

https://dspace.org/

# Antiretroviral drug susceptibility of a hinge region variant of HIV-1 subtype C protease

**Zondagh, Jake**

**2018**

https://hdl.handle.net/10539/25929

Zondagh, Jake (2018) Antiretroviral drug susceptibility of a hinge region variant of HIV-1 subtype C protease, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/25929>

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

# Antiretroviral drug susceptibility of a hinge region variant of HIV-1 subtype C protease

Jake Zondagh (336484)

28 May 2018

A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg in fulfilment of the requirements for the degree of Doctor of Philosophy

Supervisor: Professor Yasien Sayed

Co-supervisor: Doctor Ikechukwu Achilonu

# Declaration

I, Jake Zondagh (student number: 336484), am a student registered for the degree of Doctor of Philosophy (PhD) in the academic year 2018.

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 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:  Date: 28 May 2018.

# Abstract

Since their discovery, protease inhibitors continue to be an essential component of antiretroviral treatment for human immunodeficiency virus type 1 (HIV-1). However, the development of resistance to protease inhibitors remains one of the most significant challenges in the fight for sustained viral suppression in those infected with HIV-1. Studies show that specific mutations arising within the HIV-1 *gag* and *protease* genes can lead to the development of resistance. In this research, a South African HIV-1 subtype C Gag-protease variant (W1201i) was investigated. This variant was considered due to the presence of a mutation and insertion (N37T↑V), located within the hinge region of the protease enzyme. Moreover, the variant displayed the following polymorphisms: Q7K, I13V, G16E, M36T, D60E, Q61E, I62V and M89L. Genotyping of W1201i Gag revealed a previously unreported MSQAG insertion between the CA/p2 and p2/NC cleavage sites. Additionally, a mutation and insertion (I372L↑M), and multiple polymorphisms (S369N, S371N, I373M and G377S) were discovered within the p2/NC cleavage site. Single-cycle phenotypic assays were performed to determine the drug susceptibility and replication capacity of the variant. The results show that the mutations present in the N37T↑V protease conferred a replicative advantage and reduced susceptibility to lopinavir, atazanavir and darunavir. Interestingly, the mutations in W1201i Gag were found to modulate both replication capacity and protease inhibitor susceptibility. *In silico* studies were performed to understand the physical basis for the observed variations. Molecular dynamics simulations showed that the N37T↑V protease displayed altered dynamics around the hinge and flap region and highlighted the amino acids responsible for the observed fluctuations. Furthermore, induced fit docking experiments showed that the variant bound the

protease inhibitors with fewer favourable chemical interactions than the wild-type protease. Collectively, these data elucidate the biophysical basis for the selection of hinge region mutations and insertions by the HI virus and show that protease, as well as Gag, needs to be evaluated during resistance testing.

# Dedication

I dedicate this thesis to:

My Mom and Dad, without your love and support I would never have made it this far.

My brother Michael Zondagh, I can see the world clearly because you allow me to stand on your shoulders.

My ouma Daniëllina Bothma, dankie dat ouma altyd in my glo. Ek sou nie hier uitgekom het sonder ouma se ondersteuning en liefde nie.

My Kimmi, you have given me so much. Thank you for sharing a rowboat with me.

I am truly grateful to you all.

## Acknowledgements

Firstly, I would like to thank my supervisor, Professor Yasien Sayed, co-supervisor, Dr Ikechukwu Achilonu, and advisor, Professor Heini Dirr, for their excellent guidance and supervision throughout my studies.

Thank you to the South African National Research Foundation and the University of the Witwatersrand for funding.

Thank you to all the members of the Protein Structure-Function Research Unit.

Thank you to Professor Lynn Morris for welcoming me into her research unit, and for everything that I learned during my stay there.

Lastly, I would like to thank Dr Adriaan Basson for his assistance with the phenotypic drug susceptibility assays, and for always making time to answer my questions.

# Research Outputs

## Publications forming part of PhD thesis:

1. Jake Zondagh, Alison Williams, Ikechukwu Achilonu, Heini W. Dirr and Yasien Sayed. (2018) Overexpression, purification and functional characterisation of wild-type HIV-1 subtype C protease and two variants using a thioredoxin and his-tag protein fusion system. *Protein. J*., Manuscript accepted.

2. Jake Zondagh, Adriaan E. Basson, Ikechukwu Achilonu, Lynn Morris, Heini W. Dirr and Yasien Sayed. (2018) Susceptibility of an HIV-1 Subtype-C Protease Hinge Region Variant. Manuscript in preparation.

3. Jake Zondagh, Vijayakumar Balakrishnan, Ikechukwu Achilonu, Heini W. Dirr and Yasien Sayed. (2018) Molecular dynamics and ligand docking of a hinge region variant of South African HIV-1 subtype C protease, *J. Mol. Graph. Model*. 82 (2018) 1–11. doi:10.1016/j.jmgm.2018.03.006.

## Conference outputs

1. SASBMB, Goudini Spa, Rawsonville, 2014. *Poster presentation*: "Structure, Function and Thermodynamic Properties of HIV-1 Subtype C Protease Variant I36T↑T." Jake Zondagh and Yasien Sayed.

2. Biophysics in the Understanding, Diagnosis and Treatment of Infectious Diseases, Stellenbosch, South Africa, 2015. *Poster presentation*: "Antiretroviral drug susceptibility of a hinge region variant of HIV-1 subtype C protease." Jake Zondagh, Ikechukwu Achilonu and Yasien Sayed.

3. University of the Witwatersrand Cross faculty postgraduate symposium, Johannesburg, 2017. *Oral presentation:* "Antiretroviral Drug Susceptibility of a Hinge Region Variant of HIV-1 Subtype C Protease." Jake Zondagh, Adriaan E. Basson, Ikechukwu Achilonu, Lynn Morris, Heini W. Dirr and Yasien Sayed.

4. Molecular Biosciences Research Thrust, Johannesburg 2017. *Oral presentation.* "The N37T↑V hinge region insertion mutation alters the molecular dynamic landscape of HIV-1 Subtype C protease." Jake Zondagh, Vijayakumar Balakrishnan, Ikechukwu Achilonu, Heini Dirr and Yasien Sayed.

# Table of contents

Declaration....ii

Abstract.... iii

Dedication....v

Acknowledgements ....vi

Research Outputs....vii

Publications forming part of PhD thesis: ....vii

Conference outputs .... viii

List of Figures....xi

Abbreviations....xii

CHAPTER 1....1

Introduction ....1

1.1 Human immunodeficiency virus....1

1.2 HIV protease ....3

1.3 Function of HIV-1 C-SA protease ....5

1.4 Protease inhibitors....8

1.5 HIV-1 C-SA protease drug susceptibility ....12

1.6 W1201i Gag-protease ....14

1.7 Aim and objectives ....16

CHAPTER 2 ..... 17

Overexpression, purification and functional characterisation of wild-type HIV-1 subtype C protease and two variants using a thioredoxin and his-tag protein fusion system ..... 20

CHAPTER 3 ..... 49

Drug Susceptibility and Replication Capacity of a Rare HIV-1 Subtype-C Protease Hinge Region Variant ..... 50

CHAPTER 4 ..... 75

Molecular dynamics and ligand docking of a hinge region variant of South African HIV-1 subtype C protease ..... 77

CHAPTER 5 ..... 108

General discussion and conclusions ..... 108

5.1 Expression and purification of HIV-1 C-SA proteases ..... 108

5.2 Drug susceptibility and replication capacity of W1201i ..... 110

5.3 Characterisation of W1201i Gag ..... 113

5.4 Molecular dynamics of N37T↑V protease ..... 114

5.5 Computational ligand docking ..... 117

5.6 Enzyme kinetics of N37T↑V protease ..... 120

5.7 Conclusions ..... 121

References ..... 122

# List of Figures

Figure 1: Life cycle of HIV-1 .....2

Figure 2: A ribbon diagram representing the homodimeric native structure of wild-type HIV-1 C-SA protease .....4

Figure 3: Schematic representing the catalytic mechanism of HIV-1 protease .....7

Figure 4: The two-dimensional molecular structures of the FDA-approved HIV-1 protease inhibitors and the clinical use release dates.....9

Figure 5: A ribbon diagram showing the positions of the amino acid substitutions and single amino acid insertion present on the N37T↑V variant. .....15

# Abbreviations

| | |
|---|---|
| ARV: | Antiretroviral |
| ART: | Antiretroviral Therapy |
| ATV: | Atazanavir |
| AZT: | Azidothymidine |
| CD: | Circular Dichroism |
| CS: | Cleavage Site |
| C-SA: | South African HIV-1 Subtype C |
| $C_\alpha$: | Alpha Carbon |
| DRV: | Darunavir |
| $E_a$: | Activation Energy |
| ELISA: | Enzyme-Linked Immunosorbent Assay |
| FDA: | Food and Drug Administration |
| HIV-1: | Human Immunodeficiency Virus type 1 |
| IFD: | Induced Fit Docking |
| IN: | Integrase |
| IPTG: | Isopropyl β-D-1-Thiogalactopyranoside |
| ITC: | Isothermal Titration Calorimetry |
| $K_a$: | Affinity constant |
| $k_{cat}$: | Catalytic constant |

| | |
|---|---|
| $k_{cat}/K_M$: | Catalytic efficiency |
| $K_d$: | Dissociation constant |
| LPV: | Lopinavir |
| MD: | Molecular Dynamics |
| MTCT: | Mother-to-Child Transmission |
| NNRTIs: | Non-Nucleoside Reverse Transcriptase Inhibitors |
| NRTIs: | Nucleoside Reverse Transcriptase Inhibitors |
| ns: | nanoseconds |
| NVP: | Nevirapine: |
| N37T↑V: | HIV-1 subtype C protease containing asparagine 37 mutated to threonine; the upwards arrow indicates an insertion of valine at position 37 |
| PCR: | Polymerase Chain Reaction |
| PDB: | Protein Data Bank |
| PIs: | Protease Inhibitors |
| PMTCT: | Prevention of Mother-to-Child Transmission |
| PR: | Protease |
| R: | Universal gas constant |
| RC: | Replication Capacity |
| Resistance control: | Multi-drug resistant *gag-protease* isolate |
| Rg: | Radius of Gyration |

| | |
|---|---|
| RMSD: | Root Mean Square Deviation |
| RMSF: | Root Mean Square Fluctuation |
| RT: | Reverse Transcriptase |
| RTV: | Ritonavir |
| SASA: | Solvent Accessible Surface Area |
| SDS-PAGE: | Sodium dodecyl sulfate-polyacrylamide gel electrophoresis |
| Wild-type control: | MJ4GP |
| WTGagN37T↑VPR: | Chimeric construct consisting of patient-derived *protease*, combined with wild-type *gag* |
| W1201i: | Patient-derived HIV-1 *gag-protease* |
| Δ*G*: | Change in Gibbs free energy |

# CHAPTER 1

## Introduction

### 1.1 Human immunodeficiency virus

Human immunodeficiency virus (HIV) is the pathogen responsible for acquired immunodeficiency syndrome (AIDS). HIV and AIDS have a devastating impact on sociocultural frameworks and economics in regions with high infection rates [1]. In South Africa, AIDS results in the death of ~110 thousand people a year. In 2016, 7.1 million people were living with AIDS in South Africa. Of those infected, ~56% were women [2].

Two types of HIV have been identified: HIV-1 and HIV-2. HIV-1 is predominant and may be classified into the groups M, N, O and P; where classification is based on sequence data. Similarly, group M can be further classified into the subtypes A, B, C, D, F, G, H, J, and K [3]. In South Africa, HIV-1 subtype C (HIV-1 C-SA) is responsible for 95% of infections [4].

HIV-1 is a retrovirus that contains two copies of positive-sense single-stranded RNA, which code for nine genes [5]. Illustrated in Figure 1, gene products such as protease (PR), integrase (IN), reverse transcriptase (RT) and various structural proteins each have a unique function that facilitates successful viral replication [6, 7].

HIV-1 infection occurs when the envelope glycoprotein 120 (or gp120) of the mature virion recognises and binds the $CD4^+$ receptor and subsequently the CCR5 (or CXCR4) co-receptor

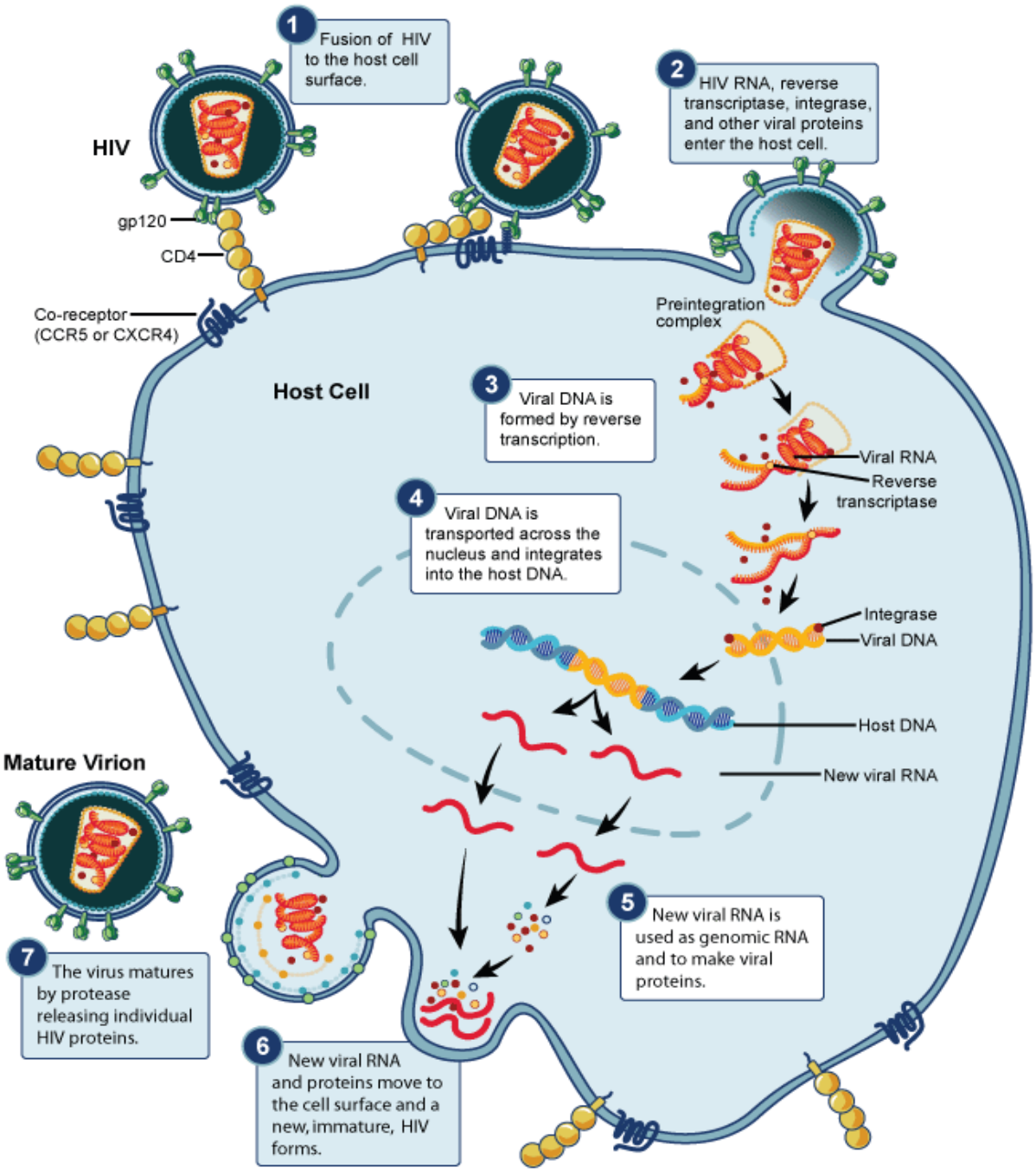



**Figure 1: Life cycle of HIV-1**

The chronology of HIV-1 replication and the functions of the viral enzymes. Step 7 indicates the involvement of HIV-1 PR. The figure was taken from NIAID (https://www.niaid.nih.gov/, 2011).

of the host T-cell [8]. This interaction results in structural modification of the envelope glycoprotein 41 (gp41) of the virion, which allows entry into the host cell through the phospholipid membrane [7, 9]. As the host cell membrane is breached, the viral capsid degrades and uncoats the two single-stranded RNA particles. Inside the host cell, the RNA is transcribed to DNA by RT. The IN enzyme incorporates the viral DNA into the host cell nucleus [9]. After incorporation, the viral genome is transcribed into mRNA which is translated by ribosomes into viral precursor proteins [7]. The resultant proteins assemble into a protoviral particle at the membrane interface, which is subsequently able to bud off from the cell. At this point, HIV PR cleaves the polyproteins to form a fully functional virion that can repeat the infection cycle [5, 7, 10].

## 1.2 HIV protease

HIV-1 PR is an attractive drug target due to its direct involvement in viral maturation, and, therefore, the viability of the viral particle [7]. Effectively targeting HIV-1 PR requires a comprehensive understating of both its structure and function. Unfortunately, conventional therapeutic agents were developed and tested predominantly on HIV-1 subtype B as it is the predominant subtype in the Americas and in Europe [11, 12]. Therefore, these compounds may not be as effective on HIV-1 C-SA and derived variants [13].

### 1.2.1 General structure of HIV-1 C-SA protease

HIV-1 C-SA PR is an obligate homodimer with a single active site [14]. Each subunit consists of one α-helix and two antiparallel β-sheets (Figure 2) [15, 16]. The monomer has a molecular weight of 11 kDa [17, 18]. There are eight amino acid polymorphisms distinguishing HIV-1

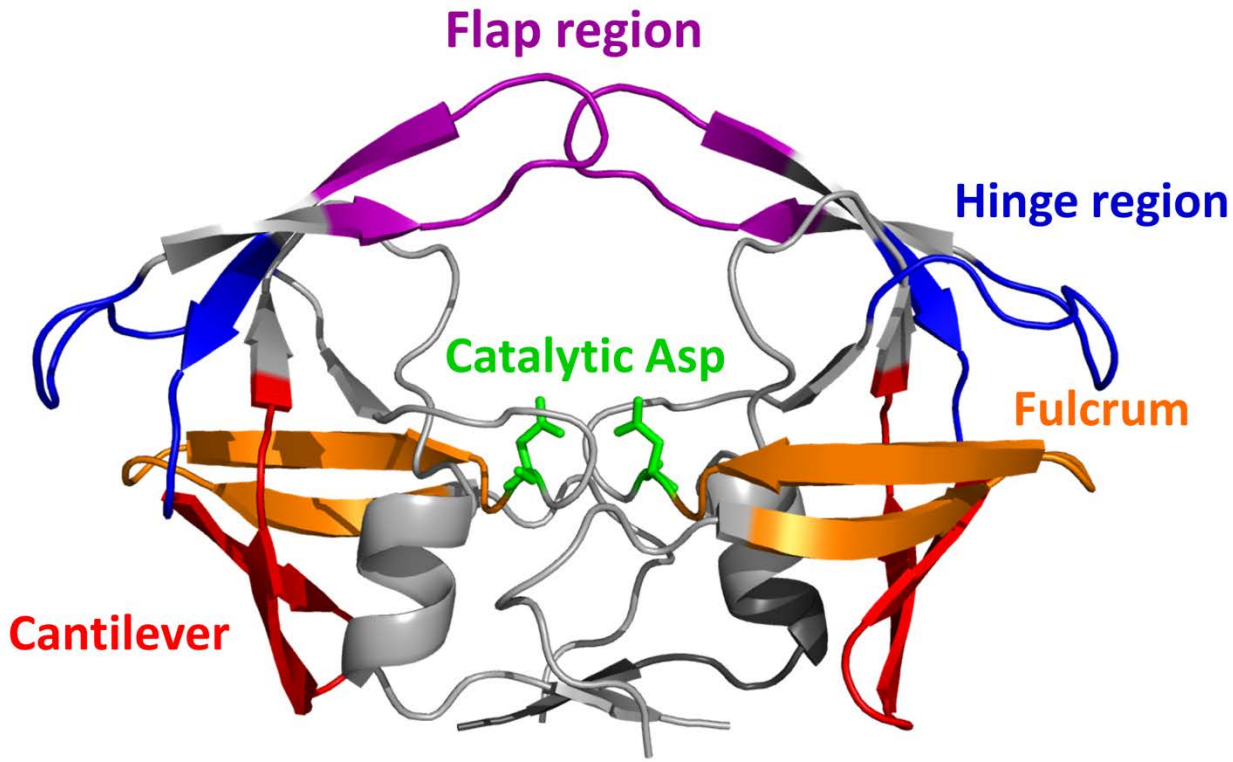



**Figure 2: A ribbon diagram representing the homodimeric native structure of wild-type HIV-1 C-SA protease**

The relative positions of the flap region, hinge region, active site Asp residues, cantilever and fulcrum are shown. This figure was generated with the molecular visualisation software PyMOL (The PyMOL Molecular Graphics System, Version 1,8 Schrödinger, LLC), using data from the Protein Data Bank (PDB ID: 3U71) [19, 20].

C-SA PR from HIV-1 B PR; namely: T12S, I15V, L19I, M36I, R41K, H69K, L89M and I93L [13, 21]. These polymorphisms may also play a role in aiding other known drug resistance mutations [13, 22].

Structurally, HIV-1 PR has five distinct regions. The hinge region is composed of the amino acid residues 35-42 and 57-61 [23, 24]. Residues 46-54 form the flap region and are integral to the specificity as well as the activity of HIV-1 PR [20]. The fulcrum region is composed of residues 10-23, and the cantilever region is composed of residues 62-75 [25]. The active site contains two catalytic Asp residues (residue 25) [26].

The hinge region remains flexible and facilitates substrate entry into the active site by extending the flap region upwards, and outwards. The PR flaps can move up to 15 Å from each other (outwards) and up to 20 Å from the active site Asp residues (upwards) [27, 28]. After recognition, the flap region encloses the bound substrate and allows chemical interactions to form [29–31]. According to Naicker *et al.*, (2014), the flexibility of the hinge region may facilitate an exaggerated movement of the flap region. Additionally, the flap region coordinates a single water molecule that is required for enzymatic catalysis [32]. Thus, the hinge region contributes to the activity and biological function of HIV-1 PR, [31, 33].

## 1.3 Function of HIV-1 C-SA protease

HIV-1 PR is responsible for the catalytic action that allows the cleavage and consequently the activation of the Gag (p55) and Gag-Pol (p160) polyproteins [6, 10, 34]. The Gag polyprotein contains the matrix (p17), capsid (p24), nucleocapsid (p7) and p6 proteins as well as two spacer peptides; namely, p1 and p2 [5]. HIV PR cleaves the Gag polyprotein at five cleavage

sites (CSs). The cleavage of Gag is an ordered and conserved process, without which the virion cannot mature into its infective form [35].

HIV-1 PR functions as a homodimeric catalyst which can cleave peptide bonds with high efficiency. The PR active site consists of a highly conserved catalytic triad consisting of residues Asp-25, Thr-26 and Gly-27 [36, 37]. The active site residues are contributed by both subunits of the dimeric PR molecule. Hydrogen bonds allow a single water molecule to bridge the two aspartate carboxyl groups as they are in proximity (< 2.4 Å) to one another [34].

The catalytic mechanism of HIV-1 PR has been extensively studied, and several similar mechanisms have been proposed. These studies have converged into a consensus mechanism that is widely accepted. HIV-1 PR has a general acid-base reaction mechanism [38, 39]. When the natural substrate binds the active site of HIV-1 PR, a covalent intermediate is formed. Only then are the products released sequentially [16, 39].

Figure 3 highlights the catalytic mechanism of the HIV-1 PR. Substrate catalysis takes place as follows: firstly, the bound water molecule present in the active site is activated by Asp-25 of a single subunit. The carboxylate group acts as a general Lewis base and the reaction results in the liberation of an OH- ion. The ion acts as a nucleophile to attack the scissile peptide bond on the substrate and results in the formation of a tetrahedral intermediate [32, 40]. The carboxyl group on the Asp-25' residue is protonated via general acid catalysis, which results in the stabilisation of the resultant oxyanion. In the final step, decomposition of the intermediate into products occurs through general acid catalysis, as well as general base catalysis, through the action of both Asp-25 and Asp-25', respectively [41]. After catalysis, the products are released from the active site.

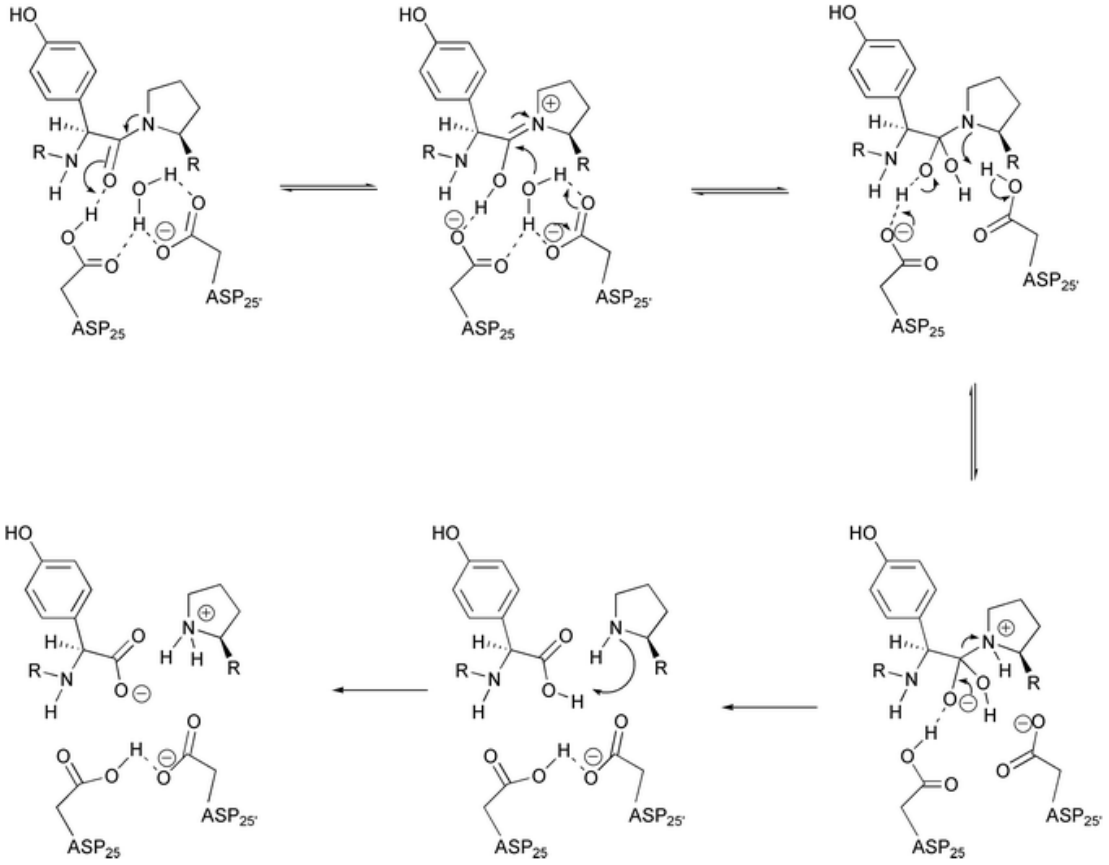



**Figure 3: Schematic representing the catalytic mechanism of HIV-1 protease**

The catalytic mechanism of HIV-1 PR is based on kinetic and structural data. The figure was taken from Brik and Wong (2002).

## 1.4 Protease inhibitors

The discovery of protease inhibitors (PIs) in the early 1990s allowed the possibility of dual-class triple combination ARV therapy, which later became known as highly active antiretroviral therapy (HAART). For more than 20 years, HIV-1 PIs have been the gold standard in therapeutic agents that target HIV PR [21, 22]. Nearly 30 years later, PIs remain part of standard health care for HIV-infected patients worldwide. There are currently nine approved (not including the prodrug fosamprenavir) PIs that explicitly target HIV PR (Figure 4) [42, 43]. Eight of the nine drugs are known as peptidomimetic inhibitors because they mimic the transition state of the natural PR substrate; namely, the viral Gag-Pol and Gag polyproteins [44, 45]. These PIs contain chemical analogues of proline and phenylalanine; amino acids which are found at positions 167 and 168 on the Gag-Pol polyprotein [46]. Thus, the PIs are competitive inhibitors that bind the active site of the HIV-1 PR molecule. The drug tipranavir (TPV) is not a peptidomimetic and it binds uniquely to HIV-1 PR; consequently, TPV presents a unique resistance profile [47].

The development of the first generation PR inhibitors; namely, saquinavir (SQV), ritonavir (RTV), indinavir (IDV), nelfinavir (NFV) and amprenavir (APV), provided essential therapeutic benefits to those infected with HIV-1 [42]. Unfortunately, resistance to these drugs developed rapidly, which led to the development of the second generation PIs: lopinavir (LPV) and atazanavir (ATV). In time, tipranavir (TPV) and darunavir (DRV), the third generation PIs were also developed [48].

The co-administration of RTV with a second or third generation PI is considered a significant step forward in HIV treatment. RTV is a potent inhibitor of the xenobiotic detoxification enzyme cytochrome P450 isoenzyme CYP3A4, which is the principal enzyme responsible for

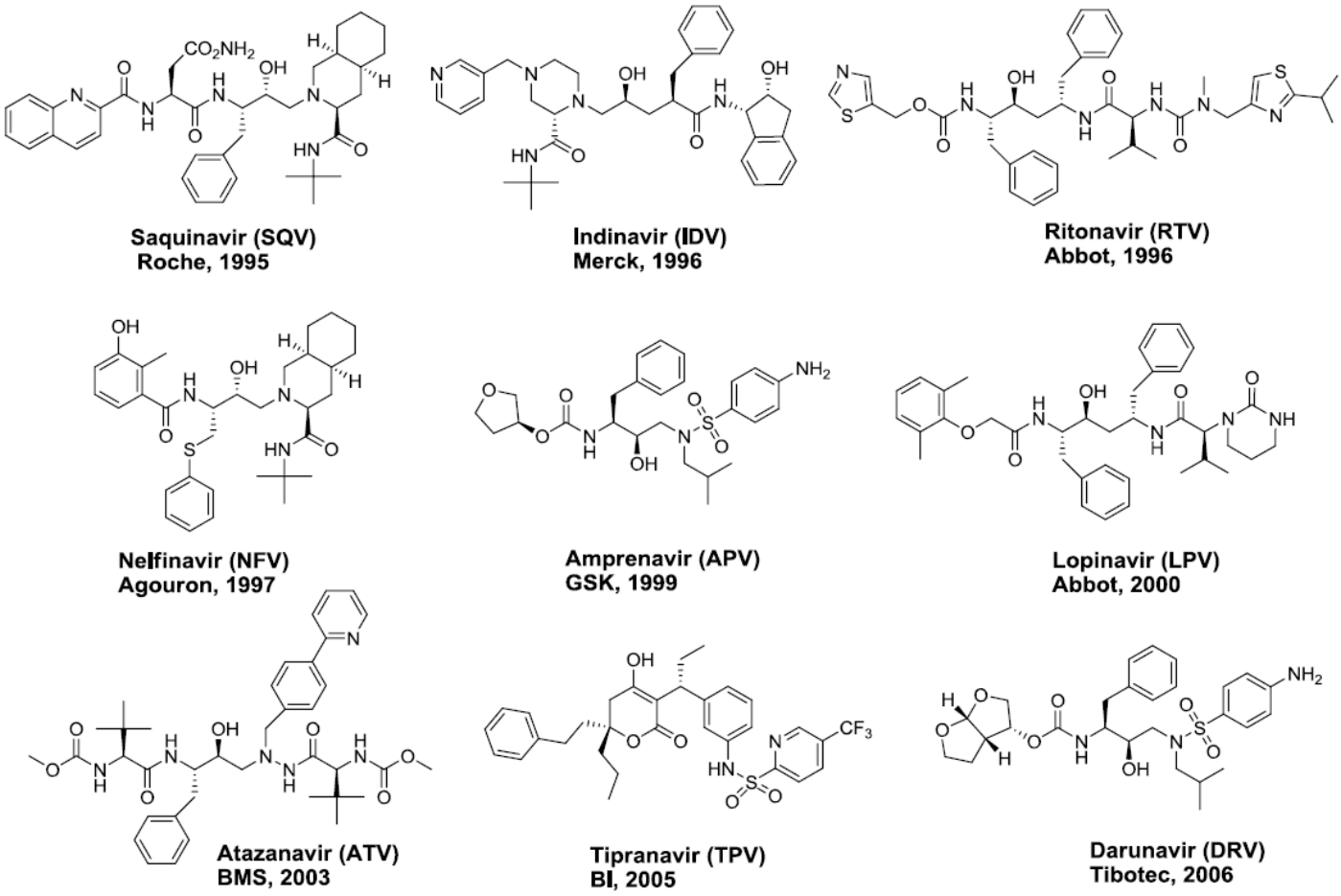



**Figure 4: The two-dimensional molecular structures of the FDA-approved HIV-1 protease inhibitors and the clinical use release dates**

The names of the pharmaceutical companies responsible for the development of the various PIs are indicated. The figure was taken from Ali *et al*., (2010).

PI metabolism [49]. The administration of RTV has been shown to compliment the effect of other PIs, a mechanism which is referred to as boosting [50]. Boosted PIs can be administered in lower concentrations, which dramatically reduces side-effects [51]. Furthermore, patients receiving boosted PIs are much less prone to develop drug resistance mutations within PR [52].

Structure-assisted drug design is the leading method used in modern pharmacology to create therapeutic agents targeting HIV-1 PR [53, 54]. The structure-based approach is multidisciplinary as it incorporates data from the biological, mathematical and physical sciences [32]. Understanding the thermodynamic principles governing drug binding is arguably one of the most important aspects of rational drug design.

The development of lead compounds (i.e. developmental candidates) relies on the optimisation of their respective binding affinities for the chosen target [43]. The binding affinity of a given lead compound is enhanced by increasing the binding enthalpy and solvation entropy, while decreasing binding entropy [55]. For instance, the binding of the PIs to HIV-1 PR may be either enthalpically or entropically driven. A favourable increase in bulk solvent entropy is predominantly responsible for entropically driven drug binding processes [56]. Usually, conformational entropy is lost in both the target and drug due to reduced flexibility upon binding [57]. Entropy loss, however, is compensated for by an increase in solvation entropy resulting from the exclusion of ordered water molecules within the active site [58]. An increase in bulk solvent entropy, therefore, drives some biological processes. The favourable entropic change is due to the hydrophobic nature of the compound [11]. Enthalpically driven drug interactions result from the optimisation of specific molecular contacts between substrate and enzyme. Interactions include ionic interactions, polar and dipolar interactions, hydrogen bonding and specific van der Waals interactions [11].

The majority of first-generation PIs were predominantly entropically driven. The binding of a specific drug be made more entropically favourable to its respective target, by increasing its overall hydrophobicity. Increasing the hydrophobicity of a drug decreases its overall solubility. Therefore, higher concentrations of the drug are needed to ensure adequate bioavailability. However, higher concentrations of the drug could result in unintended interactions with non-targeted enzymes, leading to adverse side-effects [59].

Rational drug design endeavours to create drugs that have better potency, higher selectivity, and better pharmacokinetics. Drugs posessing these atributes are often found to be enthalpically, opposed to entropically, driven [60]. However, it often takes years for enthalpically driven drug candidates to appear on the market because it is notoriously difficult to optimise binding enthalpy [61]. The second and third generation PIs were found to be enthalpically driven. In fact, all protease inhibitors with picomolar binding affinity have favourable binding enthalpies [61].

The PR enzyme has many known sites that are attractive allosteric drug targets and recently there has been a drive to develop non-peptidomimetic PIs that act on allosteric sites rather than targeting the active site of the enzyme [28, 62, 63]. The hinge and flap regions are particularly attractive targets because the opening of the PR active site is dependent on them [28]. Unfortunately, no allosteric drugs have yet been approved as a viable treatment for HIV-1 infection.

## 1.5 HIV-1 C-SA protease drug susceptibility

### 1.5.1 HIV-1 Protease variants

The leading cause of ARV failure is the development of drug resistance. HIV-1 develops drug resistance swiftly due to the action of the highly error-prone reverse transcriptase. Moreover, the HI virus displays broad genetic diversity [64, 65]. However, while some regions within the HIV-1 *protease* gene are highly mutable, other regions remain highly conserved [54, 66].

The development of PI resistance is a complicated process involving the accumulation of primary and secondary drug resistance mutations in the PR enzyme [67]. These primary resistance mutations develop in, or within proximity, to the substrate binding pocket of the PR and function by decreasing the binding affinity for PIs [68]. PR mutations that confer drug resistance occur due to specific drug pressures; however, these usually have a negative impact on the ability of the enzyme to process its natural substrate [69]. As a result, there is a reduction in the replication capacity (RC) of the virus.

Compensatory mutations, also known as secondary mutations, will develop distal to the PR active site if drug pressures persist. Secondary mutations can allow the modified enzyme to bind the substrate more efficiently, thereby restoring viral fitness [47, 70, 71]. These mutations are often able to confer reduced drug susceptibility even in the absence of primary resistance mutations within PR [72]. Furthermore, if drug pressure is ceased, strains will often revert to the wild-type genotype, in an attempt to restore viral fitness [73].

Typically, each subunit of HIV-1 PR consists of 99 amino acids. However, recently discovered PR variants had been shown to consist of 100 or even 101 amino acids per subunit [74, 75]. These unique secondary amino acid polymorphisms may determine how the HIV-1 PR

interacts with drugs specifically designed to counteract its biological function [47]. Kožíšek and colleagues proposed that amino acid insertion mutations in HIV-1 PR may contribute to drug resistance [75]. Hence, it is imperative that further work be done to characterise novel mutations in PR and to test the efficacy of current therapeutic agents on these variants [42].

At a biochemical level, the combined effects of primary and secondary mutations have important consequences on the viability of PIs. The optimisation of new drug leads against a highly variable target such as HIV-1 PR requires that each polymorphism is studied in depth. Thus, the study of clinically significant PR variants is vital for an in-depth molecular characterisation of the mechanism by which drug resistance and viral fitness occur [11].

### 1.5.2 HIV-1 protease dynamics

The flap region is critical to the activity of the PR as it allows substrate entry into the active site [62]. Studies show that increased flap dynamics result in fewer chemical interactions between the enzyme and PIs [12]. Therefore, the recognition and binding of the PIs are less thermodynamically favourable, which ultimately affects the drug susceptibility profile of the PR [76]. Flap region dynamics is partially controlled by the dynamics of the hinge region [77]. In fact, the effect of the hinge region on flap flexibility has been extensively studied experimentally (HDX-MS, NMR and EPR) as well as computationally (MD simulations) [29, 30, 78–80]. The hinge region displays a high degree of mutability. Hinge region mutations can alter drug susceptibility, enzyme activity, conformational flexibility and facilitate immune system concealment [31, 76, 81].

