HIV-1 subtype C proteases: overexpression, structural, kinetic and thermodynamic characterisation

Mihai-Silviu Tomescu

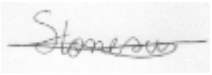
A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science.

Johannesburg, 2016

Declaration

I, Mihai-Silviu Tomescu (387090), am a student registered for the degree of Masters in the academic year II.
I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/ or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Thesis/Dissertation/Research Report has not been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

signature  16th day of February 2016

Dedication

I dedicate the entity of this work to those who have shared with me the joyous moments and the rough patches that I encountered throughout the course of this study.

Special thanks to the people I love:

My grandmother, Elena Popescu, who inspired me throughout my life. She has always shown unbelievable kindness and understanding towards others.

My mother and my father, Mihaela and Marius Tomescu, who showed me support for all the life-changing decisions that I have made.

My older brother, Dragos Tomescu, who has taken the time to guide me and encouraged me to explore the paths ahead of me.

And much love for Selisha Sooklal, who has always been there for me when I needed her. She has always encouraged me to pursue what I wanted to do and to reach my full potential. Her dedication to her work and career-choices were inspirational.

Abstract

According to UNAIDS, there are ~36.9 million people infected with HIV-1 in the world. Of those, 25.8 million live in sub-Saharan Africa and 6.8 million in South Africa. HIV-1 subtype C accounts for over 95% of HIV infections in South Africa. HIV-1 retrovirus acquires mutations rapidly because of the viral reverse transcriptase. Naturally occurring polymorphisms distinguishing wild type C-SA PR from other proteases make it less susceptible to inhibitors. E35D↑G↑S is a C-SA PR variant with a double insertion in the flap region of the protease. The insertions and background mutations may decrease susceptibility to inhibitors as well as alter kinetic parameters due to increased flap flexibility. This study intended to characterise the effect of the mutations and insertions in E35D↑G↑S on structural, kinetic activity and drug susceptibility. Chemically-synthesised E35D↑G↑S autocatalyses rapidly, impeding further characterisation. There was no detectable overexpression of the E35D↑G↑S protease in *Escherichia coli* BL21 (DE3)pLysS and Rosetta 2® cells. If the protease is catalytically enhanced, attributed cytotoxicity may prevent overexpression of the protein. Increased autocatalytic activity could also prevent crystallisation. Inactive E35D↑G↑S D25A did not overexpress either, indicating that codon harmonisation with the expression host ought to be performed. C-SA PR was shown to be a predominantly beta-sheeted protein using circular dichroism spectroscopy. The K_M of the fluorogenic substrate resembling the capsid/ p2 cleavage site for C-SA PR was 22.02 ±4.09 µM. The specific activity, catalytic turnover and catalytic efficiency of the wild-type C-SA PR protease were found to be 35.68 ±1.06 µmole.min^{-1}.mg^{-1}, 12.79 ±0.38 s^{-1} and 1.17 ±0.055 s^{-1}.µM^{-1}, respectively. The thermodynamics of binding of atazanavir, ritonavir and darunavir to C-SA PR were determined using isothermal titration calorimetry. The binding of atazanavir and ritonavir to C-SA PR is entropically driven and enthalpically opposed. However, the binding of darunavir to C-SA PR was found to be both entropically and enthalpically favourable. The dissociation constants of the inhibitors in the absence of substrate (K_d) are in the pico-molar range and increased by approximately one order of magnitude when saturating concentrations of substrate were introduced. Atazanavir, ritonavir and darunavir have dissociation constants (K_d) of 160.56 ±54.59 pM, 113.34 ±46.47 pM and 10.24 ±6.02 pM, respectively. Darunavir binds significantly tighter.

Keywords: C-SA PR, E35D↑G↑S, insertion mutations, protease, autocatalysis, ITC.

Acknowledgements

I would firstly like to thank Professor Lynn Morris (National Institute for Communicable Diseases) for kindly providing the E35D↑G↑S gene sequence. I would like to thank Professor Fernando Albericio (University of Barcelona) who supplied the chemically synthesised E35D↑G↑S. I would especially like to thank Professor Yasien Sayed for sound advice, guidance, encouragement and friendly input. I am also grateful for funding from the Postgraduate Merit Award (University of the Witwatersrand) and National Research Foundation (NRF). I would also like to thank the staff and students of the Protein Structure-Function Research Unit (University of the Witwatersrand, South Africa), namely, Dr. Sylvia Fanucchi, Dr. Ikechukwu Achilonu, Professor Heini Dirr, Previn Naicker, Alison Williams and Gavin Morris for countless conversations during our collegial relationship.

Table of Contents

List of abbreviations

3TC	Lamivudine
AIDS	Acquired immunodeficiency syndrome
ART	Antiretroviral therapy
ATV	Atazanavir
CA	Capsid protein
C-SA PR	HIV-1 South African subtype C protease
d4T	Stavudine
Da	Daltons
DNA	Deoxyribonucleic acid
DMSO	Dimethyl-sulfoxide
DRV	Darunavir
DTT	Dithiothreitol
E35D↑G↑S	HIV-1 South African subtype C protease E35D↑G↑S
EDTA	Ethylenediaminetetraacetic acid
Far-UV CD	Far ultraviolet circular dichroism
FDA	Food and Drug Administration
ΔG	Change in Gibbs free energy
ΔH	Change in enthalpy
HAART	Highly active antiretroviral therapy
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2
HPLC	High performance liquid chromatography
IC_{50}	Inhibitor concentration reducing enzyme activity by 50%
IN	Integrase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
k_{cat}	Catalytic turnover
$k_{cat/}K_M$	Catalytic efficiency, specificity constant
K_{eq}	Equilibrium constant
K_d	Dissociation constant of enzyme-inhibitor complex
kDa	Kilo-Dalton

K_i	Inhibition constant/ drug dissociation constant in the presence of substrate
K_M	Michaelis-Menten substrate-enzyme complex dissociation constant
kPa	kilo Pascals
LB	Lysogeny broth
MA	Matrix protein
mAu	milli absorbance units
mdeg	milli degrees
MRE	Mean residue ellipticity
N	Stoichiometry
NC	Nucleocapsid protein
$O.D_{600}$	Optical density at 600 nm
PCR	Polymerase chain reaction
PDB	Protein data bank
PI	Protease inhibitor
PMSF	phenylmethylsulfonyl fluoride
PR	Protease
Rev	Regulator of expression of virion protein
RNA	Ribonucleic acid
RT	Reverse transcriptase
RTV	Ritonavir
ΔS	Change in entropy
SIV	Simian immunodeficiency virus
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Tat	Trans-activator of transcription
T_m	Transition/ melting temperature
V_{max}	Maximum velocity
Vpr	Lentivirus protein R

List of figures

List of tables

1. Introduction

1.1 Discovery and identification of HIV as the causative agent of AIDS

The term *acquired immunodeficiency syndrome* (AIDS) was introduced on the 27th of July 1982 based on the nature of the illness. It was then accepted that AIDS did not discriminate between race, gender and sexual orientation (Marx, 1982). One year later, Luc Montagnier and associated researchers at the Pasteur Institute in France isolated the first strain of the *human immunodeficiency virus* (HIV) (Bare-Sinoussi *et al*., 1983). In 1984, HIV was undoubtedly proven to be the causative agent of AIDS by Robert Gallo's group (Safai *et al*., 1984). However, the virus itself was termed HIV in 1986; before that, the virus variants were classified under a different taxonomy (Coffin *et al*., 1986[1]; Coffin *et al*, 1986[2]).

1.2 Taxonomic classification of HIV

The HIV species belongs to the *Lentivirius* genus of the *Orthoretrovirinae* subfamily that is encompassed in the *Retroviridae* family (International Committee on Taxonomy of Viruses: http://ictvonline.org/virusTaxonomy.asp accessed on 22nd January 2015). There are contradicting hypotheses regarding the emergence of both types of HIV (HIV-1 and HIV-2) (Sharp *et al*., 2006). Nevertheless, it is accepted that HIV-2 originates from the *simian immunodeficiency virus* (SIV) that infects the sooty mangabey and, the more virulent HIV-1, originates from the SIV that infects chimpanzees (Sharp *et al*., 2001; Keele *et al*., 2006, Sharp and Hahn, 2011). It is speculated that the virus initially infected a human host via ingestion of SIV positive chimpanzee meat. However, the exact origin of HIV is uncertain (Fauci, 2003).

1.3 HIV-1 variants

HIV-1 is divided amongst groups M, N, O and P according to the most common recent ancestor of each group (Tebit and Arts, 2011). Most of the infections with HIV are due to the constituent subtypes of group M. There are nine subtypes making up group M; those are: A, B, C, D, F, G, H, J and K. There are also circulating recombinant forms (CRFs) that emerge due to the recombination of viruses of different subtypes within hosts infected with multiple subtypes. Additionally, there are significant primary-structure differences between variants that belong within each subtype (Hemelaar *et al*., 2006; Hemelaar, 2012).

1.4 Estimates of HIV infections

The most recent estimate made by UNAIDS was in 2014 and it revealed that the number of people infected with HIV in the world is peaking at 36.9 million (UNAIDS, 2015; http://www.unaids.org/en/resources/documents/2015/HIV_estimates_with_uncertainty_bounds_10-2014 , last updated on 12th August 2015, accessed on 11th October 2015). UNAIDS re-estimates HIV populations each year using updated modules and new assumptions (personal communication with the aidsinfo team from UNAIDS). Thus, trailing research publications that present estimates over the years, gives the impression that the number of HIV-infected patients is decreasing. On the contrary, the number of HIV infected people in the world displays a rising trend (UNAIDS, 2015). However, the infection rate has decreased (Weiss, 2008). Nevertheless, sub-Saharan Africa is currently a hotspot for the transmission of HIV (UNAIDS, 2015).

Sub-Saharan Africa is home to 25.8 million people infected with HIV and 6.8 million of those people reside in South Africa (UNAIDS, 2015). In a study done by Hemelaar, 2012, it was shown that HIV-1 subtype C is the most common of the subtypes, accounting for 48% of the global infections and for over 95% of the infections in the South African region (Hemelaar, 2012; UNAIDS 2015).

1.5 General physiology of HIV-1

The mature HIV retrovirus encloses two identical, positive-sense, single strands of RNA which dimerise. Each of the ~9 kb RNA strands encodes the full viral genome (Frankel & Young, 1998; Ahlquist *et al*., 2003; Takahashi *et al*., 2000). The viral RNA is reverse transcribed into dsDNA by the reverse transcriptase (RT) of the virus (Frankel & Young, 1998). It is the process of reverse transcription that notoriously creates numerous variations in the viral proteins that lead to altered drug susceptibility. This is because the viral reverse transcriptase (RT) does not have proofreading capabilities. As such, the enzyme is ten times less accurate than other viral polymerases (Roberts, 1988)

The viral RNA has an upstream promoter located in the 5' long terminal repeat (LTR) which is recognised by the viral protein Tat (trans-activator of transcription). Tat increases the rate of transcription of the viral genome and is also essential in the translation of the viral proteins (Jeang *et al*., 1999; Kao *et al*., 1987). The transport of RNA is assisted by the lentivirus protein R (Vpr) while the regulator of expression of virion proteins (Rev) coordinate

variations in the splicing of RNA so that the different viral proteins may be produced (Frankel & Young, 1998; Freed, 1998; Schubert & McClure, 2005; Terwilliger *et al*., 1989).

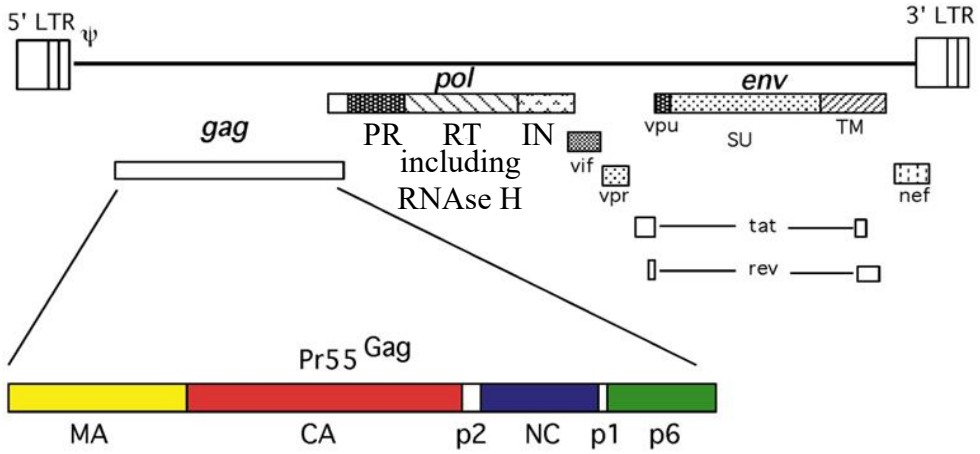
The viral genome translates into 15 proteins, three of which are enzymes (Figure 1 A and B) (Frankel & Young, 1998). The Gag polyprotein chain contains structural proteins such as the matrix protein (MA), capsid protein (CA), nucleocapsid protein (NC) and the structural proteins p1, p2 and p6 which provide structural support for the polyprotein chain. The Pol polyprotein chain contains the three enzymes: reverse transcriptase (RT), integrase (IN) and protease (PR). The envelope polyprotein (Env) consists of the surface protein gp120 (SU) and the transmembrane protein gp41 (TM). The viral protein U (Vpu) overlaps with Env at the 5' end and is translated from a different frame (Hout *et al*., 2004). The virion infectivity factor (Vif), Vpr, Tat, Rev and the negative regulatory factor (Nef) are spliced out independently (Bogerd *et al*., 2004; Frankel & Young, 1998; Freed, 1998; Niederman *et al*., 1989; Terwilliger *et al*., 1989).

The Env proteins; namely SU and TM are synthesised at an earlier stage than the rest of the viral proteins, within the endoplasmic reticulum (ER) of the host cell (Freed, 1998). Nef down-regulates surface CD4 after Env binding and viral entry, and Vpu degrades intracellular CD4 to prevent newly formed Env from binding to CD4 (Frankel & Young, 1998; Terwilliger *et al*., 1989). The Env proteins are then embedded into the host plasma membrane via the secretory pathway (Freed, 1998). The envelope proteins, SU and TM are linked in a non-covalent manner after proteolytic cleavage by cellular proteases (Binley *et al*., 2002). Ultimately, proteolytic maturation of the virus is completed after the budding of the virus (Richman, 2001). However, it is the virion infectivity factor that inhibits the human host defence factor APOBEC3G, thus, preventing detrimental viral hypermutation and ensuring the budding of infectious viral particles (Land *et al*., 2008; Bogerd *et al*., 2004).

1.6 HIV – host membrane fusion

The SU protein recognises the host cell CD4 receptor and binds it. The chemokine co-receptors, CCR5 and CXCR4, may both interact with the viral SU, thus, SU and the host receptors/ co-receptors facilitate the attachment of the virus to the host (Freed, 1998; Colman *et al.,* 2003). However, it is the TM protein that brings together and fuses the viral membrane with the host membrane after attachment (Figure 2) (Freed, 1998; Colman *et al.,* 2003, Teixeira *et al.,* 2011).

A



B

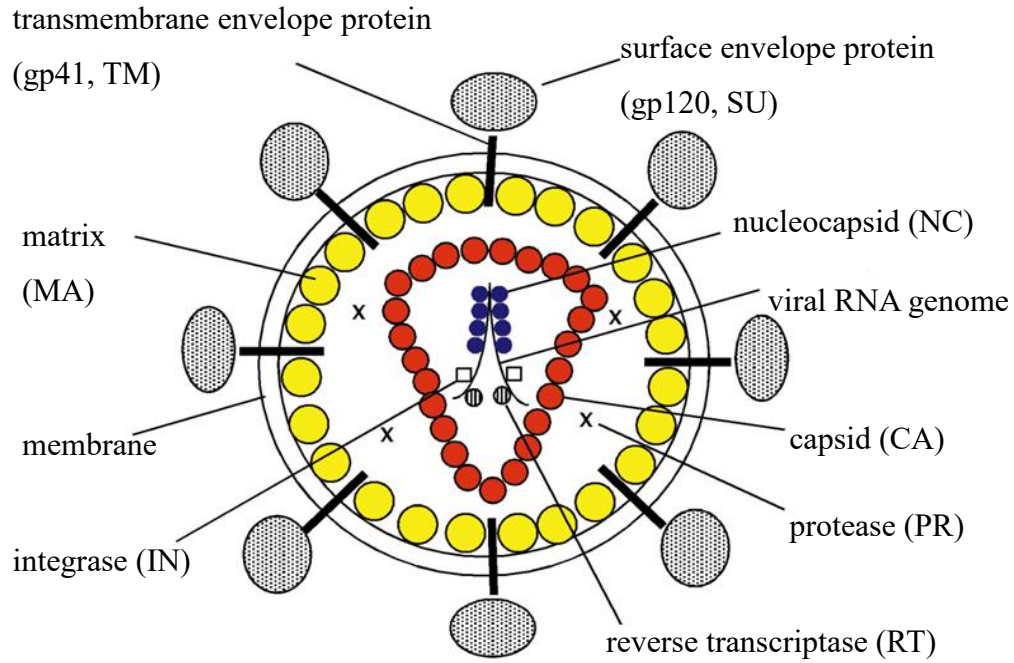


Figure 1: (A) Organisation of the HIV-1 genome and (B) schematic representation of the structure of HIV-1.

HIV-1 contains the matrix, capsid, p2, nucleocapsid, p1, p6, protease, reverse transcriptase including the RNAse H domain, integrase, Vif, Vpr, Vpu, Env proteins (SU and TM), Tat, Rev and Nef. Adapted from Freed, 1988.

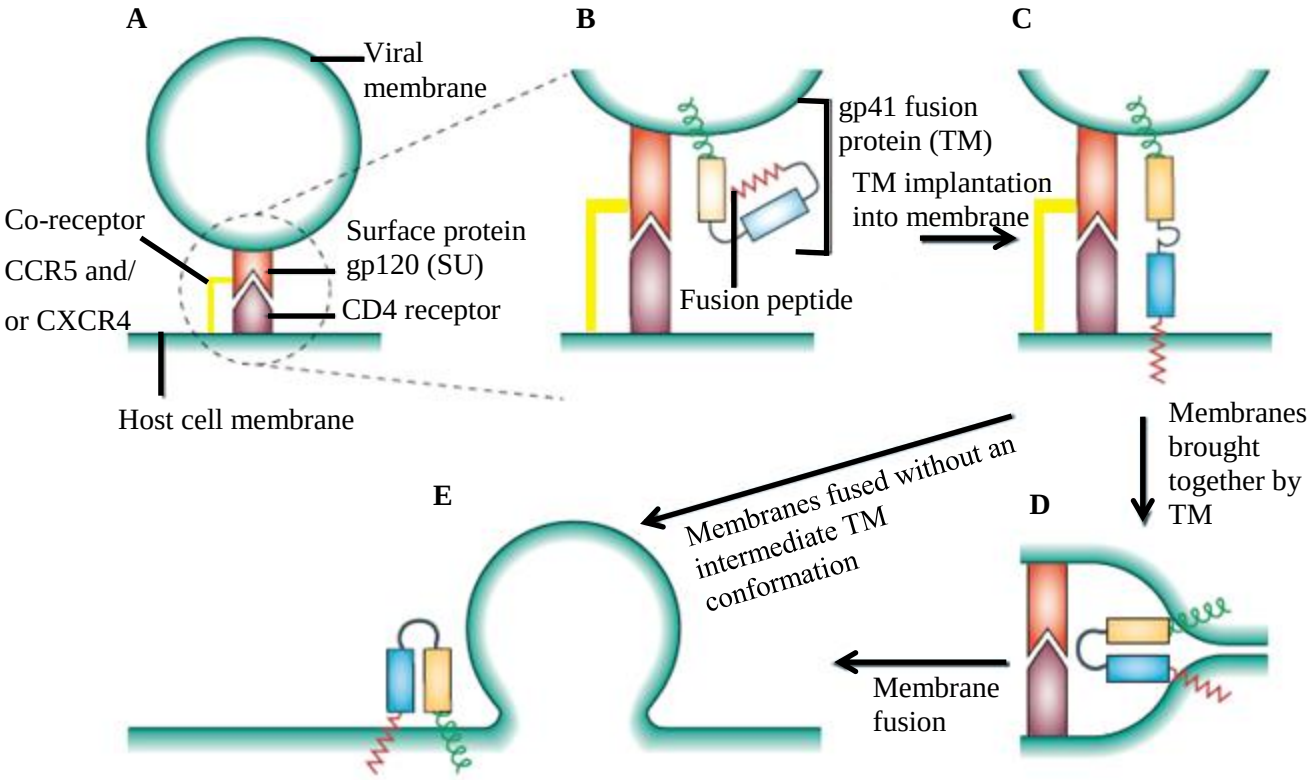


Figure 2: Fusion of the viral membrane with the host cell membrane

(**A**) The surface protein gp120 (SU), attaches to the CD4 receptor and the co-receptor(s) which may be CXCR4 or CCR5. (**B**) The transmembrane protein gp41 (TM) folds out and implants itself in the membrane. (**C**) The membranes are either brought together by a TM intermediate formation or by a direct mechanism which is not fully understood. However, it may well exist, in which case, step **D** does not occur, but only step **E** where the membranes have fused by a more direct mechanism. However, it is common that the structural conformation of TM is radically modified for the fusion of membranes to occur (Teixeira *et al.*, 2011). Adapted from Colman *et al*., 2003.

1.7 The life-cycle of HIV-1

After the viral – host membrane fusion, the matrix proteins dissipate quickly inside the host cell to release the viral core formed by numerous capsid proteins. In turn, the CA proteins scatter rapidly to release the viral RNA and the encapsulated enzymes RT, IN and PR as well as the multiple proteins necessary for the life-cycle of the virus (Frankel & Young, 1998). The MA and CA proteins are structural proteins that also assist in the intra-host transport of the viral RNA and enzymes. Moreover, the MA and CA proteins assist in the budding of the viral particles (Frankel & Young, 1998).

The three virally encoded enzymes RT, IN and PR are at the epicentre of the HIV life-cycle (Figure 3). Subsequent to the release of the three enzymes, 12 proteins and RNA into the cytoplasm of the host, the ssRNA is reverse transcribed by RT into ssDNA. The degradation of the ssDNA is prevented by the Vif protein. The ssDNA is duplexed by the polymerase ability of the RT while the RNAse H domain of the RT degrades the genomic RNA of the virus (Beilhartz and Götte, 2010). The dsDNA is transported to the nucleus where it is integrated into the DNA of the host by the viral IN. Once integrated, the viral genome (known as provirus when integrated) is transcribed by the host cell RNA polymerase II (RNA pol II). The transcribed provirus is then spliced and translated for the production of the early viral proteins such as the Env proteins and the polyprotein chains, Gag and Pol. These proteins are subsequently encapsulated in the budding virus alongside the two RNA molecules encoding the viral genome. The encapsulation of the polyproteins is assisted by Vpr. Furthermore, the encapsulation of the RNA is driven by the NC present in the polyprotein chains. During and after budding of the virus, the viral particles become infective once proteolytic maturation is fully performed by the PR.

Proteolytic maturation involves the cleavage of the polyprotein chains so that functional viral proteins can fold and assemble inside the virus so that the virus may become infective (Frankel & Young, 1998; Freed, 1998). Before the viral PR can perform the maturation function of the viral particles, the PR monomers have to dimerise. First, two Pol polyprotein chains align next to one another to dimerise, forming a functional protease which auto-processes from the polyprotein chain. The protease then cleaves the rest of the polyprotein chains into the functional proteins necessary for viral maturation (Louis *et al*., 1994).

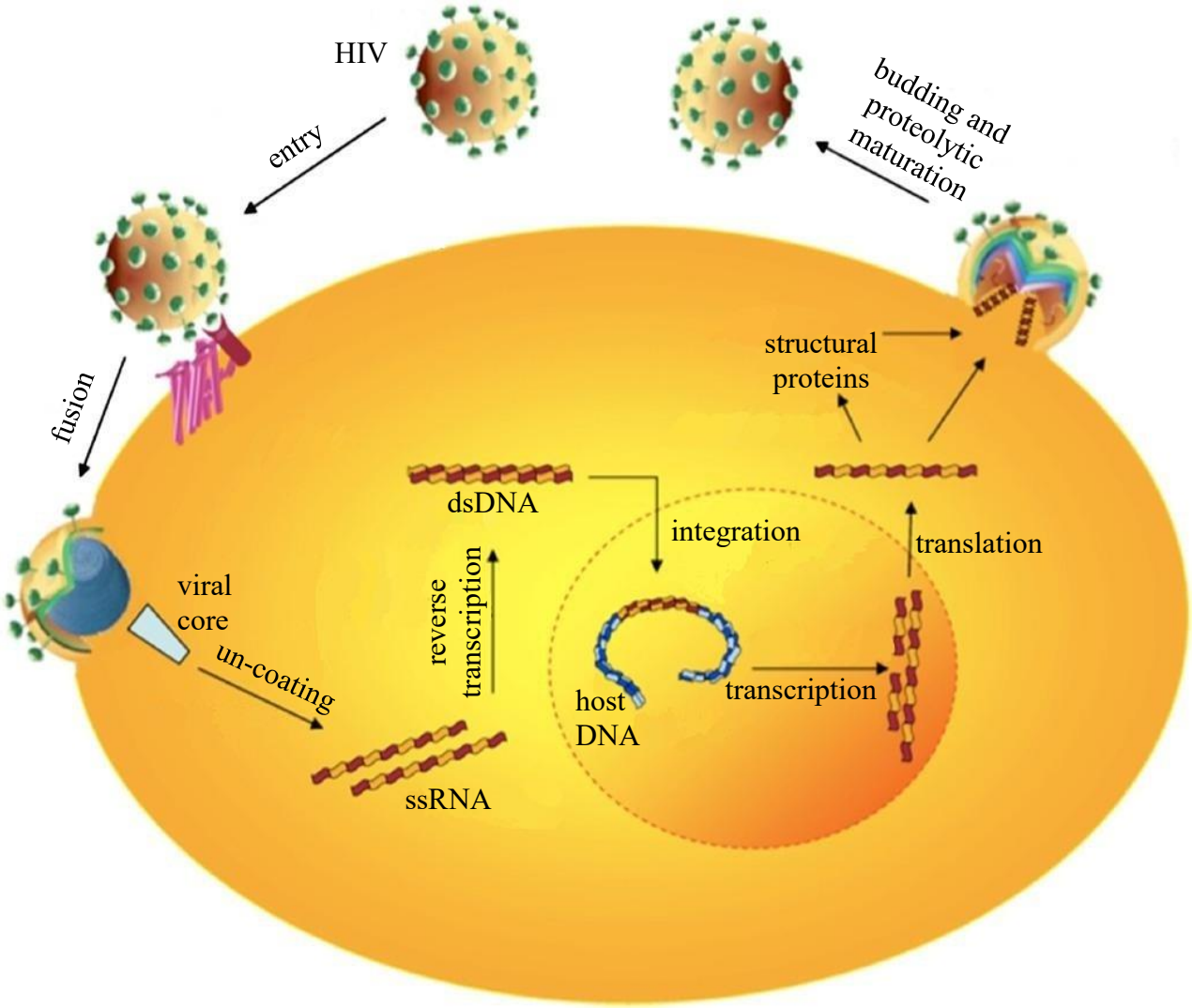


Figure 3: The life-cycle of HIV-1

The virus fuses with the host releasing the viral core which unpacks to liberate the viral RNA, enzymes and proteins. The ssRNA is converted to dsDNA by RT. The dsDNA is integrated into the host genome by IN. The provirus is transcribed and translated. The translated proteins assemble and are proteolytically cleaved for viral maturation as soon as budding occurs. Adapted from Teixeira *et al*., 2011.

The enzymes and proteins of HIV-1, including, RT, IN and PR are critical to the life-cycle of the virus. Antiretroviral therapy (ART) exploits this phenomenon (Frankel & Young, 1998; Freed, 1998; Vella *et al*., 2012).

1.8 Current therapy for AIDS

Since the discovery of HIV, there have been many advances made to deal with HIV and the chronic onset of AIDS. Antiretroviral therapy involves the administration of a combination of at least three antiretroviral drugs from at least two drug classes (FDA – Approved HIV Medicines: http://aidsinfo.nih.gov/education-materials/fact-sheets/21/58/fda-approved-hiv-medicines# last updated and reviewed on 12th October 2014; accessed on 3rd February 2015). Currently there are twenty-five antiretroviral inhibitors approved by the Food and Drug Administration (FDA) to be used in antiretroviral therapy. These are distributed amongst the following drug classes: nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), fusion inhibitors, entry inhibitors and integrase inhibitors. In addition, the FDA has approved the use of a pharmacokinetic enhancer to assist therapy. Antiretroviral therapy is also referred to as highly active antiretroviral therapy (HAART) (World Health Organisation – Antiretroviral therapy: http://www.who.int/topics/antiretroviral_therapy/en/. accessed on 9th February 2016).

1.9 HIV-1 South African subtype C protease (C-SA PR) and protease inhibitors

Central to this project is the HIV-1 South African subtype C protease (C-SA PR) and a double insertion variant of the protease. C-SA PR, like all HIV proteases, is a homodimer. Each monomer consists of 99 amino acids and has a subunit molecular mass of 10754.8 Da (~11 kDa). The isoelectric point of the protein is 9.32 (Naicker, 2013; Mosebi, 2008; Gasteiger *et al*., 2005). The secondary structure of the protease consists of 48% β-sheets, 4% α-helices and 48% random coils. The enzyme folds into a globular protein (Naicker, 2013; PDB ID: 3U71). The enzyme is an aspartic protease (EC 3.4.23.16) which has a conserved catalytic triad: Asp25, Thr26 and Gly27. Asp25 is the catalytic residue, while Thr26 and Gly27 provide attachment and support for the substrate. The tertiary structure of the protease assists the catalytic function. The flaps of the protease (residues 46 to 54 in C-SA PR) are conserved. Their function is to open and allow a substrate to enter the protease active site and close around the substrate to be cleaved. The motion of the flaps is guided by the hinge

regions (residues 35 to 42 and 57 to 61 in C-SA PR) (Wlodawer *et al*., 1989; Wlodawer and Vondrasek, 1998) (Figure 4 A).

