/F99A
ANS binding to native wild-type (●), F99A (●) and W42F/F99A (●) and unfolded forms of the wild-type (○), F99A (○) and W42F/F99A (○). The solid line represents free ANS. The arrows indicate the shift in wavelength where the maximum fluorescence intensity is observed. The wild-type and variant proteases were prepared in 10 mM sodium acetate buffer, pH 5 to final concentrations of 15 µM. The ANS was added to a final concentration of 200 µM.

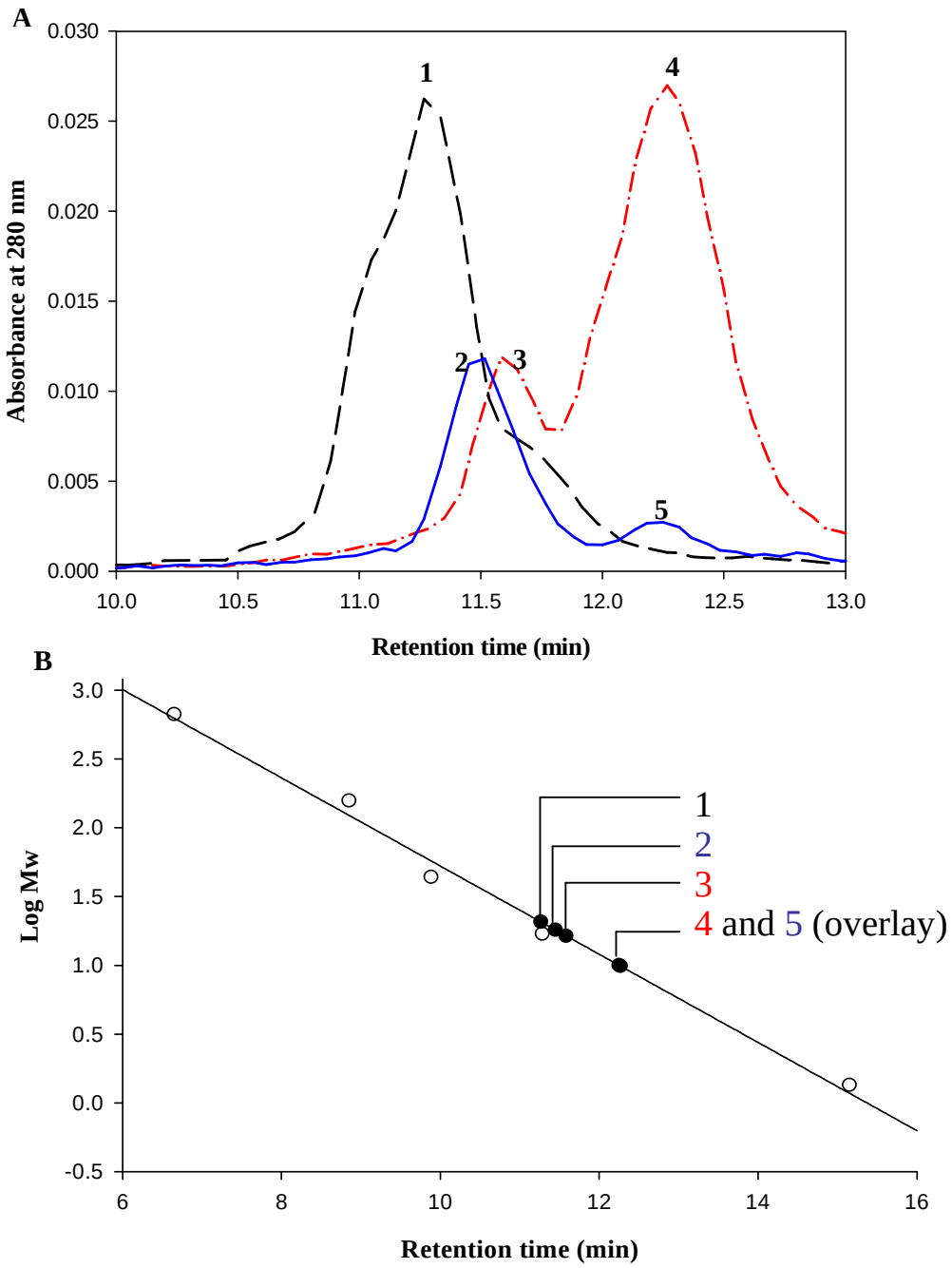


Figure 17. SE-HPLC of the wild-type, F99A and W42F/F99A
(A) The SE-HPLC elution peaks for the wild-type (3 and 4), F99A (1) and W42F/F99A (2 and 5). The flow rate was at 1 ml/min at an isocratic pressure of 65 bar for all the proteins. The equilibration buffer used 10 mM sodium acetate, 1 mM NaCl and 2 mM DTT, pH 5 (B) Standard curve includes: thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B12 (1.35 kDa). The standards and profile peak points are indicated by symbols ○ and ●, respectively. The equation for the curve is y = - 0.321 x + 4.931 and correlation coefficient, 0.98.

3.8 Thermally-induced unfolding of wild-type, F99A and W42F/F99A proteases

The thermal unfolding of the HIV-1 proteases was monitored at a wavelength of 230 nm where the differences in the local maximum was observed between circular dichroism spectra of the native wild-type, F99A and W42F/F99A. The thermal unfolding was monitored at 230 nm, which is mainly indicative of the polypeptide backbone conformation and is also postulated to be due to exciton coupling between W42 and Y59 (Noel *et al*., 2009). The unfolding was irreversible since there was an increase in turbidity concomitant with a large increase in wild-type unfolding at the same temperature (Figure 18). The wild-type shows a more defined unfolding transition between 55 ℃ and 63 ℃ in contrast to the two variants. There is an obvious overlap of the folded/native states (between 20 ℃ and 55 ℃) of the wild-type, F99A and W42F/F99A but the unfolded transition of the variants seems to approach that of the wild-type, but does not overlap with it even at 100 ℃.

3.9 Effects of F99A mutation on the catalytic activity of the protease

Evidence from the secondary structural changes observed as a result of the F99A mutation led to the postulation that it might have an adverse effect on the functionality of the protease. The assay is based on the ability of the protease to hydrolyse the chromogenic substrate, Lys-Ala-Arg-Val-Nle-*p*-nitro-Phe-Glu-Ala-Nle-NH_2. This substrate mimics the conserved cleavage site between the capsid (CA) and nucleocapsid (p2) in the Gag precursor with the sequence (Velazquez-Campoy *et al.*, 2001a; Mosebi *et al*., 2008). The first 60 seconds of the reaction fitted the linear relationship typical of progress curves. It is important to note that the F99A displayed higher levels of noise to signal ratio compared to the wild-type at enzyme concentrations below 1.8 µM. To generate reproducible data from the assays done in triplicate higher enzyme concentrations had to be used. The specific activity of the wild-type was 1.5 µmol/min/mg and that of the F99A was significantly reduced to 0.13 µmol/min/mg, which amounts to a ten-fold decrease (Figure 19). Although the substrate binding affinity (K_M) of the F99A was slightly increased, the V_{max} and k_{cat} were significantly higher for the wild-type (Figures 20, 21 and 22).

The catalytic profile of the wild-type and F99A are summarised in Table 2. These results are comparable with previous studies (Mosebi *et al.*, 2008).

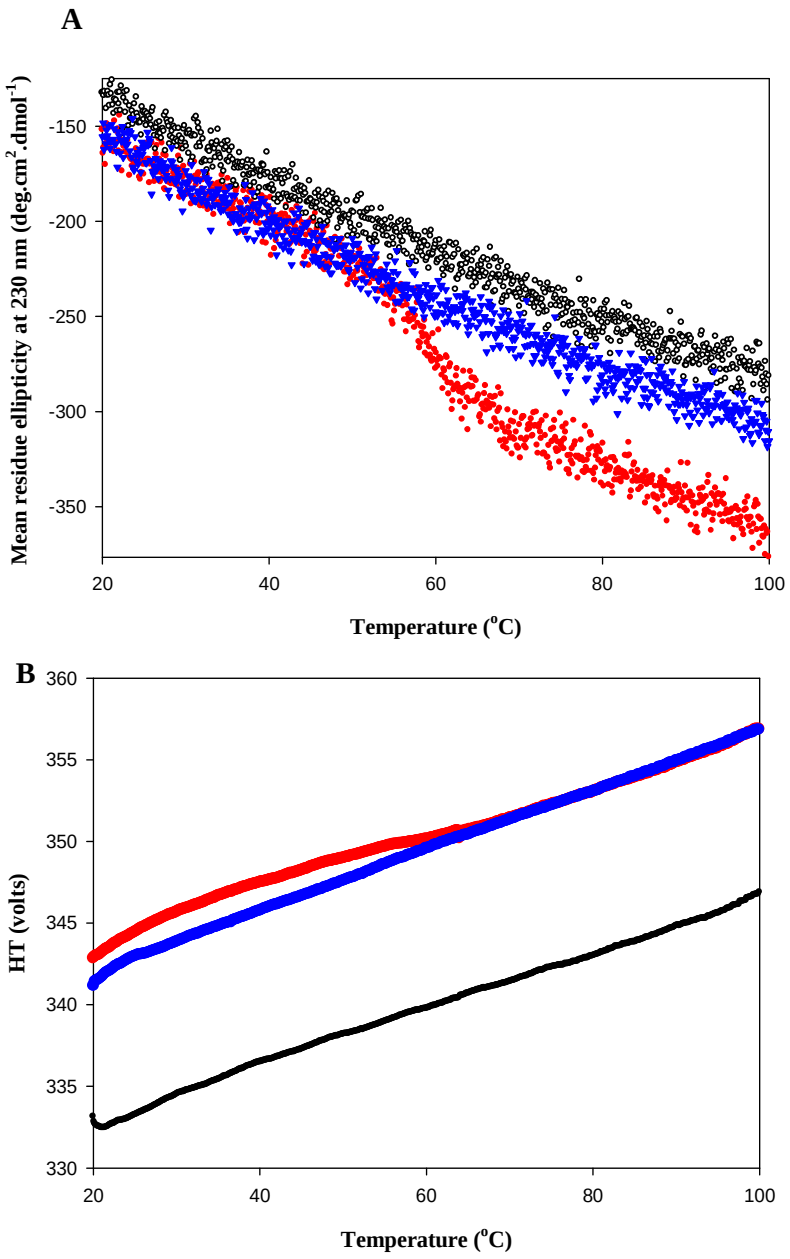


Figure 18. Thermal unfolding of the wild-type, F99A and W42F/F99A
The thermal unfolding of the wild-type (red), F99A (black) and W42F/F99A (blue) were analysed by circular dichroism. The unfolding (A) and turbidity (B) were monitored at a wavelength of 230 nm between 20 °C and 100 °C. The bandwidth was set at 1 nm and the data pitch was 0.2 °C. The proteins were each prepared to a final concentration of 15 μM in buffer containing 10 mM sodium acetate, 2 mM DTT, pH 5.