Hinge region mutations are considered secondary mutations as they occur distal to the active site and do not directly alter the molecular interactions between the PR and PIs. Instead,

secondary mutations confer their effect indirectly by altering flap flexibility [81–83]. Experiments on the apo form of HIV-1 PR have indicated that the movements of the flap region can be observed on the nanosecond timescale [84]. MD simulations have proven to be a particularly useful tool for determining the kinetics of specific regions within the PR [27].

### 1.5.3 HIV-1 Gag variants

The clinical management of PI failure is still largely based on the *protease* gene sequence [85]. Research groups are ever trying to determine the next group of PI resistance mutations. However, a body of evidence is accumulating that both drug susceptibility and PI resistance involves the viral Gag polyprotein [86, 87]. Unfortunately, the precise mechanism governing Gag's role in PI susceptibility has not been completely characterised [88]. Nijhuis and colleagues (2007) have reported that CS mutations could independently affect drug susceptibility in the absence of known PR drug resistance mutations. In fact, it was reported that non-CS mutations could reduce PI susceptibility by up to ten-fold [89]. Since then, multiple studies have eluded that CS, as well as non-CS mutations, play a key role in resistance to PIs [90–94]. Therefore, polymorphisms within Gag, both CS and non-CS mutations, should be considered when evaluating the PI susceptibility of HIV-1 variants.

## 1.6 W1201i Gag-protease

Recently, HIV-1 C-SA sequence data was obtained from a South African PI-naïve infant (NICD, South Africa). The patient displayed a high viral load and very low CD4 count despite prevention of mother-to-child transmission (PMTCT) treatment [95]. A specific Gag-PR isolate, named W1201i, was identified [96]. The PR sequence data confirmed a PR hinge

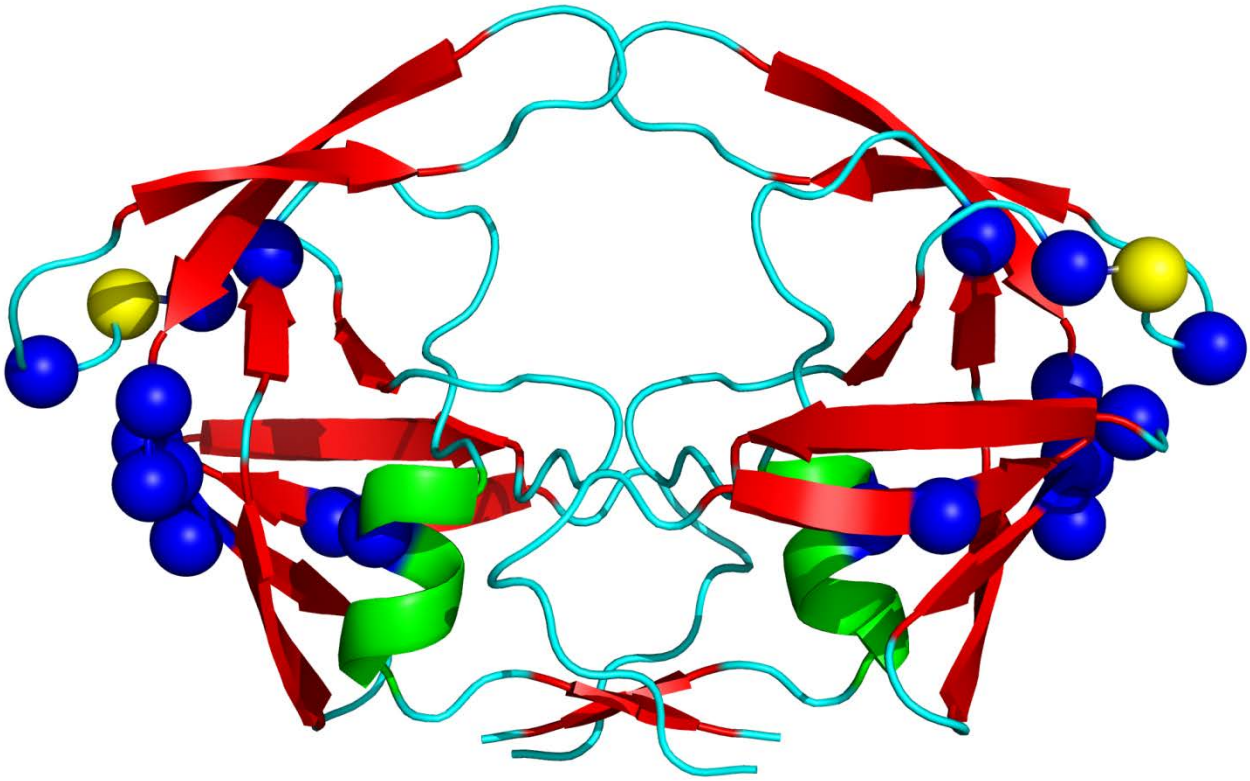


**Figure 5: A ribbon diagram showing the positions of the amino acid substitutions and single amino acid insertion present on the N37T↑V variant.**

Beta sheets and alpha helices are rendered red and green, respectively. Random coils are rendered light blue. The relative position of the Thr mutation and Val insertion is shown in yellow. The blue spheres represent the van der Waals radii of the various polymorphisms that are present in the N37T↑V variant; namely, I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L. This figure was generated with the molecular visualisation software PyMOL (The PyMOL Molecular Graphics System, Version 1,8 Schrödinger, LLC), using data from the Protein Data Bank (PDB ID: 3U71) [20].

region mutation and insertion (N37T↑V) as well as several amino acid polymorphisms; namely, I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L (Figure 5). Consequently, each monomeric subunit consists of 100 amino acids. This variant was dubbed N37T↑V as the substitution and insertion mutations were both found at residue 37. The Gag sequence of W1201i had not been characterised before this study.

## 1.7 Aim and objectives

### 1.7.1 Aim

The aim of this study was to determine how hinge region mutations and insertions, such as those found in the N37T↑V variant, affect the overall structure, function, catalytic ability, and drug susceptibility of HIV-1 C-SA PR.

### 1.7.2 Objectives

1. Overexpress and purify both the wildtype and N37T↑V variant PR using a novel purification method.

2. Probe the quaternary structural parameters of the of the N37T↑V variant using high-performance liquid chromatography.

3. Obtain the kinetic parameters of the molecular interaction between the N37T↑V variant and a fluorogenic substrate.

4. Evaluate the drug susceptibility of both the N37T↑V and wild-type proteases through phenotypic viral assays.

5. Using computational modelling and induced fit docking to evaluate the molecular dynamics and drug binding energetics of N37T↑V protease

# CHAPTER 2

## Overexpression, purification and functional characterisation of wild-type HIV-1 subtype C protease and two variants using a thioredoxin and his-tag protein fusion system

Jake Zondagh, Alison Williams, Ikechukwu Achilonu, Heini W. Dirr and Yasien Sayed.

*Protein J* (Manuscript accepted)

In this publication, a novel method for HIV-1 protease purification is described. The paper deals with the expression of wild-type HIV-1 subtype C and two variant proteases using a thioredoxin fusion protein system. The thioredoxin fusion protein system has not been used on HIV-1 proteases before.

Author contributions: Jake Zondagh performed all the experiment work on the N37T↑V protease, and Alison Williams performed all the experimental work on L38↑N↑L. The experimental work on the wild-type protease and data analysis was shared equally by both Jake Zondagh and Alison Williams. Yasien Sayed, Ikechukwu Achilonu and Heini Dirr assisted in manuscript revision. Yasien Sayed and Ikechukwu Achilonu assisted in method design and data interpretation. Yasien Sayed supervised the project.

The Protein Journal

# Overexpression, purification and functional characterisation of wild-type HIV-1 subtype C protease and two variants using a thioredoxin and his-tag protein fusion system

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | JOPC-D-18-00016 |
| **Full Title:** | Overexpression, purification and functional characterisation of wild-type HIV-1 subtype C protease and two variants using a thioredoxin and his-tag protein fusion system |
| **Article Type:** | Original Research |
| **Keywords:** | HIV-1; protease; Escherichia coli; metal ion affinity chromatography; fusion protein; hexahistidine tag |
| **Corresponding Author:** | Yasien Sayed, PhD<br>University of the Witwatersrand<br>Johannesburg, Gauteng SOUTH AFRICA |
| **Corresponding Author Secondary Information:** | |
| **Corresponding Author's Institution:** | University of the Witwatersrand |
| **Corresponding Author's Secondary Institution:** | |
| **First Author:** | Yasien Sayed, PhD |
| **First Author Secondary Information:** | |
| **Order of Authors:** | Yasien Sayed, PhD |
| | Jake Zondagh, MSc |
| | Alison Williams, MSc |
| | Ikechukwu Achilonu, PhD |
| | Heini Dirr, PhD |
| **Order of Authors Secondary Information:** | |
| **Funding Information:** | South African Medical Research Council (ZA) — Professor Yasien Sayed |
| | National Research Foundation (68898) — Professor Heini Dirr |
| **Abstract:** | In recent years, various strategies have been used to overexpress and purify HIV-1 protease because it is an essential drug target in anti-retroviral therapy. Obtaining sufficient quantities of the enzyme, however, remains challenging. Overexpression of large quantities is prevented due to the enzyme's autolytic nature and its inherent cytotoxicity in Escherichia coli cells. Here, we describe a novel HIV-1 protease purification method using a thioredoxin-hexahistidine fusion system for the wild-type and two variant proteases. The fusion proteases were overexpressed in Escherichia coli and recovered by immobilised metal ion affinity chromatography. The proteases were cleaved from the fusion constructs using thrombin. When compared to the standard overexpression and purification protocol in use in our laboratory, the expression of the fusion-derived wild-type protease was increased from 0.83 to 2.5 mg/L of culture medium. The expression levels of the two variant proteases ranged from 1.5 to 2 mg/L of culture medium. The fusion wild-type and variant proteases were inactive before the cleavage of the thioredoxin-hexahistidine fusion tag as no enzymatic activity was observed. The proteases were, however, active after cleavage of the tag. The novel thioredoxin-hexahistidine fusion system, therefore, enables the successful overexpression and purification of catalytically active HIV-1 proteases. |
| **Suggested Reviewers:** | Abidemi Kappo, PhD<br>University of Zululand<br>KappoA@unizulu.ac.za |



Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

| | |
|---|---|
| | Specializes in structural biology and protein structure-function. |
| | Khajamohiddin Syed<br>University of Zululand<br>syedk@unizulu.ac.za<br>Expert in molecular biology, gene expression and protein structure and function |
| | Thulile Ndlovu, PhD<br>Stellenbosch University<br>ndlovu@sun.ac.za<br>Biochemist with knowledge of proteins |
| | Heinrich Hoppe, PhD<br>Rhodes University<br>h.hoppe@ru.ac.za<br>Expertise in parasitic proteins and protein drug interactions, protein structure function relationships. |
| | Gillian Hunt<br>National Institute for Communicable Diseases<br>gillianh@nicd.ac.za<br>Expertise in HIV proteins |



Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

**Overexpression, purification and functional characterisation of wild-type HIV-1 subtype C protease and two variants using a thioredoxin and his-tag protein fusion system**

Jake Zondagh[1], Alison Williams[1], Ikechukwu Achilonu, Heini W. Dirr, Yasien Sayed*

Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of Witwatersrand, Johannesburg, 2050, South Africa

[1]Both authors contributed equally to this work

*Corresponding author: Yasien Sayed; e-mail address: yasien.sayed@wits.ac.za, telephone number: +27 11 717 6350.

ORCID IDs of authors:

1) Jake Zondagh: 0000-0002-7827-1244
2) Alison Williams: 0000-0003-1772-6951
3) Ikechukwu Achilonu: 0000-0002-8658-956X
4) Heini W. Dirr: 0000-0003-1455-7769
5) Yasien Sayed: 0000-0002-1781-2115

**Acknowledgements**

The research reported in this publication was supported by the South African Medical Research Council (SAMRC) under a Self-Initiated Research Grant to Yasien Sayed. The views and opinions expressed are those of the authors and do not necessarily represent the official views of the SAMRC. This work was supported by the University of the Witwatersrand, South African National Research Foundation Grant 68898 (HWD), and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation Grant 64788 (HWD). The authors would like to thank the National Research Foundation for funding and Professor Lynn Morris (Head: HIV Research, National Institute for Communicable Diseases, South Africa) for supplying the HIV-1 subtype C protease sequence information.

## Abstract

In recent years, various strategies have been used to overexpress and purify HIV-1 protease because it is an essential drug target in anti-retroviral therapy. Obtaining sufficient quantities of the enzyme, however, remains challenging. Overexpression of large quantities is prevented due to the enzyme's autolytic nature and its inherent cytotoxicity in *Escherichia coli* cells. Here, we describe a novel HIV-1 protease purification method using a thioredoxin-hexahistidine fusion system for the wild-type and two variant proteases. The fusion proteases were overexpressed in *Escherichia coli* and recovered by immobilised metal ion affinity chromatography. The proteases were cleaved from the fusion constructs using thrombin. When compared to the standard overexpression and purification protocol in use in our laboratory, the expression of the fusion-derived wild-type protease was increased from 0.8 to 2.5 mg/L of culture medium. The final concentration of the two variant proteases ranged from 1.5 to 2 mg/L of culture medium and the total wild-type protease yield from this fusion system exceeds our control purification method by 250%. The fusion wild-type and variant proteases were inactive before the cleavage of the thioredoxin-hexahistidine fusion tag as no enzymatic activity was observed. The proteases were, however, active after cleavage of the tag. The novel thioredoxin-hexahistidine fusion system, therefore, enables the successful overexpression and purification of catalytically active HIV-1 proteases.

## Keywords

HIV-1; protease; *Escherichia coli;* metal ion affinity chromatography; fusion protein; hexahistidine tag

## Abbreviations

| | |
|---|---|
| HIV-1: | Human Immunodeficiency Virus type 1 |
| PR: | Protease |
| TRX: | Thioredoxin |
| 6His: | Hexahistidine |
| TCS: | Thrombin cleavage site |
| N37T↑V: | HIV-1 subtype C protease containing asparagine 37 mutated to threonine; the upward arrow indicates an insertion of valine at position 37 |
| L38↑N↑L: | HIV-1 subtype C protease containing leucine at position 38 followed by a double insertion of asparagine and leucine |
| IMAC: | Immobilised Metal Ion Affinity Chromatography |

# 1. Introduction

Human Immunodeficiency Virus (HIV) is the etiological agent of Acquired Immunodeficiency Syndrome (AIDS). Globally, 35 million people are HIV positive, and 1.9 million people are infected each year [1]. HIV is problematic in sub-Saharan Africa because it is estimated that one in twenty adults is living with the virus and this accounts for 69% of the total global statistic [1].

HIV-1 was first isolated in 1983 [2] and, since then, it has been studied extensively. HIV-1 is the most common form of HIV and is further divided into groups and subtypes [3, 4]. Subtype B, the most studied of the subtypes, is found in America, Western Europe and Australia [5]. Subtype C, of interest to this study, is found predominantly in southern Africa, the horn of Africa and India [6, 7].

The homodimeric aspartyl protease is one of three enzymes produced by HIV-1 and is essential for the production of mature virions [8, 9], and is an important drug target. HIV protease is expressed as a Gag-Pol precursor from which it can free itself by autocatalysis after dimerisation [10]. The catalytically mature enzyme then processes the Gag and Pol polyproteins to produce viral structural proteins and reverse transcriptase and integrase enzymes [11, 12].

In-depth biochemical studies require sufficient amounts of protein. HIV-1 protease has previously been synthesised chemically [13] and expressed in heterologous systems using recombinant DNA technology [14]. Recombinant DNA technology permits the successful production of clinically significant proteins in large quantities and is, therefore, of major importance [15]. However, many expression systems do not yield adequate amounts of product necessary for specific downstream analyses such as isothermal titration calorimetry.

It is challenging to obtain HIV-1 protease in large quantities due to its cytotoxic effects when overexpressed. Bacterial and mammalian cells are primarily affected by the cytotoxic nature of HIV-1 protease [16]. In the past, various strategies have been investigated to acquire greater yields. Purification strategies include production by autocatalytic processing of a larger precursor (Gag-Pol region), recovery by refolding of *E. coli* inclusion bodies, purification of a His-tagged recombinant protein, and the use of fusion proteins such as β-lactamase, glutathione transferase and maltose binding protein [14, 17–20].

This study aimed to improve the expression of the wild-type HIV-1 subtype C protease by using a thioredoxin-fusion protein system. Additionally, this method was tested on two clinically relevant variant proteases under investigation in our laboratory. The amino acid insertions and background mutations in these variant proteases were found in protease inhibitor-naïve (PI-naïve) patients and are not prevalent in patients receiving PI therapy or failing PI therapy.

We, therefore, aimed to overexpress and purify four separate proteases; namely, a Gag-Pol derived wild-type protease as a non-fusion control (referred to as the "control wild-type"), a thioredoxin-fusion derived wild-type protease (referred to as the "fusion wild-type"), and two thioredoxin-fusion derived variants (i.e. N37T↑V and L38↑N↑L) (Fig. 1A, B). The N37T↑V protease indicates that asparagine at position 37 was mutated to threonine and the upward arrow indicates a valine amino acid was inserted. The L38↑N↑L protease represents a double insertion (asparagine and leucine) after position 38.

This system has not been used on HIV-1 protease before but has been used successfully with other human proteins [22]. In this paper, we demonstrate the successful overexpression and purification of catalytically active wild-type subtype C protease and two variants using a thioredoxin-hexahistidine fusion system.

## 2. Materials and Methods

### 2.1 Construction of the fusion plasmids

The genes coding for the fusion wild-type, N37T↑V and L38↑N↑L proteases were synthesised by GenScript (Hong Kong) and cloned into three separate pET-11a expression vectors. The sequences for the variant proteases were obtained from Professor Lynn Morris (Head of the AIDS Research Unit) at the National Institute for Communicable Diseases (NICD, South Africa). Wild-type subtype C protease was generated previously in our laboratory and contained the following polymorphisms: T12S, I15V, L19I, M36I, R41K, H69K, L89M, and I93L [21]. Fusion protein sequences were confirmed by Sanger DNA sequencing (Inqaba Biotech, South Africa). The protease sequences were aligned using the Clustal Omega tool (EMBL-EBI) [22]. Homology models were generated with the molecular visualisation software programme PyMOL, using data from the Protein Data Bank (PDB ID: 3U71) [23].

### 2.2 Expression and purification

The control wild-type protease was purified using a standard protease purification system routinely used in our laboratory [23]. Briefly, *E. coli* BL21 (DE3) pLysS cells were transformed with a plasmid encoding the control wild-type protease gene. The cells were induced at 37 °C for four hours with 1 mM isopropyl β-D-thiogalactoside (IPTG), and the protease was recovered from inclusion bodies after cell disruption. Recovery buffer contained 8 M urea, 10 mM Tris-HCl and 2 mM DTT (pH 9). The sample was incubated at 20 °C for one hour in the urea buffer before recovery by centrifugation. The sample was dialysed (and refolded) against 10 mM sodium acetate (pH 5) at 4 °C and purified using CM-Sepharose ion exchange chromatography with a 0-1 M NaCl gradient. The control protease was included to measure the success of the new purification strategy.

The three thioredoxin fusion proteases; namely, fusion wild-type, N37T↑V and L38↑N↑L, were expressed by separately transforming *E. coli* BL21 (DE3) pLysS cells with a pET-11a expression vector encoding each of the constructs. The fusion wild-type and N37T↑V fusion proteases were expressed in six litres of LB media at 37 °C for four hours using 1 mM IPTG, and expression was induced when the culture media reached an $OD_{600}$ of 0.5. Cells were harvested by centrifugation at 5000×*g*, resuspended in lysis buffer (20 mM Tris-HCl, 1 mM lysozyme, 150 mM NaCl, pH 7.5) and sonicated at 10 V for 10 cycles of 30 s.

The samples were separated into soluble and insoluble fractions by centrifugation at 24 000×*g*. The insoluble pellets were washed twice with 20 mM Tris-HCl buffer, pH 7.4, containing 1% (v/v) Triton X-100. The proteins in the insoluble fraction were unfolded using 8 M urea, and the cell debris was collected by centrifugation at 24 000×*g*. The urea concentration was decreased to 4 M by overnight dialysis against 20 mM Tris-HCl buffer (pH 7.4) at 20 °C. Fusion wild-type and N37T↑V proteases were bound to a 5 ml IMAC column charged with $Ni^{2+}$ and eluted with an imidazole gradient (0-500 mM).

Fractions containing the fusion wild-type and N37T↑V proteases were dialysed against refolding buffer (20 mM Tris-HCl, 10% (v/v) glycerol, 150 mM NaCl, pH 7.4) at 4 °C. The thioredoxin-hexahistidine tag was cleaved from the protease using thrombin (1 U/ml of sample, overnight at 20 °C). Untagged protease was collected and thrombin removed by passing the sample over a 5 ml benzamidine column (to which thrombin binds) connected in series to a 5 ml IMAC column (to which the cleaved tag and any uncleaved proteins bind). The flow-through, containing the untagged protease, was incubated in 25 mM formic acid for one hour and dialysed against 10 mM formic acid at 4 °C for 4 hours to precipitate any unwanted protein present. The pure protease sample was dialysed against 10 mM sodium acetate buffer (pH 5.0) at 4 °C overnight and stored at -80 °C until needed.

The L38↑N↑L fusion protease was overexpressed in six litres of LB media at 20 °C overnight using 1 mM IPTG. The cells were resuspended in 40 ml of 20 mM Tris-HCl buffer (pH 7.4). The cells were sonicated as described earlier and the soluble fraction was isolated by centrifugation at 24 000×*g*. The protease was purified from the soluble fraction using a 5 ml IMAC column and eluted using an imidazole gradient (0-500 mM).

Following thrombin cleavage (as described earlier), and 10 mM formic acid precipitation, the sample was dialysed against 10 mM sodium acetate buffer (pH 5) at 4 °C. The sample was passed through a CM-Sepharose column to remove any unwanted protein. The protease was eluted using a 0-1 M NaCl gradient and dialysed against 10 mM sodium acetate buffer (pH 5) at 4 °C overnight to remove any residual NaCl. The absence of the salt decreases autolysis (the ability to undergo autoproteolysis in solution). The final concentration of pure protease was determined using the absorbance value at 280 nm and the extinction coefficient of the protein according to the Beer-Lambert equation. The extinction coefficients were calculated using the ProtParam online tool [24]. The extinction coefficients used were: 25 480 $M^{1}.cm^{-1}$ for wild-type, 24 980 $M^{-1}.cm^{-1}$ for N37T↑V and 25 230 $M^{-1}.cm^{-1}$ for L38↑N↑L.

### 2.3 Structural characterisation

HIV-1 protease is functional in its homodimeric form and, therefore, it was essential to determine the quaternary structure of all the proteases. Verificatin of size was determined by size-exclusion high-performance liquid chromatography (SE-HPLC) using a TSKgel SuperSW2000 column equilibrated with 10 mM sodium acetate buffer (pH 5) containing 150 mM NaCl.

### 2.4 Functional characterisation

HIV-1 protease is prone to autolysis and, for this reason, it is important to quantify the concentration of active enzyme in a purified sample. The percentage active protease was determined by performing isothermal titration calorimetry (ITC) active site titration experiments using a VP-ITC Microcalorimeter (MicroCal Inc., Malvern Instruments, Malvern, Worcestershire, UK). Briefly, 200 µM acetyl pepstatin, a competitive inhibitor of HIV-1 protease, was titrated (6 µl injections) into a solution of 10 to 13 µM protease at 293.15 K. The percentage active protease in each sample was determined from the binding stoichiometry (N-value) after subtracting the heats of dilution and correcting baseline errors from the calorimetric data using the Origin 7.0 software package (OriginLab Corporation, Northampton, MA, USA). The N-value is used as a correction factor for the concentration of active protease in a purified sample. The ITC data were fitted using an algorithm for one set of binding sites because acetyl pepstatin binds to protease in a 1:1 ratio. An N-value of 1 is theoretically representative of 100% active enzyme in sample preparations, i.e. all the protease molecules are in their active form, and no self-cleavage has occurred.

An enzyme assay was conducted during thrombin cleavage to determine whether the protease was catalytically active. The increase in fluorescence intensity attributed to the cleavage of the fluorogenic substrate (Abz-Arg-Val-Nle-Phe($NO_2$)-Glu-Ala-Nle-$NH_2$) was measured. The protein and substrate were dissolved in buffer consisting of 20 mM Tris-HCl, 10% (v/v) glycerol and 150 mM NaCl (pH 7.4) The assay was performed with 50 nM of protein and a constant substrate concentration of 50 µM under steady state conditions (20 °C). All samples were measured for 1 minute using an excitation bandwidth of 2.5 nm and an emission bandwidth of 5 nm and all measurements were performed in triplicate. The complete cleavage of 1 nmol substrate was measured which served as a standard to convert the measured emission intensity to activity. The peptide substrate was excited at 337 nm and the fluorescence emission monitored at 425 nm. The assay was performed on a Jasco FP-6300 Spectrofluorometer.

# 3. Results

## 3.1 Construction of fusion plasmids

The fusion construct (THX-6His-TCS-PR) contained a thioredoxin (TRX) moiety followed by a hexahistidine tag (6His), thrombin cleavage site (TCS) and protease (PR) (Fig. 2). A Q7K mutation, known to decrease autolysis, was incorporated into the protease coding region of all three fusion constructs [12].

## 3.2 Overexpression of fusion proteases

Figure 3 represents whole-cell lysates. The gel shows the improved expression profile of the fusion wild-type (Fig. 3, lane 2, ~25 kDa) compared to that of the control wild-type protease (Fig. 3, lane 1, ~11 kDa) purified using ion exchange chromatography. Samples were normalised before electrophoresis to ensure that equal amounts of cell lysate were loaded onto each gel. The expression of the control wild-type was verified by separating the soluble and insoluble fractions.

The size of the fusion-product corresponds to the predicted size of the reduced, monomeric fusion protein (ProtParam tool, http://www.expasy.ch/tools/protparam.html) [24].

## 3.3 Protease purification

The control wild-type was overexpressed and purified by ion exchange chromatography as previously described by Naicker *et al.* (2014) [25]. The three fusion proteases were purified by immobilised metal ion affinity chromatography (IMAC). The steps involved in the purification of the fusion wild-type, N37T↑V and L38↑N↑L proteases are shown in Figure 4A, B and C.

The insoluble cell fractions used in this study were incubated in buffer containing 8 M urea, 10 mM Tris-HCl and 2 mM DTT (pH 9). The cell debris was collected by centrifugation, and the resultant supernatant was diluted to a final concentration of 4 M urea before the first IMAC step. The dilution was performed to prevent spontaneous crystallisation of the urea. The L38↑N↑L fusion construct was purified from the soluble fraction by metal ion affinity chromatography. The first chromatographic step yielded high concentrations of all three fusion proteases. The final pure protein samples (last lane of each gel in Fig. 4) were taken directly after the second IMAC step and were not normalised prior to performing SDS-PAGE. The final yield of the fusion derived proteases were between 150% and 250% higher than the yield of protease derived from the control method. The final concentration of free

protease is represented in milligrams per litre of culture and the data are represented in Figure 5.

### 3.4 In vitro fusion protease processing

Thrombin cleavage trials were conducted on the TRX-6His-TCS-PR construct to determine the optimal time, temperature and amount of thrombin required for optimal cleavage (Fig. 6A). We found that ideal cleavage occurred overnight at 20 °C with 1 U/ml thrombin. As thrombin cleavage progressed, protease activity (Fig. 6B) was measured by conducting enzyme assays which followed the cleavage of Abz-Arg-Val-Nle-Phe($NO_2$)-Glu-Ala-Nle-$NH_2$. This peptide mimics the cleavage site between the capsid and p2 proteins of the Gag-Pol polyprotein. To assess whether the protease samples possessed functional activity, we monitored the activity of the proteases using linear progress curves. The functional activity of the protease sample was measured concurrently with thrombin cleavage to confirm that the protein regains activity after cleavage from the TRX-6His tag. The assay was performed on a control sample containing no HIV-1 protease. The control sample possessed no activity whatsoever (data not included). Specific activity assays were performed after overnight cleavage and it was found that both mutants were active (manuscript in preparation). Both mutants were inactive prior to thrombin cleavage.

### 3.5 Structural analysis

The quaternary structures of the fusion proteases were analysed using high-performance liquid chromatography (Fig. 7). The results indicate that the dimeric sizes of the proteins were: 22 kDa, 23 kDa and 22 kDa for the fusion wild-type, N37T↑V and L38↑N↑L proteases, respectively. These sizes correspond with the expected sizes of the fully folded homodimeric molecule. Additionally, HPLC was performed on the pure fusion-derived wild-type to determine if any other molecular species were present after purification (Fig. 8).

### 3.6 Enzyme activity determination

An active site titration was performed on each purified protease sample using a VP-ITC Microcalorimeter (Fig. 9). It is important to assess the percentage active protease in a sample preparation because HIV-1 protease possesses autolytic activity [12]. This procedure, therefore, allows the experimenter to correct the concentration of active protease in a sample. Obtaining the concentration via absorbance spectroscopy at 280 nm and applying the Beer-Lambert equation is insufficient. Acetyl pepstatin is a naturally occurring weak inhibitor, and it was titrated against each protease sample. Since the stoichiometry of acetyl pepstatin

binding to HIV-1 protease is known (1:1), it is possible to determine the concentration of active protease in a sample as a function of the total measured protease concentration [26, 27]. Upon titration, the percentage of each protease in the active conformation was; 13% fusion wild-type, 32% N37T↑V and 9% L38↑N↑L.

## 4 Discussion

HIV-1 protease represents a major drug target in the treatment of HIV/AIDS. To study this enzyme, it is important to obtain sufficient quantities for use in biochemical and biophysical studies. Heterologous overexpression of the viral enzyme does not occur readily. In fact, HIV-1 protease exhibits cytotoxic effects when expressed in a variety of host cells, including bacteria, yeast and mammalian cells. The rationale of constructing a fusion protein was to reduce the cytotoxic effects observed when HIV protease is heterologously expressed in *E. coli* in the absence of a fusion tag [28–30]. In addition, incorporating the tag to the N-terminus of the HIV protease also enhances expression of soluble protein [19]. In this paper, we describe the overexpression and purification of the wild-type and two variant HIV-1 proteases using a thioredoxin-hexahistidine fusion system. A thioredoxin moiety coupled with a hexahistidine tag successfully improved the overexpression of all three proteases.

Plasmid inserts were designed to express each protease (wild-type and two variant proteases) as a fusion protease to reduce cytotoxic effects during host cell overexpression of the proteases. The fusion construct contained a thioredoxin (TRX) moiety for enhanced expression by reducing cytotoxicity [21]. This moiety was followed by a hexahistidine (6His) tag for ease of purification. A thrombin cleavage site (TCS) was included after the His-tag to allow excision of the protease molecule from the TRX-6His-TCS-PR construct.

Immobilised metal ion affinity chromatography was used to purify the fusion proteases from crude cell lysates created from each clone. Human plasma thrombin was used to cleave the TRX-6His tag from the protease, and this permitted the homodimeric assembly of the HIV-1 protease molecules. The acquisition of untagged protease from the TRX-6His tag by thrombin cleavage yielded improved amounts of pure protease. Gel analysis indicated that no autolytic activity occurred before the final thrombin cleavage step. As the thrombin cleavage assay progressed, an increasing amount of fluorogenic substrate was cleaved indicating that the dimeric protease species was active.

The fusion wild-type protease was expressed in the insoluble fraction whereas the L38↑N↑L variant was expressed in the soluble fraction of the cell lysate. Interestingly, the N37T↑V variant was expressed in roughly equal amounts in the soluble and insoluble cell fractions. It would be beneficial to transfer the expression of N37T↑V into the insoluble cell fraction completely by altering the overexpression conditions. Altering the expression profile may be achieved by varying the IPTG concentration, induction time or temperature of the induction experiment [31].

The overexpression of the fusion proteases was notably greater than that of the control wild-type protease which was purified using ion exchange chromatography. Conventionally, a large volume of culture media (6-8 L) is required to generate a sufficient amount of protease. The control method produced a final concentration of 0.8 mg/L. In this paper, we demonstrated that one litre of culture could produce 2.5 mg/L of fusion wild-type, 2 mg/L of N37T↑V and 1.5 mg/L of L38↑N↑L. Our novel HIV protease fusion purification method, therefore, produced significantly higher concentrations of pure protease than our control method (Gag-Pol derived protease).

The quaternary structure of each protease was analysed by determining the relative hydrodynamic volume using high-performance liquid chromatography. An online tool was also used to predict the sizes of the dimeric proteases for comparison. The predicted sizes of all the fusion proteases were 22 kDa (ProtParam tool) [24]. HIV-1 protease is an obligate homodimer and must be conformationally stable to function correctly. The experimentally determined sizes of the proteases were as follows: fusion wild-type, 22 kDa; N37T↑V, 23 kDa and L38↑N↑L, 22 kDa. The sizes correspond to the homodimeric size of the HIV-1 protease.

HIV-1 protease is autolytic. Therefore, it is crucial to determine the percentage of active enzyme in a prepared sample. Active site titrations, determined using ITC, showed the percentage of active enzyme in each protease sample and also verified that all the enzymes possessed enzyme activity [32]. Thirteen percent of the fusion wild-type enzyme sample was active and available to the natural ligand; whereas, N37T↑V and L38↑N↑L had 32% and 9% of active proteases in these samples, respectively. The low percentage of active proteases in the samples could be explained by high levels of autolytic activity that often occurs when proteases are incubated for an extended period (e.g. thrombin cleavage). The observed autolytic activity is particularly interesting because these proteases contain a Q7K mutation

that should minimise autolysis [33]. L38↑N↑L was expressed in the soluble fraction thus indicating that it was most likely active and able to undergo autolysis which could have contributed to the lower percentage of active protease in this sample. Additionally, HPLC on the pure protein sample showed two other molecular species present. The relative hydrodynamic volumes of these species correspond to three and four protease monomers, respectively. The presence of higher oligomeric states shows that aggregation could influence the amount of active sites in the protein sample.

Other groups have investigated the effectiveness of different HIV-1 protease fusion expression systems. In those systems, autocatalysis occurred despite the presence of the tags [17]. In our study, we postulate that the relative size of the TRX-6His-TCS moiety does not interfere with protease dimer formation – dimerisation is essential for autocatalytic activity (removal of itself from the Gag-Pol polyprotein). The thioredoxin moiety, however, sufficiently mimics the structure of the Gag protein from which HIV-1 protease cleaves itself. Here, autocatalysis (autoexcision from the Gag-Pol precursor) must not be confused with autolysis - which is the ability of a protease to undergo autoproteolytic activity in solution. Our results are, therefore, different to the studies from others [17]. In our case, we postulate that the TRX-6His-TCS moiety and the protease form higher order oligomeric states where steric hindrance effects inhibit the autocatalytic activity of the protease. This postulation is demonstrated by the observed increase substrate cleavage as a function of thrombin cleavage time as the protease is released from the TRX-6His tag (Fig. 5). The presence of higher order oligomeric states could be determined using size exclusion chromatography, analytical ultracentrifugation and static light scattering.

To prevent autolysis after cleavage of the thioredoxin tag, the protease could be incubated in a suitable concentration of inhibitor. Protease misfolding could also contribute to the presence of non-functional enzymes. Unfolding the fusion proteases in 8 M urea, before refolding in an appropriate buffer, would be expected to increase the percentage of active proteases in a prepared sample.

## 5 Conclusion

The procedure described in this study highlights a quick and easy method of HIV-1 protease purification. In addition to the smaller volume of culture media needed, the total wild-type protease yield from this fusion system exceeds our control purification method by 250%. Because the fusion proteases are autolytic, a suitable method of inhibition could be included

during the purification step so that higher yields of active protease are obtained. Although a subtype C protease was used for this study, the system could also be applied to HIV-1 proteases from other subtypes.

## 6 Compliance with ethical standards

### 6.1 Funding:

This study was funded by the University of the Witwatersrand, South African National Research Foundation Grant 68898 (HWD), the South African Research Chairs Initiative of the Department of Science and Technology, the National Research Foundation Grant 64788 (HWD) and the South African Medical Research Council (SAMRC) under a Self-Initiated Research Grant to Yasien Sayed.

### 6.2 Conflict of interest:

Jake Zondagh declares that he has no conflict of interest. Alison Williams declares that she has no conflict of interest. Ikechukwu Achilonu declares that he has no conflict of interest. Heini W. Dirr, declares that he has no conflict of interest. Yasien Sayed declares that he has no conflict of interest

### 6.3 Ethical approval:

This article does not contain any studies with human participants or animals performed by any of the authors.

## References


1. Joint United Nations Programme on HIV/AIDS (UNAIDS) (2016) Global AIDS update. Geneva, Switz.

2. Barre-Sinoussi F, Chermann J, Rey F, et al. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (80- ) 220:868–871. doi: 10.1126/science.6189183

3. Robertson DL, Anderson JP, Bradac J a, et al. (2000) HIV-1 nomenclature proposal. Science 288:492–505. doi: 10.1126/science.288.5463.55d

4. Coman RM, Robbins AH, Goodenow MM, et al. (2008) High-resolution structure of unbound human immunodeficiency virus 1 subtype C protease: Implications of flap dynamics and drug resistance. Acta Crystallogr Sect D Biol Crystallogr 64:754–763. doi: 10.1107/S090744490801278X

5. Hirsch MS, Günthard HF, Schapiro JM, et al. (2008) Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. Clin Infect Dis 47:266–285. doi: 10.1086/589297

6. McCutchan FE (2006) Global epidemiology of HIV. J Med Virol 78:S7–S12. doi: 10.1002/jmv.20599

7. Burke DS (1997) Recombination in HIV: An Important Viral Evolutionary Strategy. Emerg Infect Dis 3:253–259. doi: 10.3201/eid0303.970301

8. Debouck C, Gorniak JG, Strickler JE, et al. (1987) Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the gag precursor. Proc Natl Acad Sci U S A 84:8903–8906. doi: 10.1073/pnas.84.24.8903

9. de Oliveira T, Engelbrecht S, Janse van Rensburg E, et al. (2003) Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? J Virol 77:9422–30. doi: 10.1128/JVI.77.17.9422

10. Ermolieff J, Lin X, Tang J (1997) Kinetic properties of saquinavir-resistant mutants of human immunodeficiency virus type 1 protease and their implications in drug resistance in vivo. Biochemistry 36:12364–12370. doi: 10.1021/bi971072e

11. Tözsér J (2010) Comparative Studies on Retroviral Proteases: Substrate Specificity. Viruses 2:147–165. doi: 10.3390/v2010147

12. Mildner AM, Rothrock DJ, Leone JW, et al. (1994) The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. Biochemistry 33:9405–13.

13. Jaskólski M, Tomasselli AG, Sawyer TK, et al. (1991) Structure at 2.5-A resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. Biochemistry 30:1600–9.

