OVEREXPRESSION AND STRUCTURE-FUNCTION CHARACTERIZATION OF HIV-1 SUBTYPE C REVERSE TRANSCRIPTASE AND PROTEASE

A THESIS SUBMITTED IN FULFILLMENT OF THE DOCTOR OF PHYLOSOPHY DEGREE IN MICROBIOLOGY

BY

TSHIFHIWA TAMBANI

11531937

TO

THE DEPARTMENT OF MICROBIOLOGY

SCHOOL OF MATHEMATICAL AND NATURAL SCIENCES

UNIVERSITY OF VENDA

PROMOTER: PROF. PASCAL O. BESSONG, UNIVERSITY OF VENDA

CO-PROMOTER: PROF. ADDMORE SHONHAI, UNIVERSITY OF VENDA

February 2019

DECLARATION

I, Tshifhiwa Tambani, hereby declare that this dissertation for the award of a Doctor of Philosophy degree in Microbiology at the University of Venda, hereby submitted by me, has not previously been submitted for any examination at this or any other university, and that it is my own work in design and in execution and all the reference material contained herein has been duly acknowledged.

Signed: _____ Date: _____

Tshifhiwa Tambani (11531937)

ACKNOWLEDGEMENTS

I would like to thank God almighty, for his everlasting grace has seen me through it all.

I would also like to extend my thanks to everyone who directly and indirectly contributed to the success of this work

- My promoter: Professor Pascal Obong Bessong (Research Professor: University of Venda) for his excellent supervision, guidance, encouragement and support.

- My co- promoter: Professor Addmore Shonhai (Head, Department of Biochemistry, University of Venda) for the scientific guidance and support.

- Professor Yasien Sayed (Wits University), Professor Salerwe Mosebi (Mintek), Qasim Fish (Mintek), Dr Pravin Naicker (CSIR) and Herna De Wit (University of Johannesburg) for giving access to instrumentation and assistance with experimental protocols

- Members of the Malaria Research Group, Department of Biochemistry, University of Venda for their immense assistance with biochemical techniques.

- Fellow postgraduate colleagues of the HIV/AIDS and Global health Research Group lab for their encouragement and immense assistance.

- University of Venda Research and Ethics committee, Department of Science and Technology, and Technology Innovation Agency (Grant # BPH009) for financially supporting this work.

- National Research Foundation (Grant # 88877) for personal financial support.

DEDICATION

This work is dedicated to God Almighty and my lovely husband and children.

LIST OF ABBREVIATIONS

AIDS	Acquired immunodeficiency syndrome
Arg	Arginine
ARVs	Anti-retrovirals
Asn	Asparagine
bp	base pairs
CA	Capsid
CDC	Center for Disease Control and Prevention
CRFs	Circulating recombinant forms
DMSO	Dimethyl- sulfoxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EVG	Elvitegravir
Env	Envelope
Gag	Group specific antigen

gp	glycoprotein
HAART	Highly active antiretroviral therapy
HSRC	Human Sciences Research council
HIV	Human immunodeficiency virus
HIV-1	HIV type 1
HIV-2	HIV type 2
ICTV	International committee on Taxonomy of Viruses
IN	Integrase
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
kDa	kilodalton
LTR	Long terminal repeat
MA	Matrix
Min	Minute
ml	milliliter
mM	Millimolar
mRNA	Messenger RNA
NC	Nucleocapsid

NI-NTA	Nickel- nitrilotriacetic acid
Nef	Negative factor
NNRTI	Nonnucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PI	Protease inhibitor
Pol	Polymerase
PR	Protease
RAL	Raltegravir
Rev	Regulator of virion
RNA	Ribonucleic acid
RNaseH	Ribonuclease H
rpm	Revolutions per minute
RRE	Rev response element
RT	Reverse transcriptase
SDS	Sodium dodecyl sulphate

SIV	Simian immunodeficiency virus
SIVcp	SIV found in chimpanzees
SIVgor	SIV found in gorillas
SIVsm	SIV found in sooty mangabeys
SU	Surface glycoprotein
ssRNA	Single stranded RNA
Tat	Trans-activating regulatory protein
TM	Trans-membrane
UNAIDS	Joint United Nations Programme on HIV/AIDs
µl	Microliters
URF	Unique Recombinant form
Vif	Viral infectivity factor
Vpr	Viral protein r
Vpu	Viral protein u
Vpx	Viral protein x
WHO	World Health Organisation

TABLE OF CONTENTS

Contents **Page**

LIST OF FIGURES

Figures **Page**

LIST OF TABLES

ABSTRACT

High genetic diversity is a major contributory factor in the development of drug resistance, in addition to challenges in diagnosis and treatment monitoring in the therapeutics of human immunodeficiency virus (HIV) .Within the wide HIV-1 diversity, differences in mutational frequency, disease progression, drug response and transmission amongst HIV-1 subtypes have been shown. In spite HIV-1 subtype C (HIV-1C) being the most prevalent variant globally, none of the available drugs nor screening assays for inhibitory molecules have been developed targeting the genetics of this important subtype. This study therefore aimed to overexpress and biophysically characterize HIV-1C reverse transcriptase and protease to serve as reagents in the development of assays for routine screening of molecules inhibitory to HIV-1C. Heterologous expression of HIV-1C reverse transcriptase and protease isolates that are prevalent in South Africa was carried out in *Escherichia coli (E. coli* (BL21-DE3). The secondary and tertiary structures of the proteins were determined using, circular dichroism (CD) and fluorescence spectroscopy respectively. Thereafter, interaction studies to delineate interaction properties of natural products for possible inhibition of protease were conducted. Furthermore, *in silico* studies to determine binding interactions, further confirmed by *in vitro* binding assays of a pepsin inhibitor homolog (Bm-33) from *Brugia malayi* , against protease were also conducted. Expressed reverse transcriptase and protease from the globally prevalent HIV-1C were shown to be structurally and functionally intact for application in downstream HIV-1 inhibition assays. Interaction studies on the other hand revealed successful inhibition of the expressed HIV-1C PR with gallotanin. Furthermore, binding interactions of Bm-33 and HIV-1 PR revealed the first intermolecular interactions of the two molecules displaying possible inhibition of HIV-1 PR

Key words: HIV-1 subtype C; reverse transcriptase and protease expression; biophysical characterization; inhibition assays; molecular docking.

CHAPTER ONE: GENERAL INTRODUCTION AND LITERATURE REVIEW

1.1. GENERAL INTRODUCTION

Infection with human immunodeficiency virus type 1 (HIV-1), the causative agent of acquired immunodeficiency syndrome (AIDS) remains one of the most devastating epidemics in the world. About 36.7 million people are estimated to be living with HIV and AIDS globally, with South Africa accounting for over 18% of the HIV/AIDS epidemic (UNAIDS, 2017), making HIV/AIDS one of the global health threat and South Africa being the most affected country globally. In the effort to combat the AIDS epidemic, several anti-retroviral drugs (ARVs) that target different replication stages of the virus have been developed (Arts and Hazuda, 2012). However, none of the drugs possess curative effects but only suppress viral replication. Similar to other retroviruses, host cell infection with HIV-1 requires three key steps in the viral replication process. Firstly, reverse transcription of viral genomic RNA into viral DNA by the viral reverse transcriptase; followed by, integration of viral DNA into host cell genome by viral integrase; and lastly, cleavage of newly synthesized viral polypeptides by the viral protease into individual functional proteins for virion assembly and maturation. Following their discovery, all three viral enzymes have been considered targets for antiretroviral therapy.

Over 25 inhibitors have been approved by the US food and drug administration (FDA) for use against HIV (United States Food and Drug Administration, 2014). They function through inhibition of essential enzymes (reverse transcriptase, protease and integrase) or inhibiting viral-target cell interaction (entry inhibition). However, all these inhibitors were synthesized based on HIV-1 B genetic background and have been shown to have high barrier against resistance (Røge *et al.*, 2003; Van Houtte M, 2006), hence further research on other subtypes might give direction to the discovery of more efficient and effective treatment.

The introduction of zidovudine (AZT), a nucleoside reverse transcriptase inhibitor in the late 80s as a monotherapy witnessed high level of toxicity and poor virologic suppression, in addition to the development of multiple resistance mutations (Zhang *et al.*, 1993; Barile *et al.*, 1994). Over the years, more ARVs have been introduced toward management of HIV/AIDS. The introduction of the highly active antiretroviral therapy (HAART) as a combination therapy

comprising two nucleoside reverse transcriptase inhibitors (NRTI) and one non-nucleoside reverse transcriptase inhibitor (NNRTI) or a protease inhibitor (PI) yielded a marked reduction in morbidity and mortality due to HIV/AIDS (Hankins, 2013). Highly active antiretroviral therapy is considered the most effective because of its high reduction in viral resistance development probability (Goldman and Bao, 2004). For this reason, all treatment regimens (first, second and salvage regimens) are combination therapy.

The current ARV therapy play a major role in the treatment and management of HIV-1 infection and has dramatically reduced morbidity and mortality resulting from HIV-1 infection (McMahon and Ward, 2012; Bor *et al.*, 2013). While appreciating the role of ARVs, challenges associated forthwith cannot be ignored. These include toxicity, broad viral genetic diversity and development of viral resistance. Up to date, it has been recorded that HIV-1 develops resistance against all currently used drugs (Kozal *et al.*, 2007, 2012; van den Berg-Wolf *et al.*, 2008). Drug resistance can develop as a consequence of viral evolution in the presence or absence of drug pressure or non-adherence to antiretroviral therapy (ART) prescription. The development of drug resistant mutations is partly due to the error prone reverse transcriptase which lacks a proof-reading mechanism to correct errors made during complementary DNA synthesis from the viral RNA template (Bebenek *et al.*, 1993). The virus is also capable of evading antiretroviral drugs through establishment of latency in resting CD4+ T cells (Siliciano and Greene, 2011).

Studies have elucidated differences in mutational frequency, disease progression, drug response and transmission amongst HIV-1 subtypes (Martínez-Cajas *et al.*, 2008; Pai, Shivkumar and Cajas, 2012; Bhargava *et al.*, 2014). HIV-1 C, the most predominant variant worldwide has been shown to be prone to drug resistance and more virulent than its counterpart HIV-1B (Bessong, 2008; Hemelaar, 2013; Tebit and Arts, 2013) . Stavudine (d4T) induced K56R mutation that leads to broad nucleoside reverse transcriptase inhibitors (NRTI resistance) has been shown to occur more frequently in HIV-1C than other subtypes (Invernizzi *et al.*, 2009; Coutsinos *et al.*, 2011; van Zyl *et al.*, 2011). HIV-1C protease on the other hand, has been shown to be different from subtype B protease at eight positions(Bessong, 2008), an implication in drug resistance.

In the past years, subtype B has been the predominant subtype in Western Europe and the Americas where current drugs are designed and synthesized. Consequently, all drugs licensed for use worldwide are tailored toward HIV-1B genetic background. HIV-1C which at the time was only confined to Southern Africa, is now driving the HIV global epidemic (Tebit and Arts, 2013). More than 50% of all HIV-1 infections world-wide are caused by Subtype C virus whereas subtype B only accounts for 11% (Buonaguro *et al.*, 2007; Hemelaar, 2013). Subtype C is therefore an important variant in the study of pathogenesis, treatment and prevention of HIV infections. Consequently, studies on the inhibition of HIV-1 C enzymes including protease and reverse transcriptase are required. It is therefore imperative to develop protocols for the routine production of active and well characterized HIV-1 C proteins (protease, integrase and reverse transcriptase) as well as determining their structure-function features for the development of HIV-1 C based enzyme inhibition assays. In the current study, protease and reverse transcriptase from South African based HIV-1 C isolates were expressed and characterized for use in subsequent subtype C specific molecule screening assay development.

Since the outcome of this study will be a milestone in HIV-1 C molecule screening assay development, it was therefore imperative to setup a platform that will be used to routinely pre-screen for molecules that would exert inhibitory activities to some degree prior *in vitro* experiments. Although *in vitro* screening of molecules with inhibitory activities have been the most useful tool in drug discovery, general screening without any form of first-line confirmation could result in unsuccessful trials and escalated costs, which could be a drawback in drug discovery. The use of molecular docking facilitates drug discovery, as only a few molecules with assumed potency can be singled out from a library of molecules for further lead optimization. This has proven to be a quick and cost effective first line molecule screening platform (Morris and Lim-Wilby, 2008), an added advantage for continual molecule screening in the high viral diversity and drug resistance overwhelmed HIV-1 research field. This was achieved by subjecting to molecular docking an HIV-1 protease with a pepsin inhibitor homolog (Bm-33) that has been previously shown to have inhibitory activity against human aspartic acid protein, pepsin. *In vitro* assessment of Bm-33 and HIV protease, support the efficiency of molecular docking in the envisaged HIV protein inhibition assays.

LITERATURE REVIEW

1.1.1. Molecular epidemiology of HIV-1

HIV is a member of the genus Lentivirus, which is characterized by slow disease progression with long incubation. It falls under the Retroviridae family, comprised of single stranded, positive sense ribonucleic acid (RNA) viruses (International Committee on Taxonomy of Viruses, 2013). There are two types of HIV: HIV-1 and HIV-2 and both are transmitted either sexually, through blood transfusion or from mother to child. The two types of HIV bear some differences within their genomes. The HIV-1 genome hosts a gene that codes for viral protein 'U' (Vpu) instead of the viral protein 'X' (Vpx) which is present in the HIV-2 genome (Table 1), (Nomaguchi, Fujita and Adachi, 2008; Fujita *et al.*, 2012). Vpx targets the host cell restriction factor SAMDH1 for proteosomal degradation which is implicated in HIV-2 replication since HIV-2 reverse transcriptase is less active than HIV-1 reverse transcriptase (Lahouassa *et al.*, 2012). Both HIV types cause clinically indistinguishable AIDS. HIV-2, is however, less easily transmitted and the period between initial infection and illness is longer than that of HIV-1 (Mikako *et al*., 2012). HIV-1 is the most predominant virus worldwide, while HIV-2 is concentrated in West Africa and is rarely found elsewhere (Gilbert *et al.*, 2003).

The strains of HIV-1 can be classified into four groups: group M (the major group), group O (the outlier) and two new groups, N (non-M, non-O) and P (Thomson, Pérez-Álvarez and Nájera, 2002). Group O is restricted to West-Central Africa and group N which was discovered in 1998 in Cameroon is extremely rare. In 2009, a new group O-like variant which was closely related to simian immunodeficiency virus infecting gorillas (SIVgor) was discovered in a Cameroonian living in France and was classified as group P (Plantier *et al.*, 2009). The majority of HIV-1 infections belong to HIV-1 group M. There are at least nine known genetically distinct subtypes of HIV1 within group M namely, subtypes A, B, C, D, F, G, H, J and K and sixty-one circulating recombinant forms (CRFs), (HIV Databases, 2014). CRFs are identified by numbers in ascending order of discovery then followed by parental subtype letters (i.e. CRF01 AE: the first isolated CRF, a subtype A/E recombinant), (Thompson *et al*., 2002). Subtypes A and F have been further subdivided into sub-subtypes A1, A2, A3 and F1, F2 respectively.

Recombination can occur in different ways, one of which is when two viruses of different subtypes simultaneously infect the same cell of an infected person and when their genetic material mix, thereby resulting in a new hybrid virus or a recombinant. Most of these new strains do not survive for long, but there are those that survive and infect more than one person and are referred to as circulating recombinant forms (CRFs), (Tebit and Arts, 2013). HIV-1 subtypes and CRFs are found predominantly in different geographical regions (Table 1), subtypes A, B and C being the most widespread with C accounting for 50% of the global HIV-1 epidemic (Buonaguro *et.al* 2007, Tebit and Arts, 2013).

Table 1: HIV-1 subtype prevalence per region.

Subtype	**Region**
A	West and Central Africa, Russia (Hemelaar *et al.*, 2011)
CRF A/E	South East Asia, Central Africa (HIV Sequence databases, 2014)
B	Europe, North America, Japan, Australia (Chalmet *et al.*, 2010)
C	Eastern and Southern Africa, India and Nepal (Tebit and Arts, 2013)
D	East and Central Africa (Hemelaar, 2013)
F	Central Africa, South America, Eastern Europe (Chalmet *et al*., 2010)
G and CRF A/G	East and West Africa, central Europe (Gilbert *et al*., 2013)
H	Central Africa (Tebit and Arts, 2013)
J	Central America (Chalmet, 2010)
K	East Africa: Democratic republic of Congo and Cameroon (Plantier *et al*., 2009)

1.2.2. HIV-1 Protease

1.2.2.1. Structure of HIV-1 protease

HIV-1 protease, a 22 kDa homo dimeric protein is encoded from a 297 bp gene in the *pol* region of the HIV genome. Each monomer consists of 99 amino acids (Navia *et al.*, 1989; Kim *et al.*, 1995) and forms part of the active site pocket along the conserved sequence Asp–Thr–Gly, which provides the aspartyl group that is necessary for catalysis (Menéndez-Arias, 2010), a characteristic resulting in the classification of PR under the family of aspartates (Wlodawer and Vondrasek, 1998). Each chain contains an N-terminal Proline residue and C-terminal Phenylanine residue. The position of each monomer forms an axis of symmetry and the secondary structure of each monomer includes one α-helix, and two antiparallel β sheets. The active site cleft is in the center of the protein "covered" by flaps that extend from Methionine-46 to Lysine- 55 (Figure 1.2.2).

1.2.2.2. Catalytic activity of HIV-1 protease

The catalytic activity of HIV-1 protease (PR) is to cleave polyproteins to their component peptides during viral replication. PR hydrolyzes viral polyproteins into functional protein products that are essential for viral maturation and thereby giving rise to infectious virions. This maturation process occurs as the virion buds from the host cell. Without an effective HIV protease, newly synthesized HIV virions remain non-infectious. HIV-1 protease recognizes the asymmetric shape of the peptide substrate, rather than a particular amino acid sequence (Prabu-Jeyabalan, Nalivaika and Schiffer, 2000, 2002). The amino acid sequence of all the cleavage sites is different, but the peptides have a superimposable secondary structure, yielding a substrate “envelope” which is designed to fit within the protease substrate-binding region. The first cleavage takes place at the C-terminal part of p2 (MA-CA-p2↓NC-p1-p6) and subsequent cleavages separate MA from CA-p2 (MA↓CA-p2) and NC-p1 from p6 (NC-p1↓p6). Separation of the small spacer proteins p2 (CA↓p2) and p1 (NC↓p1) is the final catalytic step. The ordered cleavage appears to be regulated by the amino acid sequence near the actual protease cleavage site (Wending *et al*., 2010).

Aliphatic residues stabilize each monomer in a hydrophobic core. Additionally, the dimer is stabilized by noncovalent interactions, hydrophobic packing of side chains and interactions involving the catalytic residues. Each monomer contains two cysteine residues, but these do not form disulfide bonds. The active site forms at the dimer interface, in a cleft between the two domains as part of a four stranded β turn (Zhang *et al*., 2008). Protein-protein associations in the dimer include interactions between the catalytic triad residues (Asp25 to Gly27), Ile50 and Gly51 at the tip of the flaps, and the antiparallel β sheets formed by the four termini in the dimer (residues 1-5 and 95-99). Additional interactions include a complex salt bridge between Asp29 and Arg87 of one subunit and Arg8 of the other subunit (Sluis-Cremer and Tachedjian, 2002) (Figure 4). Asp25 from each monomer holds a water molecule by forming hydrogen bonds. Asp25 and water induce a general acid/base protein hydrolysis. Asp25 acts as an acid as it donates a proton to the carbonyl oxygen of the substrate and acts as a base to accept a proton from the water molecule. A nucleophilic attack of the water molecule results in proteolysis of the substrate. No covalent bonding is involved (Silva *et al*., 1996).

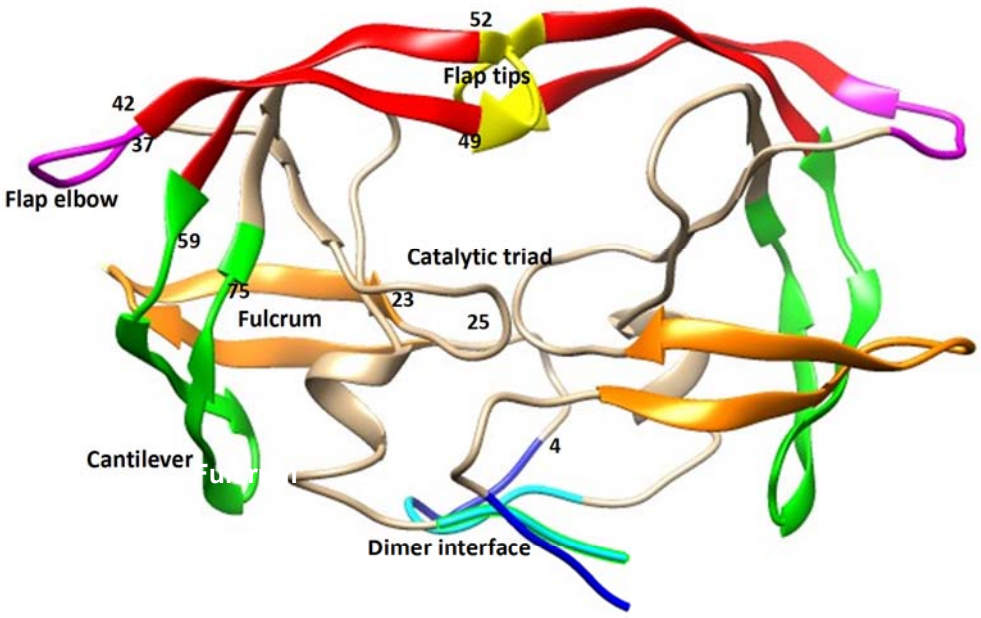


Figure 1.2.2: The crystal structure of the homodimeric HIV protease. Colours indicate distinct regions. Flaps: residues 43-58, red; flap tips: residues 49-52, yellow; flap elbow: residues 37-42, magenta; cantilever: residues 59-75, green; fulcrum: residues 10-23, orange; and interleaved-strand motif forming the dimer interface: residues 1-4 and 96-99, blue/cyan. To generate the structure, protease amino acid sequences were retrieved and submitted to Phyre2 (www.sbg.bio.ic.ac.uk/phyre2), an online tool for protein structure prediction. The model was viewed and manipulated using Chimera version 1.9 developed by Pettersen *et al* (2004), (http://www.rbvi.ucsf.edu/chimera), (https://www.pnas.org/content/108/22/9072.figures-only/ accessed on November 2018.

The mobile flap contains three characteristic regions; side chains that extend outward (Met46, Phe53), hydrophobic chains extending inward (Ile47, Ile54) and a glycine rich region. Upon substrate hydrolysis, a water molecule binds to Ile50 from the interior of the cleft when the protein is unliganded (Rose *et al*., 1998). This water plays a role in the opening and closing of the flaps as well as increasing the affinity between enzyme and substrate (Okimoto *et al.*, 2000). The flap tips (glycine rich region) are highly flexible. This is thought to be necessary for substrate binding and product release. A curling-in movement of the flap forms a hydrophobic cluster as the tips are buried against a hydrophobic wall inside the cleft. The tips curl into the cleft resulting in an open space wide enough for a substrate to enter. Upon curling, the electronegative active site is exposed. Given the electronegative active site in combination with hydrophobic walls, a neutral or positively charged substrate could potentially be guided into a conformation optimal for binding (Scott and Schiffer, 2000).

Protease undergoes substantial conformational changes as the cleft of the active site tightens around a substrate. Together, the two flap regions of protease secure the substrate within the binding cleft. The substrate is held rigidly in place by the flaps in order to cause proper cleavage. The residues in the protease cleft provide pockets into which the side chains of the substrate extend. The substrate binds in an extended β-sheet conformation along this cleft (hydrophobic, charged) through extensive hydrogen and van der Waals bonding. Threonine 26 of the active site remains shielded from the substrate interaction. The basis of substrate recognition is thought to involve the interactions of at least six subsites within the cleft, three subsites being contributed from each monomer (Perez, Fernandes and Ramos, 2010). Hydrophobic subsites closest to the scissile bond have been shown to determine substrate specificity. Residues Pro81, Val82 and Ile84 form the binding pocket. The recognition subsites are identical in each monomer, yet the amino acids on each side of the scissile bond are not the same (Short *et al*., 2000).

1.2.3. HIV-1 Reverse transcriptase

Retroviral reverse transcriptase has two separate catalytic activities, a polymerase that can copy either the RNA or DNA substrate into DNA and an RNaseH that cleaves RNA if, and only if, it is part of an RNA/DNA duplex (Sarafianos *et al.*, 2009).

1.2.3.1. Structure of HIV-1 reverse transcriptase

HIV-1 reverse transcriptase is a heterodimer; the larger subunit, p66, is 560-aa in length and contains both a polymerase and an RNaseH domain. The smaller subunit, p51, contains the first 440 amino acids of p66 that correspond closely, but not exactly, to the polymerase domain (Sarafianos *et al*., 2009). The structure of the polymerase domain of p66 resembles the human right hand, composed of “fingers”, “palm”, and “thumb” subdomains and a connection subdomain linking the polymerase and RNaseH domains (Figure 1.2.3). Although the p51 subunit contains the same four subdomains that are found in the polymerase domain of p66 (fingers, palm, thumb, and connection), the physical relationships of the subdomains differ in p66 and p51. Although p51 contains the same amino acids that form the polymerase active site in p66, there is no functional polymerase active site in p51, and the p51 subunit plays no direct role in polymerization (Ciuffi and Bushman, 2009).