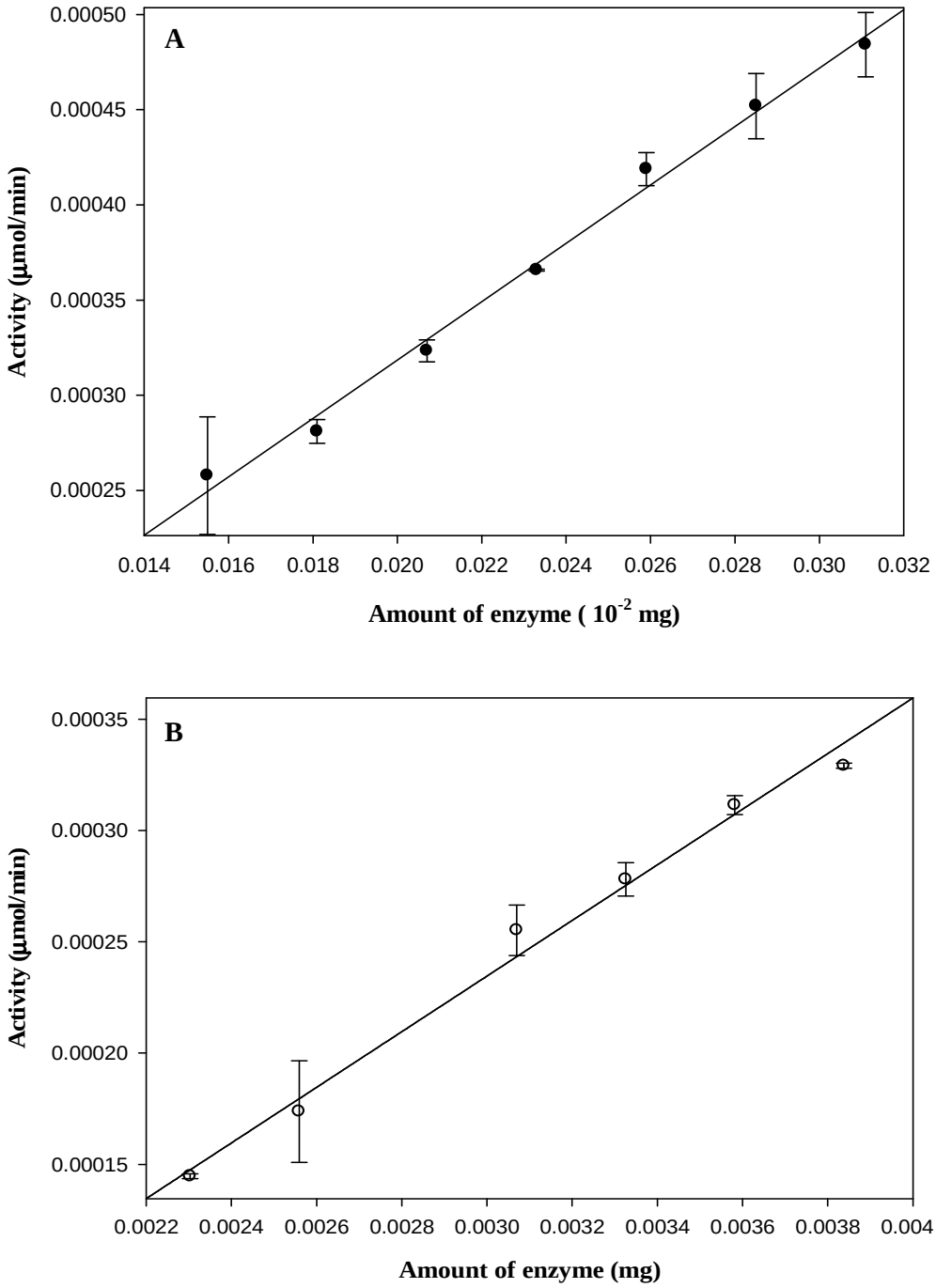


Figure 19. Specific activity of wild-type and F99A HIV-1 subtype C proteases
The specific activity of the wild-type (●) and F99A (○) were computed from the slope of the graphs to be 1.5± 6.3×10^{-4} µmol/min/mg and 0.13± 5.31×10^{-3} µmol/min/mg, respectively. The assays were performed in triplicate and data is reported as mean ± SE. The assays were performed at 20 °C in 50 mM sodium acetate, 0.1 M NaCl, pH 5. The reactions were initiated with 50 µM substrate in varying protease concentration range of 0.4 to 4.2 µM and 1.8 to 3 µM for the wild-type and F99A, respectively. The reactions were monitored at a wavelength of 300 nm. The correlation coefficient is 0.97 for both graphs.

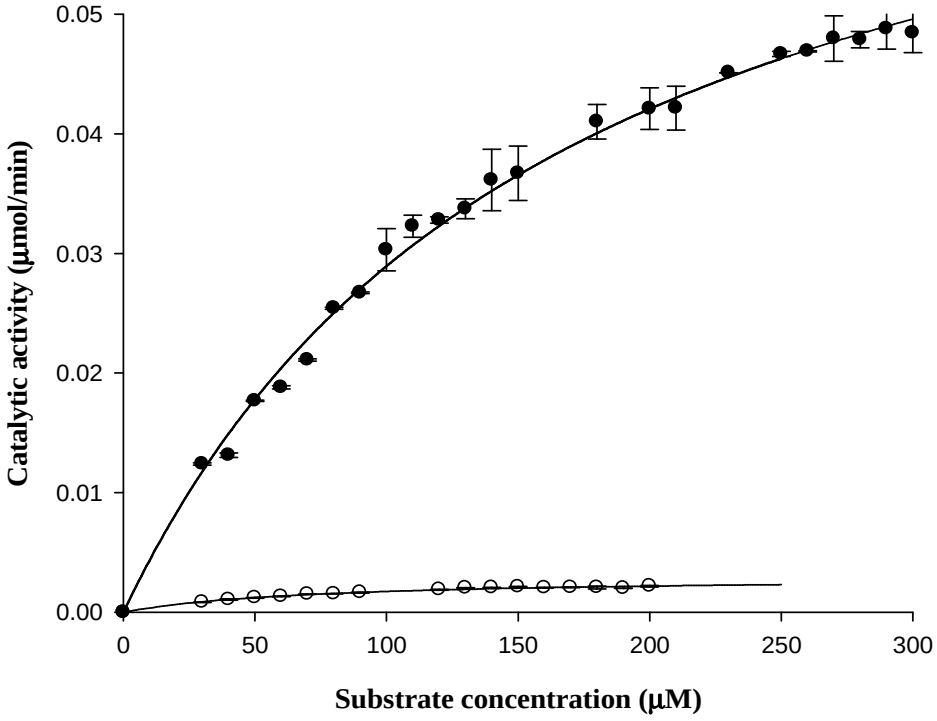


Figure 20. Determination of the Michaelis-Menten constants and maximum velocities of the wild-type and F99A proteases using 2 µM of enzyme
The wild-type (●) and F99A (○) were found to have K_M values of 166.3± 6.6 µM and 98.0± 3.9 µM, respectively. The F99A had a V_{max} of 0.003± 6.32×10^{-5} µmol/min and that of the wild-type was of 0.08± 1.47×10^{-3} µmol/min. The wild-type and F99A protease concentration were kept constant at 2 µM with the substrate concentration varied between 40 and 300 µM. The assays were conducted at 20 °C in 50 mM sodium acetate, 0.1 M NaCl, pH 5. The reactions were monitored at 300 nm. The correlation coefficient of the wild-type and F99A plots are 0.99 and 0.98, respectively. The assays were performed in triplicate and data is reported as mean ± SE.