14. Cheng Y-SE, McGowan MH, Kettner CA, et al. (1990) High-level synthesis of recombinant HIV-1 protease and the recovery of active enzyme from inclusion bodies. Gene 87:243–248. doi: 10.1016/0378-1119(90)90308-E

15. Sørensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol 115:113–128. doi: 10.1016/j.jbiotec.2004.08.004

16. Yang H, Nkeze J, Zhao RY (2012) Effects of HIV-1 protease on cellular functions and

their potential applications in antiretroviral therapy. Cell Biosci 2:32. doi: 10.1186/2045-3701-2-32

17. Louis JM, McDonald R a, Nashed NT, et al. (1991) Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in *Escherichia coli*. Eur J Biochem 199:361–369.

18. Hansen J, Billich S, Schulze T, et al. (1988) Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. EMBO J 7:1785–1791.

19. Volontè F, Piubelli L, Pollegioni L (2011) Optimizing HIV-1 protease production in *Escherichia coli* as fusion protein. Microb Cell Fact 10:53. doi: 10.1186/1475-2859-10-53

20. Leuthardt A, Roesel JL (1993) Cloning , expression and purification of a recombinant poly-histidine-linked HIV-1 protease. FEBS 326:275–280.

21. Mosebi S, Morris L, Dirr HW, Sayed Y (2008) Active-Site Mutations in the South African Human Immunodeficiency Virus Type 1 Subtype C Protease Have a Significant Impact on Clinical Inhibitor Binding: Kinetic and Thermodynamic Study. J Virol 82:11476–11479. doi: 10.1128/JVI.00726-08

22. Sievers F, Wilm A, Dineen D, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. doi: 10.1038/msb.2011.75

23. Naicker P, Achilonu I, Fanucchi S, et al. (2013) Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance. J Biomol Struct Dyn 31:1370–1380. doi: 10.1080/07391102.2012.736774

24. Gasteiger E, Hoogland C, Gattiker A, et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protoc Handb 571–607. doi: 10.1385/1-59259-890-0:571

25. Naicker P, Sayed Y (2014) Non-B HIV-1 subtypes in sub-Saharan Africa: impact of subtype on protease inhibitor efficacy. Biol Chem 395:1151–1161. doi: 10.1515/hsz-2014-0162

26. Velazquez-Campoy A, Freire E (2006) Isothermal titration calorimetry to determine association constants for high-affinity ligands. Nat Protoc 1:186–91. doi: 10.1038/nprot.2006.28

27. Velazquez-campoy A, Kiso Y, Freire E (2001) The Binding Energetics of First- and Second-Generation HIV-1 Protease Inhibitors : Implications for Drug Design. 390:169–175. doi: 10.1006/abbi.2001.2333

28. Baum EZ, Bebernitz GA, Gluzman Y (1990) Isolation of mutants of human immunodeficiency virus protease based on the toxicity of the enzyme in *Escherichia coli*. Proc Natl Acad Sci 87:5573–5577. doi: 10.1073/pnas.87.14.5573

29. Blanco R, Carrasco L, Ventoso I (2003) Cell Killing by HIV-1 Protease. J Biol Chem 278:1086–1093. doi: 10.1074/jbc.M205636200

30. Cheng YE, Lo K, Hsu H, et al. (2006) Screening for HIV protease inhibitors by protection against activity-mediated cytotoxicity in *Escherichia coli*. J Virol Methods 137:82–87. doi: 10.1016/j.jviromet.2006.06.003

31. Sawyer JR, Schlom J, Kashmiri S V (1994) The effects of induction conditions on production of a soluble anti-tumor sFv in *Escherichia coli*. Protein Eng 7:1401–1406.

32. Tomasselli AG, Olsen MK, Hui JO, et al. (1990) Substrate analogue inhibition and active site titration of purified recombinant HIV-1 protease. Biochemistry 29:264–269. doi: 10.1021/bi00453a036

33. Ro JR, Saltoliii R, Craikli CS (1993) Regulation of Autoproteolysis of the HIV-1. Mol Biol 101:11939–11945.

## Figure Captions

**Fig. 1 A** Homology models of the (1) N37T↑V and (2) L38↑N↑L proteases. The secondary structural elements of the homology models are rendered as ribbons. The relative positions of the amino acid insertions are indicated by arrows. Spheres without arrows represent background mutations present in each variant. The N37T↑V variant has the following mutations: I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L. The background mutations in L38↑N↑L include K20R, E35D, R57K and V82I. The homology models were generated with the molecular visualisation software programme PyMOL, using data from the Protein Data Bank (PDB ID: 3U71). **B** The sequence alignment data shows the positions of the mutations. The wild-type subtype C protease sequence is included as a reference. The alignment was performed using the Clustal Omega tool (EMBL-EBI)**.**

**Fig. 2** Plasmid construct of TRX-6His-TCS-PR. The abbreviation TRX-6His-TCS-PR denotes the thioredoxin-like moiety (TRX), hexahistidine (6His) tag and protease (PR) enzyme. A thrombin cleavage site (TCS) is present between the hexahistidine tag and the protease. The size of each constituent is shown and is represented in kilodalton. The entire construct is ~25.1 kDa. The figure was adapted from a figure in SnapGene® (GSL Biotech; available at snapgene.com).

**Fig. 3 A** SDS-PAGE gel showing the overexpression of control wild-type and fusion wild-type. Whole lysates are shown. Transformed BL21 (DE3) pLysS *E. coli* cells were grown to early exponential phase and induced for six hours with 1 mM IPTG. MW: molecular weight marker, lane 1: control wild-type protease overexpression, lane 2: fusion wild-type overexpression. The positions of the fusion wild-type protease (lane 2, ~25 kDa) and the control wild-type protease (lane 1, ~11 kDa) are indicated by arrows. **B** Soluble and insoluble cell fractions of control wild-type. MW: molecular weight marker, lane 1: insoluble cell fraction, lane 2: soluble cell fraction. The position of the protease is indicated by an arrow (~11 kDa). Protease was recovered from the insoluble cell fraction.

**Fig. 4** Overexpression profiles of fusion wild-type and variant proteases from IPTG-induced cell lysates. The purification steps, from cell lysis to final product, are shown from left to right. MW: molecular weight marker. **A** Fusion wild-type purification profile. **B** N37T↑V fusion purification profile. **C** L38↑N↑L fusion purification profile. Samples were stained with 0.25% Coomassie Blue R-250 and analysed by 16% SDS polyacrylamide gel electrophoresis. The last lane in each gel confirms the presence of pure protease for fusion wild-type,

N37T↑V variant and L38↑N↑L variant proteases, respectively. The positions of the fusion proteins (~25 kDa) and pure protease samples (~11 kDa) are indicated by arrows.

**Fig. 5** Quantity of fusion-derived HIV-1 protease produced per litre of culture media compared to an ion exchange purification method. Bar (a) 0.83 mg/L control wild-type protease, (b) 2.5 mg/L fusion-derived wild-type protease, (c) 2 mg/L N37T↑V protease, and (d) 1.5 mg/L L38↑N↑L protease.

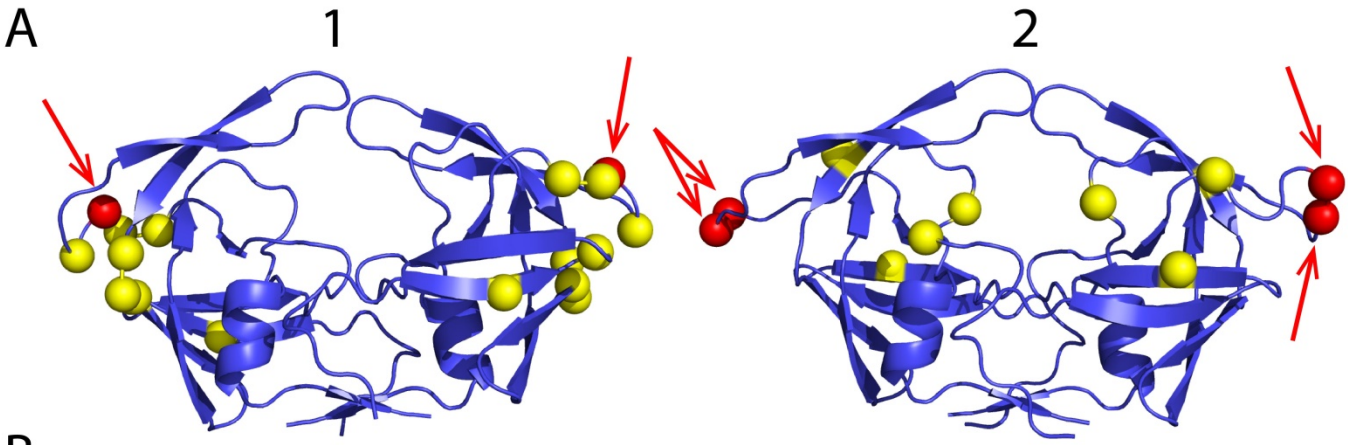
**Fig. 6 A** Time-course thrombin cleavage assay of fusion wild-type protease. **B** Fusion wild-type protease activity over time. Protease activity was monitored during thrombin (1 U/ml) cleavage by following fluorogenic substrate (Abz-Arg-Val-Nle/Phe($NO_2$)-Glu-Ala-Nle-$NH_2$) processing at a wavelength of 425 nm.

**Fig. 7** Size exclusion-HPLC retention times of **A** fusion wild-type, **B** N37T↑V and **C** L38↑N↑L proteases. The molecular standards consisted of blue dextran (2000 kDa), serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome C (12.4 kDa) and aprotinin (6.5 kDa). The retention time of each protease is indicated. The relative molecular weight of each protease was calculated from the standard curve. The elution of fusion wild-type and L38↑N↑L was 0.25 ml/min and the elution of N37T↑V was 0.20 ml/min.

**Fig. 8 A** Size exclusion-HPLC chromatogram of pure fusion wild-type. Three molecular species are present. The size of each species was calculated. (a) 22 kDa (b) 47 kDa and (c) 74 kDa. **B** Standard curve used to extrapolate the log molecular weight of each molecular species in the sample. The molecular standards consisted of blue dextran (2000 kDa), serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome C (12.4 kDa) and aprotinin (6.5 kDa). The retention time of each species is indicated. The elution elution was performed at 0.20 ml/min.

**Fig. 9** Active site titrations were performed on **A** fusion derived wild-type, **B** N37T↑V and **C** L38↑N↑L protease. Each protease was titrated against acetyl-pepstatin. The data were fitted to a one-to-one model.

**Fig 1**




B

```
N37T↑V          PQITLWKRPLVSVKVEGQIKEALLDTGADDTVLEETTV-LSGKWKPKMIGGIGGFIKVRQ 59
Wild-type       PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEEIN--LPGKWKPKMIGGIGGFIKVRQ 58
L38↑N↑L         PQITLWKRPLVSIKVGGQIREALLDTGADDTVLEDINLNLPGKWKPKMIGGIGGFIKVKQ 60
                ************:** ***:**************:    * *****************:*

N37T↑V          YEEVPIEICGKKAIGTVLIGPTPVNIIGRNLLTQLGCTLNF 100
Wild-type       YDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF 99
L38↑N↑L         YDQILIEICGKKAIGTVLVGPTPINIIGRNMLTQLGCTLNF 101
                *::: *************:****:******:**********
```

**Fig. 2**

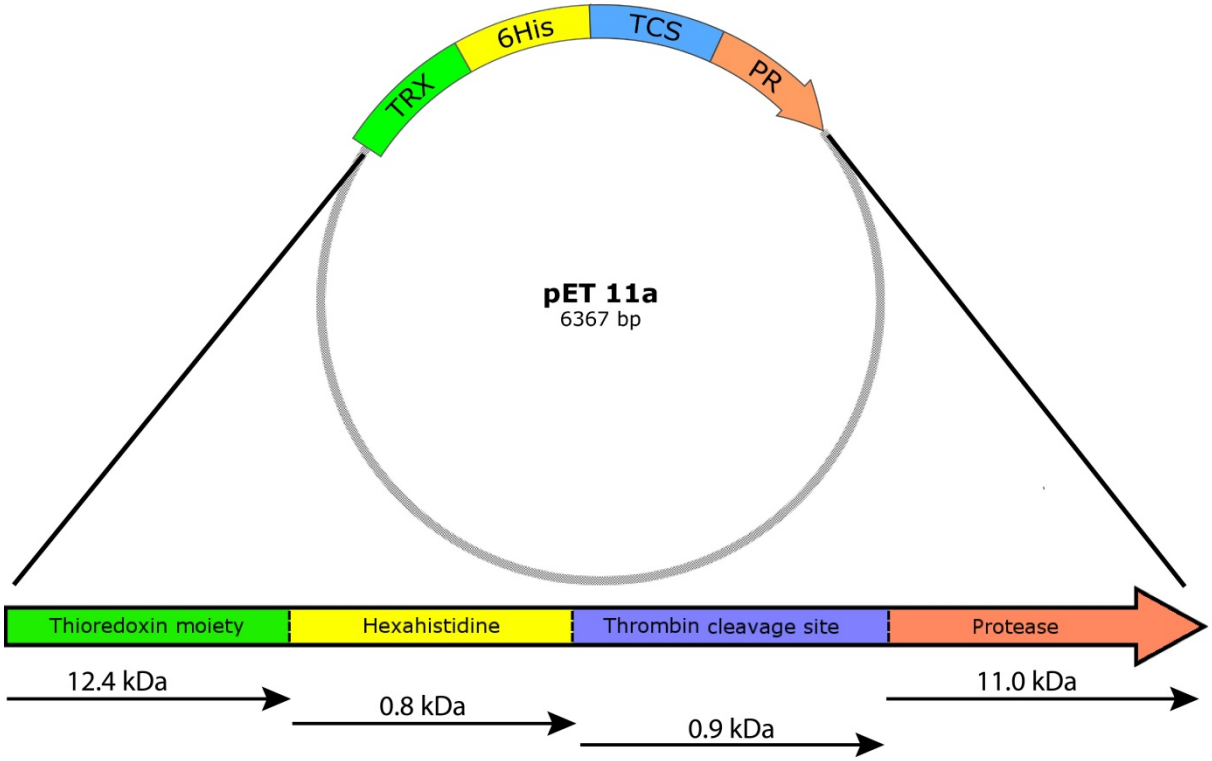

TRX
6His
TCS
PR
pET 11a
6367 bp
Thioredoxin moiety
Hexahistidine
Thrombin cleavage site
Protease
12.4 kDa
0.8 kDa
0.9 kDa
11.0 kDa

**Fig. 3**

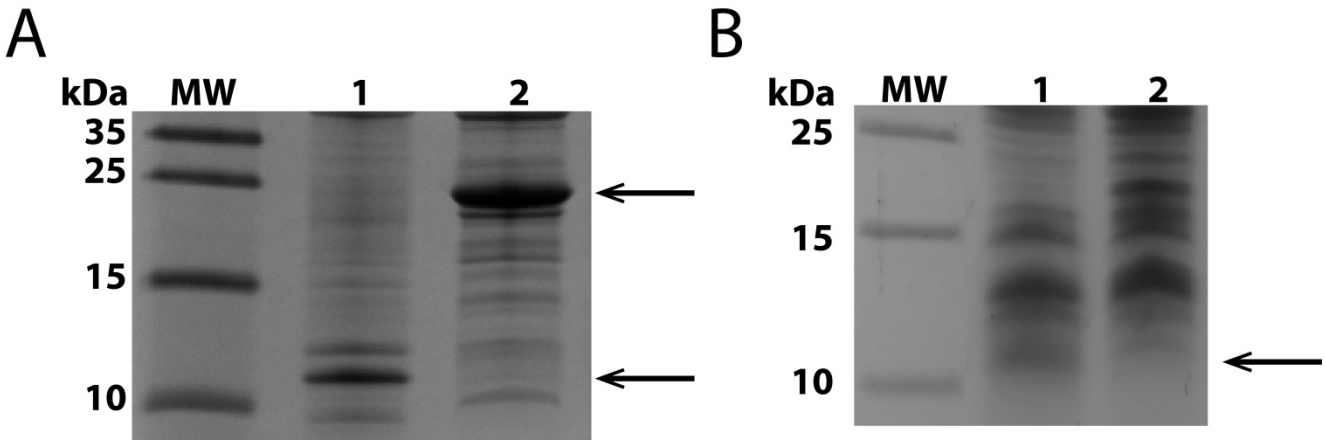


**Fig. 4**

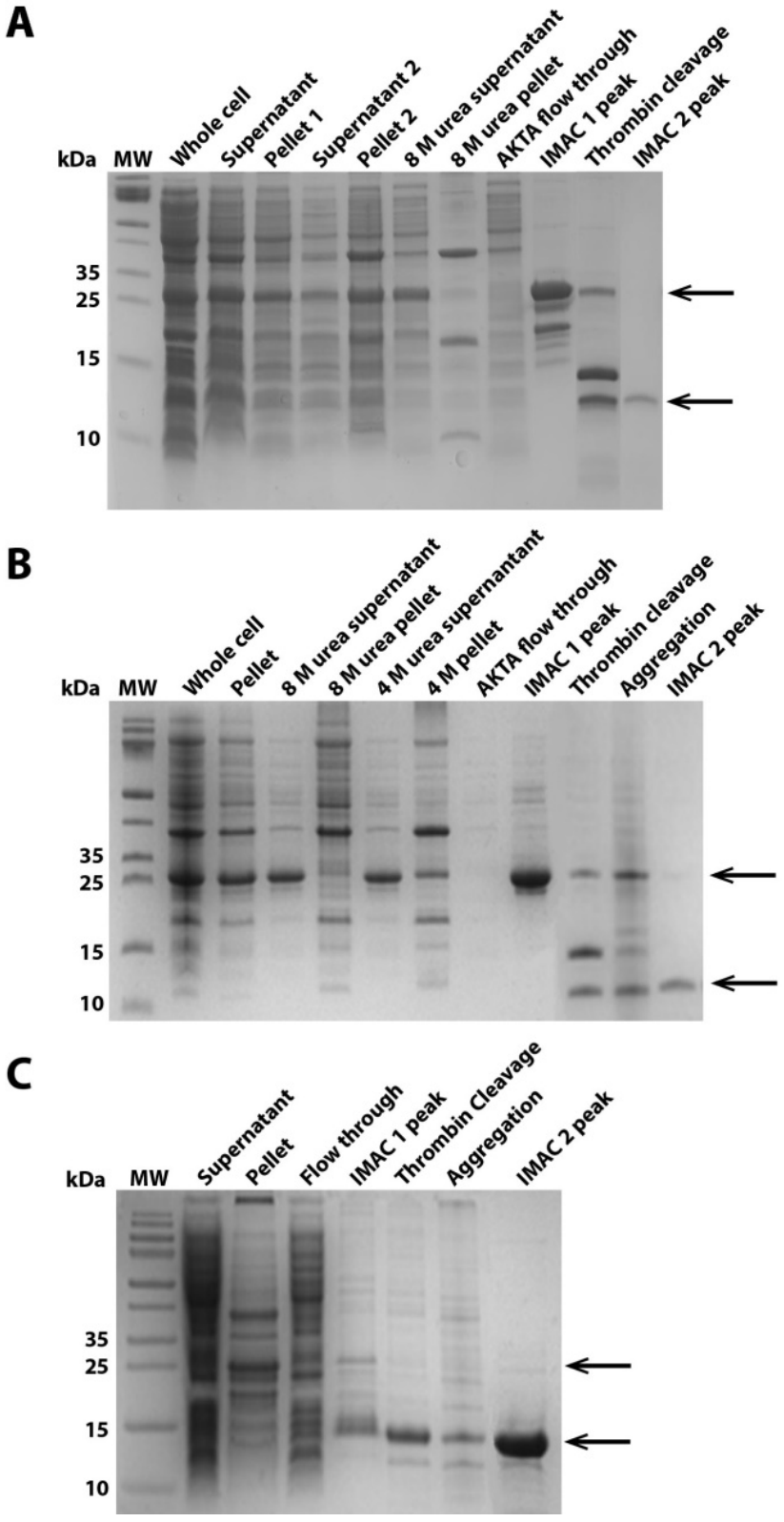

A
kDa
MW
Whole cell
Supernatant
Pellet 1
Supernatant 2
Pellet 2
8 M urea supernatant
8 M urea pellet
AKTA flow through
IMAC 1 peak
Thrombin cleavage
IMAC 2 peak
35
25
15
10
B
kDa
MW
Whole cell
Pellet
8 M urea supernatant
8 M urea pellet
4 M urea supernantant
4 M pellet
AKTA flow through
IMAC 1 peak
Thrombin cleavage
Aggregation
IMAC 2 peak
35
25
15
10
C
kDa
MW
Supernatant
Pellet
Flow through
IMAC 1 peak
Thrombin Cleavage
Aggregation
IMAC 2 peak
35
25
15
10

**Fig. 5**

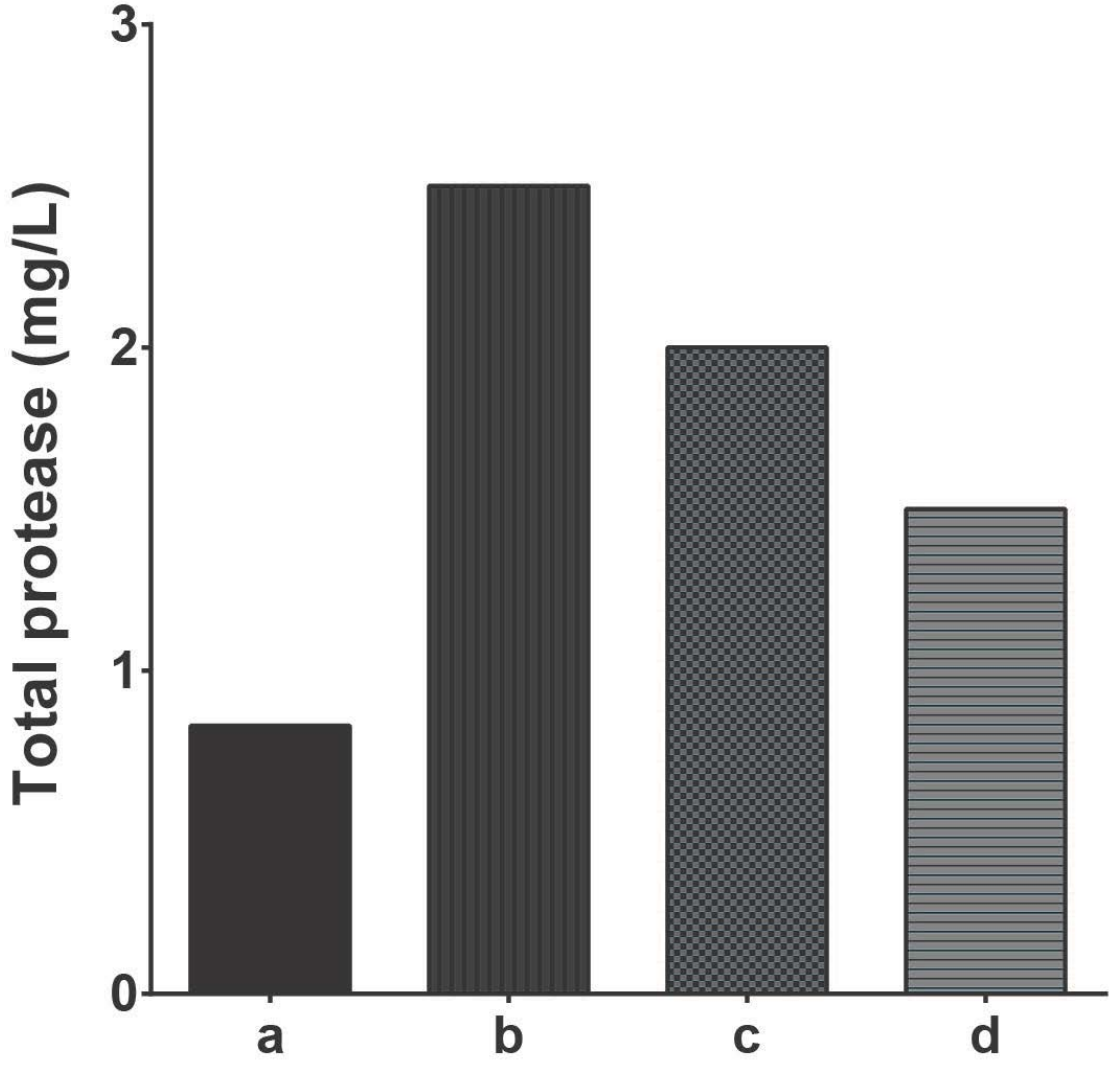


**Fig. 6**

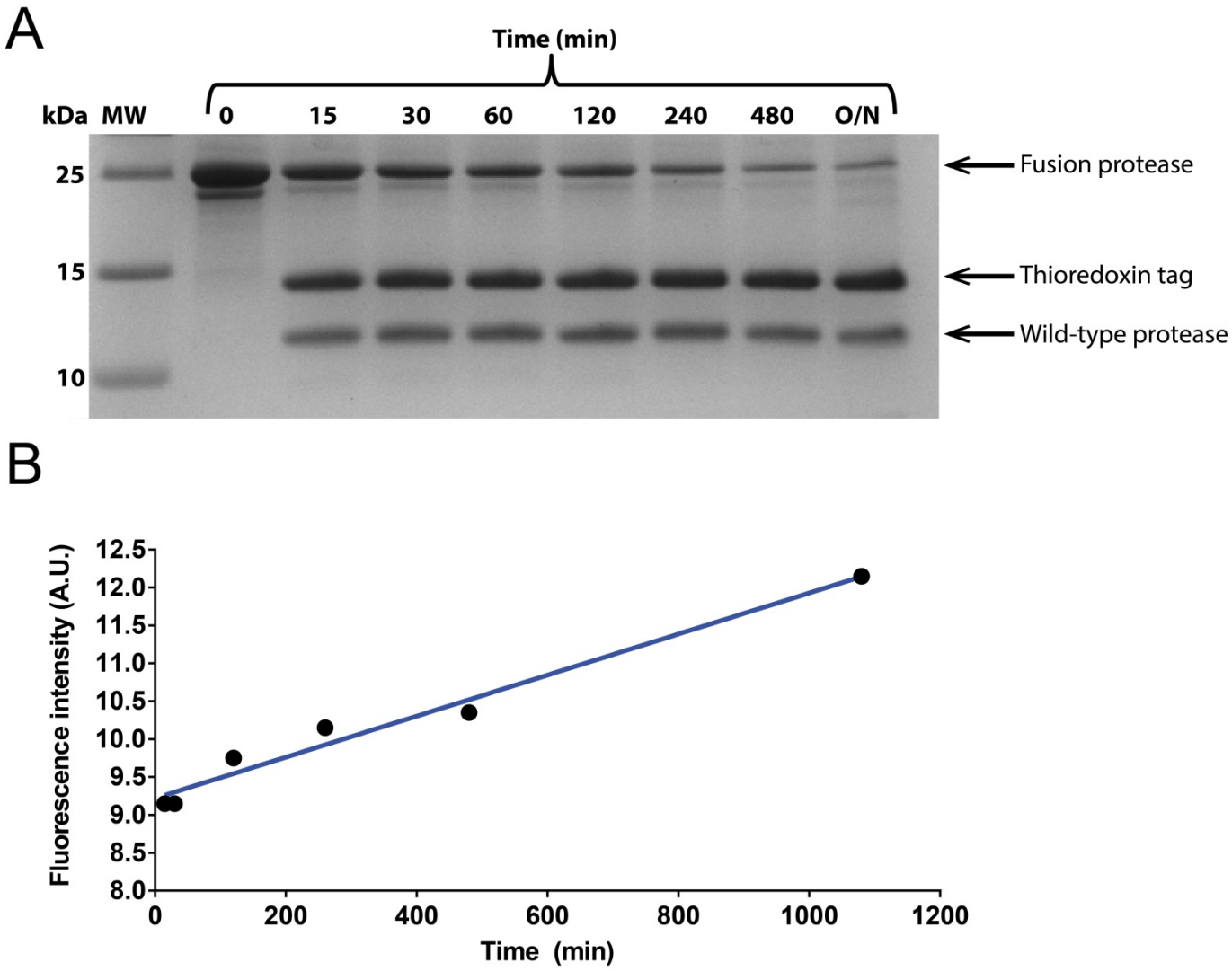

A
Time (min)
kDa MW
0
15
30
60
120
240
480
O/N
25
15
10
Fusion protease
Thioredoxin tag
Wild-type protease
B
Fluorescence intensity (A.U.)
12.5
12.0
11.5
11.0
10.5
10.0
9.5
9.0
8.5
8.0
0
200
400
600
800
1000
1200
Time (min)

**Fig. 7**

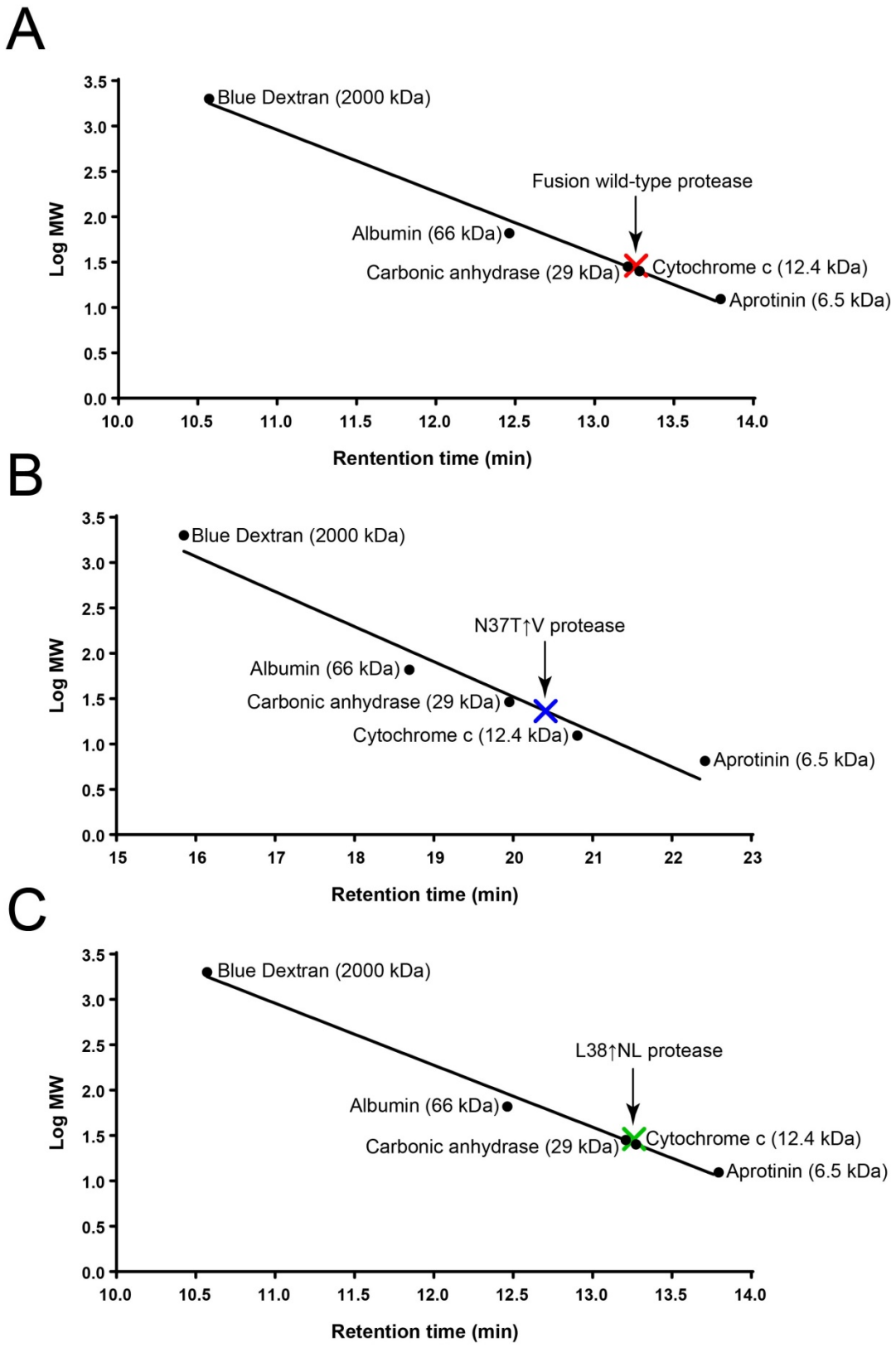

A
Blue Dextran (2000 kDa)
Fusion wild-type protease
Albumin (66 kDa)
Carbonic anhydrase (29 kDa)
Cytochrome c (12.4 kDa)
Aprotinin (6.5 kDa)
Log MW
Rentention time (min)
B
Blue Dextran (2000 kDa)
N37T↑V protease
Albumin (66 kDa)
Carbonic anhydrase (29 kDa)
Cytochrome c (12.4 kDa)
Aprotinin (6.5 kDa)
Log MW
Retention time (min)
C
Blue Dextran (2000 kDa)
L38↑NL protease
Albumin (66 kDa)
Carbonic anhydrase (29 kDa)
Cytochrome c (12.4 kDa)
Aprotinin (6.5 kDa)
Log MW
Retention time (min)

**Fig. 8**

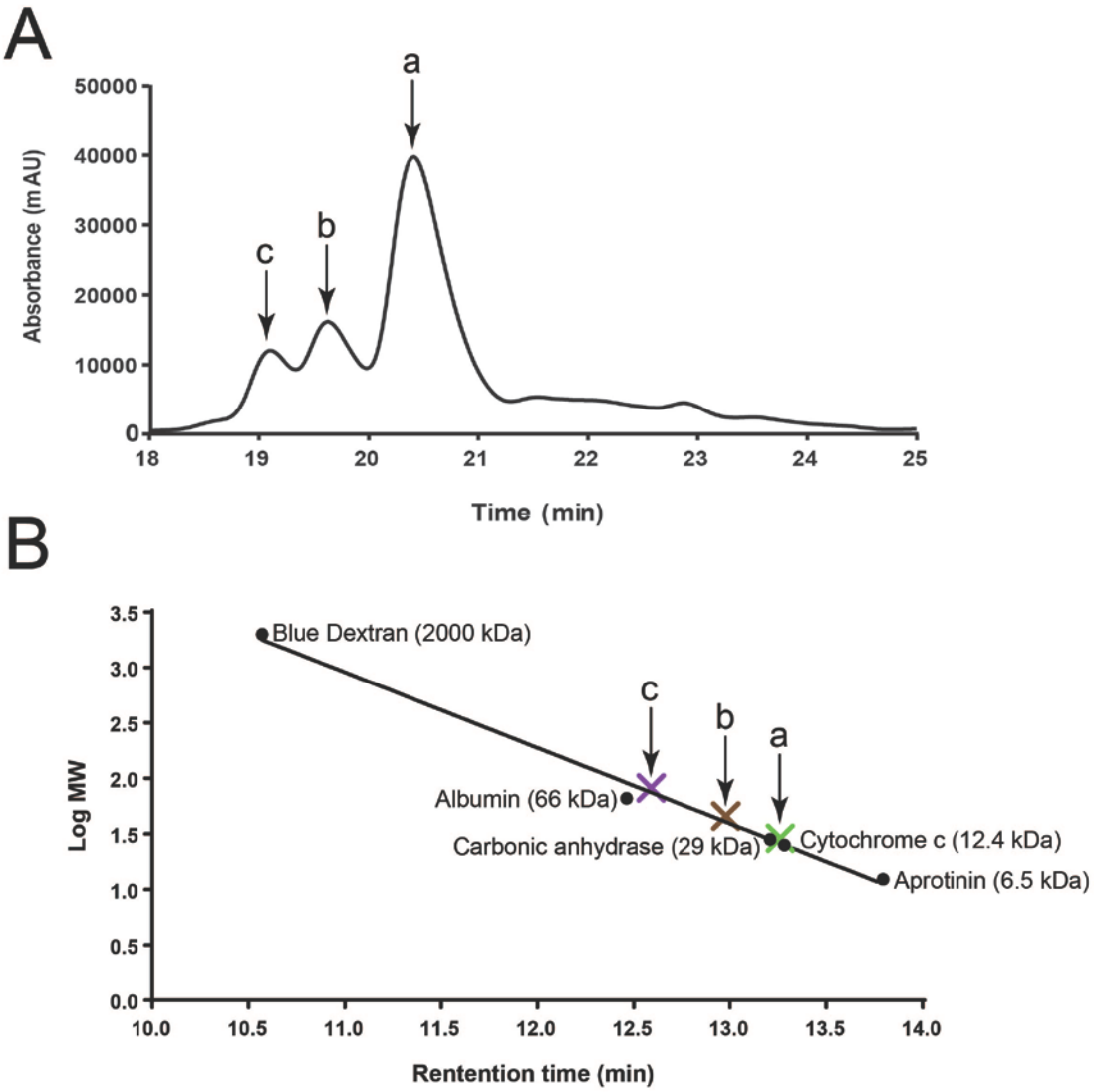
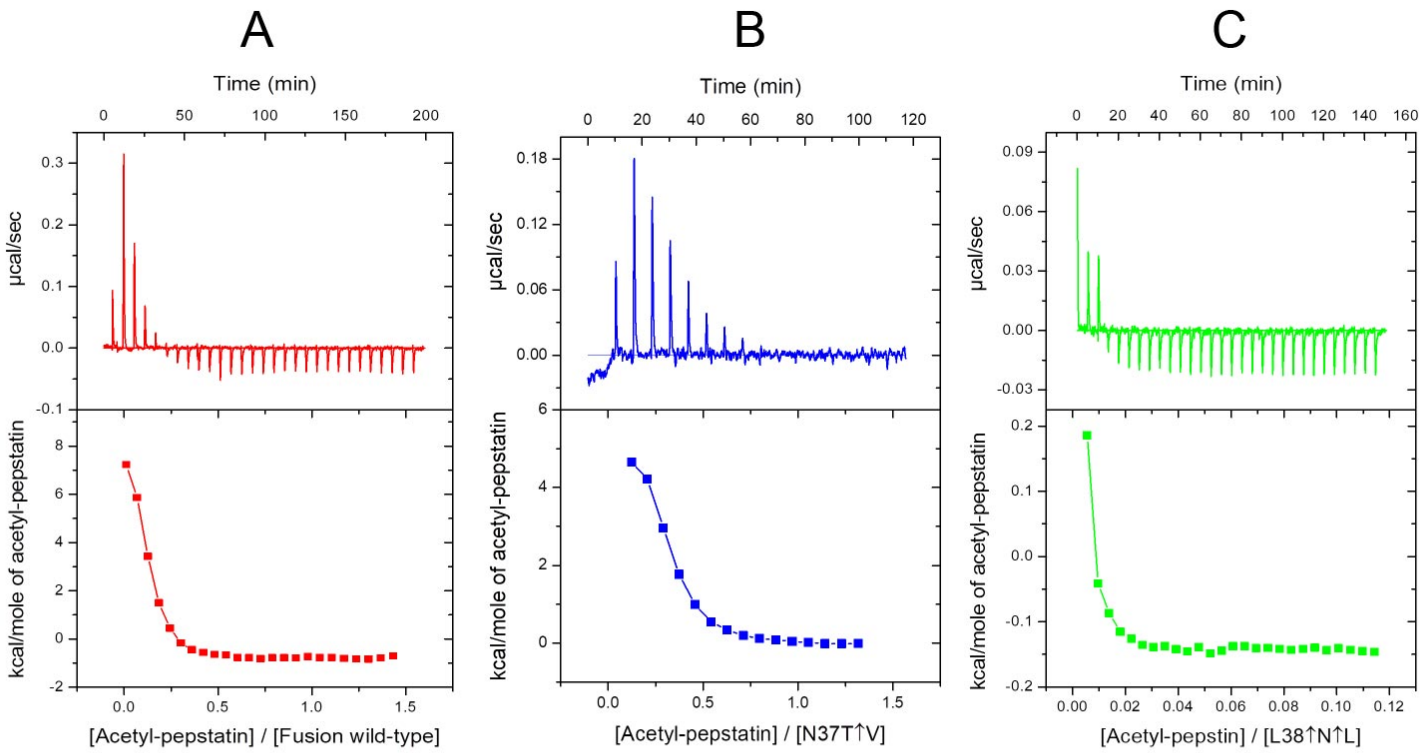

A
a
b
c
50000
40000
30000
20000
10000
0
Absorbance (mAU)
18
19
20
21
22
23
24
25
Time (min)
B
Blue Dextran (2000 kDa)
c
b
a
Albumin (66 kDa)
Carbonic anhydrase (29 kDa)
Cytochrome c (12.4 kDa)
Aprotinin (6.5 kDa)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Log MW
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
Rentention time (min)

**Fig. 9**


A
B
C
Time (min)
µcal/sec
kcal/mole of acetyl-pepstatin
[Acetyl-pepstatin] / [Fusion wild-type]
[Acetyl-pepstatin] / [N37T↑V]
[Acetyl-pepstin] / [L38↑N↑L]

# CHAPTER 3

## Drug Susceptibility and Replication Capacity of a Rare HIV-1 Subtype-C Protease Hinge Region Variant

Jake Zondagh, Adriaan E. Basson, Ikechukwu Achilonu, Lynn Morris, Heini W. Dirr and Yasien Sayed.