HIV proteases are vital in the proteolytic maturation of the virus. Without the protease, the polyprotein chains of the virus will not be cleaved into functional proteins and enzymes (Louis *et al*., 1994). The inhibition of the viral protease by protease inhibitors (PIs) prevents viral maturation (Figure 4 B). Thus far, protease inhibitors have proven to be useful in HIV treatment because they are prescribed, amongst other antiretroviral inhibitors, in highly active antiretroviral therapy (HAART) (Richman, 2001; De Clercq, 2006, 2009).

HIV-1 retroviral proteases acquire mutations frequently due to the lack of proofreading ability of the viral reverse transcriptase. Beneficial variants are naturally selected for, as they confer drug resistance and improved viability of the virus. HIV-1 South African subtype C protease contains naturally occurring polymorphisms that make it different from other subtypes, such as the B subtype protease for which PIs were designed. The polymorphisms that make C-SA PR different from other subtypes also lower the susceptibility of the virus to PIs (Mosebi *et al,* 2008). Thus, antiretroviral protease inhibitors are comparatively less effective for patients harbouring C-SA PR and the corresponding virus subtype (Velázquez-Campoy *et al,* 2003; Mosebi *et al,* 2008). Additionally, subtype C proteases have been documented to be the most drug-resistant of the subtypes (Velázquez-Campoy *et al*., 2001).

1.10 Double insertion mutant C-SA PR E35D↑G↑S

The amino acid sequence for HIV-1 South African subtype C protease double insertion mutant E35D↑G↑S (henceforth known as E35D↑G↑S) was obtained from patient #2421 by Professor Lynn Morris (Head of the AIDS Unit at the National Institute for Communicable Diseases, South Africa). The patient received treatment with the nucleoside reverse transcriptase inhibitors (NRTIs), d4T (stavudine) and 3TC (lamivudine). The patient also received treatment with the non-nucleoside reverse transcriptase inhibitor (NNRTI), efavirenz. However, the patient was drug-naïve with respect to protease inhibitors. Keeping that in mind, it is important to investigate whether E35D↑G↑S has acquired mutations that weaken the interaction of the protease with protease inhibitors. Also, questions are raised concerning the effect of the insertions on the kinetic behaviour of the protease.

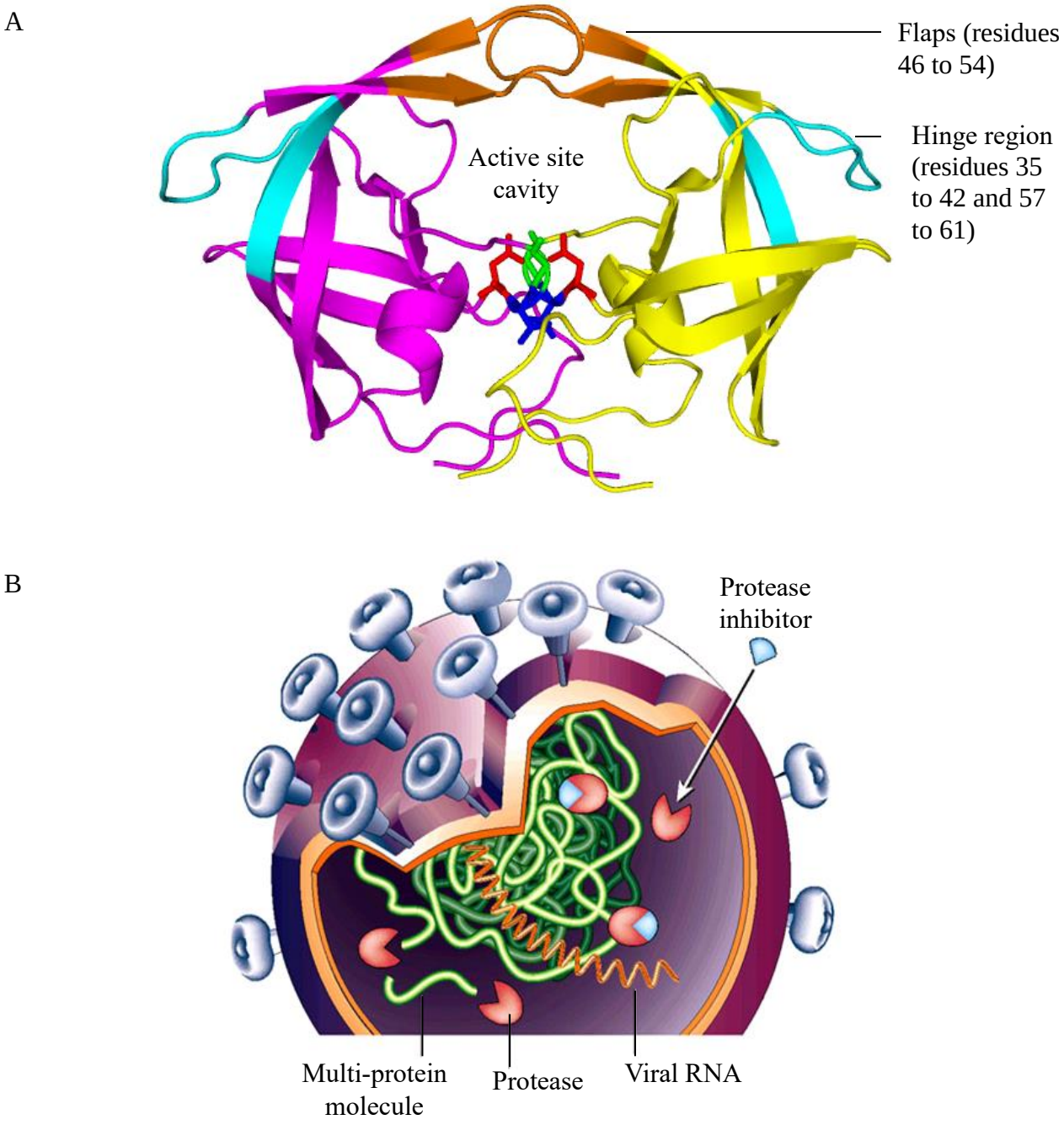


Figure 4: (A) HIV-1 South African subtype C protease and (B) Representation of an immature HIV-1 particle

(A) C-SA PR in ribbon outlay (PDB ID 3U71). The two chains are yellow and magenta respectively. The hinge (cyan) and the flap (orange) regions are shown. The catalytic triad is Asp25 (red), Thr26 (blue) and Gly27 (green). Asp25 is the catalytic residue while Thr26 and Gly27 offer structural support to the substrate. The figure was generated with PyMOL v0.99 (DeLano, 2002). (B) An immature HIV particle, consisting of unprocessed polyprotein chains due to the inhibition of the PRs by PIs. The figure was adapted from Richman, 2001.

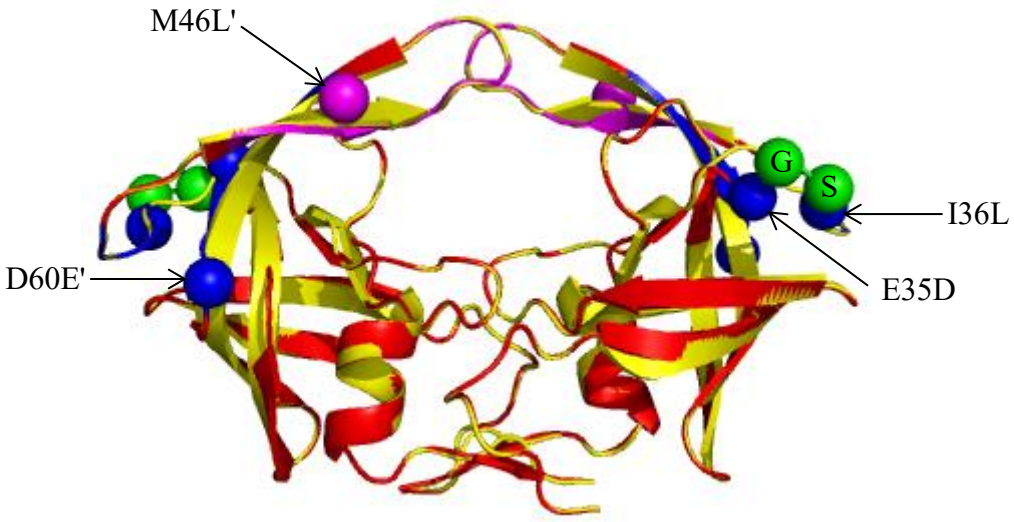
The mutations in E35D↑G↑S occur in the hinge and flap regions of the protease. The background mutation M46L occurs in the flap while E35D, I36L and D60E occur in the hinge region (Figure 5 A and B). In addition, the two insertions occur in the hinge region of the protease. The hinge regions rotate the flaps, directing the opening and closing of the active site; therefore, governing the ease of access of substrates and inhibitors (Naicker *et al*., 2014). Insertions in the hinge region, as well as, flap dynamics have been found to impact on the catalytic rate of proteases and their affinity for various substrates and inhibitors (Naicker *et al*., 2013; Naicker *et al*., 2014; Kozísek *et al*., 2008). In addition, mutations acquired in the recognition sequence of the viral polyprotein cleavage sites may provide better complementarity to the mutations in proteases. Overall, the acquisition of mutations can become beneficial to the virus (Doyon *et al*., 1996; Zhang *et al*., 1997; Mammano *et al*., 1998; Kozísek *et al*., 2008).

Three of the background mutations (I36L, M46L, and D60E) have been documented to make the binding of current protease inhibitors such as atazanavir (ATV) and ritonavir (RTV) inferior. Consequently, it is possible that E35D↑G↑S may exhibit reduced susceptibility to ATV and RTV. However, none of the mutations in E35D↑G↑S have been found to reduce susceptibility to darunavir (DRV) (Wensing *et al*., 2014; Standford University HIV Drug Resistance Database updated on the 9th March 2015, accessed on 24th August 2015 at: http://hivdb.stanford.edu/pages/download/resistanceMutations_handout.pdf).

The homology structure prediction of E35D↑G↑S was generated using SWISS-MODEL (Figure 5 A) (Biasini *et al*., 2014; Arnold *et al*., 2006; Kiefer *et al*., 2009; Guex *et al*., 2009; Schwede *et al*., 2003). The amino acid sequence of the wild-type C-SA PR was aligned with that of E35D↑G↑S using ClustalW 2.1 (Larkin *et al*., 2007) (Figure 5 B).

A Procheck (Laskowski *et al*., 1993, Laskowski *et al*., 2001) analysis of the E35D↑G↑S variant revealed that the percentage composition of beta sheets and alpha helices is 54% and 6%, respectively, which is a slightly larger proportion when compared to C-SA PR (46% β-sheeted and 4% alpha helical). E35D↑G↑S has a molecular mass of 10880.9 Da (~11 kDa) and a p*I* = 9.32 (Gasteiger *et al*., 2005).

A



B

```
C-SA        PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEE--INLPGKWKPKMIGGIGGFIKVRQ 58
E35DGS      PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEDGSLNLPGKWKPKLIGGIGGFIKVRQ 60
            **********************************:  :*********:************

C-SA        YDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF 99
E35DGS      YEQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF 101
            *:*****************************************
```

Figure 5: (A) SWISS-MODEL homology structure prediction of E35D↑G↑S alignment with C-SA PR and (B) ClustalW 2.1. sequence alignment of C-SA PR with E35D↑G↑S.

C-SA PR (PDB ID 3U71 – yellow) is the wild type protease. The model prediction of the E35D↑G↑S variant was made using the PDB ID 3U71 structure as a template. The hinge regions of E35D↑G↑S are coloured in blue and the flaps are coloured in magenta. The insertions occur in the beginning of the hinge region between E35 and I36 (labelled blue spheres). The protein structure alignment (A) was produced using PyMOL v0.99 (DeLano, 2002).

1.11 Active site of HIV-1 protease

The active site of HIV-1 protease spans across eight subsites. In a directional order, the subsites are arranged as follows: S4, S3, S2, S1, S1', S2', S3' and S4' (prime notation denominates two chains). Namely: D29, G48, G27, and a H_2O molecule co-ordinated by I50 and I50', which positions the attached substrate in reach of the two catalytic residues, D25 and D25'. The co-ordinated water molecule provides hydrogen-bond donors to make up both S1 and S1'. Thereafter, G27', G48' and D29' follow to constitute S2', S3' and S4' (Figure 6). The subsites make contacts with the following annotated positions of the peptide substrate: P4, P3, P2, P1, P1', P2' P3' and P4'. The cleavage of the substrate occurs at the catalytic site between P1 and P1' where D25 and D25' are found (Schechter and Berger, 1967; Wlodawer and Erickson, 1993; Wlodawer and Vondrasek, 1998). The number and strength of the hydrogen bonds made with the substrate may well determine the affinity of the enzyme for the substrate. Similarly, the affinity for PIs may also be affected.

1.12 Catalytic mechanism of HIV-1 protease

The active site of the protease contains Asp25 and Asp25', respectively, involved in the acid-base hydrolysis of the substrate. Each of the residues belongs to one of the two chains, respectively. Only one of the two residues is protonated. This is a requirement for the binding of substrate. The protonation of a single residue is made possible by an increase in the pK_a of the protonated aspartic acid residue from 3.1 to 5.2. Thus, the residue has a propensity to keep the bonded proton (Hyland *et al*., 1991). The protonated aspartic acid plays a role in hydrogen bonding of the substrate. The deprotonated aspartic acid residue deprotonates a water molecule which, in turn, nucleophilically attacks the carbonyl carbon of the scissile peptide bond recognised for hydrolysis. The reaction results in the formation of a tetrahedral transition-state intermediate (hydroxyethylamine isostere pivotal in drug design); whereby the carbonyl oxygen is now one electron deficient of the octet. The charged oxygen makes a double-bond with the carbonyl carbon thereby separating it from the amide end which was simultaneously protonated by the aspartic acid that has initially deprotonated the water molecule. In the end, the initially deprotonated aspartic acid which was subsequently protonated, is restored to the deprotonated state, ready to catalyse the hydrolysis of another scissile peptide bond (Figure 7) (Brik and Wong, 2003).

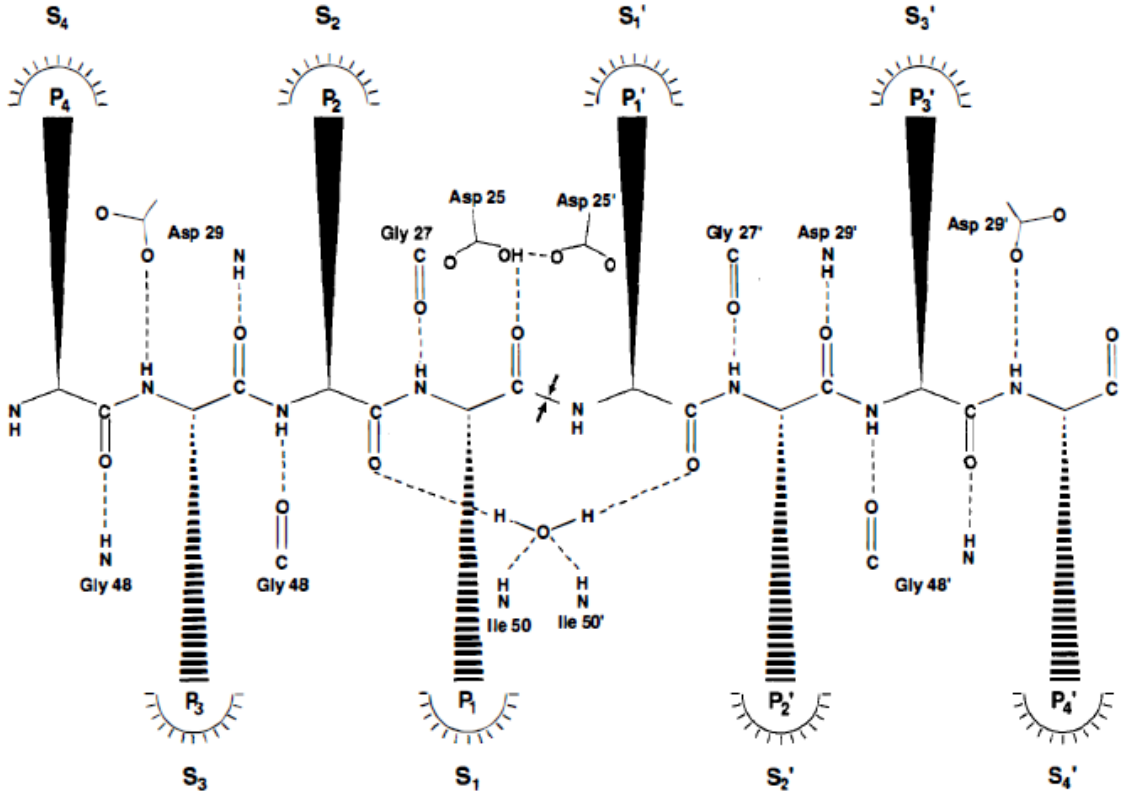


Figure 6: Schematic representation of the active site of HIV-1 protease

The subsites on the enzyme are denoted by S4, S3, S2, S1, S1', S2', S3' and S4'. The corresponding positions taken by the peptide substrate are denoted: P4, P3, P2, P1, P1', P2' P3' and P4'. The amino acids forming subsites on each side of Asp25 and Asp25' help position the substrate while the catalytic aspartic acid residues cleave the scissile peptide bond between P1 and P1'. The prime annotation distinguishes two chains in the homodimeric protein. The figure was taken from Wlodawer and Erickson, 1993.

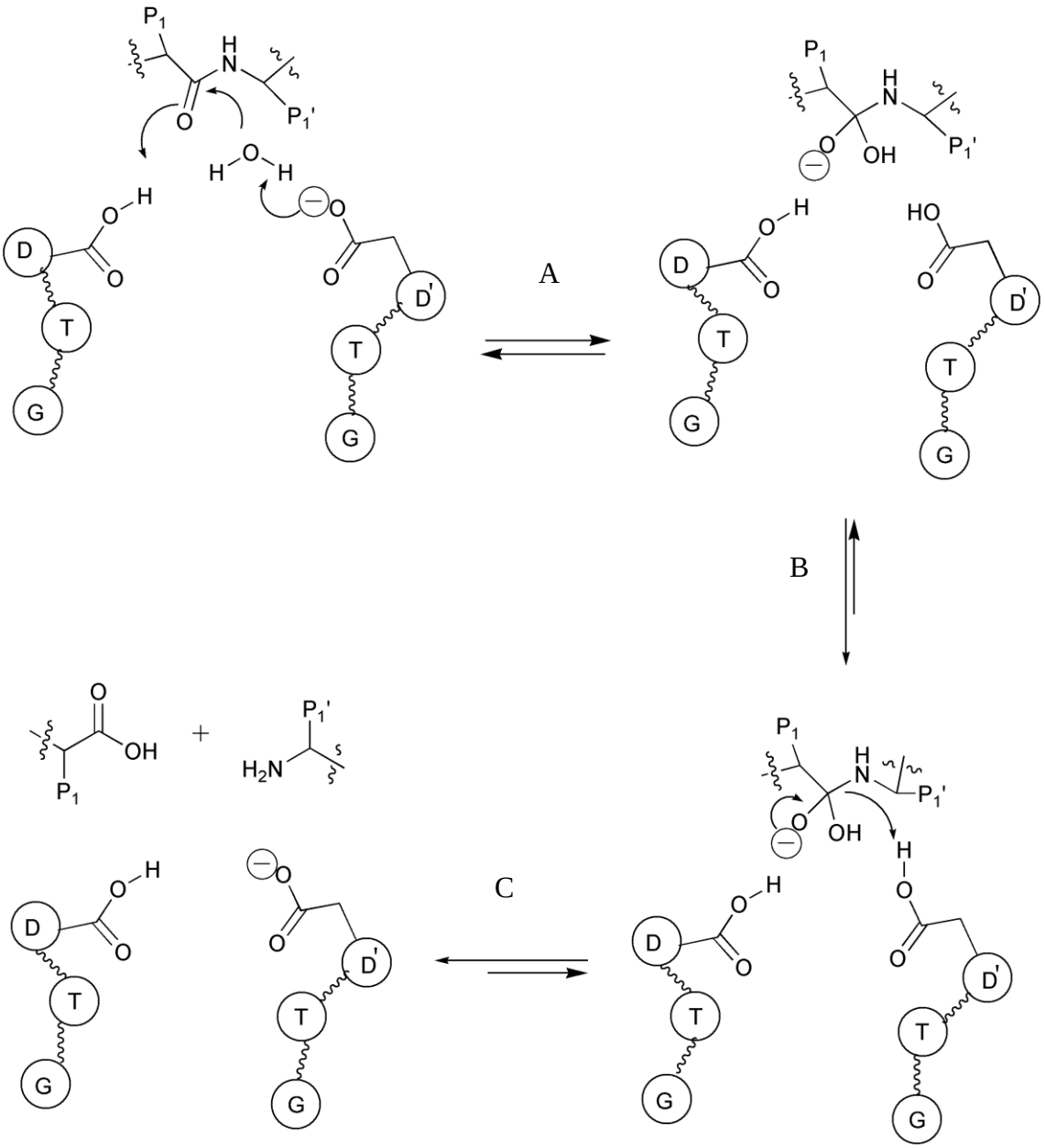


Figure 7: Catalytic mechanism of HIV-1 protease

(A) Deprotonation of H_2O by charged D', and nucleophilic attack on the carbon involved in the scissile peptide bond by the free hydroxyl group. Thus, the tetrahedral intermediate is formed. (B) Nucleophilic attack on carbonyl carbon by charged carbonyl oxygen, leaving the nitrogen atom charged, which in turn, nucleophilically attacks the hydrogen present on D' (from H_2O). (C) The hydrolysed product is released by protonation of the leaving N-terminus of the cleaved scissile peptide bond (P1'). The protease resets and is ready to catalyse again. The figure was adapted from Brik and Wong, 2003.

1.13 Protease inhibitors to be tested

Protease inhibitors were designed to mimic transition-state intermediates formed within a sequence recognised by HIV-1 proteases. However, they were designed so that the transition-state intermediates could not be cleaved by HIV-1 proteases. Accordingly, a common feature of PIs is the presence of hydroxyethylamine isosteres. Hydroxyethylamine isosteres do not undergo peptide bond cleavage and are exclusively selected to mimic cleavage sites between Tyr-Pro and Phe-Pro, in order to achieve inhibitor specificity for HIV-1 protease (Wlodawer and Vondrasek, 1998).

First generation PIs bind to proteases in an entropically driven manner ($-T\Delta S < 0$) and experience enthalpic resistance ($\Delta H > 0$). Second generation PIs were designed to bind to HIV-1 proteases both enthalpically ($\Delta H < 0$) and entropically favourable (Velázquez-Campoy *et al*., 2001). Additionally, HIV-1 proteases are less effective at acquiring resistance to second generation PIs that bind to the backbone because mutations would not change the binding sites of the PIs with the protein backbone (Ghosh *et al*., 2007[1]; Ghosh *et al*., 2007[2]). However, the distance from PIs to the binding sites on the proteases may shift due to chemical properties brought upon by mutations. More likely than that, contact sites would shift due to insertion mutations.

Due to the insertion mutations in E35D↑G↑S, second generation PIs are favoured for testing in this investigation. Darunavir binds to the backbone of HIV-1 proteases which makes it challenging for HIV-1 proteases to acquire drug resistance against darunavir (Ghosh *et al*., 2007[1]; Ghosh *et al*., 2007[2]). Darunavir (Figure 8) was shown to bind to C-SA PR enthalpically favourable (ΔH = -79.91 kJ.mol^{-1}) and entropically disfavoured ($-T\Delta S$ = 8.93 kJ.mol^{-1}) (Naicker *et al*., 2014). However, darunavir was shown to bind in an entropically favourable manner to HIV-1 proteases in multiple instances (Kožíšek *et* al., 2014; Mittal *et al*., 2013). Nevertheless, it is intriguing to discover whether the insertion mutations of E35D↑G↑S would convey resistance to backbone-binding second generation PIs. Since the double insertion mutant protease, E35D↑G↑S, has no mutations that have been documented to provide resistance to darunavir (Wensing *et al*., 2014; Standford University HIV Drug Resistance Database updated on the 9th March 2015, accessed on the 24th August 2015 at: http://hivdb.stanford.edu/pages/download/resistanceMutations_handout.pdf), the testing of darunavir may provide critical insight into drug resistance due to insertion mutations.

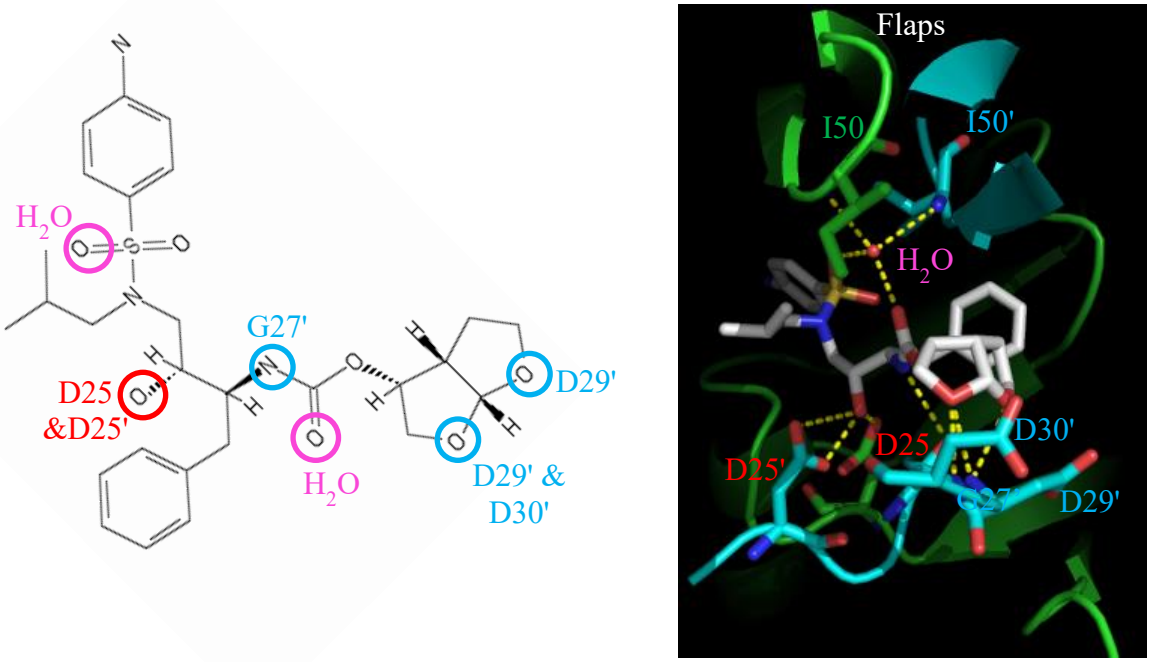


Figure 8: Darunavir complexed with HIV-1 protease (PDB ID 3OY4)

Darunavir is hydrogen bonded to the backbone of the HIV-1 proteases subtype B variant. Moreover, the drug makes contacts with a coordinated water molecule at the conserved S1 and S1' subsites (typical of PI binding). The oxygen atom in the hydroxyethylamine isostere is a hydrogen bond acceptor for the donor hydrogen on the protonated catalytic aspartic acid (D25 or D25' labelled in red). The hydrogen bond partners on darunavir are circled and annotated with the contacted residues. Darunavir makes 11 hydrogen bonds with the protease subsites in this instance. The chemical structure of darunavir was produced using PubChem Sketcher V2.4 (accessed at http://pubchem.ncbi.nlm.nih.gov/edit2/index.html, on 18th August 2015) with SMILES notation obtained from Drugbank ID DB01264. The crystal structure of complexed darunavir was analysed and displayed using PyMOL V0.99. Labels are colour-coded with the chain of the homodimeric enzyme.

Atazanavir (Figure 9 A) and ritonavir (Figure 9 B) bind exothermically to C-SA PR, as it is expected of second generation backbone binding PIs (Velázquez-Campoy *et al*., 2001; Velázquez-Campoy *et al*., 2003). However, ritonavir was also shown to bind endothermically (Naicker *et al*., 2014). This could be because the experimentally determined binding enthalpy is close to zero. Thus, data fitting errors could indicate favourable or unfavourable enthalpy. Nevertheless, there is reasonable certainty that atazanavir and ritonavir bind in an entropically driven manner.

The size of the inhibitors may affect binding energetics to proteases. Darunavir being a smaller molecule (MW = ~547 g/mol) binds enthalpically driven because it makes more contacts than it disrupts water molecules bonded to the protein (Kožíšek *et* al., 2014; Mittal *et al*., 2013). Atazanavir (MW = ~704 g/mol) and ritonavir (MW = ~720 g/mol) are relatively large inhibitors. Indeed the large size of the inhibitors contributes to the positive entropy generated upon binding, due to the displacement of ordered water molecules. At the same time though, the displacement of too many ordered water molecules around the protease active site may be enthalpically disfavoured. That is because disruption of hydrogen bonds requires energy. This can justify why the two drugs may bind endothermically.