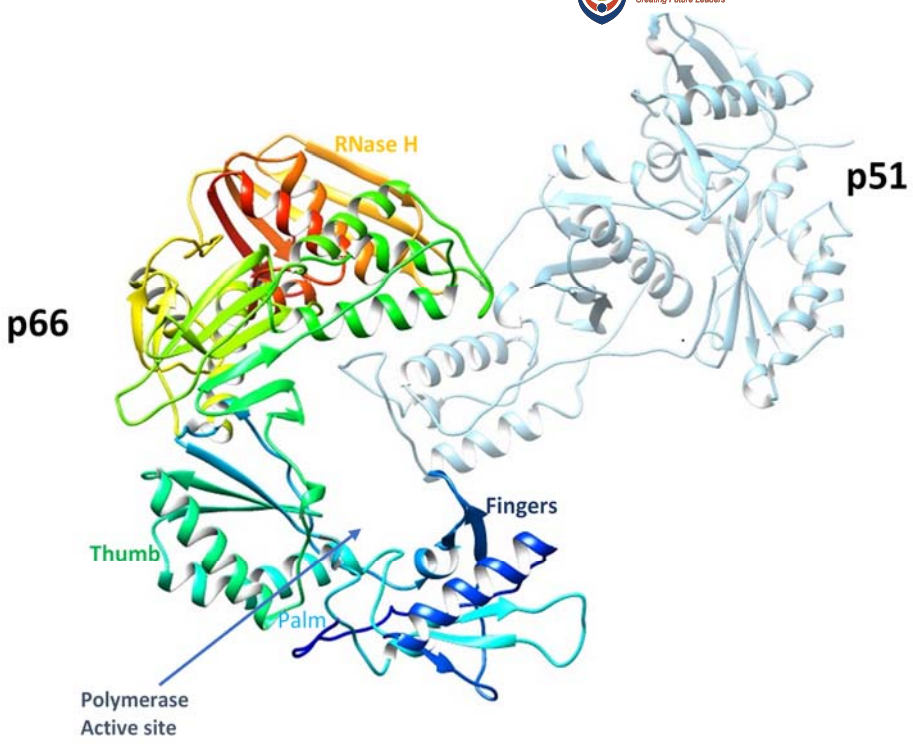


Figure 1.2.3: The crystal structure of HIV-1 reverse transcriptase *illustrating the p51 domain* (dimmed light blue) and p66. The subdomains of p66: fingers (blue), palm (cyan), thumb (lime green), connection (yellow) and RNase H (orange), (https://www.pnas.org/content/98/13/ 6991/ accessed on November 2018). The reverse transcriptase structure was generated as described (Section 1.2.2).

1.2.3.2. Catalytic activity of HIV-1 reverse transcriptase

Reverse transcriptase is used to generate complementary DNA (cDNA) from a viral RNA template (Rodgers *et al*., 1995). The process of reverse transcription is highly error-prone and that leads to high occurrence of drug resistance mutations (Beckhout *et al*., 2001). Synthesis of double-stranded DNA occurs in the host cell cytosol wherein a cellular tRNA acts as a primer and hybridizes to a complementary primer binding site (PBS) on the viral genome. Complementary DNA then binds to the U5 (non-coding region) and R region (a direct repeat found at both ends of the RNA molecule) of the viral RNA. An enzyme RNAse H degrades the 5' end of the RNA which removes the U5 and R region. The primer then "jumps" to the 3' end of the viral genome and the newly synthesized DNA strands hybridizes to the complementary R region on the RNA. The first strand of complementary DNA (cDNA) is extended and the majority of viral RNA is degraded by RNAse H. Once the strand is completed, second strand synthesis is initiated from the viral RNA. Another 'jump' occurs where the PBS from the second

strand hybridizes with the complementary PBS on the first strand. Both strands are extended further and can be incorporated into the host genome by the enzyme integrase (Kati *et al*., 1992)

1.2.4. HIV-1 treatment and drug resistance

Before the introduction of antiretroviral therapy (ART), (Figure 1.2.4), HIV treatment solely depended on prophylaxis of AIDS-related opportunistic infections (Arts and Hazuda, 2012). It has been demonstrated that antiretroviral therapy administered to patients during primary HIV infection improves the clinical course of the disease and increases the CD4+ T-cell count compared to individuals who remain untreated during the period of primary HIV infection (Menendez-Arias, 2002). Inhibitors of HIV-1 reverse transcriptase and protease are widely used in the clinical treatment of AIDS. However, the emergence of drug-resistant variants of HIV-1 severely limits the long-term effectiveness of these drugs. The optimization of current antiretroviral drug regimens and the development of new drugs are challenging issues in HIV chemotherapy (Arts and Hazuda, 2012). Today, HIV positive patients are given a combination of drugs termed highly active antiretroviral therapy (HAART) that attack HIV at various stages in its life cycle. These antiretroviral drugs target protease, reverse transcriptase, integrase and entry of the virus (Weber *et al*., 2011). Figure 1.2.4 illustrates examples of different classes of ARVs and the years in which they were approved. Many challenges are associated with establishing a course of treatment for HIV. Each effective drug has side effects, often very serious and sometimes life-threatening. Common side effects include extreme nausea and diarrhea, liver damage and failure, and jaundice (Enzo Biochem, Annual Report, 2006).

The following are classes of drugs currently being used;

Protease inhibitors (PIs): Bind to the active site of HIV protease, preventing the progression of viral *gag* and *gag-pol* polyproteins precursors and thereby resulting in the formation of immature non-infectious viral particles (Wensing, van Maarseveen and Nijhuis, 2010).

Integrase inhibitors (InSTI): These fall under a class of catalytic inhibitors because they act by inhibiting the catalytic strand transfer activity of the enzyme. They are the latest produced ARVs and are used as salvage therapy (Steigbigel *et al*., 2008).

Reverse transcriptase inhibitors (RTIs): Block reverse transcriptase enzymatic function and prevent completion of synthesis of the double-stranded viral DNA thus preventing HIV from multiplying. The process of reverse transcriptase copying the viral single stranded RNA genome into double-stranded viral DNA is then prevented (De Clercq, 2013)**.**

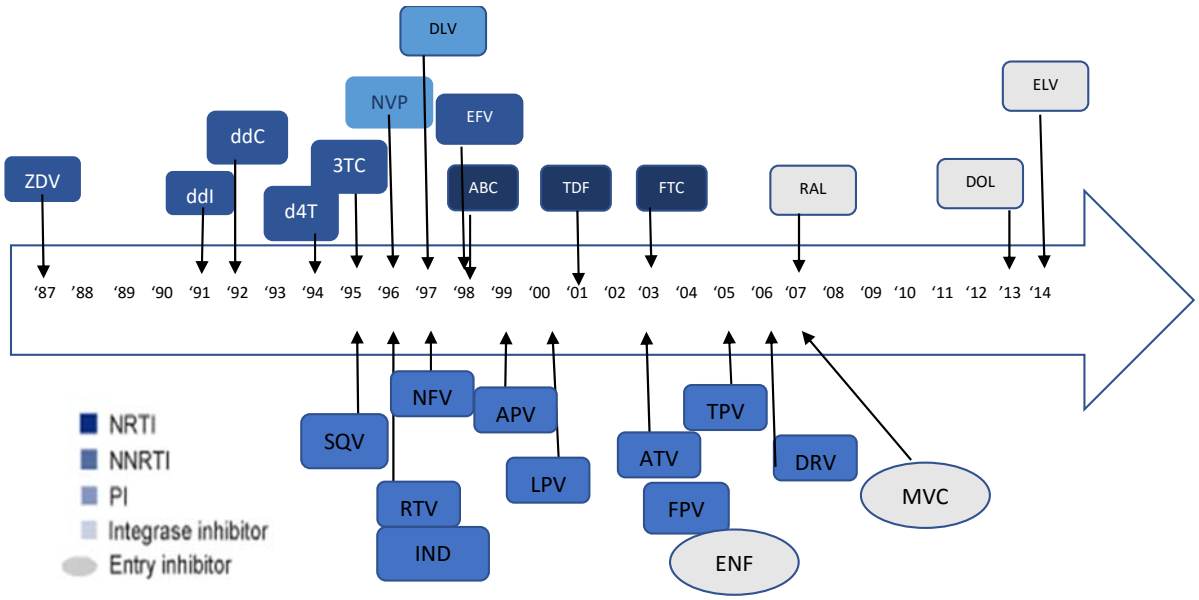
Non-nucleoside reverse transcriptase inhibitors (NNRTIs): Stop HIV production by binding directly onto reverse transcriptase, preventing RNA-dependent and DNA-dependent DNA polymerase activities (Sluis-Cremer and Tachedjian, 2002).

Nucleoside analog reverse transcriptase inhibitors (NARTIs or NRTIs): This was the first effective class of antiretroviral drugs. These inhibitors are analogs of the naturally occurring deoxynucleotides needed to synthesize the viral DNA and they compete with the natural deoxynucleotides for incorporation into the growing viral DNA chain (De Clercq, 2013).

Nucleotide analog reverse transcriptase inhibitors (NtARTIs or NtRTIs): The mode of action of NtARTIs or NtRTIs are the same. Normally, nucleoside analogs are converted into nucleotide analogs by the body (Cihlar and Ray, 2010)**.**

Entry inhibitors: Entry inhibitors block viral entry into the cell, interacting directly with the viral receptor and avoiding fusion of the viral membrane with the target cell membrane (De Clercq, 2013).

Antisense drugs: These drugs are mirror images of parts of the HIV genetic code. The drug locks onto the virus genome to prevent it from functioning (Enzo Biochem, Annual Report, 2006).



3TC, lamivudine; ABC, abacavir; APV; amprenavir; ATV; atazanavir; d4T, stavudine; ddC, zalcitabine; ddI, didanosine; DLV, delaviridine; DRV, duranavir; EFV, efavirenz; ENF; enfurvitide; FPV, fosamprenavir, FTC, emtricitabine; IDV, indinavir; LPV/r, lopinavir/ritonavir; MCV, maraviroc; NFV, nelfinavir; NVP, nevirapine; RAL, raltegravir; RTV, ritonavir; SQV, saquinavir; TDF, tenofovir; TPV, tripranavir; ZDV, zidovudine.

Figure 1.2.4: An Illustration of antiretroviral drugs indicating different classes, examples for each class and the year each drug was licensed. (http://www.clinicaloptions.com/HIV/Treatment%252520Updates/ accessed on November 2018).

1.3. PROBLEM STATEMENT

Since the discovery of human immunodeficiency virus (HIV) in the early 1980's (Gallo and Montagnier, 2003), several drugs have been developed in the quest to combat infection with the virus (Arts and Hazuda, 2012). However, none of the available drugs has the ability to clear the virus out of the infected host. One of the reasons being the capability of HIV-1 to establish the state of latency in resting CD4+ T cells, where it is out of ARVs reach (Siliciano and Greene, 2011), since the drugs only target replicating viruses. Another reason being viral drug resistance development (Sunpath *et al.*, 2012), a consequence of the activity of a high error prone HIV-1 reverse transcriptase which lacks proof reading mechanism to correct errors made during cDNA synthesis (Sarafianos *et al*., 2009). Such factors further contribute to the complication of treatment options for HIV-1.

Broad viral genetic diversity, high drug toxicity and viral resistance remain the main challenges in HIV-1 treatment (Sunpath *et al.,* 2012). The fact that all available antiretroviral agents and assays have been tailored for HIV-1 subtype B genetic background cannot be overlooked. In the past years, subtype B has been the predominant subtype in Western Europe and the Americas. However, subtype C which was previously confined to Southern Africa is now driving the HIV epidemic globally (Tebit and Arts, 2013; Buonaguro, 2007). More than 50% of HIV-1 infections world-wide are caused by HIV-1C whereas subtype B only accounts for 11% . Differences in disease progression, response to ART and transmission amongst HIV-1 subtypes have also been clearly defined (Tylor *et al*., 2008) and subtype C is considered more virulent than B (Hemelaar, 2013).

This study was therefore aimed at developing protocols for routine production of active HIV-1C proteins (reverse transcriptase and protease) as well as determining their structure-function features for application in HIV-1C inhibition assays.

1.4. AIM AND OBJECTIVES OF THE THESIS

AIM: This study aimed to develop protocols that will be used for the overexpression and biophysical characterization of HIV-1C reverse transcriptase and protease towards the development of assays for routine screening of compounds that inhibit their function.

OBJECTIVES:

1. To overexpress and characterize HIV-1 subtype C reverse transcriptase for application in HIV-1 inhibition
2. To overexpress and characterize HIV-1 subtype C protease for application in HIV-1 inhibition
3. To determine binding affinities of HIV-1 Protease and a *Brugia malayi* pepsin inhibitor homolog (Bm-33) for possible inhibition of protease

1.5. STRUCTURE OF THE THESIS

This thesis comprises **5 chapters** compiled as follows:

Chapter one comprises the general introduction, literature review, problem statement as well as objectives for the thesis.

Chapters two to four address the three specific objectives of the thesis and they each comprise an abstract, introduction, objectives, materials and methods, results, discussion and references.

Chapter five outlines the general conclusion and recommendations generated from the results obtained from each chapter.

CHAPTER TWO: EXPRESSION AND CHARACTERIZATION OF HIV-1 SUBTYPE C REVERSE TRANSCRIPTASE

ABSTRACT

Reverse transcriptase has been one of the most targeted enzymes in HIV-1 therapeutics. Its key role in the conversion of the virally encoded RNA to DNA has made it the hallmark in HIV replication. The very first licensed antiretrovirals were targeting the activity of reverse transcriptase. However, high toxicity, adverse side effects and viral drug resistance development has limited the effectiveness of reverse transcriptase inhibitors. Continuous synthesis of newer drugs with reduced toxicity and resistance barrier is of dire need. This chapter focuses on the production as well as structure-function determination of HIV-1 subtype C reverse transcriptase. Plasmids harboring the p51 and p66 HIV-1 reverse transcriptase subunits were co-transformed in *E. coli* and subsequently expressed. Following a nickel affinity purification step of the histidine tagged heterodimeric reverse transcriptase, the protein was evaluated for its catalytic activity. The active protein was then structurally characterized by fluorescence spectroscopy, circular dichroism (CD) spectroscopy and high-pressure liquid column chromatography. Results obtained revealed that the expressed reverse transcriptase was catalytically active to over 2.5% than the commercial HIV reverse transcriptase when analyzed by the calorimetric assay. Confirmation of activity was noted when the expressed reverse transcriptase was seen to produce a HeLa transcript from which, an 80 bp gene fragment was successfully amplified. The CD spectrum of reverse transcriptase at 200-280 nm represented a typical α- helical molecule. An intrinsic fluorescence with a shift of emission spectrum from approximately 338 to 345 nm of the native to unfolded forms of reverse transcriptase were representative of an optimal reverse transcriptase tryptophan residue fluorescence emission. The two peaks representing either one of the two reverse transcriptase monomers (p51 and p66) eluting at 15 minutes and the heterodimeric form at 24 minutes were obtained. The resulting reverse transcriptase was therefore concluded to have been optimally expressed, functionally active and well characterized for use in HIV-1C -based inhibitory molecule screening assay.

2.1. INTRODUCTION

Reverse transcriptase is one of the most targeted enzymes in HIV-1 therapeutics. Its key role in the conversion of virally encoded RNA to DNA has made it the hallmark in HIV replication (Hu & Hughes, 2012). The very first drugs to be produced in HIV therapeutics targeted the activity of reverse transcriptase (Mitsuya *et al.*, 1985). High toxicity, adverse side effects and viral resistance development has since been a limiting factor to the efficacity of reverse transcriptase inhibitors. Side effects such as mitochondrial myopathy, lipoatrophy, lactic acidosis and Liver dysfunction has been associated with some of the more commonly used reverse transcriptase inhibitors (Shikuma *et al*., 2001; Payne *et al*., 2011). Continual synthesis of newer drugs with reduced toxicity and drug resistance barrier is of importance.

Several molecule screening assays have been developed for HIV reverse transcriptase (Suzuki *et al.*, 1993). However, their common disadvantage is that they are based on the HIV-1B genetic background as they are produced in Western countries where B is the most prevalent HIV-1 variant. With an overwhelming HIV genetic diversity, subtype specific molecule screening assays would be relevant for the discovery of drugs with a reduced resistance barrier. HIV-1 variants other than B, with an inclusion of subtype C and circulating recombinant form A-G (CRF AG) are predominantly spreading globally (Babosa *et al*., 2019). Subtype C has now predominated over 50% of the global HIV epidemic with a prevalence of over 90% in the Southern African region (Wilkinson *et al.*, 2016). With such statistics, development of subtype C based molecule screening assays is amicable.

This chapter focuses on producing a catalytically active HIV-1C reverse transcriptase that is well characterized for application in the development of HIV-1C specified molecule screening assay. This application will be of great importance in countries in the Sothern African region where HIV-1C predominates.

2.2. STUDY OBJECTIVES

2.2.1. Aim

To overexpress and purify an HIV-1C reverse transcriptase.

2.2.2. Objectives

- To express and purify an HIV-1 C reverse transcriptase in *E. coli*
- To determine the secondary and tertiary structures of the expressed reverse transcriptase
- To evaluate the proteolytic activity of the expressed reverse transcriptase

2.3. MATERIAL AND METHODS

2.3.1. Sources of reagents

Plasmids p6H-RT51 and pET-RT66 were obtained from Professor Yasien Sayed of the Witwatersrand University (Johannesburg, South Africa) bearing South African HIV -1 subtype C reverse transcriptase subunits p51 and p66 respectively.

2.3.2. Buffers

Buffer A (Resuspension and Wash): 50 mM Tris–HCl, pH 7.9, 60 mM NaCl, 10% glycerol, and 1 mM 2-mercaptoethanol. Buffer B (Elution): same as Buffer A but contained 500 mM instead of 60 mM NaCl. Buffer C (Refolding): same as Buffer A, but Tris–HCl pH was at 6.5 instead of 7.9. Buffer D (Storage): same as buffer A but Tris–HCl, pH was reduced to 7.5, and NaCl concentration increased to 100 mM NaCl. Glycerol and 2-mercaptoethanol concentrations remained the same in all the buffers. All buffers were freshly prepared and chilled to 4°C prior use to avoid possible protein denaturation. Phenyl methylsulfonyl (PMSF) to a final concentration of 1mM was added to each buffer aliquot before use.

2.3.3. Expression and purification of reverse transcriptase

2.3.3.1. Expression in *E. coli*

Vectors bearing p51 and p66 subunits (p6H-RT51 and pET-RT66) of RT were co-transformed in BL21(DE3) *E. coli* competent cells (Merck, New Jersey, USA). The two plasmids, p6H-RT51 and pET-RT66 contained ampicillin and kanamycin resistant genes respectively. The vectors use a T7 promoter regulated expression system, controlled by IPTG (Isopropyl-β-D-1-thiogalactopyranoside) induction. A primary culture was prepared by inoculating the transformants in ampicillin (100 µg) and kanamycin (30 µg) enriched media (10 g Tryptone, 5 g Yeast extract, and 5 g NaCl in 500 ml distilled water). The media constituents and antibiotics concentrations remained the same through the experiment unless indicated. The overnight culture was diluted 50 times in fresh LB broth supplemented with ampicillin and kanamycin and allowed to grow at 37°C until OD_{600} of 0.5 was reached. Expression of the protein under lac UV

promoter regulation in *E. coli* cells was completely suppressed and activated with 1mM IPTG (Merck, New Jersey, USA). The culture was allowed to grow for 3 hours at 37°C with 220 rpm agitation. Induced bacteria were harvested by centrifugation at 5000 xG (Beckman coulter, Atlanta, USA) for 30 minutes at 4°C, resuspended in Buffer A and then stored at -80°C prior protein purification.

2.3.3.2. Purification of the heterodimeric HIV-1C reverse transcriptase

Reverse transcriptase induced bacterial cells were lysed by homogenization using a Bandelin sonicator (Bandelin, Anschrift, Berlin). To reduce viscosity, 1µg/ml DNAse was added to the solution and stirred on ice , followed by 10 bursts, 30 seconds' sonication cycles at 80 Hz, with 30 seconds' pulse. The sonicate was centrifuged at 5000 xG for 30 minutes at 4°C. The cleared supernatant containing the heterodimeric p51/66 RT was then bound to Nickel-Nitrilotriacetic acid-agarose (Ni-NTA) beads overnight at 4°C. Thereafter, the bound protein was loaded into a silica gel filter fitted gravity column (Bio-Rad, California, USA) and washed with 500 ml buffer. The protein was eluted with 10 ml buffer B (containing 500 mM Sodium chloride). To refold the protein, eluent fractions were pooled and dialyzed overnight in a snakeskin dialysis tubing (10 kDa cut off (Separations, Rodepoort, South Africa) at 4°C in 5L of buffer C. This was followed by an overnight buffer exchange dialysis of the protein against 5 L of storage buffer D at 4°C. The purified reverse transcriptase was resolved on a 12% Tricine polyacrylamide gel (Figure 2.4.1B). The protein was then stored at -80°C in aliquots for further characterization.

2.3.4. Structural and functional characterization of HIV-1C reverse transcriptase

2.3.4.1. Activity of reverse transcriptase determined by colorimetric assay

To evaluate if the expressed reverse transcriptase was catalytically active, a reverse transcriptase, colorimetric assay kit (Roche, Basel, Switzerland) was used following the manufacturer's instruction. For test experiments, the test reverse transcriptase (RTT) was used instead of the commercial reverse transcriptase (RTC) enclosed within the kit, which was then used as a positive control. The kit uses a principle of reverse transcription of an enclosed template RNA that comes together with a primer hybrid of Poly A and Oligo dT as well as biotin and digoxigenin labelled nucleotides. The kit is supplied with a test microplate coated with streptavidin (which has affinity for biotin), an anti-digoxigenin antibody conjugated to

peroxidase and the peroxidase substrate ABTS. Two nanograms of RTT was assayed against 2 ng of the commercial RTC. Into the streptavidin coated plate, RNA template together with the primer hybrid poly A and Oligo dT were added followed by the addition of RT (both and test separately). After a 1-hour incubation, an anti-digoxigenin antibody conjugated to peroxidase was added and allowed to bind to digoxigenin. Peroxidase substrate, ABTS was added into the reaction and reverse transcriptase activity was measured as the absorbance reading from the colored enzyme-substrate catalysis reaction product released when measured by the Versa Max microplate reader (Molecular devices, San Jose, USA; Figure 2.4.2). The reaction was ran in duplicates and two independent determinations were carried out

2.3.4.2. Reverse transcription of HeLa RNA by the expressed HIV-1C reverse transcriptase

The Invitrogen superscript IV (SS IV) RT-PCR kit protocol (Thermofischer Scientific, Massachusetts, USA) was adopted for the synthesis of a cDNA from a template HeLa RNA using the expressed reverse transcriptase instead of the superscript reverse transcriptase as well as 4 µl of the in-house prepared reverse transcriptase buffer (50 mM Tris-HCl, 4 mM $Mgcl_2$, 50 mM KCl) instead of the SSIV buffer. The reaction was initiated by addition of HeLa RNA at 5 µl, Oligo dT primer and DNTps both at 1 µl in a 200 µl PCR tube. With the exception of reverse transcriptase, primers were allowed to anneal at 65°C for 5 minutes. The reaction was stopped by incubating the mixture on ice for 1 minute. The enzyme was then added together with the in-house reverse transcriptase buffer at 4 µl, DTT (100 mM) and RNase inhibitor both at 1 µl. At a final reaction volume of 13 µl, the mixture was incubated for 1 hour at 37°C for reverse transcription to take place. Upon completion of the reaction, the enzyme was inactivated by raising the incubation temperature to 80°C for 10 minutes and RNA degraded by adding 1 µl RNase A with the temperature reduced to 37°C for 20 minutes. The same experiment was repeated using SS IV, Superscript III (SS III) and *Avian myeloblastosis* (AMV) reverse transcriptases' as controls. Resulting amplicons were presented (Figure 2.4.3).