(Manuscript in preparation)

In this manuscript, the drug susceptibility and replication capacity of an HIV-1 Subtype C isolate is described. The N37T↑V protease confers reduced drug susceptibility to three commonly used protease inhibitors. Furthermore, the N37T↑V protease was observed to increase viral replication capacity.

Author contributions: Jake Zondagh performed all the experimental work, analysed the data and wrote the manuscript. Adriaan E. Basson assisted with the phenotypic drug susceptibility assays and viral replication assays. Yasien Sayed, Adriaan E. Basson, Lynn Morris, Ikechukwu Achilonu and Heini Dirr assisted in manuscript revision. Yasien Sayed supervised the project and assisted in data analysis and interpretation.

# Drug Susceptibility and Replication Capacity of a Rare HIV-1 Subtype-C Protease Hinge Region Variant

Jake Zondagh[†], Adriaan E. Basson[§±], Ikechukwu Achilonu[†], Lynn Morris[§‡], Heini W. Dirr[†], Yasien Sayed[†]*

† Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of Witwatersrand, Johannesburg, 2050, South Africa

± HIV Pathogenesis Research Unit, Department of Molecular Medicine and Haematology, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa

§ Centre for HIV and STIs, National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service (NHLS), Johannesburg, South Africa

‡ Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

* Corresponding author: Yasien Sayed; e-mail address: Yasien.Sayed@wits.ac.za.

## Abstract

Protease inhibitors form the main component of second-line antiretroviral treatment in South Africa. Despite their efficacy, mutations arising within the HIV-1 *gag* and *protease* genes contribute to the development of resistance against this class of drug. In this paper, we investigate the drug susceptibility, replication capacity and catalytic activity of a South African HIV-1 subtype C Gag-protease (W1201i) that contained a mutation and insertion (N37T↑V), located within the hinge region of the protease. An *in vitro* single-cycle drug susceptibility assay showed a small (3-fold), but significant ($p<0.0001$) reduction in drug susceptibility to darunavir when compared to a wild-type control (MJ4). Substitution of W1201i-Gag with MJ4-Gag resulted in an additional small (2-fold), but significant ($p<0.01$) reduction in susceptibility to lopinavir and atazanavir. The W1201i pseudovirus had a significantly ($p<0.01$) reduced replication capacity (16.4%) compared to the wild-type control. However, this was dramatically increased to 160% ($p<0.05$) when W1201i-Gag was substituted with wild-type control-Gag. Furthermore, the N37T↑V protease displayed reduced catalytic processing power when assayed against a fluorogenic substrate that mimics the wild-type CA/p2 cleavage site. Collectively, these data suggests that the N37T↑V mutation and insertion, marginally increases viral infectivity and decreases drug susceptibility. This is contrary to most other secondary mutations which do not increase viral infectivity and are usually only able to confer reduced drug susceptibility when modified by active site resistance mutations. Additionally, polymorphisms arising in Gag can modify the impact of protease with regards to viral replication and susceptibility to protease inhibitors.

**Keywords:**

HIV-1; Subtype C; protease; hinge region; Gag; drug susceptibility; phenotypic assay; drug resistance; replication capacity, catalytic efficiency, specific activity.

**Abbreviations**

| | |
|---|---|
| ART: | Antiretroviral Therapy |
| ARV: | Antiretroviral |
| ATV: | Atazanavir |
| AZT: | Azidothymidine |
| DRV: | Darunavir |
| ELISA: | Enzyme-Linked Immunosorbent Assay |
| HIV-1: | Human Immunodeficiency Virus type 1 |
| LPV: | Lopinavir |
| MD: | Molecular Dynamics |
| NVP: | Nevirapine: |
| NNRTIs: | Non-Nucleoside Reverse Transcriptase Inhibitors |
| NRTIs: | Nucleoside Reverse Transcriptase Inhibitors |
| N37T↑V: | HIV-1 subtype C protease containing asparagine 37 mutated to threonine; the upwards arrow indicates an insertion of valine at position 37 |
| PMTCT: | Prevention of Mother-to-Child Transmission |
| PCR: | Polymerase Chain Reaction |
| PIs: | Protease Inhibitors |
| PR: | Protease |
| RC: | Replication Capacity |
| Resistance control: | Multi-drug resistant *gag-protease* isolate |
| RTV: | Ritonavir |
| Wild-type control: | MJ4GP |
| WTGagN37T↑VPR: | Chimeric construct consisting of patient-derived *protease*, combined with wild-type *gag* |
| W1201i: | Patient-derived HIV-1 *gag-protease* |

## Introduction

HIV-1 subtype C is responsible for the majority of HIV infections in Southern Africa, the epicentre of the global HIV pandemic [1]. The HIV-1 protease enzyme remains an attractive drug target because it is among the three viral enzymes involved in viral replication and is responsible for the maturation of an infective virion [2,3]. There are currently nine FDA approved protease inhibitors (PIs) available to those infected with HIV [4,5].

PIs are substrate transition state analogues that bind competitively to the protease active site [3]. PIs are prescribed as part of the second-line regimen in South Africa, after failure on a non-nucleoside reverse transcriptase inhibitor (NNRTI)-based regimen [6]. Of the nine approved PIs, lopinavir (LPV) or atazanavir (ATV) are preferred. Darunavir (DRV) is exclusively prescribed for salvage therapy after failure on a PI-based second-line regimen [7]. ATV and LPV are co-administered with ritonavir (RTV) as it functions as a potent inhibitor of the xenobiotic detoxification enzyme cytochrome P450 Isoform 3A4, thereby increasing PI bioavailability [8]. However, ritonavir (RTV) is avoided as a standalone drug due to the severity of its side-effects at higher concentrations, and due to the selection of drug resistance mutations [8–10].

The development of antiretroviral (ARV) drug resistance remains one of the most significant hurdles in the fight for sustained viral suppression within HIV-1 infected patients on antiretroviral therapy (ART) [11]. PI-resistance is mediated by primary mutations located within the active site of the enzyme. These mutations ordinarily give rise to low levels of resistance when not enhanced by secondary mutations distal to the active site. Furthermore, studies indicate that the drug resistance profile of HIV-1 protease is enhanced when compensatory changes arise within the Gag polyprotein, which can modulate the replicative capacity of the virus to further decrease the likelihood of successful treatment [8,12–16].

A recent study has reported on an HIV-1 isolate (W1201i) from an HIV-1 infected South African drug-naïve infant with an N37T↑V hinge region mutation and insertion in the protease enzyme [17]. An increasing number of insertion mutations in protease are being reported, although their impact on drug susceptibility has not been well characterised [18]. The hinge (residues 35-42 and 57-61) of protease is highly variable and controls the molecular dynamics (MD) of the flap region (residues 46-54) [19]. The flap region, in turn, controls substrate entry into the active site and influences drug binding [8]. MD simulations suggest that the N37T↑V hinge region mutation increases the dynamics of the flap region

which alters its kinetic profile [20]. The altered kinetics of the N37T↑V protease may result in reduced PI drug susceptibility [8,21]. Here, we report on the *in vitro* drug susceptibility, viral replication capacity and enzymatic catalytic efficiency of the W1201i isolate.

## Materials and Methods

### 2.1 Cohort study

Plasma samples from newly diagnosed PI-naïve children younger than two years of age were investigated in a previous study that examined drug resistance following short-course treatment to prevent mother-to-child HIV transmission [17]. A sample from an infant (W1201i) was selected for further study because the isolate was shown to contain a protease hinge region mutation and insertion (N37T↑V). No ethics clearance was required for this research. The sequences are available on GenBank (personal communication from Professor Lynn Morris).

Prenatally, the mother of W1201i was PI-naïve but had been exposed to nevirapine (NVP) for 42 days before labour to prevent mother-to-child transmission. The infant had been treated with azidothymidine (AZT) as prophylaxis after birth and was PI-naïve at the time blood samples were taken [17].

### 2.2 Vector construction

Total viral RNA was extracted from patient plasma using the QIAmp Viral RNA Mini kit (QIAGEN, Belgium) and reverse transcribed using a Thermoscript RT-PCR kit (Invitrogen, CA). The 1.8 kb *gag-protease* amplicon was amplified by nested polymerase chain reaction (PCR) using an Expand High Fidelity PCR kit (Roche Applied Science, Basel, Switzerland). Population-based Sanger sequencing of *gag-protease* and the construction of the patient-derived HIV-1 expression vector was performed as previously described by Giandhari *et al.*, (2016).

A wild-type reference (p8.9MJ4GP) containing the *gag-protease* of a subtype C reference isolate (MJ4, GenBank: AF321523.1) and a known multi-drug resistant reference (resistance control) were included in this study [22,23]. Moreover, an artificial chimeric construct was created by exchanging W1201i-*gag* in the patient-derived HIV-1 expression vector with the MJ4 wild-type-*gag* to form WTGagN37T↑V. The W1201i *gag-protease*, chimeric construct and the resistance control were each cloned into a p8.9NSX+ expression vector to use in the single cycle phenotypic assays.

## 2.3 Phenotypic drug susceptibility

A single-cycle non-replication assay was used for phenotypic drug susceptibility testing [24]. Briefly, HEK293T cells were transfected the HIV-1 expression vector, the pMDG vector that expresses vesicular stomatitis virus protein G for entry, and the pCSFLW vector that encodes the firefly luciferase reporter gene. The transfected cells were harvested after incubation at 37 °C under 5% $CO_2$ for 36 hours. The harvested cells were seeded in 96-well culture plate with serially diluted PIs. After 24 hours, the supernatants were transferred to the corresponding wells of an indicator plate that contained uninfected HEK293T cells. The degree of pseudoviral infection was measured 48 hours post-transfer, as determined by the expression of firefly luciferase in infected target cells. Luciferin containing Bright-Glo (Promega, CA, USA) was added to each well and incubated for 2 minutes before reading the luminescence on the Victor 3 Luminometer (Perkin Elmer, Massachusetts).

The half maximal inhibitory concentration ($IC_{50}$) values were calculated for each sample and drug (DRV, ATV, LPV). The percentage luciferase activity was plotted against the log of the drug concentration to determine the relative half maximal inhibitory concentration ($IC_{50}$). Experiments were performed in duplicate and three experiments were averaged for the final result. The drug susceptibility (represented in fold change) of the W1201i isolate, resistance control, and the WTGagN37T↑VPR construct was expressed relative to MJ4. One-way analysis of variance (ANOVA) of the $IC_{50}$ values of MJ4 was used to identify fold change values. The lower biological cut-off value for each drug was set at the $99^{th}$ percentile of $IC_{50}$ replicates for MJ4: LPV (1.2 FC), ATV (1.4 FC) and DRV (1.7 FC). Values above these levels indicate a decrease in drug susceptibility. Figures and statistics were compiled in GraphPad Prism 5 (GraphPad Software, Inc. La Jolla, CA, USA).

## 2.4 Viral replication capacity

Replication capacity (RC) was determined by harvesting pseudovirions from transfected drug-free HEK293T cells and infecting fresh HEK293T cells with the neat viral stocks. The subsequent expression of firefly luciferase was quantified 48 hours later, as described in section 2.3. Input virus was quantified using a chemiluminescent p24-antigen ELISA assay (Protocol 2, Aalto Bio Reagents Ltd., Dublin, Ireland). The RC was calculated by referring to the ratio of input virus (nanogram p24) against the level of luciferase expression.

### 2.5 Protease characterisation

The N37T↑V protease derived from the W1201i isolate was purified as a recombinant protein through immobilised metal ion affinity chromatography [25]. The wild-type protease was purified as described by Naicker *et al*., (2013). A Q7K mutation, known to decrease autolysis without affecting catalysis, was incorporated into the protease coding region of both enzymes.

Enzymatic parameters were determined following the hydrolysis of the HIV-1 protease fluorogenic substrate (Abz-Arg-Val-Nle-Phe($NO_2$)-Glu-Ala-Nle-$NH_2$), which mimics the CA/p2 cleavage site in the HIV-1 Gag polyprotein. The Nle-Phe($NO_2$) residue efficiently quenches the substrate aminobenzoyl (Abz) group. Quenching is abolished when the substrate is cleaved, which allows Abz to fluoresce at a wavelength of 425 nm when excited at 337 nm [26,27].

For the specific activity, a variable protein concentration (10 to 50 nM) and constant substrate concentration (50 µM) was assayed under steady state conditions. The activity was determined from the slope of the progress curve. The catalytic efficiency ($k_{cat}/K_M$) was determined with a variable substrate concentration (1 to 10 µM) and a constant protein concentration (50 nM) assayed during steady state. All samples were measured for 1 minute using an excitation bandwidth of 2.5 nm and an emission bandwidth of 5 nm. The complete cleavage of 1 nmol substrate was measured and used to convert the emission intensity to activity. The catalytic constant ($k_{cat}$) for each protease was determined from their respective specific activity values.

## Results

### 3.1 Genotypic analysis of gag-protease from W1201i

The protease encoding region of the W1201i amplicon was previously shown to contain a mutation and a rare insertion at position 37 in the hinge region [10]. A homology model of the W1201i protease containing the N37T↑V as well as a sequence alignment with the consensus wild-type, MJ4 and resistance control proteases is shown in Figure 1. Compared to the MJ4, the variant protease displayed the following background mutations: Q7K, I13V, M36T, Q61E and M89L. Furthermore, three minor ATV resistance mutations were also found: G16E, D60E, and I62V [28]. Both the wild-type and N37T↑V exhibited L63P, a minor LPV resistance mutation; however, the effect of L63P was found to be negligible. The

resistance control displayed the six major multi-drug resistance mutations; namely, L10I, K20R, M46I, I54V, I62V, and V82A. Additionally, five background mutations were present: E35D, S39P Q61H, T74S and I77V. Resistance mutations were verified by the Stanford HIV drug resistance database (https://hivdb.stanford.edu/) and by Wensing *et al*., (2017).

Given the close link between Gag and protease we also sequenced the Gag region of W120li. All polymorphisms and their locations in Gag are indicated in Figure 2, which shows a sequence alignment between W1201i and a subtype C wild-type consensus sequence. Genotypic analysis (Figures 3) of W1201i *gag* showed several insertions in, or near, cleavage sites, which could potentially influence drug susceptibility and RC [29,30]. Mutations include a single PTAPP duplication and LE insertion in $p6^{Gag}$, as well as an I372L↑M mutation and insertion in the p2/NC cleavage site. In addition to I372L↑M, the p2/NC cleave site showed the following polymorphisms: S369N, S371N, I373M and G377S. Moreover, a previously unreported MSQAG duplication was found between the CA/p2 and p2/NC cleavage sites.

### 3.2 Phenotypic drug susceptibility

The W1201i Gag-proteaseThe W1201i sample-derived pseudovirus was susceptible to both LPV (FC 2.0) and ATV (FC 1.6), and no significant difference ($p>0.05$) was observed compared to the wild-type control (Figure 4). However, a reduced susceptibility was observed to DRV (FC 4.6), which was significantly higher ($p<0.0001$) than the wild-type control. With the substitution of W1201i-Gag, WTGagN37T↑VPR showed a small, but significant ($p<0.001$), reduction in susceptibility for both LPV (FC 3.4) and ATV (FC 3.7). No further significant reduction ($p>0.05$) was observed for DRV (FC 4.7). The resistance control exhibited a significant ($p<0.0001$) reduction in susceptibility (FC>10) towards all three PIs. Error bars could not be included for the resistance control assayed against ATV and DRV because the drug susceptibility of these samples surpassed the sensitivity of the assay. These samples show only a qualitative relative difference between it and the other isolates.

### 3.3 Viral replication capacity

When assessed for RC, the W1201i pseudovirus showed a significant ($p<0.05$) reduction in RC (16.4%), compared to the wild-type control (100%) (Figure 5). However, with the substitution of the Gag, the resulting WTGagN37T↑VPR pseudovirus showed a drastic increase in RC (164%, $p<0.01$), compared to the wild-type virus suggesting that the polymorphisms in the W120li Gag caused the observed reduction in RC.

### 3.4 Protease characterisation

The enzyme kinetic parameters of the wild-type and N37T↑V proteases were determined via fluorogenic substrate cleavage (Table 1). The specific activity was found to be 12 ± 1 μmol.min$^{-1}$.mg$^{-1}$ for wild-type and 1.8 ± 0.2 μmol.min$^{-1}$.mg$^{-1}$ for N37T↑V, indicating that the variant shows reduced specificity toward the fluorogenic substrate. The catalytic constant (substrate turnover per second) was determined from the specific activity data and was found to be 6.4 ± 0.2 and 1.5 ± 0. 2 s$^{-1}$ for wild-type and N37T↑V, respectively. The wild-type protease was five times more catalytically efficient than the N37T↑V variant with a $k_{cat}/K_M$ value of 2.4 ± 0.3 s$^{-1}$.μM$^{-1}$ compared to 0.5 ± 0.1 s$^{-1}$.μM$^{-1}$ for wild-type. Cumulatively, the data indicate that the N37TV variant processes the fluorogenic substrate less efficiently.

## Discussion

Nucleotide insertions occur most commonly between codon 32 and 42 in the *protease* gene [18]. However, the prevalence of insertion mutation are rare and their prevalence range from 0.1 to 4.55% [15,31,32]. It has been reported that insertion mutations can modulate the activity of the enzyme and impact the RC [15,18,33,34]. Frequently, insertion mutations are found to be duplications of neighbouring amino acids generated due to the error-prone nature of reverse transcriptase. Furthermore, two-thirds of protease insertion variants contain one or more of the major resistance mutations [15,34]. We have previously shown that the N37T↑V mutation and insertion increases the molecular dynamics of both the flap and the hinge regions in HIV-1 protease [20]. Here, we investigated the *in vitro* phenotypic drug susceptibility, replication capacity and enzymatic properties of this insertion variant to understand its impact on HIV-1 protease.

As anticipated, W1201i displayed a significant reduction in RC compared to the wild-type control pseudovirus. However, when the W1201i Gag was substituted with wild-type Gag, a significant increase in RC, relative to the wild-type control, was observed. Thus, N37T↑V appears to catalyse the cleavage of wild-type Gag more effectively. We postulate that the selection of N37T↑V is dependent on the variations within Gag, and not vice versa since natural selection would favour a virus with an increased RC. The increase in RC was complemented by an additional small decrease in drug susceptibility to all three PIs.

In the context of its native Gag, a small, but statistically significant decrease in susceptibility to DRV was observed, while replication capacity was significantly impaired. However, in the presence of an unrelated wild-type Gag, a small but significant decrease in susceptibility was observed to all three PIs while replication capacity was significantly improved. Cumulatively, the data show that the W1201i Gag confers no advantage to RC and PI susceptibility. However, genotyping of Gag revealed notable polymorphisms and insertions.

The W1201i Gag displayed a PTAPP duplication and variations in the p2/NC cleavage site (S369N, S371N, I372L↑M, I373M and G377S). Gag insertion variants are usually co-selected with other polymorphisms to help maintain viral fitness [35]. However, these polymorphisms can accumulate in the absence of drug pressures [36]. It has been shown that nucleoside-based ART positively correlates with PTAP duplications in subtype B strains [36]. However, the PTAP duplication occurs at a higher frequency in subtype C variants [36,37]. Studies suggest that the PTAP motif influences the packaging of Pol proteins during late assembly, and may be involved in improving viral fitness [14,38]. However, a study performed by Martins *et al*., (2015) concluded that PTAP duplications had little effect on the viral infectivity in wild-type strains, but only led to increased viral infectivity, and decreased drug susceptibility, when modified by specific drug resistance mutations. Our results are in accordance with the findings of Martins *et al*., (2015), in that the W1201i Gag conferred no replicative advantage. In fact, W1201i Gag appears to be disadvantageous with regards to viral infectivity. Similarly, it has been reported that p2/NC cleave site mutations will accumulate under PI pressure [39]. It is currently believed that cleavage site mutations accumulate to restore lost viral fitness due to the development of mutations within protease [40]. Our results do not corroborate these findings because neither the PTAP duplication nor the cleavage site mutations were able to restore the RC of W1201i. However, we did not investigate the RC of W1201i in the absence of these polymorphisms.

Not all the insertions within W1201i Gag are thought to be advantageous. The W1201i isolate presented a previously unreported MSQAG insertion in proximity to the both the CA/p2 and p2/NC cleavage sites (Figure 3). Tamiya and colleagues suggested that insertions near the CA/p2 cleavage may compromise RC; however, the mechanism has not been established [14]. It is believed that insertions alter the conformation of the cleavage site and limit access to the protease enzyme [14]. It is reasonable to expect that insertions could impact RC, as the cleavage of Gag is thought to be controlled by the shape of the cleavage site and not a particular amino acid sequence [41]. Therefore, it is possible that the observed

reduction in the RC of W1201i is due to a deleterious effect conferred by the MSQAG insertion. Moreover, these data are in agreement with the low specific activity, catalytic turnover and catalytic efficiency displayed by the N37T↑V PR towards a substrate mimicking the CA/p2 cleavage site. If the N37T↑V protease displays lowered catalytic efficiency toward the substrate, then further structural changes such as an MSQAG insertion could possibly enhance the detrimental effect on specificity.

The fold change of susceptibility to LPV, ATV and DRV was analysed (Figure 4 A, B and C) because these drugs are the most commonly used ARVs in South Africa, with DRV reserved for third line ART. W1201i displayed a minimal, but statistically significant decrease in drug susceptibility to ATV, whereas the chimeric construct displayed a minimal but statistically significant decrease in drug susceptibility to LPV, ATV and DRV. Under normal conditions, protease insertion variants are usually fully susceptible to all PIs [34]. However, here the variant is modified by only three minor ATV resistance mutations, compared to the wild-type. Despite this, the N37T↑V protease displayed decreased susceptibility to both LPV and DRV. Considering the drug resistance profile of N37T↑V, it is possible that the observed reduction in susceptibility to LTV and DRV is due to the presence of the hinge region insertion and mutation.

Analysis of the kinetic properties of the N37T↑V protease indicates that the hinge region insertion and mutation impacts the activity of the enzyme. The specific activity, catalytic activity and catalytic turnover were decreased compared with the wild-type, as a result of the combined effects of the mutations present in the N37T↑V protease (Table 1). These data confirm that the catalytic processing power of the N37T↑V protease is less than the wild-type, at least concerning the fluorogenic substrate used. These results are to be expected as the fluorogenic substrate is a mimic of the wild-type Gag CA/p2 cleavage site, and insertion mutations within HIV-1 protease (particularly hinge region insertions) are known to decrease the rate of substrate processing [18].

Gag and protease appear to be inextricably linked with regards to PI resistance and viral fitness. A body of evidence is accumulating in support of the fact that Gag polymorphisms can modulate PI susceptibility, even in the absence of protease drug resistance mutations [23,42–44]. Despite the evidence, phenotypic testing is still focused predominantly on the protease sequence [44]. In this study, we have once again shown a clear link between protease and Gag based on phenotypic testing.

## Acknowledgements

The research reported in this publication was supported by the South African Medical Research Council under a Self-Initiated Research Grant to Yasien Sayed. The views and opinions expressed are those of the authors and do not necessarily represent the official views of the SAMRC. This work was supported by the University of the Witwatersrand, the South African National Research Foundation Grant 68898 (HWD), and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation Grant 64788 (HWD). AZT, DRV and LPV were supplied by the National Institutes of Health (NIH) AIDS Reagent and Reference Program. Plasmids p8.9 and pMDG were supplied by Didier Trono (École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland), and plasmid pCSFLW was supplied by Nigel Temperton (University College London, London, United Kingdom). We thank Johanna Ledwaba and Dr Gillian Hunt for identifying this rare protease sequence and Dr Jennifer Giandhari for help with Gag-Pro sequencing and providing the protocols used to test drug susceptibility and viral infectivity.

## References


[1] Joint United Nations Programme on HIV/AIDS (UNAIDS), Global AIDS update, Geneva, Switz. (2016).

[2] A. Wlodawer, M. Miller, M. Jaskólski, B.K. Sathyanarayana, E. Baldwin, I.T. Weber, L.M. Selk, L. Clawson, J. Schneider, S.B. Kent, Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease., Science. 245 (1989) 616–21. http://www.ncbi.nlm.nih.gov/pubmed/2548279.

[3] M.J. Todd, N. Semo, E. Freire, The structural stability of the HIV-1 protease., J. Mol. Biol. 283 (1998) 475–88. doi:10.1006/jmbi.1998.2090.

[4] S.M. Ahmed, H.G. Kruger, T. Govender, G.E.M. Maguire, Y. Sayed, M. a a Ibrahim, P. Naicker, M.E.S. Soliman, Comparison of the Molecular Dynamics and Calculated Binding Free Energies for Nine FDA-Approved HIV-1 PR Drugs Against Subtype B and C-SA HIV PR, Chem. Biol. Drug Des. 81 (2013) 208–218. doi:10.1111/cbdd.12063.

[5] Y. Cai, N.K. Yilmaz, W. Myint, R. Ishima, C.A. Schiffer, Differential Flap Dynamics in Wild-type and a Drug Resistant Variant of HIV-1 Protease Revealed by Molecular Dynamics and NMR Relaxation., J. Chem. Theory Comput. 8 (2012) 3452–3462. doi:10.1021/ct300076y.

[6] South African National Department of Health, National Consolidated Guidelines for the Prevention of Mother-To-Child Transmission of HIV (PMTCT) and the Managment of HIV in Children, Adolescents and Adults, Dep. Heal. Repub. South Africa. (2015) 1–128. www.doh.gov.za.

[7] G. Meintjes, J. Black, F. Conradie, V. Cox, S. Dlamini, J. Fabian, G. Maartens, T. Manzini, M. Mathe, C. Menezes, M. Moorhouse, Y. Moosa, J. Nash, C. Orrell, Y. Pakade, F. Venter, D. Wilson, S.A.H.I.V.C. Society, Adult antiretroviral therapy guidelines 2014, South. Afr. J. HIV Med. 15 (2014) 121–143. doi:10.7196/SAJHIVMED.1130.

[8] A. Ali, R.M. Bandaranayake, Y. Cai, N.M. King, M. Kolli, S. Mittal, J.F. Murzycki, M.N.L. Nalam, E. a Nalivaika, A. Özen, M.M. Prabu-Jeyabalan, K. Thayer, C. a Schiffer, Molecular Basis for Drug Resistance in HIV-1 Protease, Viruses. 2 (2010)

2509–2535. doi:10.3390/v2112509.

[9] R.K. Zeldin, R.A. Petruschke, Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients, J. Antimicrob. Chemother. 53 (2004) 4–9. doi:10.1093/jac/dkh029.

[10] J. Giandhari, A.E. Basson, A. Coovadia, L. Kuhn, E.J. Abrams, R. Strehlau, L. Morris, G.M. Hunt, Genetic Changes in HIV-1 Gag-Protease Associated with Protease Inhibitor-Based Therapy Failure in Pediatric Patients, AIDS Res. Hum. Retroviruses. 31 (2015) 776–782. doi:10.1089/aid.2014.0349.

[11] J.L. Martinez-Cajas, M.A. Wainberg, M. Oliveira, E.L. Asahchop, F. Doualla-Bell, I. Lisovsky, D. Moisi, E. Mendelson, Z. Grossman, B.G. Brenner, The role of polymorphisms at position 89 in the HIV-1 protease gene in the development of drug resistance to HIV-1 protease inhibitors., J. Antimicrob. Chemother. 67 (2012) 988–94. doi:10.1093/jac/dkr582.

[12] H. Gatanaga, Y. Suzuki, H. Tsang, K. Yoshimura, M.F. Kavlick, K. Nagashima, R.J. Gorelick, S. Mardy, C. Tang, M.F. Summers, H. Mitsuya, Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors., J. Biol. Chem. 277 (2002) 5952–61. doi:10.1074/jbc.M108005200.

[13] L. Myint, M. Matsuda, Z. Matsuda, Y. Yokomaku, T. Chiba, A. Okano, K. Yamada, W. Sugiura, Gag Non-Cleavage Site Mutations Contribute to Full Recovery of Viral Fitness in Protease Inhibitor-Resistant Human Immunodeficiency Virus Type 1, Antimicrob. Agents Chemother. 48 (2004) 444–452. doi:10.1128/AAC.48.2.444-452.2004.

[14] S. Tamiya, S. Mardy, M.F. Kavlick, K. Yoshimura, H. Mistuya, Amino Acid Insertions near Gag Cleavage Sites Restore the Otherwise Compromised Replication of Human Immunodeficiency Virus Type 1 Variants Resistant to Protease Inhibitors, J. Virol. 78 (2004) 12030–12040. doi:10.1128/JVI.78.21.12030-12040.2004.

[15] E.Y. Kim, M.A. Winters, R.M. Kagan, T.C. Merigan, Functional correlates of insertion mutations in the protease gene of human immunodeficiency virus type 1 isolates from patients., J. Virol. 75 (2001) 11227–33. doi:10.1128/JVI.75.22.11227-

11233.2001.

[16] A. Fun, A.M.J. Wensing, J. Verheyen, M. Nijhuis, Human Immunodeficiency Virus Gag and protease: partners in resistance., Retrovirology. 9 (2012) 63. doi:10.1186/1742-4690-9-63.

[17] L. Kuhn, G. Hunt, K.-G. Technau, A. Coovadia, J. Ledwaba, S. Pickerill, M. Penazzato, S. Bertagnolio, C.A. Mellins, V. Black, L. Morris, E.J. Abrams, Drug resistance among newly diagnosed HIV-infected children in the era of more efficacious antiretroviral prophylaxis, AIDS. 28 (2014) 1673–1678. doi:10.1097/QAD.0000000000000261.

[18] M. Kozisek, K.G. Saskova, P. Rezacova, J. Brynda, N.M. van Maarseveen, D. De Jong, C.A. Boucher, R.M. Kagan, M. Nijhuis, J. Konvalinka, Ninety-Nine Is Not Enough: Molecular Characterization of Inhibitor-Resistant Human Immunodeficiency Virus Type 1 Protease Mutants with Insertions in the Flap Region, J. Virol. 82 (2008) 5869–5878. doi:10.1128/JVI.02325-07.

[19] P. Naicker, I. Achilonu, S. Fanucchi, M. Fernandes, M.A.A. a Ibrahim, H.W. Dirr, M.E.S.S. Soliman, Y. Sayed, Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance., J. Biomol. Struct. Dyn. 31 (2013) 1370–1380. doi:10.1080/07391102.2012.736774.

[20] J. Zondagh, V. Balakrishnan, I. Achilonu, H.W. Dirr, Y. Sayed, Molecular dynamics and ligand docking of a hinge region variant of South African HIV-1 subtype C protease, J. Mol. Graph. Model. 82 (2018) 1–11. doi:10.1016/j.jmgm.2018.03.006.

[21] J.E. Foulkes-Murzycki, W.R.P. Scott, C.A. Schiffer, Hydrophobic Sliding: A Possible Mechanism for Drug Resistance in Human Immunodeficiency Virus Type 1 Protease, Structure. 15 (2007) 225–233. doi:10.1016/j.str.2007.01.006.

[22] T. Ndung'u, B. Renjifo, M. Essex, Construction and Analysis of an Infectious Human Immunodeficiency Virus Type 1 Subtype C Molecular Clone, J. Virol. 75 (2001) 4964–4972. doi:10.1128/JVI.75.11.4964-4972.2001.

[23] J. Giandhari, A.E. Basson, K. Sutherland, C.M. Parry, P.A. Cane, A. Coovadia, L. Kuhn, G. Hunt, L. Morris, Contribution of Gag and Protease to HIV-1 Phenotypic Drug Resistance in Paediatric Patients Failing Protease-Inhibitor Based Therapy,

Antimicrob. Agents Chemother. 60 (2016) 2248–56. doi:10.1128/AAC.02682-15.

[24] C.M. Parry, A. Kohli, C.J. Boinett, G.J. Towers, A.L. McCormick, D. Pillay, Gag Determinants of Fitness and Drug Susceptibility in Protease Inhibitor-Resistant Human Immunodeficiency Virus Type 1, J. Virol. 83 (2009) 9094–9101. doi:10.1128/JVI.02356-08.

[25] J. Zondagh, A. Williams, I. Achilonu, H.W. Dirr, Y. Sayed, Overexpression, purification and functional characterisation of wild-type HIV-1 subtype C protease and two variants using a thioredoxin and his-tag protein fusion system (submitted), Protein J. (2018).

[26] Z. Szeltner, L. Polgár, Rate-determining Steps in HIV-1 Protease Catalysis, J. Biol. Chem. 271 (1996) 32180–32184. doi:10.1074/jbc.271.50.32180.

[27] A. Carmel, A. Yaron, An intramolecularly quenched fluorescent tripeptide as a fluorogenic substrate of angiotensin-I-converting enzyme and of bacterial dipeptidyl carboxypeptidase., Eur. J. Biochem. 87 (1978) 265–73. doi:10.1111/j.1432-1033.1978.tb12375.x.

[28] A.M. Wensing, V. Calvez, H.F. Günthard, V.A. Johnson, R. Paredes, D. Pillay, R.W. Shafer, D.D. Richman, 2017 Update of the Drug Resistance Mutations in HIV-1., Top. Antivir. Med. 24 (2017) 132–133. http://www.ncbi.nlm.nih.gov/pubmed/28208121.

[29] S. Ibe, N. Shibata, M. Utsumi, T. Kaneda, Selection of human immunodeficiency virus type 1 variants with an insertion mutation in the p6(gag) and p6(pol) genes under highly active antiretroviral therapy., Microbiol. Immunol. 47 (2003) 71–9. http://www.ncbi.nlm.nih.gov/pubmed/12636256.

[30] A.N. Martins, A.A. Waheed, S.D. Ablan, W. Huang, A. Newton, C.J. Petropoulos, R.D.M. Brindeiro, E.O. Freed, Elucidation of the Molecular Mechanism Driving Duplication of the HIV-1 PTAP Late Domain., J. Virol. 90 (2015) 768–79. doi:10.1128/JVI.01640-15.

[31] J. Pereira-Vaz, V. Duque, L. Trindade, J. Saraiva-da-Cunha, A. Meliço-Silvestre, Detection of the protease codon 35 amino acid insertion in sequences from treatment-naïve HIV-1 subtype C infected individuals in the Central Region of Portugal, J. Clin. Virol. 46 (2009) 169–172. doi:10.1016/j.jcv.2009.06.019.

[32] M. Stürmer, S. Staszewski, H.W. Doerr, K. Hertogs, A 6-Base Pair Insertion in the Protease Gene of HIV Type 1 Detected in a Protease Inhibitor-Naive Patient Is Not Associated with Indinavir Treatment Failure, AIDS Res. Hum. Retroviruses. 19 (2003) 967–968. doi:10.1089/088922203322588314.

[33] S. Paolucci, F. Baldanti, L. Dossena, G. Gerna, Amino acid insertions at position 35 of HIV-1 protease interfere with virus replication without modifying antiviral drug susceptibility, Antiviral Res. 69 (2006) 181–185. doi:10.1016/j.antiviral.2005.12.005.

[34] M.A. Winters, T.C. Merigan, Insertions in the human immunodeficiency virus type 1 protease and reverse transcriptase genes: clinical impact and molecular mechanisms., Antimicrob. Agents Chemother. 49 (2005) 2575–82. doi:10.1128/AAC.49.7.2575-2582.2005.

[35] F. Bally, R. Martinez, S. Peters, P. Sudre, A. Telenti, Polymorphism of HIV Type 1 Gag p7/p1 and p1/p6 Cleavage Sites: Clinical Significance and Implications for Resistance to Protease Inhibitors, AIDS Res. Hum. Retroviruses. 16 (2000) 1209–1213. doi:10.1089/08892220050116970.

[36] S. Sharma, S.G. Aralaguppe, M. Abrahams, C. Williamson, C. Gray, P. Balakrishnan, S. Saravanan, K.G. Murugavel, S. Solomon, U. Ranga, The PTAP sequence duplication in HIV-1 subtype C Gag p6 in drug-naive subjects of India and South Africa, BMC Infect. Dis. 17 (2017) 95. doi:10.1186/s12879-017-2184-4.

[37] T. Flys, N. Marlowe, J. Hackett, N. Parkin, M. Schumaker, V. Holzmayer, P. Hay, S.H. Eshleman, Analysis of PTAP duplications in the gag p6 region of subtype C HIV type 1., AIDS Res. Hum. Retroviruses. 21 (2005) 739–41. doi:10.1089/aid.2005.21.739.

[38] M.F. Maguire, R. Guinea, P. Griffin, S. Macmanus, R.C. Elston, J. Wolfram, N. Richards, M.H. Hanlon, D.J.T. Porter, T. Wrin, N. Parkin, M. Tisdale, E. Furfine, C. Petropoulos, B.W. Snowden, J.-P. Kleim, Changes in Human Immunodeficiency Virus Type 1 Gag at Positions L449 and P453 Are Linked to I50V Protease Mutants In Vivo and Cause Reduction of Sensitivity to Amprenavir and Improved Viral Fitness In Vitro, J. Virol. 76 (2002) 7398–7406. doi:10.1128/JVI.76.15.7398-7406.2002.

[39] H.C. Côté, Z.L. Brumme, P.R. Harrigan, Human immunodeficiency virus type 1

protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir., J. Virol. 75 (2001) 589–94. doi:10.1128/JVI.75.2.589-594.2001.