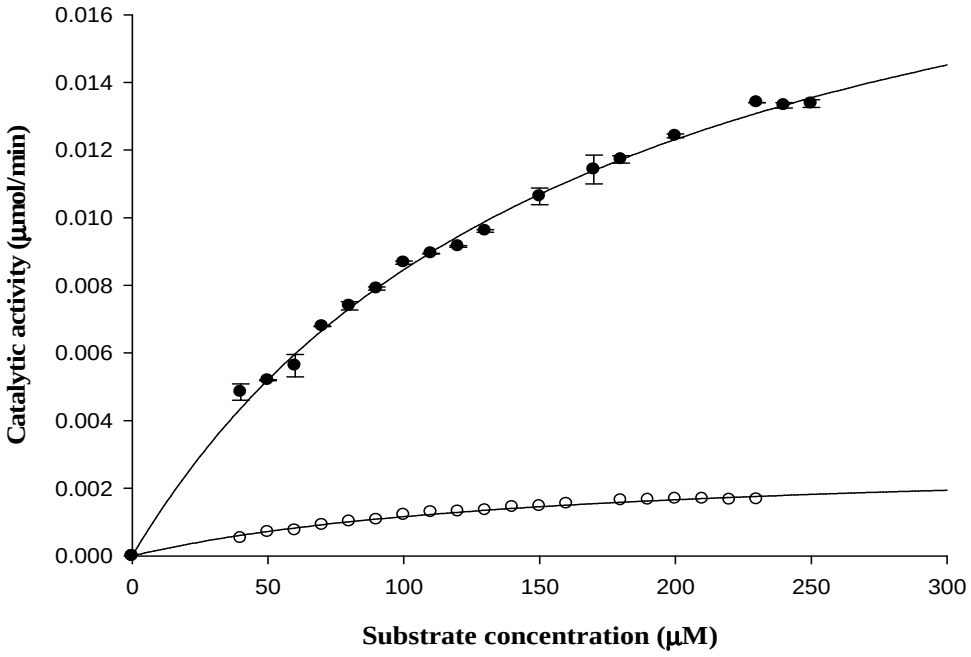


Figure 21. Michaelis-Menten constant and maximum velocity determination of the wild-type and F99A using 0.24 µM enzyme

The K_M for the wild-type (●) and F99A (○) had comparable values of 166.9±5.54 µM and 143.6±1.7 µM, respectively. The V_{max} for the wild-type were determined to be $0.023 \pm 3.95 \times 10^{-4}$ µmol/min and that of the F99A was $0.003 \pm 9.68 \times 10^{-5}$ µmol/min. The reactions were monitored at a wavelength of 300 nm. The assays were performed at 20 °C in 50 mM sodium acetate, 0.1 M NaCl, pH 5. Each assay was done in triplicate. The reactions were monitored at a wavelength of 300 nm. The correlation coefficient of the wild-type and F99A plots are 0.99 and 0.97, respectively. The assays were performed in triplicates and data is reported as mean ± SE.

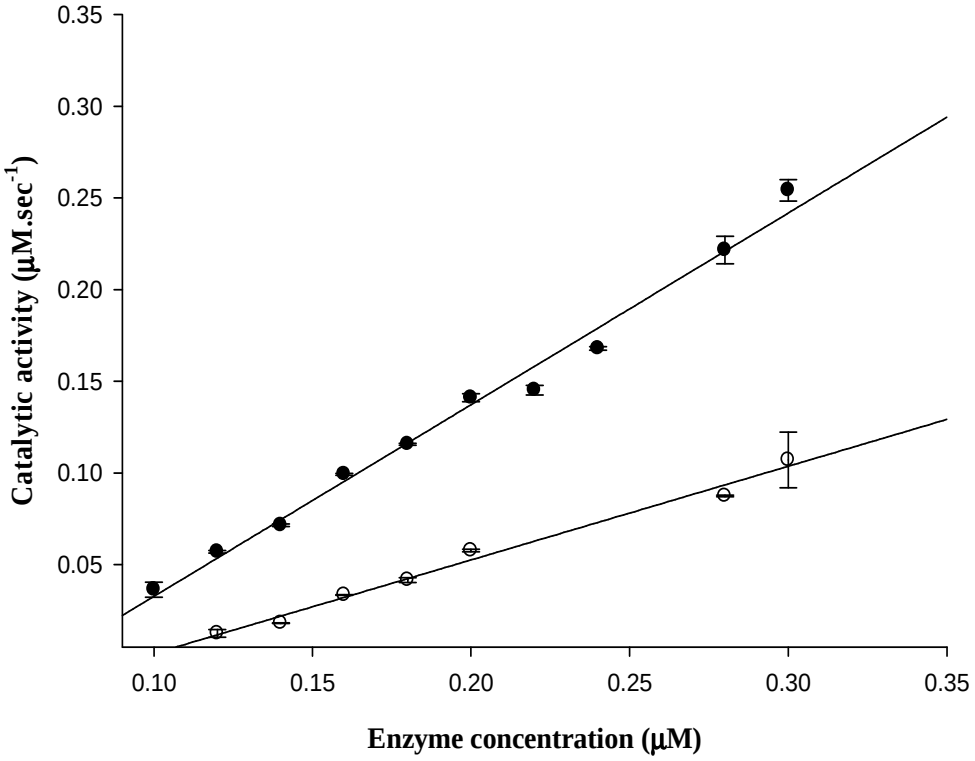


Figure 22. Turn-over determination of the wild-type and F99A proteases
The amount of substrate converted to product per active site (or turn over number or k_{cat}) was calculated from the slopes of the linear curves. The wild-type (●) and F99A (○) k_{cat} were found to be $1.04\pm0.23\times10^{-1}$ sec^{-1} and 0.51 ± 0.0215 sec^{-1}, respectively. Experiments were done at 20 °C in 50 mM sodium acetate, 0.1 M NaCl, pH 5. The protease concentrations were varied between 0.1 to 0.3 µM while the substrate was kept constant at 250 µM. Each reaction was done in triplicate and monitored at 300 nm. The data is reported as mean ± SE.

Table 2: Catalytic profiles of the wild-type and F99A
The catalytic parameters of the wild-type and F99A were computed and summarised below:

Enzyme concentration (µM)	Parameter	Wild-type	F99A
	Specific activity (µmol/min/mg)	1.5	0.13
	k_{cat} (sec^{-1})	1.04	0.51
2	K_M (µM)	166.3	98.0
	V_{MAX} (µmol/min)	0.08	0.003
0.24	K_M (µM)	166.9	143.6
	V_{MAX} (µmol/min)	0.023	0.003

CHAPTER 4 DISCUSSION

The C-terminal residue, F99, is conserved across the HIV-1 protease subtypes including drug resistant strains. The bulky, hydrophobic ring of this residue was replaced with the smaller side-chain of alanine to create a cavity (Figures 23 and 24) within the hydrophobic cleft around the F99. The mutation could also disrupt the 'lock-and-key' motif formed by the F99 and the residues forming the hydrophobic cleft (Figure 6). Engineering of the F99A and W42F/F99A mutations was successful and this was confirmed by Inqaba Biotec sequencing results. The wild-type, F99A and W42F/F99A possess a short peptide upstream (Figure 11) of the protease amino acid sequence. This short upstream peptide mimics the transframe/protease (TF/PR) cleavage site and has been shown to be crucial for the HIV-1 protease self-processing during heterologous expression in *E. coli* cells (Hostomsky *et al*., 1989; Valverde *et al.*, 1992). The F99A and W42F/F99A showed much higher expression yield compared to the wild-type under the same conditions. A decrease in the catalytic activity accompanying the F99A mutation could have caused the enzyme to be less toxic to the *E. coli* cells (Hostomsky *et al*., 1989; Miroux and Walker, 1996; Dumon-Seignovert *et al*., 2004). The role of this residue in the structure, function and thermal stability of the HIV-1 subtype C protease was investigated.

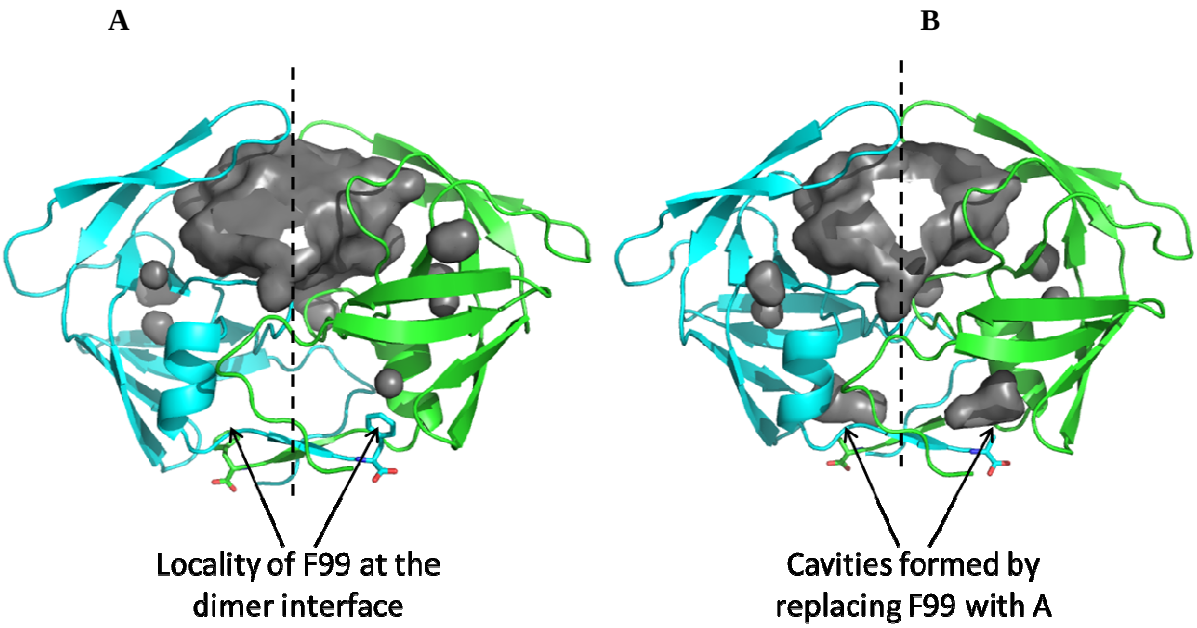


Figure 23 Global view of the F99 position and cavity formed by F99A mutation in the HIV-1 protease

Each subunit is represented by the ribbon structure coloured blue and green, respectively. The wild-type (A) contains the phenyl ring of F99 while in the F99A (B) the phenyl ring is replaced by cavities (grey) in each subunit. The symmetrical axis is indicated by the dashed lines. Image was generated using PyMOL v0.99 (DeLano Scientific, 2006). PDB code 3U71 (Naicker, P. *et al*., unpublished data).