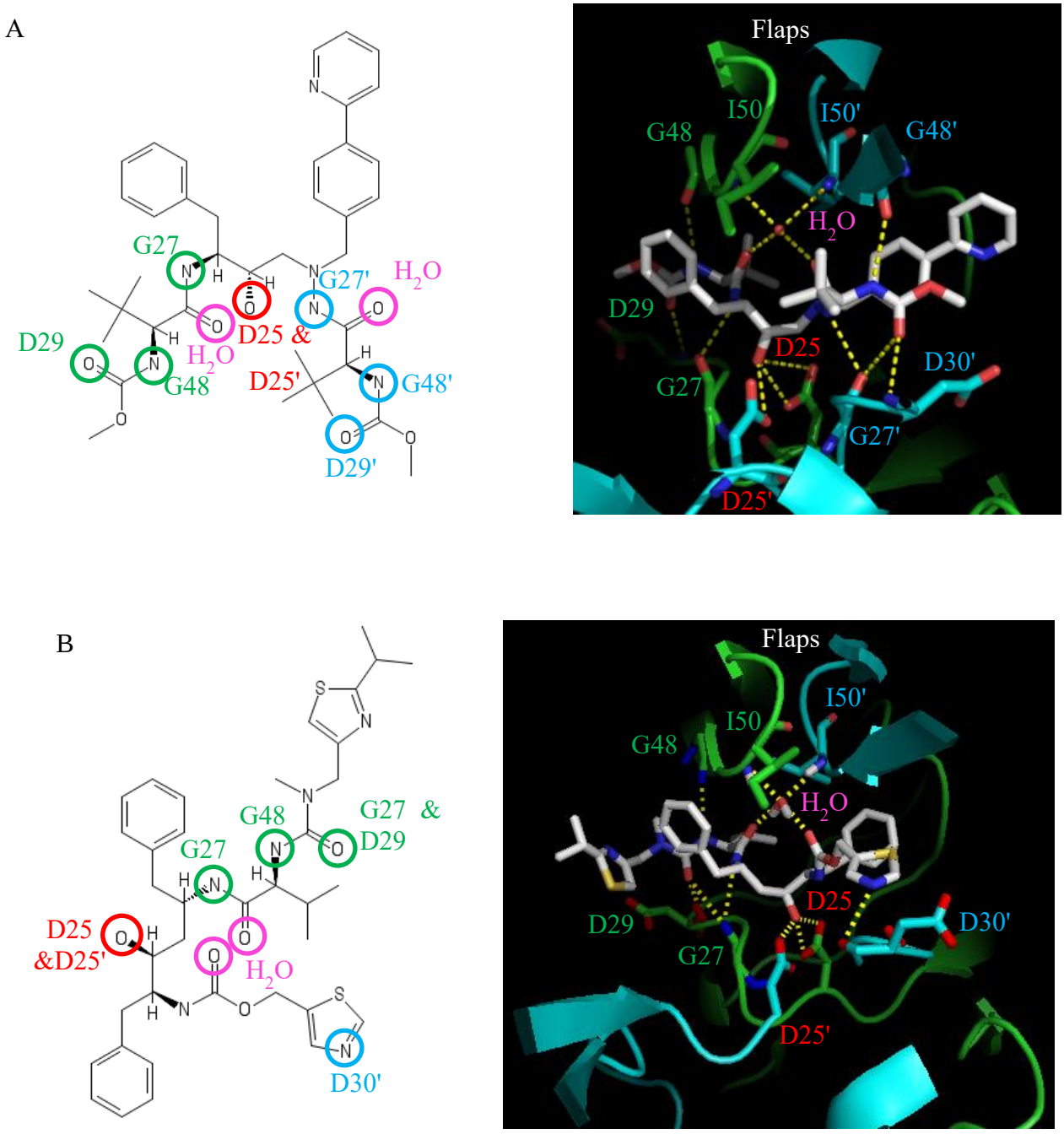


Figure 9: Binding of atazanavir (A) and ritonavir (B) to HIV-1 protease

Atazanavir (Drugbank ID DB01072; PDB ID 3EKY) and ritonavir (Drugbank ID DB00503; PDB ID 1HXW) bind the backbone of HIV-1 protease subtype B. The drugs are larger in comparison to darunavir. Therefore, the drugs may displace more water molecules bonded to the protein, requiring energy, making the binding less favourable enthalpically than darunavir. Atazanavir makes 12 hydrogen bonds with the subsites while ritonavir makes 10. The hydrogen bonding partners of the two inhibitors are circled and annotated with the bonded residues. The water molecule providing the S1 and S1' subsites, is labelled in magenta. The labels are colour-coded with the chains and catalytic residues.

2 Aim and Objectives

2.1 Aim

The aim of the proposed experimental research was to determine the effect of the mutations observed in E35D↑G↑S in comparison to the wild type C-SA PR with regards to structure, kinetic activity and drug resistance. This would shed light on the effect of hinge insertions on the drug binding mechanism of HIV-1 protease subtype C as well as HIV-1 proteases in general.

2.2 Objectives

- Solubilisation of lyophilised E35D↑G↑S
- Secondary structure characterisation using far-UV circular dichroism
- Tertiary structure characterisation using intrinsic fluorescence spectroscopy
- Kinetic characterisation (specific activity; K_M; k_{cat}; k_{cat} / K_M) monitoring cleavage of substrate over time
- IC_{50} and K_i determination monitoring cleavage of substrate over time
- Binding thermodynamic characterisation using isothermal titration calorimetry

3 Methods and Materials

3.1 Materials

The protease inhibitors used in this research, namely, atazanavir, ritonavir and darunavir were obtained from the NIH AIDS Reagent Programme. The aspartic protease inhibitor, acetyl pepstatin, was obtained from MINTEK, South Africa. The fluorogenic substrate used (Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2) was obtained from Synpeptide®, China. Ampicillin, chloramphenicol and PMSF were obtained from Sigma-Aldrich®, South Africa.

3.2 Solubilisation of chemically synthesised, lyophilised E35D↑G↑S and degradation studies

A lyophilised sample of chemically synthesised E35D↑G↑S was kindly provided by Professor Fernando Albericio (University of Barcelona, Spain). The lyophilised sample did not dissolve in deionised water or activity buffer (10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide). It was assumed that the chemically synthesised protein was misfolded. Subsequently, the protein was unfolded in 10 mM Tris buffer, pH 8.00, containing 0.02% sodium azide, 2 mM dithiothreitol (DTT) as a reducing agent and 6 M guanidinium chloride or 8 M urea as a denaturant. Refolding was achieved via overnight dialysis. However, the amount of enzyme present was barely detectable on SDS-PAGE (described in section 3.2) following overnight dialysis. Moreover, degradation was visible suggesting autocatalytic activity.

Degradation studies were set up to determine the shortest dialysis period required for the E35D↑G↑S protease to refold and become active, in order to limit the time permitting autocatalysis. Degradation studies involved the variation of the period allowed for dialysis-assisted refolding. The protease was unfolded in 10 mM Tris-HCl buffer, pH 8.00, containing 2 mM DTT and 8 M urea. The unfolded protease was dialysed against 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide. Dialysed protease samples were taken at 0, 0.25, 0.5, 0.75, 1, 2, 3, 4, 5, 6, 7 and 24 hours and immediately placed in reducing sample buffer for SDS-PAGE (described in section 3.2). Densitometry analyses of the electrophoresed gels were performed using GelAnalyzer. The dialysis time was reduced to 1 hour and 15 minutes. It was then necessary to determine the percentage of protein that remained active after 1 h and 15 minutes of dialysis. Thus, the enzyme was titrated with acetyl pepstatin using active site titration calorimetry (described in section 3.12.1).

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

SDS-PAGE is a common analytical technique described by Laemmli (Laemmli, 1970). SDS-PAGE separates proteins by size since the sodium-dodecyl sulfate imparts proportional weight and negative charge to proteins based on the size of each protein species. Proteins are thus pulled to the anode when electric current is applied. It has been documented in a modified SDS-PAGE protocol, published by Schägger, 2006, that tricine makes a better trailing ion than glycine as it achieves better separation of proteins in the lower molecular weight range (smaller than 20 kDa) (Schägger, 2006; Haider *et al*., 2011). The monomeric molecular weight of the HIV-1 proteases used in this study is approximately 11 kDa (Gasteiger *et al*., 2005). Therefore, tricine SDS-PAGE was used to assess the presence and to a lesser extent, the purity of proteins according to Schägger (Schägger, 2006). Gels (10% and 16%), sample buffer (12% SDS, 6% β-mercaptoethanol, 30% glycerol, 0.05% Coomassie blue G-250, 150 mM Tris-HCl pH 7.00), fixing and staining solution (0.025% Coomassie G-250, 50% methanol and 10% acetic acid) and de-stain solution (10% acetic acid) were prepared. The cathode buffer contained 100 mM Tris, 100 mM Tricine and 0.1% SDS. The anode buffer contained 100 mM Tris-HCl pH 8.9. A 10% polyacrylamide gel was initially used to electrophorese the lyophilised E35D↑G↑S sample before and after overnight dialysis-assisted refolding. Thereafter, 16% polyacrylamide gels were used for further studies. The Page Ruler Prestained Protein Ladder (Thermoscientific®) was used as the molecular weight maker.

3.4 *Escherichia coli* BL21(DE3)pLysS cells and pET-11b vector encoding C-SA PR variants

Escherichia coli BL21(DE3)pLysS is a protein expression cell line particularly useful when expressing cytotoxic proteins. That is because the T7 lysozyme, encoded by the pLysS lysogen, interferes with the elongation function of the T7 RNA polymerase encoded by the DE3 lysogen. Thus, basal expression of (cytotoxic) proteins is minimised (Moffatt & Studier, 1987; Stano & Patel, 2004; Studier & Moffatt, 1986). The pLysS lysogen also contains a selectable marker that confers resistance to chloramphenicol. Chloramphenicol (35 μg/ μl) was used in all experimental cultures inoculated with *E. coli* BL21(DE3)pLysS.

The pET-11b vector (Novagen®) is used for the expression of recombinant proteins. The pET-11b expression vector confers resistance to ampicillin. Therefore, all the broth and agar plates that were used for inoculating or plating BL21(DE3)pLysS transformed with the

pET-11b vector were supplemented with 100 μg/ μl ampicillin. The vector also houses a T7 promoter upstream from the recombinant gene inserted into the multiple cloning site. Protein overexpression in BL21(DE3)pLysS transformed with pET-11b occurs by inducing overexpression of the T7 RNA polymerase (harboured in the DE3 lysogen) with isopropyl β-D-1-thiogalactopyranoside (IPTG).

The C-SA PR construct, incorporating the Q7K autocatalysis-resistant mutation (Mildner *et al*., 1994) was produced in pET-11b plasmid in our laboratory via site-directed mutagenesis by Mosebi (Mosebi *et al*., 2008). However, the Q7K mutation does not completely prevent autocatalysis. The Q7K mutation has been kept in both HIV-1 protease variants used in this study. Prior to the protease sequence, there is a 12 amino acid long gag-pol recognition sequence that starts with a methionine residue (start codon) since the protease lacks a start codon. As such, the protease autoprocesses out of the gag-pol recognition sequence after folding and dimerisation.

The lyophilised sample of E35D↑G↑S was insufficient; therefore, a gene construct was synthesised in order to continue with the proposed aims and objectives. The E35D↑G↑S variant gene sequence was designed using the C-SA PR gene sequence as a template so that codon differences between the two variant protease gene constructs were minimised (shown is supplementary material S1). The pET-11b vector encoding E35D↑G↑S was synthesised by GenScript (Piscataway, USA). Site-directed mutagenesis was not a suitable technique for the production of the E35D↑G↑S variant due to numerous mutations and insertions in the variant which would be time consuming and costly.

A D25A active site mutation of C-SA PR and E35D↑G↑S were generated through site-directed mutagenesis. The active site mutants were generated as a result of the unsuccessful expression of the E35D↑G↑S variant. The D25A mutation inactivates the aspartic HIV-1 protease (Koh *et al*., 2011). As such, the active site mutant was expected to express better because the autocatalytic function of the protease is muted.

3.5 Primer design and site-directed mutagenesis used to create the D25A active site mutant

Site-directed mutagenesis is a technique involving the use of polymerase chain reaction (PCR). However, the primers used contain a mismatched base where a single mutation will be introduced in the amplified product by the polymerase. Site-directed mutagenesis was

performed using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies) according to the manufacturer's instructions. The initial denaturation step was carried out at 95 °C for 2 minutes. Thereafter, 18 cycles of denaturation (95 °C for 20 s), annealing (60 °C for 10 s) and elongation (68 °C for 3.5 minutes) were performed. A final elongation step was carried out a t 68 °C for 5 minutes.

The forward and reverse primers used to introduce the D25A mutation were designed manually. The primers were designed separately for C-SA PR and E35D↑G↑S based on the sequence identified by Inqaba Biotec™ (Pretoria, South Africa) for each variant. The two sets of primers were aligned with ClustalX 2.0 (Larkin, 2007) in order to verify if it was necessary to obtain two different sets of primers or if one set would be suitable for both variants. It was deduced that one set of primers would suit both variants since they were identical. In the forward primer, the adenine nucleotide was changed to a cytosine nucleotide. In the reverse primer, the thymine nucleotide was changed to a guanine nucleotide underlined in the sequence. The changed nucleotides allow for the aspartic acid residue at position 25 to be mutated to an alanine residue. The changed nucleotides are underlined in the sequences of the primers. The forward (5' CAAAGAAGCTCTGCTGGCCACTGGCGCTGACGAC 3') and reverse (5' GTCGTCAGCGCCAGTGGCCAGCAGAGCTTCTTTG 3') primers were checked with OligoCalc for high melting temperature (Tm). The Tm was found to be 80 ˚C (> 78 °C, as recommended by the manufacturer of the kit). Primers were synthesised by Inqaba Biotec™ (Pretoria, South Africa).

3.6 Transformation of *Escherichia coli* BL21(DE3)pLysS with pET-11b encoding protease variants

Transformation was done using the heat-shock method. Eighty microlitres of chemically competent BL21(DE3)pLysS cells (provided in the laboratory) were thawed on ice. Upon complete thawing, 2 µl of 189 ng/ µl DNA (5 µl of 1 ng/ µl DNA in the case of site-directed mutagenesis product) was added to the cells and mixed gently by tapping. The cells were then incubated on ice for 30 minutes. Thereafter, the cells were heat-shocked at 42 ˚C for 45 seconds followed by immediate incubation on ice for 2 minutes. Four hundred microlitres of SOC media (pre-warmed at 37 ˚C) was then added. SOC media contains 2% (w/ v) tryptone, 0.5% (w/ v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose. SOC media improves transformation efficacy (Hanahan, 1983). After the addition of SOC media, the cells were incubated at 37 ˚C with shaking at 230 rpm for 1 hour.

One hundred microlitres of the culture were plated on LB plates (1.5 % agar) supplemented with the appropriate antibiotics. The remaining volume of culture was pelleted for 2 minutes at 10 000 x g using a Spectrafuge 24D microcentrifuge (Labnet International Inc.). Two hundred microlitres of the supernatant were discarded and 100 µl of the re-suspended cells were plated. Single colonies were selected and inoculated overnight in LB-Lennox media which contains 1% (w/ v) tryptone, 0.5% (w/ v) yeast extract and 0.5% (w/ v) NaCl (Lennox, 1955). Glycerol stocks were made by an equal-volume dilution of the overnight culture with a sterile 80% glycerol solution. Glycerol stocks were stored at -80 ˚C. The plasmid was extracted from an overnight culture with the GeneJET Plasmid Miniprep Kit (Thermoscientific®) and sent for sequencing. All cells stocks used have been verified and confirmed to be transformed with the pET-11b plasmid containing the correct DNA sequence. All sequencing was done by Inqaba Biotec™ (Pretoria, South Africa). DNA sequences were translated using ExPASy – Translate (Gasteiger *et al*., 2003). Sequence analysis was performed using ClustalX 2.0 (Larkin, 2007). There were no amino acid mismatches between the expected sequence and the sequence contained within the cell stocks.

3.7 Expression trials of E35D↑G↑S and E35D↑G↑S D25A

Expression trials were performed in order to identify the appropriate conditions for expression of E35D↑G↑S. Briefly, 1.5 ml of bacterial cultures grown in LB-Lennox were pelleted using a Spectrafuge 24D microcentrifuge (Labnet International Inc.) at 10, 000 x g for 2 minutes. The cells were re-suspended in 175 µl of 10 mM Tris-HCl buffer, pH 8.00, containing 2 mM EDTA and 1 mM PMSF. Thereafter, the samples were frozen at -20 ˚C. The cells were thawed at room temperature to aid cell lysis. The thawed cells were incubated for ten minutes at room temperature with 1 µg/ ml DNAse and 10 mM $MgCl_2$. The cells were further lysed by sonication using a XL-2000 series sonicator manufactured by Misonix Ultrasonic Liquid Processors. The cells were sonicated once for ten seconds at 12 Watts. The lysed cells were centrifuged for 10 minutes at 12, 000 x g using a Spectrafuge 24D microcentrifuge (Labnet International Inc.). The supernatant was mixed with an equal volume of reducing SDS-PAGE sample buffer and electrophoresed as per the protocol described by Schägger (Schägger, 2006). The pellet was re-suspended in 165 µl of Tris-HCl buffer, pH 8.00, containing 2 mM EDTA and 1 mM PMSF, thereafter, mixed with an equal volume of reducing sample buffer. Ten microlitres of the expression trials samples containing reducing sample buffer were electrophoresed using 16% polyacrylamide gels.

In the expression trials of E35D↑G↑S, the effect of temperature (20 and 37 ˚C), IPTG concentration (0, 0.3, 0.6 and 1.0 mM) and induction periods (4 hours, 6 hours and, overnight) were tested in *E. coli* BL21(DE3)pLysS grown in LB-Lennox. Since overexpression was unsuccessful, a E35D↑G↑S D25A active site mutant was created. The inactive protein was subjected to expression trials using *E. coli* BL21(DE3)pLysS and *E. coli* Rosetta 2® cells. The Rosetta 2® cell line contains seven rare codons. In light of that, the Rosetta 2® cell line was tested to verify whether codon incompatibility prevents the expression of the E35D↑G↑S protease. The Rosetta 2® cell line was tested in LB-Lennox and LB-Miller media. LB-Miller media contains 1% (w/ v) tryptone, 0.5% (w/ v) yeast extract and 1% (w/ v) NaCl. The *E. coli* Rosetta 2® cell line was a kind gift from Dr. Sylvia Fanucchi (University of the Witwatersrand, South Africa).

3.8 Overexpression of C-SA PR

Glycerol stocks were inoculated and grown in LB-Lennox media overnight. Fresh media was inoculated with 1% (v/ v) of the overnight culture. At $OD_{600\ nm}$ = 0.6, overexpression was induced with 1 mM IPTG for 4 hours at 37 °C. Cells were pelleted at 4 ˚C via centrifugation at 5, 000 x g with Sorval Lynx 4000 Centrifuge (Thermoscientific®). The cells were re-suspended in 1% (v/ v) of the culture volume, in 10 mM Tris-HCl buffer, pH 8.00, containing 2 mM EDTA and 1 mM PMSF. The cells were frozen at -20 ˚C. The frozen cells were thawed at room temperature to assist with cell lysis. The cells were further lysed by sonication at 12 Watts using a XL-2000 series sonicator manufactured by Misonix Ultrasonic Liquid Processors. Ten sonication cycles were performed. The cycles were 1 minute long alternating with 1 minute rest periods. Prior to sonication, the samples were pre-chilled on ice for 30 minutes and kept on ice throughout the sonication procedure.

3.9 Pre-modified purification of C-SA PR

After lysing the cells containing C-SA PR, the insoluble fraction was pelleted for 30 minutes at 24, 000 x g at 4 °C using a Sorval Lynx 4000 Centrifuge (Thermoscientific®). The pellet, which contained C-SA PR inclusion bodies, was re-suspended (twice) in 10 mM Tris-HCl buffer, pH 8.00, containing 2 mM EDTA and 1 mM PMSF with added 1% (v/ v) Triton-X 100. The solution was then centrifuged for 30 minutes at 24, 000 x g at 4 °C. This process removes the remaining lipid bilayer and any other weakly soluble contaminants with hydrophobic properties. The supernatant was discarded. The pellet was then re-suspended in 40 ml of 10 mM Tris-HCl buffer, pH 8.00, containing 8 M urea and 2 mM DTT for complete

denaturation of the inclusion bodies containing C-SA PR. The denatured inclusion bodies were refolded via dialysis in 10 mM formic acid with 10% glycerol and 0.02% sodium azide. The refolded protease was then activated by dialysis into activity buffer (10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide). A major part of contaminant precipitated when dialysing in activity buffer. The precipitant was removed by centrifugation for 30 minutes at 24, 000 x g at 4 °C. The active protease was purified using a 5 ml CM-sepharose column (GE Healthcare) with the use of an AKTA Prime Plus® chromatographic system manufactured by GE Healthcare. The eluted protease was concentrated using a 50 ml amicon® 8050 pressure cell. A 3, 000 Da cut-off regenerated cellulose membrane (Millipore) was used for concentrating the protein sample. The protein was concentrated at a pressure of 300 kPa using nitrogen gas, with constant stirring at 200 rpm.

3.10 Modified purification of C-SA PR

Steps were followed as per the pre-modified purification of C-SA PR. However, the activity buffer used for dialysis was modified by adding 2 mM DTT and 2 mM NaCl. The added components presumably disrupt non-specific protein-protein interactions. In all other aspects, the protocol remained unchanged when compared to the pre-modified purification protocol (described in section 3.8).

In the course of optimising the purification of C-SA PR and decreasing autocatalytic cleavage various approaches were taken. The rate of re-folding was increased using 1% SDS as a denaturant of inclusion bodies (in the presence of 2 mM DTT) (Schlager *et al*., 2012). The denatured protein sample was refolded in dialysis buffer (10 mM formic acid with 10% glycerol and 0.02% sodium azide). The rapid refolding procedure produced soluble protease. However, increased amounts of contaminating proteins were salting in as well. In addition, various pHs were used in activation buffers (10 mM Tris-HCl, pH 7.00, pH 7.5 and pH 8.00 containing 0.02% sodium azide) in order to reduce autocatalytic activity of the protease. Varying pHs changes the protonation state of the catalytic aspartic acid, taking advantage of the catalytic mechanism of HIV-1 protease (Hyland *et al*., 1991). However, these conditions did not prove to be effective because the protease produced SDS-PAGE profiles typical of autocatalysis (data not shown).

3.11 Protein concentration determination

The concentration of protease was determined spectrophotometrically. The absorbance of protein samples was taken at 280 nm using a Jasco V-630 Spectrophotometer. The concentration was calculated using the Beer-Lambert law (equation 1).

$$A = \varepsilon c l \qquad \text{(equation 1)}$$

Where A is the absorbance, ε is the molar extinction coefficient, c is the molar concentration and l is the path length (1 centimetre).

The theoretical extinction coefficient of the dimeric C-SA PR (25, 230 $M^{-1}.cm^{-1}$) was determined using ProtParam (Gasteiger *et al.*, 2005). The absorbance at 280 nm of the dialysate was subtracted as a blank reading. In addition, light scattering due to aggregates was corrected for by subtracting the absorbance reading at 340 nm. The concentration of protein was determined for every experiment.

3.12 Size-exclusion ultra-fast liquid chromatography (UFLC)

The purity of C-SA PR was analysed using size-exclusion ultra-fast liquid chromatography using the ultra-fast liquid chromatograph manufactured by SHIMADZU®. The samples were chromatographed using a TSKGEL SUPER SW2000 (Sigma-Aldrich®). The column has a separation range from 5 kDa to 150 kDa. The column combines the properties of reverse-phase and normal phase resins. That is, the resin is silica-based with hydrophilic components. The hydrophilic phase prevents the formation of hydrophobic interactions between proteins and the silica gel, allowing proteins to migrate with the mobile phase.

Samples were chromatographed in 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide and either 150 mM or 300 mM NaCl. The concentration of protein used was 20 µM (matching the concentration used for active site titration experiments). The samples were eluted at a flow rate of 0.2 ml/ min. The molecular weight standard used was MWGF70 (Sigma-Aldrich®) containing aprotinin (6.5 kDa), cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa) and bovine serum albumin (66 kDa).

3.13 Isothermal titration calorimetry

Isothermal titration calorimetry is a highly sensitive calorimetric technique used for determining binding energetics of interacting macromolecules. The micro-calorimeter detects enthalpic changes (ΔH) during titrations. Changes in enthalpy due to chemical interactions in

the sample cell trigger the power feedback system to compensate for emerging heat changes. It is in this manner that the reference cell and sample cell are kept at isothermic conditions. The system comprising of the two cells is isolated by an adiabatic jacket (Holdgate & Ward, 2005; Núñez *et al*., 2012) (Figure 10). A titration entails a series of injections into the sample cell. The reference cell contains the solvent alone which is usually deionised water. That is, if the buffer system does not have a very large osmolarity or ionic strength (Duff *et al*., 2011). In which case, the buffer should be loaded into the reference cell instead of solvent alone.

The injection volumes are user-defined. The purity of the reactants must be of highest attainable quality. The concentrations of the reactants must be prepared and known with the highest possible accuracy. The data-points that are obtained are equivalent to the heat compensated by the instrument into the adiabatic system so that the temperature of the system is kept constant. The concentration of titrand is best measured spectrophotometrically using a sample extracted from the sample cell after loading it, in order to account for dilution by residual dialysis buffer used to wash the sample cell.

The injected solution (titrant) is usually a ligand and it is at a higher concentration than the reacting counterpart in the sample cell (titrand) which is usually a protein or enzyme. This is because the protein or enzyme must be saturated with ligand several injections prior to the end of the titration experiment. That is, in order to accurately determine the reaction equilibrium constant (K_a) and the heats of dilution of the ligand into buffer.

The heats of dilution of titrants are often measured in blank titrations of ligand into buffer. However, the heats of each injection are due to several factors: binding, ligand dilution and, macromolecule dilution and mixing of components due to stirring (Holdgate & Ward, 2005). It is for this reason that heats of dilution are more accurate at the end of the experimental titration as opposed to the blank titration.

Optimally, a thermogram is sigmoidal in shape. For that, experimental design must be planned to render a Wiseman C-value ($C = K_a$ x [titrand] x N) between 1 and 1000 for the data set to fit a sigmoidal equation (Wiseman *et al*., 1989; Turnbull *et al*., 2003) where K_a is the binding constant at equilibrium and N is the binding stoichiometry. However, a C-value between 10 and 500 is optimal. For strong binders ($K_a \geq 10^8\ M^{-1}$), the C-value is commonly between 100 and 500. This implies that the titrant concentration may not be much higher or lower than the concentration of titrand (Freyer and Lewis, 2008). Moreover, the concentration of protein for ITC experiments is set to be 10 to 50 times larger than the

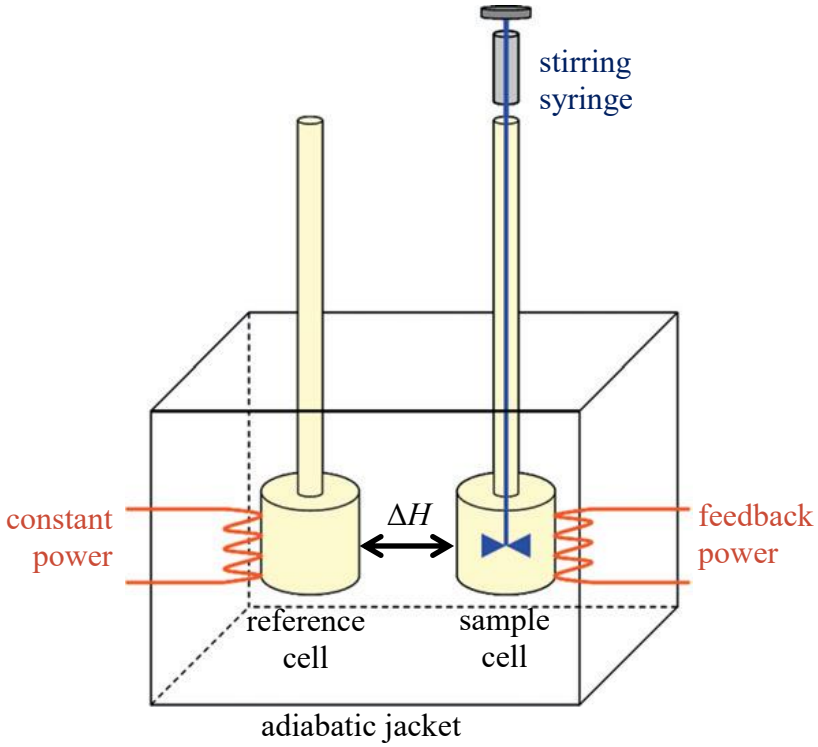


Figure 10: Schematic representation of a typical micro-calorimeter

The reference and sample cells are insulated by the adiabatic jacket. Changes in enthalpy (ΔH) trigger thermal compensation of power feedback to the sample cell. The two cells are brought back to an isothermal state in this manner. The syringe injects titrant into the sample cell containing the titrand. Heat is generated by binding, titrant dilution, titrand dilution and, mixing of titrant with titrand. Figure adapted from Núñez *et al.*, 2012.

dissociation constant of the ligand. Similarly, the concentration of ligand is set to be 10 to times larger than the concentration of protein. However, the concentration of ligand may often be set to be over 100 times larger than the concentration of protein for weak binders ($K_a \leq 10^4\ M^{-1}$) (Freyer and Lewis, 2008).

The pre- and post-transition regions of a titration curve directly determine the change in enthalpy (ΔH) of the reaction/ binding-event monitored. Knowing the molar enthalpy, the concentration of complexed molecules is determined. In light of that, by knowing the concentration of un-complexed molecules, the equilibrium constant (K_a) can be established (Figure 11) (Núñez *et al*., 2012). The K_a is established from the average of the K_a values at each point on the transition region of the titration curve. It is for this reason that the transition region, ideally, should be well-populated by data points. Ultimately, the change in entropy (ΔS) and change in Gibbs free-energy (ΔG) can be determined using equation 2 (Freyer and Lewis, 2008).

$$\Delta G = \Delta H - T\Delta S = -\,RT\ln K_a \qquad \text{(equation 2)}$$

Where ΔG is the Gibbs free-energy of the reaction, ΔH is the enthalpy of the reaction, T is the absolute temperature (K) of the reaction, ΔS is the entropy of the reaction, R is the gas constant (8.314 J mol^{-1} K^{-1}) and K_a is the equilibrium constant of the reaction. The inverse of K_a is K_d which is the dissociation constant of the enzyme-inhibitor complex.