[40] G. Teto, C.T. Tagny, D. Mbanya, J.Y. Fonsah, J. Fokam, E. Nchindap, L. Kenmogne, A.K. Njamnshi, G.D. Kanmogne, Gag P2/NC and pol genetic diversity, polymorphism, and drug resistance mutations in HIV-1 CRF02-AG- and non-CRF02-AG-infected patients in Yaoundé, Cameroon, Sci. Rep. 7 (2017) 1–14. doi:10.1038/s41598-017-14095-4.

[41] M. Prabu-Jeyabalan, E. Nalivaika, C.A. Schiffer, Substrate shape determines specificity of recognition for HIV-1 protease: Analysis of crystal structures of six substrate complexes, Structure. 10 (2002) 369–381. doi:10.1016/S0969-2126(02)00720-7.

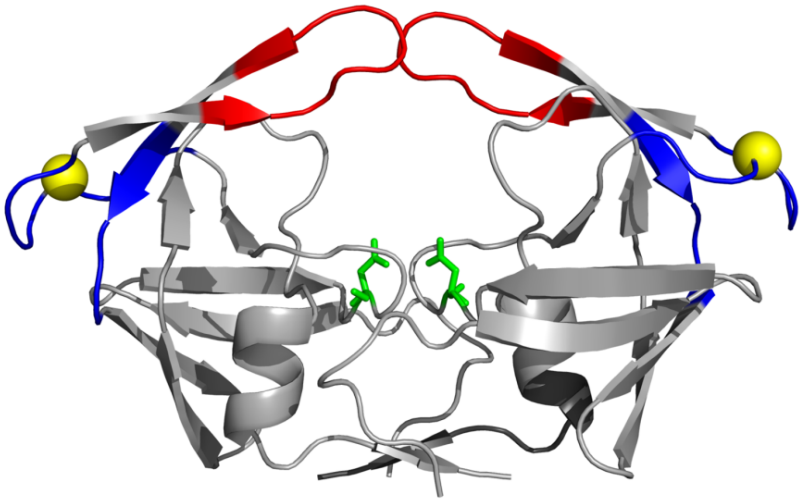
[42] M. Nijhuis, N.M. Van Maarseveen, S. Lastere, P. Schipper, E. Coakley, B. Glass, M. Rovenska, D. De Jong, C. Chappey, I.W. Goedegebuure, G. Heilek-Snyder, D. Dulude, N. Cammack, L. Brakier-Gingras, J. Konvalinka, N. Parkin, H.G. Kräusslich, F. Brun-Vezinet, C.A.B. Boucher, A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism, PLoS Med. 4 (2007) 0152–0163. doi:10.1371/journal.pmed.0040036.

[43] C.M. Parry, A. Kohli, C.J. Boinett, G.J. Towers, A.L. McCormick, D. Pillay, Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1., J. Virol. 83 (2009) 9094–9101. doi:10.1128/JVI.02356-08.

[44] R.K. Gupta, A. Kohli, A.L. McCormick, G.J. Towers, D. Pillay, C.M. Parry, Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays., AIDS. 24 (2010) 1651–5. http://www.ncbi.nlm.nih.gov/pubmed/20597164.

[45] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega., Mol. Syst. Biol. 7 (2011) 539. doi:10.1038/msb.2011.75.

A



B

```
Wild-type           PQITLWKRPLVSIKVGGQIKEALLDTGADDTVLEEIN-LPGKWKPKMIGGIGGFIKVRQY  59
N37T↑V              PQITLWKRPLVSVKVEGQIKEALLDTGADDTVLEETTVLSGKWKPKMIGGIGGFIKVRQY  60
MJ4                 PQITLWQRPLVSIKVGGQIKEALLDTGADDTVLEEMS-LSGKWKPKMIGGIGGFIKVRQY  59
Resistance_control  PQITLWQRPIVSIKVGGQIREALLDTGADDTVLEDIN-LPGKWKPKIIGGIGGFVKVRQY  59
                    ••••••:••:••:•• •••:•••••••••••••••:   • ••••••:•••••••:•••••

Wild-type           DQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF  99
N37T↑V              EEVPIEICGKKAIGTVLIGPTPVNIIGRNLLTQLGCTLNF  100
MJ4                 DQIPIEICGKKAIGTVLIGPTPVNIIGRNMLTQLGCTLNF  99
Resistance_control  DHVPIEICGKKAIGSVLVGPTPANIIGRNMLTQLGCTLNF  99
                    :.: ••••••••••:••:••••.••••••:••••••••••
```

**Figure 1:** (A) Homology model of the N37T↑V protease. The relative positions of the amino acid insertions (N37T↑V) are indicated by yellow spheres. The hinge regions (residues 35-42 and 57-61) are denoted in blue, and the flap regions (residues 46-54) are denoted in red. The homology model was generated with the molecular visualisation software programme PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.), using data from the Protein Data Bank (PDB ID: 3U71) [19]. (B) The sequence alignment data shows the positions of the mutations of both N37T↑V and the resistance control protease. The N37T↑V variant has the following mutations: Q7K, I13V, G16E, M36T, D60E, Q61E, I62V and M89L. The resistance control protease has the following mutations: L10I, K20R, E35D, S39P, M46I, I54V, Q61H, I62V, T74S, I77V and V82A. The wild-type subtype C protease consensus and MJ4 sequences are included as a reference. The alignment was performed using the Clustal Omega online tool (EMBL-EBI) [45].

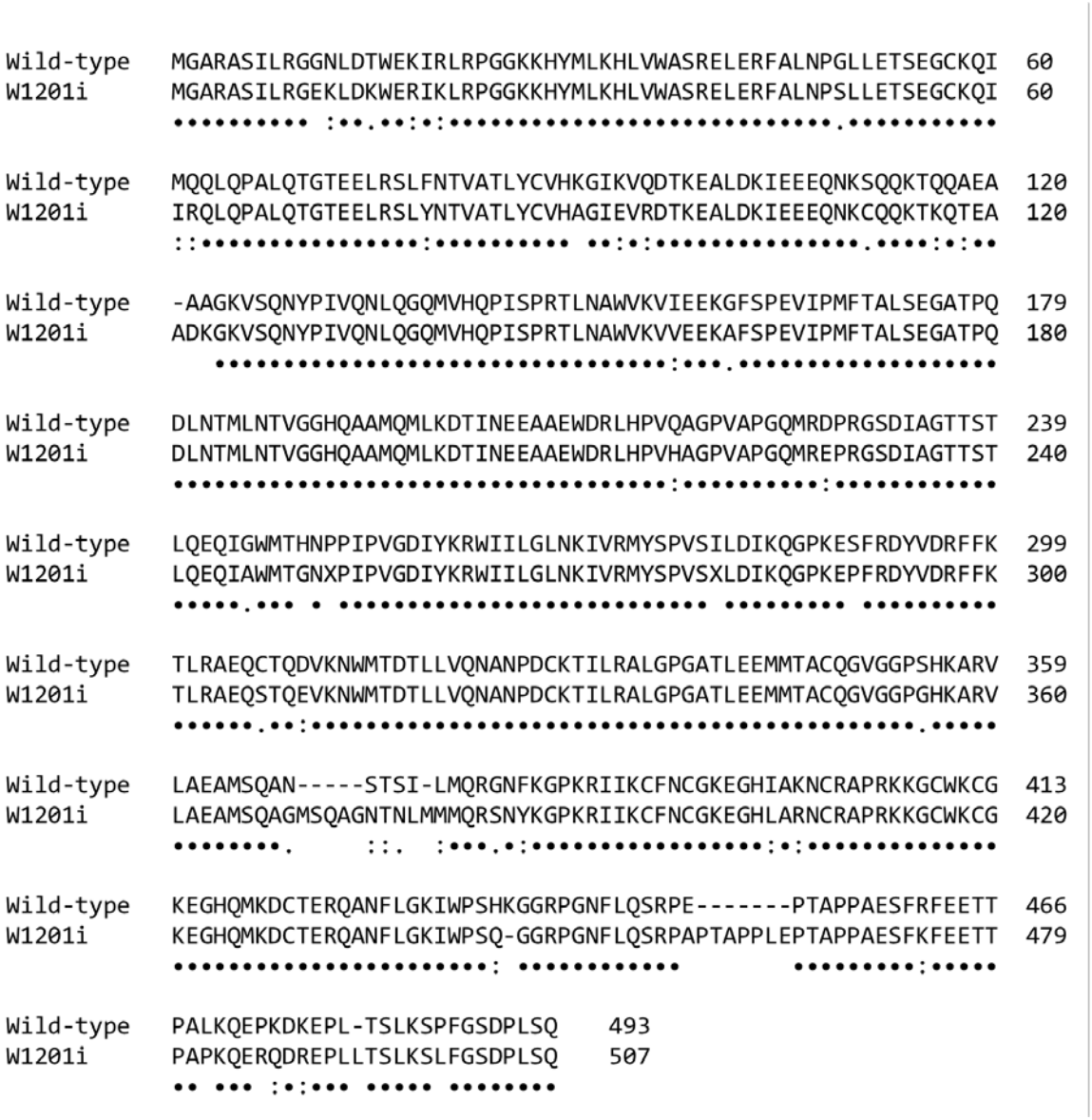



**Figure 2:** Alignment of the consensus wild-type Gag sequence and W1201i-Gag. Sequence alignment was performed on the Clustal Omega online tool (EMBL-EBI) [45].

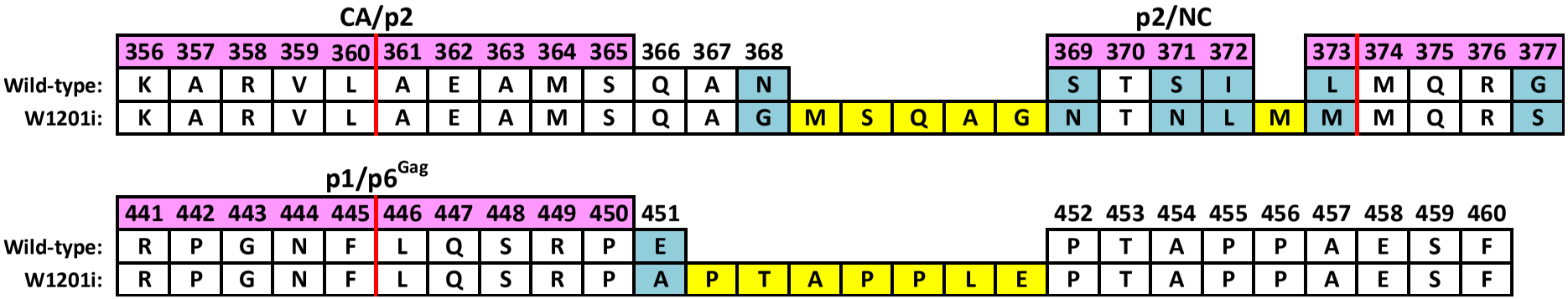



**Figure 3:** Alignment of notable regions within Gag of wild-type reference and W1201i. Amino acid insertions are shown in yellow, and blue depicts the amino acid polymorphisms. Cleavage sites are indicated above the sequence in purple.

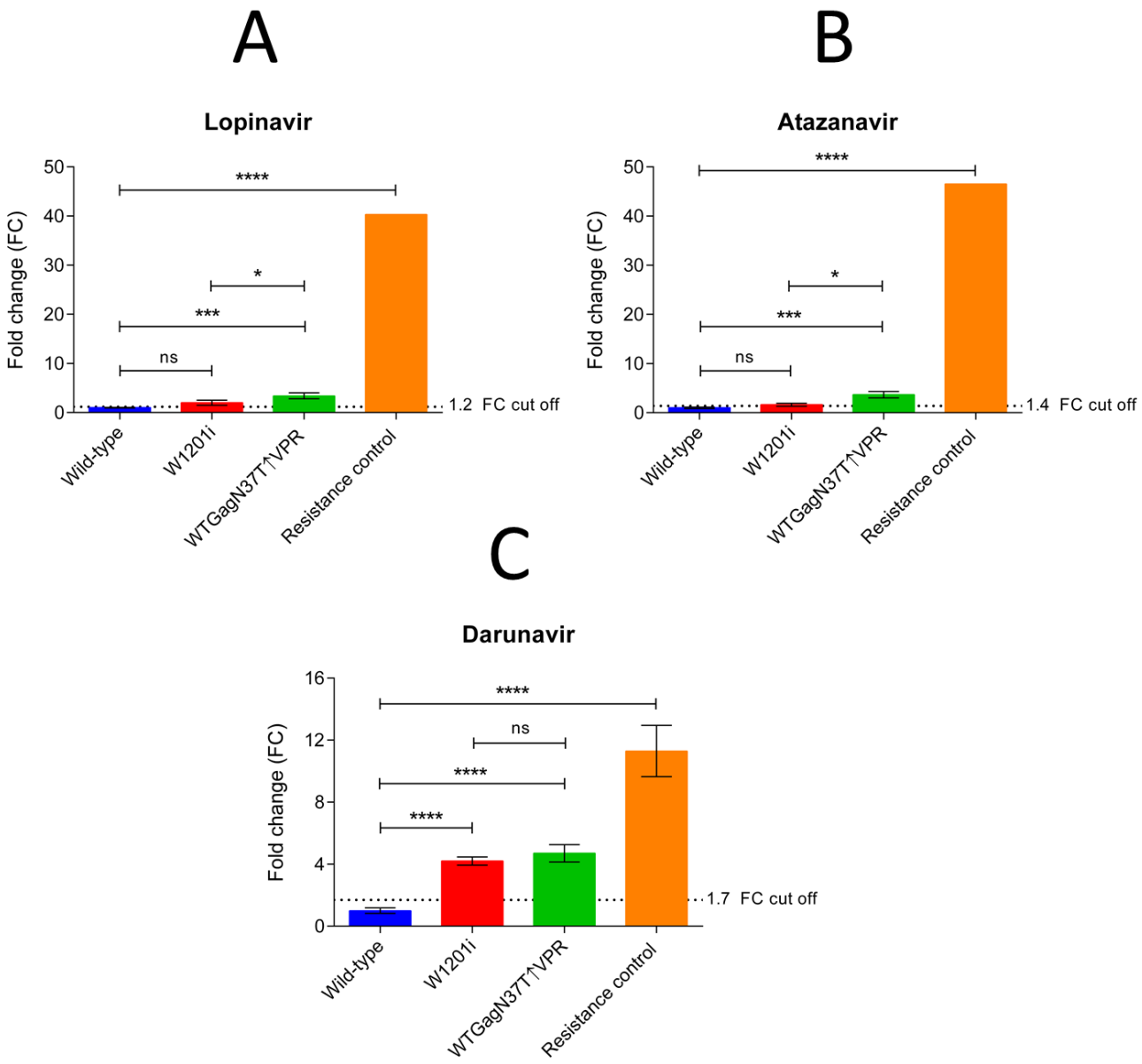



**Figure 4:** The phenotypic drug susceptibility of four pseudoviruses. The *in vitro* susceptibilities of the W120i, WTGagN37TVPR and resistance control are shown relative to the wild-type control for (A) LPV, (B) ATV and (C) DRV. The error bars indicate the standard error of the mean (SEM). The horizontal bars represent the statistical significance (P value). The lower biological cut-off values are shown for each drug.

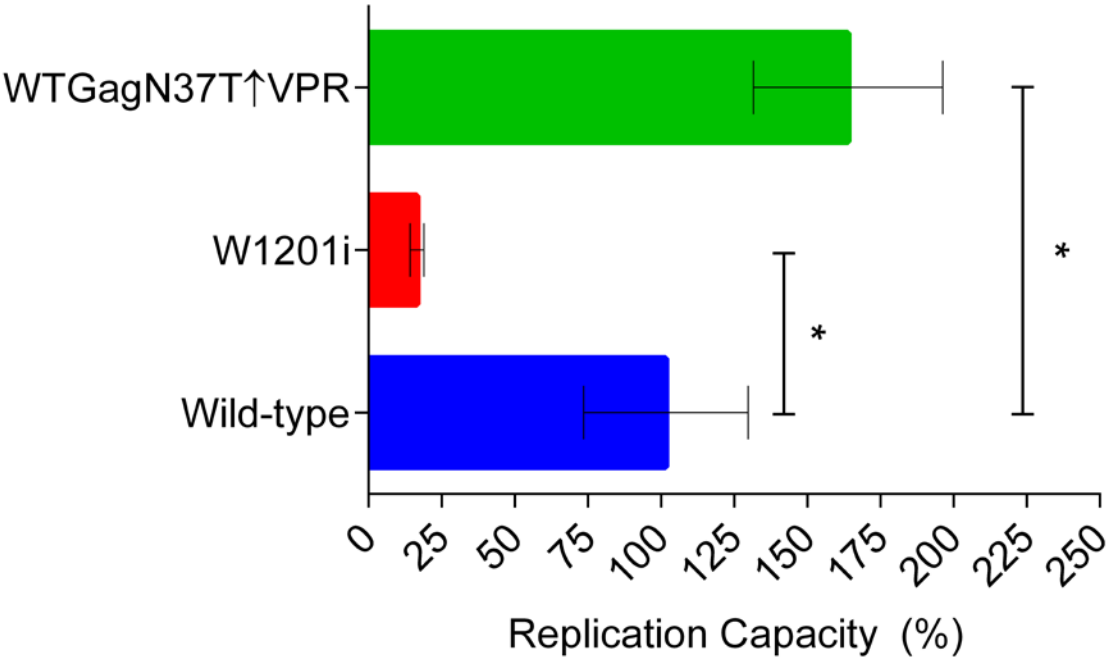



**Figure 5:** Viral infectivity of the isolates/constructs. The replication capacity of W1201i and WTGagN37T↑VPR are shown relative to the wild-type control. The error bars indicate the standard error of the mean (SEM). The horizontal bars represent the statistical significance (P value)

**Table 1:** Enzymatic parameters of the wild-type subtype C and N37T↑V proteases.

| | Specific Activity ($\mu mol.min^{-1}.mg^{-1}$) | Catalytic efficiency ($s^{-1}.\mu M^{-1}$) | Catalytic constant ($s^{-1}$) |
|---|---|---|---|
| Wild-type | 12 ± 1 | 2.4 ± 0.1 | 6.4 ± 0.2 |
| N37T↑V | 1.8 ± 0.2 | 0.5 ± 0.1 | 1.5 ± 0.2 |

# CHAPTER 4

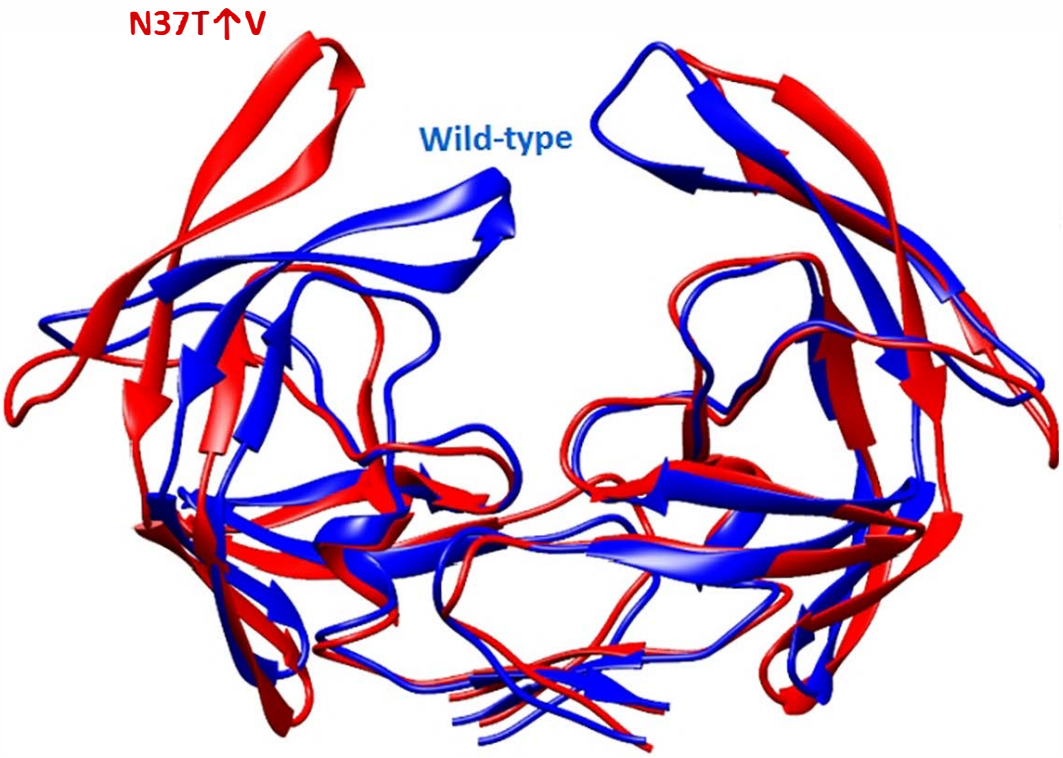
## Molecular dynamics and ligand docking of a hinge region variant of South African HIV-1 subtype C protease

Jake Zondagh, Vijayakumar Balakrishnan, Ikechukwu Achilonu, Heini W. Dirr and Yasien Sayed.

*J. Mol. Graph. Model.* 82 (2018) 1–11. doi:10.1016/j.jmgm.2018.03.006.

In this manuscript, the molecular dynamics and drug binding characteristics of the N37T↑V protease were examined computationally. It was found that the N37T↑V variant displayed altered dynamics around the hinge and fap regions. The altered dynamics of the variant resulted in higher binding free energies when docked to three commonly used protease inhibitors.

Author contributions: Jake Zondagh performed all the Molecular dynamics simulations, analysed the data and wrote the manuscript. Vijayakumar Balakrishnan assisted with induced fit docking experiments. Yasien Sayed, Ikechukwu Achilonu and Heini Dirr and assisted in manuscript revision. Yasien Sayed supervised the project and assisted in data analysis and interpretation.




Graphical abstract

# Molecular dynamics and ligand docking of a hinge region variant of South African HIV-1 subtype C protease

Jake Zondagh, Vijayakumar Balakrishnan, Ikechukwu Achilonu, Heini W. Dirr, Yasien Sayed*

Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2050, South Africa

*Corresponding author: Yasien Sayed; e-mail address: Yasien.Sayed@wits.ac.za.

## Abstract

HIV-1 protease is an important antiretroviral drug target due to its key role in viral maturation. Computational models have been successfully used in the past to understand the dynamics of HIV-1 protease variants. We performed molecular dynamics simulations and induced fit docking on a wild-type South African HIV-1 subtype C protease and an N37T↑V hinge region variant. The simulations were initiated in a cubic cell universe and run in explicit solvent, with the wild-type and variant proteases in the fully closed conformation and under periodic boundary conditions. The trajectory for each simulation totalled 20 nanoseconds. The results indicate that the N37T↑V hinge region mutation and insertion alter the molecular dynamics of the flap and hinge regions when compared to the wild-type protease. Specifically, the destabilisation of the hinge region allowed a larger and protracted opening of the flap region due to the formation of two key hinge/cantilever salt-bridges, which are absent in the wild-type protease. Domain-domain anti-correlation was observed between the flap and hinge region for both models. However, the N37T↑V variant protease displayed a lower degree of anti-correlation. The mutations affected the thermodynamic landscape of inhibitor binding as there were fewer observable chemical contacts between the N37T↑V variant protease and lopinavir, atazanavir and darunavir, respectively. These data elucidate the biophysical basis for the selection of hinge region insertion mutations by the HI virus.

**Keywords**: HIV-1; Subtype C; protease; flap dynamics; hinge region; ligand docking; molecular dynamics; induced fit.

## Abbreviations

| | |
|---|---|
| ART: | Antiretroviral Treatment |
| ATV: | Atazanavir |
| $C_\alpha$ | Alpha Carbon |
| DRV: | Darunavir |
| FDA: | Food and Drug Administration |
| HIV-1: | Human Immunodeficiency Virus type 1 |
| IFD: | Induced Fit Docking |
| LPV: | Lopinavir |
| MD: | Molecular Dynamics |
| ns: | Nanoseconds |
| N37T↑V: | HIV-1 subtype C protease containing asparagine 37 mutated to threonine; the upwards arrow indicates an insertion of valine at position 37 |
| PDB: | Protein Data Bank |
| PI: | Protease Inhibitors |
| PR: | Protease |
| $R_g$: | Radius of Gyration |
| RMSD: | Root Mean Square Deviation |
| RSMF: | Root Mean Square Fluctuation |
| SASA: | Solvent Accessible Surface Area |

# 1. Introduction

HIV continues to be a socioeconomic problem in sub-Saharan Africa. In South Africa alone, more than six million people are HIV positive [1]. The virus encodes several proteins and enzymes, one of which is a homodimeric aspartyl protease that is directly involved in viral maturation [2]. The function of HIV-1 protease is the cleavage and subsequent activation of the viral Gag and Gag-Pol polyproteins [3,4]. Thus, it remains an attractive target for rational drug design [5,6].

Currently, there are nine United States Food and Drug Administration (FDA) approved protease inhibitors (PIs) [5]. The PIs are mainly developed and tested in countries affected by HIV-1 subtype B and are known to be less effective against some subtype C variants [7]. Numerous subtype C variants modulate drug susceptibility via a reduction in the binding affinity to the PIs. An investigation into PI/protease interaction is required to understand the biochemical basis underlying PI susceptibility [8].

The flap, hinge, cantilever and fulcrum regions make up the four domains of an HIV-1 protease monomer. The hinge region is composed of amino acid residues 35-42 and 57-61 (figure 1A). Residues 46-54 form the flap region and are integral to the specificity and activity of HIV-1 protease [9]. The fulcrum region consists of residues 10-23 (figure 1A). According to Gustchina and Weber (1990), the hinge region regulates the stability and movement of the flap region, which in turn controls access to the substrate binding site. Flap region dynamics affect the binding energetics of both PIs and natural substrates and are responsible for coordinating a single water molecule required for proteolytic cleavage [11]. Liu et al., (2006) suggests that increased flap region flexibility can markedly reduce the rate of substrate proteolysis and drug susceptibility. It is, therefore, important to understand how hinge region mutations, and more specifically the insertion mutations, modulate PI-susceptibility.

Recently, South African HIV-1 subtype C protease (HIV-1 C-SA) sequence data were obtained from a South African PI-naïve infant (Professor Lynn Morris, AIDS Virus Research Unit, NICD, Johannesburg, South Africa) [13]. The sequence data confirmed a hinge region mutation and insertion (N37T↑V) as well as several background mutations: I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L (figure 1B and 1C) [14]. As a result, each monomeric subunit is comprised of 100 amino acids instead of the conventional 99

amino acids in the canonical wild-type sequence [15]. In this paper, the variant is referred to as N37T↑V, as the substitution and insertion mutations are both present at residue 37. The N37T↑V nomenclature was developed by Yasien Sayed and Ikechukwu Achilonu (Protein Structure-Function Research Unit, University of the Witwatersrand, South Africa) because no convenient taxonomy for describing such mutations currently exists.

Previous studies have compared the flap dynamics of subtype C and subtype B. However, mutations within subtype C need to be explored, particularly if they can potentially confer reduced drug susceptibility [9]. Several molecular dynamics (MD) simulation studies have shown the spontaneous opening of the flap region during specific solvation states [16]. However, PI bound structures are only observed in the closed conformation as opposed to the semi-open conformation of the apo-enzyme [17].

Recent advances in computational biology allow detailed insight into molecular interactions, particularly those controlled by dynamics. A well-known example of this is the visualisation of the spontaneous opening of the protease flaps, which was proven through NMR studies and reproduced by MD simulations [18]. Accurate homology models need to be created to study proteins computationally. Homology models are created from data derived from previously solved crystal structures of homologous proteins. Currently, the Protein Data Bank (PDB: www.rcsb.org) contains numerous solved X-ray crystal structures of HIV-1 protease, including structures that are ligated with the nine FDA-approved protease inhibitors (PIs) [19,20]. In this study, the behaviour of the hinge region was assessed because it plays a significant role in flap region dynamics and drug susceptibility [21]. Our findings may assist in shaping the development of more selective protease inhibitors.

## 2. Materials and methods

### 2.1 Computational data acquisition

Homology models were constructed from the coordinates of the wild-type South African HIV-1 subtype C protease, resolved at a resolution of 2.72 Å (PDB ID: 3u71) [9,20]. This reference structure was chosen because the South African subtype C protease sequence has the closest homology to the N37T↑V clinical variant. Figure 1B depicts the homology model of N37T↑V, where the relative locations of the polymorphisms are indicated. Sequence alignment was performed on the Clustal Omega online tool [22]. Homology models were generated using the SWISS-model online tool [23,24]. Models were validated by PROCHECK [25,26].

The wild-type and variant (N37T↑V) system each contained the Q7K substitution that is used in our laboratory to decrease autoproteolysis. The Q7K mutation is present in many experimental protease constructs, particularly in structural, kinetic, and binding affinity studies because it does not affect the kinetic properties of the protease [27]. In addition to the N37T↑V mutations, the variant enzyme has the following polymorphisms: I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L. Individually, these polymorphisms represent background mutations as confirmed on the Stanford HIV database (http://hivdb. stanford.edu/).

MD simulation studies were performed on Groningen Machine For Chemical Simulations (GROMACS) version 5.0 [28], running on two 3.68 GHz Intel core i7 5960x computers implemented on LINUX architecture. The AMBER99sb force field was used for both models [29,30]. Simulations were conducted with explicit solvent in a cubic box universe. Long-range electrostatics was handled by the particle-mesh Ewald (PME) method [31]. Wild-type and variant systems were neutralised by the addition of 8 (charge -8) and 4 (charge -4) chloride ions, respectively.

The solvated systems were relaxed with energy minimisation using the steepest descent method. Minimisation was continued until the systems reached a maximum atomic force of 1000 kJ/mol/nm. Following minimisation, 5 ns of MD simulation was performed under the NPT ensemble (constant number of particles (N), constant pressure (P) and constant temperature (T)). Afterwards, the temperatures were linearly increased from 10 to 300 degrees kelvin. Positional restraints were also calculated, and the original values annealed to

relax the models and ensure a stable temperature. Next, the temperature-stable models were subjected to 5 ns of simulation under the NVT ensemble (constant number of particles (N), constant volume (V) and constant temperature (T)), where the pressure of each system was equilibrated. These trajectories were used in subsequent data acquisition. After equilibration, MD simulations were performed for 20 ns under a constant temperature of 300 K with a Berendsen thermostat, and an average pressure of 1 atm maintained by the Parrinello–Rahman barostat algorithm [32].

2.2 Data analysis

The resultant trajectories were analysed by both the script-based utilities in GROMACS version 5.0 and with the molecular visualisation software UCSF Chimera version 1.11.2 [28,33,34]. Models were viewed with the molecular visualisation software PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). Sequence alignment was performed using the Clustal Omega tool (EMBL-EBI) [22]. Principal component analyses were performed using the Bio3D package in the language and environment for statistical computing known as R [35–37]. Figures were compiled on GraphPad Prism 6 (GraphPad Software, Inc. La Jolla, CA, USA).

2.3 Ligand docking

Glide software (Schrödinger LLC 2009, USA) was used to perform the induced fit docking (IFD) and calculate the binding free energies. The docking simulations were executed on a CentOS EL-5 workstation. Graphical visualisation was achieved using the PyMOL software (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

The wild-type subtype C reference structure (PDB ID: 3u71) was downloaded from the PDB [9]. The homology models were generated with the SWISS-model online tool [23,24]. Models were validated by PROCHECK [25,26]. Schrödinger modules Glide, Prime, QSite, Liaison and MacroModel were used for protein preparation. Each model was modified by correcting bond orders while ionizable residues were assigned a charge corresponding to a solution at pH 5.0 (experimentally determined optimal pH). The models were subject to energy minimisation until the average RMSD reached 0.3 Å.

The Schrödinger suite contains various tools for ligand preparation. The LigPrep tool was used to convert 2D drug structures to 3D structures and to add hydrogens, correct bond lengths and bond angles, correct chirality and to perform energy minimisation. The Epik tool

was used to choose the lowest energy tautomers and ring structures. The energy-minimised protein models were loaded into the workspace, and a PI was selected and specified to the active site of each. IFD calculations were carried out for lopinavir (LPV), atazanavir (ATV) and darunavir (DRV). Twenty conformational poses (system default setting) were calculated for each drug. The best conformations were chosen for further evaluation based on their respective IFD scores and their similarity to known PI bound HIV-1 protease crystallographic structures (PDB: www.rcsb.org).

## 3. Results and discussion

MD simulations have advanced sufficiently to model macromolecular structure-to-function relationships [38]. Our goal was to understand how an HIV-1 protease hinge region insertion and mutation affects the behaviour of the protease by using the N37T↑V variant as a representative model. Solvated MD simulations were performed on the apo-structures of both the wild-type protease and the clinically relevant N37T↑V hinge region variant. The simulations were initiated with each structure in a closed, energy-minimised conformation. The starting structures were aligned with the reference crystallographic structure (PDB ID: 3u71). The alignment RMSD for the wild-type and N37T↑V models were 0.4 Å and 0.41 Å, respectively. Simulations were conducted for 20 ns to observe the complete opening of the flaps and their subsequent return to the closed conformation. The closed (< 7 Å), semi-open (7-12 Å) and fully-open (>12 Å) conformations are defined based on the distance between the alpha carbon ($C_\alpha$) of the two flap tip residues (Gly 51/52 to Gly-51′/52′). Representative snapshots of the simulations were taken to show the various conformations of the proteases over time (figure 2A). Both proteases sampled the closed, semi-open and fully-open conformations that are known to occur in HIV proteases. The switching of flap handedness is known to occur during MD simulations. Figure 2B shows the switching of handedness in both models. Flap switching occurred at and 14.98 ns in the wild-type simulation and 11.45 ns in the variant simulation. The switching only occurred when the protease flaps were in the semi-open conformation. This observation has been previously reported by others as well [9,17].

### 3.1 RMSD and RMSF

The RMSD infers the overall dynamic motion concerning the $C_\alpha$ backbone structure. The data show that the RMSD variation of N37T↑V was much greater than that of the wild-type

protease (figure 3A). The wild-type exhibited a relatively low RMSD throughout the simulation, with values remaining well below 3 Å. The highest RMSD reached by the wild-type model was 2.85 Å at approximately 13.12 ns. Conversely, the N37T↑V model exhibited a higher RMSD overall. The highest RMSD reached by the variant model was 4 Å at approximately 14.05 ns. Both simulations follow a similar trajectory until approximately 12 ns at which point the variant protease samples a larger dynamic field (figure 3A, 12-19 ns). The data highlights that the dynamics of the wild-type trajectory differ considerably from that of the N37T↑V trajectory.

The root mean square fluctuation (RMSF) was calculated to understand which residues were responsible for the altered dynamics. The RMSF indicates the specific motion of each amino acid throughout the MD simulation. Figure 3B shows the residue-based RMSF of the wild-type and N37T↑V variant over a 20 ns trajectory. Regions of high mobility were defined as regions that fluctuated more than the average RMSF of the wild-type protease (~1.5 Å). Residues 34-60 (34′-60′) exhibit high mobility. These residues illustrate that the N37T↑V variant has enhanced flap and hinge region dynamics when compared to the wild-type enzyme. Similarly, residues 78-85 (78′-85′) and residues 16-24 (16′-24′) show prominent fluctuations. These two groups of residues correspond to a random coil and a portion of the fulcrum region, respectively.

3.2 Radius of gyration and solvent accessibility

The relative radius of gyration ($R_g$) and differences in the global solvent accessible surface area (SASA) were calculated to assess the degree of compactness of each protease. Figure 4A shows that the highest $R_g$ attained by the wild-type model was 1.86 nm at 13.12 ns. The $R_g$ of the variant model exceeded 1.95 nm at 13.39 ns. Relative to the wild-type, the variant displayed a noticeably larger $R_g$ from 3 to 9 ns and from 12 to 20 ns. The $R_g$ data indicate that the N37T↑V is relatively unstable compared to the wild-type model because the wild-type adopts a more constrained conformation for the duration of the simulation. Therefore, the variant has fewer interactions between adjacent amino acids and the stability the flap and hinge domains are particularly affected [39].

Figure 4B shows that the global SASA of the variant is greater than the wild-type throughout the simulations. The average SASA for the wild-type is 110 $nm^2$, whereas the average SASA for N37T↑V is 116.75 $nm^2$. The higher degree of solvent accessibility infers that the variant protease adopts a more open conformation than the wild-type protease. The increase in the

SASA of the N37T↑V protease further confirms that the variant protease is more dynamic relative to the wild-type protease. These findings may assist us in understanding the biochemical basis for an altered kinetic profile of the N37T↑V variant (manuscript in preparation).

### 3.3 Essential dynamics

The essential dynamics refer to the correlated and anti-correlated motions between different domains within a protein and were examined to understand which motions are related and to identify the pattern of movement for each trajectory [40]. The essential dynamics analyses were restricted to the backbone $C_\alpha$ atoms. The essential dynamics within a model are usually due to only a few eigenvectors with large eigenvalues [41]. The eigenvectors describe the degree of atomic fluctuation of each model. The highest eigenvalues correspond to the most relevant dynamics of the molecule, represented as dynamics cross-correlation maps (figure 5). Analysis of these maps revealed that the greatest degree of flexibility occurs around the flap region for both the wild-type and the variant protease (figure 5). Usually, the opening of the flap region is anti-correlated (directionally opposing) with the hinge region trajectory for wild-type HIV-1 protease [42]. Although some anti-correlation is observed in the variant model, the opening of the variant flaps is less reliant on the downward trajectory of the hinge region.

### 3.4 Flap opening and curling

The trajectory data show that the MD of the N37T↑V protease is altered. Figure 6 demonstrates the differences in the MD profiles between the two models. The hinge region stability was determined by measuring the relative distance between Gly-16 (fulcrum) and Gly-40/41 (hinge). Experimental evidence suggests that the motion of the flaps correlate with hinge region fluctuations [42]. The flaps cover the substrate upon protease-substrate association and open again to release the products. Therefore, the exposure of the substrate binding site (drug binding site) is proportional to the distance of the flap tips from one another, as well as the distances of the flap tips from the catalytic site. The distance between both flap tips (Gly 51/52) and the catalytic Asp-25 residue as well as the relative distance between the two flap tips Gly-51/51 to Gly-50′/51′ were both calculated. The corresponding regions between the two chains behaved differently in the wild-type and N37T↑V models, as shown by the large variance in flap tip to active site distance between chain A and B. However, the N37T↑V protease exhibits larger dynamic fluctuations than the wild-type

because the hinge region insertion (Val-38) allows the protease to sample regions of newly allowed space, thereby altering the dynamic profile of the flap region. Inhibitors that alter the behaviour of the hinge region may provide key therapeutic benefits and be a viable drug design strategy. An example would be an inhibitor that can lock the hinge in place and prevent the flaps from opening [43]

Flap curling acts as a trigger for the opening of the HIV-1 protease flaps because the opening of the protease, both from "closed to semi-open" and "semi-open to fully-open" is preceded by the curling of the flap tips [43,44]. Perryman *et al*., (2003) reported that flap curling is due to the increased conformational flexibility of the two glycine residues at position 49 and 51 in the wild-type protease. The mechanism is important because it allows the burial of a conserved hydrophobic Ile-50/50′ residue which ensures that the opening of the flap region remains thermodynamically favourable. We analysed the angle produced between three adjacent $C_\alpha$: residues Gly-49, Ile-50 and Gly-51 (Gly-49′, Ile-50′ and Gly-51′) for wild-type and residues Gly-50, Ile-51 and Gly-52 (Gly-50′, Ile-51′ and Gly-52′) for N37T↑V. The angle between these three residues (henceforth referred to as tri-$C_\alpha$ angles) served as a metric to measure flap curling. For both models, the opening of the flap region was preceded by the positive curling (a decrease in the tri-$C_\alpha$ angles) early during the simulation (0-2 ns). Despite this, the N37T↑V protease extended both flaps at a faster rate and was observed to sample multiple open and closed conformations before the wild-type was able to complete a single open-to-closed cycle. Interestingly, no flap curling was observed when the variant protease was in the fully open conformation. At ~12.3 ns the variant flaps remained uncurled; thereby, exposing the hydrophobic Ile-51 to the bulk solvent. In theory, this would cause an ordering of water molecules around Ile-51, leading to a decrease in solvation entropy associated with this residue. Therefore, flap curling is not the driving force behind the extended opening of the variant flaps.