The resulting cDNA was used as template for a normal DNA polymerase reaction containing the following reagents: 2.5 µl in-house reverse transcriptase buffer, 0.25 µl kit enclosed sense and anti-sense HeLa primers, 0.25 µl of both DNTP mix and $Mgcl_2$, 0.1 µl Taq polymerase, 5 µl template DNA and water to bring the reaction volume to 25 µl. Cycling conditions were set following the manufacturer's instructions and PCR carried out in a Bio-Rad thermal cycler (Bio-

Rad, California, USA). To ensure reproducibility, an independent run of reverse transcription PCR was conducted using only the expressed reverse transcriptase (Appendix B)

2.3.4.3. Far-UV circular dichroism spectroscopy

Circular dichroism for determining α; β constituents of the purified reverse transcriptase was determined using Chira-scan CD-spec at the Center for Scientific and Industrial Research, Pretoria, South Africa (CSIR). Reverse transcriptase was diluted to a final concentration of 10μM. Measurements were performed using a 1mm path-length cuvette at the wavelength range of 180-250nm at 293K at 1nm resolution. Baseline reading was taken using the enzyme storage buffer and the machine was set to auto-subtract the baseline from the final absorbance readings. Three readings were taken for each measurement, averaged and smoothed using the Chira-scan pro-data viewer and then converted to excel using APL data converter. To normalize the obtained data, mean residue ellipticity (MRE) was determined from obtained ellipticity using the following equation: MRE = MRWθ/ 10d.c where MRW is the mean residue weight, θ is the observed ellipticity (degrees), d is cuvette path-length and c is the protein concentration (g/ml). The MRW was obtained by dividing the protein molecular mass by N-1 where N is the number of amino acid residues. Secondary structure of the protein was illustrated by plotting the wavelength against the MRE.

2.3.4.4. Fluorescence spectroscopy

Fluorescence spectra of reverse transcriptase was determined using a PC interface connected Synergy Mx Spectrofluorometer (Biotek, Vermont, USA) and the GenS software was used for result analysis. Tryptophan emission of reverse transcriptase was measured from 300-700 nm where the excitation wavelength was set at 280 nm. Fluorescence intensities of reverse transcriptase was determined in both native and unfolded forms. For native experiments, reverse transcriptase was diluted to 13 μM in PBS and fluorescence was measured shortly after that. To perform unfolded experiments of reverse transcriptase, Urea was included in PBS to a final concentration of 8 M while maintaining the concentration of reverse transcriptase at 13 μM. Fluorescence emission of the unfolded reverse transcriptase was read following a 1-hour incubation in 8 M Urea-PBS buffer at room temperature. The obtained data was analyzed using Gen5 data analysis software (Biotek, Vermont, USA).

2.3.4.5. Size exclusion chromatography

The size exclusion chromatography (SEC) analyses of reverse transcriptase were conducted using the Shimadzu LC-20AB chromatograph instrument (Shimadzu, Kyoto, Japan). Absorbance was measured at 205nm and separations were performed at 27°C at the assay run time of 30 min with a 1.3 ml/min flow rate. The mobile phase constituted 50% Phosphate buffered saline (PBS) at pH 4.5 with a sample injection volume of 10 µl and reverse transcriptase concentration of 6.5 µM. Retention factor average (Kav) was calculated and data presented in Table 2.4.6. The formula used thereof was as follows: Kav = (Ve - Vo)/ (Vc – Vo), where Ve is the elution volume; Vc, geometric volume of the column; Vo, void volume. Protein standards of Aprotinin (10 mg), Ribonuclease A (50 mg), Ovalbumin (50 mg) and Aldolase (950 mg) GE Healthcare, Björkgatan 30, Sweden) ran to determine protease molecular weight from the SEC and the data generated is represented on Appendix A1 and calculations thereof on Appendix A2. To calculate the molecular weight (Mr), a standard curve was plotted (Appendix A3) to generate the equation where x would be the molecular weight and y being the protein Kav. The standards were provided with known molecular weights, presented on Appendix A2.

2.4. RESULTS

2.4.1. Expression and purification of South African HIV-1C reverse transcriptase isolate

Distinctive 66 and 51 kDa bands appearing on a 12% Bis-Tris polyacrylamide gel (lanes 2 to 4; Figure 2.4.1A) are indicative of successful expression of p51/66 subunits of reverse transcriptase over a 3-hour induction period. The presence of faint bands of the two heterodimers before induction (lane 1; Figure 2.4.1A) indicates leakage of the T7 polymerase prior IPTG induction. This is normally prevented by the presence of a T7 lysozyme gene, a phenomenon that was lacking in the vectors used in this regard. Two distinct bands with molecular weights of 66 and 51 kDa were indicative of purified reverse transcriptase subunits, p66 and p51 by nickel affinity gradient column chromatography (lane 1-3 of Figure 2.4.1B). The bands resulted from plasmids pDMI51 and p6HRT66 harboring p51 and p66 subunits of reverse transcriptase respectively, both with N-terminally attached histidine tags that facilitated purification with affinity column chromatography.

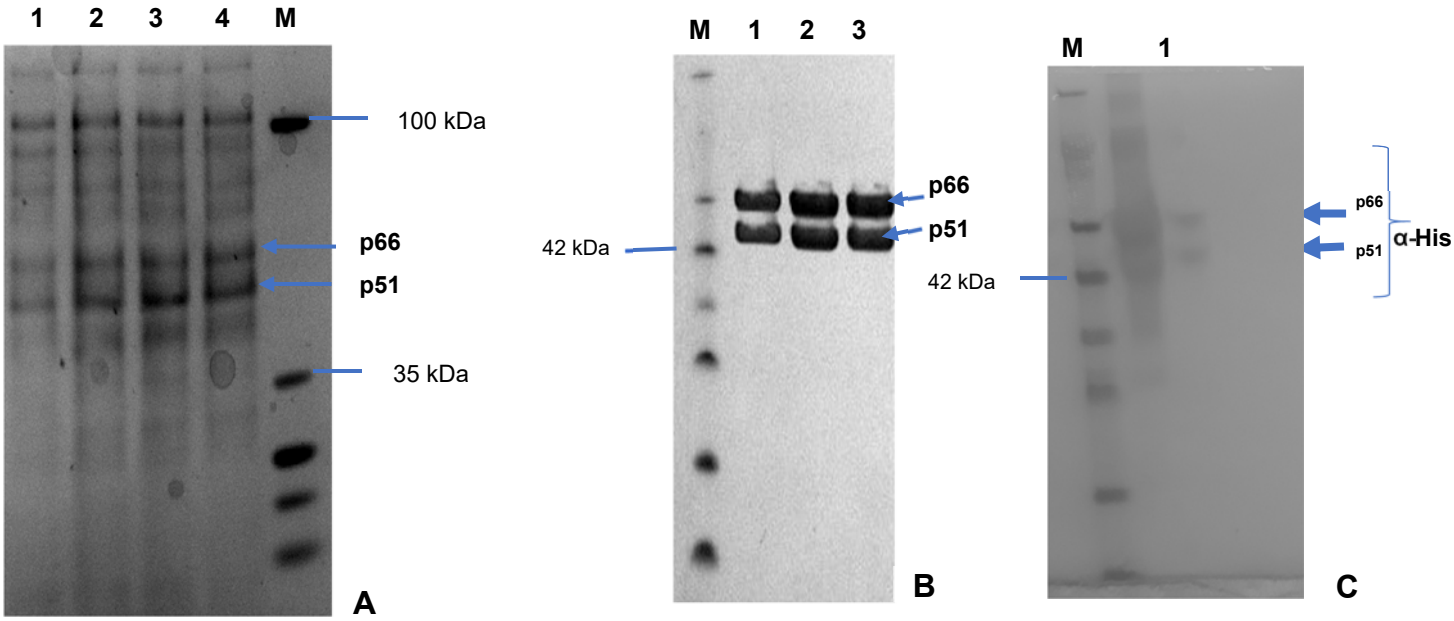


Figure 2.4.1: Representative 12% Tricine polyacrylamide gels. A: time course expression of reverse transcriptase showing the presence of p51 and p66 subunits over a 3-hour induction with IPTG (lanes 1-4, representing 0-3 hours respectively). Leakage was noted before induction (lane 1. B: Distinct bands of p66 and p51 are seen after purification of the protein with nickel affinity column that has affinity to histidine. C: Western blot showing successful transfer of the two bands (Lane 1. C). Confirming expression of reverse transcriptase.

2.4.2. Activity of the expressed HIV-1C RT determined by reverse transcription PCR

The catalytic efficiency of the expressed HIV-1C RT was determined by a reverse transcription polymerase chain reaction. At 15 ng the expressed enzyme was optimally active at 32°C upon amplifying an 80bp HeLa gene (Figure 2.4.2). The amplicon from the reverse transcript (lane 4; Figure 2.4.2) of the expressed enzyme indicated the competitiveness of the expressed reverse transcriptase when compared to positive controls of HeLa DNA amplified from a SS IV, SS III, and AMV reverse transcriptase run and comparable results were obtained. (Lanes 1-3; Figure 2.4.2, respectively). To ensure reproducibility of the RT-PCR outcomes the experiment was repeated in three independent runs (Appendix B)

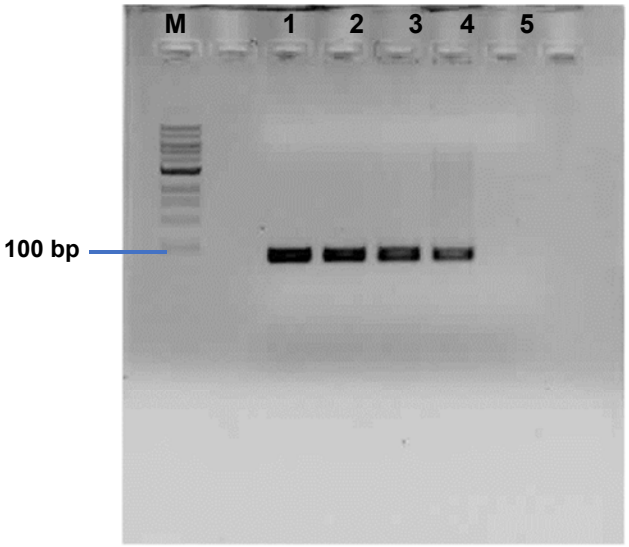


Figure 2.4.2: A 1% agarose gel electrophoresis representing PCR products amplified from commercial reverse transcriptases Superscript IV, Superscript III, AMV reverse transcriptases (lanes 1, 2, 3 respectively) and the expressed reverse transcriptase (lane 4) transcripts of HeLa RNA. Lane 5 represents a negative control where water was used instead of the enzyme.

2.4.3. Far UV- circular dichroism (CD) spectroscopy of reverse transcriptase

In determining the secondary conformation of reverse transcriptase, a far-UV CD spectrum of the enzyme was determined. The wavelength ranges at which the ellipticity was mostly minimum for the reverse transcriptase, spectra were split into two negative transitions at approximately 208 and 220 nm representing β-sheets and α-helices respectively (Figure 2.4.3). the α-helices content was seen to be slightly higher than that of β-sheets, a known characteristic of HIV reverse transcriptase (Jacobo-Molina *et al.*, 1993).

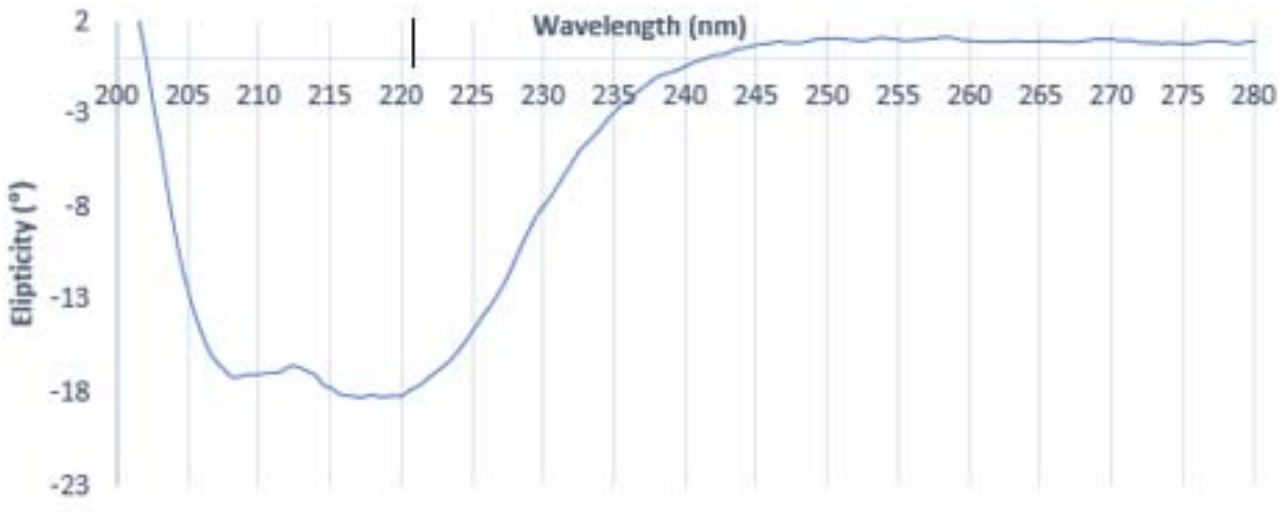


Figure 2.4.3: A far UV-CD spectrum of reverse transcriptase. The spectra indicate a high α helical content with negative transition splitting into two transitions, ~208 and 210 nm respectively. Data is presented in units of molar ellipticity per residue. Spectra were recorded at pH 7.0.

2.4.4. Fluorescence spectrometry of reverse transcriptase

For tertiary structure determination, a fluorescence spectroscopy based on Tryptophan residues fluorescence was conducted. Reverse transcriptase contains 37 Tryptophan residues (18 in p51 and 19 in p66). To determine the conformation of reverse transcriptase, Urea was used to assume the denatured configuration distinguished from the native form. An intrinsic reduction of emission maxima with a shift from approximately 338 to 340 nm of the native to 8M Urea unfolded reverse transcriptase respectively was obtained (Figure 2.4.4). A decrease in fluorescence emission with a shift in absorption is indicative of Urea denaturation of reverse transcriptase. When unfolded, the protein exposes its tryptophan residues that normally buried in the protein's native conformation (Divita, Restle and Goody, 1993). As a result, the quantum yield of the unfolded protein decreases leading to a decrease in fluorescence yield. This observation therefore confirms the expressed reverse transcriptase to be in its native conformation.

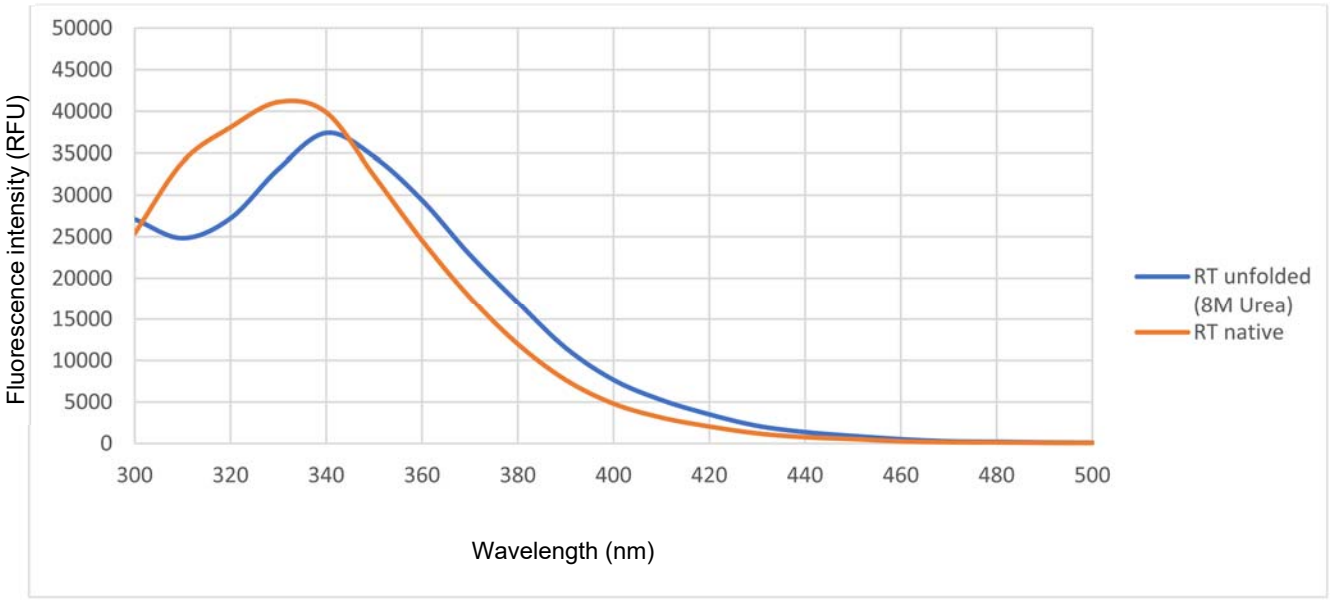


Figure 2.4.4: An illustration of the fluorescence spectrophotometry emission spectra profiles of a native (orange) and an 8M Urea unfolded (blue) reverse transcriptase. A decrease in fluorescence intensity with a shift in absorption from 338 to 340 nm from the native to the unfolded form of reverse transcriptase respectively, an indication of tryptophan fluorescence changes in the enzyme's transition from its native dimeric form to an unfolded state.

2.4.5. Size exclusion chromatography

Standard SEC peaks were indicative of the times where a protein of a particular weight was expected to elute (Appendix A1). The two peaks representing either one of the two reverse transcriptase monomers (p51 and p66) eluting at 15 minutes and the heterodimeric form at 24 minutes (Figure 2.4.5). The first peak further indicates the strong interactions between the two subunits (p51 and p66) as the peak is a bit extended allowing the two subunits to elute one after the other but at almost the same time.

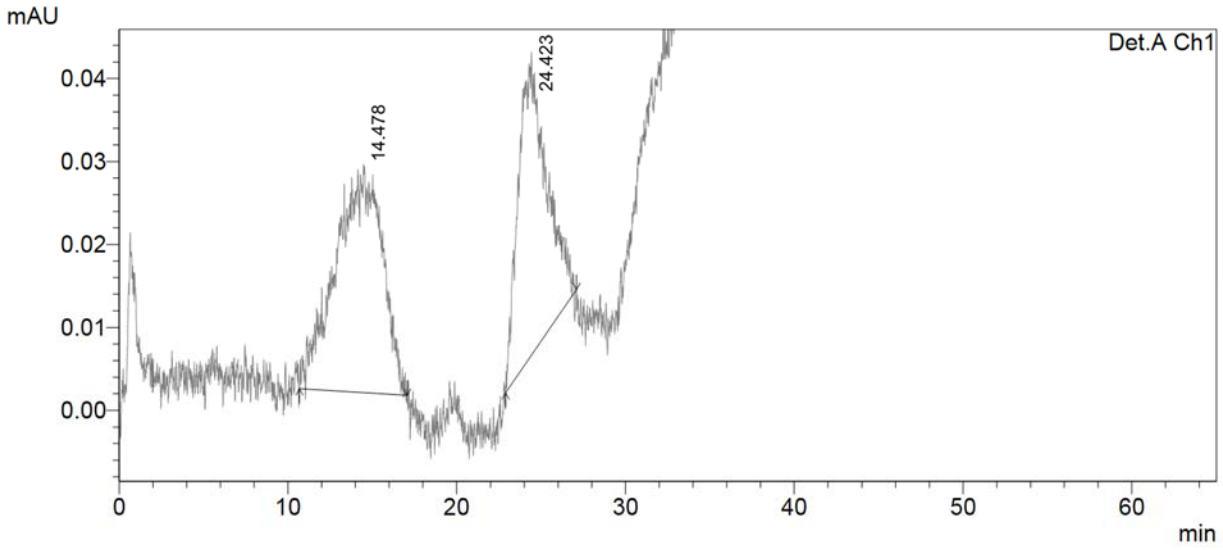


Figure 2.4.5: An HIV reverse transcriptase size exclusion chromatogram, illustrating elution of either one of the two monomers (p51 and p66) and heterodimeric form at 15 and 24 minutes

Table 2.4.5: Reverse transcriptase retention factor and apparent molecular weights

Peaks	Time (min)	Ve (ml)	Kav	Mr (Da)
Peak 1	14.478	7	0.3507	100123.3
Peak 2	24.423	4	0.009492	479444.1

Formula used: Kav = (Ve - Vo)/ (Vc – Vo), where Ve is the sample elution volume; Vc and Vo, geometric and void volumes of the column respectively (all in ml). Where Vo = 4.2477; Vc =

14.33. Molecular weight equation, x = (2.1707- y)/ (0.6122), Where x is the molecular weight and y being the protein Kav.

From the molecular weight obtained for the first reverse transcriptase, the peak appeared within the void volume, and it is likely the unfolded reverse transcriptase in its monomeric form. For the second peak suspected to be the heterodimeric form of reverse transcriptase, there was about a 10 kDa difference. The slightly smaller calculated Mr (Table 2.4.5) may be because of strong molecular interactions between subunits, and due to the tighter arrangement, the heterodimeric reverse transcriptase elutes later than expected.

2.5. DISCUSSION AND CONCLUSION

The current study demonstrated the expression of recombinant HIV-1C reverse transcriptase protein isolate from South Africa. The protein was successfully expressed in BL21 (DE3) *E. coli* cells (Figure 2.4.1A). The protein which was produced as a polyhistidine tagged species was successfully purified by nickel chromatography (Figure 2.4.1B). There is need to conduct structure function analyses of this medically important protein. To this end biochemical analyses to determine the protein functionality and the protein conformation were conducted.

To secondary structure of the expressed HIV-1C RT as shown by Far-UV CD spectrometry the protein to be constituted by both Beta-sheets and alpha-helices. However, the alpha-helix content was seen to be higher than that of beta-sheets, confirming reverse transcriptase to be mostly alpha-helical as previously described (Jacobo-Molina *et al*., 1993). This was shown by the appearance of two negative transitions at 208 and 220 nm (Figure 2.4.3). A decrease in fluorescence intensity of tryptophan with an emission maxima shift from 338 to 340 nm was observed upon denaturation of the expressed HIV-1C RT. Reverse transcriptase has been documented to undergo fluorescence changes in association with dimerization. This is attributed to the repeat motif of Tryptophan residues buried in the dimer connection subdomain residues ranging from positions 398 to 414 (Braz *et al*., 2010; Divita *et al*., 1993). Dissociation of reverse transcriptase dimer should therefore expose Tryptophan residues, leading to reduction in the tryptophan quantum yield which then results in a decreased fluorescence intensity as observed in the current work. This served as confirmation of the compact status of reverse transcriptase tertiary structure. Though reverse transcriptase is known to exist in a dimeric form, the p51 monomer in reverse transcription only serves as a scaffold. The p66 monomer has been shown to be capable of executing reverse transcription independent of p51 (Ciuffi *et al.*, 2009).

Catalytic activity of the expressed reverse transcriptase was validated through successful synthesis of a HeLa transcript that amplified to an 80bp gene fragment upon PCR analysis. The reverse transcription activity of the test enzyme was competitive when compared to commercial enzymes SS II, SS IV and AMV reverse transcriptase's. Though HIV reverse transcriptase has been shown to be highly error prone when compared to AMV reverse

transcriptase (1/17000 error rate per detectable nucleotide as compared to 1/30 000 respectively (Roberts, Bebenek and Kunkel, 1988), the main focus of this experiment was only to determine its catalytic activity for possible application in downstream development of inhibitory assays, which was successfully accomplished.

Reverse transcriptase produced in this work was proven to be both structurally intact and catalytically active. The protein can therefore be routinely produced in substantial amounts for applications such as anti-HIV-1C drug screening assay development. The activity evaluation assay used in this work can also be employed to screen for molecules with inhibitory activity against HIV-1 C reverse transcription. Successful inhibition would be recorded by reduced absorbance in the case of the colorimetric assay and the absence of an 80bp amplicon upon PCR analysis of HeLa transcript. The protocol developed herein can be applied for the analyses of homologues of reverse transcriptase as well as production of mutant forms of the proteins. This could set up a platform for studies of HIV drug resistance mechanisms, subsequently leading to the development of assays to screen for inhibitory molecules with an increased drug resistance barrier.

CHAPTER THREE: MOLECULAR DOCKING AND *IN VITRO* BINDING AFFINITIES OF HUMAN PEPSIN INHIBITOR HOMOLOG (BM-33) WITH HIV-1 PROTEASE

ABSTRACT

The use of molecular docking is crucial in facilitating drug discovery. Due to its proven ability to virtually screen and single out molecules with assumed potency, molecular docking is recently becoming a gold standard pre-screening tool in the pharmaceutical field. However, molecular docking results can only be validated by experimental data. To set up an in-silico modelling platform to virtually screen for molecules that can be used in downstream *in vitro* assays, a molecular docking approach was used. An HIV-1 protease (HIV-1 PR) sequence was used with a *Brugia malayi* pepsin inhibitor homolog (Bm-33) that has been previously shown to have inhibitory characteristics against an aspartic, human pepsin as an assumed inhibitor. The crystal structures of the two proteins were constructed using the online tool PHYRE and then subjected to molecular docking using AutoDock Vina. The resulting docking clusters were analyzed using Autodock tools and a cluster that yielded the lowest binding energy was selected for further analysis. Upon docking results analysis, up to 15 interactions of Bm-33 with protease were identified. Comparable interactions were also identified with a control (human pepsin) where up to 19 interactions were observed. Molecular docking results yielded good assumed binding affinities of protease and Bm-33 and this prompted *in vitro* interaction analysis of the molecules. As a result, four peptides were synthesized from 19 interactions observed from residues interacting with protease. Upon *in vitro* analyses of the four peptides with HIV-1 PR, dissociation constants (K_d) ranging from 0.2 to 0.002 were confirmatory to binding affinities observed upon docking experiments. This validated the interaction of HIV-1 PR and Bm-33, qualifying the use of molecular docking as a pre-screening tool of molecules to be used in protein inhibitory assays.