The inflection point of the titration curve directly indicates the point in the titration where the concentration of ligand is equal to that of the macromolecule. With the aid of this, the binding stoichiometry (N) and percentage of active sites are directly established in active site titration experiments (Figure 11) (Freyer and Lewis, 2008, Núñez *et al*., 2012; Holdgate & Ward, 2005).

3.13.1 Active site titration calorimetry

In order to account for autocatalysis of the protease and misfolded specimens, active-site titration experiments were performed. The data was used to determine the concentration of active protein for subsequent analytical experiments. Active site titration experiments are performed with ligand-protein systems that have well-established binding energetics parameters. The binding energetics parameters confirm whether the titration experiment was performed appropriately; thus, indicating whether the determined binding stoichiometry (N)

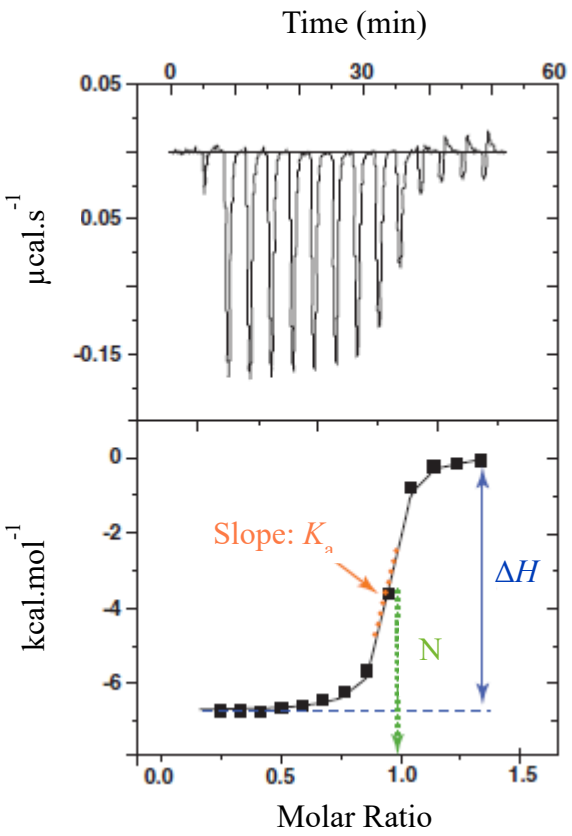


Figure 11: Parameters determined directly using isothermal titration calorimetry

The pre- and post-transition regions of the titration curve are used in determining the molar enthalpy (ΔH) of the reaction monitored. The molar enthalpy is used to determine the amount of protein-ligand complex formed at each injection. Thus, the transition region is used to determine the equilibrium constant (K_a) of the binding event. The inflection point of the sigmoidal curve occurs when the concentration of ligand is equal to that of the macromolecule. The inflection point is used to determine the stoichiometry (N) and percentage of active sites. Figure adapted from Núñez *et al*., 2012.

was accurate. More than that, the binding energetic parameters obtained in these typical titrations are used as fixed parameters when fitting displacement titration data sets.

Acetyl-pepstatin is a well-documented HIV-1 protease inhibitor (Seelmeier *et al*., 1988). Acetyl-pepstatin (8 mM) stocks were made by dissolving in 9 mM NaOH. The active-site titration experiment was conducted as follows: PR (20 µM) was titrated with 35 injections of 7 µl acetyl-pepstatin (200 µM). The dialysis buffer contained 10 mM sodium acetate, pH 5.00, containing 0.02% sodium azide. The DTT was dialysed out particularly for ITC because the oxidisation of DTT introduces noise into the baseline of ITC data. The reference cell contained deionised water. All solutions were degassed. The volume of the syringe used in all ITC experiments was 250 µl. All titrations were performed using a standard volume Nano ITC (TA Instruments). Data were fitted with NanoAnalyze (TA Instruments) using the independent model, also known as the single-site model or one-to-one model, as the fitting model.

3.13.2 Displacement titration calorimetry (DTC)

Protease inhibitors bind with high affinity to HIV-1 proteases, having dissociation constants ($1/ K_a = K_d$) in the nanomolar and picomolar range. As such, the determination of energetic parameters of inhibitors, especially K_a, is impaired in terms of accuracy when directly titrated because the protein saturates with proportionally very low amounts of inhibitor. Thus, the transition regions of the titration curves obtained are underpopulated by data points. In order to resolve this problem, it is necessary to titrate PIs into titrands composed of PR complexed with a weaker inhibitor like acetyl-pepstatin. The high affinity protease inhibitors displace acetyl-pepstatin. These experiments are known as displacement titration experiments (Velazquez-Campoy & Freire, 2006)

Stocks of atazanavir, ritonavir and darunavir had a 10 mM concentration dissolved in 100% DMSO. Atazanavir, ritonavir and darunavir were titrated against C-SA PR complexed with acetyl-pepstatin in displacement titration calorimetry experiments. The protease was initially incubated with 400 µM acetyl-pepstatin in a 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide and 2% dimethyl sulfoxide (DMSO). The DMSO was used to solubilise the hydrophobic inhibitors. Buffers were matched for all reactants. The concentration of protease and inhibitor used was altered for individual experiments in order to obtain well-defined, data-populated regions of titration curves. In general, the stronger the

affinity of a ligand for a macromolecule, the lower the concentration and/ or injection volume of ligand was set to be. This prevents premature saturation of the protease.

The injection volumes and concentrations of ligand and protein for each experiment were as follows: seven microlitre injections of 200 µM atazanavir were titrated into 18 µM C-SA PR. Six microliter injections of 110 µM ritonavir were titrated into 18 µM C-SA PR. Five microlitre injections of 110 µM darunavir were titrated into 15 µM C-SA PR. The number of injections was maximised for a 250 µl syringe. Data were fitted with NanoAnalyze (TA Instruments) using the competitive replacement fitting model.

3.14 Homology modelling and secondary structure prediction

Homology modelling is a technique used to generate 3D structures from primary-structure of proteins using experimentally determined structures as templates. The generated structures compensate for the absence of experimentally determined structures (Biasini *et al*., 2014).

The 3D structure of E35D↑G↑S was generated using SWISS-MODEL and the amino acid sequence of E35D↑G↑S. The model was generated using C-SA PR (PDB 3U71) as a template based on the highest sequence identity (96%). The Global Model Quality Estimation (ranging from 0 to 1) was 0.98, indicating a fairly reliable prediction of the structure based on the template used (Biasini *et al*., 2014; Arnold *et al*., 2006; Kiefer *et al*., 2009; Guex *et al*., 2009; Schwede *et al*., 2003). The structure was used to depict the location of the insertions in Figure 5 A.

The secondary structure prediction of the E35D↑G↑S model was made using PROCHECK (Laskowski *et al*., 1993, Laskowski *et al*., 2001). The predicted secondary structure content was used in the introduction section to elaborate the predicted aspects of E35D↑G↑S as well as to compare with the secondary structure of C-SA PR, thus, allowing expectations of circular dichroism experiments.

3.15 Secondary structure characterisation using far-UV circular dichroism

Circular dichroism (CD) spectroscopy is used to quantify the absorbance of left and right circularly polarised light. The differential absorbance of left and right circularly polarised light is known as ellipticity. Chiral molecules, such as those forming peptide bonds, absorb differentially polarised light in the far-UV region (240 - 180 nm). This is characteristic of how amino acids fold to form particular secondary structure elements (Woody, 1995; Kelly & Price, 2000). Typical ellipticity profiles have been documented. Anti-parallel β-sheets

produce a peak at 190 nm and a trough at ~216 nm. Alpha helices produce a relatively higher peak at ~190 nm, a trough at 222 nm and a trough at ~209 nm. Random coils produce a trough at ~195 nm (Greenfield, 1996, 2006; Kelly & Price, 2000).

In the event that a protein consists mostly of beta sheets and random coils, it is likely that destructive interference would occur in the region between 190 - 195 nm. HIV-1 protease is such a protein (Figure 12). Therefore, the ellipticity profile of the proteases being studied should characteristically trough at ~216 nm. It is also expected that the ellipticity profile of E35D↑G↑S would reveal more random coil content than C-SA PR. This expectation is based on a homology model (Figure 5 A) built with SWISS-MODEL (Biasini *et al*., 2014; Arnold *et al*., 2006; Kiefer *et al*., 2009; Guex *et al*., 2009; Schwede *et al*., 2003) which depicts the insertions as a random coil extension of the flap region of the protease.

The far-UV CD spectrum of C-SA PR was obtained using a Jasco J-1500 CD Spectrophotometer. The experimental parameters included a 0.5 nm band width and a data pitch of 0.2 nm. The path length of the cuvette was 0.2 mm. The ellipticity spectrum was taken from 180 to 250 nm (average of five accumulations). The mean residue ellipticity (MRE, deg.cm^2.dmol^{-1}) was calculated using equation 3.

$$\text{MRE} = \frac{100 * \theta}{Cnl} \qquad \text{(equation 3)}$$

Where θ is the ellipticity in millidegrees (mdeg), C is the mM concentration of protein (0.015 mM for C-SA PR), n is the number of amino acids in the protein (198 per dimeric C-SA PR) and l is the path length in cm (0.2 cm). The CD absorbance was multiplied by 32.98 (correlation coefficient) in order to be converted to θ (Woody, 1996; Kelly & Price, 2000). The buffer (10 mM sodium acetate, pH 5.0,) blank was subtracted from the C-SA PR data. The data were smoothed using SigmaPlot® v 11.0. The smoothing of the data was done using SigmaPlot® v 11.0 with a negative exponential smoothing technique. Accordingly, smoothing the data with a polynomial regression and weights computed from the Gaussian density function with a sampling proportion of 0.1.

The change in structure could also be used to determine structural modifications upon binding of ligands to the proteases. However, signal-to-noise ratio is very low due to the presence of DMSO used to solubilise the highly insoluble protease inhibitors. Samples containing DMSO produced excessive noise (data not shown).

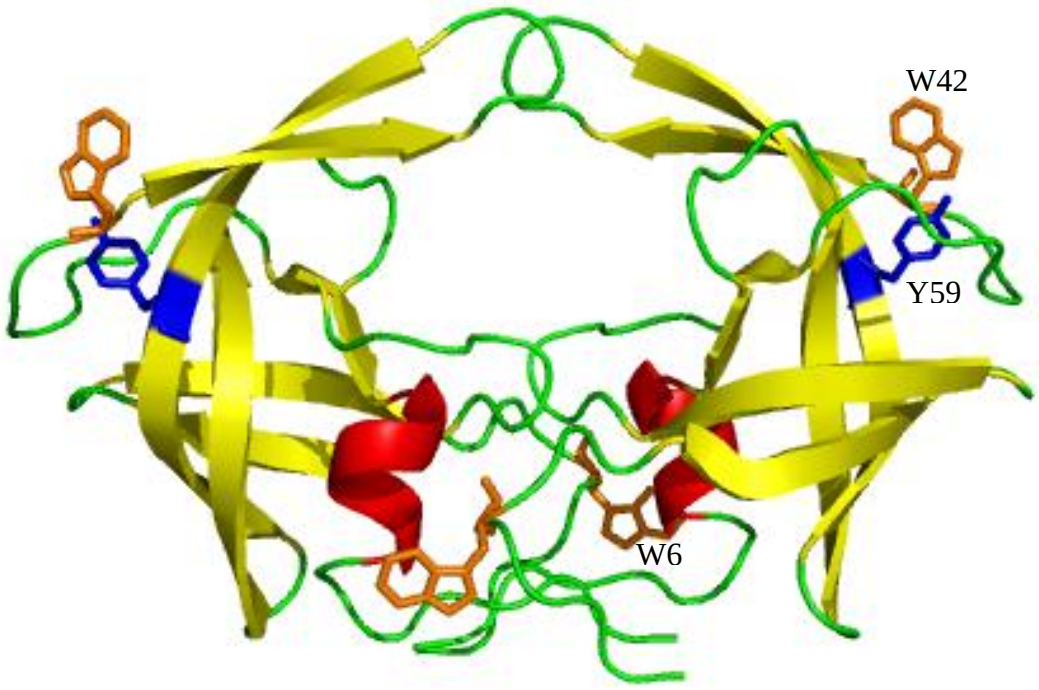


Figure 12: Crystal structure of C-SA PR (PDB I.D. 3U71) displaying secondary structural elements and fluorophores.

Secondary structural elements: B-sheets (yellow), α-helices (red), random coils (green). Fluorophores: W6 and W42 (orange) are solvent exposed. Y59 protrudes through the flap of the protease. Figure produced using PyMOL v0.99 (DeLano, 2002).

3.16 Tertiary structure characterisation – intrinsic fluorescence

Aromatic compounds are optically active. They absorb light in the UV region at a wavelength characteristic of the compound. The absorbance process mitigates the transition of electrons into excited states. Fluorescence emission is the phenomenon of photon release upon the return of excited electrons to the ground state. The emission wavelength is longer than the excitation/ absorption wavelength. That is because the energy is lost to the solvent uponemission, producing a red shift (Lakowicz, 2006).

Proteins contain several intrinsic fluorophores. The fluorophores most commonly excited are tyrosine and tryptophan at 280 nm and 295 nm, respectively. Fluorescence spectra provide information regarding the location of fluorophores on proteins. That is, solvent-exposed fluorophores lose energy to the solvent. As a result, the emission wavelength undergoes a red shift. Moreover, the presence of quenchers can be confirmed by quantifying the change in emission intensity after the addition of a quencher. A quencher can be a mutated amino acid as well (Royer, 1995; Lakowicz, 2006).

There are two tryptophan residues and one tyrosine residue conserved in the C-SA PR and E35D↑G↑S. Namely, W6, W42 and Y59. The two tryptophan residues are completely exposed to the solvent while the tyrosine residue protrudes through the flap of the protease (Figure 12). The double insertion found in E35D↑G↑S may further expose Y59 to solvent and may cause a red shift in the emission spectra. However, the shift may not be prominent; thus, the fluorescence spectra of the two protease variants may not be different. Fluorescence spectroscopy can be used to determine the relative structural differences between C-SA PR and E35D↑G↑S.

Fluorescence spectra were generated at 280 nm and 295 nm using the Jasco FP-6300 Spectrofluorometer. The experimental parameters were set as follows: the excitation and emission bandwidths were 2.5 nm and 5 nm, respectively. The emission spectrum was measured from five accumulations using a data pitch of 0.5 nm and a scanning speed of 200 nm/ min. The concentration of C-SA PR used was 5 µM. The data were plotted using Microsoft® Office Excel 2010.

3.17 Fluorometrically determined enzyme kinetics of C-SA PR

The kinetic profile of C-SA PR was determined fluorometrically using a Jasco FP-6300 Spectrofluorometer. The substrate, Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2, which

harbours a sequence preferred by HIV-1 proteases (Polgar *et al*., 1994), was used to determine the enzyme kinetics of C-SA PR. The substrate mimics the capsid/ p2 cleavage site found within the HIV-1 polyprotein chain. The sequence is recognised and hydrolysed between Nle (norleucine) and Phe(NO2) (nitro-phenylalanine) by the aspartic protease (Carmel & Yaron, 1978). The aminobenzoic acid (Abz) is excited at 337 nm and emits at 425 nm; however, it is quenched by the proximal nitro-phenylalanine. Thus, time-dependent cleavage of the scissile peptide bond (product formation) can be monitored fluorometrically and used to determine the catalytic rates of HIV-1 protease variants. This method can also serve as a basic comparative study involving the activity and affinity of different HIV-1 protease variants for this particular substrate. Of course, the kinetic parameters obtained are specific to the fluorogenic substrate used and to the sequence it mimics. To establish a dependence of substrate-sequence effects on HIV-1 protease affinity and activity, a broader study was performed involving various substrates (Polgar *et al*., 1994).

Fluorescence intensity standards were done in triplicate using 1.25 nanomoles of completely cleaved substrate, allowing for the conversion of intensity units to amounts of fluorogenic substrate cleaved over time. The excitation and emission bandwidth for all kinetic experiments were 2.5 nm and 5 nm, respectively. The experiments were monitored using a medium sensitivity with a 0.5 s response and a 1 s data pitch. The activity buffer contained 1 M NaCl and 100 mM sodium acetate, pH 5.00. The reaction volume was always kept constant at 400 µl. Cleavage rates were converted to appropriate units for each experiment. All experiments were performed in triplicate.

3.17.1 Michaelis-Menten determination of K_M for C-SA PR

For a Michaelis-Menten plot to be produced, the concentration of substrate must span a broad range. The substrate concentrations used were 0, 5, 10, 25, 50, 100, 125, 150 and 200 µM. The enzyme was added last to a final concentration of 10 nM. Thereafter, the reaction was monitored at 337 nm for 30 s. The enzymatic activity (µmol.min^{-1}) was plotted versus the concentration of substrate used (µM). The hyperbolic curve was fitted with the Michaelis-Menten equation (Johnson & Goody, 2011; Briggs & Haldane, 1925) (equation 4) using SigmaPlot® v 11.0:

$$v_0 = \frac{V\max\,[S]}{K_M + [S]} \qquad \text{(equation 4)}$$

Where v_0 is the initial velocity (before any significant amount of product is formed), V_{max} is the maximum velocity of the enzyme, [S] is the concentration of substrate and, K_M is the Michaelis-Menten substrate dissociation constant. The K_M corresponds to the concentration of substrate that induces the enzyme to have half maximal velocity. K_M can be used to identify whether the mutations and insertions in E35D↑G↑S alter the affinity of the fluorogenic substrate (and capsid/ p2 cleavage site) for the enzyme. Importantly, the K_M value is needed for the design of specific activity/ catalytic turnover and, catalytic efficiency experiments.

3.17.2 Determination of specific activity and catalytic turnover (k_{cat}) of C-SA PR

Specific activity is the catalytic activity of an enzyme standardised per milligram of enzyme. Data for specific activity can be used to determine the catalytic turnover of an enzyme. The catalytic turnover of an enzyme is the number of completed catalytic cycles by an active site per second (s^{-1}). In order for specific activity and catalytic turnover to be determined, the enzyme must be saturated with substrate. Thus, allowing for the steady-state assumption to be made. The steady-state assumption dictates that all the enzymes in a solution are complexed with substrate at any given time. This assumption implies that the loss of enzyme-substrate complex is instantly countered by the formation of the complex. Moreover, the assumption that the formation of product is negligible is also made so that the factor of product inhibition can be eliminated (Segel, 1988). As a result, the initial velocity is equal to the maximum velocity which becomes a function of enzyme-substrate complex concentration and the turnover number. The catalytic turnover is calculated using equation 5.

$$v_0 = V_{max} = k_{cat}[ES] \qquad \text{(equation 5)}$$

Where v_0 is the initial velocity, V_{max} is the maximum velocity, k_{cat} is the catalytic turnover and, [ES] is the total concentration of enzyme (assumed to be complexed with substrate). Therefore, the turnover number can be determined as long as either the substrate depletion or product formation can be monitored. In this case the formation of product was monitored fluorometrically (described in section 3.15).

The substrate concentration was kept constant at 50 µM (steady-state). Enzyme was added last, in increasing increments of concentrations (10, 20, 30, 40 and 50 nM). The initial velocity (µmoles.min^{-1}) was plotted against the amount of C-SA PR in milligrams, producing a gradient equivalent to the specific activity of C-SA PR (µmoles.min^{-1}.mg^{-1}). In order to determine k_{cat}, the initial velocity (µmoles.s^{-1}) was plotted against the amount of C-SA PR in

μmoles. The data were fitted using a linear equation resulting in a gradient equal to the catalytic turnover of C-SA PR (s^{-1}). The data were fitted using SigmaPlot® v 11.0.

3.17.3 Determination of catalytic efficiency (k_{cat}/ K_M) of C-SA PR

The catalytic efficiency of enzyme isoforms for a particular substrate is equated as a ratio of the turnover number (k_{cat}) to the Michaelis-Menten substrate dissociation constant (K_M). This proportion incorporates the affinity of enzymes for substrate (K_M) as well as the rate of product released (k_{cat}). The catalytic efficiency is directly used to compare the effectiveness of enzymes to catalyse the same substrate (Eisenthal *et al*., 2007).

The catalytic efficiency or, specificity constant (k_{cat}/ K_M) cannot be calculated as a proportion of the independently determined k_{cat} and K_M values. That is because it is usually not the case for the catalysis of substrate to be limited only by the diffusion of the substrate. There are other factors such as conformational changes, solvent effects and free-energy barriers that contribute to the rate limiting step of substrate catalysis (Kuo-Chen, 1976; for review see Lin & Lapointe, 2013). As a result, the catalytic efficiency of enzymes is determined experimentally.

Experimental determination of k_{cat}/ K_M is performed with concentrations of substrate significantly lower than the K_M so that the rate of formation of the enzyme-substrate complex is accounted for. The concentration of enzyme was kept constant at 50 nM while the concentration of substrate was ranged from 2 μM to 8 μM at 1 μM intervals. Data were fitted as catalytic turnover (s^{-1}) versus the concentration of substrate with a linear equation using SigmaPlot® v 11.0. The gradient of the fit was equal to the enzyme-substrate specificity constant ($s^{-1}.\mu M^{-1}$).

3.17.4 IC_{50} determination and calculation of K_i

The concentration of protease inhibitor that effectively inhibits 50% of the enzyme (IC_{50}) was determined separately for atazanavir, ritonavir and darunavir in the presence of 50 nM active enzyme and 50 μM fluorogenic substrate (saturating conditions, steady-state assumption). The concentrations of atazanavir and ritonavir used were: 0, 20, 25, 40, 50, 75, 150 and 200 nM. The concentrations of darunavir used were optimised to accommodate for the stronger inhibitory ability of the inhibitor. The concentrations of darunavir used were: 0, 5, 10, 20, 25, 40, 50, 75, and 100 nM.

The enzyme was pre-incubated with inhibitor for 1 hour. The substrate was added last to initiate the reaction. The reaction was monitored over thirty seconds. The buffer used was 100 mM sodium acetate buffer, pH 5.00, containing 1 M NaCl and 2% DMSO. The IC_{50} experiments were done in triplicate. Each singlet data set was fitted independently to determine the IC_{50} value, thus, enabling the production of a standard deviation from the triplicates. The triplicate plot was used to determine IC_{50} values. Data were fitted with a two-parameter exponential decay model using SigmaPlot® v 11.0.

Obtained IC_{50} values and standard deviations were used to calculate K_i and the standard deviation of the determined K_i value. The K_i was calculated according to equation 6.

$$K_i = \left(IC_{50} - \frac{[E]}{2}\right) / \left(\frac{[S]}{K_M} + 1\right) \qquad \text{(equation 6)}$$

Where K_i is the dissociation constant of the enzyme-inhibitor complex, in the presence of substrate; [E] is the concentration of active enzyme (50 nM), [S] is the concentration of substrate (50 µM). K_M is the Michaelis-Menten constant determined experimentally (described in section 3.15.1) and IC_{50} is the concentration of inhibitor reducing the activity of the enzyme to 50%. IC_{50} was determined experimentally for atazanavir, ritonavir and darunavir. Equation 6 was developed for tight binding inhibitors (Copeland *et al*., 1995).

Irreversible binders may have IC_{50} values equal to half of the concentration of enzyme. The concentration of reversible inhibitors should exceed half the concentration of the enzyme by a factor relative to the K_i of the inhibitor. Indeed, the stronger the binder, the lower the K_i value; therefore, lower concentration of inhibitor is needed to achieve IC_{50}.

The calculation of K_i requires the accurate determination of K_M and IC_{50} values. The K_i value of inhibitors can also be determined accurately using the Dixon method (Dixon, 1953). However, this method requires the production of Michaelis-Menten plots using multiple concentrations of inhibitor. The plots would then be analysed using the Lineweaver-Burk method. However, this method is very expensive with respect to substrate. As a result, the Dixon method was not feasible considering the limited quantities of substrate available.

3.18 Crystallisation trials

Crystallisation trials were performed on both C-SA PR and E35D↑G↑S using the Hampton® screening kit. Crystal trials were set up using the hanging-drop vapour diffusion method. Crystallisation droplets were composed of equal volumes of protein solution and screening

solution. The final volume of the crystallisation droplets were 2 µl and 4 µl. Crystal trials were set with apo-enzymes and drug-complexed enzymes. The concentrations of enzyme used were 6 mg/ ml and 9 mg/ ml. The enzymes were prepared in a 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide, prior to mixing equal volumes with the Hampton® screening kit reagents. Crystal trials were set with C-SA PR with the aim to obtain a structure of a higher resolution than the 3U71 structure available. Lyophilised E35D↑G↑S was set up for crystallisation trials because a crystal structure contains highly valuable information. Although the lyophilised E35D↑G↑S sample contained contaminating peptides, the lack of any other sample has limited the options for crystal trials to this sample.

4 Results

4.1 Solubilisation of lyophilised E35D↑G↑S and confirmation of activity

Lyophilised E35D↑G↑S was refolded by removing denaturants (6 M guanidinum hydrochloride and 8 M urea in the presence of 2 mM DTT) using dialysis. The pre- and post-dialysis samples were electrophoresed on a 10% polyacrylamide gel. The lyophilised sample of protease contained an unknown protein/ peptide contaminant with a molecular weight lower than 10 kDa as seen in lanes 1, 2, 3 and, 4 (Figure 13). Following overnight dialysis, autocatalysis was apparent as the amount of E35D↑G↑S decreased significantly (lanes 2 and 4, Figure 13).

A degradation study was set up in order to identify an optimal time for dialysis-assisted refolding that would produce refolded, active E35D↑G↑S within a minimum time period to limit autocatalysis. The time allocated for dialysis-assisted refolding was monitored with the aid of SDS-PAGE at various times over a 24-hour period. Densitometry analyses were performed on the electrophoresed gels. Bands typical of autocatalytic degradation emerged after 45 minutes of dialysis (Figure 14 A and B). An optimal dialysis period of 1 hour and 15 minutes was chosen.

The denatured sample of E35D↑G↑S was dialysed out of 8 M urea and 2 mM DTT for 1 hour and 15 minutes. After dialysis, the protease was shown to hydrolyse the fluorogenic substrate, confirming that the protease was active for further experiments (data not shown). Furthermore, as indicated by a 16% polyacrylamide gel (Figure 15), the protease was refolded and active since prominent bands of lower molecular weight than the E35D↑G↑S

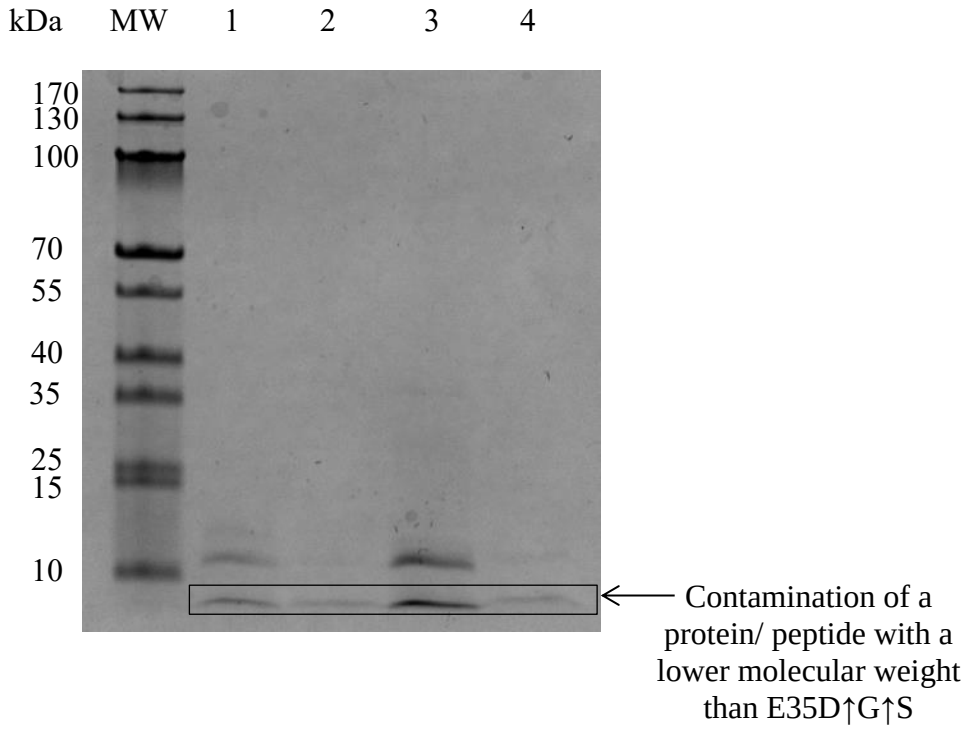
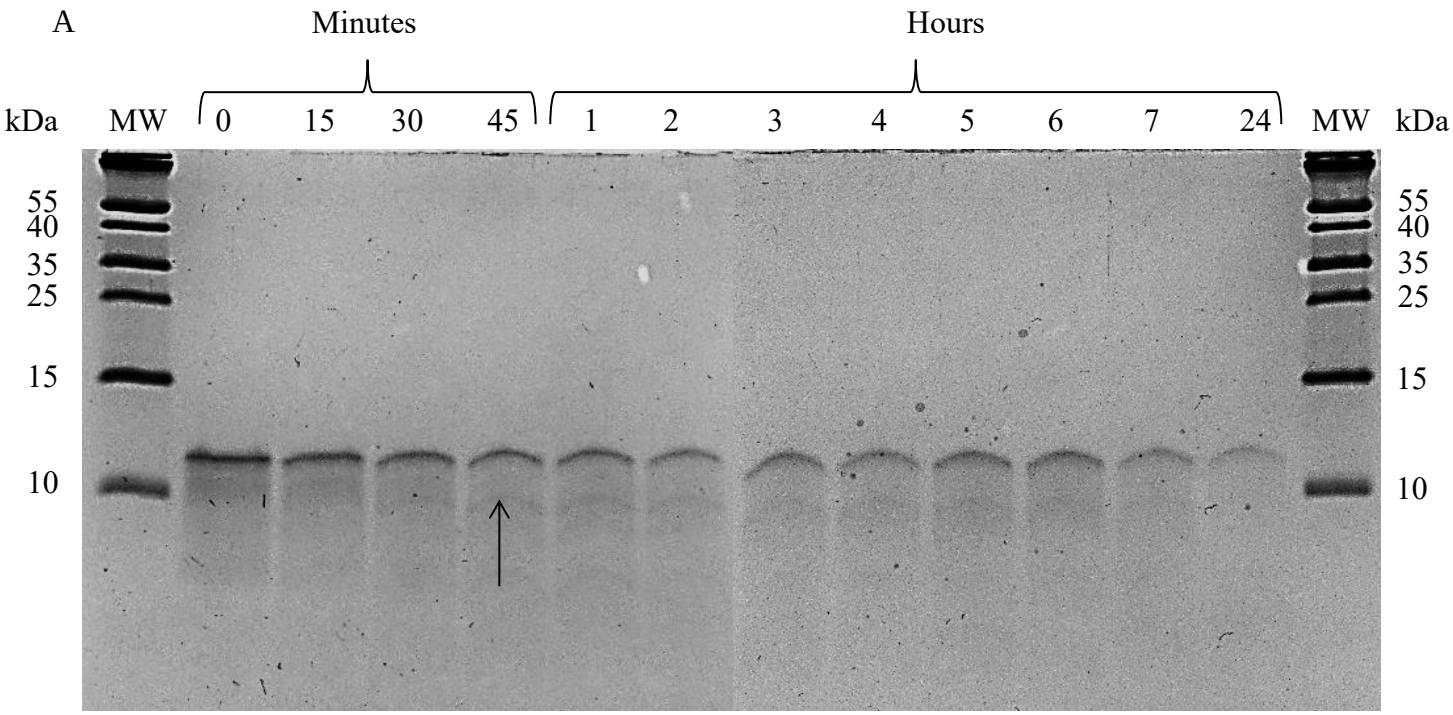


Figure 13: Contaminated lyophilised E35D↑G↑S and autocatalysis of the refolded protease.