### 3.5 Analysis of residue interaction networks

Residue interaction networks describe the non-covalent interactions between amino acid residues within a simulation. We analysed the residue interaction networks of the two models to understand how the N37T↑V protease flaps are able to extend further away from the active site than the wild-type flaps (see section 3.4). Specifically, the formation and dissolution of salt-bridges were examined. A salt-bridge can form when two oppositely charged residues fall within 4 Å of one another [45]. HIV-1 protease contains several charged residues that

form stabilising salt-bridges [9,46]. Our data reveal that the N37T↑V protease has an altered network of salt-bridge formation that modify the dynamics of the flap and hinge regions. Figure 7A illustrates an important inter-hinge salt-bridge between Glu-35 and Arg-57 that remains intact for the majority of the wild-type simulation. However, Glu-35 and Arg-58 in the variant protease do not fall within 4 Å of each other. The enhanced mobility derived from the Val-38 insertion weakens the interaction, increasing the probability that the bond will not form. Taken alone, these data are inadequate at explaining the extensive motion observed for the variant flaps. This salt-bridge can similarly be disrupted in the wild-type protease (figure 7B at ~18 ns) without any effect on flap dynamics.

Other biochemical interactions also likely contribute to the extensive opening of the variant flaps. Consider the wild-type cantilever region that contains an aspartic acid residue (Asp-60) in close proximity (<10 Å) to two lysine residues (Lys-41 and Lys-43) (figure 8A). Both lysine residues are located in the hinge region. At the start of the simulation, the wild-type protease is in the fully closed conformation, and Asp-60 falls within 4 Å of Lys-43. However, once disrupted, this contact is never re-established (figure 8B). At no point during the simulation does Asp-60 and Lys-41 come within 4 Å of one another (data not shown).

Conversely, the N37T↑V protease has two charged residues *in lieu* of Asp-60; namely, Glu-61 and Glu-62. As the simulation progresses, the destabilised variant hinge region fluctuates until the Lys-44/Glu-61 and Lys-42/Gly-62 salt-bridges can form. Upon formation of the salt-bridges, the hinge-cantilever contacts can leverage the flaps open allowing them to sample areas of space that are non-permissible in the wild-type protease. We postulate that the formation of the Lys-44/Glu-61 and Lys-42/Gly-62 salt-bridges provide the thermodynamically favourable conditions that allow the variant flaps to open to this extent.

### 3.6 Ligand docking

In a parallel study, we have shown that the N37T↑V protease displays a 3 fold increase in $IC_{50}$ to DRV (manuscript in preparation). We set out to understand the biochemical basis for the experimentally observed reduction in DRV susceptibility using IFD simulations. Additionally, LPV and ATV were also considered. Computational molecular docking methods such as IFD are routinely used to predict protein-ligand interactions. IFD considers the possible binding modes and the associated conformational changes between a ligand and protein upon ligand binding. Table 1 shows the calculated free energy ($\Delta G$) of docking for the proteases complexed to the selected PIs. The wild-type/LPV, wild-type/ATV and wild-

type/DRV complexes show binding free energies of -12 kcal/mol, -12 kcal/mol and -10 kcal/mol, respectively. The N37T↑V/LPV, N37T↑V/ATV and N37T↑V/DRV complexes had binding free energies of -8 kcal/mol, -9 kcal/mol and -7 kcal/mol, respectively. It is evident that there are diffrences between the binding energies of the PIs bidning to the the active site of the wild-type and variant proteases.

Ahmed *et al*., (2013) performed a comparative study using experimentally determined binding free energy data for proteases from HIV-1 B and C-SA subtypes. The data were derived from isothermal titration calorimetry experiments. The authors found that LPV, ATV and DRV bind the wild-type C-SA protease with higher affinities due to strong interactions between the PIs and protease flaps. The experimental binding free energies for LPV, ATV and DRV are -15.1 kcal/mol, -14.3 kcal/mol and -15.2 kcal/mol, respectively. Given the lower resolution of computational binding, our results are reasonably consistent with what is reported in literature.

Figure 9 shows a ligand interaction diagram of DRV bound to the proteases. The diagram shows that the wild-type/DRV complex possesses more hydrophobic contacts relative to the N37T↑V/DRV complex. Therefore, more water molecules are excluded from the SASA within the binding pocket which entropically favours the association of the PI and wild-type protease. Similarly, the association between the wild-type/DRV complex is more enthalpically favourable due to the relative abundance of favourable contacts such as van der Waals interactions, hydrogen bonds and salt bridges [48]. Similar conclusions can be drawn from the association of the proteases with LPV and ATV. Ligand interaction plots of the proteases bound to LPV and ATV are included in the supplementary material (figure 1).

### 3.7 Study limitations

In this study, there are several limitations to consider. The duration of the simulations was relatively short. Simulations were conducted to observe the opening and closing of the flap regions only. Longer simulation run times could provide further insight into the mechanisms that govern the opening of the variant flaps. The variant displays an altered trajectory; however, most individual analyses in the current work have a high correlation for the majority of the trajectories. Therefore, care must be taken to prevent over-interpretation of the results (e.g. RMSD, RMSF and $R_g$). Secondly, the present study focusses on the cumulative effect of all the mutations present within the N37T↑V clinical variant. Another

manuscript currently in preparation focuses on key residues that play a role in flap and hinge dynamics.

## 4. Conclusion

The precise impact of hinge region mutations and insertions in the HIV-1 protease is not yet understood. The prevalence of hinge region insertions is gradually increasing, and it is important to understand the basis of this selection process. Previously, we have shown that the N37T↑V mutation and insertion confers reduced drug susceptibility to DRV, a third-line antiretroviral drug (manuscript in preparation). In this article, we report how a hinge region mutation and insertion, such as N37T↑V, affects the dynamics of the HIV-1 protease. The findings here correlate with previously published *in vitro* experiments performed on HIV-1 protease mutations as well as other *in silico* based studies, indicating that the motion of the flaps is partially controlled by the hinge region. The data show that hinge region flexibility was greatly enhanced by the N37T↑V mutation and insertion, which results in a faster rate of active site opening. Moreover, the flap regions of the variant protease open to a greater extent due to the action of two key salt-bridges between the hinge and cantilever regions. IFD studies showed that the variant displayed a reduced binding affinity for LPV, ATV and DRV, respectively. The N37T↑V model and resultant kinetic scheme presented here provides an explanation for the experimentally determined decrease in drug susceptibility that should be further explored using conventional biochemical techniques.

## 5. Acknowledgements

The research reported in this publication was supported by the South African Medical Research Council under a Self-Initiated Research Grant to Yasien Sayed. The views and opinions expressed are those of the authors and do not necessarily represent the official views of the SA MRC. This work was supported by the University of the Witwatersrand, South African National Research Foundation Grant 68898 (HWD), and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation Grant 64788 (HWD). The authors would like to thank the National Research Foundation and Professor Lynn Morris (Head: HIV Research, National Institute for Communicable Diseases, South Africa).

## 6. References


[1] Joint United Nations Programme on HIV/AIDS (UNAIDS), Global AIDS update, Geneva, Switz. (2016).

[2] E.S. Furfine, E. D'Souza, K.J. Ingold, J.J. Leban, T. Spector, D.J.T. Porter, Two-step binding mechanism for HIV protease inhibitors, Biochemistry. 31 (1992) 7886–7891. doi:10.1021/bi00149a020.

[3] B.R. Cullen, Human immunodeficiency virus as a prototypic complex retrovirus., J. Virol. 65 (1991) 1053–6. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC239870/.

[4] C.J. Li, B.J. Dezube, D.K. Biswas, C.M. Ahlers, A.B. Pardee, Inhibitors of HIV-1 transcription, Trends Microbiol. 2 (1994) 164–169. doi:10.1016/0966-842X(94)90666-1.

[5] A. Ali, R.M. Bandaranayake, Y. Cai, N.M. King, M. Kolli, S. Mittal, J.F. Murzycki, M.N.L. Nalam, E. a Nalivaika, A. Özen, M.M. Prabu-Jeyabalan, K. Thayer, C. a Schiffer, Molecular Basis for Drug Resistance in HIV-1 Protease, Viruses. 2 (2010) 2509–2535. doi:10.3390/v2112509.

[6] A. Wlodawer, J. Vondrasek, INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design 1, Annu. Rev. Biophys. Biomol. Struct. 27 (1998) 249–284. doi:10.1146/annurev.biophys.27.1.249.

[7] A. Velázquez-Campoy, S. Vega, E. Freire, Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes., Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 6062–7. doi:10.1073/pnas.111152698.

[8] A. Velazquez-Campoy, S. Vega, E. Freire, Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes, Biochemistry. 41 (2002) 8613–8619. doi:10.1021/bi020160i.

[9] P. Naicker, I. Achilonu, S. Fanucchi, M. Fernandes, M.A.A. a Ibrahim, H.W. Dirr, M.E.S.S. Soliman, Y. Sayed, Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance., J. Biomol. Struct. Dyn. 31 (2013) 1370–1380. doi:10.1080/07391102.2012.736774.

[10] A. Gustchina, I.T. Weber, Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap, FEBS Lett. 269 (1990) 269–272. doi:10.1016/0014-5793(90)81171-J.

[11] H. Meiselbach, A.H.C. Horn, T. Harrer, H. Sticht, Insights into amprenavir resistance in E35D HIV-1 protease mutation from molecular dynamics and binding free-energy calculations., J. Mol. Model. 13 (2007) 297–304. doi:10.1007/s00894-006-0121-3.

[12] F. Liu, A.Y. Kovalevsky, J.M. Louis, P.I. Boross, Y.-F. Wang, R.W. Harrison, I.T. Weber, Mechanism of Drug Resistance Revealed by the Crystal Structure of the Unliganded HIV-1 Protease with F53L Mutation, J. Mol. Biol. 358 (2006) 1191–1199. doi:10.1016/j.jmb.2006.02.076.

[13] L. Kuhn, G. Hunt, K.-G. Technau, A. Coovadia, J. Ledwaba, S. Pickerill, M. Penazzato, S. Bertagnolio, C.A. Mellins, V. Black, L. Morris, E.J. Abrams, Drug resistance among newly diagnosed HIV-infected children in the era of more efficacious antiretroviral prophylaxis, AIDS. 28 (2014) 1673–1678. doi:10.1097/QAD.0000000000000261.

[14] S.-Y. Rhee, M.J. Gonzales, R. Kantor, B.J. Betts, J. Ravela, R.W. Shafer, Human

immunodeficiency virus reverse transcriptase and protease sequence database, Nucleic Acids Res. 31 (2003) 298–303. doi:10.1093/nar/gkg100.

[15] M. Kozisek, K.G. Saskova, P. Rezacova, J. Brynda, N.M. van Maarseveen, D. De Jong, C.A. Boucher, R.M. Kagan, M. Nijhuis, J. Konvalinka, Ninety-Nine Is Not Enough: Molecular Characterization of Inhibitor-Resistant Human Immunodeficiency Virus Type 1 Protease Mutants with Insertions in the Flap Region, J. Virol. 82 (2008) 5869–5878. doi:10.1128/JVI.02325-07.

[16] S. Karthik, S. Senapati, Dynamic flaps in HIV-1 protease adopt unique ordering at different stages in the catalytic cycle, Proteins Struct. Funct. Bioinforma. 79 (2011) 1830–1840. doi:10.1002/prot.23008.

[17] V. Hornak, A. Okur, R.C. Rizzo, C. Simmerling, HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations, Proc. Natl. Acad. Sci. 103 (2006) 915–920. doi:10.1073/pnas.0508452103.

[18] J. Trylska, V. Tozzini, C. a Chang, J.A. McCammon, HIV-1 Protease Substrate Binding and Product Release Pathways Explored with Coarse-Grained Molecular Dynamics, Biophys. J. 92 (2007) 4179–4187. doi:10.1529/biophysj.106.100560.

[19] J.M. Louis, Y. Zhang, J.M. Sayer, Y. Wang, R.W. Harrison, I.T. Weber, The L76V Drug Resistance Mutation Decreases the Dimer Stability and Rate of Autoprocessing of HIV-1 Protease by Reducing Internal Hydrophobic Contacts, Biochemistry. 50 (2011) 4786–4795. doi:10.1021/bi200033z.

[20] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank., Nucleic Acids Res. 28 (2000) 235–42. doi:10.1093/nar/28.1.235.

[21] a Velazquez-Campoy, M.J. Todd, E. Freire, HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity., Biochemistry. 39 (2000) 2201–7. http://www.ncbi.nlm.nih.gov/pubmed/10694385.

[22] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega., Mol. Syst. Biol. 7 (2011) 539. doi:10.1038/msb.2011.75.

[23] K. Arnold, L. Bordoli, T. Schwede, Structural bioinformatics The SWISS-MODEL workspace : a web-based environment for protein structure homology modelling, 22 (2006) 195–201. doi:10.1093/bioinformatics/bti770.

[24] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T.G. Cassarino, M. Bertoni, L. Bordoli, T. Schwede, SWISS-MODEL : modelling protein tertiary and quaternary structure using evolutionary information, 42 (2014) 252–258. doi:10.1093/nar/gku340.

[25] R. Laskowski, J.A. Rullmann, M. MacArthur, R. Kaptein, J. Thornton, AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR, J. Biomol. NMR. 8 (1996) 477–486. doi:10.1007/BF00228148.

[26] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291. doi:10.1107/S0021889892009944.

[27] J.M. Louis, R. a McDonald, N.T. Nashed, E.M. Wondrak, D.M. Jerina, S. Oroszlan, P.T. Mora, Autoprocessing of the HIV-1 protease using purified wild-type and mutated

fusion proteins expressed at high levels in Escherichia coli., Eur. J. Biochem. 199 (1991) 361–369.

[28] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. 1–2 (2015) 19–25. doi:10.1016/j.softx.2015.06.001.

[29] O.F. Lange, D. Van Der Spoel, B.L. De Groot, Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR Data, Biophys. J. 99 (2010) 647–655. doi:10.1016/j.bpj.2010.04.062.

[30] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, Proteins Struct. Funct. Genet. 65 (2006) 712–725. doi:10.1002/prot.21123.

[31] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089. doi:10.1063/1.464397.

[32] H. Okumura, S.G. Itoh, Y. Okamoto, Explicit symplectic integrators of molecular dynamics algorithms for rigid-body molecules in the canonical, isobaric-isothermal, and related ensembles, J. Chem. Phys. 126 (2007) 84103. doi:10.1063/1.2434972.

[33] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56. doi:10.1016/0010-4655(95)00042-E.

[34] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera-A visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612. doi:10.1002/jcc.20084.

[35] L. Skjaerven, X. Yao, B.J. Grant, Beginning Trajectory Analysis with Bio3D, 2014 (2013) 1–15.

[36] B.J. Grant, A.P.C. Rodrigues, K.M. ElSawy, J.A. McCammon, L.S.D. Caves, Bio3d: an R package for the comparative analysis of protein structures, Bioinformatics. 22 (2006) 2695–2696. doi:10.1093/bioinformatics/btl461.

[37] L. Skjaerven, X. Yao, G. Scarabelli, B.J. Grant, Integrating protein structural dynamics and evolutionary analysis with Bio3D, BMC Bioinformatics. 15 (2014) 399. doi:10.1186/s12859-014-0399-6.

[38] J.L. Klepeis, K. Lindorff-Larsen, R.O. Dror, D.E. Shaw, Long-timescale molecular dynamics simulations of protein structure and function, Curr. Opin. Struct. Biol. 19 (2009) 120–127. doi:10.1016/j.sbi.2009.03.004.

[39] M.I. Lobanov, N.S. Bogatyreva, O. V Galzitskaia, Radius of gyration is indicator of compactness of protein structure, Mol. Biol. (Mosk). 42 (2008) 701–706. doi:10.1134/S0026893308040195.

[40] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Essential dynamics of proteins, Proteins Struct. Funct. Genet. 17 (1993) 412–425. doi:10.1002/prot.340170408.

[41] S. Haider, G.N. Parkinson, S. Neidle, Molecular Dynamics and Principal Components Analysis of Human Telomeric Quadruplex Multimers, Biophys. J. 95 (2008) 296–311. doi:10.1529/biophysj.107.120501.

[42] A.L. Perryman, J. Lin, J.A. McCammon, HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: Possible contributions to drug resistance and a potential new target site for drugs, Protein Sci. 13 (2004) 1108–1123.

doi:10.1110/ps.03468904.

[43] W.R.P. Scott, C. a. Schiffer, Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance, Structure. 8 (2000) 1259–1265. doi:10.1016/S0969-2126(00)00537-2.

[44] D.I. Freedberg, R. Ishima, J. Jacob, Y. Wang, I. Kustanovich, J.M. Louis, D.A. Torchia, Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations., Protein Sci. 11 (2002) 221–32. doi:10.1110/ps.33202.

[45] S. Kumar, R. Nussinov, Close-range electrostatic interactions in proteins, ChemBioChem. 3 (2002) 604–617. doi:10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X.

[46] X. Huang, M.D. Britto, J.L. Kear-Scott, C.D. Boone, J.R. Rocca, C. Simmerling, R. Mckenna, M. Bieri, P.R. Gooley, B.M. Dunn, G.E. Fanucci, The role of select subtype polymorphisms on HIV-1 protease conformational sampling and dynamics, J. Biol. Chem. 289 (2014) 17203–17214. doi:10.1074/jbc.M114.571836.

[47] S.M. Ahmed, H.G. Kruger, T. Govender, G.E.M. Maguire, Y. Sayed, M. a a Ibrahim, P. Naicker, M.E.S. Soliman, Comparison of the Molecular Dynamics and Calculated Binding Free Energies for Nine FDA-Approved HIV-1 PR Drugs Against Subtype B and C-SA HIV PR, Chem. Biol. Drug Des. 81 (2013) 208–218. doi:10.1111/cbdd.12063.

[48] X. Du, Y. Li, Y.-L. Xia, S.-M. Ai, J. Liang, P. Sang, X.-L. Ji, S.-Q. Liu, Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods, Int. J. Mol. Sci. 17 (2016) 144. doi:10.3390/ijms17020144.

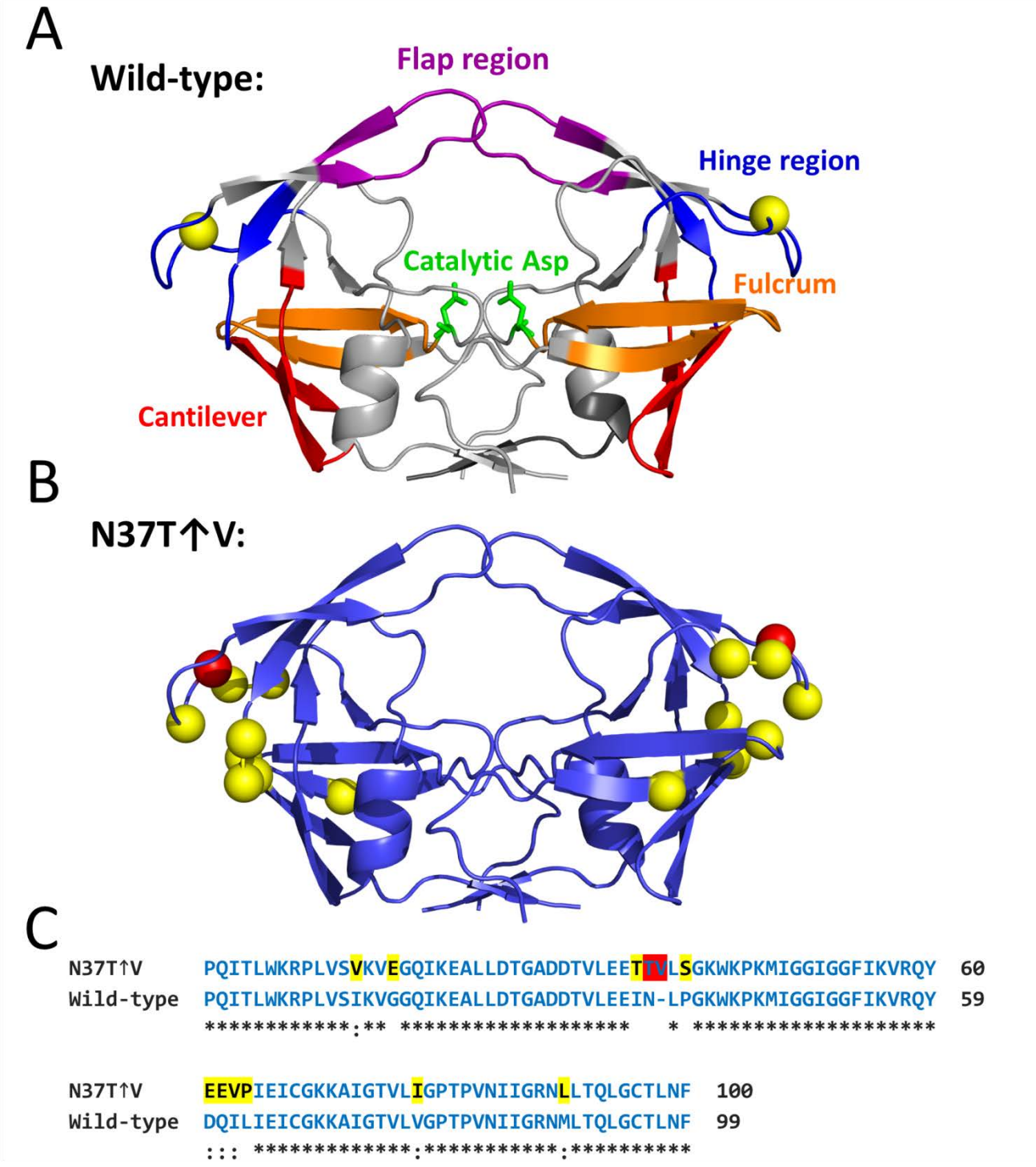



Figure 1. A) The structural architecture of the wild-type HIV-1 protease enzyme (PDB ID: 3u71). The flap regions (46-54 and 46′-54′) are rendered in purple, the hinge regions (residues 35-42, 35′-42′, 57-61, and 57′-61′) are rendered in blue, the fulcrum regions (residues 10-23 and 10′-23′) are rendered in orange, the cantilever regions (residues 62-75 and 62′-75′) are rendered in red and the two catalytic aspartic acids (resides 25 and 25′) are rendered in green. B) Homology model of N37T↑V showing the secondary structural elements of the homology model in blue. The relative position of the insertion mutation is

indicated by red spheres. The yellow spheres represent various background mutations. C) Amino acid sequence alignment of the wild-type and N37T↑V proteases. The sequence alignment data show the positions of the polymorphisms coloured in accordance with Figure 1B. Wild-type subtype C consensus sequence is included as a reference.

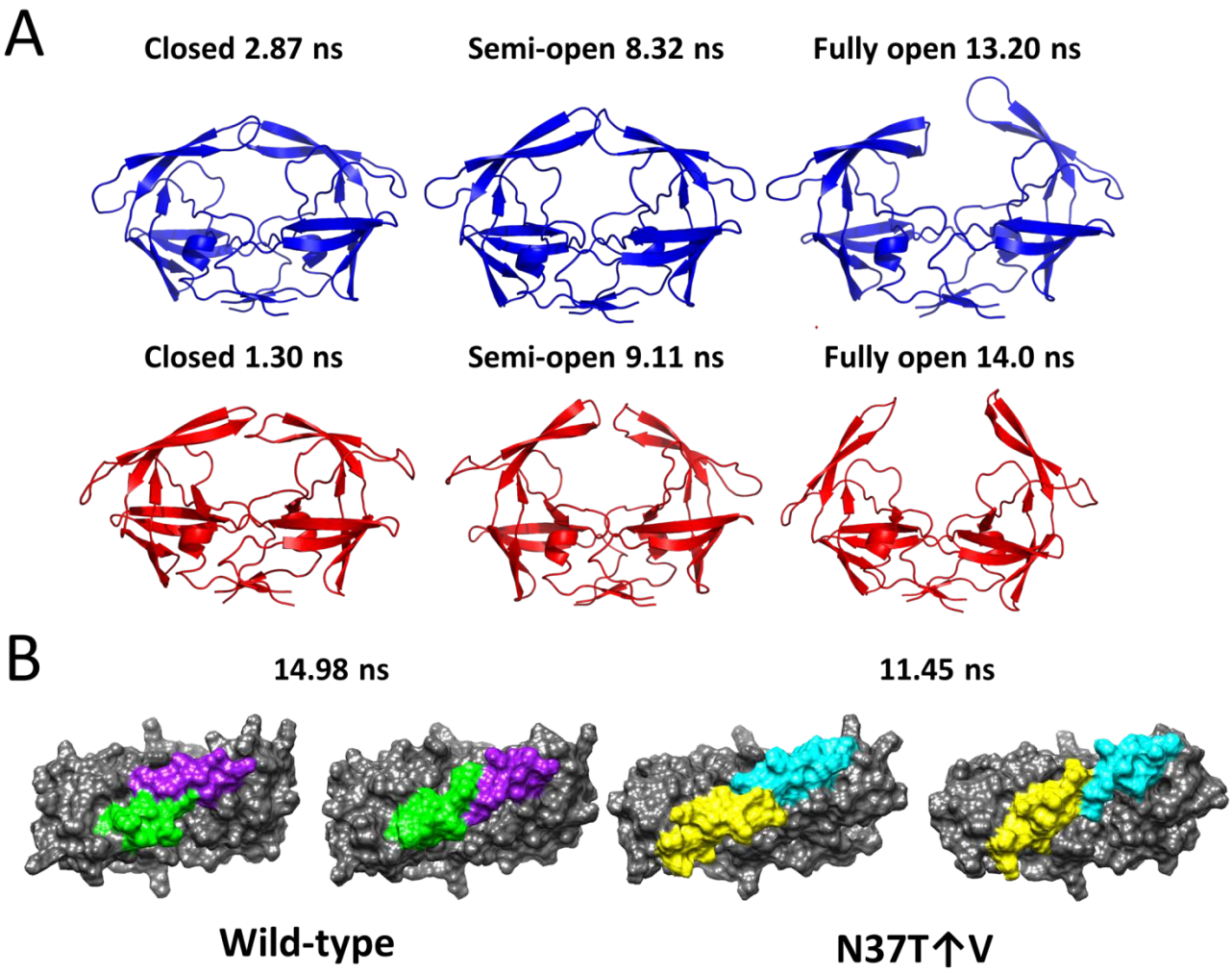



Figure 2. The topologies of the proteases are in a state of flux throughout the simulations. A) Trajectory snapshots of the Wild-type (blue) and N37T↑V (red) at various time points. The snapshots were taken when proteases were in closed, semi-open and fully-open conformations. Models are depicted in side view as cartoon diagrams. B) Atomic surface representations show the switching of the flap handedness. The wild-type flaps are depicted in green and purple, and the N37T↑V flaps are depicted in yellow and cyan.

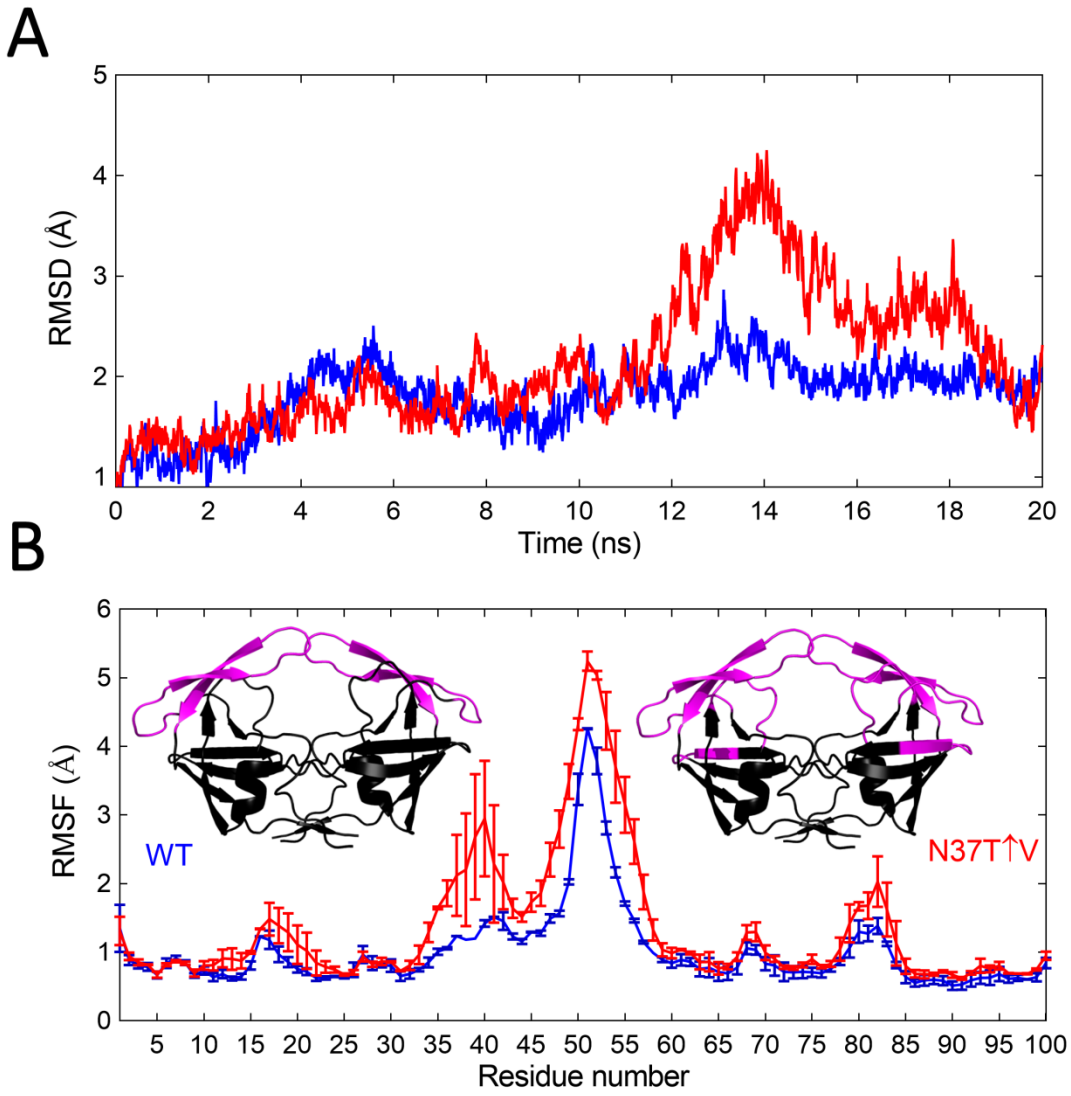



Figure 3. A) RMSD of the peptide backbone with respect to time. The RMSD of the backbone $C_\alpha$ atoms were calculated over the entire trajectory for both the wild-type (blue) and the N37T↑V (red) proteases. All simulations were initiated in the closed conformation. B) RMSF of the structure with respect to amino acid number. The $C_\alpha$ fluctuations were measured over 20 ns. Wild-type data are depicted in blue and N37T↑V data are depicted in red. For each simulation, the values for the two subunits were averaged per corresponding amino acid. The error bars reflect the average values of the two monomers. Included are ribbon diagrams of each protease, wild-type (left inset) and N37T↑V (right inset), colour coded to denote areas of low (<1.5 Å, depicted in black) and high (>1.5 Å, depicted in magenta) atomic fluctuations during the simulations.

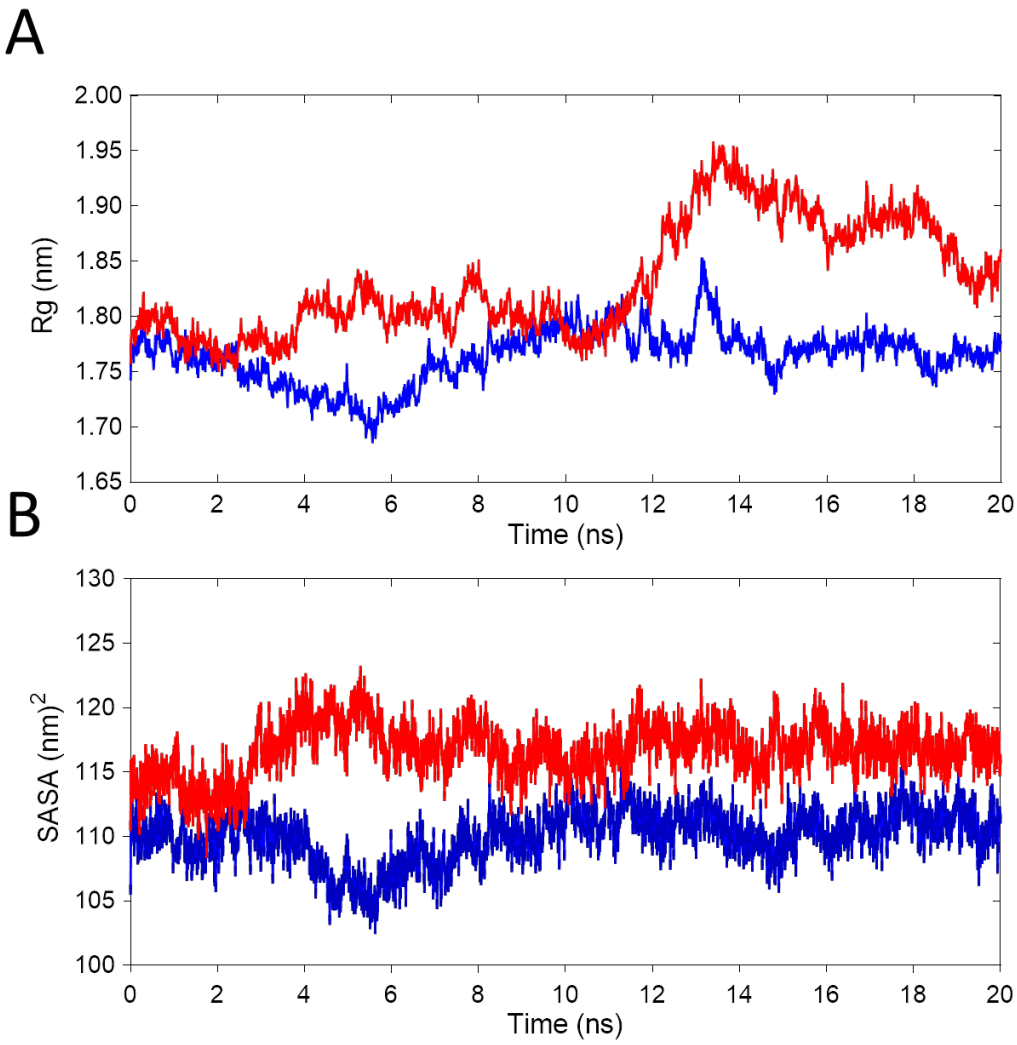



Figure 4. The wild-type protease is more compact than the N37T↑V protease. A) The radius of gyration was measured over the course of each simulation. The blue and red lines highlight the deviations in radius of gyration for wild-type and N37T↑V, respectively. B) The global solvent accessible surface area was calculated for each model. N37T↑V is depicted in red, and the wild-type is depicted in blue.

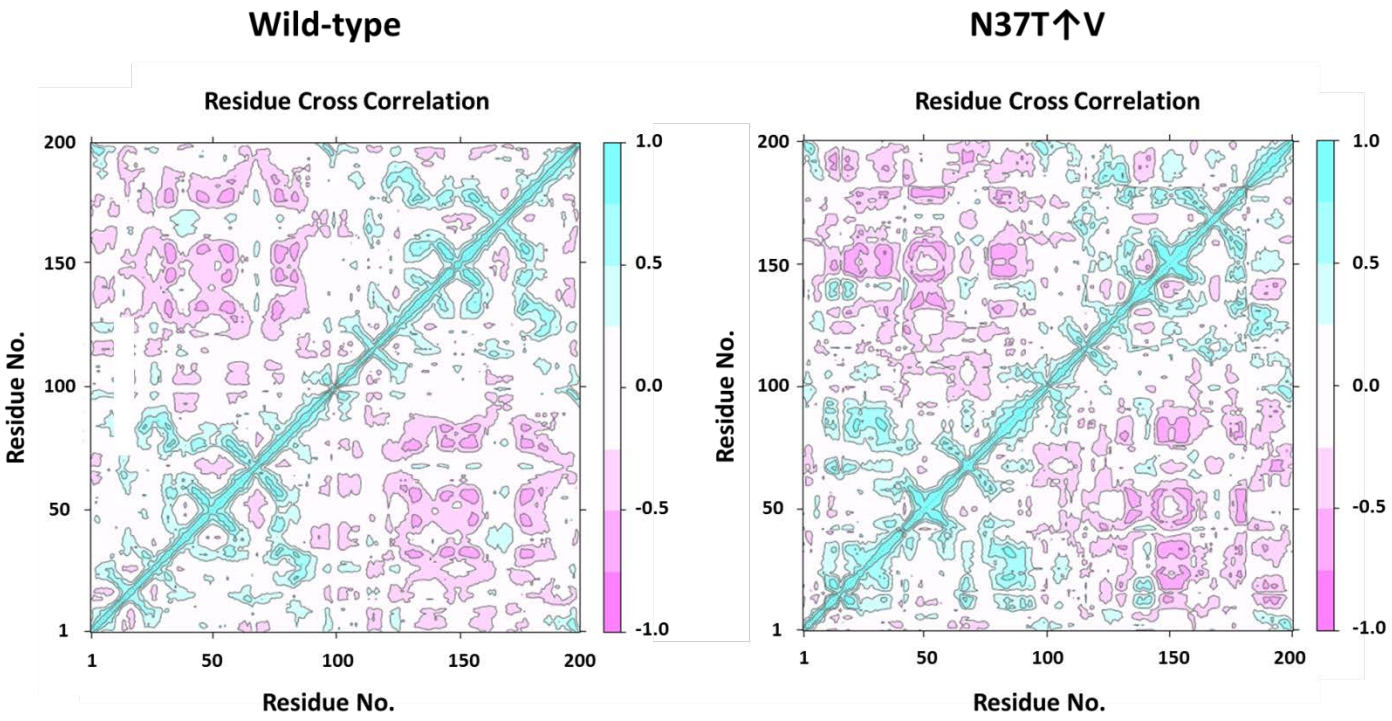



Figure 5. Dynamical cross-correlation maps showing the vectors governing the $C_\alpha$ backbone atoms of the two simulations. The spectrum of positive (1) to negative correlations (-1) is shown on the right. Blue represents positively correlated amino acid vectors, indicating that they move in the same direction, and pink represents anti-correlated amino acid vectors, indicating that they move in opposite directions.