3.1. STUDY OBJECTIVES

3.1.1. Aim

To explore binding interactions of HIV-1 PR and a Brugia malayi pepsin inhibitor homolog (Bm33) for possible inhibition of protease.

3.1.2. Objectives

- To determine binding interactions of HIV-1 PR and Bm-33 using molecular docking
- To determine binding efficiency of HIV-1 protease and Bm-33 in vitro

3.2. INTRODUCTION

3.2.1. Background and rationale

Escalating genetic divergence and drug resistance is lately becoming a setback in antiviral therapeutics. Toxicity and resistance by HIV to antiretroviral drugs remains problematic. Continual screening and modification of new potent drugs is eminent. *In vitro* screening of molecules with inhibitory activities have been the most useful tool in drug discovery. However, screening for molecules with just suspected potency without any form of first-line confirmation could result in unsuccessful trials and escalated cost, a serious setback in drug discovery. The use of molecular docking plays an important role in drug discovery, due to its proven ability to virtually screen and single out only a few molecules with assumed potency from a pool of a dozen others that can be taken further for *in vitro* assays and characterization. Several docking software packages have been developed and many improvements have been employed over the years. This has given rise to recent and improved versions of docking software packages that yield better, most accurate and cost effective first line molecule screening platforms such as Autodock Vina (Trott and Olson, 2010; Handoko *et al.*, 2012), an added advantage for continual molecule screening in the high viral diversity and drug resistance overwhelmed research fields such of HIV-1.

An Aspin, identified as a pepsin inhibitor homolog (Bm-33) from *Brugia malayi* was shown to have inhibitory characteristics against 4 important human aspartic proteases [Pepsin, Renin, Cathepsin-D and Cathepsin-E), (Krishna *et al.*, 2013). Since HIV-1 protease and pepsin share the same active site binding mechanism, exploring protease interaction with the newly identified inhibitor of pepsin (Bm-33) would be of great interest. *Brugia malayi* pepsin inhibitor homolog-33, a 33 kDa Aspin (parasitic produced aspartic protease inhibitor) comprising 153 amino acids, is produced by *Brugia malayi*, a filarial parasitic nematode as its defense mechanism in the human host. Together with other protease inhibitors (Serpins; Serine protease inhibitors and Cystatins; Cysteine protease inhibitors), Aspins are believed to play a role in human immune response modulation during the lymph-dwelling parasite's long-life cycle within the host (Martenz, 1985). Size

exclusion chromatography revealed over 80% pepsin inhibition by the 33 kDa refolded Bm-33, expressed in *E. coli* (Krishna *et al.*, 2011). The mostly α-helical Bm-33, peptidic in nature could open new grounds for natural products research in HIV-1 protease inhibition which could lead to the production of protease inhibitors with reduced toxicity. Further exploration of HIV-1 protease interaction with known and newly identified pepsin inhibitors would be an added advantage in HIV-1 therapeutics.

Owing to structural homology of human pepsin and HIV protease, cross-interaction of HIV protease with pepsin inhibitors have already been documented in several studies. Acetyl-pepstatin, a small molecular weight hexa-peptide comprising a unique amino acid statin, have been shown to exhibit marked inhibitory activity against HIV-1 protease (Sayer and Louis, 2009; Matúz *et al.*, 2012). Although inhibition of HIV-1 protease is the main focus, pepsin level monitoring in patients would be of immense importance when protease-pepsin interacting inhibitors are used to avoid conditions such as vitamin B-12 deficiency resulting from amide bonds cleavage failure of ingested foods due to insufficient gastric pepsin in humans. Reduced levels of pepsin on the other hand can however have a positive impact in reduction of inflammation of the laryngeal mucosa caused by high accumulation of pepsin in the event of laryngopharyngeal reflux.

Protease has been shown to be the most targeted in HIV therapeutics and subtype C shown to bear a high proportion of polymorphisms. This therefore led to consideration of HIV-1C protease as an experimental subject for *in silico* modelling experiments in the pursuit of a pre-screening tool that can be applied in the downstream drug screening assay development for inhibition of the globally predominant HIV-1 C. To achieve this, molecular docking of HIV protease with Bm-33 was conducted. Confirmatory *in vitro* assessments were done to further validate the outcomes of *in silico* modelling.

3.3. MATERIAL AND METHODS

3.3.1. HIV-1 protease and human pepsin homology modelling and sequence alignment

Amino acid sequences of HIV proteases and human pepsin retrieved from National center for biotechnology information (NCBI) sequence databank (https://www.ncbi.nlm.nih.gov/nucleotide/) were submitted to the MAFT MSA tool (http://www.ebi.ac.uk/Tools/msa/mafft/) for multiple sequence alignment. The output was submitted to Boxshade (http://www.ch.embnet.org/software/BOX_form.html) for similarity and identity shading. The aligned sequences with highlighted similarities are shown in figure 3.3.1.

For comparison of HIV-1 protease and human pepsin secondary structures, HIV-1 proteases and human pepsin, were selected as consensus. The amino acid sequences were retrieved from NCBI database (https://www.ncbi.nlm.nih.gov/nucleotide/) and submitted to the Protein Homology/analogY Recognition Engine V 2.0 (PHYRE2),(http://www.sbg.bio.ic.ac.uk/phyre2)) for homology modeling. The crystal structure of Bm-33 was also generated through PHYRE2 and was used together with those of protease and pepsin for molecular docking experiments. Figure 3.4.1 shows models for both protease and pepsin and the predicted secondary structures.

3.3.2. Molecular docking and analyses of Bm-33 with protease and pepsin.

The binding conformations and affinities of Bm-33 with HIV-1 protease as well as Pepsin were explored using molecular docking program, Autodock Vina (Trott and Olson, 2010). Autodock tools was used to analyze the docking conformations. Clusters presenting the lowest binding energies were selected for further analysis. Binding affinities were then determined using LigPlot (Wallace, Laskowski and Thornton, 1995) docking analyzer. Duranavir, Lopinavir, Ritonavir and Acetyl pepstatin bound to protease were used as references (Figure 3.3.2A). Protein databank (PDB) codes of protease bound to the afore mentioned inhibitors were retrieved from the RSCB PDB website (https://www.rcsb.org).

The PDB codes for all structures retrieved from RCSB PDB website were recorded in Table 3.3.2.

The same results analysis algorithms used for protease were applied for Pepsin docking results. Standard inhibitors used herein were acetyl pepstatin and a phosphonate inhibitor both bound to human pepsin. PDB codes of the crystal structures used were also recorded in Table 3.3.2.

Table 3.3.2: Crystal structures of inhibitor bound proteins and their PDB codes

Crystal structure description	**PDB code**
HIV protease bound to duranavir	2b60
HIV protease bound to lopinavir	1mui
HIV protease bound to acetylpepstatin	5hvp
Human pepsin bound to acetylpepstatin	1pso
Human pepsin bound to a phosphonate inhibitor	1qrp

Following molecular docking analyses, ranges of residues interacting with protease were selected for short peptide generation. Synthesis of the peptides was carried out by GenScript (GenScript, New Jersey, USA).

3.3.3. Determination of protease and Bm33 peptides binding affinities

To confirm molecular docking findings, an ELISA assay was conducted to determine the affinity of protease to bind the four peptides generated from protease-Bm33 *in silico* interaction using the microplate solid-phase protein binding assay (Biesiadecki and Jin, 2011). Five micrograms of peptides in buffer A (50mM $NaHCO_3$; pH 9.5) was immobilized onto the wells of a 96-well polystyrene microtitering plate at 100µl/well. Bovine serum albumin (BSA) at the same concentration as peptides was used as a control. The plate

was covered and incubated overnight at 4ºC. The contents of the plate were discarded. The plate was then quickly washed with 150 µl/ well buffer T (50 mM Tris, 150 mM $NaCl_2$, 0.1 % Tween-20, pH 7.5), using a multichannel pipette. Wash buffer was discarded, and the wells blocked with 1% BSA in buffer T for 1 hour at room temperature with gentle agitation. Following blocking, contents of the plate were discarded, and the plate was washed 3 times with 150 µl buffer T. The first wash was carried out as previous washes and the last two washes were accompanied by 3 minutes incubations post buffer addition to ensure thorough removal of residual BSA. To bind protease to the immobilized peptides, 3-fold serial dilutions of the protein were added to the wells at 100 µl /well and incubated for 2 hours at room temperature with gentle agitation. The plate wash was conducted similar to the previous wash. An anti-Histidine antibody was added to detect the 6X Histidine tagged protease at 5 µg 100 µl / well. The reaction was incubated for 40 minutes and then followed by a three times wash. To detect the bound anti-histidine antibody, 100 µl of TMB /well was added and the absorbance was read at 370 nm for 30 minutes. Obtained data was analyzed using GraphPad prism software (Version 7).

3.4. RESULTS

3.4.1. Protease and Pepsin homology modeling

In the 3D models, protease is shown to possess seven *β*-strands (in green) and these seem extended as compared to the eight shorter *β*-strands in pepsin (Figure 3.4.1A and B). Both proteins possess a single *α*-helix (orange) with several loop segments (blue). A similar active site is located in a disordered loop region (yellow). Loops are generally considered to be disordered and flexible to allow substrate entry into the active site. Both proteases possess the structurally significant triad, (DTG, in red), which constitutes the only *α*-helix in these homologues. Although both homologues possess the conserved GIG (magenta), in the HIV-1 PR this region is partly on a *β*-strand and a loop region whereas in the human PR it's mainly on a looped region. Secondary structure differences could be functional variations such as unique interaction with substrate and/or inhibitors.

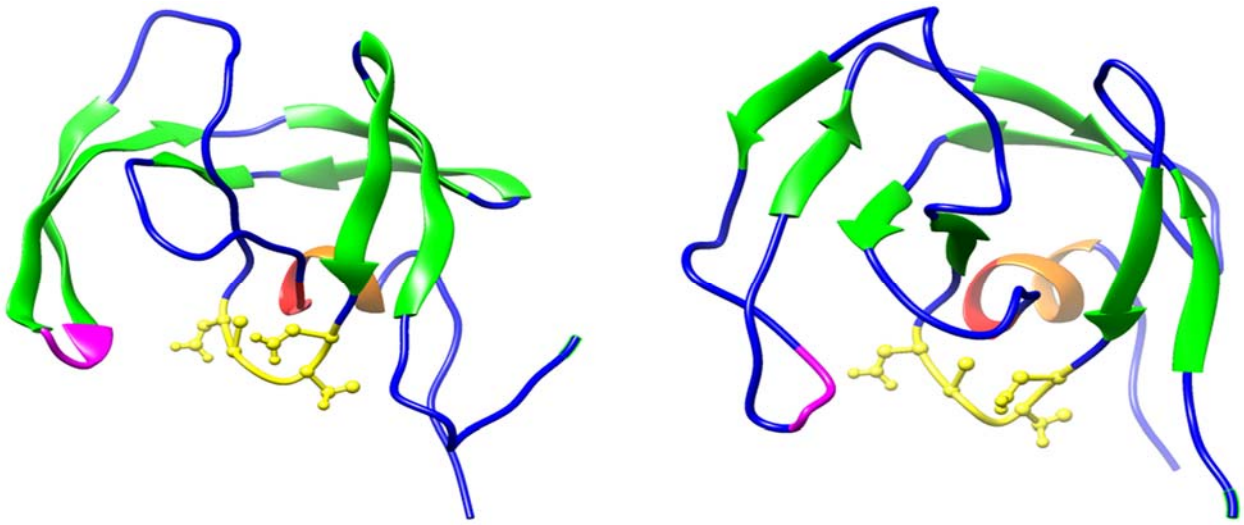


Figure 3.4.1. 3D homology models of HIV-1 PR Subtype C protease monomer (A) and human pepsin (B). Protease amino acid sequences were retrieved and submitted to Phyre2 (www.sbg.bio.ic.ac.uk/phyre2). The models were viewed and manipulated using Chimera version 1.9. Protein regions are highlighted as follows: beta sheets/strands in green, alpha helices in orange and loops in blue. Also highlighted is the active site in yellow, the GR(N/D) triad in red and the conserved GIG in magenta.

3.4.2. Protease and Pepsin sequence alignment

The first crystal structure of HIV-1 protease was obtained in 1989 (Navia *et al.*, 1989). Protease is a 99-amino acid aspartyl protease, falling in the aspartic acid group of proteases. Aspartates possess two aspartic acid residues in their active site. HIV-1 protease is active as a homodimer made up of two 99 residue chains harboring a single active site (Brik and Wong, 2003). The active site, made up of residues 25-29, forms part of the dimer interface with the residues Asp-Thr-Gly (**DTG**) forming a network of hydrogen bonds known as the "fireman's grip" (Mager, 2001). These residues are conserved in aspartic acid proteases including the human protease homologs (Figure 3.4.2). Each of the two monomers is said to contribute one DTG triad towards active site formation.

Also forming part of the interface are four stranded antiparallel *β*-sheets comprising N-terminal residues 1-4 (**PQIT**) and C-terminal residues 96-99 (**TLNF**), (Weber, 1990) and these account for 50% of the dimer interface. According to Ishima *et al.*, 2010, the HIV-1 protease structure is well defined, being made up of seven *β*-strands, one *α*-helix with some disordered terminal and loop segments (Figure 3.4.2). The single *α*-helix, amino acids 84-94, contains the well conserved Gly 86 Arg 87 (N/D) only in retroviral proteases. This triad is thought to have an important structural role and is responsible for connecting the main body of the protein to the last *β*-strand of the terminal *β*-sheet (Weber, 1990). Conservative substitution of the Arg 87 with Lys is said to cause disruption to the formation of dimer (Louis *et al.*, 1990, Ishima *et al.*, 2001) highlighting the vital role of this residue in the dimerization process of the protease. Arg 87 forms intermonomer hydrogen bonds with Asp 29 near the active site and this is crucial for the stability of the dimer (Ferentz and Wagner, 2000). Interestingly substitution of Asp 29 with N or the deletion of terminal residues (N- and C-) was seen to destabilize dimer formation (Ishima *et al.*, 2003). The HIV-1 protease monomer is also known to possess a glycine rich loop referred to as a flap, residues 47-52, **GGIGG**. The human homologs are also shown to possess **VGIG.**

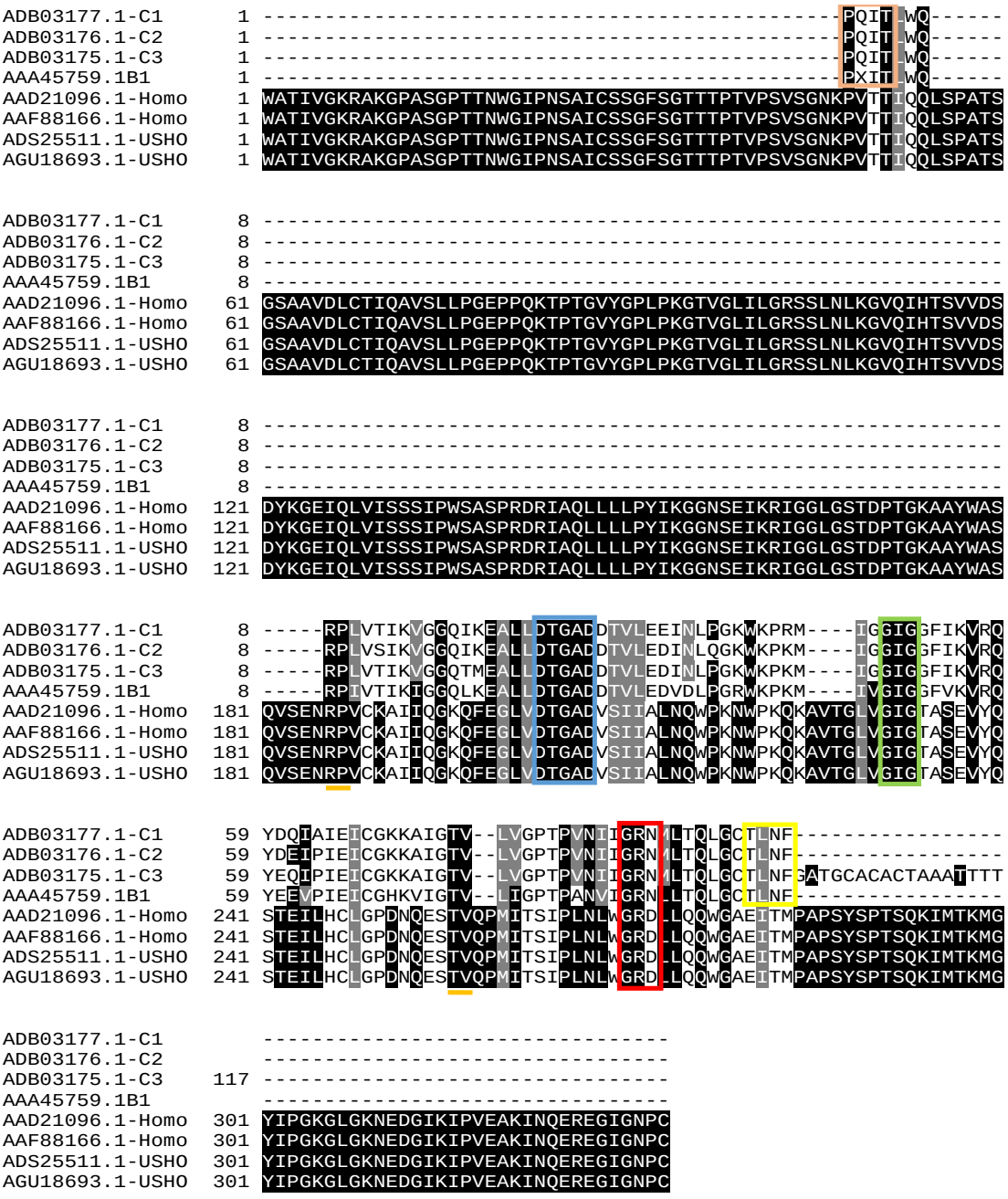
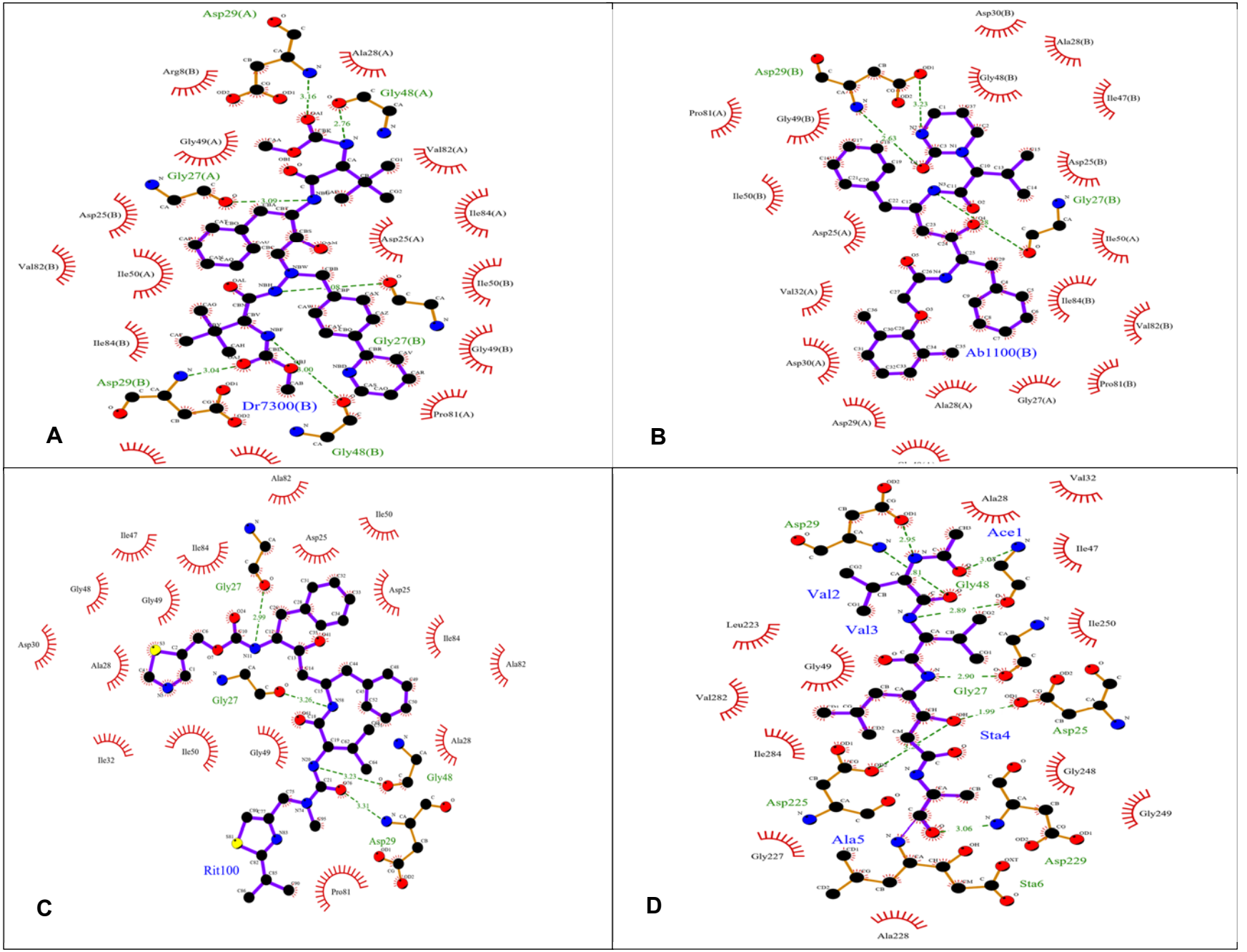


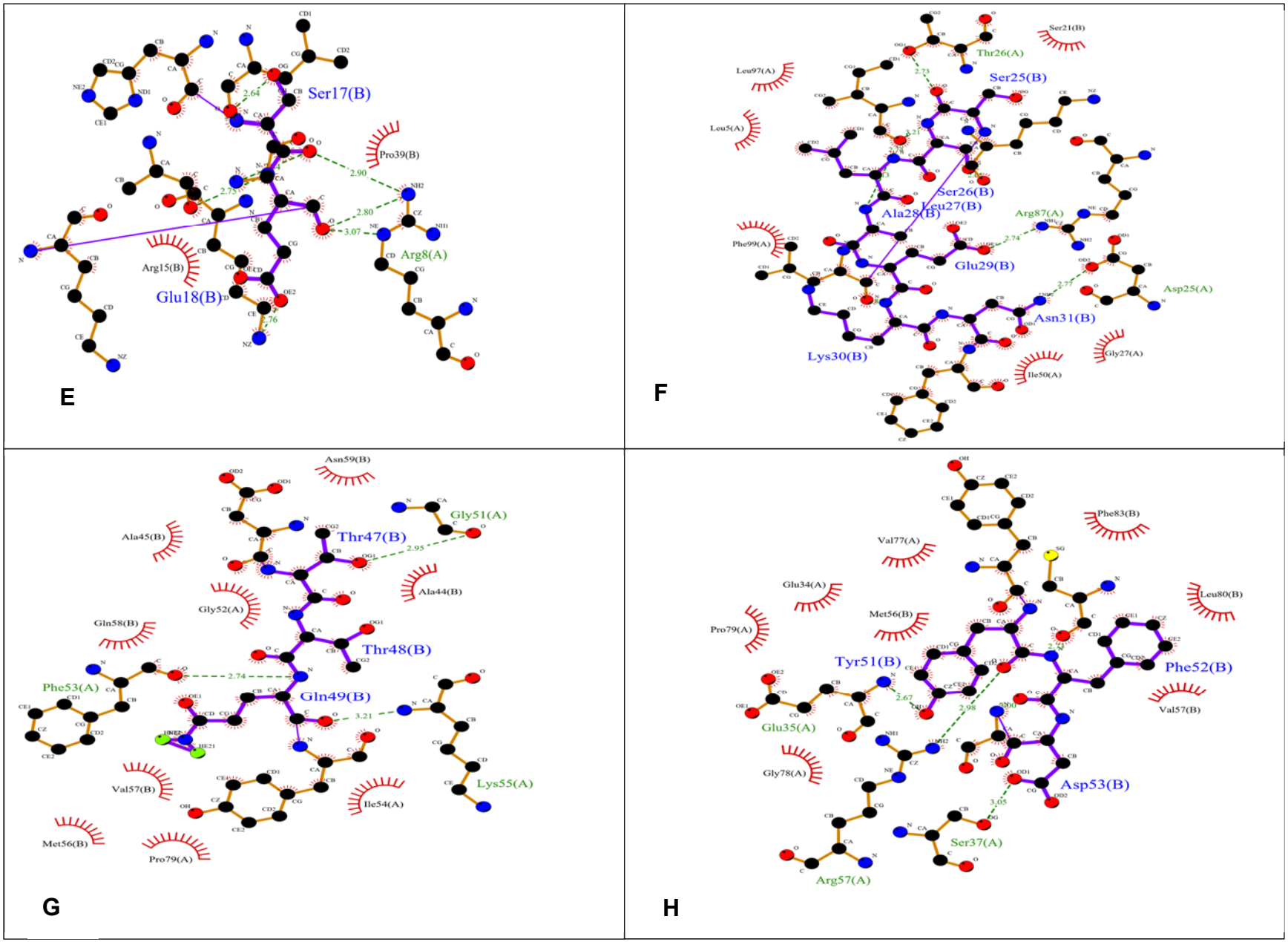
Figure 3.4.2: Multiple sequence alignment of HIV-1 PR and human pepsin homologues. Sequences of HIV-1 PR subtype C homologues (NCBI accession numbers: ADB03177.1, ADB03176.1, ADB03175.1), subtype B (NCBI accession number: AAA45759.1B) and human PR homologues (NCBI accession numbers: AAD21096.1, AAF88166.1, ADS25511.1, AGU18693.1). Terminal residues that form part of the interface are highlighted as follows: N-terminal (orange box) and C-terminal (yellow box). Amino acid residues that are predicted to form 1.the active site (blue box), (Ishima *et al*., 2001) 2. The flap (green box) and 3. The GR (N/D) triad (red box). Residues showing a high degree of conservation across species are indicated with a black background. Residues predicted to be crucial for the fireman's grip are indicated with a gray background (Louise *et al* 2002).