E35D↑G↑S was dialysed overnight from different denaturants: Lane 1 shows E35D↑G↑S unfolded in 6 M guanidinium hydrochloride and 2 mM DTT; following dialysis-assisted refolding from 6 M guanidinium hydrochloride (lane 2). Lane 3 shows E35D↑G↑S unfolded in 8 M urea and 2 mM DTT; following dialysis-assisted refolding from 8 M urea (lane 4). The gel contained 10% polyacrylamide. Contaminating protein/ peptide present in the lyophilised sample (lanes 1, 2, 3 and, 4) indicated by the arrow.



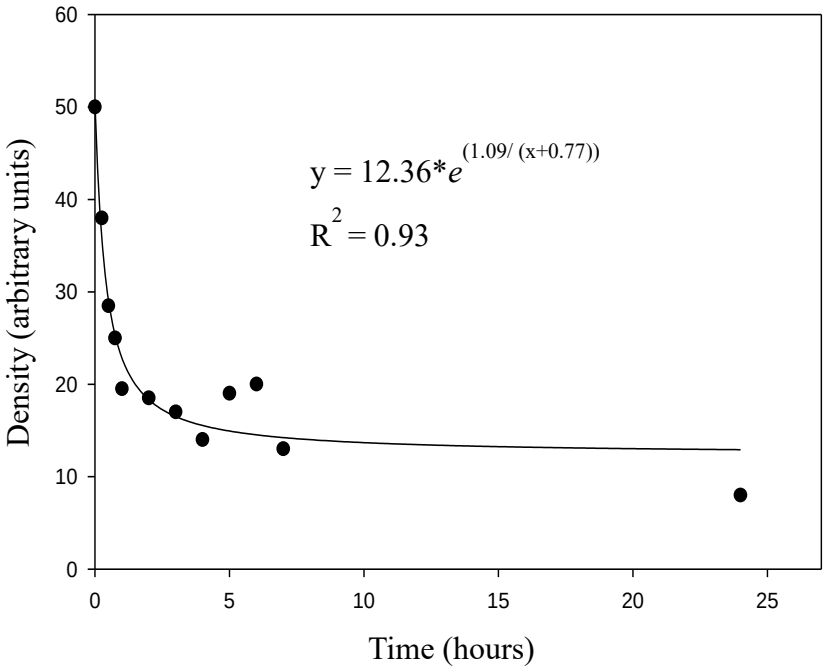


Figure 14: Degradation trial of lyophilised E35D↑G↑S

(A) The E35D↑G↑S protease, electrophoresed on a 16% polyacrylamide gel, autocatalyses rapidly. Bands of lower molecular weight than 10 kDa, characteristic of autocatalysis, emerged after 45 minutes of dialysis (indicated by a black arrow). (B) Densitometry analyses were performed on the 16% polyacrylamide gel using GelAnalyzer. The densitometry intensity points were fitted with Sigma Plot® 11.0.

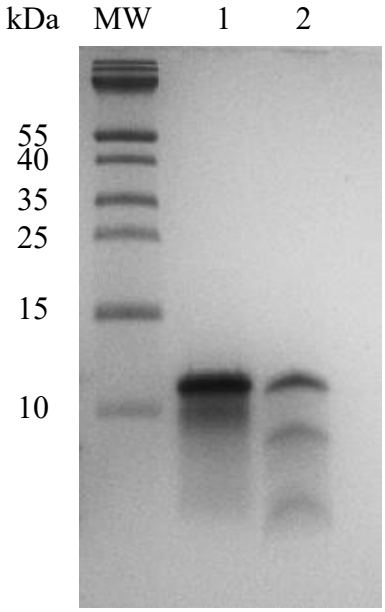


Figure 15: Refolding of lyophilised E35D↑G↑S

E35D↑G↑S was refolded via dialysis (for 1 hour and 15 minutes) from 8 M urea and 2 mM DTT (10 mM Tris-HCl buffer, pH 8.00) (lane 1) into 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide (lane 2). Prominent bands of lower molecular weight than 10 kDa emerged (characteristic of autocatalysis), indicating that the E35D↑G↑S protease was successfully refolded and activated.

protease emerged – a phenomenon typical of autocatalysis. The refolded E35D↑G↑S was titrated with acetyl-pepstatin in order to determine the percentage of active protease (Figure 16). The titration did not reach saturation as expected. The thermal signals for the injections were endothermic ($\Delta H > 0$).

4.2 Expression trials of E35D↑G↑S in *Escherichia coli* BL21(DE3)pLysS cells at 20 ˚C and 37 ˚C

Expression trials of E35D↑G↑S were set up using *E. coli* BL21(DE3)pLysS cells at 20 ˚C (Figure 17 A) and 37 ˚C (Figure 17 B). Varying concentrations of IPTG and induction periods were tested. Induction was done at $O.D._{600}$ = 0.6 (mid-log phase of growth curve). The media used was LB-Lennox. Overexpression was not observed in either the soluble fraction or the pelleted insoluble fraction.

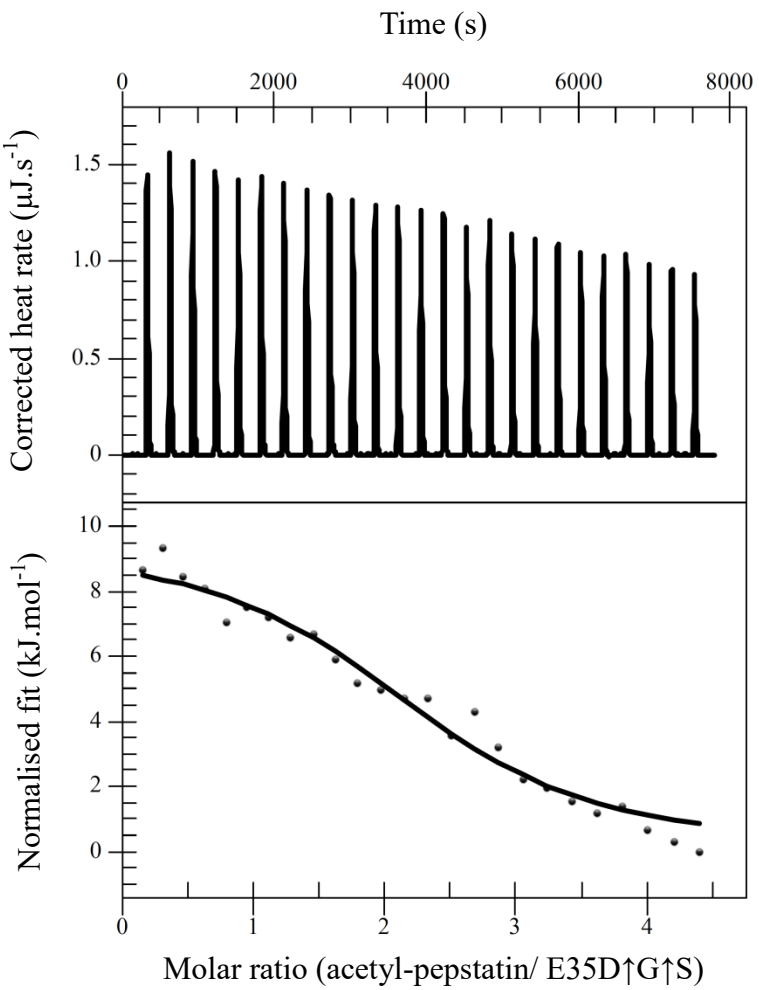
4.3 Expression trials of the inactive mutant E35D↑G↑S D25A in *E. coli* BL21(DE3)pLysS and *E. coli* Rosetta 2® cells

Expression trials of E35D↑G↑S D25A were performed using *E. coli* BL21(DE3)pLysS (Figure 18) and *E. coli* Rosetta 2® (Figure 19 A) cells at 37 ˚C in LB-Lennox. Additionally, expression trials using *E. coli* Rosetta 2® cells were performed using LB-Miller (Figure 19 B). Induction was done at an $O.D._{600}$ = 0.6 with various concentrations of IPTG. The induction period was also varied. Overexpression of the E35D↑G↑S D25A protein was not detectable in either the soluble or the pelleted insoluble fractions.

4.4 Pre-modified purification of C-SA PR

C-SA PR was expressed as inclusion bodies. The refolded protease was purified using CM-sepharose cation exchange chromatography (Figure 20 A). When the protease was activated via dialysis from refolding buffer (10 % glycerol) into 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide, without 2 mM NaCl and 2 mM DTT, the eluted protease sample was contaminated with a protein of ~15 kDa. Moreover, autocatalysis products below 10 kDa were also present in the sample (Figure 20 B). The purified C-SA PR sample was shown to contain contaminating proteins/ peptides using UFLC chromatography (Figure 21). Moreover, the C-SA PR peak contributes 56% to the sum total intensity of the identified peaks in the UFLC chromatogram. The protease sample was titrated with acetyl-

A



B

Thermodynamic parameters	Δ*G* (kJ/ mol)	Δ*H* (kJ/ mol)	-*T*Δ*S* (kJ/ mol)	N (stoichiometry)	K_d (µM)
Value	-30.79 ±1.97	9.44 ±1.29	-40.23 ±0.68	2.36 ±0.18	3.26 ±3.95

Figure 16: Active site titration of refolded E35D↑G↑S with acetyl-pepstatin

(A) E35D↑G↑S was titrated with acetyl-pepstatin. The titration did not reach saturation. Thermal injections were endothermic. The data were fitted with a one-to-one fitting model using NanoAnalyze (TA Instruments). (B) The thermodynamic parameters of the active site titration are outlined.

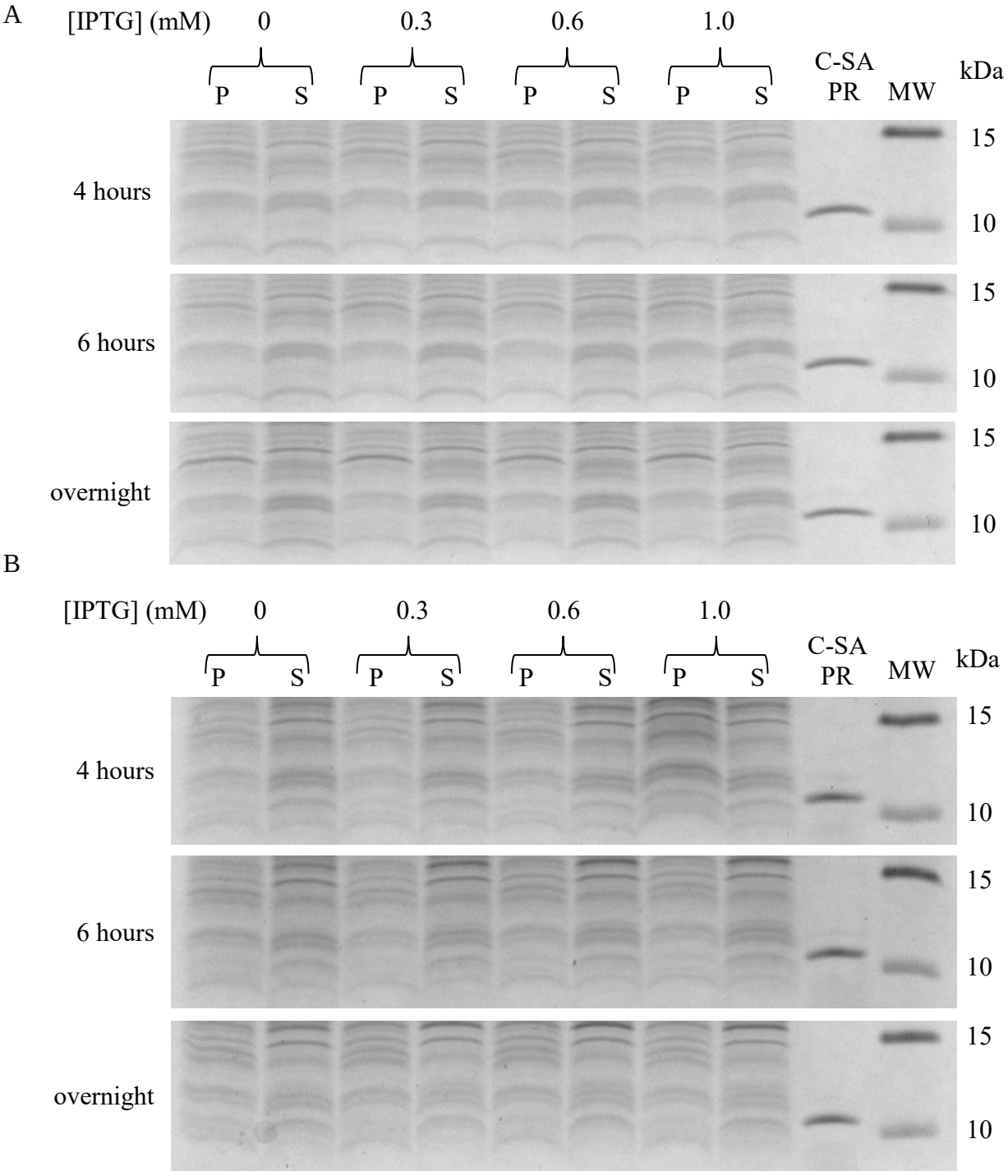


Figure 17: Expression trials of E35D↑G↑S in *E. coli* BL21(DE3)pLysS at (A) 20 ˚C and (B) 37 ˚C at varying concentrations of IPTG and induction periods in LB-Lennox

Expression of E35D↑G↑S was induced at $O.D._{600}$ = 0.6 with various concentrations of IPTG for 4 hours, 6 hours and overnight. C-SA PR was electrophoresed as a molecular weight marker indicating accurately where E35D↑G↑S should be expected. Overexpression was not observed in either the soluble supernatant (S) or in the insoluble pellet (P).

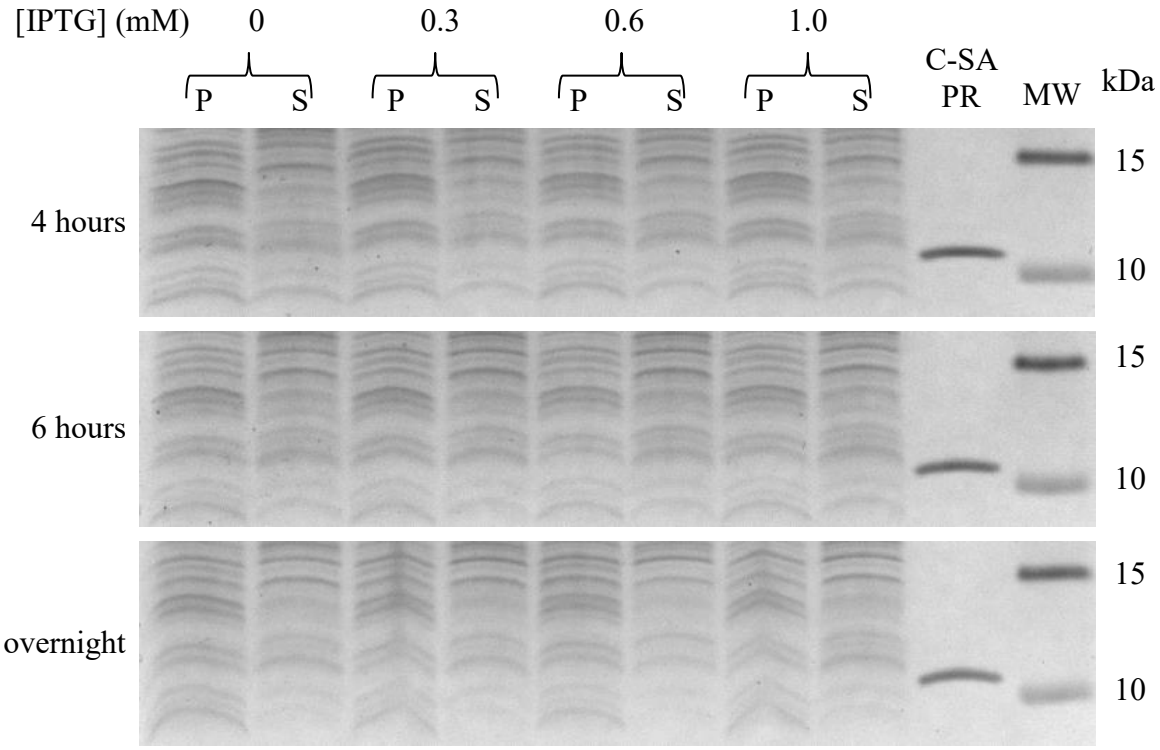


Figure 18: Expression trials of E35D↑G↑S D25A in *E. coli* BL21(DE3)pLysS at 37 ˚C in LB-Lennox using various concentrations of IPTG and induction periods

E35D↑G↑S D25A is an inactive variant of the E35D↑G↑S protease. Overexpression of E35D↑G↑S D25A was induced at $O.D._{600}$ = 0.6 with various concentrations of IPTG for 4 hours, 6 hours and overnight. C-SA PR was electrophoresed as a molecular weight marker indicating accurately where E35D↑G↑S should be expected. Overexpression was not observed in the soluble supernatant (S) or in the insoluble pellet (P).

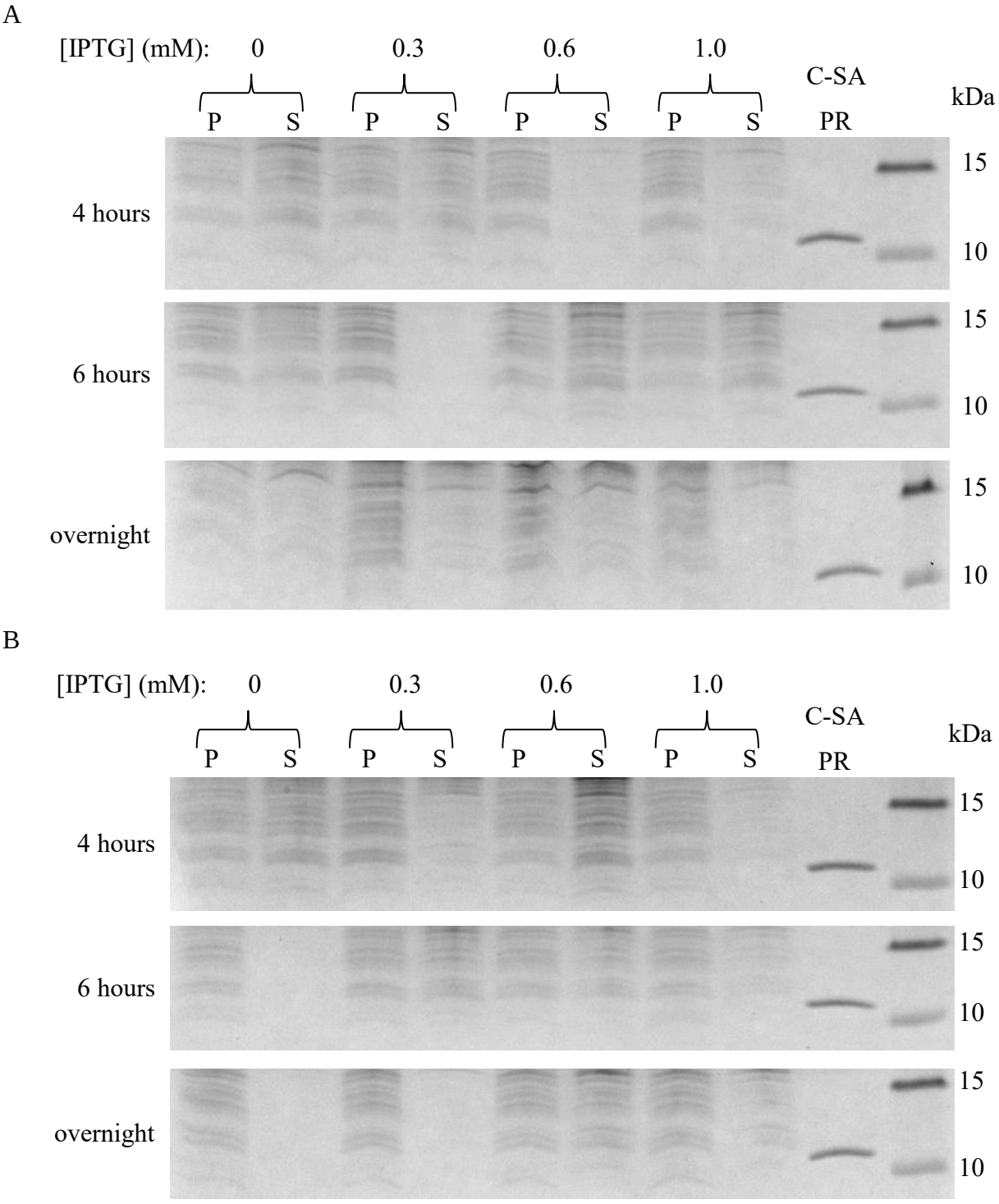


Figure 19: Expression trials of E35D↑G↑S D25A in *E. coli* Rosetta 2® cells in (A) LB-Lennox and (B) LB-Miller

Overexpression of the inactive protease variant E35D↑G↑S D25A was induced at $O.D._{600}$ = 0.6 with various concentrations of IPTG for 4 hours, 6 hours and overnight. C-SA PR was electrophoresed as a molecular weight marker indicating accurately where E35D↑G↑S should be expected. Overexpression was not observed in either the soluble supernatant (S) or in the insoluble pellet (P). Feint S lanes indicate insufficient proteins in the soluble fraction.

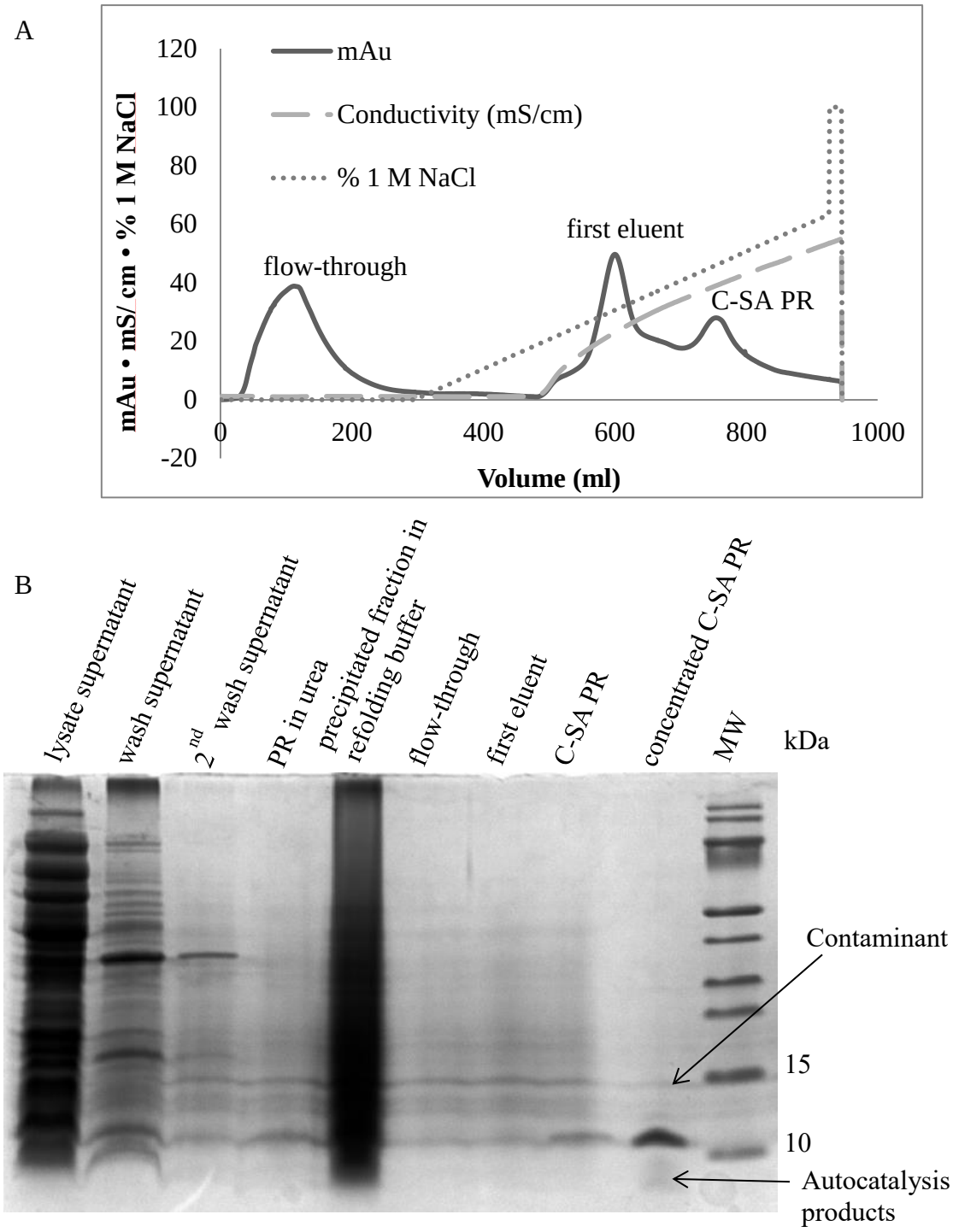


Figure 20: (A) Elution and (B) purification profile of C-SA PR using pre-modified protocol

C-SA PR was activated via dialysis into 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide, without 2 mM NaCl and 2 mM DTT. The eluted protease sample contains contaminating proteins and autocatalysis products (concentrated C-SA PR). (A) Fractions were eluted from a CM-sepharose column using a NaCl gradient. (B) Eluted fractions were electrophoresed on 16% polyacrylamide gel.

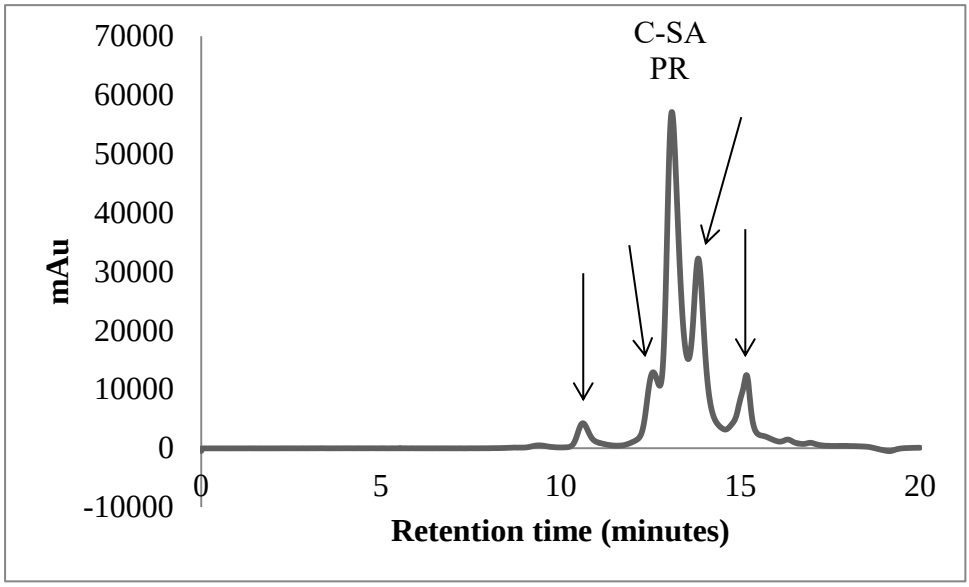


Figure 21: UFLC chromatogram of purified C-SA PR

The concentrated C-SA PR was analysed using UFLC. The protease was activated via dialysis into 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide, without 2 mM NaCl and 2 mM DTT. Contaminating protein peaks are indicated by arrows. The C-SA PR peak contributes 56% to the sum total intensity of the identified peaks.

pepstatin in an active site titration experiment and was determined that 55% of the C-SA PR protease sample was active (Figure 22 A and B).