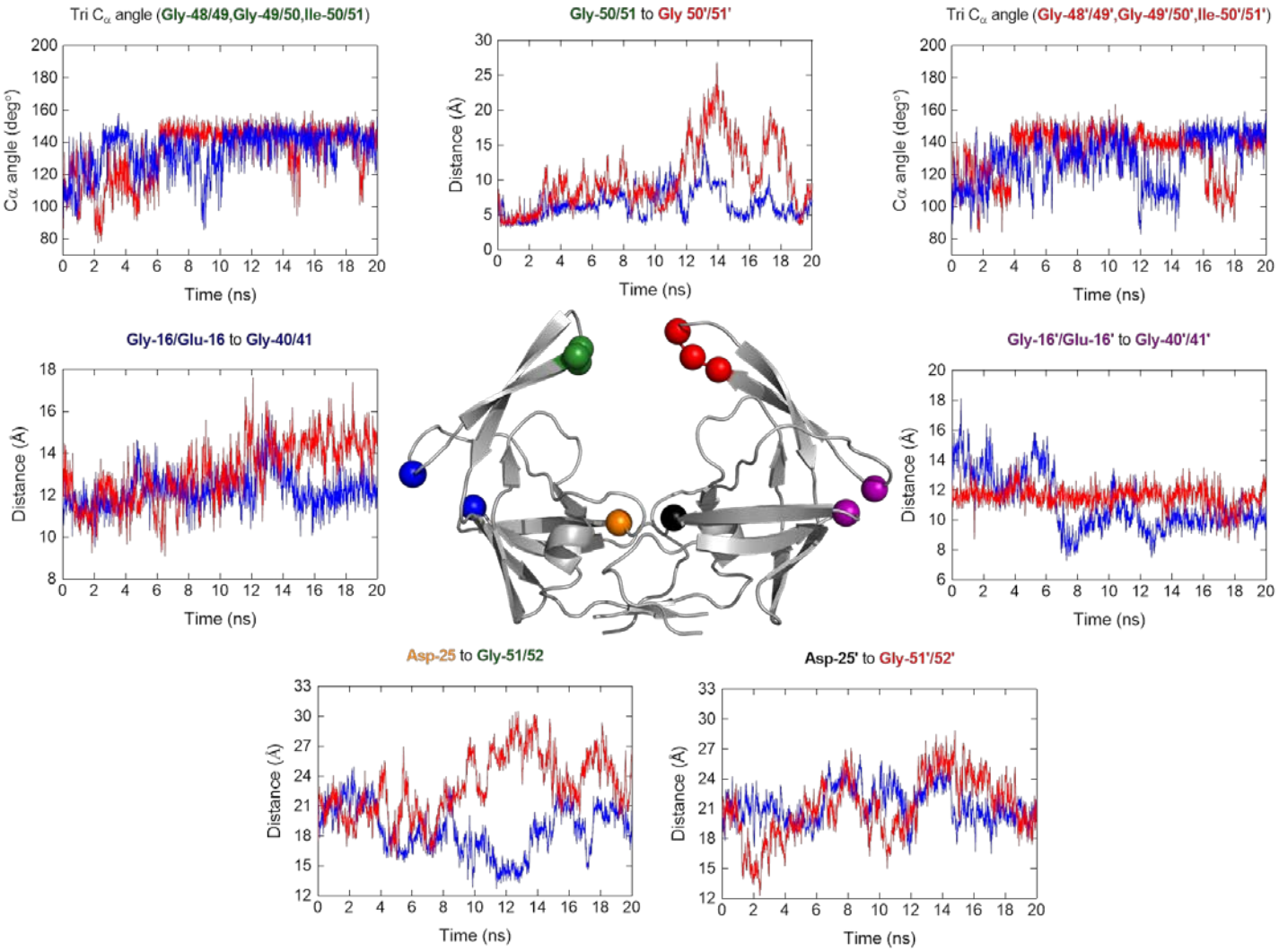



Figure 6. The N37T↑V hinge region mutation and insertion change the dynamic profile of the protease enzyme. The MD data are mapped around the N37T↑V protease which is depicted in the open conformation (centre). Wild-type data are depicted in blue and N37T↑V data are depicted in red. The $C_\alpha$ atoms of key residues are denoted by spheres. For simplicity, the spheres were colour labelled according to the figure headings. Residues of chain B are represented by a prime symbol. Equivalent residues between the variant and wild-type are numbered differently after amino acid 37 due to the presence of an insertion mutation in the variant.

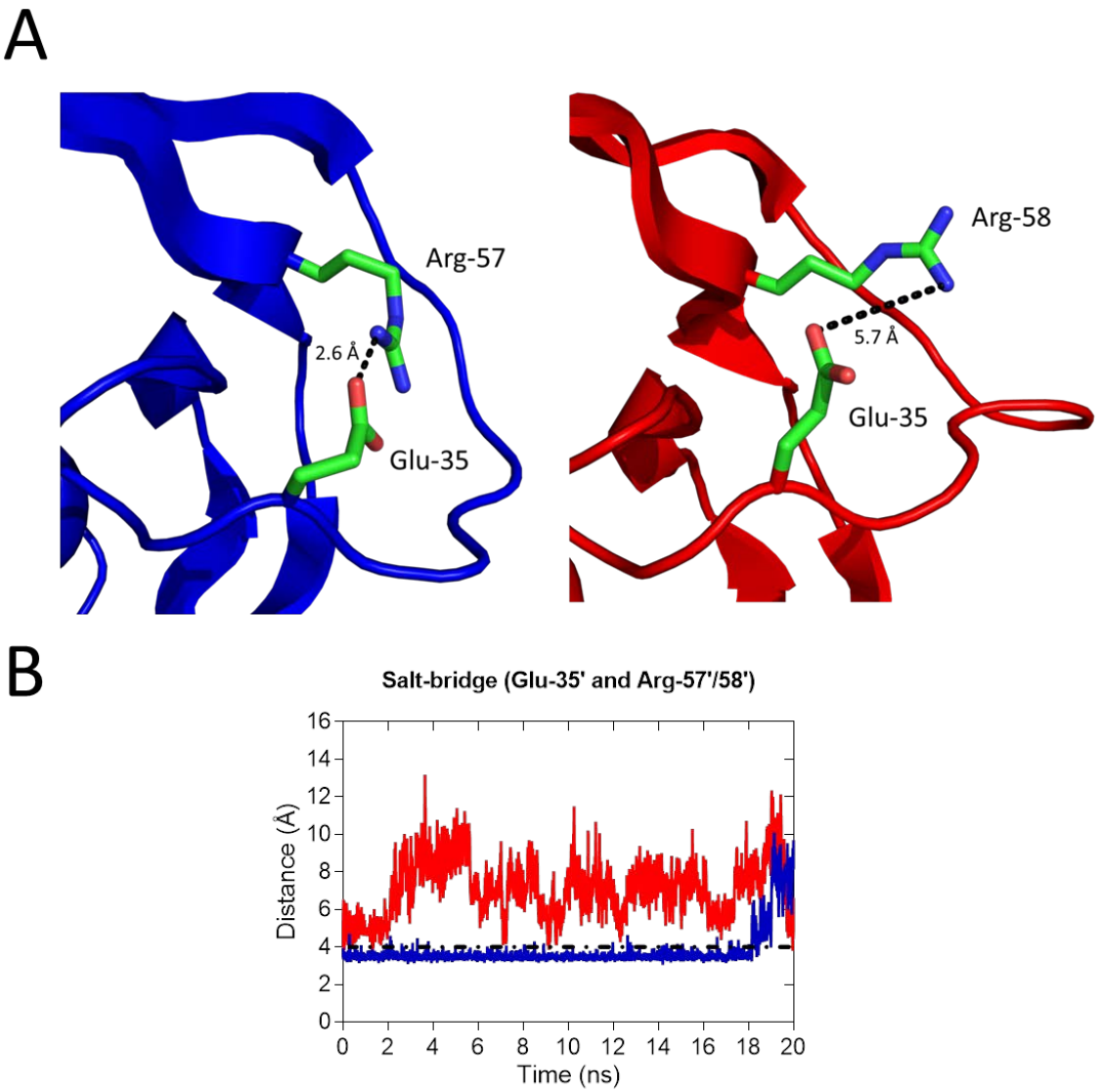



Figure 7. The wild-type protease hinge region is stabilised by a salt-bridge formed between Glu-35 and Arg-57 A) Snapshots of wild-type (blue) and N37T↑V (red) proteases in the fully open conformation. Only the hinge and cantilever regions are illustrated. Charged residues are depicted as sticks and labelled. B) The distance between charged residues residing in the hinge and cantilever regions were measured over the trajectory of both the wild-type (blue) and N37T↑V (red) proteases. The dotted black line marks the maximum distance (4 Å) at which salt-bridges may form.

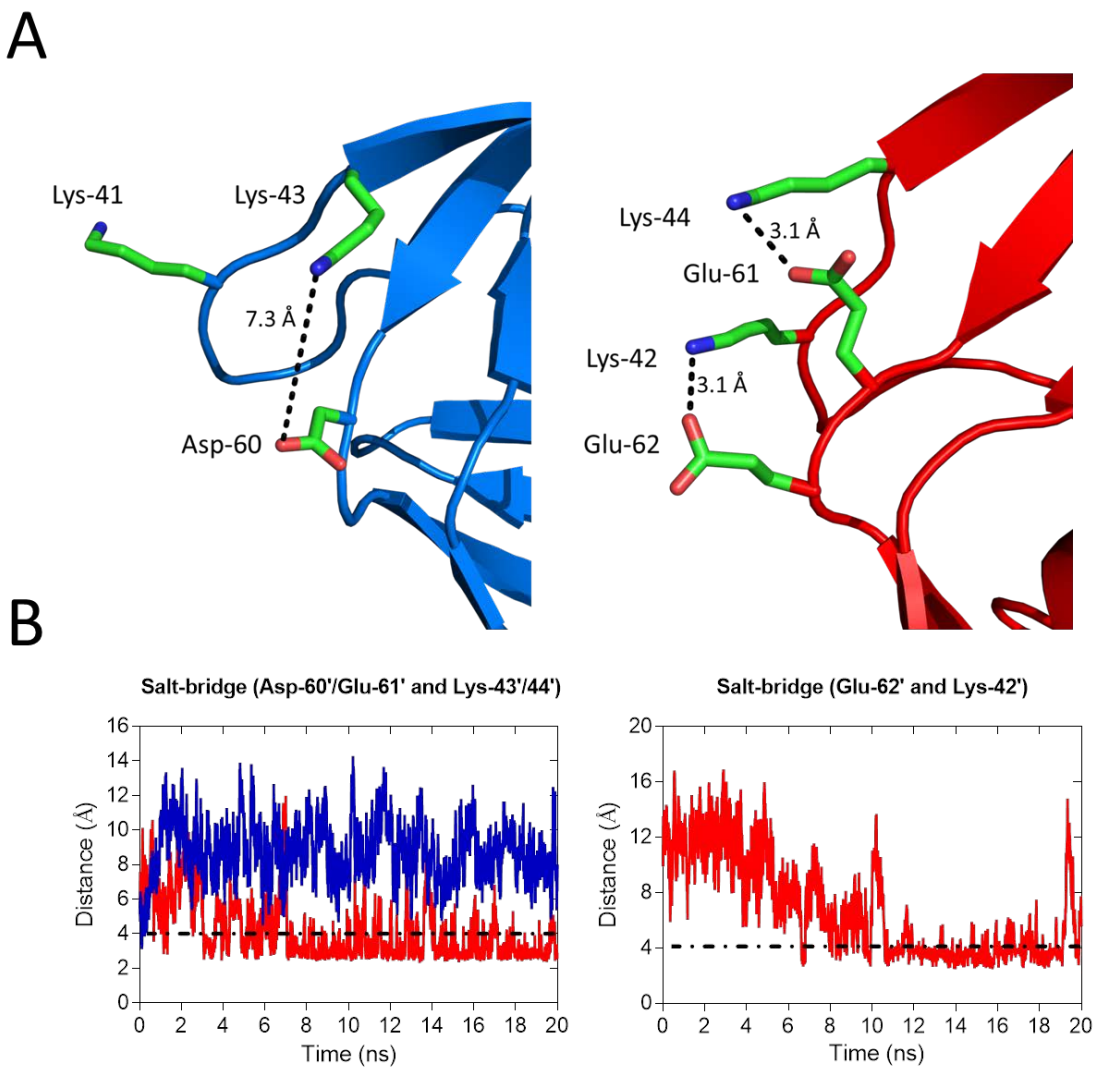



Figure 8. The N37T↑V protease exhibits an altered salt-bridge network between the hinge and cantilever regions. A) Snapshots of the wild-type (blue) and N37T↑V (red) proteases in the fully open conformation. Only the hinge and cantilever regions are illustrated. Charged residues are depicted as sticks and labelled. B) The distance between charged residues residing in the hinge and cantilever regions were measured over the trajectory of both the wild-type (blue) and N37T↑V (red) proteases. The dotted black line marks the maximum distance (4 Å) at which salt-bridges may form.

Table 1. Computationally determined binding docking energy of the wild-type (blue) and N37T↑V (red) proteases. The Glide energy and Glide E-model values, used to for pose selection, are included.

| kcal/mol | LPV | | ATV | | DRV | |
|---|---|---|---|---|---|---|
| **Docking score** | -12 | -8 | -12 | -9 | -10 | -7 |
| **Glide energy** | -85 | -64 | -140 | -119 | -71 | -72 |
| **Glide E-model** | -135 | -83 | -85 | -78 | -109 | -56 |

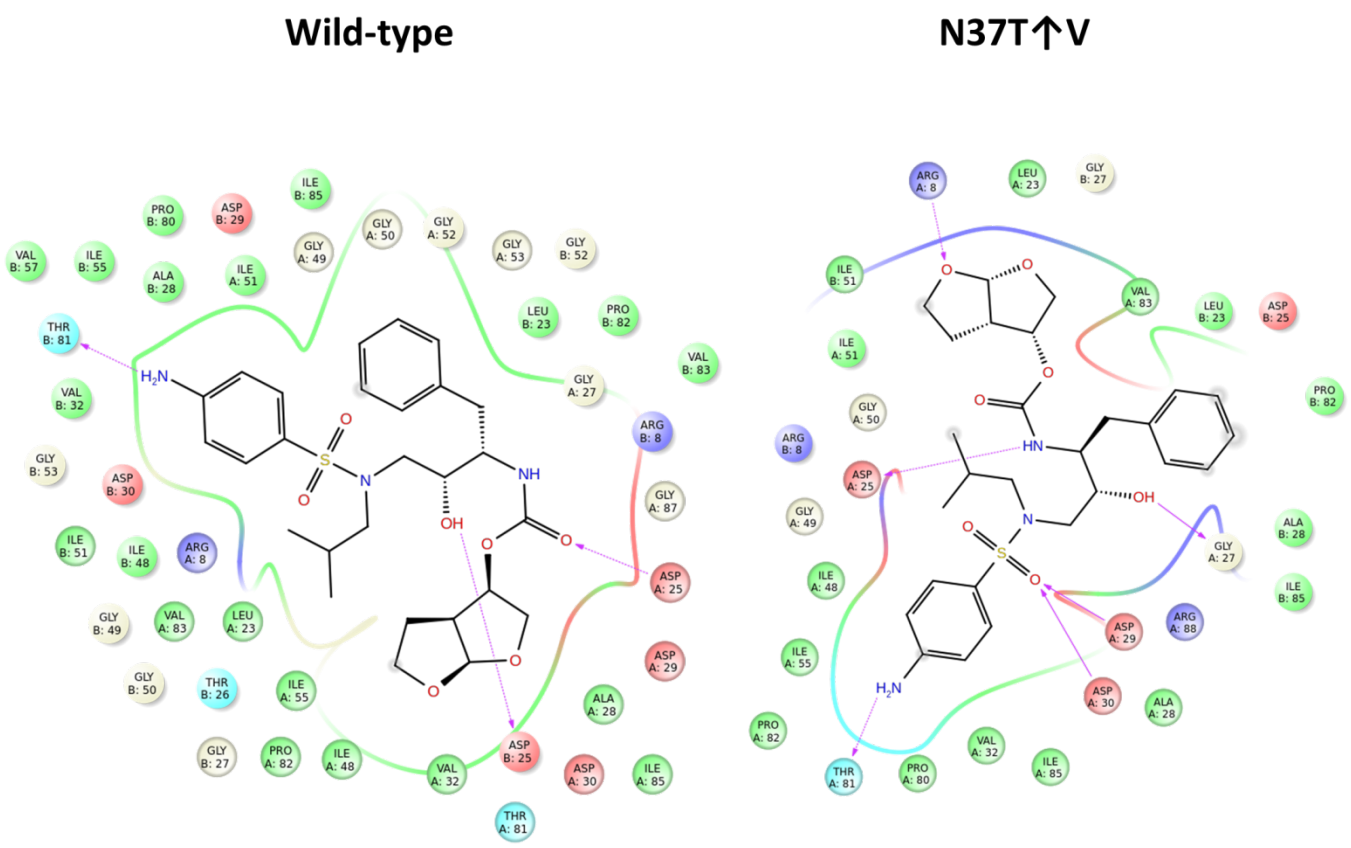



Figure 9. 2-D ligand interaction plot highlighting the differences in DRV binding between the wild-type (left) and N37T↑V (right) protease. In the figures above, hydrophobic interactions are depicted in green, polar interactions are depicted in cyan, positively charged depicted as red/blue lines, and hydrogen bonds are depicted as red arrows.

## 7. Supplemental figure

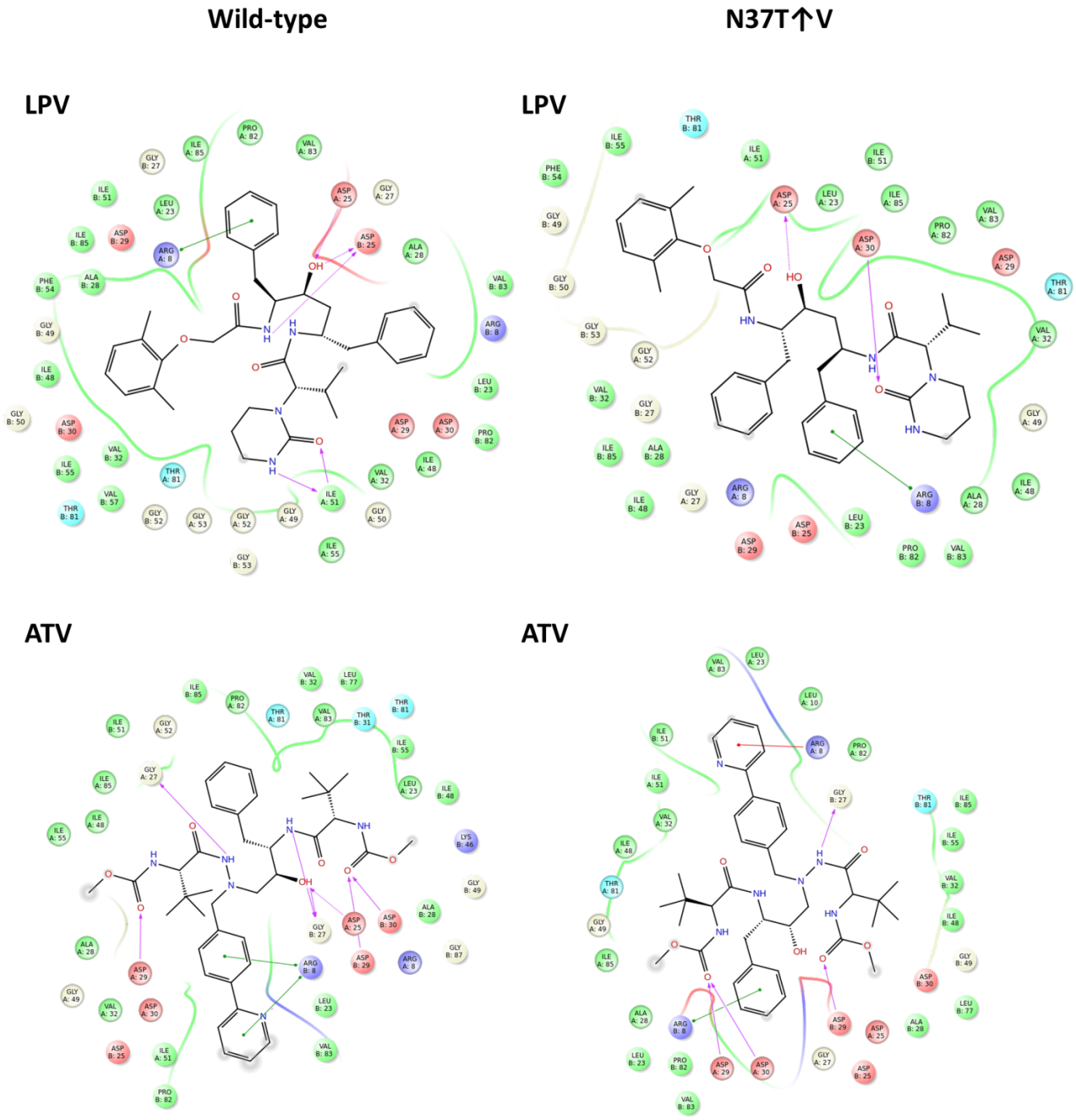



Figure 1. 2-D ligand interaction plot high highlighting the differences in ATV and LPV binding between the wild-type (left inset) and N37T↑V (right inset) proteases. In the figures above, hydrophobic interactions are depicted in green, polar interactions are depicted in cyan, positively charged residues are depicted in red, negatively charged residues are depicted in blue, salt-bridges are depicted as red/blue lines, and hydrogen bonds are depicted as red arrows.

# CHAPTER 5

## General discussion and conclusions

### 5.1 Expression and purification of HIV-1 C-SA proteases

Recombinant DNA technology allows for the successful overexpression of heterologous proteins that can be used for scientific investigation [97]. However, some proteins remain difficult to overexpress as a result of their inherent biochemical properties. The overexpression of HIV-1 PR is particularly challenging due to its cytotoxicity and autolytic nature [98]. The cytotoxicity of HIV-1 PR is derived from its natural function, because it performs detrimental non-specific proteolysis within the host cell, which interferes with normal proliferation [99]. PR expression in bacterial host cells ultimately leads to a slower growth rate, higher death rate and decreased final cell density [99]. Autolysis refers to the ability of HIV-1 PR to perform proteolysis on other HIV-1 PR molecules in solution and this process negatively impacts the final concentration of active PR [100, 101]. For these reasons, current expression systems do not always produce enough PR for subsequent downstream experiments that require high yields, e.g. isothermal titration calorimetry (ITC).

In Chapter 2, a novel method for overexpressing and purifying HIV-1 PR was preposed. The method involved the overexpression of HIV-1 PR as a fusion protein containing a hexahistidine tag for purification purposes. The final construct (TRX-6His-TCS-PR), contained a thioredoxin (TRX) moiety followed by a hexahistidine (6His) tag, thrombin cleavage site (TCS) and protease (PR). The results were compared to a non-fusion Gag-Pol

derived wild-type purified using ion-exchange chromatography. Fusion constructs were prepared for the wild-type PR as well as two variants, one of which was the N37T↑V PR.

A relatively large initial yield of the fusion protein was recovered after the first immobilised metal ion affinity chromatography step (first IMAC step), indicating that the fusion protein displayed a much lower cytotoxic profile than its non-fusion PR counterpart. Additionally, the host cell density and growth rate were improved using this system, signifying that the fusion protein does not disrupt normal host cell activity to the extent of the non-fusion PR control.

The fusion protein was then subjected to an overnight thrombin cleavage assay to free the HIV-1 PR from the construct. The fusion protein was not catalytically active when assayed prior to cleavage of the 6His tag; however, upon cleavage using thrombin, the activity of PR increased over time. These data confirm that the proteolytic activity of HIV-1 PR is virtually abolished when expressed as a TRX fusion protein. However, the activity of the PR is recovered after cleavage from the fusion construct.

After cleavage, pure PR was recovered through a second IMAC purification step. In addition to the smaller volume of culture media needed (1 litre vs 6 litres), the total yield (mg/litre of culture) fusion-derived wild-type and N37T↑V exceeded the control purification method by 250% and 200%, respectively (Chapter 2-Results, Figure 5). The fusion system described in this work solves a fundamental problem of HIV-1 PR overexpression, in that the intrinsic cytotoxicity of the PR is reduced significantly. This method, however, is not without limitations as active site titrations using ITC revealed a relatively low quantity of active protein. Unfortunately, there is a significant loss of active PR during thrombin cleavage due to autolysis. Additionally, there may be some misfolded PR molecules as a result of the refolding

conditions employed during purification, as fusion proteins were recovered from the insoluble cell fraction.

Possible steps that can be taken to improve the quantity of active PR include the following: firstly, thrombin cleavage can be performed at a lower temperature, which would minimise autolysis by decreasing the relative kinetic energy of the PR. Secondly; thrombin cleavage could be performed for a shorter period of time, thereby decreasing the contact time among proteases. Lastly, pure PR could be unfolded in a suitable denaturant and subsequently be refolded; this would increase the number of active proteases in the event of misfolding. These methods have previously been used to recover HIV-2 PR [102, 103].

## 5.2 Drug susceptibility and replication capacity of W1201i

The drug susceptibility and RC of the W1201i isolate were evaluated by performing *in-vitro* singe cycle drug susceptibility assays [104–107]. Additionally, a chimeric construct, consisting of W1201i PR (N37T↑V) and wild-type Gag was assessed to account for the effects conferred by the W1201i Gag. A wild-type subtype C Gag-PR isolate (MJ4), and a known multidrug-resistant isolate served as controls [108]. The isolates were assayed against LPV, ATV and DRV, the most commonly used PIs in South Africa [52].

Results were calculated as a fold-change (FC) in $IC_{50}$ for each drug, and the results were compared to the wild-type and resistance control data. The interpretation of the drug susceptibility tests was based on individually defined cut-offs. The cut-offs were calculated from repeated experiments on the same sample and serve as a measure of variation between experiments [109, 110]. Therefore, cut-offs are an indication of the reproducibility of the experiments and values below the experimentally determined cut-off indicate that an isolate is

fully susceptible to a given PI. Conversely, values above the cut-off indicate that the isolate confers reduced susceptibility to a given PI.

At baseline, FC cut-offs were 1.2 for LPV, 1.4 for ATV and 1.7 for DRV. The W1201i isolate conferred a 4-fold (FC: 4.2) reduction in susceptibility to DRV. No significant differences were observed for LPV and ATV. The chimeric construct showed a 3-fold (FC: 3.4) reduced susceptibility to LPV, 4-fold (FC: 3.6) reduced susceptibility to ATV and 5-fold (FC: 4.7) reduced susceptibility to DRV (Chapter 3-Results, Figure 4). A one-way ANOVA and Bonferroni's multiple comparisons test confirmed that the differences in FC with respect to calculated baseline were significant for all values.

The chimeric construct displayed a similar reduction in susceptibility to DRV compared to the W1201i isolate. Remarkably, the chimeric construct displayed an additional, small but significant, decrease in susceptibility to both LPV and ATV. These results indicate that the observed reduction in susceptibility is primarily due to the N37T↑V PR and not the patient-derived Gag. In fact, the W1201i Gag confers a small, but significant, increase in PI susceptibility. These data add to the body of research highlighting the role of Gag in PI susceptibility [111, 112].

As expected, the W1201i isolate showed a significant 5-fold reduction in RC compared to the wild-type (Chapter 3-Results, Figure 5). Interestingly, the RC of the chimeric construct was increased by more than 60% compared to the wild-type. These data indicate that W1201i Gag is accountable for the observed decrease in RC and not N37T↑V. Additionally, N37T↑V PR is the likely candidate liable for the observed increased in RC.

The co-evolution of Gag and PR has been well-documented [94, 111]. Mutations arising within Gag are able to influence the PR gene sequence and vice versa. This begs the question: did the N37T↑V variant arise in response to mutations accumulating in Gag?

Despite PMTCT treatment, the patient from which the W1201i isolate was recovered contracted HIV. Prenatally, the patient's mother had been exposed to nevirapine (NVP), for 42 days before labour and the patient had been treated with azidothymidine (AZT) as prophylaxis after birth. Both the patient and the mother were PI-naïve at the time blood samples were taken. It has been well-established that both Gag CS, as well as non-CS mutations, can accumulate under NNRTI (e.g. NVP) and NRTI (e.g. AZT) drug pressure [113]. Furthermore, if the mutations seen in N37T↑V developed prior to the Gag mutations, then viral fitness would have markedly increased and natural selection would favour a virus with a replicative advantage. Therefore, the mutations in N37T↑V likely developed as a response to mutations arising in Gag and not vice versa.

In conclusion, the results presented here demonstrate that hinge region mutations and insertions are capable of modifying viral fitness. However, further studies on PR insertions are needed to determine the full extent of the conferred fitness, not only *in vitro* but *in vivo* as well [114]. Currently, genotypic and phenotypic testing often do not take into account the sequence of *gag* during PI resistance testing. This study, however, confirms that the effects of Gag need to be considered if meaningful conclusions are to be drawn about PI resistance [85, 111, 112, 115].

## 5.3 Characterisation of W1201i Gag

Many pressures exist that shape the evolution of HIV-1 Gag. Firstly, specific drug pressures, particularly during suboptimal PI-based therapy, can alter both the CS and non-CS sequences. ARVs such as NNRTIs and NRTIs are known to cause the accumulation of mutations within Gag [113]. Secondly, immune system responses can cause mutations due to the presence of multiple CD8 epitopes along the Gag sequence [116]. Hence, PR and Gag are deemed to co-evolve with respect to drug resistance [111].

In Chapter 3, genotypic evaluation of the W1201i Gag-PR isolate was performed. The W1201i Gag sequence contained notable amino acids polymorphisms and insertions. Firstly, a PTAP motive duplication (PTAPP duplication and LE insertion in $p6^{Gag}$) was found after the p1/p6 cleavage site. Additionally, a previously unreported MSQAG insertion was found within the p2 domain. Within the p2/NC cleavage site, an I372L↑M mutation and insertion was found as well as the following polymorphisms: S369N, S371N, I373M and G377S (Chapter 3-Results, Figure 3).

Some of these polymorphisms and insertions have previously been reported, and are known to affect RC and PI susceptibility. It has been shown that PTAP duplications are more common in subtype C isolates (e.g. W1201i), and are known to influence the virological response to ART [71, 113, 117–122]. Additionally, they may confer a replicative advantage [120, 123]. Interestingly, a positive correlation exists between nucleoside-based ART and PTAP duplications [111, 117]. Furthermore, the p2 domain is highly involved in the sequential processing by HIV-1 PR [124]. However, mutations within p2/NC are not known to directly affect viral loads or RC of the virus [125].

The current study focused predominantly on the effect of a PR hinge region mutation and insertion on PI susceptibility and RC. Therefore, the precise impact of the individual point mutations and insertions within W1201i Gag has not been established here and further investigation needs to be done to ascertain the effects of each of these polymorphisms.

## 5.4 Molecular dynamics of N37T↑V protease

Molecular dynamics simulations have proven to be an ideal tool for studying the dynamics of HIV-1 PR in full atomic detail [27, 55]. To perform MD simulations, accurate homology models need to be created from X-ray crystallographic or nuclear magnetic resonance (NMR) data [30]. Fortunately, there are numerous crystal structures available of the apo and drug bound forms of HIV-1 PR [126]. Crystal structures provide a wealth of information about the overall structure of the enzyme and the binding modes of specific PIs [127]. Additionally, they provide insight into explicit biochemical interactions between the PR and PIs. However, crystal structures do not provide adequate information on the kinetics (i.e. the fluctuations and flexibility of specific regions) of a PR presenting functional polymorphisms [27].

Both computational and experimental studies show that the dynamics of the PR flap region is essential to its function [20, 76, 79, 128–130]. Mutations that alter the flap region behaviour can ultimately lead to lower PI susceptibility and increased viral fitness [27]. The factors that control flap mobility, therefore, have implications in the rational design of new PIs [131].

It has been well-documented that the motion of the PR flap region is interrelated with the motion of the hinge and fulcrum regions [25, 132–134]. In this study, MD simulations were performed to analyse a hinge region variant of C-SA HIV-1 PR; namely, the N37T↑V PR (a video of the simulations can be viewed at: https://youtu.be/V4aHhznWxW8) [135]. Insertion

mutations can cause conformational changes in the geometry of highly mobile regions within HIV-1 PR [136]. For this reason, it was anticipated that the N37T↑V model would display an altered kinetic profile.

Molecular dynamics simulations showed that the N37T↑V variant indeed presented altered dynamics compared to the wild-type PR because conformers of N37T↑V were able to sample regions of space not permitted in the wild-type model. Analysis of RMSD and RMSF data revealed that the majority of the fluctuations corresponded to residues within the flap and hinge regions. Consequently, the inter-flap and inter-hinge distances were measured and a residue interaction network was created.

It was found that the observed variations in the N37T↑V PR dynamics were due to the presence of the hinge region mutation and insertion as well as several other mutations. Therefore, it is not a single process that governs the larger opening of the flap region, but rather the collective action of many amino acid polymorphisms [84]. An altered salt-bridge network, which involved residues of the flap, hinge and fulcrum regions, proved to be a key player in contributing to the larger and protracted opening of the N37T↑V PR flap region.

The data reveal that the N37T↑V PR does not form the stabilising Glu-35/Arg-57 inter-hinge salt-bridge seen in the wild-type model. The resultant destabilisation, in conjunction with the added mobility derived from the N37T↑V insertion, allows the association of two key salt-bridges; namely, Glu-61/Lys-44 and Glu-62/Lys-42. These key salt-bridges are absent within the wild-type PR, and are directly responsible for the larger opening of the variant flaps. Additionally, the increased mobility lessens the probability that the stabilising Glu-35/Arg-57 salt-bridge will form thereby allowing the N37T↑V flaps to remain open for longer relative to the wild-type.

The semi-open conformation is usually the dominant species of wild-type HIV-1 PR in solution, but occasionally the PR will sample the fully-open and closed conformations [30, 134, 137]. Contact with the substrate stimulates the PR to open its flap region fully to allow substrate entry into the active site. It is only in the fully-open configuration that PR readily binds to the substrate because when the flaps are open there is adequate space to accommodate the substrate within the active site, enabling the PR to form the optimal number of chemical contacts required for recognition [28, 62].

Scott and Schiffer have suggested that the semi-open conformation would disallow substrate entry into the active site of the PR [62]. Substrate entry would require that the flaps move at least 15 Å from each other, starting from the closed position, or 7 Å from the semi-open position [138]. In contrast, product release does not require the opening of the flaps to the fully-open conformation, as the products can slide out of the binding cleft on either side of the enzyme [139]. Variants that select for predominantly closed species lead to lower substrate/PR association rates whereas preference for the fully-open conformer is known to improve the rate of substrate entry [28]. An improved rate of entry would contribute positively to the enzyme kinetics of the PR by increasing the catalytic efficiency ($k_{cat}/K_M$) and the catalytic turnover ($k_{cat}$) [77].

However, an optimum flap distance that allows substrate recognition and entry appears to exist, and if this distance is surpassed a marked decrease in catalytic efficiency will result. If a PR variant conformer can sample regions of space further than the established 15 Å from the closed position, then lower rates of substrate processing results due to a decrease in favourable chemical contacts [76, 139]. The alteration of the kinetic parameters is possibly due to increased rates of dissociation between the substrate and N37T↑V PR (Chapter 3-Results, Table 1).

Although the MD models of the wild-type and N37T↑V proteases were moderately different, similarities were also observed. Both structures exhibited a change in flap handedness, where the flaps spontaneously rearrange from the closed to semi-open form (including reversal of flap handed-ness). This phenomenon has been reported by others as well; however, understanding of the precise impact of flap handedness has not yet been discovered [36, 63, 134]. Furthermore, both models sampled multiple opening and closing events in the nanosecond range, which corresponds to the timescale reported by others [28, 31, 32, 134, 140, 141]. The similarities show that there are some conserved molecular dynamics between the two models. Indeed, throughout the simulation the tertiary structure of the variant remains largely the same a result of the high sequence similarity (89 % identity) to the wild-type.

The data represented here could aid in understanding how natural hinge region polymorphisms modulate PI susceptibility and perhaps even provide insight into the development of new allosteric PIs [63]. Recent improvements in macromolecular simulation technology can complement laboratory experiments by providing information on the processes that govern the MD of a system as well as the chemistry behind protein-ligand interactions.

## 5.5 Computational ligand docking

In Chapter 4, the dynamic regions of the HIV-1 PR enzyme were highlighted. The PR flaps undergo considerable structural changes that are crucial not only for enzyme activity, but for inhibitor binding [34]. Similarly, PIs are peptidomimetic and, have many possible conformers [142]. Understandably, the PIs do not rigidly bind to the HIV-1 PR. It is well-established that conformational changes occur in both the drug and ligand upon binding [11, 127, 141].

Therefore, the flexibility of the drug and protein must be taken into account when performing computational drug binding [143].

The goal of computational docking is to predict the complex of two or more biological molecules, e.g. receptor and ligand. Once docked, the binding modes and energetics of molecules can be calculated. At a minimum, one should possess structural information about the molecules of interest; these data can be determined experimentally (e.g. NMR and X-ray crystallography), and predicted computationally (e.g. homology models and MD simulations) [79]. Understandably, accurate input information such as working pH conditions (charge-state) and using statistically likely conformers will greatly improve the binding calculations. Ultimately, the goal is to achieve computationally determined binding free energies that are comparable to experimentally observed binding energies.

Classical computational docking methods treated proteins as rigid bodies, i.e. lock and key approach, which allowed a reduction in both the search space and computational energy, so that the calculations could be performed [144, 145]. With the advent of modern computationally-based methods for ligand docking such as IFD, a broader view of protein-ligand interactions is possible [146]. Modern computers have enough computing power to analyse thousands of possible conformers, where both the protein and ligand are conformationally sampled to find the best poses and the lowest possible binding free energies. Thus, the accuracy of computationally determined energies has improved drastically [147]. IFD is particularly useful if the protein is known to have highly dynamic regions surrounding the substrate binding site [148].

In this study, IFD experiments were carried out on both the N37T↑V as well as the wild-type proteases so as to understand the differences in their respective PI binding modes and energies.