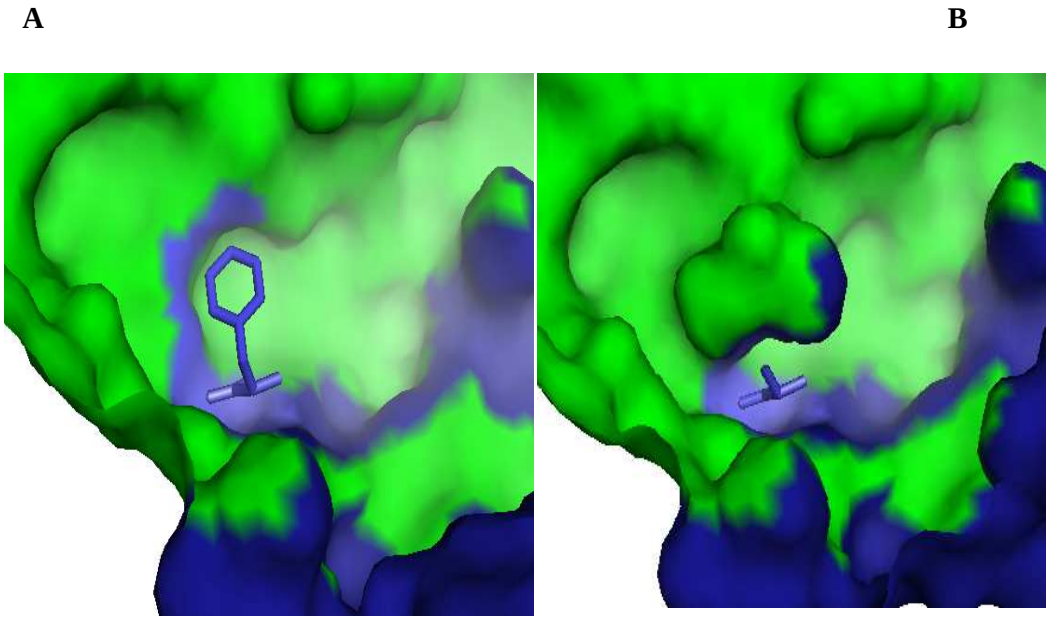


Figure 24. Simulation of the cavity created by substituting phenylalanine with alanine.
The molecular surface area around the (A) F99 and (B) A99 residues is indicated by circles. The pocket occupied by the phenyl ring may be replaced by water molecules in the F99A hence the bulging of the surface area. The green and blue represent the surface area computed for the respective subunits. Image generated using PyMOL v0.99 (DeLano Scientific, 2006). PDB code 2R5Q (Coman *et al*., 2008b).

4.1 Role of F99 on the structure of HIV-1 subtype C protease

The W42F/F99A mutation was created so that tertiary structural changes that have occurred at the N- and C-termini antiparallel β-sheet region can be monitored exclusively. The W42 is solvent exposed, thus, the indole ring does not form any hydrogen bonds (Swiss-pdb Viewer) with any core residues which could be essential for maintaining the native structure of the protease. The bulky side chain of tryptophan was replaced with an equally, but non-fluorescent, bulky phenylalanine. This was done to compensate for the van der Waals interactions that the W42 side chain makes with neighbouring residues in its native wild-type conformation. Thus, any structural changes observed for the W42F/F99A mutant should be due to the F99A mutation.

4.1.1 Secondary structural characterisation

The wild-type, F99A and W42F/F99A were prepared to the same final concentrations following dialysis against DTT-containing buffer to avoid oxidation of sulfur-containing residues. The HIV protease contains two cysteine residues, C67 and C95 per monomer. The C95 residue is located at the dimer interface and its oxidation causes the inhibition of protease catalytic activity by preventing dimerisation (Davis *et al.*, 2003). The proteins were monitored under the same set of parameters. The same results were obtained following every round of purification.

The observed trough at 203 nm implies that the engineered F99A mutation caused a decrease but not complete loss in β-sheet content. It is possible that this loss is localised at the N- and C-termini β-sheet since this is where the mutation was engineered. Synthetic homopolypeptides are often used as models for the basic characterisation of the ‘pure’ spectra of α-helices, β-sheets and random coils (Pauling and Corey, 1951; Manning *et al.*, 1988; Bannister and Bannister, 1974). In contrast to proteins, homopolypeptides form longer chained α-helices and β-sheets and the random coils are extended. This poses a complexity when analysing the more complicated structures of globular proteins.

Therefore, different wavelengths often assigned to a single structural element, might in fact be more multifaceted (Manning *et al.*, 1988; Bannister and Bannister, 1974). Using PyMOL as a simulation tool (Figures 23 and 24), this mutation leaves a cavity within the N- and C-termini antiparallel β-sheet resulting in loss of inter-monomeric van der Waals interactions with the highly conserved C67 (Coman *et al*., 2008b). This presumed perturbation could have diminished the anchoring force between the loop where C67 is positioned and the N- and C-termini antiparallel β-sheet, making these regions less structured (Figure 25). This could have led to marked untwisting of the β-sheets perhaps forming random coils, apparent as the negative band at 203 nm. This phenomenon has been reported for sheets which are less twisted (Manning *et al*., 1988). It has also been shown that increasing the side chain bulk in homopolymers increases the twisting of β-sheets (Bannister and Bannister, 1974) Therefore, loss of the bulky phenyl ring may have caused disruption of the β-sheets. But, whether these changes were local to the N- and C-termini antiparallel β-sheet and if they were significant enough to induce the observed spectral shifts can only be deduced conclusively with a crystal structure.

Although the aforementioned explanation makes good and valid arguments for the significant differences of the far-UV spectra for the wild-type and mutants, the contribution of the aromatic amino acids must also be considered. A series of phenylalanine mutations at different sites of the bovine pancreatic trypsin inhibitor (BPTI) has been previously analysed using far-UV CD (Sreerama *et al.*, 1999). In the study, and several others, it was concluded that mutation of an aromatic residue, depending on its local environment, can significantly alter the far-UV CD spectra (Sreerama *et al*., 1999; Woody, 1995; Boren *et al.*, 1999; Krittanai and Johnson, 1997).