3.4.3. Molecular docking

3.4.3.1. Molecular docking analysis of HIV-1 Protease with Bm-33

Up to 100 conformations were generated for each run during molecular docking. Clusters with the lowest energies upon analysis with Autodock tools were selected for further analysis. When using LigPlot analysis, up to 15 interactions of protease to Bm-33 residues were identified, of which 2 are directed to 2 of the 3 active site residues of protease; Asp 25 and Thr 26 essential for substrate catalysis (Figure 3.4.3.1 E-J). There was also a marked number of hydrophobic interactions of protease residues with some of Bm-33 residues that are expected to contribute in the interaction between the two molecules. Apart from the active site residues, other amino acids involved in the protease - Bm-33 interaction included Arg 8, Glu 21, Glu 31, Ser 37, Gly 51, Phe 53, Lys 55, Arg 57 and Arg 87. Residues of the 33 kDa Bm-33 that are involved in the interaction with protease include Ser 17, Glu 18, Ser 25, Glu 29, Asn 31, Thr 47, Gln 49, Tyr 51, Asp 53, Arg 66, Asp 84 and Met 87.





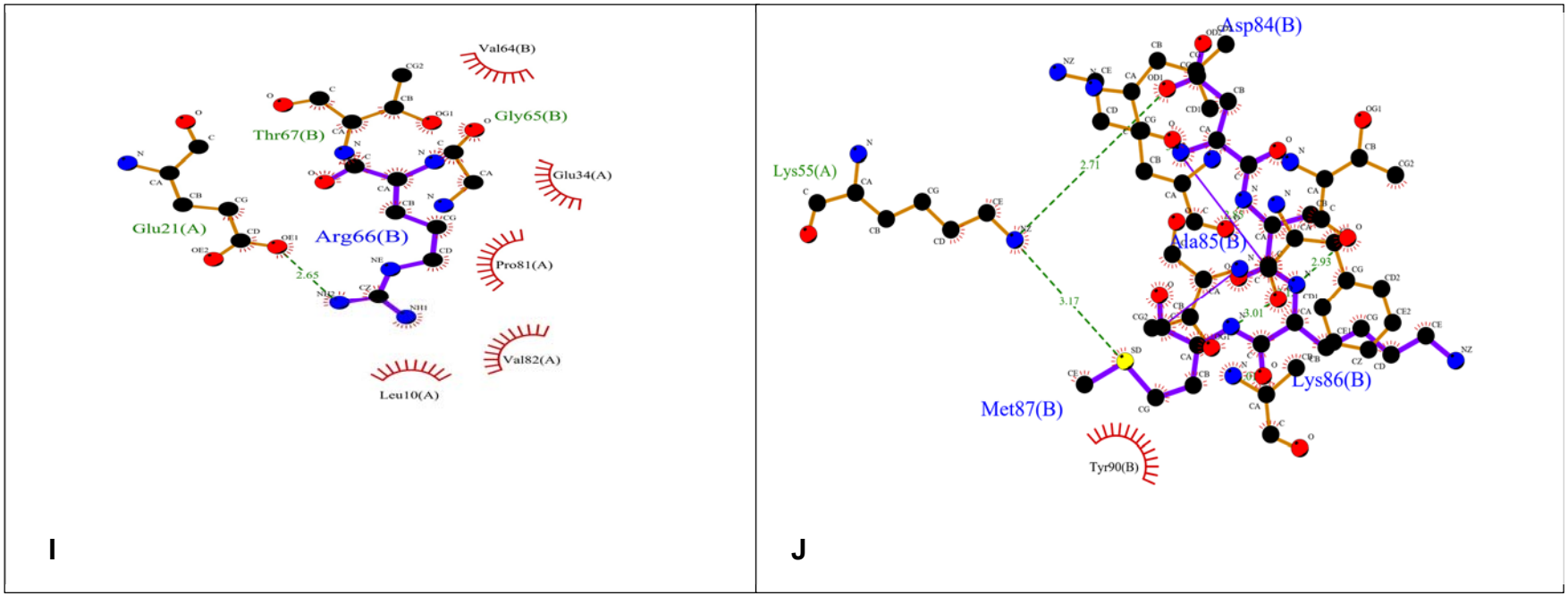
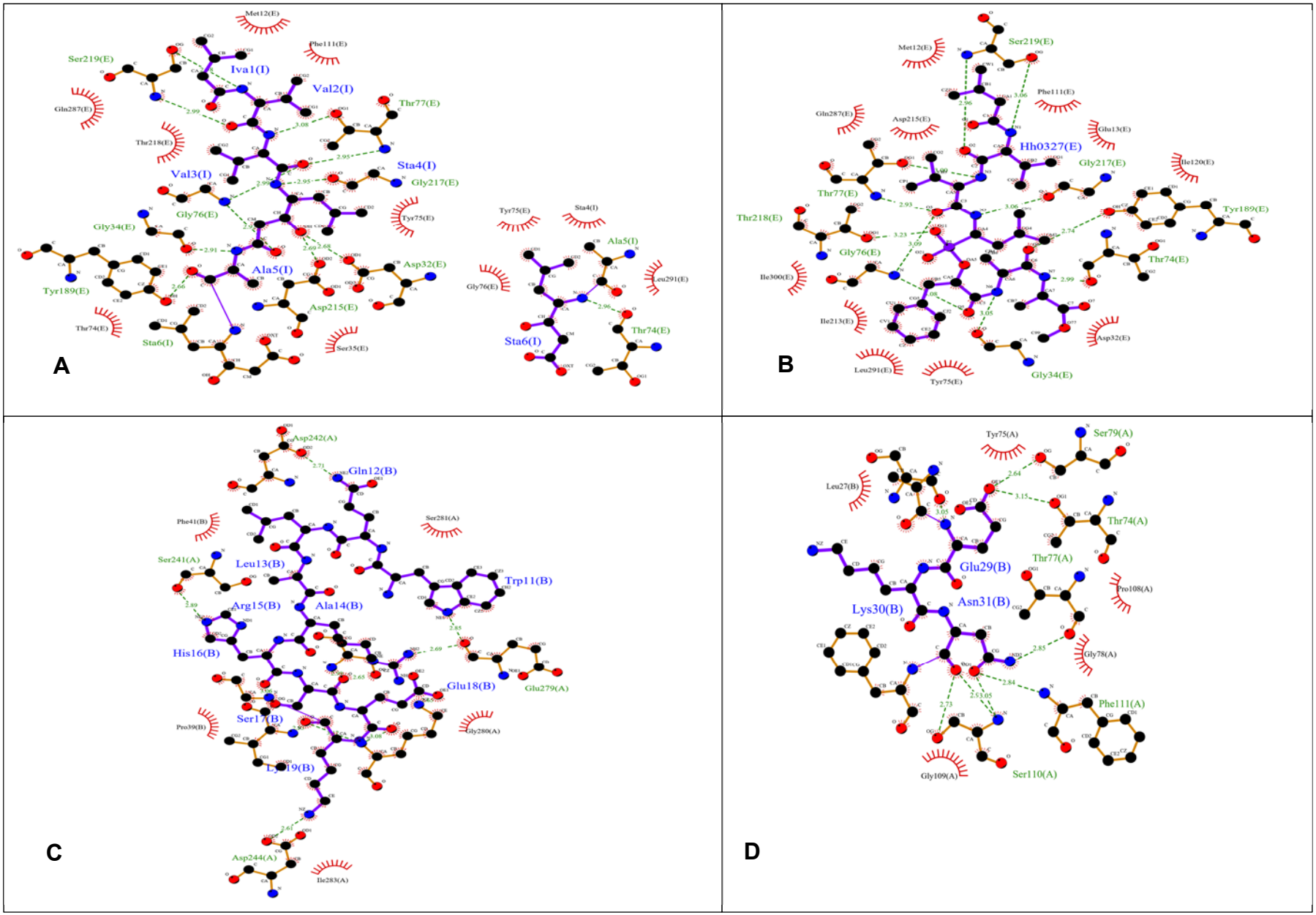


Figure 3.4.3.1: Geometry optimized structures of intermolecular H-bonds of Bm-33 with HIV-1 protease analyzed by LigPlot. Ligands are indicated by co-jointed residues with purple bonds, thin purple bonds are either covalent bonds between protein and ligand or elastic bonds between ligand residues. Protein residues are indicated with brown bonds between atoms. Protease residue names labelled in green and identified as chain A (A) and ligand residue names are in blue and identified as chain B (B). Red atom is Oxygen; Blue, Nitrogen; Black, Carbon; Yellow, Sulphate and Green, Hydrogen. Green dotted lines indicate hydrogen bonds and their distances. Sunrise-like side residues are an indication of hydrophobic interacting amino acids and the spikes indicate where the interaction is directed. Figures A-D are standards of HIV-1 protease interacting with known inhibitors, Duranavir, Lopinavir, Ritonavir and acetyl pepstatin respectively; E-J showing HIV-1 Protease residues interacting with Bm-33 at various positions.

3.4.3.2. Molecular docking analysis of human pepsin with Bm-33

Upon analysis of Pepsin to Bm-33 docking results, up to 19 interactions were revealed (Figure 3.4.3.2C-G). The first two plots in figure 3.4.3.2 are standards [Pepsin with pepstatin(A) and with phosphonate inhibitor (B)] and their interactions are almost similar (showing similarities of up to 6 interactions). However, of all the 19 interactions between Pepsin and Bm-33, Thr77 is the only common residue as in the other two interactions of the standards. The active site residues on Pepsin essential for inhibitor binding are Asp32, Gly76, Thr77 and Asp215. In a study conducted by Krishna *et al* (2013) Pepsin has been shown to interact with BM-33 *in vitro*. This serves to confirm that Bm-33 interacts with residues outside of the active site, inducing conformational changes to the protein active site and thereby deterring the protein activity. Most of the interactions of Bm-33 are seen directed to pepsin residues around position 241 to 244.



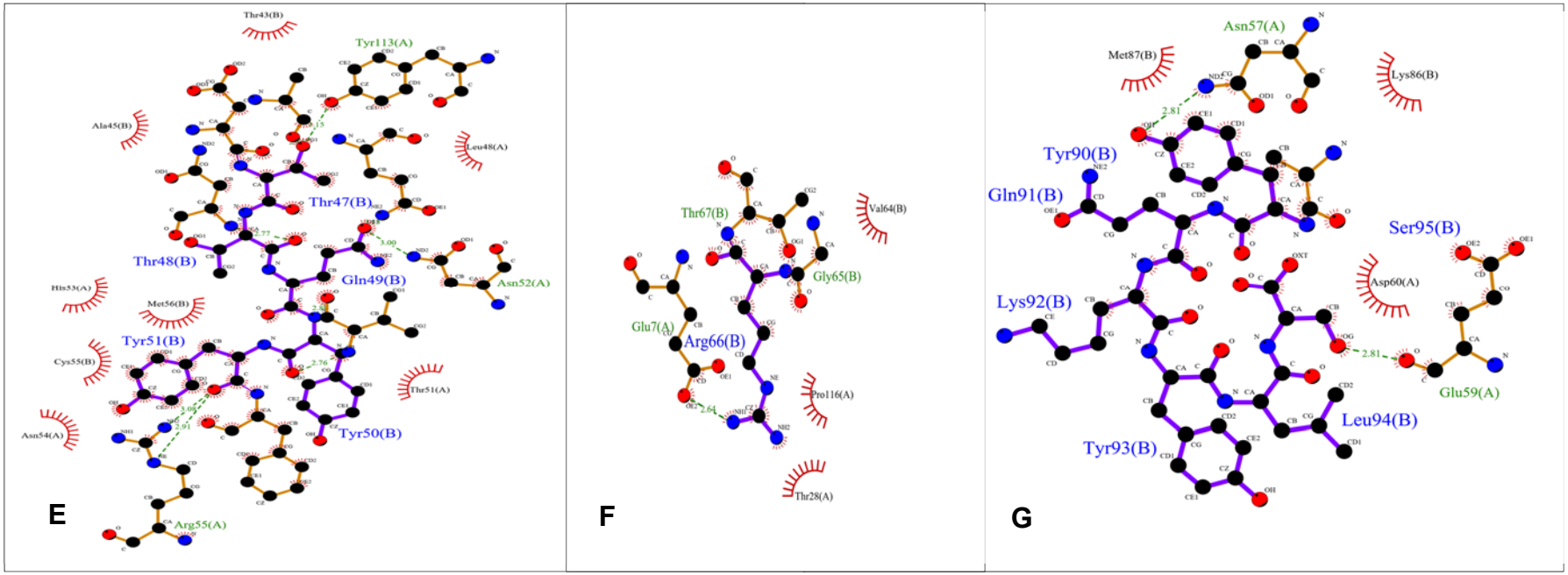


Figure 3.4.3.2: Geometry optimized structures of intermolecular H-bonds of Bm-33 with human pepsin analyzed by LigPlot. A-B representing the standards of Pepsin with Acetylpepstatin and phosphonate inhibitors respectively; C-G representing residues of Bm-33 [In blue, identified as chain B (B)], and pepsin residues in green identified as chain A (A) with a total of up to 19 interactions.

3.5. ELISA based *in vitro* binding affinities of protease and Bm33 peptides

Binding affinity was determined between HIV protease and the four peptides resulting from molecular docking outcomes as possible candidate inhibitors. The peptides were barcoded as, Bm1; Bm3; Bm5; and Bm7 (synthesized by GenScript) with amino acid sequences, WQLARHSEK; SSLAEKN; TTQYYFD and AKMTAYQKYLS respectively. The BSA absorbance background for each serial dilution was averaged and subtracted from absorbance of each respective data reading. Resultant absorbance represented peptides affinities to bind protease at concentrations ranging from 0.0004 to 0.3 µM (Figure 3.5.1). The interaction of peptides with protease was shown to be concentration dependent, increasing with the increase in protease concentration. For all four peptides the highest interaction was observed at 0.3 µM and decreased with the decrease in protease concentration.

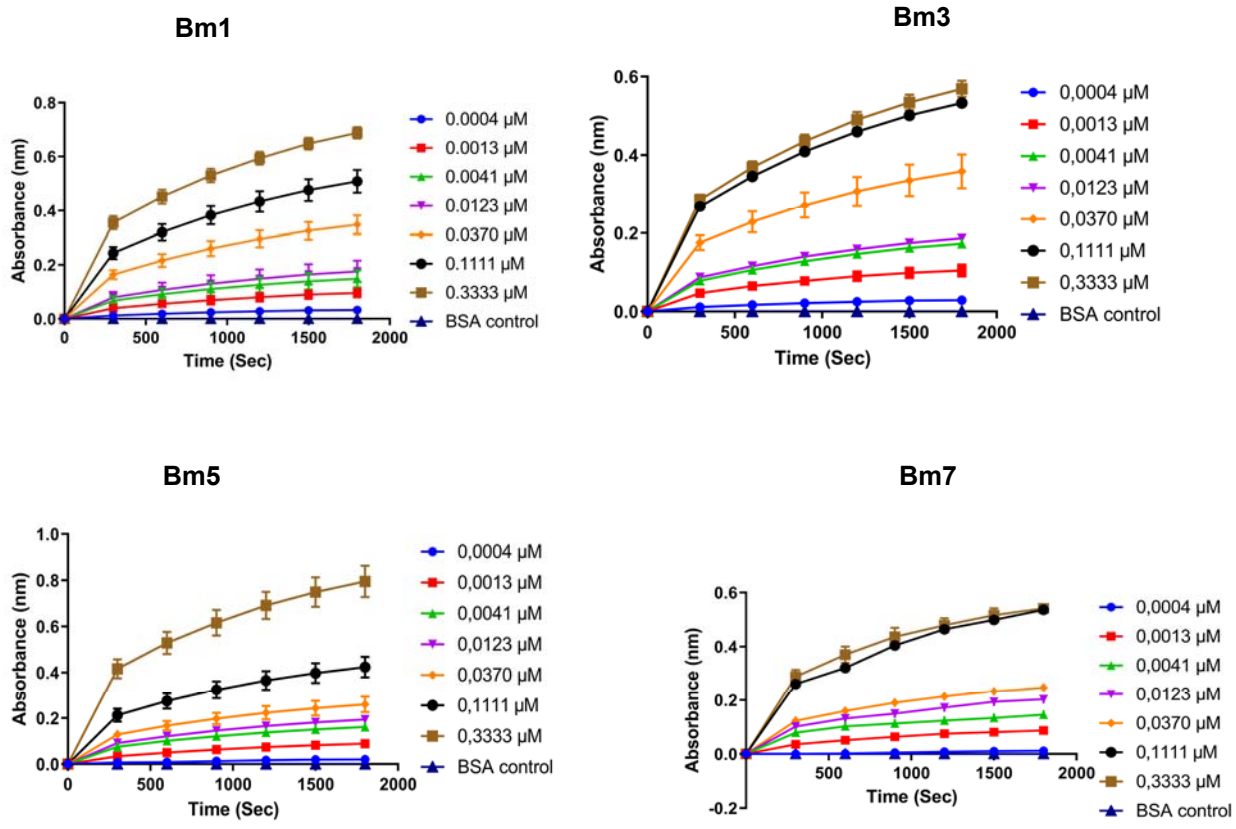


Figure 3.5.1: Representative graphs illustrating concentration dependent binding between protease and four Bm33 peptides, Bm1-Bm7 as a measure of absorbance at 370 nm in 30 minutes. All peptides showed the same pattern of binding which was seen to be dependent on the increase of protease concentration with the highest absorbance of 0.78 nm which shows specific binding of protease and

To further analyze binding affinities of protease and the peptides, data obtained in Figure 3.5.1 was used to generate protein-protein binding curves (Figure 3.5.2), which when fitted yielded estimated equilibrium binding constants (Kd), (Table 3.5). When calculated, the binding kinetics of the four peptides revealed, Bm3 to be bound with highest affinity followed by Bm7, Bm1 then lastly Bm5. However, maximal binding revealed that the difference in each peptide binding protease was not significant ($p < 0.05$ (Figure 3.5.2, insert)).

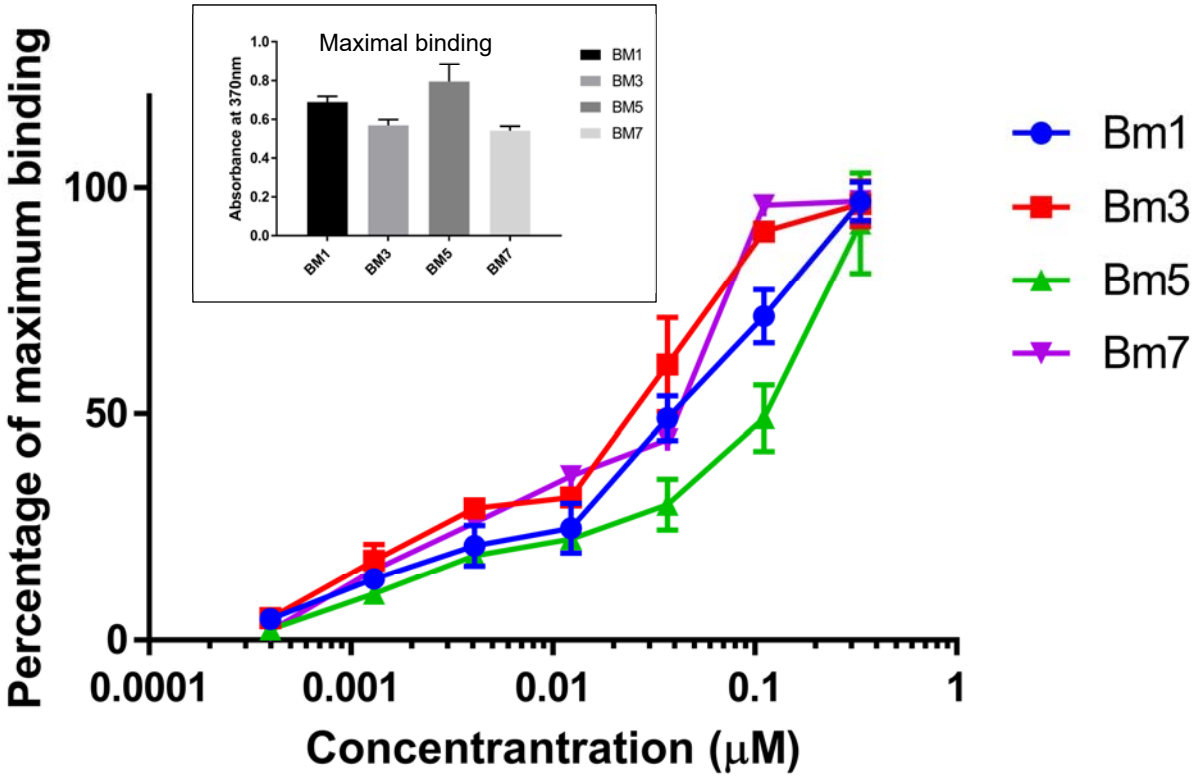


Figure 3.5.2: A representative set of typical protease-Bm33 binding curves illustrating the binding affinities of Bm33 peptides Bm1 (blue), Bm3 (red), Bm5 (green) and Bm7 (purple) with HIV protease. The peptides are configured as the immobilized protein and protease the soluble protein. The binding curves demonstrate Bm3 to be bound with the highest affinity to protease followed by Bm1 then Bm7 and lastly Bm5. This is evident by the lower concentration required for 50% maximal binding. Absolute absorbance of maximal binding for all the peptides (insert bar graph) demonstrate coupling strength.

Table 3.5: Comparative kinetics of peptides binding protease

Peptides	Bm1	Bm3	Bm5	Bm7
Bmax	102.9	102.2	117.1	107.3
Kd	0.03725	0.02016	0.1101	0.02793

*Bmax being the total number of receptors expressed in the same units as the Y values (fmol/mg) and Kd the equilibrium dissociation constant (nM).

3.6. DISCUSSION AND CONCLUSION

In order to assess the binding possibilities of HIV-1 protease to the pepsin inhibitor homolog (Bm-33) which has been previously shown to inhibit pepsin (Krishna *et al.*, 2011, 2013) , structural similarities of the two aspartates were explored through sequence alignment and 3D structure comparison. The aspartates were shown to share catalytic site similarities with Asp 32 as well as 215 in pepsin and Asp 25 in protease being located similarly in the catalytic triad DTG as already depicted in literature (Pearl and Taylor 1987; Gilski *et al*. 2011).

Molecular docking has been widely used for the determination of molecule binding modes and affinities to target proteins in an effort to identify and modify lead molecules in drug discovery. This has led to a successful virtual screening of molecules with potency against target proteins leading to the synthesis of some known inhibitors. Owing to structural homology of human pepsin and HIV protease (Pearl and taylor 1987; Gilski et a., 2011) as well as the history of cross-interactivity of HIV protease and pepsin inhibitors (Sayer and Louis, 2009; Matúz *et al.*, 2012), exploration of the mostly α-helical, peptidic Bm-33 was seen as an important step towards the discovery of possible lead for HIV-1 PR inhibition.

In the current study, docking analysis of Bm-33 with both protease and pepsin, revealed its efficiency of binding to both the proteins at several residues and not only just to the active site residues as it is known for inhibitors that are currently in use against protease (Lv, Chu and Wang, 2015) and Pepsin. When binding affinities of the four inhibitors of protease [Duranavir, Lopinavir, Ritonavir and Acetyl pepstatin) were explored using Ligplot analysis, the highest number of hydrogen bonds were only 6, obtained from the interaction of protease with Duranavir (Figure 3.4.3.1 A). This is very much limited when compared to 15 hydrogen bond interactions of Bm-33 with protease and over 10 hydrophobic contacts (Figure 3.4.3.1 E-J). Up to 19 interactions with Pepsin, a good number of interactions to validate Bm-33 as a good possible lead for protease inhibitor development. Due to the length of Bm-33, only residues involved in hydrogen bonding were grouped in ranges that would allow possible visualization of the interaction between the two molecules. Consequently, 6 plots were generated comprising residue ranging in the following manner: 17-18, 25-30, 47-49, 51-53,

66 and 84-87 (Figure 4.3.3.1. E-J). In the Bm-33 to Pepsin interaction, plots generated comprised residue ranges of 11-19, 29-31, 47-50, 66 and 90-95 (Figure 3.4.3.2. C-G).