4.5 Modified purification of C-SA PR

C-SA PR was eluted from a CM-sepharose column with an approximate 380 mM concentration of NaCl (Figure 23 A and B). The purification profile is similar to the one obtained for the pre-modified protocol. However, the protease sample obtained did not contain the contaminants that were present when using the pre-modified protocol as can be observed on SDS-PAGE and HPLC (Figure 24 A and B). However, there is a protein band immediately above the C-SA PR band that could be contaminant, or C-SA PR that did not autoprocess out of the gag-pol recognition sequence (Figure 24 A). The C-SA PR peak contributes 69% to the sum total intensity of the identified peaks in the UFLC chromatogram (Figure 24 B). In addition, the concentration of active C-SA PR was 74% when using the modified protocol, as determined with an active site titration of C-SA PR with acetyl-pepstatin (Figure 25). The activation of the protease in 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide, 2 mM DTT and 2 mM NaCl yields a higher percentage of active C-SA PR.

4.6 Structural characterisation of C-SA PR

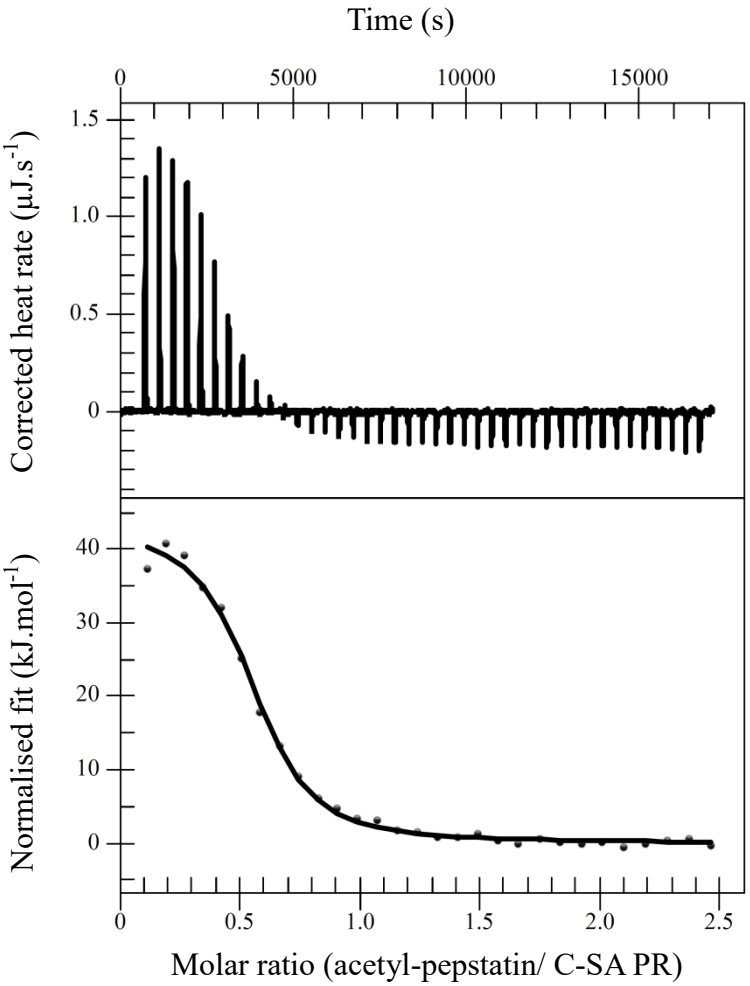
4.6.1 Far-UV circular dichroism – secondary structural characterisation

The far-UV CD mean residue ellipticity spectrum of C-SA PR (Figure 26) had a trough at 216 nm and a peak at 203.5 nm. The far-UV CD spectrum of C-SA PR indicates that the protein is predominantly β sheeted with a high proportion of random coils (Kelly & Price, 2000). The data in Figure 26 were smoothed using SigmaPlot® v 11.0. Data smoothing was done using the negative exponential smoothing technique which is a local smoothing technique using polynomial regression and weights computed from the Gaussian density function.

4.6.2 Intrinsic fluorescence – tertiary and quaternary structural characterisation

Excitation of C-SA PR at 280 nm and 295 nm wavelengths resulted in emission peaks at 346 nm in both instances (Figure 27). The fluorescence intensity is approximately 3 times higher when exciting at 280 nm when compared to exciting at 295 nm. The data were plotted using Microsoft® Excel 2010.

A



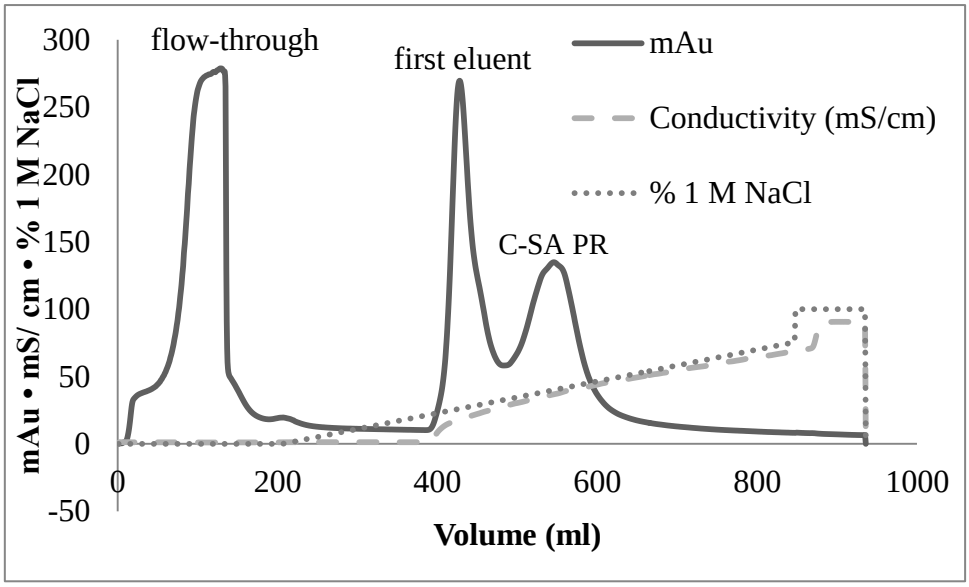
B

Thermodynamic parameters	Δ*G* (kJ/ mol)	Δ*H* (kJ/ mol)	-*T*Δ*S* (kJ/ mol)	N (stoichiometry)	K_d (µM)
Value	-35.60 ±0.54	43.04 ±1.62	-78.64 ±1.08	0.55 ±0.02	0.45 ±0.10

Figure 22: Active-site titration with acetyl-pepstatin against C-SA PR

(A) The C-SA PR purified using the pre-modified purification protocol was titrated with acetyl-pepstatin. The binding parameters of the titration are outlined in B. The parameters are as expected for the titration. The stoichiometry is 55%. Data were fitted with a one-to-one fitting model using NanoAnalyze (TA Instruments). The errors of the thermodynamic parameters are a direct resultant of the standard deviation of the fit (1.00).

A



B

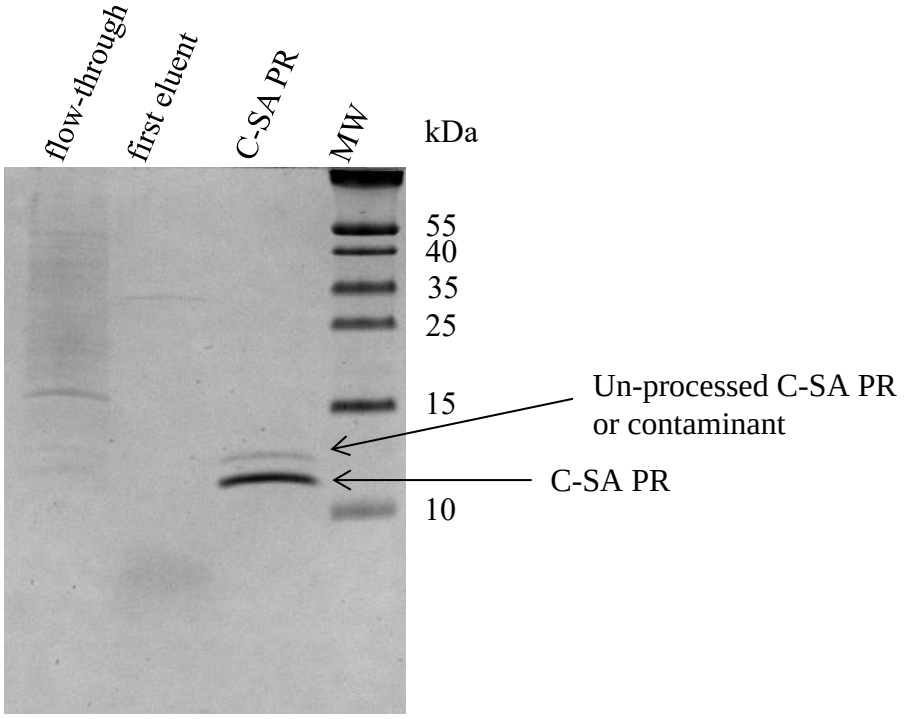


Figure 23: Elution profile of C-SA PR using modified purification protocol

(A) Elution chromatogram of C-SA PR and (B) 16% polyacrylamide gel of chromatographed components. C-SA PR was activated via dialysis into 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide, 2 mM NaCl and 2 mM DTT. The purification flow-through contained proteins co-precipitated with the protease when overexpressed. The first eluent contained autocatalysis products. The second eluent contained mostly C-SA PR. The band above C-SA PR could be either contaminating protein unable to be removed using CM-sepharose or C-SA PR with the gag-pol recognition sequence that may not have been cleaved off yet.

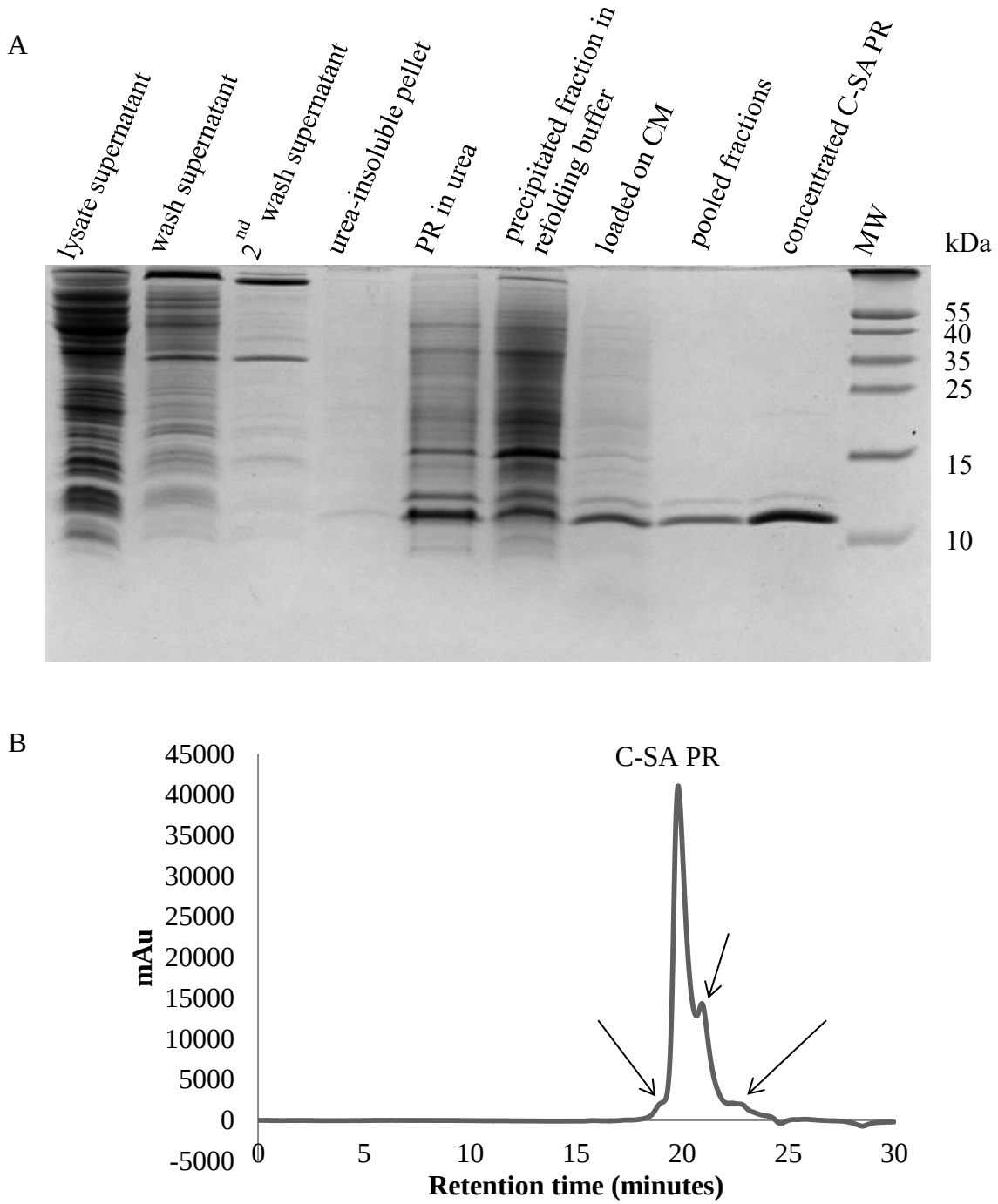
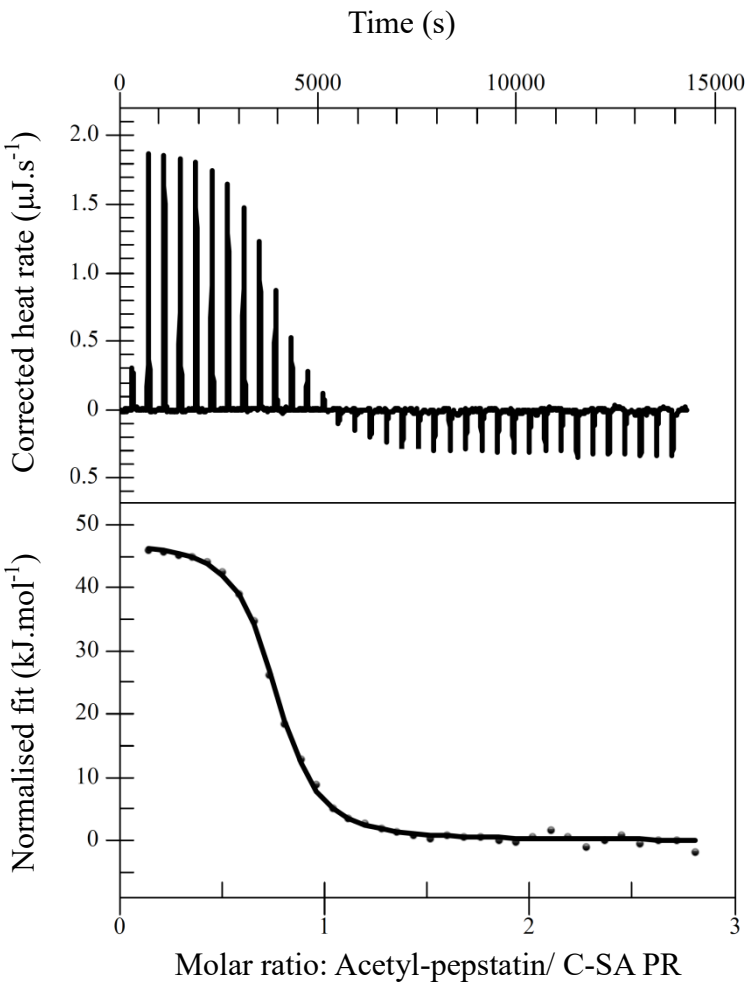


Figure 24: Purification profile of C-SA PR using modified purification protocol

(A) 16% polyacrylamide gel and (B) UFLC chromatogram of purified C-SA PR sample activated via dialysis into 10 mM sodium acetate buffer, pH 5.00, containing 0.02% sodium azide, 2 mM NaCl and 2 mM DTT. The electrophoresed band above C-SA PR could be either contaminating protein that were not removed during the purification or C-SA PR with the gag-pol recognition sequence that may not have been cleaved off yet. Contaminating protein peaks are indicated by arrows. C-SA PR contributes 69.0% to the total intensity of the UFLC chromatogram.

A



B

Thermodynamic parameters	Δ*G* (kJ/ mol)	Δ*H* (kJ/ mol)	-*T*Δ*S* (kJ/ mol)	N (stoichiometry)	K_d (µM)
Value	-36.58 ±0.31	47.62 ±0.70	-84.20 ±0.39	0.74 ±0.01	0.30 ±0.04

Figure 25: Active-site titration with acetyl-pepstatin against C-SA PR

The C-SA PR sample used in this titration (A) was purified using modified purification protocol. The binding parameters of the titration are outlined in B. The stoichiometry is 74%. Data was fitted with a one-to-one fitting model using NanoAnalyze (TA Instruments). The errors of the thermodynamic parameters are a direct resultant of the standard deviation of the fit (0.87).

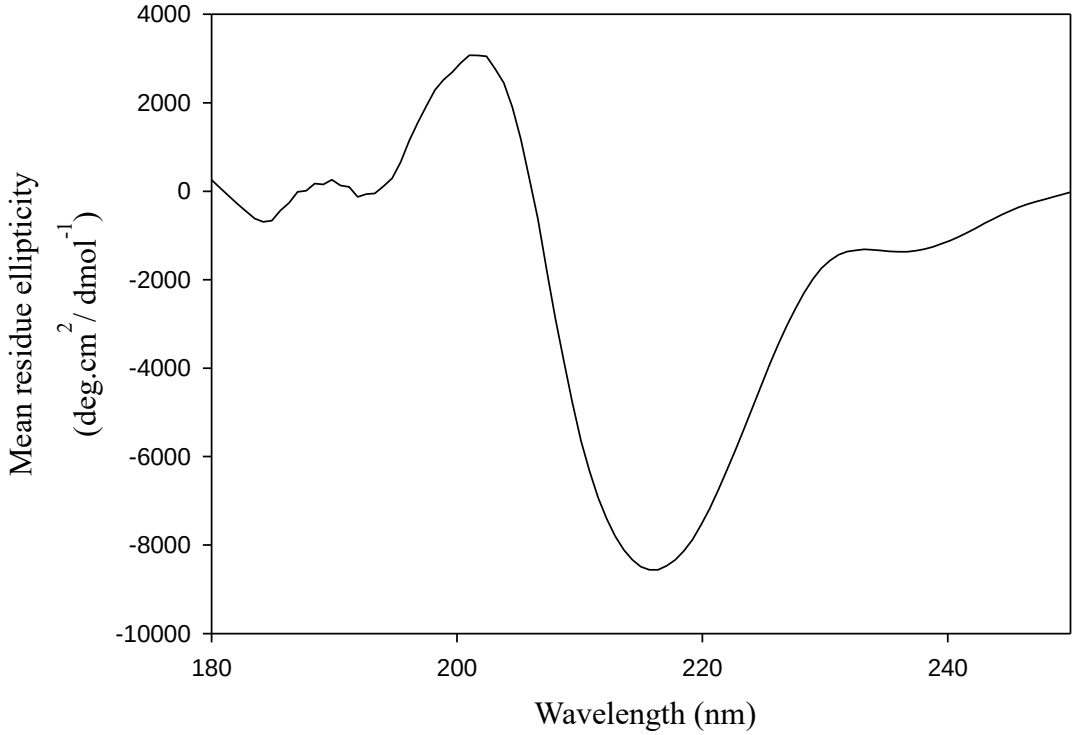


Figure 26: Far-UV circular-dichroism spectra of C-SA PR

The mean residue ellipticity of the far-UV CD spectrum of C-SA PR has a trough at 216 nm and a peak at 203.5 nm. The trough at 216 nm indicates that the protein contains a large proportion of β sheets. The peak at 203.5 further confirms that the protein is β-sheeted. However, the peak at 203.5 nm is typical of β-sheeted proteins with a high content of random coils. The data was fitted and smoothed with SigmaPlot® v 11.0. The data were smoothed using the negative exponential smoothing technique which is a local smoothing technique using polynomial regression and weights computed from the Gaussian density function.

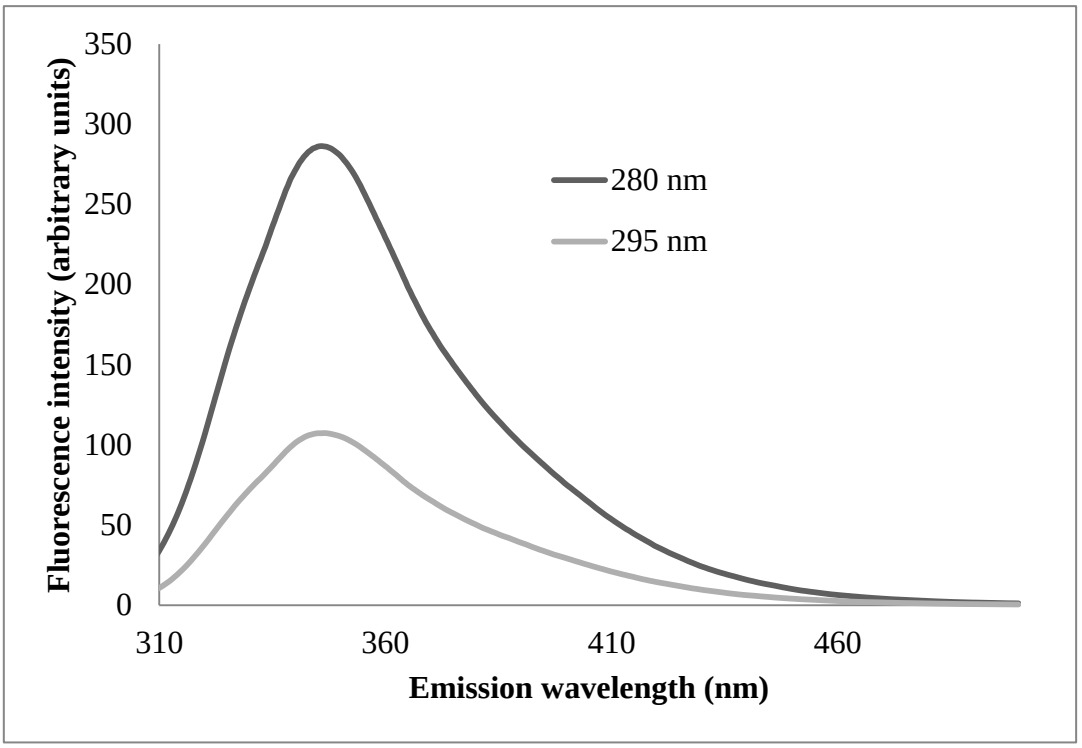


Figure 27: Intrinsic fluorescence of C-SA PR

C-SA PR was excited at 280 nm and 295 nm. The emission maximum is at 346 nm at both excitation wavelengths. Excitation at 280 nm produces an emission signal 3 times higher than excitation at 295 nm. Data were plotted using Microsoft® Excel 2010.

4.7 Enzyme kinetics of C-SA PR

4.7.1 Michaelis-Menten K_M determination

The dissociation constant (K_M) of the fluorogenic substrate, resembling the capsid/ p2 cleavage site, from C-SA PR was determined to be 20.22 ±4.09 µM. The data were fitted using SigmaPlot® v 11.0 with the Michaelis-Menten equation. The K_M was obtained from the equation of the plot (Figure 28). The correlation coefficient (R^2) of the fitted data fit was 0.97.

4.7.2 Specific activity and catalytic turnover

The specific activity and catalytic turnover of C-SA PR were determined from the same data set under steady-state conditions (enzyme saturated with substrate). However, the units were converted accordingly. The specific activity of C-SA PR was determined to be 35.68 ±1.06 µmoles.min^{-1}.mg^{-1} (Figure 29 A) and the catalytic turnover, k_{cat} = 12.79 ±0.38 s^{-1} (Figure 29 B). Data were fitted linearly using SigmaPlot® v 11.0. The correlation coefficient of the fitted data set was R^2 = 0.99.

4.7.3 Catalytic efficiency of C-SA PR

The catalytic efficiency/ specificity constant ($k_{cat}.K_M^{-1}$) of C-SA PR for the fluorogenic substrate and, therefore, for the capsid/ p2 cleavage site was determined under non-steady-state conditions ($S << K_M$). The data were plotted as catalytic turnover against substrate concentration and fitted linearly using SigmaPlot® v 11.0. The catalytic efficiency of the reaction was determined to be 1.17 ±0.06 s^{-1}.µM^{-1} (Figure 30). The correlation coefficient of the data fitting was R^2 = 0.98.

4.8 Inhibition kinetics of atazanavir, ritonavir and darunavir against C-SA PR

The inhibitor concentration that inhibits 50% of the enzyme (IC_{50}) is specific to the experimental conditions used, with particular regards to the enzyme concentration (50 nM). However, the concentration of substrate (50 µM) is equally important. The enzyme-substrate complex dissociation constant in the presence of substrate (K_i) was calculated from the IC_{50} values obtained for atazanavir, ritonavir and, darunavir using equation 6.

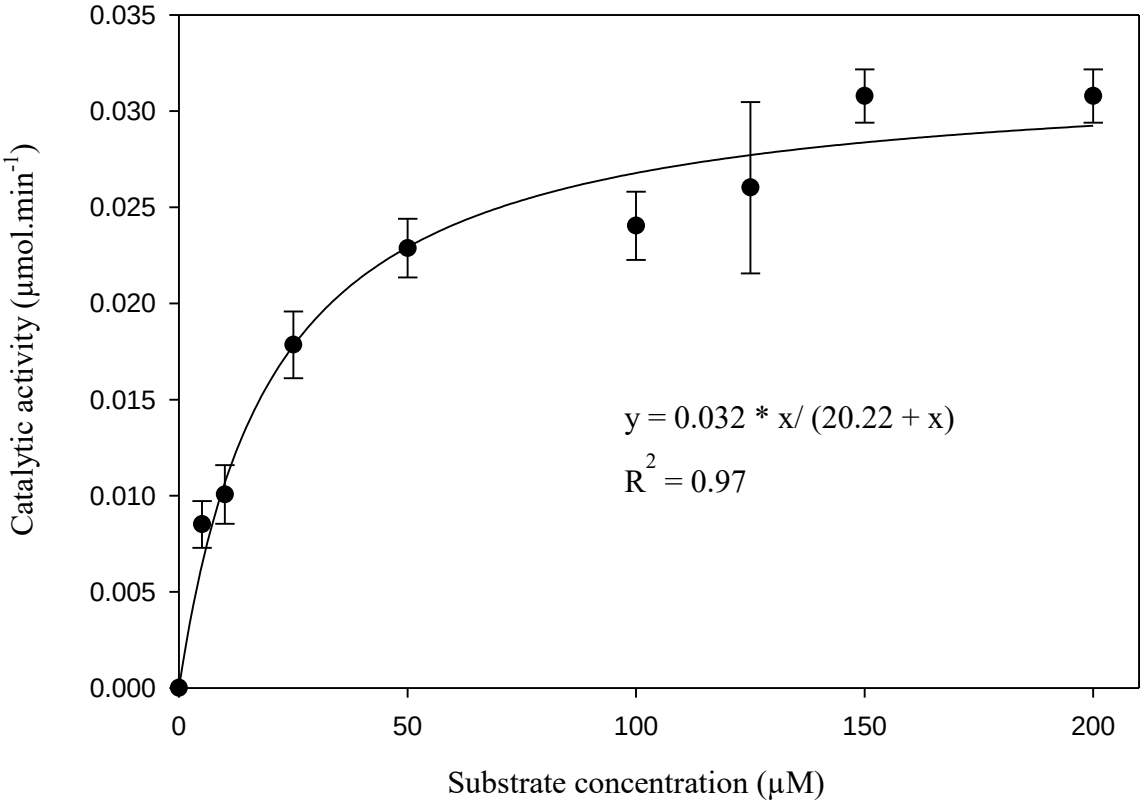
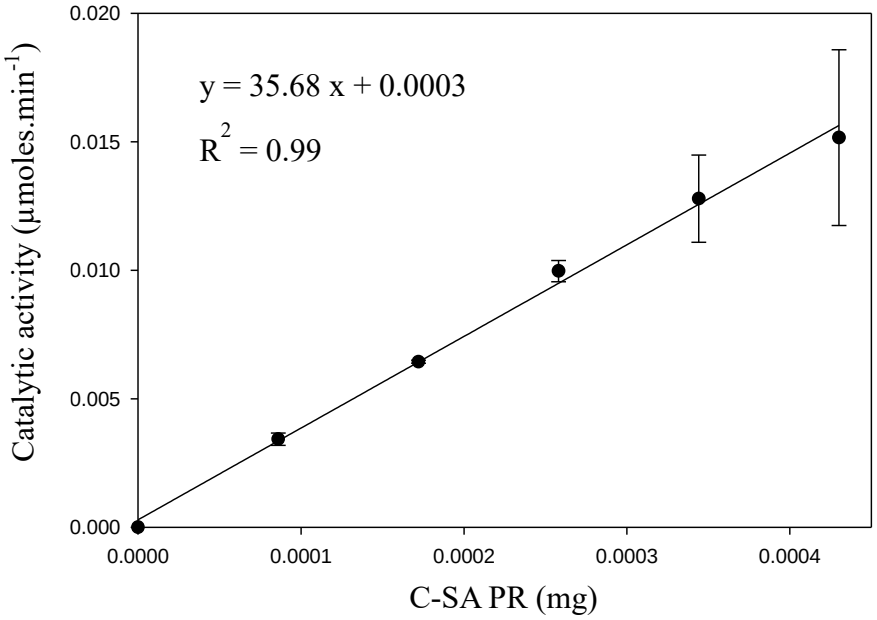


Figure 28: Michaelis-Menten plot of the hydrolysis of fluorogenic substrate catalysed by C-SA PR

The data was fitted with the Michaelis-Menten equation using SigmaPlot® v 11.0. The correlation coefficient of the fit was R^2 =0.97. The K_M was determined from the Michaelis-Menten equation. K_M = 20.22 ±4.09 µM. The error bars are standard deviation values obtained from the experiment performed in triplicate.