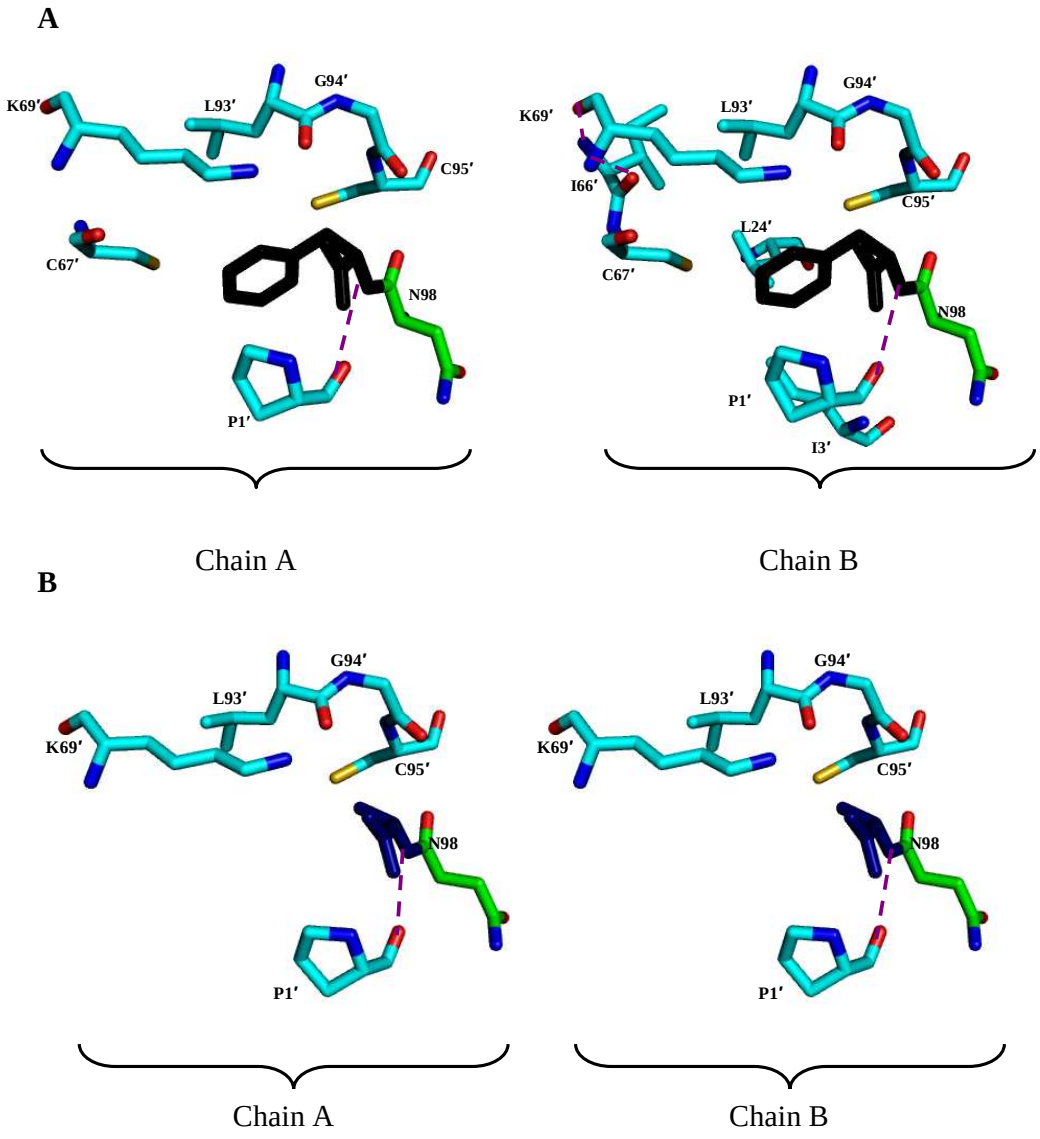


Figure 25. Local environment around F99 and F99A
The wild-type F99 (black) and F99A (blue) with the residues that are within 4 Å depicted in A and B, respectively. The F99, from the respective chains, was used to compute the neighbouring residues. The phenylalanine ring is positioned between C67 and C95. Absence of this ring in F99A causes loss of van der Waals inter-monomer interactions with C67. The prime indicates residues from the paired subunit. Hydrogen bonds are shown by dashed lines. Figure generated using PyMOL v0.99 (DeLano Scientific, 2006) using the PDB code 2R5Q (Coman *et al*., 2008b)

In most cases, the overall arrangement of the backbone of the wild-type and variants was similar in the crystal structures. The amount of aromatic side chains coupled with their local environment can interfere with secondary structural analysis in the far-UV region (Woody, 1994; Boren *et al*., 1999; Krittanai and Johnson, 1997). Even minute modifications in the orientations of the aromatics can cause major differences in CD spectra (Boren *et al*., 1999). In the current study, the F99A mutation might have induced local unfolding at the N- and C-termini antiparallel β-sheet perhaps changing the environment around the nearby W6 and/or its orientation. This is also an early indication that the local environment around the tryptophan(s) has definitely changed as a result of the F99A mutation. The loss of an additional aromatic ring in the W42F/F99A may explain the slight difference in the mean residue ellipticity compared to the F99A (Figure 13). Thus, either the loss of the aromatic phenyl ring itself and/or loss of the conserved interactions with the neighbouring residues (Figure 25), which might have induced partial unfolding of the protease, could account for these results.

The maximum at 230 nm detected in the wild-type far-UV spectra is postulated to be due to exciton coupling between W42 and Y59 which are about 3.4 Å apart (Sreerama *et al*., 1999; Noel *et al*., 2009). The F99 is positioned approximately 34 Å to 39 Å from the postulated exciton coupling in the same monomer and 25 Å to 31 Å in the other monomer. So, it is plausible that the effects of F99 are also global causing the hinge region to be less ordered making the W42 and Y59 less prone to contact and diminishing the chances of the exciton coupling. This maximum has also been attributed to exciton coupling between an aromatic transition and amide transition (Siezen and Argos, 1983). Therefore, either the F99A has induced local unfolding at the N- and C-termini antiparallel β-sheet only, or its unfolding effects were also global by perturbing the presumed W42 and Y59 exciton coupling. This perturbation is represented by absence of the local maxima at 230 nm for the mutant proteases.

4.1.2 Tertiary structural characterisation

Fluorescence spectroscopy was used as a technique for comparing the tertiary structure of the wild-type and mutants. The fluorescence of tryptophan and tyrosine residues as well as their location was used as probes. The samples that were used for the far-UV CD analysis were not discarded but were used for fluorescence spectroscopy analysis. This was done to eliminate any discrepancies due to sample preparation and protein concentration between the CD and fluorescence data. A 2 nm red shift is observed after every round of purification. This suggests that the tryptophan(s), being the major probes, may be slightly solvent exposed. Also, inferring from the loss of the secondary structure, the protease native structure may have become less compact especially at the N- and C-termini antiparallel β-sheet exposing the W6 residue to the solvent. But why do the mutants have a higher intensity versus the wild-type and why do they overlay?

The lower fluorescence intensity exhibited by the wild-type may be due to tryptophan fluorescence quenching. There are several factors that influence the quantum yield but is a poorly understood concept (Lakowicz, 1983). Fluorescence quenching of tryptophan depends on the chemical nature of its neighbouring residues. In its excited state, the indole ring can be quenched by a nearby aromatic, proton from a charged amino group, electron acceptors such as protonated carboxyl group and electron transfer from disulfides, amides or protein backbone (Lakowicz, 1983). The W42 and W6 residues in the wild-type are located such that they make van der Waals contacts with neighbouring amino acids, some of which have been implicated in tryptophan fluorescence quenching. These include R41, K43, P44, R57, Q58 and Y59 near the W42 and making contact with the W6 residue are T4, L5, Q7, R87, T91 and Q92 (Figure 26). Thus, there are possible tertiary structural changes that have occurred at the N- and C-termini antiparallel β-sheet and hinge region.

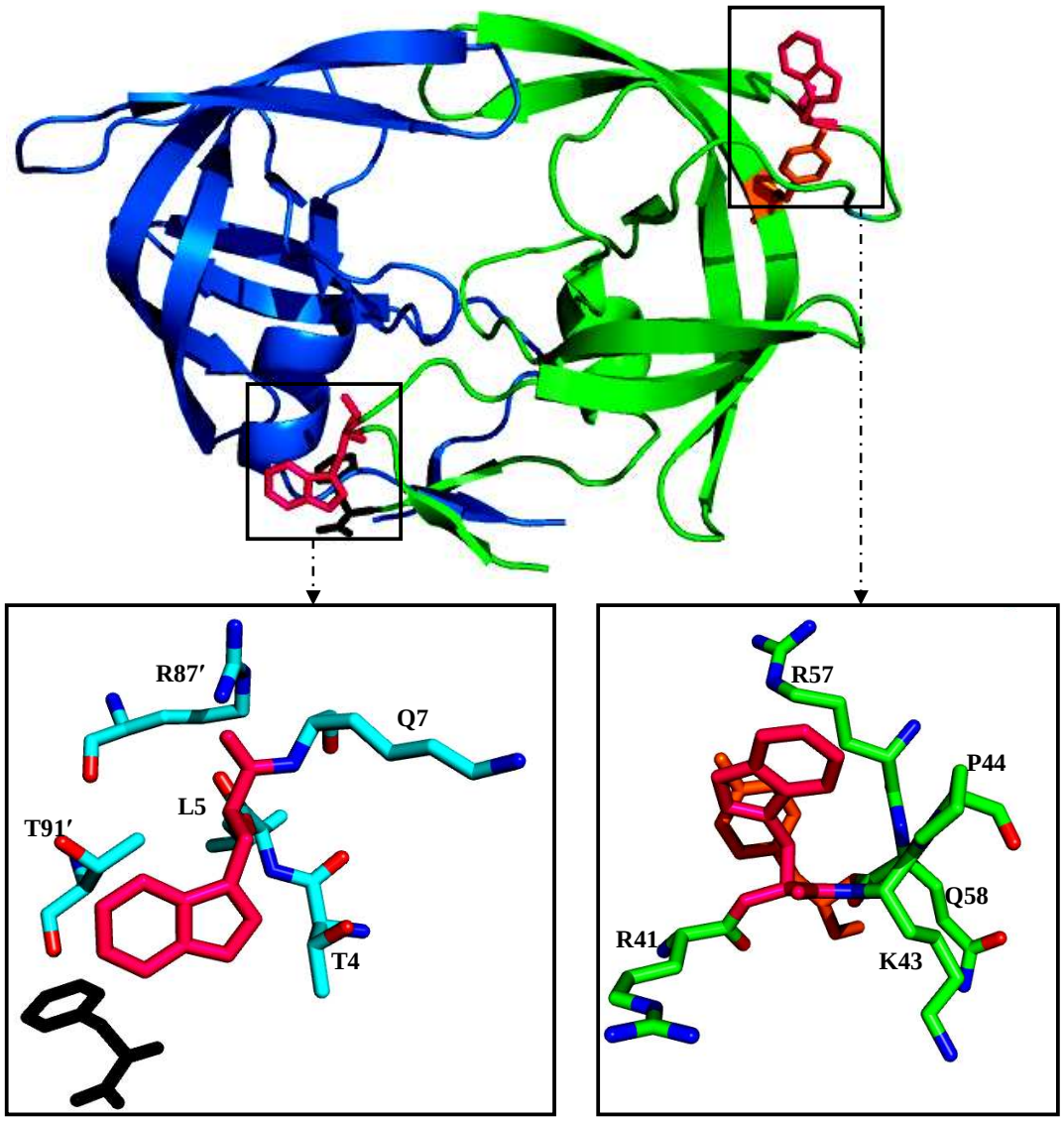


Figure 26. Illustration of the W6 and W42 neighbouring residues

The symmetrical subunits are represented in blue and green. The fluorescent probes W6 and W42, Y59 and F99 are highlighted in pink, orange and black, respectively. The residues that are 4 Å of W6 and W42 are depicted in box A and B, respectively. Image was generated using PyMOL v0.99 (DeLano Scientific, 2006). PDB code 2R5Q (Coman *et al*., 2008b).