From the observed hydrogen bonding to both protease and Pepsin, Bm-33 displays characteristics of an allosteric inhibitor and not an active site inhibitor as compared to the standards whose interactions are competitive of the substrate, fitting right within the protease active site pocket. In this study Bm-33 showed no interactions with any of the active site residues but it has however been shown in previous studies to inhibit pepsin and other human aspartic proteases (Krishna *et al.*, 2011, 2013). This confirms it as an allosteric inhibitor binding elsewhere in the protein away from the active site but inducing conformational changes that alters protein activity

Interestingly, as Bm-33 was shown to exert hydrogen bonds towards the active site residues Asp 25 and Thr26, a hydrophobic interaction towards Gly 27, the third residue forming the conserved catalytic triad Asp, Thr, Gly (Pearl and Taylor, 1987; Gilski *et al.*, 2011) was also noted. Furthermore, Bm-33 is as well seen interacting with residues involved in secondary facilitation of protein activity outside of the active site such as Arg 87. Together with Gly 86 and Asn/Asp 88, Arg 87 form the second conserved triad in the α helices of each monomer of the dimeric protease. Studies have elucidated complete obliteration of protease activity owing to substitutions on the highly conserved Arg 87 (Louis *et al.*, 1989; Ishima *et al.*, 2010), thereby depicting the residue's importance in the protein's proteolytic activity. Furthermore, Arg 87 and Arg 8 have been reported to form inter-monomer and intra-monomer hydrogen bonds respectively with Asp 29 (Weber, 1990; Ishima *et al.*, 2003). In this study, Bm-33 wash shown to have up to 3 hydrogen bonds with Arg 8 alone, further validating its possibility to retard protease activity through disruption of the intra-monomer interaction mentioned above. Other residues involved in protease- Bm-33 interactions include Gly 51 which is vital in protease structural stability through its involvement in protein-protein interactions for the facilitation of dimer stability (Todd, Semo and Freire, 1998). Glycine 51 also forms part of the glycine rich region, essential for protease flap flexibility (Scott and Schiffer, 2000); interference thereof could have a negative impact in substrate reception and retention for proteolysis, a good retort of inhibition. Over 50% of Bm-33 interactions with protease

involved residues forming part of the salt-bridge, a phenomenon crucial for hinge region stability (Naicker *et al.*, 2013).

In vitro binding assay further confirmed interactions of protease with the peptides as noted in molecular docking. Though there was no significant difference in the affinities at which the peptides bound protease, Bm3 was seen to exert the highest affinity, followed by Bm7, Bm1 then lastly Bm5. Although the peptides could not produce inhibitory activities against protease (as seen in Section 3.5.7) the binding assay has shown reliability of docking results produced in this work. It should be noted that peptides used for protease inhibition were only small portions generated from interacting residues selected from Bm-33 and not the complete molecule. Synergistic interactions might have played a role in the noted Bm-33 and protease interactions upon molecular docking and that would have been a shortfall for the peptides. A complete Bm-33 polypeptide has previously been shown to be inhibitory to human pepsin (Krishna *et al.*, 2011), a homolog of protease. Therefore, generation of the complete Bm-33 polypeptide would be an important milestone towards possible protease inhibition.

In conclusion, this work revealed the first direct evidence of molecular interaction of Bm-33 and HIV-1 PR, setting a platform for future exploration of Bm-33 as a possible lead for HIV-1 PR inhibition.

CHAPTER FOUR: EXPRESSION, CHARACTERIZATION AND EVALUATION OF HIV-1 SUBTYPE C PROTEASE

ABSTRACT

Studies on inhibition of HIV-1C enzymes such as protease, which exhibits high variability compared to subtype B protease, are required. In the current study, the South African HIV-1 C protease isolate was successfully expressed in *E. coli*, characterized and its activity evaluated. Inhibition of the produced protease with peptides generated from molecular docking outcomes as well as natural products previously isolated from the methanol stem-bark extract of *Peltophorum africanum* and tested against HIV reverse transcriptase was determined. Although none of the peptides generated from molecular docking outcomes showed any activity against protease, from the three isolated molecules, gallotanin exhibited noted protease inhibition (IC_{50} = 1.937 µg/ml). Neither bergenin nor catechin elicited any inhibition against protease. Inhibition of protease by gallotanin, has however set a platform for the refinement and characterization for its possibility in the development of a natural broad-spectrum inhibitor of HIV-1 since it has as well been shown to inhibit HIV reverse transcriptase and integrase in previous studies. HIV-1C protease was successfully expressed and characterized. The applicability of protease activity evaluation assay used herein, in inhibition assessment has set a platform for continual screening of molecules that can inhibit HIV-1C protease.

4.1. STUDY OBJECTIVES

4.1.1. Aim

To overexpress and purify an HIV-1C protease isolate that is prevalent in South Africa for application in the development of subtype specific inhibition assays.

4.1.2. Objectives

- To express and purify an HIV-1C protease isolate (prevalent in South Africa) in *E. coli*
- To determine the secondary and tertiary structures of the expressed protease
- To evaluate the proteolytic activity of the expressed protease
- To assess protease catalytic activity by inhibition with peptides generated from molecular docking outcomes as well as previously isolated natural compounds

4.2. INTRODUCTION

Cleavage of polypeptide chains into functional proteins is very crucial for viral maturation, thereby making HIV-1 protease one of the most important enzymes in the life cycle of HIV. Protease inhibition has been shown to be highly essential in HIV/AIDS therapy. However, adverse side effects of protease inhibitors are in contrast with their potency (Bozzette, Ake, Tam, Chang, & Louis, 2003; Kotler, 2008; V. *et al*., 2014). Multiple amino acid substitutions in HIV-1 protease have also been shown to be a cause of reduced sensitivity of the virus to available protease inhibitors (Barrie *et al.*, 1996).

The complex genetic diversity in HIV-1 is as a consequence of the viral error prone reverse transcriptase which lacks proof reading mechanism, leading to the production of progeny virions different from the parental viruses. This makes HIV-1 highly prone to polymorphisms and resistance mutations (Mata-Munguía *et al*., 2014). Studies have elucidated differences in mutational frequency, disease progression, drug response and transmission amongst HIV-1 subtypes (Martinez-Cajas *et al*., 2008 ; Pant Pai *et al*., 2011, 2012; Bhargava *et al*., 2014). HIV-1C, the most predominant variant worldwide (Hemelaar, 2012; Tebit and Arts, 2014) has been shown to be more prone to drug resistance and more virulent than its counterpart HIV-1B (Häggblom *et al*., 2017). However, none of the drugs currently in use address subtype differences. All drugs available for use worldwide are tailored toward the HIV-1B genetic background. With known variation in resistance profiles amongst HIV subtypes, it is therefore imperative to design drugs with high increased subtype specificity and a lower resistance barrier.

A high proportion of polymorphisms related to drug resistant sites have been noted on HIV-1C than on HIV-1B proteases (Toor *et al.*, 2011). Mutations such as K20R, M36I and I93L, known to occur as a result of drug pressure in HIV-1B protease, have been seen at high proportions in drug naïve HIV-1C infected individuals (Bessong *et al*., 2008). Studies conducted on samples from South Africa, Botswana and Zimbabwe, where 90% of HIV-1 infection is accounted for by HIV-1C, also confirmed high mutational frequency of HIV-1C protease over subtype B protease (Bessong and Obi, 2006). HIV-1C protease was also

shown to be different from subtype B protease at eight positions, T12S, I15V, L19I, M36I, R41K, H69K, L89M and I93L (Bessong *et al.,* 2008).

Polymorphisms I15V, E35D, R41K and R57K, in subtypes C, F and CRF_01 A/E together with mutation S37N have been shown to alter the protease flap orientation resulting to an alternative “open” flap conformation (Huang *et al*., 2014). This is also referred to as the curled flap conformation, responsible for drug resistance (Kear *et al.*, 2009). The above-mentioned substitutions occur naturally in subtypes C, F and CRF_ 01 A/E. In other subtypes, including HIV-1B, such substitutions occur as secondary mutations to compensate active site primary mutations for restoration of enzyme activity (Weber and Agniswamy, 2009; Agniswamy *et al.*, 2012). South African HIV-1C protease flaps have been shown to exhibit a wider range of orientation as compared to subtype B (Naicker *et al*., 2013); thereby increasing the potential for a lower resistance barrier for HIV-1C protease.

With the stipulated mutational frequencies and therapeutic outcomes in subtype C proteases and the worldwide prevalence of HIV-1C, availability of well characterized subtype c proteases and recombinant forms would therefore be of importance. This would allow for continual screening and development of updated molecules with subtype C specified anti-protease activity. This study was therefore focused on the expression and characterization of HIV-1C proteases which can be applied in the subsequent development of subtype C based anti-HIV enzyme assays.

4.3. MATERIAL AND METHODS

4.3.1. Sources of reagents

Glycerol stocks of BL21 (DE3) *E.coli* cells transformed with South African HIV-1C based protease plasmid (pET11b-HIV-1C protease) were obtained from Professor Yasien Sayed (Protein Structure and Function Research Unit, Witwatersrand University, Johannesburg, South Africa). The plasmid is a T7 promoter-based system that is regulated by the expression of a T7 polymerase.

4.3.2. Protease Expression

Glycerol stocks of BL21 (DE3) pLysS *E. coli* competent cells transformed with HIV-1 C protease, were resuscitated in 100 ml overnight culture at a ratio 1µl: 1ml, glycerol stock: broth) enriched with 100µg/ml of both Ampicillin as well Chloramphenicol (Melford, Chelsworth, United Kingdom). The culture was then grown for 16 hours at 30°C with agitation at 100 rpm. The same concentrations of Ampicillin and Chloramphenicol were used throughout the study unless otherwise stated. Expression was carried out the same way as in Section 2.1, except that the resuspension of the bacterial pellet was done in a protease binding buffer (10mM Tris-HCL pH 8.0; 2mM EDTA disodium; 1mM PMSF; Sigmaaldrich, Missouri, USA) to a final volume of 50ml and stored at -20°C.

4.3.3. Bacterial cell lysis and protein fractionation by Ion-exchange Column Chromatography

Bacterial cell lysis was done as reported in Section 2.3.3.1. Centrifugations were done at 5000 xG for 30 minutes at 4°C, unless otherwise stated. The retrieved protease formed inclusion bodies. To solubilize the protein, the pellet was washed in ice cold binding buffer (with 2% Triton-X100) adjusting the volume to 25ml per each 50ml pellet. Each resuspension was homogenized and centrifuged, pellets resuspended in denaturing buffer (10mM Tris-HCL pH 8.0; 8M Urea; 5M DTT ; Sigmaaldrich, Missouri, USA) at room temperature. Resuspended pellets were homogenized and incubated for 30 minutes at

room temperature. This was followed by centrifugation at 20°C and subsequent collection of the unfolded protein in the supernatant. To refold the protein, Formic acid was added to the retrieved supernatant to a final concentration of 25mM and then dialyzed in a snakeskin dialysis tubing (3 kDa cut off; Separations, Rodepoort, South africa,) overnight at 4°C in 5L of refolding buffer (10mM Formic acid; 10% Glycerol; 0.02% Sodium azide *w/v* ; Sigmaaldrich, Missouri, USA). Buffer exchange of the protein into storage buffer (10mM Sodium acetate pH 5.0; 2mM DTT; 0.02% Sodium azide *w/v* ; Sigmaaldrich, Missouri, USA) was carried out in an overnight dialysis of the protein sample in a snakeskin dialysis tubing against 5L of storage buffer at 4°C. The protein sample was transferred to 50ml centrifuge tubes, centrifuged and filtered with a 0.45µm filter for immediate subsequent purification.

The protein was allowed to flow through a CM-Sepharose cation exchange column connected to the AKTA prime purification system (GE Healthcare, Illinois, USA) which was regenerated with 1M NaCl and equilibrated with degassed fresh storage buffer (without DTT). This was followed by gradient elution of the protein with 0 – 1M NaCl at 4ml/minute flow rate (Figure 4.5.1B). Fractions containing the right size of the protease were pooled together and dialyzed overnight against 5L of storage buffer at 4°C. The resulting protease was concetrated using a 3 kDa Centricon membrane (Millipore, Billerica, MA, USA; Figure 4.5.1A) Following resolution on a 12% Tricine gel, the protein was quantified using the VersaMax spectrophotometric reader at 280 nm absorbance and then stored at -80°C for subsequent biophysical characterization.

4.3.4. Structural and functional characterization of the purified protease

4.3.4.1. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) measurements to determine the concentration of an active (dimeric form) of protease was obtained by determining the binding constants of protease with Acetyl-pepstatin-A (Acep-A) using TA-ITC calorimeter (TA-ITC, Delaware, USA). The calorimeter was sterilized with 250ml Conrad, rinsed with 750ml distilled water and equilibrated with 50ml of protease storage buffer. Acetyl-pepstatine-A (Acep-A), was dialyzed against protease storage buffer as well. Protease and Acep-A concentrations

were both adjusted using the PR storage buffer. Protease was diluted to 25µM and Acep-A to 20µM in a final titration volume of 25µl. The reaction mixture was injected into the calorimeter and titrations ran at 28°C. Data analysis was conducted using the TA-instruments ITC software, Tros Version 4.3.0 (TA-ITC, Delaware, USA). The percentage of the active protein was calculated by multiplying n (number of active sites/percentage of active protein) with the actual final concentration of the protein injected into the calorimeter (Figure 4.5.2).

4.3.4.1 Far-UV circular dichroism spectroscopy

Circular dichroism for determining α; β constituents of the purified (native) protease was determined using Chira-scan CD- spectrometer (Applied Photophysics, Surrey, United Kingdom). Protease was diluted with milliQ distilled water to a final concentration of 10µM. Water was considered for dilution to reduce background noise. Measurements were performed using a 1mm path-length cuvette at the wavelength range of 180-250nm at 293K at 1nm resolution. Baseline reading was taken using water and the machine was set to auto-subtract the baseline from the final absorbance readings. Five readings were taken for each measurement, averaged and smoothed using the Chira-scan pro-data viewer and then converted to excel using APL data converter. To normalize the obtained data, mean residue ellipticity (MRE) was determined from obtained ellipticity using the following equation: MRE = MRWθ/ 10d.c where MRW is the mean residue weight, θ is the observed ellipticity (degrees), d is cuvette path-length and c is the protein concentration (g/ml). The MRW is obtained by dividing the protein molecular mass by N-1 where N is the number of amino acid residues. Secondary structure of the protein was illustrated by plotting the wavelength against the MRE.

4.3.4.3. Fluorescence spectroscopy

Fluorescence spectra of protease was determined the same way as done in Section 2.3.4.4 for reverse transcriptase. Tryptophan emission of protease was measured from 300-700 nm where the excitation wavelength was set at 280 nm. Fluorescence intensities of protease was determined in both native and unfolded forms. For native experiments, protease was diluted to 51µM in PBS and fluorescence was measured shortly after that. To perform unfolded experiments of protease, Urea was included in PBS to a final

concentration of 8M while maintaining the protease concentration of 51µM. Fluorescence emission of the unfolded protease was read following a 1-hour incubation in 8M Urea-PBS buffer at room temperature. The obtained data was analyzed using GenS software.

4.3.4.4. Size exclusion chromatography

The SEC analyses of protease were conducted the same way as done in Section 2.3.4.5, except that the protease concentration was at 25.5 µM. Protein standards of Aprotinin (10 mg), Ribonuclease A (50 mg), Ovalbumin (50 mg) and Aldolase (950 mg (GE Healthcare, Björkgatan 30, Sweden) were ran to determine protease molecular weight from the SEC.

4.4.4.5. Catalytic activity evaluation of protease

Protease activity was evaluated in a reaction buffer containing 2 M NaCl, 60 mM NaOAc (pH 5.0) at a final reaction volume of 200 µl in non-clear 96 well plates. Constant concentrations of 0.02 and 2 µM for protease and the substrate respectively were maintained. Protease was added first, and the reaction was initiated by addition of the fluorogenic substrate (Sigmaaldrich, Missouri, USA). Fluorescence was read in a VersaMax microplate reader (Molecular devices, California, USA) over 15 minutes at 37ºC. Excitation and emission monochromators of the microplate reader were set at 350 and 490 nm respectively (as per substrate specifications). Reactions were run in duplicates and two independent determinations were carried out. Protease activity was determined as a measure of increase in emitted fluorescence upon substrate proteolysis. Lopinavir was used as a negative control to inhibit the reaction at concentrations ranging from 1, serially diluted to 0.25 µM. To further validate substrate proteolysis, the substrate was incubated in the absence of protease and the fluorescence thereof was determined.

4.3.5. Protease inhibition assay

To test the applicability of the above conducted protease activity evaluation experiment in inhibition assay development, natural products previously isolated from methanol stem-bark extract of *P. africanum* were assayed. The isolated molecules included two compounds of known molecular weights, bergenin and catechin as well as a hydrolyzable tannin (gallotanin (Bessong e*t al.,* 2005). Other molecules tested included the four peptides, generated based on molecular docking outcomes in Section 4.2.2.

Inhibitory characteristics of test molecules were determined by the inclusion of each of the molecules at varying concentrations in the protease activity evaluation assay conducted in Section 4.4.4.5 above. Lopinavir was used as a negative control in this case. Each of the test molecules was assayed individually. However, the four peptides were later combined and assayed as one treatment (referred to as Bm33). Inhibitory molecules were serially diluted in reaction buffer, followed by addition of protease. The mixture was preincubated for 10 minutes prior reaction initiation by addition of the substrate.

4.5. RESULTS

4.5.1. Expression and Purification of protease

Overexpression of protease was shown by the presence of a 22 kDa band on 16% SDS gel which was seen appearing upon IPTG induction (lanes 2-4, Figure 4.5.1A). The absence of a band in lane 1 was an indication of successful suppression of the T7 polymerase phage which should only be activated upon IPTG addition. During purification the IMAC profile, using a 0-1 M NaCl gradient wash (Figure 4.5.1B), showed elution of protease at 10-150 mM NaCl. The presence of protease in the pooled fractions was confirmed by appearance of a 22 kDa band (lane 7 Figure 4.5.1C) during phoretic analysis. Lanes 1 to 6 (Figure 4.5.1B) are representative of washes of protease at an increasing concentration of NaCl buffer (0-100 mM) . This showed successful expression of protease in its homodimeric form. A faint band appearing above protease might have been a polymerized form of protease as this was also seen appearing in a confirmatory western blot (Figure 4.5.1D).

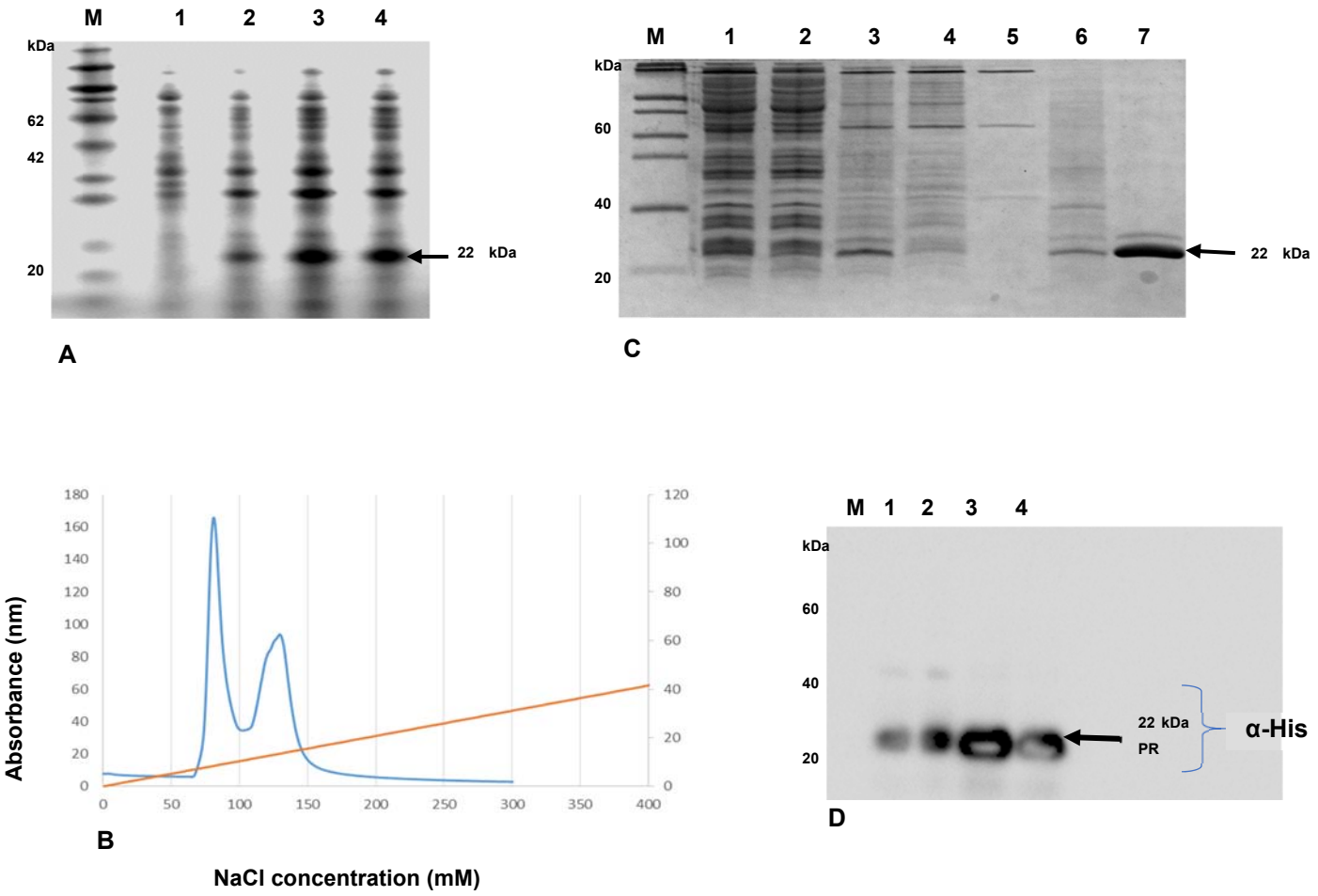


Figure 4.5.1 HIV protease expression and purification profiles, A**:** A 4-hour time course expression of HIV protease showing overtime accumulation of protease at 1-hour interval following IPTG induction, seen by appearance of an ~22 kDa band appearing just above the 20 kDa band of a SeeBlue protein marker (Lane M). B: An IMAC purification profile of a 0- 400 mM NaCl gradient elution of the expressed protein. C: An SDS gel showing fractions pooled from IMAC during purification. Appearance of a 22 kDa band eluting at 100-150 mM NaCl is indicative of protease on lane 7 of the gel, with lanes 1-6 being indicative of . D: A Western blot, confirmatory of the 22 kDa HIV protease.

4.5.2. Isothermal calorimetry titration

The observed endothermic heat signal observed during active- site titration displayed the ability of protease to bind acetyl pepstatin. The stoichiometry value of 0.655 was indicative of approximately 65% of the protein that is in the active (dimeric) conformation (Figure 4.5.2).