A



B

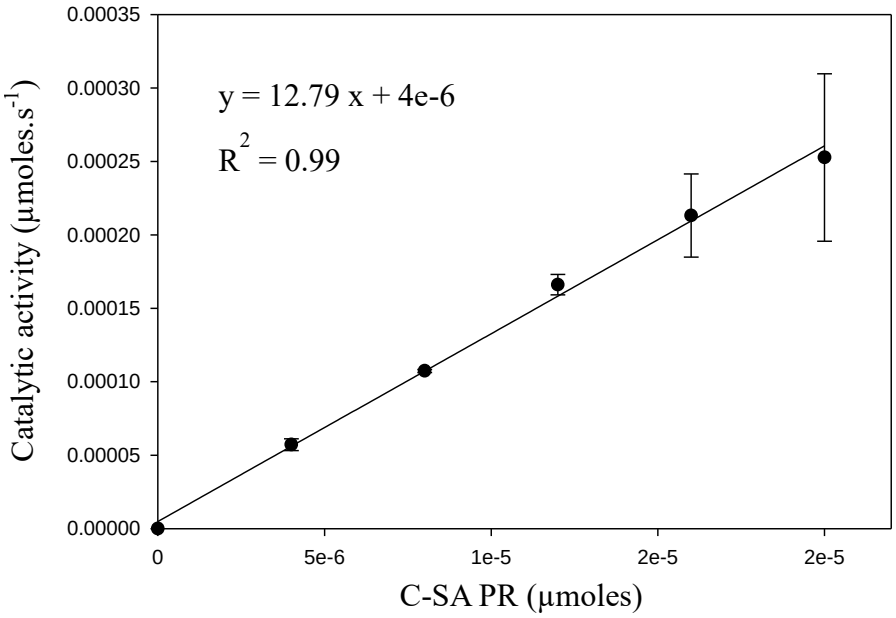


Figure 29: (A) specific activity and (B) catalytic turnover of C-SA PR

The plots were generated using the same data points. However, the units were converted so that the gradient of each plot gives specific activity and k_{cat}, respectively. The goodness of fit is $R^2 = 0.99$ for both plots. The specific activity of C-SA PR was determined experimentally to be 35.68 ±1.06 µmoles.min^{-1}.mg^{-1}. Likewise, the catalytic turnover of C-SA PR was determined to be k_{cat} = 12.79 ±0.38 s^{-1}. The error bars are standard deviation values obtained from the experiment performed in triplicate.

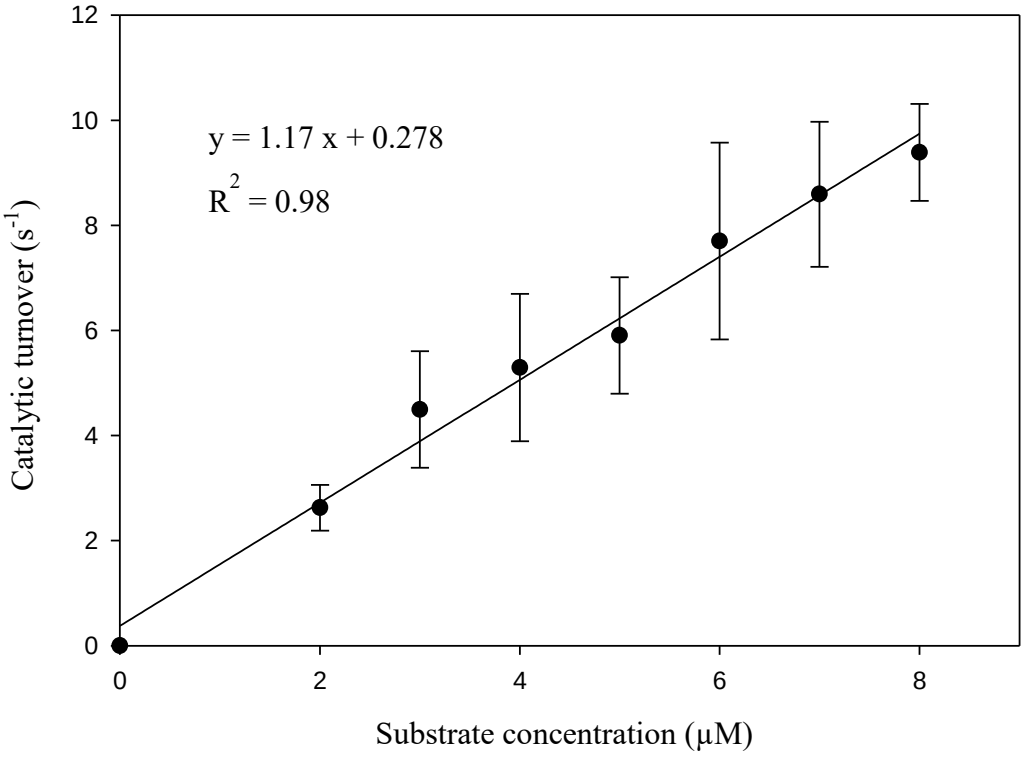


Figure 30: Catalytic efficiency of C-SA PR

The catalytic efficiency/ specificity constant of C-SA PR for the fluorogenic substrate which mimics the capsid/ p2 cleavage site was determined to be $k_{cat}.K_M^{-1}$ = 1.17 ±0.06 $s^{-1}.\mu M^{-1}$. The correlation coefficient of the fit was R^2 = 0.98. Data were fitted with a linear fit using SigmaPlot® v 11.0. The gradient of the fit is equal to the catalytic efficiency. The error bars are standard deviation values obtained from the experiment performed in triplicate.

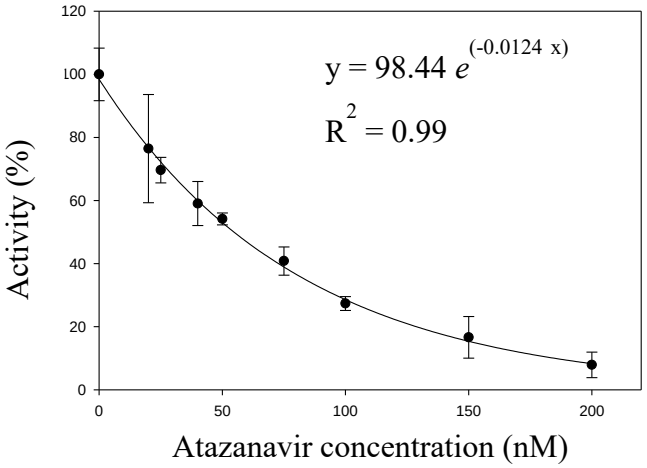
The fitted data sets for the IC_{50} determination of atazanavir, ritonavir and darunavir are displayed in Figure 31 A, B and C, respectively. The correlation coefficients of the data sets of atazanavir, ritonavir and darunavir were R^2 = 0.99, 0.98 and 0.98, respectively. The IC_{50} values were determined in triplicate. Data were fitted with a two-parameter exponential decay model using SigmaPlot® v 11.0. The error bars are standard deviation values obtained from the experiment performed in triplicate.

The IC_{50} values of atazanavir, ritonavir and darunavir were determined to be 54.63 ±12.60 nM, 49.63 ±9.85 nM and 25.99 ±0.13 nM, respectively. The calculated K_i values for atazanavir, ritonavir and darunavir were 8.53 ±2.06 nM, 7.09 ±2.84 nM and 285.07 ±36.00 pM, respectively. The IC_{50} and K_i values of the three inhibitors are graphically represented in Figure 32.

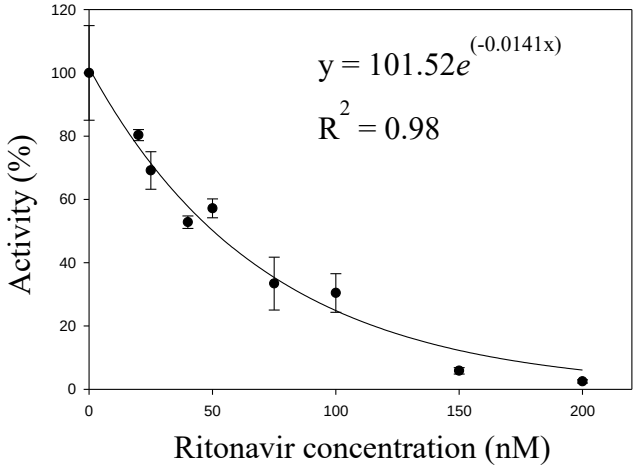
4.9 Thermodynamics of binding of atazanavir, ritonavir and darunavir to C-SA PR

The energetics of binding of atazanavir, ritonavir and darunavir were determined in displacement-titration calorimetry experiments. The three inhibitors were titrated against C-SA PR complexed with acetyl-pepstatin. The thermograms of the displacement titrations are displayed in Figure 33. The standard deviation values of the data-fits were 1.83, 1.24 and 1.50 for atazanavir, ritonavir and darunavir, respectively. The thermodynamics of binding of atazanavir, ritonavir and darunavir to C-SA PR are depicted and specified in Figure 34. Atazanavir and ritonavir have similar thermodynamic parameters as well as comparable dissociation constants in the pico-molar region. Atazanavir has a K_d of 160.56 ±54.59 pM and ritonavir has a K_d of 113.34 ±46.47 pM. The binding of the two inhibitors to C-SA PR is enthalpically unfavourable. Darunavir binds tighter than atazanavir and ritonavir, having a dissociation constant, K_d = 10.24 ±6.02 pM. The binding of darunavir is also enthalpically favourable. Darunavir has the lowest dissociation constant, in the low pico-molar region. The data were fitted using NanoAnalyze (TA Instruments).

A



B



C

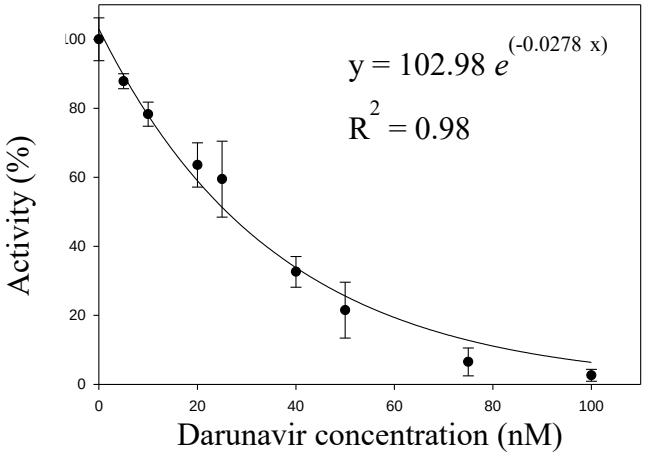


Figure 31: IC_{50} determination of (A) atzanavir, (B) ritonavir and (C) darunavir for C-SA PR

The correlation coefficient of the data sets, were 0.99, 0.98 and 0.98, for atazanavir, ritonavir and darunavir, respectively. Data was fitted with a two parameter exponential decay model using SigmaPlot® v 11.0. The error bars are standard deviation values obtained from the experiment performed in triplicate.

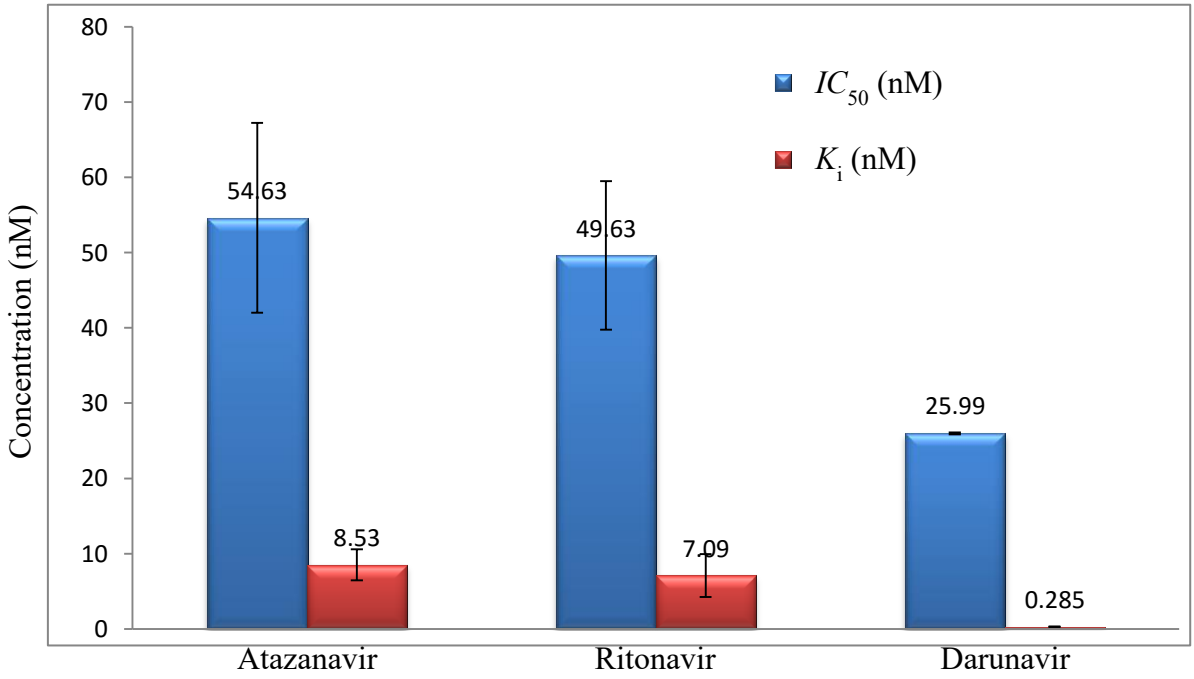


Figure 32: Graphical representation of the IC_{50} and K_i values of atazanavir, ritonavir and darunavir for C-SA PR

Atazanavir and ritonavir have comparable inhibitory potencies. However, ritonavir is a stronger inhibitor of C-SA PR than atazanavir. Darunavir inhibits C-SA PR with the highest potency between the three inhibitors. The binding of darunavir is significantly tighter, having an IC_{50} values close to 25 (value expected for irreversible inhibitors). The error bars are standard deviation values obtained from the experiments performed in triplicate. The graph was produced using Microsoft® Office Excel 2010.

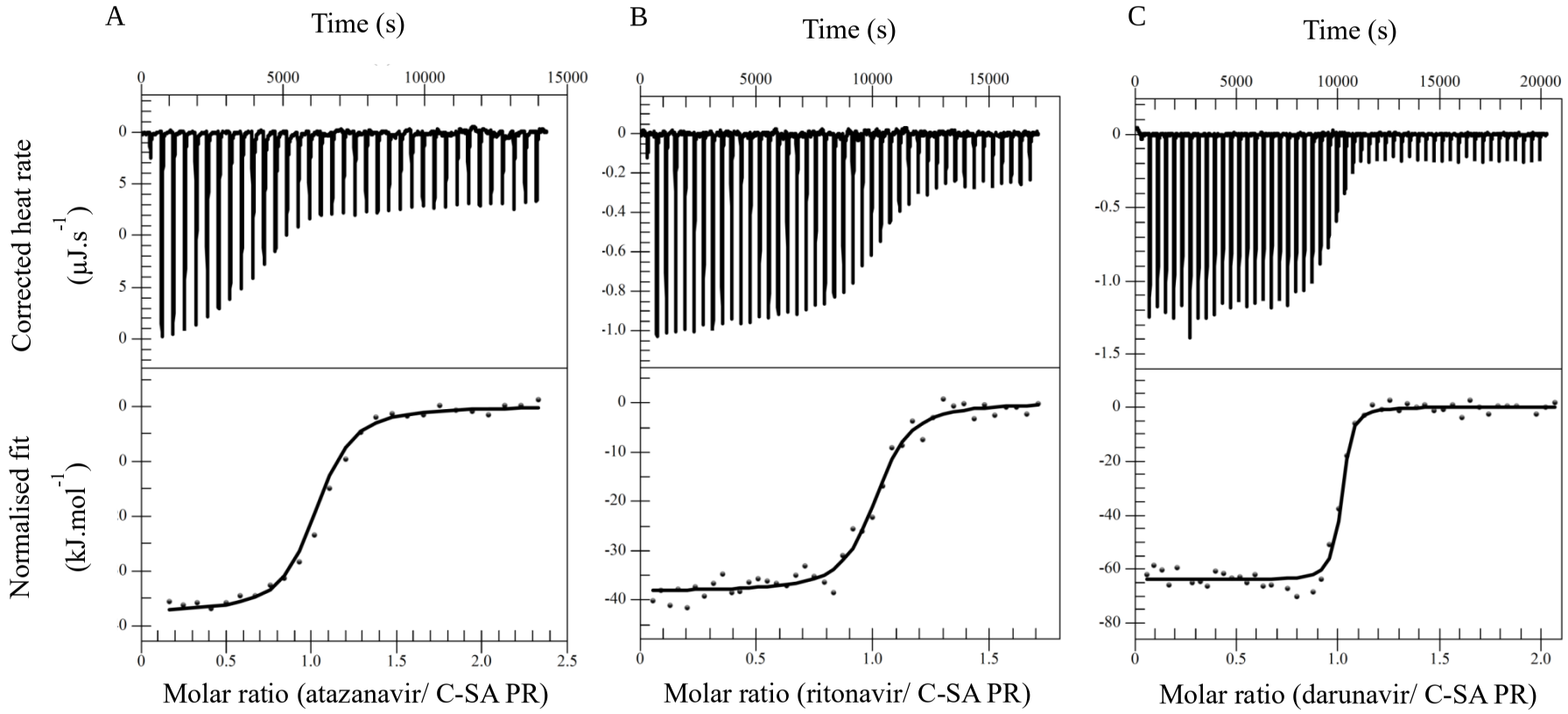


Figure 33: Displacement titration thermograms of (A) atazanavir, (B) ritonavir and (C) darunavir against C-SA PR complexed with acetyl-pepstatin

The data were fitted with a competitive-replacement model using NanoAnalyze (TA Instruments). The thermograms display the heat signals as well as the goodness of the fit. The standard deviation values of the data-fits were 1.83, 1.24 and 1.50 for atazanavir, ritonavir and darunavir, respectively. The errors of the thermodynamic parameters are displayed in Figure 34.

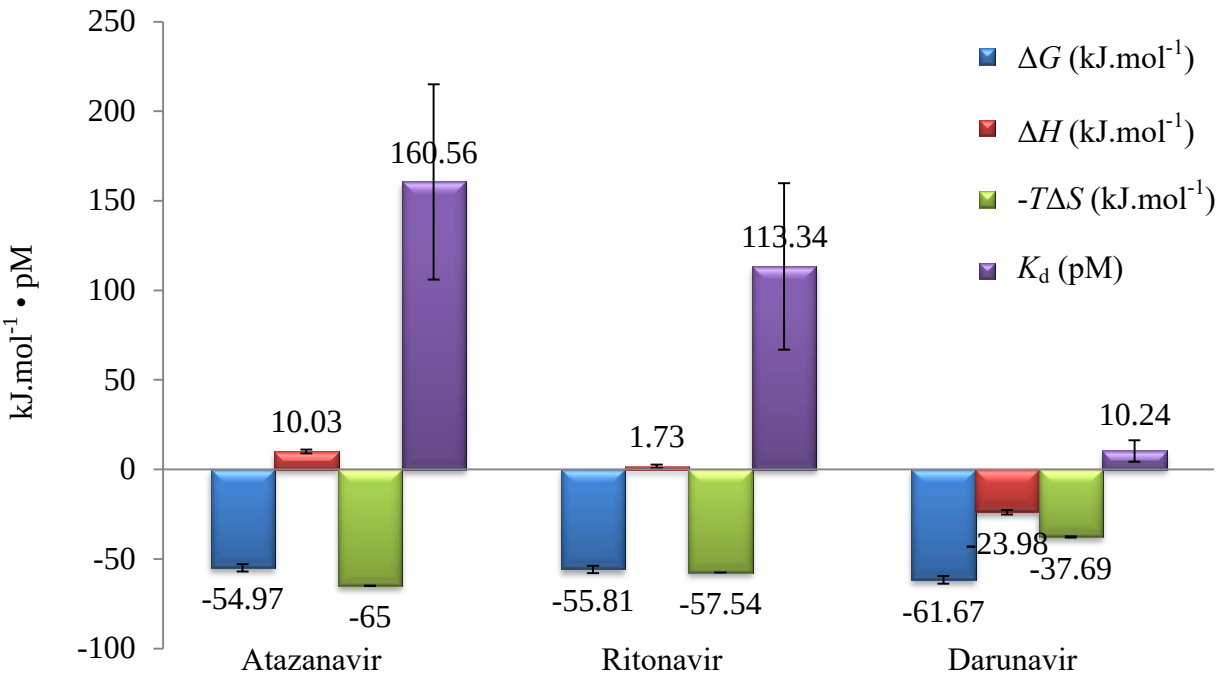


Figure 34: Thermodynamics of binding of atazanavir, ritonavir and darunavir to C-SA PR

Atazanavir and ritonavir have similar binding energetics. However, darunavir binds tighter than atazanavir and ritonavir. In addition, the binding of darunavir to C-SA PR is enthalpically favourable, as opposed to atazanavir and ritonavir. The units of ΔG, ΔH and $-T\Delta S$ are in kJ.mol^{-1}. The dissociation constant, K_d, is standardised in pM units. Data was fitted with NanoAnalyze. The error bars are resultant of the standard deviations of fit. The graph was produced using Microsoft® Office Excel 2010.

5 Discussion

5.1 Solubilisation of lyophilised E35D↑G↑S and confirmation of activity

E35D↑G↑S is a naturally occurring double insertion variant of C-SA PR. Chemical synthesis is an alternate and viable option of obtaining proteins that may prove problematic to express in biological systems (Nillson *et al*., 2005). A chemically synthesised, lyophilised sample of E35D↑G↑S was obtained. The lyophilised E35D↑G↑S sample contained smaller peptides; this is not uncommon for chemically synthesised amino-acid chains (Mueller *et al*., 2012). The protease seemed to be prone to autocatalysis since most of the sample appeared to have been degraded after overnight dialysis-assisted refolding (Figure 13; Figure 14 and Figure 15). Some of the lyophilised E35D↑G↑S was active after 30 minutes of dialysis since electrophoresed protein bands, typical of autocatalysis, were observed (Figure 14 A and Figure 14). Figure 14 B is representative of the proteolytic autocatalysis exhibited by the E35D↑G↑S protease. Although the correlation co-efficient of the data fit in Figure 14 B is only 0.93, it is clear that autocatalysis occurs upon refolding of the protease.

The Q7K autocatalysis-resistant mutation was present in both C-SA PR and E35D↑G↑S protease variants. However, there are two other mutations (L33I and L63I) that have been shown to prevent autocatalysis (Mildner *et al*., 1994). These mutations were not incorporated into the protease sequences resulting in autocatalytic activity. It is recommended that the L33I and L63I mutations are introduced in the protease sequences before further experimentation.

Reducing the period of dialysis-assisted refolding from overnight to just 1 hour and 15 minutes at 4 °C, has afforded a decrease in the period allowed for autocatalysis. However, the E35D↑G↑S remained autocatalytically active as observed on the 16% polyacrylamide gel in Figure 15. Additionally, the experimental preparation of the active-site titration experiment (Figure 16) involved the thermal equilibration of the E35D↑G↑S sample to 20 °C. The thermal equilibration process was approximately 2 hours. Consequently, the protease may have been more prone to autocatalysis at the elevated temperature. As a result, the amount of active protease remaining in the sample cell may have been considerably reduced. Thus, the failure to reach saturation during the active site titration (Figure 16) may be an effect of the lack of active E35D↑G↑S. On the other hand, 1 hour and 15 minutes of dialysis may have been insufficient for the removal of residual urea. The solvation of urea in water is endothermic (Della Gatta *et al*., 2007). Residual urea may have contributed to the

endothermic signal of the active site titration, thus, interfering with the thermodynamic profile. Furthermore, the apparent entropically driven binding of acetyl-pepstatin to E35D↑G↑S (Figure 16), can be explained by the chaotropic properties of urea. The solvation of urea increases the entropy of the system (Moelbert, *et al*., 2004).

5.2 Expression trials of E35D↑G↑S in *Escherichia coli* BL21(DE3)pLysS cells at 20 ˚C and 37 ˚C

Escherichia coli is a commonly used host for protein expression. *E. coli* has minimal requirements for growth, replicates rapidly and the genetic constitution of *E. coli* is well-understood. Overexpression of proteins in *E. coli* is a preferred method of protein production as it is cost-effective and reproducible once conditions are established (Baneyx, 1999). The overexpression of E35D↑G↑S became an objective in this study due to the limited supply of the lyophilised E35D↑G↑S sample. The wild type variant, C-SA PR can be overexpressed and recovered from *E. coli* BL21(DE3)pLysS. Therefore, expression trials of the E35D↑G↑S protease were set up using *E. coli* BL21(DE3)pLysS cells.

Expression trials of E35D↑G↑S were unsuccessful (Figure 17 A and B) since expression was not observed at varying temperatures (20 ºC and 37 ºC), concentrations of IPTG (0 - 1 mM) and induction periods (4 hours, 6 hours and overnight). It was speculated that overexpression was prevented by the autocatalytic and cytotoxic properties of the E35D↑G↑S protease. In order to verify that, an inactive E35D↑G↑S protease incorporating the active site mutation, D25A, was produced using site-directed mutagenesis.

5.3 Expression trials of the inactive mutant E35D↑G↑S D25A in *E. coli* BL21(DE3)pLysS and *E. coli* Rosetta 2® cells

A D25A active site mutation was produced for both C-SA PR and E35D↑G↑S. Expression trials were only performed with the E35D↑G↑S D25A protease because E35D↑G↑S, unlike C-SA PR, did not overexpress.

The aspartic acid residue at position 25 was selected for mutation because it is the catalytic residue responsible for the acid-base hydrolysis of the substrate. It has been shown that a D25N mutation in HIV-1 proteases impairs dimerisation (Sayer *et al.*, 2008). Therefore, the aspartic acid at position 25 was substituted with an alanine residue. The D25A mutation inactivates HIV-1 proteases without affecting dimerisation (Koh *et al*., 2011).

Expression trials did not produce any detectable quantities of E35D↑G↑S D25A in *E. coli* BL21(DE3)pLysS at 37 ºC regardless of the IPTG concentrations and induction periods (Figure 18). Hence, it was deduced that the codons present in *E. coli* BL21(DE3)pLysS are not compatible with the E35D↑G↑S gene sequence used. Although C-SA PR does express, the gene sequence of C-SA PR was not codon harmonised with the expression host. Since E35D↑G↑S was designed using the C-SA PR gene sequence template, codon mismatches were inherent. Codon mismatches are known to hinder protein expression (Angov *et al*., 2011). For that reason, *E. coli* Rosetta 2® cells were used in expression trials of E35D↑G↑S D25A as the Rosetta 2® cell line contains seven rare codons.

Overexpression of E35D↑G↑S D25A was not detected in Rosetta 2® cells (Figure 19 A and B). The lack of expression of the inactive E35D↑G↑S D25A protease reiterates the codon incompatibility of the E35D↑G↑S gene sequence with the expression host. The sequence of E35D↑G↑S ought to be harmonised with the expression host. Moreover, codon harmonisation may also improve the yield of both wild type C-SA PR and E35D↑G↑S variants recovered from expression hosts (Angov *et al*., 2011).

5.4 Pre-modified purification of C-SA PR

The wild type C-SA PR was refolded in 10% glycerol with 10 mM formic acid and 0.02% sodium azide. Contaminating cellular contents from the *E. coli* expression host precipitated when dialysed in activation buffer (10 mM sodium acetate, pH 5.00, containing 0.02% sodium azide). A significant amount of remaining contaminating proteins were eluted in the flow through using a CM-sepharose column (Figure 20 A and B). The NaCl continuous gradient used to elute C-SA PR removed a significant amount of a contaminant with a molecular weight of 15 kDa (Figure 20 A and B). However, the eluted C-SA PR still contained some contaminating proteins, probably because of the similar charge of the contaminant allowing simultaneous binding and elution of the contaminant and C-SA PR from the CM-sepharose column. The gel displayed in Figure 20 B appears to have been overloaded with highly concentrated precipitated fraction in refolding buffer. The overloaded sample may have spilled over into neighbouring wells. Nevertheless, the UFLC chromatogram in Figure 21 confirms the presence of contaminating proteins. The identity of the contaminating proteins cannot be determined with the technique. However, a western blot analysis of SDS-PAGE gels could reveal if the contaminating protein is unprocessed PR from the gag-pol region or other protein contaminant.

The percentage contribution of the intensity of the peak labelled C-SA PR to the sum total of the intensities of all the peaks in the UFLC chromatogram (Figure 21) is 56%. This value is strikingly similar to the percentage of active sites in the same sample (~55%) determined calorimetrically with an active site titration (Figure 22 A and B). This correlation between the data sets confirms that acetyl-pepstatin does not bind non-specifically. Attesting further to this fact, are the similar thermodynamic parameters (ΔH = 38.1 kJ/ mol; K_d = 0.25 μM) of the active site titration found in literature (Naicker *et al.*, 2014).

5.5 Modified purification of C-SA PR

Activating C-SA PR in sodium acetate buffer, pH 5.00, containing 0.02% sodium azide, 2mM DTT and 2 mM NaCl, allowed for the improved removal of contaminating proteins from the C-SA PR sample. The effect is observed when comparing the purification profile obtained using the modified protocol (Figure 23 and Figure 24) with the purification profile obtained using the pre-modified protocol (Figure 20 and Figure 21). A significant amount of contaminant has been eliminated using the modified protocol.

The band immediately above the C-SA PR band in the "concentrated C-SA PR" lane in Figure 24 A, may be C-SA PR that has not been autocatalysed out of the gag-pol region of the recombinant protein, which may contribute to the percentage of active C-SA PR. This could be confirmed by performing a western blot analysis of gels revealing a similar profile. Alternatively, contaminant may be present. The contribution of C-SA PR to the sum total of the intensities of all the peaks in the UFLC chromatogram (Figure 24 B) is 69% which exceeds the 56% contribution when using the pre-modified purification protocol. In addition, the 69% contribution is close to the 74% active C-SA PR determined calorimetrically (Figure 25). This data further verifies that acetyl-pepstatin does not bind non-specifically.