The lower fluorescence intensity observed for the wild-type when the tyrosine and tryptophan residues are excited (Figure 14) may be mainly, due to, fluorescence quenching of W6. This residue would be more sensitive to local structural changes at N- and C- termini antiparallel β-sheet since it is positioned near this region and also the fluorescence spectra of F99A and W42F/F99A overlay/superimpose. The W42 and W6, which are both present in the wild-type and F99A proteases, were exclusively excited, and only W6 in W42F/F99A protease was exclusively excited. The W42F/F99A exhibited a 2 nm wavelength red shift compared to the wild-type and this confirms the solvent exposure of the W6 suggesting that the local tertiary structure at the N- and C-termini antiparallel β-sheet region has been disrupted. The intensity difference between the F99A and W42F/F99A, in this case, is due to the absence of W42 residue in the W42F/F99A mutant.

4.1.3 Role in quaternary structural characterisation

The HIV-1 protease quaternary structure was assessed with SE-HPLC (Figure 17). The relative molecular weight, at least for the wild-type, was anomalous. The wild-type is 1.3-fold less than the expected/theoretical value for the dimer and 1.5-fold larger than the expected value for the monomer. Both the wild-type and F99A were active even at concentrations that were 60-fold lower than the final concentrations used for the SE-HPLC experiments. This strongly suggests that the proteins are dimeric. Also the wild-type enzyme kinetic parameters are comparable to previous studies for the wild-type dimer (Mosebi *et al*., 2008; Mpye, 2010). There is still evidence of an altered hydrodynamic volume exhibited by the mutants. The mutants might be slightly less compact than the wild-type suggesting that the F99A might have caused ‘partial’ unfolding at the N- and C-termini antiparallel β-sheet. However, it is not obvious using only this technique to assert if the exposed residues, if any, are hydrophobic or polar.

When ANS binds to hydrophobic patches it displays higher fluorescence intensity and a blue shift (Stryer, 1968; Xie *et al.,* 1999). This property of the ANS was used in probing if there were any exposed hydrophobic patches accompanying the F99A mutation.

But, it has also been shown that the sulfonic group of ANS can bind to positively charged residues of a protein. This also leads to enhanced fluorescence intensity (Matulis and Lovrien, 1998; Andujar-Sanchez *et al.*, 2011). The native and urea-unfolded forms of the wild-type and mutants were incubated with ANS. The native forms of the wild-type and mutants both bind ANS displaying similar fluorescence intensities. In contrast, when they are in the denatured state, binding is not observed. In fact, others showed competitive binding between ANS and acetyl pepstatin implying that ANS binds to the active site of the protease (Xie *et al*., 1999). The SE-HPLC result suggests that dimeric species might still exist in both the F99A and W42F/F99A, although some distortion in the shape is evident. If the ANS does bind at the dimer interface, this further confirms that the dimer is still intact. Therefore, the F99 may be crucial during folding and/or dimerisation of the protease to its fully folded dimeric form which is required for catalysis.

4.2 Importance of F99 on the thermal stability of HIV-1 subtype C protease

Given that the secondary structure and catalytic profiles of the F99A are different from the wild-type, it was imperative to investigate if the stability had been altered. The thermal unfolding was monitored at 230 nm, which is mainly indicative of the polypeptide backbone conformation and is postulated to be due to exciton coupling between W42 and Y59 (Noel *et al*., 2009). Therefore, unfolding might be largely localised at the hinge region. This wavelength has been utilised in other studies to monitor unfolding (Arakawa *et al.*, 2001; Noel *et al*., 2009). It has been shown that at this wavelength the effects of solvent absorption are lower than at shorter wavelengths (Arakawa *et al*., 2001). The transition observed at this wavelength is from the n-π^* transitions of the amide peptide bond (Siezen and Argos, 1983).

The wild-type shows a more defined transition-state in contrast to the mutants. This implies that, unlike with the mutants, an increase in temperature changes the conformation of the wild-type resulting in a clear difference between species that were populated under native versus unfolded conditions.

The transition-state of mutants is independent of changes in temperature suggesting that there is no free energy barrier between the ensemble of conformations through-out the melting curve (Naganathan *et al.*, 2005). This suggests that the conformational changes that are induced by denaturing conditions may already have occurred before thermal denaturation. This is evidence that the same species present at the pre transition-state of the mutants is populated through-out the melting curve becoming less structured with increase in temperature. Therefore, the transition-state of the wild-type is representative of the thermal energy that was required to induce conformational change. The urea-induced equilibrium unfolding of the protease monitored at this wavelength has been described by a two-state mechanism (Noel *et al.*, 2009): $N_2 \Leftrightarrow 2U$. Therefore, the energy barrier presented by the wild-type transition-state may represent the change from the dimeric form to unfolded monomers. The less defined transition-state of the mutants may mean that the dissociation of monomers may already have been induced by the F99A mutation. What could also be happening in the wild-type is that the increase in temperature increases distance between the W42 and Y59 by increasing the flexibility of the hinges. This would decrease the contact between these residues abolishing the exciton coupling as unfolding occurs. The absence of this phenomenon in the mutants means the interaction may have already been perturbed by the F99A mutation.

4.3 Effects of the F99A mutation on the catalytic activity of HIV-1 subtype C protease

It was important to investigate the catalytic activity of F99A since the F99 is essential in maintaining the structural integrity of the protease, hence, it may influence its ability to function efficiently. The assays were performed by following cleavage of the chromogenic substrate which mimics the *in vivo* Gag-Pol precursor cleavage site. The specific activity was significantly reduced by more than ten-fold when compared to that of the wild-type. Interestingly, F99A has a slightly lower K_M value of 98.0 µM versus 166.3 µM for the wild-type. This implies that the mutant binds the substrate slightly tighter. But, the V_{max} of F99A has a significantly lower maximum velocity of 0.0030 µmol/min compared to the wild-type with V_{max} of 0.08 µmol/min. The protein concentration that was used in this case was 2 µM.

When the protein concentration of 0.24 µM was used, the binding affinity was comparable. Nonetheless the maximum velocity was ten-fold higher for the wild-type. The k_{cat} of the mutant was also compromised as it is 2-fold less than the wild-type. The presence of activity in the F99A even at concentrations as low as 0.24 µM proves that it is still dimeric.

The peptide mimicking the CA-p2 cleavage site and protease were co-crystallized in a study by Prabu-Jeyabalan and colleagues (Prabu-Jeyabalan *et al*., 2000). It was found that the protease makes 24 hydrogen bonds with the substrate and only one direct hydrogen bond with the substrate side-chain. This lead to the conclusion that the specificity of the HIV-1 protease is mainly attributed to the shape complementarity induced upon substrate binding (Prabu-Jeyabalan *et al*., 2000). There was a loss in the secondary structural content and changes in the tertiary and quaternary structure accompanying the F99A mutation. Thus, the conformation of the F99A may have altered the positioning of residues that are involved in catalysis hence the adverse effects on the specific/catalytic activity. Whether these residues interact directly or indirectly with the substrate cannot be deduced without a crystal structure.

It was postulated by Singh and colleagues that the HIV-1 protease binds its substrate weakly. This is to its advantage because *in vivo*, the HIV-1 protease is more likely to function under high substrate where weak binding lowers the activation barrier, thus, increasing its catalytic activity (Singh *et al.*, 2011). The weak substrate binding has also been implicated in the so called 'promiscuity' of the protein (Singh *et al*., 2011). Tighter binding to substrate may affect catalysis by destabilising the transition-state binding thereby lowering the V_{max}. This substantiates why the lower K_M of F99A resulted in a reduced maximum velocity. From the tertiary structural studies it is evident that there are slight modifications in the local environment of the W6 and W42 of the F99A mutant. The W6 is 3.05 Å and 4.76 Å from the R87 (of second subunit) and R8 (in the same subunit), respectively. Co-crystallization of the HIV protease and the tetrahedral intermediate showed three water molecules bridging R87 and R8 to the catalytic triad/residues D25-T26-G27 suggesting they play a role in the proper positioning of these catalytic triad/residues during catalysis (Das *et al.*, 2010).

The R8 residue makes direct van der Waals interactions with P3 side-chain of the peptide which mimics the CA-p2 substrate. Therefore, the change in the W6 local environment might have compromised the function of R8 in catalysis possibly by changing its position. This, in turn, could have caused a compromise in the positioning the catalytic residues and the interaction of R8 with P3 side-chain. The W42, on the other hand, is positioned at the hinge region of the flaps. It has been established that the hinge controls the flexibility of the flaps (Kozisek *et al*., 2008; Mosebi *et al*., 2008). The altered W42 environment may have modified the flexibility (more likely to have increased the flexibility) at the flaps, increasing the substrate binding affinity of the F99A but compromised the catalysis. Although there may be an increase in F99A autocleavage, the catalytic activity of the non-precursor-associated protease was severely reduced. Others (Pettit *et al*., 2003) have suggested that in the case of the precursor-associated protease the P1 and F99 side-chains cause stereochemical hindrance, hence, the alanine smaller side-chain increases cleavage at the N-terminus of the protease (Pettit *et al*., 2003).

4.4 Conclusions

The conserved F99 makes interactions that are crucial for the structural integrity of the HIV-1 subtype C protease. These structural modifications have greatly influenced catalysis by perturbing the pairing of the N- and C-termini antiparallel β-sheet and in turn perturbing dimerisation. This also confirmed the importance of this residue in the mechanism of *in vitro* self-processing which simulates the processing in the host cell. The thermal stability of the protein has been compromised and This study has, therefore, demonstrated the importance of F99 and its potential role as a target in the design of small molecule that inhibit dimerisation at the N- and C-termini antiparallel β-sheet.