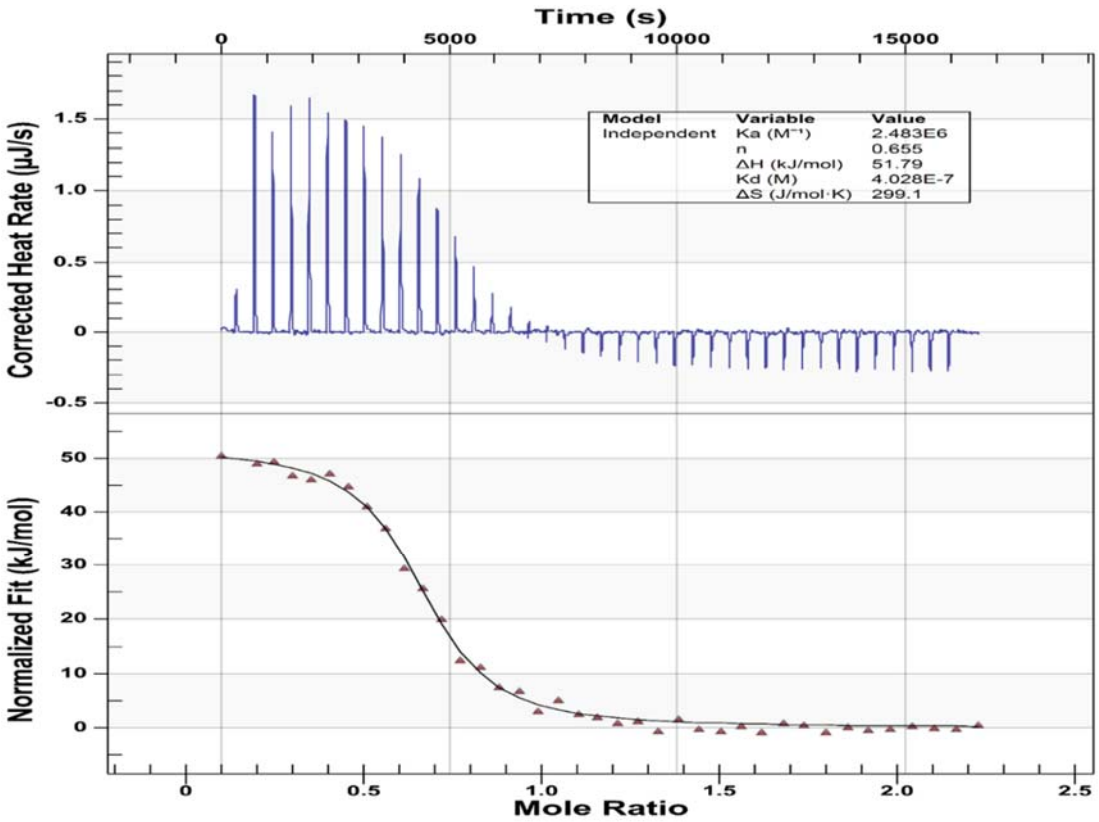


Figure 4.5.2: Active- site titration of HIV-1 Protease binding with 20µM Acetyl pepstatin A. Corrected calorimetric peaks are shown on the upper panel. Integrated heats for the corrected calorimetric data, against acetyl pepstatin in molar ratio. The binding parameters shown in the insert, where K_a is the binding constant; n, number of binding sites; ΔH, enthalpy of binding; ΔS, entropy of binding. The actual concentration of the active/ dimeric protein was determined by multiplying the concentration of the injected protein by n.

4.5.3. Far UV- circular dichroism (CD) spectroscopy of Protease

The characteristic profile of the protein secondary structure is determined by the wavelength at which the ellipticity is mostly minimum. For the structure to be considered as either mostly β or α, the wavelength at which the ellipticity is minimum should either be 210 - 215 nm or 220 respectively. The CD spectrum of protease (Figure 4.5.3) at 200-280 nm represented a typical β- helical molecule as it is known for HIV protease to be comprised mostly of β-helices, with a reduced α-sheet content (Louis *et al*., 2007).

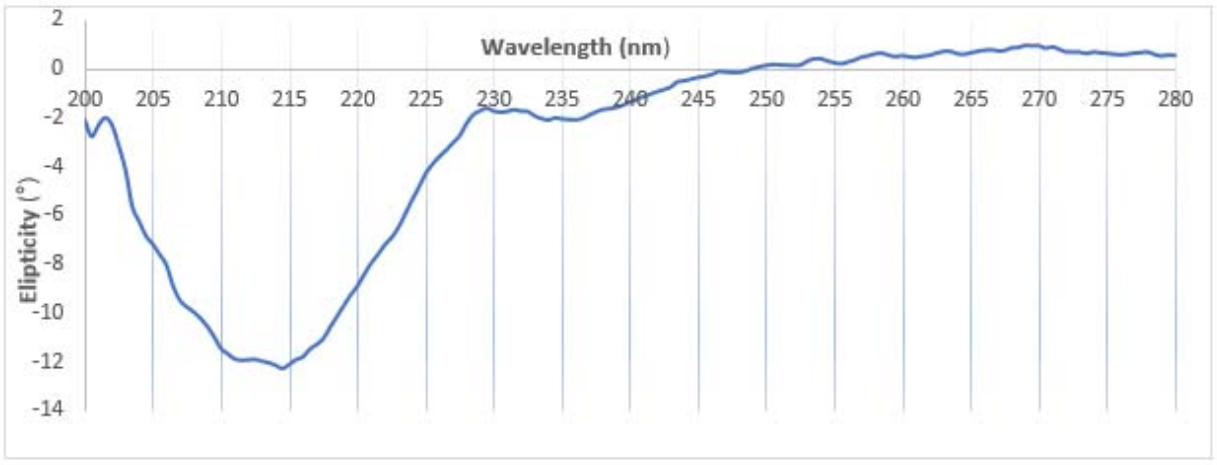


Figure 4.5.3: A far UV-CD spectrum of purified protease (native), recorded at a Protease concentration of 10 μM at pH 7.0. The minimum ellipticity obtained at 215 nm indicating protease high constituent of β sheets, a characteristic of HIV protease.

4.5.4. Fluorescence spectroscopy

Selective excitation of tryptophan and tyrosine residues in protease yielded intrinsic fluorescence emission. The emission maxima shift from ~338 to 340 nm (Figure 4.5.4) from the native to unfolded conformation of protease respectively, were . Protease has only two tryptophan residues that are buried, at the flap tips of each protease monomer (at positions 6 and 42) in its folded form. Reduction in fluorescence intensity is consequence of dimer dissociation by treatment with Urea, which supposedly exposes

the Tryptophan residues and thereby reducing the quantum yield. This served to confirm that the expressed protease was in its folded functional conformation

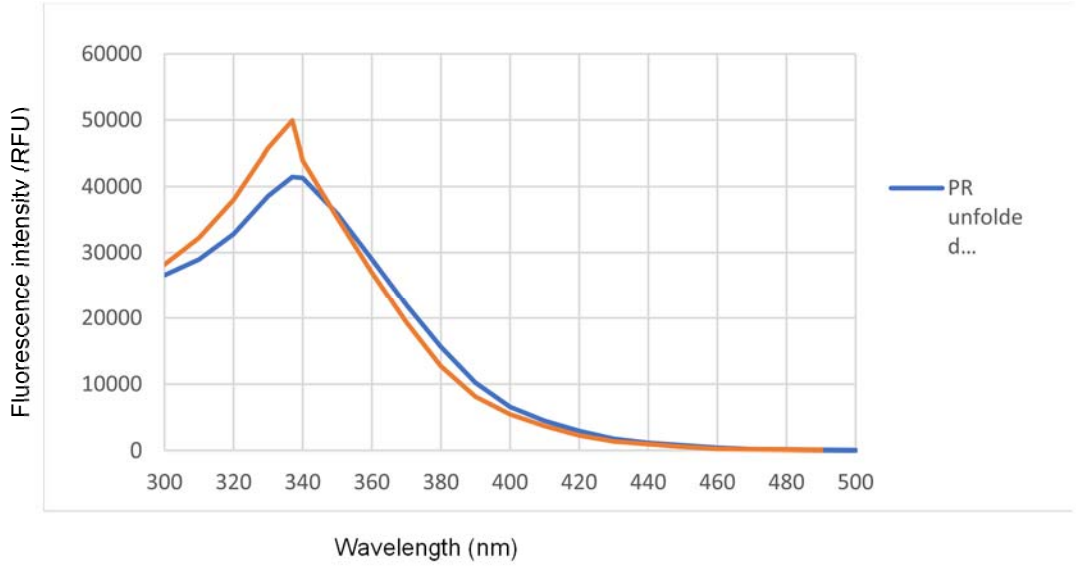


Figure 4.5.4: An illustration of the fluorescence spectrophotometry emission spectra profiles of a native (blue) and an 8M Urea unfolded (orange) protease.

4.5.5. Size exclusion chromatography of Protease

Two peaks that seemed like one stretched peak from 34 to 36.51 minutes (Figure 4.5.5) were indicative of protease elution. The two peaks were indicative of the monomeric and the dimeric forms of protease respectively. The standards used, standard curve and calculations for comparison used in this section are shown in the appendix section (Appendices A1 and A2).

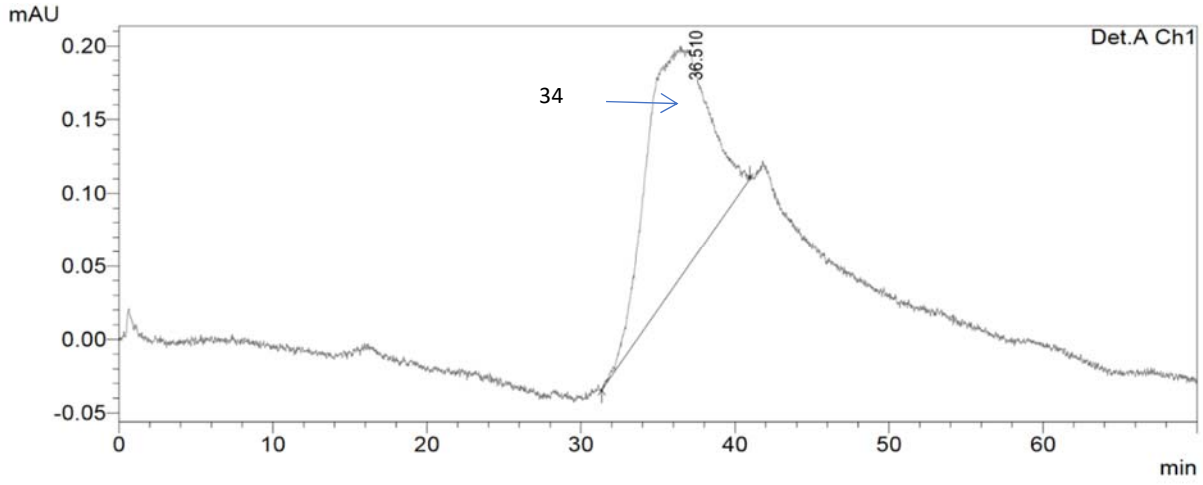


Figure 4.5.5: An HIV protease size exclusion chromatogram, illustrating elution profiles of the monomeric and dimeric forms of protease. Fractions were collected at 34 and 36 minutes respectively.

Molecular weights (Mr) for the two peaks obtained from the protease run were shown to be 11 kDa and 18 kDa for the first and second peaks, respectively (Table 4.5.5). Peak 1 has the exact molar mass for a protease monomer (11 kDa) and the second peak appeared apparently lower than the expected 22 kDa dimer size at 18 kDa.

Table 4.5.5: Protease retention factor and molecular weight calculations

Peaks	Time (min)	Ve (µl)	Kav	Mr (Da)
Peak 1	34	10	0.590371	18343.22
Peak 2	36.51	11	0.665057	10873.62

Formula used: Kav = (Ve - Vo)/ (Vc – Vo), where Ve is the sample elution volume; Vc and Vo, geometric and void volumes of the column respectively (all in µl). Where Vo = 4.2477; Vc = 14.33. Molecular weight equation was generated from the standard curve (plotted from standard runs, graph represented in the appendix section), x = (2.1707- y)/ (0.6122), where x is the molecular weight and y being the protein Kav.

4.5.6. Catalytic activity evaluation of protease

Protease (PR) activity, when determined as a measure of emitted fluorescence upon catalysis of the fluorogenic substrate was shown to be optimal. The blue plot in Figure 4.5.6 shows successful proteolysis of 2 µM of the fluorogenic substrate by 20 nM protease expressed in Section 4.3.2. Fluorescence was seen increasing in a 15 minutes incubation from 24 to 94 relative fluorescence units (RFU). This was validated by observation of a constant fluorescence along 12 RFU (Figure 4.5.6, red plot) emitted when 0.0625 µM of lopinavir was included in the reaction. Further validation was shown by a constant emission of fluorescence below 2 RFU of the substrate incubated in the absence of protease.

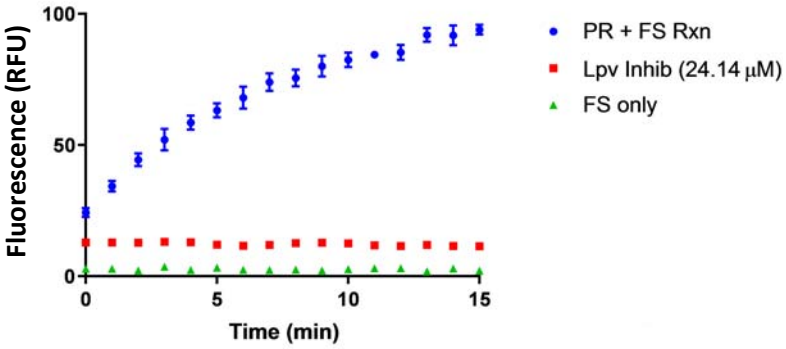


Figure 4.5.6: A time dependent curve showing an increase in fluorescence emission (blue), an indication of successful proteolysis of the fluorogenic substrate (FS) by the expressed protease (PR). Non-reactivity of the substrate in the absence of PR (green) validated the observed results. Inhibition of the reaction was achieved with 0.0625 µM of lopinavir (red).

4.5.7. Bioactivity of test molecules against protease

From the three molecules isolated from *P. africanum* in a previous study (Bessong et al., 2005), only gallotanin exhibited a marked inhibition of protease. A clear illustration of substrate cleavage disruption indicating protease inhibition by gallotanin was shown in a time dependent graph showing fluorescence emission upon cleavage by protease (Figure 4.5.7.1). This showed the same trend as lopinavir inhibition in protease catalytic activity analysis (Section 4.5.6).

When assayed with protease for inhibition efficiency assessment, none of the four peptides synthesized on the bases of molecular docking outcomes (Section 3.5) showed

no activity against protease even at 100 µM (results shown in the appendix section, Appendix B).

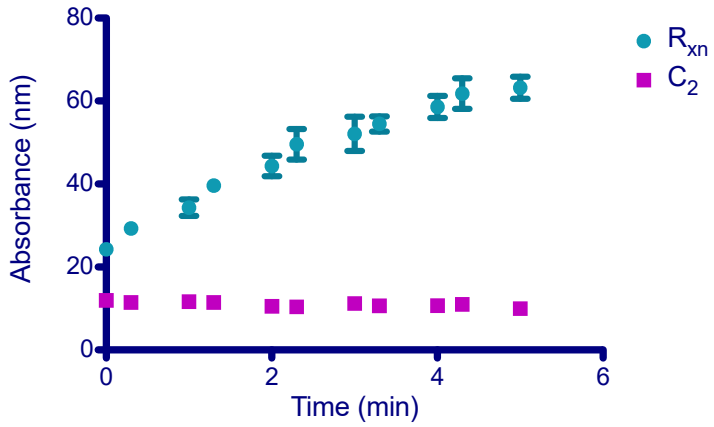


Figure 4.5.7: A time dependent curve where protease inhibition by gallotanin is indicated by constant fluorescence emission (C_2/purple) against normal protease activity where fluorescence emission increases (R_{xn}/green) over time upon substrate proteolysis.

A calculated LogIC_{50} of 0.2871 and an IC_{50} of 1.937 µg/ml (Figure 4.5.7.2A) were obtained with gallotanin when compared to 0.015 µg/ml of the positive control, lopinavir (Figure 4.5.7.2B). Similar to the peptides, bergenin and catechin showed no noted activity against protease (graphs shown in the appendix section).

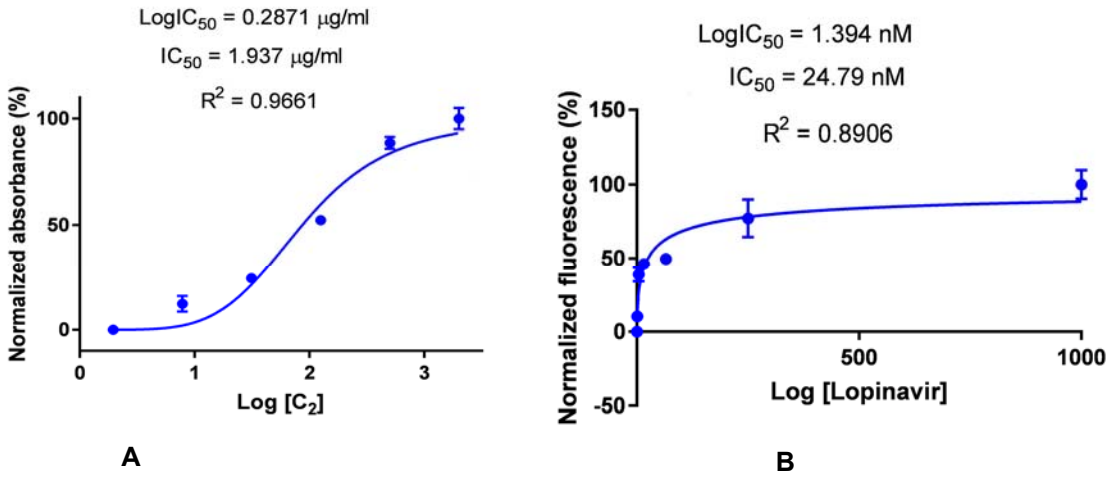


Figure 4.5.7: A dose response sigmoid curve representing concentration of gallotanin inhibiting 50% of protease activity at an IC_{50} of 1.937 µg/ml (A). The R^2 value of 0.9661 indicated a good fit of the obtained data. The positive control, a known protease inhibitor (lopinavir) is also shown (B) to inhibit 50% protease activity at an IC_{50} of 0.024 µM (0.015 µg/ml).

4.6. DISCUSSION AND CONCLUSION

Successful expression of a dimeric 22 kDa protease was achieved. The protease was intended for use in molecule screening assay development and therefore its structural characteristics should be known. That was achieved by determining the protease tertiary structure by Far-UV CD spectrometry, Fluorescence spectrometry as well as SEC. Documenting the normal behavioural patterns of the protein in its native form is of importance as the very same parameters could be used to observe changes when an inhibitor or mutational changes are introduced to the protein.

Far-UV CD spectral results revealed that the expressed HIV-1C protease is comprised of 99% β-sheets, similar to previous reports (Weber, 1990; Louis *et al.*, 2007). Disruption of the C-terminal residues forming up the β-sheet has been associated with loss of protease ability to dimerize (Naicker *et al.*, 2013), thereby depicting the importance of the β-sheet residues in protease dimerization. Each one of the two protease monomers is known to comprise seven β sheets and just one α helix (Lee and Chmielewski, 2010).

Fluorescence spectral profiles yielded the emission maxima of 338nm for the native protease which was seen increasing to approximately 340nm when the protein was denatured with 8M urea. This confirmed the intactness of the produced protease. An increase in fluorescence was indicative of exposed tryptophan residues which resulted from denaturation of protease by Urea. In the active dimeric form of protease, the two Tryptophan residues on the flap tips of each monomer remain buried and thereby reducing the emission spectra of the residues.

Upon analysis with SEC, two peaks were identified which seemed to have been the monomeric, 11 kDa and an 18 kDa, assumed to be the dimeric form of protease. Although the 18 kDa band seemed to be a bit lower than the known 22 kDa dimeric form of protease (Kim *et al.*, 1995), the protein is expected to have folded to its a quaternary structure a known characteristic in proteins (Cieplaka and Niewieczerza, 2009). The monomer and dimer percentage of collected fractions (11 and 10 µl of dimer and monomer respectively)

in size exclusion chromatography confirmed the presence 65% of the active protease as determined by the active site titration calorimetry (ITC).

To assess possible application of the protease activity evaluation assay in inhibitory molecule screening, plant derived molecules and synthetic peptides were assayed. Inhibition assay yielded no inhibitory activities of the four Bm-33 peptides against protease even at 100 µM.

Plant derived molecules assayed herein for protease inhibition were isolated in a previous study conducted by Bessong *et al* (2005). Of the three molecules, only gallotanin was found to be inhibitory to protease at an IC50 of 1.937 µg/ml as compared to a modified known protease inhibitor, lopinavir (IC50 = 0.015 µg/ml). In previous studies, gallotanin has been shown to elicit HIV reverse transcriptase inhibition by interfering with both RNA/DNA-dependent polymerase and ribonuclease H activities (Andreola, 2001; Shaw-Reid *et al*., 2003 and Bessong *et al*., 2005). To a lower extent, gallotanin has also been shown to have some effect on the 3' end processing of HIV integrase (Bessong *et al*., 2005). This qualifies gallotanin as a potent candidate for broad-spectrum drug synthesis.

In conclusion, protocols developed herein for expression and characterization of protease have set a platform for routine HIV-1C protease production and inhibitory molecule screening. With just a few additional modifications, the protocol developed in this study can also be applied for expression, characterization and inhibition of mutant forms of the protein.

CHAPTER FIVE: GENERAL CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

The current work demonstrated heterologous expression of proteins from the globally prevalent HIV-1C. Structural and functional features of the expressed proteins were shown to be intact for application in downstream HIV-1 inhibition assays. With minor modifications, activity evaluation assays employed for both reverse transcriptase and protease in the current work, have been shown to have the capability for application as inhibition assays. Furthermore, the PCR based protocol developed herein for determining the activity of HIV-1C reverse transcriptase has set a platform in the development of a cost effective HIV-1C inhibition assay that would be used in resource limited settings. *In silico* and *in vitro* characterization of protease interaction with natural compounds revealed convincing possible inhibition of HIV-1C protease. Thus, setting a platform for future exploration of gallotanin as well as the Bm-33 for possible inhibition of HIV-1C protease. The outcomes of this work made a significant contribution in setting up a platform for possible inhibition of the globally prevalent HIV-1C.

Envisaged plan for future work would involve scaling up the expression of the enzymes to ensure sufficiency for routine use in envisaged inhibition assays. Modification of the developed reverse transcriptase activity assay developed herein and development ELISA for reverse transcriptase inhibition will be the following step.

REFERENCES

Agniswamy, J. *et al.* (2012) 'Terminal interface conformations modulate dimer stability prior to amino terminal autoprocessing of HIV-1 protease.', *Biochemistry*, 51(5), pp. 1041–50. doi: 10.1021/bi201809s.

Arts, E. J. and Hazuda, D. J. (2012) 'HIV-1 antiretroviral drug therapy', *Cold Spring Harbor Perspectives in Medicine*, 2(4). doi: 10.1101/cshperspect.a007161.

Barile, M. *et al.* (1994) 'Mechanisms of toxicity of 3'-azido-3'-deoxythymidine. Its interaction with adenylate kinase', *Biochem Pharmacol*, 48(7), pp. 1405–1412. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7945440.

Barrie, K. A. *et al.* (1996) 'Natural Variation in HIV-1 Protease, Gag p7 and p6, and Protease Cleavage Sites within Gag/Pol Polyproteins: Amino Acid Substitutions in the Absence of Protease Inhibitors in Mothers and Children Infected by Human Immunodeficiency Virus Type 1', *Virology*, 219(2), pp. 407–416. doi: 10.1006/viro.1996.0266.

Bebenek, K. *et al.* (1993) 'Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots', *Journal of Biological Chemistry*, 268(14), pp. 10324–10334.

van den Berg-Wolf, M. *et al.* (2008) 'Virologic, immunologic, clinical, safety, and resistance outcomes from a long-term comparison of efavirenz-based versus nevirapine-based antiretroviral regimens as initial therapy in HIV-1-infected persons.', *HIV clinical trials*, 9(5), pp. 324–36. doi: 10.1310/hct0905-324.

Bessong, P. O. (2008) 'Polymorphisms in HIV-1 subtype C proteases and the potential impact on protease inhibitors', *Tropical Medicine and International Health*, 13(2), pp. 144–151. doi: 10.1111/j.1365-3156.2007.01984.x.

Bessong, P. O. and Obi, C. L. (2006) 'Ethnopharmacology of human immunodeficiency virus in South Africa - A minireview', *African Journal of Biotechnology*.

Bhargava, M. *et al.* (2014) 'Do HIV-1 non-B subtypes differentially impact resistance mutations and clinical disease progression in treated populations? Evidence from a systematic review', *Journal of the International AIDS Society*. doi: 10.7448/IAS.17.1.18944.

Biesiadecki, B. J. and Jin, J. P. (2011) 'A high-throughput solid-phase microplate protein-binding assay to investigate interactions between myofilament proteins', *Journal of Biomedicine and Biotechnology*. doi: 10.1155/2011/421701.

Bor, J. *et al.* (2013) 'Increases in adult life expectancy in rural South Africa: Valuing the scale-up of HIV treatment', *Science*, 339(6122), pp. 961–965. doi: 10.1126/science.1230413.

Bozzette, S. a *et al.* (2003) 'Cardiovascular and cerebrovascular events in patients treated for human immunodeficiency virus infection.', *The New England journal of medicine*, 348(8), pp. 702–710. doi: 10.1056/NEJMoa022048.

Brik, A. and Wong, C.-H. (2003) 'HIV-1 protease: mechanism and drug discovery', *Organic & Biomolecular Chemistry*, 1(1), pp. 5–14. doi: 10.1039/b208248a.

Buonaguro, L. *et al.* (2007) 'DNA-VLP prime-boost intra-nasal immunization induces cellular and humoral anti-HIV-1 systemic and mucosal immunity with cross-clade neutralizing activity', *Vaccine*, 25(32), pp. 5968–5977. doi: 10.1016/j.vaccine.2007.05.052.

Chalmet, K. *et al.* (2010) 'Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections', *BMC Infectious Diseases*. doi: 10.1186/1471-2334-10-162.

Cihlar, T. and Ray, A. S. (2010) 'Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine', *Antiviral Research*, pp. 39–58. doi: 10.1016/j.antiviral.2009.09.014.