The use of DTT in the purification of C-SA PR has a beneficial effect. However, it is important that DTT is removed from samples prior to experimentation because HIV-1 protease has been shown to have elevated catalytic activity in the presence of the reducing agent (Davis *et al.*, 1999). The reason why the sample purified in the presence of DTT may yield protein of higher purity despite elevated autocatalytic activity conferred by DTT is as follows: Autocatalysed specimens in samples without DTT may be held together by the disulphide bonds created by the non-reduced cysteine residues, thus, retaining the isoelectric point of the protease and eluting together with the intact specimens from the CM-sepharose column. However, autocatalysed specimens will not be held together by disulphide bonds in

the presence of DTT, thus, their isoelectric points would change and would not elute together with the intact specimens from the CM-sepharose column. After purification, DTT is removed immediately through dialysis. Nevertheless, the mutation of the two conserved cysteine residues to alanine residues (C67A and C95A) may be cost effective by eliminating the need for a reducing agent. However, the C67A and C95A have been shown to increase the catalytic activity of HIV-1 protease by as high as 5 fold (Davis *et al*., 1999). Therefore, it may not be solely beneficial to introduce the C67A and C95A mutations as this may result in excessive use of substrate.

5.6 Structural characterisation of C-SA PR

5.6.1 Far-UV circular dichroism – secondary structural characterisation

The far-UV CD spectra of C-SA PR reveals that the secondary structure of the protein is characteristically beta-sheeted (Woody, 1995; Mosebi *et al*., 2008; Corrêa & Ramos, 2009). More specifically, the trough at 216 nm (Figure 26) indicates that the beta strands form antiparallel beta sheets (Kelly & Price, 2000; Ishima *et al*., 2001; Frutos *et al*., 2007). The peak at 203.5 nm is somewhat out of the range of a beta sheeted profile. This is likely due to ellipticity contributed by the large proportion of random coils and the 4-amino-acid alpha helix in each monomer of C-SA PR (Kelly & Price, 2000).

Considering the far-UV CD spectra characteristics, a thermal unfolding curve was produced using a 216 nm wavelength (data not shown). However, the signal-to-noise ratio was too low for the data to display a sigmoidal trend. This made the identification of an inflection point difficult. Therefore, the melting point (T_m) of the protein could not be determined with significant accuracy.

5.6.2 Intrinsic fluorescence – tertiary and quaternary structural characterisation

The fluorescence intensity emitted by C-SA PR (Figure 27) at 295 nm is lower than that emitted at 280 nm (both peaking at 346 nm). That is because at 280 nm, both the tyrosine and tryptophan residues are excited. Moreover, thermal unfolding studies should be performed exciting the sample at 280 nm because; firstly, the signal is more intense, and secondly, because all the tryptophan residues are already solvent-exposed while the tyrosine residue is partially covered by the flap of the protease (Figure 12). Therefore, a red shift in emission wavelength when exciting at 295 nm is unlikely. Moreover, a red shift in the emission wavelength when exciting at 280 nm may not be prominent; thus, the fluorescence spectra of the two protease variants may not be different. Due to the lack of a digitally regulated peltier, a

thermal unfolding study could not be performed using intrinsic fluorescence spectroscopy. The intrinsic fluorescence profile of C-SA PR could have been used to compare quaternary structural differences with E35D↑G↑S, if the fluorescence spectroscopy profile of E35D↑G↑S was obtained.

5.7 Enzyme kinetics of C-SA PR

The kinetic parameters of C-SA PR obtained experimentally in this study are comparable to the kinetic parameters of C-SA PR and subtype B PR found in literature using the same fluorogenic substrate (Naicker *et al*., 2013; Naicker *et al*., 2014) (Table 1). The K_M, specific activity, k_{cat} and k_{cat}/K_M of C-SA PR determined in this study do not differ greatly from that of published studies with C-SA PR and subtype B PR using the same fluorogenic substrate. However, due to the lack of literature revolving around the fluorogenic substrate, it is difficult to conclude on the significance of the small differences between the data sets. Nevertheless, the accurate determination of K_M is necessary in calculating the K_i of the inhibitors (Copeland *et al*., 1995). The K_M, specific activity, k_{cat} and k_{cat}/K_M of C-SA PR (Figure 28, Figure 29 and Figure 30) would have been used to compare the affinity, specificity and catalytic activity of the C-SA PR and E35D↑G↑S protease variants for the capsid/ p2 cleavage site incorporated in the fluorogenic substrate.

5.8 Inhibition kinetics of atazanavir, ritonavir and darunavir against C-SA PR

The concentration of atazanavir, ritonavir and darunavir that reduces the activity of C-SA PR by 50% was determined (Figure 31 and Figure 32). The error of the IC_{50} values for atazanavir (~23%) and ritonavir (~20%) are relatively large. However, this is not uncommon when monitoring the cleavage of substrate fluorometrically. Experimental errors of fluorometrically determined IC_{50} values in literature have been noted to be ranging from 9% to 43% (Naicker *et al*., 2014). In addition, the K_i values of atazanavir, ritonavir and darunavir were calculated using equation 6. The errors of the K_i values are a consequence of the IC_{50} errors.

Atazanavir and ritonavir have similar inhibitory potencies. However, darunavir is a significantly more potent C-SA PR inhibitor. This was also shown by the finding of Naicker *et al*., 2014 regarding both C-SA PR and subtype B PR (Naicker *et al*., 2014) (Table 1). Although the literature indicates that darunavir is a significantly more potent inhibitor of subtype B PR than of C-SA PR, this study indicates the contrary. Once again, the lack of literature in this are containing data generated using the same fluorogenic substrate does not allow for thorough comparisons to be made. However, it is certain that darunavir is the most

Table 2: Comparison of the kinetic parameters of C-SA PR obtained experimentally in this study with the kinetic parameters of C-SA PR and subtype B PR found in literature

	Experimentally determined in this study	C-SA PR (Naicker *et al*., 2013)	C-SA PR (Naicker *et al*., 2014)	Subtype B PR (Naicker *et al*., 2014)
K_M (μM)	20.22 ±4.09	–	29.4 ±4.4	38.7 ±6.0
Specific activity (μmoles.min^{-1}.mg^{-1})	35.68 ±1.06	79	–	–
k_{cat} (s^{-1})	12.79 ±0.38	–	20.6 ±0.8	9.0 ±0.2
k_{cat}/ K_M (s^{-1}.μM^{-1})	1.17 ±0.06	–	0.46 ±0.5	0.32 ±0.1
k_i ATV (nM)	8.53 ±2.06	–	5.53 ±1.88	10.96 ± 4.29
k_i RTV(nM)	7.09 ±2.84	–	18.25 ±1.72	6.47 ±2.17
k_i DRV (nM)	0.285 ± 0.036	–	3.30 ±1.42	0.49 ±0.20

potent PI that can be used in ART.

The thermodynamics of binding of darunavir to C-SA PR provides insight into the molecular mechanism that makes darunavir a stronger-binding protease inhibitor than atazanavir and ritonavir.

5.9 Thermodynamics of binding of atazanavir, ritonavir and darunavir to C-SA PR

The thermodynamics of binding of atazanavir, ritonavir and darunavir to C-SA PR were determined in displacement titration calorimetry experiments (Figure 33 and Figure 34). The errors of the K_d values are proportionally large, although the standard deviation of the fitted data sets were not excessive (Figure 33). It is likely that the K_d errors are large because of the standard deviation of the fit across the transition regions of the sigmoidal thermograms of the titrations (Figure 33). Slight deviations in the fitted gradient of the transition region can reflect as large errors of the K_d, especially since the K_d values are in the picomolar range. Moreover, it has been noted that for tight binding systems (K_d in the nanomolar range), data fitted with NanoAnalyze (TA Instruments) has produced errors as high as 97% of the K_d value (Bechara *et al*., 2013).

The binding of atazanavir and ritonavir to C-SA PR is entropically driven and enthalpically opposed. This is probably due to the large size of the inhibitors. As seen in Figure 9 A and B, the two inhibitors protrude extensively out of the active site of HIV-1 protease, unlike darunavir (Figure 8). Indeed, the large molecules would displace a significant amount of water molecules contributing positively to the solvent entropy. Thus, the binding of atazanavir and ritonavir to C-SA PR is entropically favourable. However, the energy required for displacing the water molecules hydrogen-bonded with the protease may not be compensated by the energy released by the hydrogen bonds made by the inhibitors with the protease. Thus, the enthalpy of binding of atazanavir and ritonavir to C-SA PR is unfavourable (Figure 34). Similarly, the binding of ritonavir to C-SA PR was determined to be unfavourable by Naicker *et al*., 2014 (8.37 $kJ.mol^{-1}$), conversely, favourable by Mosebi *et al*., 2008 (-17.57 $kJ.mol^{-1}$). Nevertheless, the enthalpic opposition is not severe. It is possible that the enthalpy of the binding of atazanavir and ritonavir to C-SA PR have low enough magnitudes to make it difficult to determine whether the binding enthalpy is favourable or not.

Darunavir fits into the active site of HIV-1 protease without excessively protruding outwards. Not surprisingly, darunavir binds to C-SA PR in an entropically and enthalpically favourable manner (Figure 34). By far, darunavir has the lowest dissociation constant, in the picomolar region. The favourable entropy of binding of darunavir to C-SA PR is most probably caused by displacement of water molecules from the active site region of the HIV-1 protease. Moreover, it has been elucidated that differences in enthalpic binding between inhibitors is due to the packing of water molecules around the PI-PR complex rather than the number and location of hydrogen bonds made by PIs with the PR (Mittal *et al*., 2013). The smaller size of darunavir may allow more water molecules to pack around the PI-PR complex. Nonetheless, the hydrogen bonding profiles of atazanavir and darunavir are very similar (Mittal *et al*., 2013).

The thermodynamic data for the three inhibitors tested correlates with the determined K_i values. That is, the isothermally determined dissociation constant, K_d, is approximately one order of magnitude lower than the K_i. That indicates that the fluorogenic substrate used competes with the inhibitors, increasing their dissociation constants.

5.10 Crystallisation of HIV-1 proteases

Numerous crystallisation trials for C-SA PR and E35D↑G↑S were attempted. Although the first crystal structure of HIV-1 protease was obtained from chemically synthesised HIV-1 protease (Wlodawer *et al*., 1989), the synthetic E35D↑G↑S sample used in this research did not crystallise. Crystal trials require high concentrations of protein. The process of crystallisation of proteases is often made challenging due to the inherent autocatalytic properties of these enzymes. Although the quantities of proteases available were a limiting factor there is a severe lack of structural data regarding C-SA PR and variants of the subtype. Therefore, efforts need to be made to produce crystallographic data albeit resource expensive.

It is recommended that the L33I and L63I autocatalysis-resistant mutations (Mildner *et al*., 1994) are introduced in the protease sequences so that the formation of crystals is not as limited by autocatalysis.

6 Conclusion

The use of lyophilised E35D↑G↑S has revealed that the protease is exceptionally autocatalytic, which is problematic for further experimentation. Failure to overexpress

E35D↑G↑S as well as the inactive E35D↑G↑S D25A, indicates the codons are incompatible with the cells. Codon optimisation of the C-SA PR and E35D↑G↑S protease gene sequences is necessary in order to clarify the reasons for the inability of E35D↑G↑S to be overexpressed. The experimentally determined kinetic parameters, including those of inhibition kinetics, as well as thermodynamic parameters of C-SA PR were determined successfully. Darunavir has the advantage of making enough hydrogen bonds (7 hydrogen bonds) with HIV-1 protease to bind enthalpically and is small enough to fit in the active site without protruding outwards and displacing too many water molecules. Thus, the binding of darunavir was established to be entropically favourable. The larger size of atazanavir and ritonavir may be the factor that compromises the enthalpically favourable binding of the two inhibitors to C-SA PR.

7 References

Angov, E., Legler, P. M., & Mease, R. M. (2011). Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Heterologous Gene Expression in E. coli*, 1-13. Humana Press.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22:** 195-201.

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, **10:** 411-421.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rozioux, C., Rozenbaum, W., Montagnier, L. (1983). Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science*, **220:** 868-871.

Bechara, C., Pallerla, M., Zaltsman, Y., Burlina, F., Alves, I. D., Lequin, O., & Sagan, S. (2013). Tryptophan within basic peptide sequences triggers glycosaminoglycan-dependent endocytosis. *The FASEB Journal*, **27:** 738-749.

Beilhartz, G. L., & Götte, M. (2010). HIV-1 ribonuclease H: structure, catalytic mechanism and inhibitors. *Viruses*, **2:** 900-926.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, **42:** W252-W258.

Binley, J. M., Sanders, R. W., Master, A., Cayanan, C. S., Wiley, C. L., Schiffner, L., Travis, B., Kuhmann, S., Burton, D. R., Hu, S. L., and Olson, W. C. (2002). Enhancing the proteolytic maturation of human immunodeficiency virus type 1 envelope glycoproteins. *Journal of virology*, **76:** 2606-2616.

Bogerd, H. P., Doehle, B. P., Wiegand, H. L., & Cullen, B. R. (2004). A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. *Proceedings of the National Academy of Sciences of the United States of America*, **101:** 3770-3774.

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

Carmel, A., & Yaron, A. (1978). An Intramolecularly Quenched Fluorescent Tripeptide as a Fluorogenic Substrate of Angiotensin-I-Converting Enzyme and of Bacterial Dipeptidyl Carboxypeptidase. *European Journal of Biochemistry*, **87:** 265-273.

Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., & Vogt, P. (1986[1]). What to call the AIDS virus? *Nature*, **321:** 10.

Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Orsozlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogot, P. (1986[2]). Human immunodeficiency viruses. *Science*. **232:** 697.

Colman, P. M., Lawrence, M. C., Walter, T., & Hall, E. (2003). The structural biology of type I viral membrane fusion. *Nature Reviews Molecular Cell Biology*, **4:** 309-319.

Corrêa, D. H., & Ramos, C. H. (2009). The use of circular dichroism spectroscopy to study protein folding, form and function. *African Journal of Biochemistry Research*, **3:** 164-173.

Copeland, R. A., Lombardo, D., Giannaras, J., & Decicco, C. P. (1995). Estimating K_I values for tight binding inhibitors from dose-response plots. *Bioorganic & Medicinal Chemistry Letters*, **5:** 1947-1952.

Davis, D. A., Yusa, K., Gillim, L. A., Newcomb, F. M., Mitsuya, H., & Yarchoan, R. (1999). Conserved cysteines of the human immunodeficiency virus type 1 protease are involved in regulation of polyprotein processing and viral maturation of immature virions. *Journal of Virology*, **73:** 1156-1164.

De Clercq, E. (2006). Anti-HIV drugs. *Verhandelingen-Koninklijke Academie voor Geneeskunde van Belgie*, **69:** 81-104.

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

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

Della Gatta, G., Badea, E., Józwiak, M., & Del Vecchio, P. (2007). Thermodynamics of solvation of urea and some monosubstituted N-alkylureas in water at 298.15 K. *Journal of Chemical & Engineering Data*, **52:** 419-425.

Dixon, M. (1953). The determination of enzyme inhibitor constants. *Biochemical Journal*, **55:** 170-171.

Duff Jr, M. R., Grubbs, J., & Howell, E. E. (2011). Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *Journal of Visualized Experiments: JoVE*, **55:** e2796.

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

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

Fauci, A. S. (2003). HIV and AIDS: 20 years of science. *Nature Medicine*, **9:** 839-843.

Frankel, A. D., & Young, J. A. (1998). HIV-1: fifteen proteins and an RNA.*Annual review of biochemistry*, **67:**1-25.

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

Freyer, M. W., & Lewis, E. A. (2008). Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ ligand binding and kinetic interactions. *Methods in Cell Biology*, **84:** 79-113.

Frutos, S., Rodriguez-Mias, R. A., Madurga, S., Collinet, B., Reboud-Ravaux, M., Ludevid, D., & Giralt, E. (2007). Disruption of the HIV-1 protease dimer with interface peptides: Structural studies using NMR spectroscopy combined with [2-13C]-Trp selective labeling. *Peptide Science*, **88:** 164-173.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, **31:** 3784-3788.

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

Ghosh, A. K., Chapsal, B. D., Weber, I. T., & Mitsuya, H. (2007[1]). Design of HIV protease inhibitors targeting protein backbone: an effective strategy for combating drug resistance. *Accounts of Chemical Research*, **41:** 78-86.

Ghosh, A. K., Dawson, Z. L., & Mitsuya, H. (2007[2]). Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drug-resistant HIV. *Bioorganic & Medicinal Chemistry*, **15:** 7576-7580.

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

Guex, N., Peitsch, M. C., Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, **30:** S162-S173.

Haider, S. R., Sharp, B. L., & Reid, H. J. (2011). A comparison of Tris-glycine and Tris-tricine buffers for the electrophoretic separation of major serum proteins. *Journal of Separation Science*, **34:** 2463-2467.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology*, **166:** 557-580.

Hemelaar, J., Gouws, E., Ghys, P. D., & Osmanov, S. (2006). Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *Aids*, **20:** W13-W23.

Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine*, **18**: 182-192.

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

Hout, D. R., Gomez, M. L., Pacyniak, E., Mulcahy, E. R., Gomez, L. M., Jackson, M., Flick, M., Fegley, B., McCormick, C., Wisdom, B. J. & Culley, N. (2004). Fusion of the upstream vpu sequences to the env of simian human immunodeficiency virus (SHIV KU-1bMC33) results in the synthesis of two envelope precursor proteins, increased numbers of virus particles associated with the cell surface and is pathogenic for pig-tailed macaques. *Virology*, **323:** 91-107.

Hu, W., Kaminski, R., Yang, F., Zhang, Y., Cosentino, L., Li, F., Luo, B., Alvarez-Carbonell, D., Garcia-Mesa, Y., Karn, J., Mo, X., and Khalili, K. (2014). RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proceedings of the National Academy of Sciences*, **111:** 11461-11466.

Hyland, L. J., Tomaszek Jr, T. A., Roberts, G. D., Carr, S. A., Magaard, V. W., Bryan, H. L., Fakhoury, S. A., Moore, M. L., Minnich, M. D., Culp, J. S., DesJarlais, R. L., & Minnich, M. D. (1991). Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by oxygen-18 isotope exchange. *Biochemistry*, **30:** 8441-8453.

Ishima, R., Ghirlando, R., Tözsér, J., Gronenborn, A. M., Torchia, D. A., & Louis, J. M. (2001). Folded monomer of HIV-1 protease. *Journal of Biological Chemistry*, **276:** 49110-49116.

Jeang, K. T., Xiao, H., & Rich, E. A. (1999). Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *Journal of Biological Chemistry*, **274:** 28837-28840.

Johnson, K. A., & Goody, R. S. (2011). The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry*, **50:** 8264-826.

Keele, B. F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M. L., Bibollet-Ruche, F., Chen, Y., Wain, L. V., Liegeois, F., Loul, S., Ngole, E. M., Bienvenue, Y., Delaporte, E., Brookfield, J. F. Y., Sharp, P. M., Shaw, G. M., Peeters, M., Hahn, B. H. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, **313:** 523-526.

Kelly, S. M., & Price, N. C. (2000). The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science*, **1**: 349-384.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*, **37:** D387-D392.

Koh, Y., Aoki, M., Danish, M. L., Aoki-Ogata, H., Amano, M., Das, D., Shafer, R. W., Ghosh, A. K., & Mitsuya, H. (2011). Loss of protease dimerization inhibition activity of darunavir is associated with the acquisition of resistance to darunavir by HIV-1. *Journal of Virology*, **85:** 10079-10089.

Kožíšek, M., Lepšík, M., Grantz Šašková, K., Brynda, J., Konvalinka, J., & Řezáčová, P. (2014). Thermodynamic and structural analysis of HIV protease resistance to darunavir–analysis of heavily mutated patient-derived HIV-1 proteases. *FEBS Journal*, **281:** 1834-1847.

Kožíšek, M., Šašková, K. G., Řezáčová, P., Brynda, J., van Maarseveen, N. M., De Jong, D., Boucher, C. A., Kagan, R. M., nijhuis, M., & Konvalinka, J. (2008). Ninety-nine is not enough: molecular characterization of inhibitor-resistant human immunodeficiency virus type 1 protease mutants with insertions in the flap region. *Journal of Virology*, **82:** 5869-5878.

Kuo-Chen, C. (1975). The kinetics of the combination reaction between enzyme and substrate. *Scientia Sinica*, **19:** 505-528.

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

Lakowicz, J. R. (2006). Principles of Fluorescence Spectroscopy. 3rd edition. *Springer*. New York.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23:** 2947-2948.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, **26:** 283-291.

Laskowski, R. A., MacArthur, M. W., & Thornton, J. M. (2001). PROCHECK: validation of protein structure coordinates. *International Tables of Crystallography, Vol. F. Crystallography of Biological Macromolecules. Kluwer Academic Publishers, The Netherlands*, 722-725.

Lennox, E. S. (1955). Transduction of linked genetic characters of the host by bacteriophage P1. *Virology*, **1:** 190-206.

Lin, S. X., & Lapointe, J. (2013). Theoretical and experimental biology in one—A symposium in honour of Professor Kuo-Chen Chou's 50th anniversary and Professor Richard Giegé's 40th anniversary of their scientific careers. *Journal of Biomedical Science and Engineering*, **6:** 435-442.

Mammano, F., Petit, C., & Clavel, F. (1998). Resistance-Associated Loss of Viral Fitness in Human Immunodeficiency Virus Type 1: Phenotypic Analysis of Protease and gag Coevolution in Protease Inhibitor-Treated Patients. *Journal of Virology*, **72:** 7632-7637.

Marx, J. L. (1982). New disease baffles medical community. *Science*, **217:** 618-621.

Mildner, A. M., Rothrock, D. J., Leone, J. W., Bannow, C. A., Lull, J. M., Reardon, I. M., Sarchich, J. L., Howe, W. J., Tomich, C. S. C., Smith, C. W., Heinrikson, R. L., & Tomasselli, A. G. (1994). The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. *Biochemistry*, **33:** 9405-9413.

Miller, J. H. (1972). Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Mittal, S., Bandaranayake, R. M., King, N. M., Prabu-Jeyabalan, M., Nalam, M. N., Nalivaika, E. A., Yilmaz, C. A., & Schiffer, C. A. (2013). Structural and thermodynamic basis of amprenavir/ darunavir and atazanavir resistance in HIV-1 protease with mutations at residue 50. *Journal of Virology*, **87:** 4176-4184.

Moelbert, S., Normand, B., & De Los Rios, P. (2004). Kosmotropes and chaotropes: modelling preferential exclusion, binding and aggregate stability. *Biophysical Chemistry*, **112:** 45-57.

Moffatt, B. A., & Studier, F. W. (1987). T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell*, **49:** 221-227.

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

Mueller, K., Chinchilla, D., Albert, M., Jehle, A. K., Kalbacher, H., Boller, T., & Felix, G. (2012). Contamination risks in work with synthetic peptides: flg22 as an example of a pirate in commercial peptide preparations. *The Plant Cell*, **24:** 3193-3197.

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

Naicker, P., Stoychev, S., Dirr, H. W., & Sayed, Y. (2014). Amide hydrogen exchange in HIV–1 subtype B and C proteases–insights into reduced drug susceptibility and dimer stability. *FEBS Journal*, **281:** 5395-5410.

Niederman, T. M., Thielan, B. J., & Ratner, L. (1989). Human immunodeficiency virus type 1 negative factor is a transcriptional silencer. *Proceedings of the National Academy of Sciences*, **86:** 1128-1132.

Nilsson, B. L., Soellner, M. B., & Raines, R. T. (2005). Chemical synthesis of proteins. *Annual Review of Biophysics and Biomolecular Structure*, **34:** 91-118.

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

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

Richman, D. D. (2001) Review article HIV chemotherapy. *Nature*, **410:** 995-1001.

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

Royer, C. A. (1995). Fluorescence spectroscopy. *Protein Stability and Folding*, 65-89. Humana Press.

Safai, B., Groopman, J. E., Popovic, M., Schurpbach, J., Sarangadharan, M. G., Arnett, K., Sliski, A., Gallo, R. C. (1984). Seroepidemiological studies of human T-lymphotropic retrovirus type III in acquired immunodeficiency syndrome. *The Lancet*, **323:** 1438-1440.

Sayer, J. M., Liu, F., Ishima, R., Weber, I. T., & Louis, J. M. (2008). Effect of the active site D25N mutation on the structure, stability, and ligand binding of the mature HIV-1 protease. *Journal of Biological Chemistry*, **283**: 13459-13470.

Schägger, H. (2006). Tricine–SDS-PAGE. *Nature Protocols*, **1:** 16-22.

Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and biophysical research communications*, **27:** 157-162.

Schlager, B., Straessle, A., & Hafen, E. (2012). Use of anionic denaturing detergents to purify insoluble proteins after overexpression. *BMC Biotechnology*, **12:** 95.

Schubert, U., & McClure, M. (2005). Human immunodeficiency virus. *Topley and Wilson's Microbiology and Microbial Infections*.

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, **31:** 3381-3385.

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

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

Sharp, P. M., Bailes, E., Chaudhuri, R. R., Rodenburg, C. M., Santiago, M. O., Hahn, B. H. (2001). The origins of acquired immune deficiency syndrome viruses: where and when?. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **356:** 867-876.

Sharp, P. M., & Hahn, B. H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine*, **1:** 1-23.

Stano, N. M., & Patel, S. S. (2004). T7 lysozyme represses T7 RNA polymerase transcription by destabilizing the open complex during initiation. *Journal of Biological Chemistry*, **279:** 16136-16143.

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, **189:** 113-130.

Teixeira, C., Gomes, J. R., Gomes, P., & Maurel, F. (2011). Viral surface glycoproteins, gp120 and gp41, as potential drug targets against HIV-1: brief overview one quarter of a century past the approval of zidovudine, the first anti-retroviral drug. *European Journal of Medicinal Chemistry*, **46:** 979-992.

Terwilliger, E. F., Cohen, E. A., Lu, Y. C., Sodroski, J. G., & Haseltine, W. A. (1989). Functional role of human immunodeficiency virus type 1 vpu. *Proceedings of the National Academy of Sciences*, **86:** 5163-5167.

Turnbull, W. B., & Daranas, A. H. (2003). On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *Journal of the American Chemical Society*, **125:** 14859-14866.

UNAIDS (2015). HIV estimates with uncertainty bounds 1990-2014.

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

Velázquez-Campoy, A., Vega, S., Fleming, E., Bacha, U., Sayed, Y., Dirr, H. W., & Freire, E. (2003). Protease inhibition in African subtypes of HIV-1. *Aids Reviews*, **5:** 165-171.

Velázquez-Campoy, A., & Freire, E. (2006). Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nature Protocols-Electronic*, **1:** 186-191.

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

Weiss, R. A. (2008). Special anniversary review: twenty-five years of human immunodeficiency virus research: successes and challenges. *Clinical & Experimental Immunology*, **152:** 201-210.

Wensing, A. M., Calvez, V., Günthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., Shafer, R. W., & Richman, D. D. (2014). Special Contribution 2014 Update of the Drug Resistance Mutations in HIV-1. *Topics in Antiviral Medicine*, **22:** 642.

Wiseman, T., Williston, S., Brandts, J. F., & Lin, L. N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical Biochemistry*, **179:** 131-137.

Wlodawer, A., & Erickson, J. W. (1993). Structure-based inhibitors of HIV-1 protease. *Annual Review of Biochemistry*, **62:** 543-585.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, **245:** 616-621.

Wlodawer, A., & Vondrasek, J. (1998). INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design 1. *Annual Review of Biophysics and Biomolecular Structure*, ***27*:** 249-284.

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

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

8 Supplementary material

```
CLUSTAL W (1.83) multiple sequence alignment

                    Start codon
C-SA                ATGGACCGTCAGGGTACTGTATCTTTCAACTTCCCGCAGATCACTCTGTGGAAACGTCCG
E35DGS              ATGGACCGTCAGGSTWYTGTATCTTTCAACTTCCCGCAGATCACTCTGTGGAAACGTCCG
                    ************* *  ****************************************

C-SA                CTGGTATCCATCAAAGTCGGTGGTCAGATCAAAGAAGCTCTGCTGGACACTGGCGCTGAC
E35DGS              CTGGTATCCATCAAAGTCGGTGGTCAGATCAAAGAAGCTCTGCTGGACACTGGCGCTGAC
                    ************************************************************
                                  E35D↑G ↑S  I36L                          M46L
C-SA                GACACTGTTCTGGAAGAA------ATCAATCTGCCGGGTAAATGGAAGCCGAAAATGATC
E35DGS              GACACTGTTCTGGAAGATGGTAGCTTGAATCTGCCGGGTAAATGGAAGCCGAAATTGATC
                    ****************        * ************************** ******
                                                      D60E
C-SA                GGTGGCATCGGCGGTTTTATCAAAGTTCGTCAGTATGATCAGATCCTGATCGAAATCTGC
E35DGS              GGTGGCATCGGCGGTTTTATCAAAGTTCGTCAGTATGAACAGATCCTGATCGAAATCTGC
                    ************************************ ********************

C-SA                GGTAAAAAAGCTATCGGTACCGTTCTGGTTGGTCCGACTCCGGTTAACATCATCGGCCGT
E35DGS              GGTAAAAAAGCTATCGGTACCGTTCTGGTTGGTCCGACTCCGGTTAACATCATCGGCCGT
                    ************************************************************

C-SA                AACATGCTGACTCAGCTGGGTTGCACTTTGAACTTCCCGATCTCTCCGATTGAAACTGTT
E35DGS              AACATGCTGACTCAGCTGGGTTGCACTTTGAACTTCCCGATCTCTCCGATTGAAACTGTT
                    ************************************************************
                           Stop codon
C-SA                CCGGTTTAA
E35DGS              CCGGTTTAA
                    *********
```

S1: Nucleotide sequence alignment of C-SA PR and E35D↑G↑S

Sequence aligned with Clustal W 1.83. Start and start codons are underlined and annotated. The mutated codons are underlined. The two insertions are highlighted in grey.