Future work

The crystal structure and singular value decomposition (SVD) could be used to further investigate the secondary structural changes. Urea-induced stability studies could be employed to specify and compare the thermodynamic parameters for the wild-type and variant proteases. To investigate if the F99A induces changes in the flexibility of the protease Thermolysin digestive studies may be used.

CHAPTER 5: REFERENCES

Alves, C. S., Kuhnert, D. C., Sayed, Y. & Dirr, H. W. 2006. The intersubunit lock-and-key motif in human glutathione transferase A1-1: role of the key residues Met51 and Phe52 in function and dimer stability. *Biochem J,* **393,** 523-8.

Andujar-Sanchez, M., Jara-Perez, V., Cobos, E. S. & Camara-Artigas, A. 2011. A thermodynamic characterization of the interaction of 8-anilino-1-naphthalenesulfonic acid with native globular proteins: the effect of the ligand dimerization in the analysis of the binding isotherms. *J Mol Recognit,* **24,** 548-56.

Arakawa, T., Philo, J. S. & Kita, Y. 2001. Kinetic and thermodynamic analysis of thermal unfolding of recombinant erythropoietin. *Biosci Biotechnol Biochem,* **65,** 1321-7.

Balzarini, J., Stevens, M., De Clerqk, E., Schols, D. and Pannecouque, C. 2005. Pyridine N-oxide derivaves: Unusual anti-HIV compounds with multiple mechanisms of antiviral action. *J of Antimicrobial chemotherapy,* **55**,135-138.

Bannister, W. H. & Bannister, J. V. 1974. Circular dichroism and protein structure. *International Journal of Biochemistry,* **5,** 673-677.

Bannwarth, L., Rose, T., Dufau, L., Vanderesse, R., Dumond, J., Jamart-Gregoire, B., Pannecouque, C., De Clercq, E. & Reboud-Ravaux, M. 2009. Dimer disruption and monomer sequestration by alkyl tripeptides are successful strategies for inhibiting wild-type and multidrug-resistant mutated HIV-1 proteases. *Biochemistry,* **48,** 379-87.

Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H. & Kunkel, T. A. 1989. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J Biol Chem,* **264,** 16948-56.

Benjwal, S., Verma, S., Rohm, K. H. & Gursky, O. 2006. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Sci,* **15,** 635-9.

Bogan, A. A. & Thorn, K. S. 1998. Anatomy of hot spots in protein interfaces. *J Mol Biol,* **280,** 1-9.

Boren, K., Andersson, P., Larsson, M. & Carlsson, U. 1999. Characterization of a molten globule state of bovine carbonic anhydrase III: loss of asymmetrical environment of the aromatic residues has a profound effect on both the near- and far-UV CD spectrum. *Biochim Biophys Acta,* **1430,** 111-8.

Bowman, M. J. & Chmielewski, J. 2002. Novel strategies for targeting the dimerization interface of HIV protease with cross-linked interfacial peptides. *Biopolymers,* **66,** 126-33.

Braman, J., Papworth, C. & Greener, A. 1996. Site-directed mutagenesis using double-stranded plasmid DNA templates. *Methods Mol Biol,* **57,** 31-44.

Broder, S. 2010. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. *Antiviral Res,* **85,** 1-18.

Broglia, R., Levy, Y. & Tiana, G. 2008. HIV-1 protease folding and the design of drugs which do not create resistance. *Curr Opin Struct Biol,* **18,** 60-6.

Burgoyne, N. J. & Jackson, R. M. 2006. Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics,* **22,** 1335-42.

Caffrey, D. R., Somaroo, S., Hughes, J. D., Mintseris, J. & Huang, E. S. 2004. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci,* **13,** 190-202.

Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M. D., Desrosiers, R. C., Tiollais, P. & Sonigo, P. 1987. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature,* **328,** 543-7.

Chothia, C. & Janin, J. 1975. Principles of protein-protein recognition. *Nature,* **256,** 705-8.

Chung, C. T., Niemela, S. L. & 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,** 2172-5.

Coffin, J. M. 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science,* **267,** 483-9.

Coman, R. M., Robbins, A. H., Fernandez, M. A., Gilliland, C. T., Sochet, A. A., Goodenow, M. M., Mckenna, R. & Dunn, B. M. 2008a. The contribution of naturally occurring polymorphisms in altering the biochemical and structural characteristics of HIV-1 subtype C protease. *Biochemistry*, **47,** 731-43.

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

Das, A., Mahale, S., Prashar, V., Bihani, S., Ferrer, J. L. & Hosur, M. V. 2010. X-ray snapshot of HIV-1 protease in action: observation of tetrahedral intermediate and short ionic hydrogen bond SIHB with catalytic aspartate. *J Am Chem Soc*, **132,** 6366-73.

Davis, D. A., Brown, C. A., Newcomb, F. M., Boja, E. S., Fales, H. M., Kaufman, J., Stahl, S. J., Wingfield, P. & Yarchoan, R. 2003. Reversible oxidative modification as a mechanism for regulating retroviral protease dimerization and activation. *J Virol,* **77,** 3319-25.

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

Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry,* **29,** 7133-55.

Dumon-Seignovert, L., Cariot, G. & Vuillard, L. 2004. The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif,* **37,** 203-6.

Finzel, B. C., Weber, P. C., Hardman, K. D. & Salemme, F. R. 1985. Structure of ferricytochrome c' from Rhodospirillum molischianum at 1.67 A resolution. *J Mol Biol,* **186,** 627-43.

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

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

Galiano, L., Ding, F., Veloro, A. M., Blackburn, M. E., Simmerling, C. & Fanucci, G. E. 2009. Drug pressure selected mutations in HIV-1 protease alter flap conformations. *J Am Chem Soc,* **131,** 430-1.

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

Han, J. H., Kerrison, N., Chothia, C. & Teichmann, S. A. 2006. Divergence of interdomain geometry in two-domain proteins. *Structure,* **14,** 935-45.

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

Hoetelmans, R. M., Meenhorst, P. L., Mulder, J. W., Burger, D. M., Koks, C. H. & Beijnen, J. H. 1997. Clinical pharmacology of HIV protease inhibitors: focus on saquinavir, indinavir, and ritonavir. *Pharm World Sci,* **19,** 159-75.

Hostomsky, Z., Appelt, K. & Ogden, R. C. 1989. High-level expression of self-processed HIV-1 protease in Escherichia coli using a synthetic gene. *Biochem Biophys Res Commun,* **161,** 1056-63.

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

Ishima, R., Torchia, D. A., Lynch, S. M., Gronenborn, A. M. & Louis, J. M. 2003. Solution structure of the mature HIV-1 protease monomer: insight into the tertiary fold and stability of a precursor. *J Biol Chem,* **278,** 43311-9.

Janin, J. & Chothia, C. 1990. The structure of protein-protein recognition sites. *J Biol Chem,* **265,** 16027-30.

Jones, S. & Thornton, J. M. 1995. Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol,* **63,** 31-65.

Jones, S. & Thornton, J. M. 1996. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A,* **93,** 13-20.

Keskin, O., Ma, B. & Nussinov, R. 2005. Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol,* **345,** 1281-94.

Koh, Y., Matsumi, S., Das, D., Amano, M., Davis, D. A., Li, J., Leschenko, S., Baldridge, A., Shioda, T., Yarchoan, R., Ghosh, A. K. & Mitsuya, H. 2007. Potent inhibition of HIV-1 replication by novel non-peptidyl small molecule inhibitors of protease dimerization. *J Biol Chem,* **282,** 28709-20.

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

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

Krittanai, C. & Johnson, W. C. 1997. Correcting the circular dichroism spectra of peptides for contributions of absorbing side chains. *Anal Biochem,* **253,** 57-64.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* **227,** 680-5.

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

Lambert, D. M., Petteway, S. R., Jr., Mcdanal, C. E., Hart, T. K., Leary, J. J., Dreyer, G. B., Meek, T. D., Bugelski, P. J., Bolognesi, D. P., Metcalf, B. W. & Et Al. 1992. Human immunodeficiency virus type 1 protease inhibitors irreversibly block infectivity of purified virions from chronically infected cells. *Antimicrob Agents Chemother,* **36,** 982-8.

Lancaster, D. E., Mcneill, L. A., Mcdonough, M. A., Aplin, R. T., Hewitson, K. S., Pugh, C. W., Ratcliffe, P. J. & Schofield, C. J. 2004. Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *Biochem J,* **383,** 429-37.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., Mcwilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* **23,** 2947-8.

Lesk, A. M. & Chothia, C. 1980. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol,* **136,** 225-70.

Li, Q., Hanzlik, R. P., Weaver, R. F. & Schonbrunn, E. 2006. Molecular mode of action of a covalently inhibiting peptidomimetic on the human calpain protease core. *Biochemistry,* **45,** 701-8.

Liu, F., Kovalevsky, A. Y., Tie, Y., Ghosh, A. K., Harrison, R. W. & Weber, I. T. 2008. Effect of flap mutations on structure of HIV-1 protease and inhibition by saquinavir and darunavir. *J Mol Biol,* **381,** 102-15.

Logsdon, B. C., Vickrey, J. F., Martin, P., Proteasa, G., Koepke, J. I., Terlecky, S. R., Wawrzak, Z., Winters, M. A., Merigan, T. C. & Kovari, L. C. 2004. Crystal structures of a multidrug-resistant human immunodeficiency virus type 1 protease reveal an expanded active-site cavity. *J Virol,* **78,** 3123-32.