Ciuffi, A. *et al.* (2009) 'Methods for integration site distribution analyses in animal cell genomes', *Methods*. doi: 10.1016/j.ymeth.2008.10.028.

De Clercq, E. (2013) *The nucleoside reverse transcriptase inhibitors, nonnucleoside reverse*

transcriptase inhibitors, and protease inhibitors in the treatment of HIV infections (AIDS)., *Advances in pharmacology (San Diego, Calif.)*. doi: 10.1016/B978-0-12-405880-4.00009-3.

Coutsinos, D. *et al.* (2011) 'A template-dependent dislocation mechanism potentiates K65R reverse transcriptase mutation development in subtype C variants of HIV-1', *PLoS ONE*, 6(5). doi: 10.1371/journal.pone.0020208.

Divita, G., Restle, T. and Goody, R. S. (1993) 'Characterization of the dimerization process of HIV-1 reverse transcriptase heterodimer using intrinsic protein fluorescence', *FEBS Letters*. doi: 10.1016/0014-5793(93)81383-B.

Ferentz, A. E. and Wagner, G. (2000) 'NMR spectroscopy: A multifaceted approach to macromolecular structure', *Quarterly Reviews of Biophysics*. doi: 10.1017/S0033583500003589.

Fujita, M. *et al.* (2012) 'SAMHD1-dependent and -independent functions of HIV-2/SIV Vpx protein', *Frontiers in Microbiology*. doi: 10.3389/fmicb.2012.00297.

Gallo, R. C. and Montagnier, L. (2003) 'The Discovery of HIV as the Cause of AIDS', *New England Journal of Medicine*. doi: 10.1056/NEJMp038194.

Gilbert, P. B. *et al.* (2003) 'Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal', *Statistics in Medicine*. doi: 10.1002/sim.1342.

Gilski, M. *et al.* (2011) 'High-resolution structure of a retroviral protease folded as a monomer', *Acta Crystallographica Section D: Biological Crystallography*, 67(11), pp. 907–914. doi: 10.1107/S0907444911035943.

Goldman, D. P. and Bao, Y. (2004) 'Effective HIV Treatment and the Employment of HIV+ Adults', *Health Services Research*, 39(6p1), pp. 1691–1712. doi: 10.1111/j.1475-6773.2004.00313.x.

Handoko, S. D. *et al.* (2012) 'QuickVina: Accelerating AutoDock Vina using gradient-based

heuristics for global optimization', *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(5), pp. 1266–1272. doi: 10.1109/TCBB.2012.82.

Hankins, C. (2013) 'Overview of the current state of the epidemic', *Current HIV/AIDS Reports*, pp. 113–123. doi: 10.1007/s11904-013-0156-x.

Hemelaar, J. *et al.* (2011) 'Global trends in molecular epidemiology of HIV-1 during 2000-2007', *AIDS*, 25(5), pp. 679–689. doi: 10.1097/QAD.0b013e328342ff93.

Hemelaar, J. (2013) 'Implications of HIV diversity for the HIV-1 pandemic', *Journal of Infection*, pp. 391–400. doi: 10.1016/j.jinf.2012.10.026.

Van Houtte M, S. M. (2006) 'NRTI resistance associated with the RT mutation K70E in HIV-1', in *Drug Resistance Mutations*, pp. 31-33 (Special Contribution). doi: 10.1016/j.bbcan.2008.08.003.

Hu, W. S. and Hughes, S. H. (2012) 'HIV-1 reverse transcription.', *Cold Spring Harbor perspectives in medicine*. doi: 10.1101/cshperspect.a006882.

International Committee on Taxonomy of Viruses (2013) 'ICTV-Master-Species-List-2013', *Taxaonomy Releases*.

Invernizzi, C. F. *et al.* (2009) 'Signature nucleotide polymorphisms at positions 64 and 65 in reverse transcriptase favor the selection of the K65R resistance mutation in HIV-1 subtype C.', *The Journal of infectious diseases*, 200(8), pp. 1202–1206. doi: 10.1086/605894.

Ishima, R. *et al.* (2003) 'Solution Structure of the Mature HIV-1 Protease Monomer: Insight into the tertiary fold and stability of a precursor', *Journal of Biological Chemistry*, 278(44), pp. 43311–43319. doi: 10.1074/jbc.M307549200.

Ishima, R. *et al.* (2010) 'Highly conserved glycine 86 and arginine 87 residues contribute differently to the structure and activity of the mature HIV-1 protease', *Proteins: Structure, Function and Bioinformatics*, 78(4), pp. 1015–1025. doi: 10.1002/prot.22625.

Jacobo-Molina, A. *et al.* (1993) 'Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.90.13.6320.

Kear, J. L. *et al.* (2009) 'Subtype polymorphisms among HIV-1 protease variants confer altered flap conformations and flexibility', *Journal of the American Chemical Society*. doi: 10.1021/ja907088a.

Kim, E. E. *et al.* (1995) 'Crystal Structure of HIV-1 Protease in Complex with VX-478, a Potent and Orally Bioavailable Inhibitor of the Enzyme', *Journal of the American Chemical Society*, 117(3), pp. 1181–1182. doi: 10.1021/ja00108a056.

Kotler, D. P. (2008) 'HIV and antiretroviral therapy: lipid abnormalities and associated cardiovascular risk in HIV-infected patients.', *Journal of acquired immune deficiency syndromes (1999)*. doi: 10.1097/QAI.0b013e318186519c.

Kozal, M. J. *et al.* (2007) 'The Incidence of HIV drug resistance and its impact on progression of HIV disease among antiretroviral-naive participants started on three different antiretroviral therapy strategies', *HIV Clin Trials*, 8(6), pp. 357–370. doi: 55673120R0H2Q724 [pii]\r10.1310/hct0806-357.

Kozal, M. J. *et al.* (2012) 'A Nucleoside- and Ritonavir-Sparing Regimen Containing Atazanavir Plus Raltegravir in Antiretroviral Treatment-Naïve HIV-Infected Patients: SPARTAN Study Results', *HIV Clinical Trials*, 13(3), pp. 119–130. doi: 10.1310/hct1303-119.

Krishna, N. R. S. *et al.* (2011) 'Expression, purification and characterization of refolded rBm-33 (pepsin inhibitor homolog) from Brugia malayi: A human Lymphatic Filarial parasite', *Protein Expression and Purification*. Elsevier Inc., 79(2), pp. 245–250. doi: 10.1016/j.pep.2011.06.014.

Krishna, N. R. S. *et al.* (2013) 'Physicochemical characterization of an aspin (rBm-33) from a filarial parasite Brugia malayi against the important human aspartic proteases', *Journal of Enzyme Inhibition and Medicinal Chemistry*, 28(5), pp. 1054–1060. doi:

10.3109/14756366.2012.710849.

Lahouassa, H. *et al.* (2012) 'SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates', *Nature Immunology*. doi: 10.1038/ni.2236.

Lee, S. G. and Chmielewski, J. (2010) 'Cross-linked peptoid-based dimerization inhibitors of HIV-1 protease', *ChemBioChem*, 11(11), pp. 1513–1516. doi: 10.1002/cbic.201000248.

Life technology (no date) 'Reverse Transcription'. Available at: https://www.lifetechnologies.com/de/de/home/life-science/pcr/reverse-transcription.html.

Louis, J. M. *et al.* (1989) 'Substitution mutations of the highly conserved arginine 87 of HIV-1 protease result in loss of proteolytic activity', *Biochemical and Biophysical Research Communications*, 164(1), pp. 30–38. doi: 10.1016/0006-291X(89)91678-1.

Louis, J. M. *et al.* (2007) 'HIV-1 Protease: Structure, Dynamics, and Inhibition', *Advances in Pharmacology*, pp. 261–298. doi: 10.1016/S1054-3589(07)55008-8.

Lv, Z., Chu, Y. and Wang, Y. (2015) 'HIV protease inhibitors: a review of molecular selectivity and toxicity.', *HIV/AIDS (Auckland, N.Z.)*, 7, pp. 95–104. doi: 10.2147/HIV.S79956.

Mager, P. P. (2001) 'The active site of HIV-1 protease.', *Medicinal research reviews*, 21(4), pp. 348–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11410934.

Martínez-Cajas, J. L. *et al.* (2008) 'Role of genetic diversity amongst HIV-1 non-B subtypes in drug resistance: A systematic review of virologic and biochemical evidence', *AIDS Reviews*.

Mata-Munguía, C. *et al.* (2014) 'Natural polymorphisms and unusual mutations in HIV-1 protease with potential antiretroviral resistance: A bioinformatic analysis', *BMC Bioinformatics*, 15(1). doi: 10.1186/1471-2105-15-72.

Matúz, K. *et al.* (2012) 'Inhibition of XMRV and HIV-1 proteases by pepstatin A and acetyl-pepstatin', *FEBS Journal*, 279(17), pp. 3276–3286. doi: 10.1111/j.1742-4658.2012.08714.x.

McMahon, T. and Ward, P. R. (2012) 'HIV among immigrants living in high-income countries: A realist review of evidence to guide targeted approaches to behavioural HIV prevention', *Systematic Reviews*, 1(1). doi: 10.1186/2046-4053-1-56.

Menéndez-Arias, L. (2010) 'Special issue: Retroviral enzymes', *Viruses*. doi: 10.3390/v2051181.

Mitsuya, H. *et al.* (1985) '3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro.', *Proceedings of the National Academy of Sciences*, 82(20), pp. 7096–7100. doi: 10.1073/pnas.82.20.7096.

Morris, G. M. and Lim-Wilby, M. (2008) 'Molecular docking.', *Methods in molecular biology (Clifton, N.J.)*, 443, pp. 365–382. doi: 10.1007/978-1-59745-177-2_19.

Naicker, P. *et al.* (2013) 'Structural insights into the South African HIV-1 subtype C protease: Impact of hinge region dynamics and flap flexibility in drug resistance', *Journal of Biomolecular Structure and Dynamics*, 31(12), pp. 1370–1380. doi: 10.1080/07391102.2012.736774.

Navia, M. A. *et al.* (1989) 'Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1.', *Nature*, pp. 615–620. doi: 10.1038/337615a0.

Nomaguchi, M., Fujita, M. and Adachi, A. (2008) 'Role of HIV-1 Vpu protein for virus spread and pathogenesis', *Microbes and Infection*. doi: 10.1016/j.micinf.2008.07.006.

Okimoto, N. *et al.* (2000) 'Protein hydrolysis mechanism of HIV-1 protease: Investigation by the ab initio MO calculations', *RIKEN Review*.

Pai, N. P., Shivkumar, S. and Cajas, J. M. (2012) 'Does genetic diversity of HIV-1 non-B subtypes differentially impact disease progression in treatment-naive HIV-1-infected individuals? A systematic review of evidence: 1996-2010', *Journal of Acquired Immune Deficiency Syndromes*. doi: 10.1097/QAI.0b013e31824a0628.

Payne, B. A. I. *et al.* (2011) 'Mitochondrial aging is accelerated by anti-retroviral therapy through the clonal expansion of mtDNA mutations', *Nature Genetics*, 43(8), pp. 806–810. doi: 10.1038/ng.863.

Pearl, L. H. and Taylor, W. R. (1987) 'A structural model for the retroviral proteases.', *Nature*, 329(6137), pp. 351–4. doi: 10.1038/329351a0.

Perez, M. A. S., Fernandes, P. A. and Ramos, M. J. (2010) 'Substrate recognition in HIV-1 protease: a computational study.', *The journal of physical chemistry. B*, 114(7), pp. 2525–2532. doi: 10.1021/jp910958u.

Plantier, J. C. *et al.* (2009) 'A new human immunodeficiency virus derived from gorillas', *Nature Medicine*. doi: 10.1038/nm.2016.

Prabu-Jeyabalan, M., Nalivaika, E. and Schiffer, C. A. (2000) 'How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease', *Journal of Molecular Biology*. doi: 10.1006/jmbi.2000.4018.

Prabu-Jeyabalan, M., Nalivaika, E. and Schiffer, C. A. (2002) 'Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes', *Structure*. doi: 10.1016/S0969-2126(02)00720-7.

Roberts, J. D., Bebenek, K. and Kunkel, T. A. (1988) 'The accuracy of reverse transcriptase from HIV-1.', *Science (New York, N.Y.)*, 242(4882), pp. 1171–3. doi: 10.1126/science.2460925.

Røge, B. T. *et al.* (2003) 'Drug resistance mutations and outcome of second-line treatment in patients with first-line protease inhibitor failure on nelfinavir-containing HAART', *HIV Medicine*, 4(1), pp. 38–47. doi: 10.1046/j.1468-1293.2003.00133.x.

Sarafianos, S. G. *et al.* (2009) 'Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition', *Journal of Molecular Biology*. Elsevier Ltd, 385(3), pp. 693–713. doi: 10.1016/j.jmb.2008.10.071.

Sayer, J. M. and Louis, J. M. (2009) 'Interactions of different inhibitors with active-site aspartyl residues of HIV-1 protease and possible relevance to pepsin', *Proteins: Structure, Function and Bioinformatics*, 75(3), pp. 556–568. doi: 10.1002/prot.22271.

Shikuma, C. M. *et al.* (2001) 'Mitochondrial DNA decrease in subcutaneous adipose tissue of HIV-infected individuals with peripheral lipoatrophy', *AIDS*, 15(14), pp. 1801–1809. doi: 10.1097/00002030-200109280-00009.

Siliciano, R. F. and Greene, W. C. (2011) 'HIV latency', *Cold Spring Harbor Perspectives in Medicine*, 1(1). doi: 10.1101/cshperspect.a007096.

Sluis-Cremer, N. and Tachedjian, G. (2002) 'Modulation of the oligomeric structures of HIV-1 retroviral enzymes by synthetic peptides and small molecules', *European Journal of Biochemistry*. doi: 10.1046/j.1432-1033.2002.03216.x.

Sunpath, H. *et al.* (2012) 'High rate of K65R for antiretroviral therapy-naive patients with subtype C HIV infection failing a tenofovir-containing first-line regimen', *AIDS*, 26(13), pp. 1679–1684. doi: 10.1097/QAD.0b013e328356886d.

Suzuki, K. *et al.* (1993) 'Colorimetric reverse transcriptase assay for HIV-1', *Journal of Virological Methods*, 41(1), pp. 21–28. doi: 10.1016/0166-0934(93)90159-O.

Tebit, D. M. and Arts, E. J. (2013) 'From Simian to Human Immunodeficiency Viruses (SIV to HIV): Emergence from Nonhuman Primates and Transmission to Humans', in *The Role of Animals in Emerging Viral Diseases*. doi: 10.1016/B978-0-12-405191-1.00009-0.

Thomson, M. M., Pérez-Álvarez, L. and Nájera, R. (2002) 'Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy', *Lancet Infectious Diseases*. doi: 10.1016/S1473-3099(02)00343-2.

Todd, M. J., Semo, N. and Freire, E. (1998) 'The structural stability of the HIV-1 protease.', *Journal of molecular biology*, 283(2), pp. 475–488. doi: 10.1006/jmbi.1998.2090.

Toor, J. S. *et al.* (2011) 'Prediction of drug-resistance in HIV-1 subtype C based on protease sequences from ART naive and first-line treatment failures in North India using genotypic and docking analysis', *Antiviral Research*. Elsevier B.V., 92(2), pp. 213–218. doi: 10.1016/j.antiviral.2011.08.005.

Trott, O. and Olson, A. J. (2010) 'AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, EfficientOptimization, and Multithreading', *Journal of Computational Chemistry*, 31(2), pp. 455–61. doi: 10.1002/jcc.

UNAIDS (2017) 'Global Hiv Statistics — July 2017 Unaids', *Fact Sheet*, (NOVEMBER), pp. 18–25.

United States Food and Drug Administration (2014) *FDA-Approved HIV Medicines*, *aidsinfo.nih.gov*.

V., S. *et al.* (2014) 'HIV protease inhibitor exposure predicts cerebral small vessel disease', *AIDS*, pp. 1297–1306.

Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995) 'Ligplot - a Program To Generate Schematic Diagrams of Protein Ligand Interactions', *Protein Engineering*, 8(2), pp. 127–134. doi: 10.1093/protein/8.2.127.

Weber, I. T. (1990) 'Comparison of the crystal structures and intersubunit interactions of human immunodeficiency and rous sarcoma virus proteases', *Journal of Biological Chemistry*, pp. 10492–10496.

Weber, I. T. and Agniswamy, J. (2009) 'HIV-1 Protease: Structural perspectives on drug resistance', *Viruses*, pp. 1110–1136. doi: 10.3390/v1031110.

Wensing, A. M. J., van Maarseveen, N. M. and Nijhuis, M. (2010) 'Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance', *Antiviral Research*, pp. 59–74. doi: 10.1016/j.antiviral.2009.10.003.

Wilkinson, E. *et al.* (2016) 'Origin, imports and exports of HIV-1 subtype C in South Africa: A historical perspective', *Infection, Genetics and Evolution*, 46, pp. 200–208. doi: 10.1016/j.meegid.2016.07.008.

Wlodawer, A. and Vondrasek, J. (1998) 'INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design', *Annual Review of Biophysics and Biomolecular Structure*, 27(1), pp. 249–284. doi: 10.1146/annurev.biophys.27.1.249.

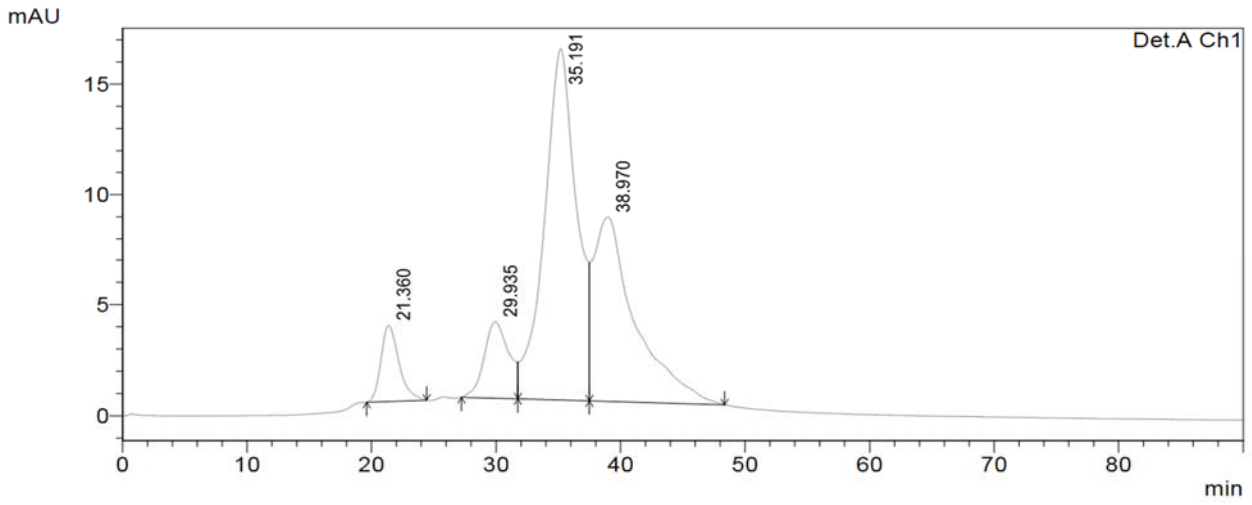
Zhang, R. *et al.* (1993) 'The time of administration of 3'-azido-3'-deoxythymidine (AZT) determines its host toxicity with possible relevance to AZT chemotherapy', *Antimicrobial Agents and Chemotherapy*, 37(9), pp. 1771–1776. doi: 10.1128/AAC.37.9.1771.

van Zyl, G. U. *et al.* (2011) 'Antiretroviral resistance patterns and factors associated with resistance in adult patients failing NNRTI-based regimens in the western cape, South Africa', *Journal of Medical Virology*. doi: 10.1002/jmv.22189.

APPENDIX

Appendix A

Standard SEC peaks, indicative of the times where a protein of a particular weight is expected to be eluted (Figure A1) were obtained from a SEC run of standard proteins.



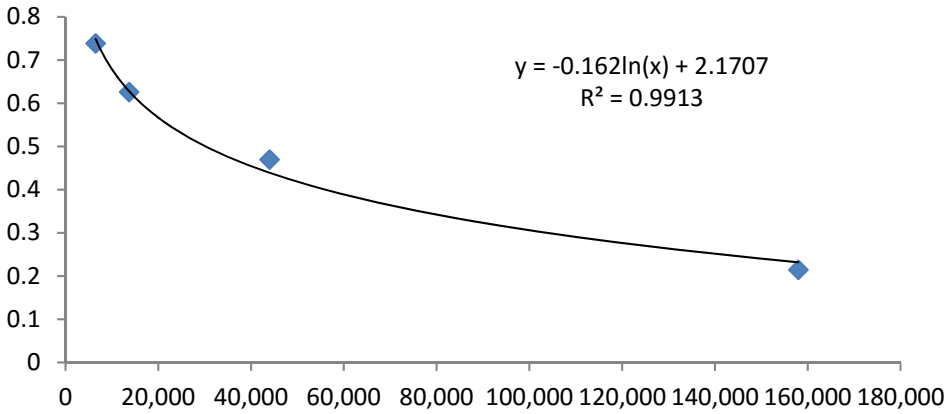
Appendix A1: A typical size exclusion chromatogram of protein standards: Aprotinin, 10 mg; Ribonuclease A, 50 mg; Ovalbumin, 50 mg; and Aldolase, 950 mg, eluting at 21, 30, 35 and 40 minutes

Molecular weights of standards presented on Appendix A2 were provided with the standards from the manufacturer. The standards Kav values were calculated from the elution volumes from the run.

Appendix A2: Retention factor and molecular weights for the standards

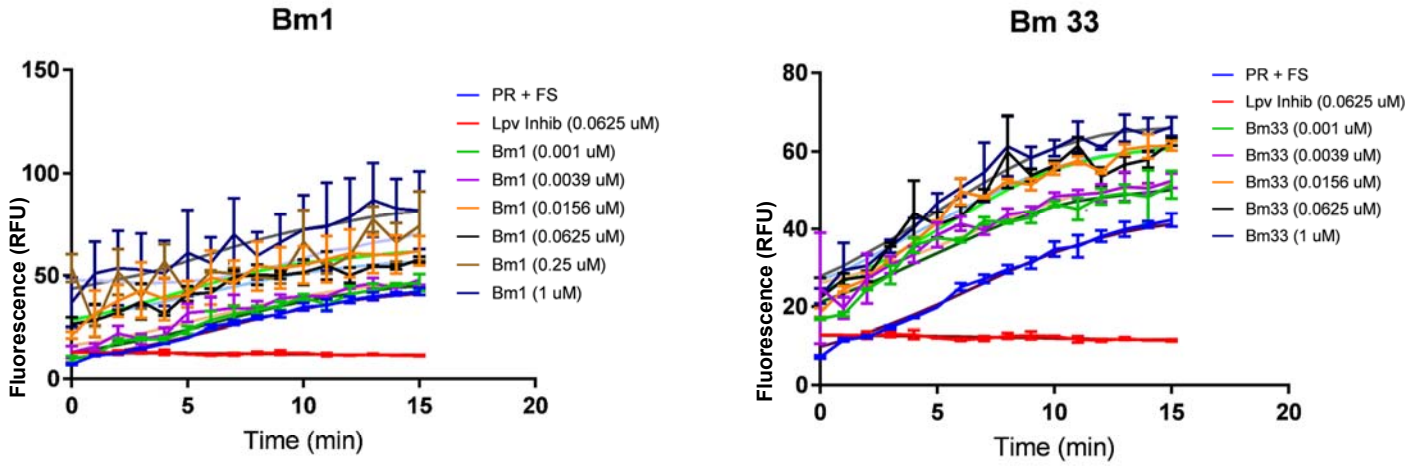
Standards	Time (min)	Ve (µl)	Kav	Mr (Da)
Aprotinin	21.36	6.408	0.738254168	6, 500
Ribonuclease A	29.935	8.9805	0.625809587	13, 700
Ovalbumin	35.191	10.5573	0.469416701	43, 000
Aldolase	38.97	11.691	0.214266586	158, 000

The equation, x = (2.1707- y)/ (0.6122) was generated from a standard curve plotted from the standards run (Figure A1) and was used to determine the molecular weights for the expressed enzymes in chapters 2 and 3 of this work.



Appendix A3: A size exclusion chromatogram standard curve plotted from the standards run from figure 3.5.5A. Aprotinin at 6500 Da; Ribonuclease A, 13700 Da; Ovalbumin, 43000 Da and Aldolase, 158000 Da. Molecular weights are shown in Table 3.5.5A.

Appendix B



Appendix B: Inhibition plots of Bm1 and Bm33 peptides to illustrate non-reactiveness of the peptides against HIV protease. Bm1 represented peptide 1 and Bm33 representing all four peptides combined. Plots of peptides 3,5 and 7 not included here since similar results were obtained. The peptides were seen to enhance rather than inhibiting protease reaction with an increase in their concentration. The red plot is indicative of protease inhibition by lopinavir at 0.0625 µM. The reaction of substrate catalysis by protease increases over time of incubation (blue). The same is also observed even in the presence of the peptides (all plots appearing above the reaction).