Louis, J. M., Ishima, R., Nesheiwat, I., Pannell, L. K., Lynch, S. M., Torchia, D. A. & Gronenborn, A. M. 2003. Revisiting monomeric HIV-1 protease. Characterization and redesign for improved properties. *J Biol Chem,* **278,** 6085-92.

Mandel-Gutfreund, Y., Zaremba, S. M. & Gregoret, L. M. 2001. Contributions of residue pairing to beta-sheet formation: conservation and covariation of amino acid residue pairs on antiparallel beta-strands. *J Mol Biol,* **305,** 1145-59.

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

Masso, M. & Vaisman, Ii 2003. Comprehensive mutagenesis of HIV-1 protease: a computational geometry approach. *Biochem Biophys Res Commun,* **305,** 322-6.

Matulis, D. & Lovrien, R. 1998. 1-Anilino-8-naphthalene sulfonate anion-protein binding depends primarily on ion pair formation. *Biophys J,* **74,** 422-9.

Merkel, J. S., Sturtevant, J. M. & Regan, L. 1999. Sidechain interactions in parallel beta sheets: the energetics of cross-strand pairings. *Structure,* **7,** 1333-43.

Mirny, L. A. & Shakhnovich, E. I. 1999. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J Mol Biol,* **291,** 177-96.

Miroux, B. & Walker, J. E. 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol,* **260,** 289-98.

Mosebi, S., Morris, L., Dirr, H. W. & Sayed, Y. 2008. Active-site mutations in the South african human immunodeficiency virus type 1 subtype C protease have a significant impact on clinical inhibitor binding: kinetic and thermodynamic study. *J Virol,* **82,** 11476-9.

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

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

Naganathan, A. N., Perez-Jimenez, R., Sanchez-Ruiz, J. M. & Munoz, V. 2005. Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. *Biochemistry,* **44,** 7435-49.

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

Nelson, M. & Mcclelland, M. 1992. Use of DNA methyltransferase/endonuclease enzyme combinations for megabase mapping of chromosomes. *Methods Enzymol,* **216,** 279-303.

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

Oroszlan, S. & Luftig, R. B. 1990. Retroviral proteinases. *Curr Top Microbiol Immunol,* **157,** 153-85.

Osmanov, S., Pattou, C., Walker, N., Schwardlander, B. & Esparza, J. 2002. Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *J Acquir Immune Defic Syndr,* **29,** 184-90.

Pace, C., Shirley, B., Mcnutt, M. & Gajiwala, K. 1996. Forces contributing to the conformational stability of proteins. *The FASEB Journal,* **10,** 75-83.

Pace, C. N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol,* **131,** 266-80.

Papathanasopoulos, M. A., Hunt, G. M. & Tiemessen, C. T. 2003. Evolution and diversity of HIV-1 in Africa-a review. *Virus Genes,* **26,** 151-63.

Parbhoo, N., Stoychev, S. H., Fanucchi, S., Achilonu, I., Adamson, R. J., Fernandes, M., Gildenhuys, S. & Dirr, H. W. 2011. A conserved interdomain interaction is a determinant of folding cooperativity in the GST fold. *Biochemistry,* **50,** 7067-75.

Pauling, L. & Corey, R. B. 1951. The structure of synthetic polypeptides. *Proc Natl Acad Sci U S A,* **37,** 241-50.

Peeters, M., Piot, P. & Van Der Groen, G. 1991. Variability among HIV and SIV strains of African origin. *AIDS,* **5 Suppl 1,** S29-36.

Perkins, S. J. 1986. Protein volumes and hydration effects. The calculations of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur J Biochem,* **157,** 169-80.

Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. 1960. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature,* **185,** 416-22.

Pettit, S. C., Gulnik, S., Everitt, L. & Kaplan, A. H. 2003. The dimer interfaces of protease and extra-protease domains influence the activation of protease and the specificity of GagPol cleavage. *J Virol,* **77,** 366-74.

Pinkner, J. S., Remaut, H., Buelens, F., Miller, E., Aberg, V., Pemberton, N., Hedenstrom, M., Larsson, A., Seed, P., Waksman, G., Hultgren, S. J. & Almqvist, F. 2006. Rationally designed small compounds inhibit pilus biogenesis in uropathogenic bacteria. *Proc Natl Acad Sci U S A,* **103,** 17897-902.

Pitt-Rivers, R. & Impiombato, F. S. 1968. The binding of sodium dodecyl sulphate to various proteins. *Biochem J,* **109,** 825-30.

Plantier, J. C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemee, V., Damond, F., Robertson, D. L. & Simon, F. 2009. A new human immunodeficiency virus derived from gorillas. *Nat Med,* **15,** 871-2.

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

Robbins, A. H., Coman, R. M., Bracho-Sanchez, E., Fernandez, M. A., Gilliland, C. T., Li, M., Agbandje-Mckenna, M., Wlodawer, A., Dunn, B. M. & Mckenna, R. 2010. Structure of the unbound form of HIV-1 subtype A protease: comparison with unbound forms of proteases from other HIV subtypes. *Acta Crystallogr D Biol Crystallogr,* **66,** 233-42.

Sayed, Y., Wallace, L. A. & Dirr, H. W. 2000. The hydrophobic lock-and-key intersubunit motif of glutathione transferase A1-1: implications for catalysis, ligandin function and stability. *FEBS Lett,* **465,** 169-72.

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

Schagger, H. & Von Jagow, G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem,* **166,** 368-79.

Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. And & Gilmanshin, R. I. 1991. Study of the "molten globule" intermediate state in protein
folding by a hydrophobic fluorescent probe. *Biopolymers,* **31,** 119-28.

Siezen, R. J. & Argos, P. 1983. Structural homology of lens crystallins. III. Secondary structure estimation from circular dichroism and prediction from amino acid sequences. *Biochim Biophys Acta,* **748,** 56-67.

Singh, M. K., Streu, K., Mccrone, A. J. & Dominy, B. N. 2011. The evolution of catalytic function in the HIV-1 protease. *J Mol Biol,* **408,** 792-805.

Sreerama, N., Manning, M. C., Powers, M. E., Zhang, J. X., Goldenberg, D. P. & Woody, R. W. 1999. Tyrosine, phenylalanine, and disulfide contributions to the circular dichroism of proteins: circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor. *Biochemistry,* **38,** 10814-22.

Stryer, L. 1968. Fluorescence spectroscopy of proteins. *Science,* **162,** 526-33.

Tavassoli, A. 2011. Targeting the protein-protein interactions of the HIV lifecycle. *Chem Soc Rev,* **40,** 1337-46.

Todd, M. J., Semo, N. & Freire, E. 1998. The structural stability of the HIV-1 protease. *J Mol Biol,* **283,** 475-88.

Tsai, C. J., Xu, D. & Nussinov, R. 1997. Structural motifs at protein-protein interfaces: protein cores versus two-state and three-state model complexes. *Protein Sci,* **6,** 1793-805.

Unaids. 2009. *AIDS Epidemic Update: Sub-Saharan Africa* [Online]. Available: http://www.unaids.org/en/dataanalysis/epidemiology/epidemiologyslides/ [Accessed 06 December 2011].

Valverde, V., Lemay, P., Masson, J. M., Gay, B. & Boulanger, P. 1992. Autoprocessing of the human immunodeficiency virus type 1 protease precursor expressed in Escherichia coli from a synthetic gene. *J Gen Virol,* **73 (Pt 3),** 639-51.

Velazquez-Campoy, A., Kiso, Y. & Freire, E. 2001a. The binding energetics of first- and second-generation HIV-1 protease inhibitors: implications for drug design. *Arch Biochem Biophys,* **390,** 169-75.

Velazquez-Campoy, A., Todd, M. J., Vega, S. & Freire, E. 2001b. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc Natl Acad Sci U S A,* **98,** 6062-7.

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

Walton, K. L., Makanji, Y., Wilce, M. C., Chan, K. L., Robertson, D. M. & Harrison, C. A. 2009. A common biosynthetic pathway governs the dimerization and secretion of inhibin and related transforming growth factor beta (TGFbeta) ligands. *J Biol Chem,* **284,** 9311-20.

Wan, M., Takagi, M., Loh, B. N., Xu, X. Z. & Imanaka, T. 1996. Autoprocessing: an essential step for the activation of HIV-1 protease. *Biochem J,* **316 (Pt 2),** 569-73.

Weber, I. T. 1990. Comparison of the crystal structures and intersubunit interactions of human immunodeficiency and Rous sarcoma virus proteases. *J Biol Chem,* **265,** 10492-6.

Who. 2010. *Global HIV/AIDS statistics* [Online]. Available: http://www.who.org [Accessed 06 December 2011].

Wlodawer, A. & Gustchina, A. 2000. Structural and biochemical studies of retroviral proteases. *Biochim Biophys Acta,* **1477,** 16-34.

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

Wlodawer, A. & Vondrasek, J. 1998. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu Rev Biophys Biomol Struct,* **27,** 249-84.

Woody, R. W. 1994. Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins. *Eur Biophys J,* **23,** 253-62.

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

Xie, D., Gulnik, S., Gustchina, E., Yu, B., Shao, W., Qoronfleh, W., Nathan, A. & Erickson, J. W. 1999. Drug resistance mutations can effect dimer stability of HIV-1 protease at neutral pH. *Protein Sci,* **8,** 1702-7.

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

Yan, M. C., Sha, Y., Wang, J., Xiong, X. Q., Ren, J. H. & Cheng, M. S. 2008. Molecular dynamics simulations of HIV-1 protease monomer: Assembly of N-terminus and C-terminus into beta-sheet in water solution. *Proteins,* **70,** 731-8.