The variant and wild-type proteases were both computationally ligated to LPV, ATV and DRV. The final docked structures were chosen for comparison based on their respective GLIDE E-model scoring functions [149, 150]. In addition, crystallographic data of known drug bound PR structures were examined to ensure that the selected poses conformed to the established natural binding modes of LPV, ATV and DRV, respectively.

It was found that the three PIs bound to the variant with slightly higher binding free energies, suggesting that the affinity between the PIs and the variant is reduced (Chapter 4-Results and Discustion, Table 1). The 3-dimensional bound structures were then plotted as 2-dimensional ligand interaction diagrams (Schrödinger, Maestro ligand interaction plot, LLC 2009, USA) to provide a snapshot of the PR/PI complex at its lowest energy point, highlighting the biochemical basis for the calculated increase in the binding free energies. The data indicate that the variant possessed fewer van der Waals contacts, ionic interactions, hydrogen bonds and hydrophobic contacts when bound to the three PIs, respectively. Thus, there were fewer thermodynamically favourable interactions between the PIs and the variant than between the PIs and the wild-type-type PR.

It is comprehensible that the N37T↑V PR would display a lower affinity toward the PIs. MD simulations show that the variant PR possesses some structural adaption around the flap and hinge regions which confer improved flexibility and range of movement (Chapter 4-Results and Discussion, Table 1). The resultant conformational changes of the N37T↑V PR invariably impact the drug binding landscape by decreasing the number of favourable interactions between the drug and PR.

## 5.6 Enzyme kinetics of N37T↑V protease

The enzymatic parameters of the proteases were determined following the hydrolysis of a fluorogenic substrate (Abz-Arg-Val-Nle-Phe($NO_2$)-Glu-Ala-Nle-$NH_2$), which mimics the CA/p2 cleavage site of the Gag polyprotein precursor [26, 151]. The activity of the N37T↑V variant PR was confirmed by performing a specific activity assay. Although relatively comparable, the variant displayed a marginally lower specific activity than the wild-type PR. Similarly, the variant PR displayed a lower catalytic processing ability ($k_{cat}/K_M$) and catalytic turnover ($k_{cat}$).

The evidence presented in Chapter 4, suggests that N37T↑V PR maintains a highly flexible flap region with increased flap and hinge region dynamics. The impaired rate of substrate processing discussed here likely stems from the altered molecular dynamics of the variant PR because the variant samples "fully-open" flap conformers that open to a larger extent than the "fully-open" wild-type conformers. Such a configuration could result in fewer chemical contacts upon substrate/PR association, leading to lower kinetic rates concerning cleavage of the fluorogenic substrate [152]. Decreased substrate processing rates are linked to the observed loss of viral fitness as the PR cannot cleave its natural substrate as well its wild-type counterpart, leading to lower rates of virion production (Chapter 3-Results, Figure 5) [69].

## 5.7 Conclusions

Non-active site polymorphisms can profoundly affect the structure, function, drug susceptibility, and replication capacity of HIV-1 protease. In this research, a hinge region variant of South African HIV-1 subtype C protease was analysed. The results provide a detailed and plausible mechanism by which hinge region insertion mutations can modulate protease inhibitor susceptibility and viral replication capacity. The natural structural determinants responsible for the observed changes were also established and it was found that protease drug susceptibility is reliant on the viral Gag polyprotein sequence. A holistic approach to protease drug resistance would, therefore, include studying the effects of variations within both protease and Gag. Lastly, the data suggest that the mutations within protease arose to compensate for the mutations already present in Gag and not vice versa. Cumulatively, this work adds to the growing knowledge base detailing the mechanisms that govern protease inhibitor drug susceptibility.

# References


1. Naicker, P. and Sayed, Y. (2014) Non-B HIV-1 subtypes in sub-Saharan Africa: impact of subtype on protease inhibitor efficacy. *Biol Chem* 395:1151–1161. doi: 10.1515/hsz-2014-0162

2. Joint United Nations Programme on HIV/AIDS (UNAIDS). (2016) Global AIDS update. Geneva, Switz

3. Ndung'u, T., Renjifo, B. and Essex, M. (2001) Construction and Analysis of an Infectious Human Immunodeficiency Virus Type 1 Subtype C Molecular Clone. *J Virol* 75:4964–4972. doi: 10.1128/JVI.75.11.4964-4972.2001

4. McCutchan, F.E. (2006) Global epidemiology of HIV. *J Med Virol* 78:S7–S12. doi: 10.1002/jmv.20599

5. Freed, E.O. (2001) HIV-1 replication. *Somat Cell Mol Genet* 26:13–33. doi: 10.1023/A:1021070512287

6. Cullen, B. (1991) Human immunodeficiency virus as a prototypic complex retrovirus. *J Virol* 65:1053–1056.

7. Freed, E.O. (1998) HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* 251:1–15. doi: 10.1006/viro.1998.9398

8. Hirsch, M.S., Günthard, H.F., Schapiro, J.M., Brun-Vézinet, F., Clotet, B., Hammer, S.M., Johnson, V.A., Kuritzkes, D.R., Mellors, J.W., Pillay, D., Yeni, P.G., Jacobsen, D.M. and Richman, D.D. (2008) Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. *Clin Infect Dis* 47:266–285. doi: 10.1086/589297

9. Emerman, M. (1998) HIV-1 Regulatory/Accessory Genes: Keys to Unraveling Viral and Host Cell Biology. *Science (80- )* 280:1880–1884. doi: 10.1126/science.280.5371.1880

10. Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A., Scolnick, E.M. and Sigal, I.S. (1988) Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci* 85:4686–4690. doi:

10.1073/pnas.85.13.4686

11. Velázquez-Campoy, A., Vega, S. and Freire, E. (2001) Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc Natl Acad Sci U S A* 98:6062–7. doi: 10.1073/pnas.111152698

12. Ahmed, S.M., Kruger, H.G., Govender, T., Maguire, G.E.M., Sayed, Y., Ibrahim, M. a a., Naicker, P. and Soliman, M.E.S. (2013) Comparison of the Molecular Dynamics and Calculated Binding Free Energies for Nine FDA-Approved HIV-1 PR Drugs Against Subtype B and C-SA HIV PR. *Chem Biol Drug Des* 81:208–218. doi: 10.1111/cbdd.12063

13. Mosebi, S., Morris, L., Dirr, H.W. and Sayed, Y. (2008) Active-site mutations in the South african human immunodeficiency virus type 1 subtype C protease have a significant impact on clinical inhibitor binding: kinetic and thermodynamic study. *J Virol* 82:11476–9. doi: 10.1128/JVI.00726-08

14. Hyland, L.J., Tomaszek, T.A. and Meek, T.D. (1991) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry* 30:8454–8463. doi: 10.1021/bi00098a024

15. Weber, I.T., Miller, M., Jaskólski, M., Leis, J., Skalka, A.M. and Wlodawer, A. (1989) Molecular modeling of the HIV-1 protease and its substrate binding site. *Science* 243:928–31.

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

17. Turner, B.G. and Summers, M.F. (1999) Structural biology of HIV. *J Mol Biol* 285:1–32. doi: 10.1006/jmbi.1998.2354

18. Furfine, E.S., D'Souza, E., Ingold, K.J., Leban, J.J., Spector, T. and Porter, D.J.T. (1992) Two-step binding mechanism for HIV protease inhibitors. *Biochemistry* 31:7886–7891. doi: 10.1021/bi00149a020

19. Berman, H.M. (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242. doi: 10.1093/nar/28.1.235

20. Naicker, P., Achilonu, I., Fanucchi, S., Fernandes, M., Ibrahim, M. a a., Dirr, H.W.,

Soliman, M.E.S. and Sayed, Y. (2012) Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance. *J Biomol Struct Dyn* 37–41. doi: 10.1080/07391102.2012.736774

21. Coman, R.M., Robbins, A.H., Fernandez, M. a., Gilliland, C.T., Sochet, A. a., Goodenow, M.M., McKenna, R. and Dunn, B.M. (2008) The contribution of naturally occurring polymorphisms in altering the biochemical and structural characteristics of HIV-1 subtype C protease. *Biochemistry* 47:731–43. doi: 10.1021/bi7018332

22. Velázquez-Campoy, A., Vega, S., Fleming, E., Bacha, U., Sayed, Y., Dirr, H.W. and Freire, E. (2003) Protease inhibition in African subtypes of HIV-1. *AIDS Rev* 5:165–71.

23. Babé, L.M., Rosé, J. and Craik, C.S. (1992) Synthetic “interface” peptides alter dimeric assembly of the HIV 1 and 2 proteases. *Protein Sci* 1:1244–1253. doi: 10.1002/pro.5560011003

24. Shao, W., Everitt, L., Manchester, M., Loeb, D.D., Hutchison, C.A. and Swanstrom, R. (1997) Sequence requirements of the HIV-1 protease flap region determined by saturation mutagenesis and kinetic analysis of flap mutants. *Proc Natl Acad Sci U S A* 94:2243–8. doi: 10.1073/pnas.94.6.2243

25. Liu, Z., Huang, X., Hu, L., Pham, L., Poole, K.M., Tang, Y., Brian, P. and Fanucci, G.E. (2016) Effects of Hinge Region Natural Polymorphisms on Human Immunodeficiency Virus-1 Protease Structure, Dynamics and Drug-Pressure Evolution. doi: 10.1074/jbc.M116.747568

26. Szeltner, Z. and Polgár, L. (1996) Rate-determining Steps in HIV-1 Protease Catalysis. *J Biol Chem* 271:32180–32184. doi: 10.1074/jbc.271.50.32180

27. Cai, Y., Yilmaz, N.K., Myint, W., Ishima, R. and Schiffer, C.A. (2012) Differential Flap Dynamics in Wild-type and a Drug Resistant Variant of HIV-1 Protease Revealed by Molecular Dynamics and NMR Relaxation. *J Chem Theory Comput* 8:3452–3462. doi: 10.1021/ct300076y

28. Perryman, A.L. and Lin, J. (2004) HIV-1 protease molecular dynamics of a wild-type and of the V82F / I84V mutant : Possible contributions to drug resistance and a potential new target site for drugs. 1108–1123. doi: 10.1110/ps.03468904.Importance

29. Soares, R.O., Torres, P.H.M., da Silva, M.L. and Pascutti, P.G. (2016) Unraveling HIV

protease flaps dynamics by Constant pH Molecular Dynamics simulations. *J Struct Biol* 195:216–226. doi: 10.1016/j.jsb.2016.06.006

30. Ishima, R., Freedberg, D.I., Wang, Y., Louis, J.M. and Torchia, D.A. (1999) Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure* 7:1047-S12. doi: 10.1016/S0969-2126(99)80172-5

31. Meiselbach, H., Horn, A.H.C., Harrer, T. and Sticht, H. (2007) Insights into amprenavir resistance in E35D HIV-1 protease mutation from molecular dynamics and binding free-energy calculations. *J Mol Model* 13:297–304. doi: 10.1007/s00894-006-0121-3

32. Wlodawer, A. and Vondrasek, J. (1998) INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design 1. *Annu Rev Biophys Biomol Struct* 27:249–284. doi: 10.1146/annurev.biophys.27.1.249

33. Wittayanarakul, K., Aruksakunwong, O., Saen-oon, S., Chantratita, W., Parasuk, V., Sompornpisut, P. and Hannongbua, S. (2005) Insights into saquinavir resistance in the G48V HIV-1 protease: quantum calculations and molecular dynamic simulations. *Biophys J* 88:867–879. doi: 10.1529/biophysj.104.046110

34. Wlodawer, A. and Erickson, J.W. (1993) Structure-Based Inhibitors of HIV-1 Protease. *Annu Rev Biochem* 62:543–585. doi: 10.1146/annurev.bi.62.070193.002551

35. Tozser, J., Yin, F.H., Cheng, Y.-S.E., Bagossi, P., Weber, I.T., Harrison, R.W. and Oroszlan, S. (1997) Activity of Tethered Human Immunodeficiency Virus 1 Protease Containing Mutations in the Flap Region of One Subunit. *Eur J Biochem* 244:235–241. doi: 10.1111/j.1432-1033.1997.00235.x

36. Leonis, G., Czyżnikowska, Ż., Megariotis, G., Reis, H. and Papadopoulos, M.G. (2012) Computational Studies of Darunavir into HIV-1 Protease and DMPC Bilayer: Necessary Conditions for Effective Binding and the Role of the Flaps. *J Chem Inf Model* 52:1542–1558. doi: 10.1021/ci300014z

37. Wang, W. and Kollman, P. a. (2001) Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *Proc Natl Acad Sci U S A* 98:14937–42. doi: 10.1073/pnas.251265598

38. Silva, A.M., Cachau, R.E., Sham, H.L. and Erickson, J.W. (1996) Inhibition and

catalytic mechanism of HIV-1 aspartic protease. *J Mol Biol* 255:321–46. doi: 10.1006/jmbi.1996.0026

39. Brik, A. and Wong, C. (2003) HIV-1 protease: mechanism and drug discovery. *Org Biomol Chem* 1:5–14. doi: 10.1039/b208248a

40. Polgár, L., Szeltner, Z. and Boros, I. (1994) Substrate-dependent mechanisms in the catalysis of human immunodeficiency virus protease. *Biochemistry* 33:9351–7. doi: 10.1021/bi00197a040

41. Davies, D.R. (1990) The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem* 19:189–215. doi: 10.1146/annurev.bb.19.060190.001201

42. Velazquez-campoy, A., Kiso, Y. and Freire, E. (2001) The Binding Energetics of First- and Second-Generation HIV-1 Protease Inhibitors : Implications for Drug Design. 390:169–175. doi: 10.1006/abbi.2001.2333

43. Leavitt, S. and Freire, E. (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol* 11:560–566. doi: 10.1016/S0959-440X(00)00248-7

44. Tomasselli, A.G., Olsen, M.K., Hui, J.O., Staples, D.J., Sawyer, T.K., Heinrikson, R.L. and Tomich, C.S.C. (1990) Substrate analogue inhibition and active site titration of purified recombinant HIV-1 protease. *Biochemistry* 29:264–269. doi: 10.1021/bi00453a036

45. Rabi, S.A., Laird, G.M., Durand, C.M., Laskey, S., Shan, L., Bailey, J.R., Chioma, S., Moore, R.D. and Siliciano, R.F. (2013) Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. *J Clin Invest* 123:3848–3860. doi: 10.1172/JCI67399

46. Clavel, F. (2004) Mechanisms of HIV Drug Resistance: A Primer. *PRN Noteb* 9:3–7.

47. Ali, A., Bandaranayake, R.M., Cai, Y., King, N.M., Kolli, M., Mittal, S., Murzycki, J.F., Nalam, M.N.L., Nalivaika, E. a., Özen, A., Prabu-Jeyabalan, M.M., Thayer, K. and Schiffer, C. a. (2010) Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses* 2:2509–2535. doi: 10.3390/v2112509

48. Weber, I.T. and Agniswamy, J. (2009) HIV-1 Protease: Structural Perspectives on Drug Resistance. *Viruses* 1:1110–36. doi: 10.3390/v1031110

49. Jayakanthan, M., Chandrasekar, S., Muthukumaran, J. and Mathur, P.P. (2010) Analysis of CYP3A4-HIV-1 protease drugs interactions by computational methods for Highly Active Antiretroviral Therapy in HIV/AIDS. *J Mol Graph Model* 28:455–463. doi: 10.1016/j.jmgm.2009.10.005

50. Burgos, J., Crespo, M., Falcó, V., Curran, A., Imaz, A., Domingo, P., Podzamczer, D., Mateo, M.G., Van den eynde, E., Villar, S. and Ribera, E. (2012) Dual therapy based on a ritonavir-boosted protease inhibitor as a novel salvage strategy for HIV-1-infected patients on a failing antiretroviral regimen. *J Antimicrob Chemother* 67:1453–1458. doi: 10.1093/jac/dks057

51. Zeldin, R.K. and Petruschke, R.A. (2004) Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients. *J Antimicrob Chemother* 53:4–9. doi: 10.1093/jac/dkh029

52. Meintjes, G., Black, J., Conradie, F., Cox, V., Dlamini, S., Fabian, J., Maartens, G., Manzini, T., Mathe, M., Menezes, C., Moorhouse, M., Moosa, Y., Nash, J., Orrell, C., Pakade, Y., Venter, F., Wilson, D. and Society, S.A.H.I.V.C. (2014) Adult antiretroviral therapy guidelines 2014. *South Afr J HIV Med* 15:121–143. doi: 10.7196/SAJHIVMED.1130

53. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B., Baldwin, E., Weber, I., Selk, L., Clawson, L., Schneider, J. and Kent, S. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science (80- )* 245:616–621. doi: 10.1126/science.2548279

54. Perez, M.A.S., Fernandes, P.A. and Ramos, M.J. (2007) Drug design: New inhibitors for HIV-1 protease based on Nelfinavir as lead. *J Mol Graph Model* 26:634–642. doi: 10.1016/j.jmgm.2007.03.009

55. Jenwitheesuk, E. and Samudrala, R. (2003) Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations. *BMC Struct Biol* 3:2. doi: 10.1186/1472-6807-3-2

56. Velazquez-Campoy, A., Luque, I. and Freire, E. (2001) The application of thermodynamic methods in drug design. *Thermochim Acta* 380:217–227. doi: 10.1016/S0040-6031(01)00671-2

57. Ershov, P. V., Gnedenko, O. V., Molnar, A.A., Lisitsa, A. V., Ivanov, A.S. and Archakov, A.I. (2012) Kinetic and thermodynamic analysis of dimerization inhibitors binding to HIV protease monomers by surface plasmon resonance. *Biochem Suppl Ser B Biomed Chem* 6:94–97. doi: 10.1134/S1990750812010039

58. Deng, N., Zhang, P., Cieplak, P. and Lai, L. (2011) Elucidating the Energetics of Entropically Driven Protein–Ligand Association: Calculations of Absolute Binding Free Energy and Entropy. *J Phys Chem B* 115:11902–11910. doi: 10.1021/jp204047b

59. Tellinghuisen, J. and Chodera, J.D. (2011) Systematic errors in isothermal titration calorimetry: concentrations and baselines. *Anal Biochem* 414:297–9. doi: 10.1016/j.ab.2011.03.024

60. Freire, E. (2005) A Thermodynamic Guide to Affinity Optimization of Drug Candidates. In: Proteomics and Protein-Protein Interactions. Springer US, Boston, MA, pp 291–307

61. Freire, E. (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov Today* 13:869–874. doi: 10.1016/j.drudis.2008.07.005

62. Scott, W.R.P. and Schiffer, C. a. (2000) Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure* 8:1259–1265. doi: 10.1016/S0969-2126(00)00537-2

63. Hornak, V. and Simmerling, C. (2007) Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discov Today* 12:132–138. doi: 10.1016/j.drudis.2006.12.011

64. Winters, M.A. and Merigan, T.C. (2005) Insertions in the human immunodeficiency virus type 1 protease and reverse transcriptase genes: clinical impact and molecular mechanisms. *Antimicrob Agents Chemother* 49:2575–82. doi: 10.1128/AAC.49.7.2575-2582.2005

65. Rambaut, A., Posada, D., Crandall, K.A. and Holmes, E.C. (2004) The causes and consequences of HIV evolution. *Nat Rev Genet* 5:52–61. doi: 10.1038/nrg1246

66. Ceccherini-Silberstein, F., Erba, F., Gago, F., Bertoli, A., Forbici, F., Bellocchi, M.C., Gori, C., D'Arrigo, R., Marcon, L., Balotta, C., Antinori, A., Monforte, A. d'Arminio. and Perno, C.-F. (2004) Identification of the minimal conserved structure of HIV-1 protease in the presence and absence of drug pressure. *AIDS* 18:11–19. doi:

10.1097/01.aids.0000131394.76221.02

67. Bally, F., Martinez, R., Peters, S., Sudre, P. and Telenti, A. (2000) Polymorphism of HIV Type 1 Gag p7/p1 and p1/p6 Cleavage Sites: Clinical Significance and Implications for Resistance to Protease Inhibitors. *AIDS Res Hum Retroviruses* 16:1209–1213. doi: 10.1089/08892220050116970

68. D'Aquila, R.T., Conway, B., Demeter, L.M., Grant, R.M., A, V., Kuritzkes, D.R., Loveday, C., Shafer, R.W. and D, D. (2002) Drug Resistance Mutations in HIV-1. *Top HIV Med* 10:11–15.

69. Zennou, V., Mammano, F., Paulous, S., Mathez, D. and Clavel, F. (1998) Loss of viral fitness associated with multiple Gag and Gag-Pol processing defects in human immunodeficiency virus type 1 variants selected for resistance to protease inhibitors in vivo. *J Virol* 72:3300–3306.

70. Buckheit Jr, R.W. (2004) Understanding HIV resistance, fitness, replication capacity and compensation: targeting viral fitness as a therapeutic strategy. *Expert Opin Investig Drugs* 13:933–958. doi: 10.1517/13543784.13.8.933

71. Maguire, M.F., Guinea, R., Griffin, P., Macmanus, S., Elston, R.C., Wolfram, J., Richards, N., Hanlon, M.H., Porter, D.J.T., Wrin, T., Parkin, N., Tisdale, M., Furfine, E., Petropoulos, C., Snowden, B.W. and Kleim, J.-P. (2002) Changes in Human Immunodeficiency Virus Type 1 Gag at Positions L449 and P453 Are Linked to I50V Protease Mutants In Vivo and Cause Reduction of Sensitivity to Amprenavir and Improved Viral Fitness In Vitro. *J Virol* 76:7398–7406. doi: 10.1128/JVI.76.15.7398-7406.2002

72. Maldarelli, F. (2003) HIV-1 fitness and replication capacity: What are they and can they help in patient management? *Curr Infect Dis Rep* 5:77–84. doi: 10.1007/s11908-003-0068-9

73. Bi, X., Gatanaga, H., Koike, K., Kimura, S. and Oka, S. (2007) Reversal periods and patterns from drug-resistant to wild-type HIV type 1 after cessation of anti-HIV therapy. *AIDS Res Hum Retroviruses* 23:43–50. doi: 10.1089/aid.2005.0029

74. Pereira-Vaz, J., Duque, V., Trindade, L., Saraiva-da-Cunha, J. and Meliço-Silvestre, A. (2009) Detection of the protease codon 35 amino acid insertion in sequences from

treatment-naïve HIV-1 subtype C infected individuals in the Central Region of Portugal. *J Clin Virol* 46:169–172. doi: 10.1016/j.jcv.2009.06.019

75. Kozisek, M., Saskova, K.G., Rezacova, P., Brynda, J., van Maarseveen, N.M., De Jong, D., Boucher, C.A., Kagan, R.M., Nijhuis, M. and Konvalinka, J. (2008) Ninety-Nine Is Not Enough: Molecular Characterization of Inhibitor-Resistant Human Immunodeficiency Virus Type 1 Protease Mutants with Insertions in the Flap Region. *J Virol* 82:5869–5878. doi: 10.1128/JVI.02325-07

76. Coman, R.M., Robbins, A.H., Goodenow, M.M., Dunn, B.M. and McKenna, R. (2008) High-resolution structure of unbound human immunodeficiency virus 1 subtype C protease: Implications of flap dynamics and drug resistance. *Acta Crystallogr Sect D Biol Crystallogr* 64:754–763. doi: 10.1107/S090744490801278X

77. Naicker, P., Stoychev, S., Dirr, H.W. and Sayed, Y. (2014) Amide hydrogen exchange in HIV-1 subtype B and C proteases - insights into reduced drug susceptibility and dimer stability. *FEBS J* 281:5395–410. doi: 10.1111/febs.13084

78. Gustchina, A. and Weber, I.T. (1990) Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap. *FEBS Lett* 269:269–272. doi: 10.1016/0014-5793(90)81171-J

79. Karthik, S. and Senapati, S. (2011) Dynamic flaps in HIV-1 protease adopt unique ordering at different stages in the catalytic cycle. *Proteins Struct Funct Bioinforma* 79:1830–1840. doi: 10.1002/prot.23008

80. Harte, W.E., Swaminathan, S. and Beveridge, D.L. (1992) Molecular dynamics of HIV-1 protease. *Proteins Struct Funct Genet* 13:175–94. doi: 10.1002/prot.340130302

81. Velazquez-Campoy, A., Vega, S. and Freire, E. (2002) Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. *Biochemistry* 41:8613–8619. doi: 10.1021/bi020160i

82. Aruksakunwong, O., Wolschann, P., Hannongbua, S. and Sompornpisut, P. (2006) Molecular Dynamic and Free Energy Studies of Primary Resistance Mutations in HIV-1 Protease−Ritonavir Complexes. *J Chem Inf Model* 46:2085–2092. doi: 10.1021/ci060090c

83. Huang, X., Britto, M.D., Kear-Scott, J.L., Boone, C.D., Rocca, J.R., Simmerling, C.,

Mckenna, R., Bieri, M., Gooley, P.R., Dunn, B.M. and Fanucci, G.E. (2014) The role of select subtype polymorphisms on HIV-1 protease conformational sampling and dynamics. *J Biol Chem* 289:17203–17214. doi: 10.1074/jbc.M114.571836

84. Tozzini, V., Trylska, J., Chang, C. en. and McCammon, J.A. (2007) Flap opening dynamics in HIV-1 protease explored with a coarse-grained model. *J Struct Biol* 157:606–615. doi: 10.1016/j.jsb.2006.08.005

85. Giandhari, J., Basson, A.E., Sutherland, K., Parry, C.M., Cane, P.A., Coovadia, A., Kuhn, L., Hunt, G. and Morris, L. (2016) Contribution of Gag and Protease to HIV-1 Phenotypic Drug Resistance in Paediatric Patients Failing Protease-Inhibitor Based Therapy. *Antimicrob Agents Chemother* 60:2248–56. doi: 10.1128/AAC.02682-15

86. Parry, C.M., Kohli, A., Boinett, C.J., Towers, G.J., McCormick, A.L. and Pillay, D. (2009) Gag Determinants of Fitness and Drug Susceptibility in Protease Inhibitor-Resistant Human Immunodeficiency Virus Type 1. *J Virol* 83:9094–9101. doi: 10.1128/JVI.02356-08

87. Gupta, R.K., Kohli, A., McCormick, A.L., Towers, G.J., Pillay, D. and Parry, C.M. (2010) Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays. *AIDS* 24:1651–5.

88. Manasa, J., Varghese, V., Pond, S.L.K., Rhee, S.Y., Tzou, P.L., Fessel, W.J., Jang, K.S., White, E., Rögnvaldsson, T., Katzenstein, D.A. and Shafer, R.W. (2017) Evolution of gag and gp41 in Patients Receiving Ritonavir-Boosted Protease Inhibitors. *Sci Rep* 7:1–11. doi: 10.1038/s41598-017-11893-8

89. Nijhuis, M., Van Maarseveen, N.M., Lastere, S., Schipper, P., Coakley, E., Glass, B., Rovenska, M., De Jong, D., Chappey, C., Goedegebuure, I.W., Heilek-Snyder, G., Dulude, D., Cammack, N., Brakier-Gingras, L., Konvalinka, J., Parkin, N., Kräusslich, H.G., Brun-Vezinet, F. and Boucher, C.A.B. (2007) A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med* 4:0152–0163. doi: 10.1371/journal.pmed.0040036

90. Robinson, L.H., Myers, R.E., Snowden, B.W., Tisdale, M. and Blair, E.D. (2000) HIV type 1 protease cleavage site mutations and viral fitness: implications for drug susceptibility phenotyping assays. *AIDS Res Hum Retroviruses* 16:1149–56. doi: 10.1089/088922200414992

91. Gatanaga, H., Suzuki, Y., Tsang, H., Yoshimura, K., Kavlick, M.F., Nagashima, K., Gorelick, R.J., Mardy, S., Tang, C., Summers, M.F. and Mitsuya, H. (2002) Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J Biol Chem* 277:5952–61. doi: 10.1074/jbc.M108005200

92. Tamiya, S., Mardy, S., Kavlick, M.F., Yoshimura, K. and Mistuya, H. (2004) Amino Acid Insertions near Gag Cleavage Sites Restore the Otherwise Compromised Replication of Human Immunodeficiency Virus Type 1 Variants Resistant to Protease Inhibitors. *J Virol* 78:12030–12040. doi: 10.1128/JVI.78.21.12030-12040.2004

93. Myint, L., Matsuda, M., Matsuda, Z., Yokomaku, Y., Chiba, T., Okano, A., Yamada, K. and Sugiura, W. (2004) Gag Non-Cleavage Site Mutations Contribute to Full Recovery of Viral Fitness in Protease Inhibitor-Resistant Human Immunodeficiency Virus Type 1. *Antimicrob Agents Chemother* 48:444–452. doi: 10.1128/AAC.48.2.444-452.2004

94. Dam, E., Quercia, R., Glass, B., Descamps, D., Launay, O., Duval, X., Kräusslich, H.-G., Hance, A.J. and Clavel, F. (2009) Gag Mutations Strongly Contribute to HIV-1 Resistance to Protease Inhibitors in Highly Drug-Experienced Patients besides Compensating for Fitness Loss. *PLoS Pathog* 5:e1000345. doi: 10.1371/journal.ppat.1000345

95. South African National Department of Health. (2015) National Consolidated Guidelines for the Prevention of Mother-To-Child Transmission of HIV (PMTCT) and the Managment of HIV in Children, Adolescents and Adults. *Dep Heal Repub South Africa* 1–128.

96. Kuhn, L., Hunt, G., Technau, K.-G., Coovadia, A., Ledwaba, J., Pickerill, S., Penazzato, M., Bertagnolio, S., Mellins, C.A., Black, V., Morris, L. and Abrams, E.J. (2014) Drug resistance among newly diagnosed HIV-infected children in the era of more efficacious antiretroviral prophylaxis. *AIDS* 28:1673–1678. doi: 10.1097/QAD.0000000000000261

97. Davis, G.D., Elisee, C., Newham, D.M. and Harrison, R.G. (1999) New fusion protein systems designed to give soluble expression in Escherichia coli. *Biotechnol Bioeng* 65:382–8.

98. Volontè, F., Piubelli, L. and Pollegioni, L. (2011) Optimizing HIV-1 protease production in Escherichia coli as fusion protein. *Microb Cell Fact* 10:53. doi:

10.1186/1475-2859-10-53

99. Blanco, R., Carrasco, L. and Ventoso, I. (2003) Cell Killing by HIV-1 Protease. *J Biol Chem* 278:1086–1093. doi: 10.1074/jbc.M205636200

100. Mildner, A.M., Rothrock, D.J., Leone, J.W., Bannow, C.A., Lull, J.M., Reardon, I.M., Sarcich, J.L., Howe, W.J., Tomich, C.S. and Smith, C.W. (1994) The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties. *Biochemistry* 33:9405–13.

101. Louis, J.M., McDonald, R. a., Nashed, N.T., Wondrak, E.M., Jerina, D.M., Oroszlan, S. and Mora, P.T. (1991) Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in Escherichia coli. *Eur J Biochem* 199:361–369.

102. Hui, J.O., Tomasselli, A.G., Reardon, I.M., Lull, J.M., Brunner, D.P., Tomich, C.-S.C. and Heinrikson, R.L. (1993) Large scale purification and refolding of HIV-1 protease fromEscherichia coli inclusion bodies. *J Protein Chem* 12:323–327. doi: 10.1007/BF01028194

103. Ishima, R., Ghirlando, R., Tözsér, J., Gronenborn, A.M., Torchia, D.A. and Louis, J.M. (2001) Folded Monomer of HIV-1 Protease. *J Biol Chem* 276:49110–49116. doi: 10.1074/jbc.M108136200

104. Boden, D., Hurley, A., Zhang, L., Cao, Y., Guo, Y., Jones, E., Tsay, J., Ip, J., Farthing, C., Limoli, K., Parkin, N. and Markowitz, M. (1999) HIV-1 drug resistance in newly infected individuals. *JAMA* 282:1135–41.

105. Aiken, C. (1997) Pseudotyping human immunodeficiency virus type 1 (HIV-1) by the glycoprotein of vesicular stomatitis virus targets HIV-1 entry to an endocytic pathway and suppresses both the requirement for Nef and the sensitivity to cyclosporin A. *J Virol* 71:5871–7.

106. Bartz, S.R. and Vodicka, M.A. (1997) Production of High-Titer Human Immunodeficiency Virus Type 1 Pseudotyped with Vesicular Stomatitis Virus Glycoprotein. *Methods* 12:337–342. doi: 10.1006/meth.1997.0487

107. Soneoka, Y., Cannon, P.M., Ramsdale, E.E., Griffiths, J.C., Romano, G., Kingsman, S.M. and Kingsman, A.J. (1995) A transient three-plasmid expression system for the

production of high titer retroviral vectors. *Nucleic Acids Res* 23:628–633. doi: 10.1093/nar/23.4.628

108. Ndung, T., Renjifo, B., Essex, M. and Essex, M.A.X. (2001) Construction and Analysis of an Infectious Human Immunodeficiency Virus Type 1 Subtype C Molecular Clone Construction and Analysis of an Infectious Human Immunodeficiency Virus Type 1 Subtype C Molecular Clone. *J Virol* 75:4964–4972. doi: 10.1128/JVI.75.11.4964

109. Sen, S., Tripathy, S.P. and Paranjape, R.S. (2006) Antiretroviral drug resistance testing. *J Postgrad Med* 52:187–93.

110. Gupta, R.K., Gibb, D.M. and Pillay, D. (2009) Management of paediatric HIV-1 resistance. *Curr Opin Infect Dis* 22:256–263. doi: 10.1097/QCO.0b013e3283298f1f

111. Fun, A., Wensing, A.M.J., Verheyen, J. and Nijhuis, M. (2012) Human Immunodeficiency Virus Gag and protease: partners in resistance. *Retrovirology* 9:63. doi: 10.1186/1742-4690-9-63

112. Giandhari, J., Basson, A.E., Coovadia, A., Kuhn, L., Abrams, E.J., Strehlau, R., Morris, L. and Hunt, G.M. (2015) Genetic Changes in HIV-1 Gag-Protease Associated with Protease Inhibitor-Based Therapy Failure in Pediatric Patients. *AIDS Res Hum Retroviruses* 31:776–782. doi: 10.1089/aid.2014.0349

113. Sen, S., Tripathy, S.P. and Paranjape, R.S. (2017) Prevalence of Human Immunodeficiency Virus Type 1 Drug Resistance Associated Mutations in Protease Cleavage and Non-Cleavage Sites and PTAP Motif Duplications in Indian HIV-1 Subtype C gag Regions. *Int J Curr Microbiol Appl Sci* 6:1322–1330. doi: 10.20546/ijcmas.2017.605.143

114. Ermolieff, J., Lin, X. and Tang, J. (1997) Kinetic properties of saquinavir-resistant mutants of human immunodeficiency virus type 1 protease and their implications in drug resistance in vivo. *Biochemistry* 36:12364–12370. doi: 10.1021/bi971072e

115. Malet, I., Roquebert, B., Dalban, C., Wirden, M., Amellal, B., Agher, R., Simon, A., Katlama, C., Costagliola, D., Calvez, V. and Marcelin, A.G. (2007) Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. *J Infect* 54:367–374. doi: 10.1016/j.jinf.2006.06.012

116. Geldmacher, C., Currier, J.R., Herrmann, E., Haule, A., Kuta, E., McCutchan, F.,

Njovu, L., Geis, S., Hoffmann, O., Maboko, L., Williamson, C., Birx, D., Meyerhans, A., Cox, J. and Hoelscher, M. (2007) CD8 T-Cell Recognition of Multiple Epitopes within Specific Gag Regions Is Associated with Maintenance of a Low Steady-State Viremia in Human Immunodeficiency Virus Type 1-Seropositive Patients. *J Virol* 81:2440–2448. doi: 10.1128/JVI.01847-06

117. Gallego, O., Mendoza, C. De. and Soriano, V. (2003) Changes in the Human Immunodeficiency Virus p7-p1-p6 gag Gene in Drug-Naive and Pretreated Patients. 41:1245–1247. doi: 10.1128/JCM.41.3.1245

118. Freed, E.O. (2002) Viral Late Domains. 76:4679–4687. doi: 10.1128/JVI.76.10.4679

119. Martins, A.N., Waheed, A.A., Ablan, S.D., Huang, W., Newton, A., Petropoulos, C.J., Brindeiro, R.D.M. and Freed, E.O. (2015) Elucidation of the Molecular Mechanism Driving Duplication of the HIV-1 PTAP Late Domain. *J Virol* 90:768–79. doi: 10.1128/JVI.01640-15

120. Sharma, S., Aralaguppe, S.G., Abrahams, M., Williamson, C., Gray, C., Balakrishnan, P., Saravanan, S., Murugavel, K.G., Solomon, S. and Ranga, U. (2017) The PTAP sequence duplication in HIV-1 subtype C Gag p6 in drug-naive subjects of India and South Africa. *BMC Infect Dis* 17:95. doi: 10.1186/s12879-017-2184-4

121. Ibe, S., Shibata, N., Utsumi, M. and Kaneda, T. (2003) Selection of human immunodeficiency virus type 1 variants with an insertion mutation in the p6(gag) and p6(pol) genes under highly active antiretroviral therapy. *Microbiol Immunol* 47:71–9.

122. Neogi, U., Rao, S.D., Bontell, I., Verheyen, J., Rao, V.R., Gore, S.C., Soni, N., Shet, A., Schülter, E., Ekstrand, M.L., Wondwossen, A., Kaiser, R., Madhusudhan, M.S., Prasad, V.R. and Sonnerborg, A. (2014) Novel tetra-peptide insertion in Gag-p6 ALIX-binding motif in HIV-1 subtype C associated with protease inhibitor failure in Indian patients. *AIDS* 28:2319–2322. doi: 10.1097/QAD.0000000000000419

123. Flys, T., Marlowe, N., Hackett, J., Parkin, N., Schumaker, M., Holzmayer, V., Hay, P. and Eshleman, S.H. (2005) Analysis of PTAP duplications in the gag p6 region of subtype C HIV type 1. *AIDS Res Hum Retroviruses* 21:739–41. doi: 10.1089/aid.2005.21.739

124. Pettit, S.C., Moody, M.D., Wehbie, R.S., Kaplan, A.H., Nantermet, P. V., Klein, C.A.

and Swanstrom, R. (1994) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J Virol* 68:8017–27.

125. Teto, G., Tagny, C.T., Mbanya, D., Fonsah, J.Y., Fokam, J., Nchindap, E., Kenmogne, L., Njamnshi, A.K. and Kanmogne, G.D. (2017) Gag P2/NC and pol genetic diversity, polymorphism, and drug resistance mutations in HIV-1 CRF02-AG- and non-CRF02-AG-infected patients in Yaoundé, Cameroon. *Sci Rep* 7:1–14. doi: 10.1038/s41598-017-14095-4

126. Sasková, K.G., Kozísek, M., Lepsík, M., Brynda, J., Rezácová, P., Václavíková, J., Kagan, R.M., Machala, L. and Konvalinka, J. (2008) Enzymatic and structural analysis of the I47A mutation contributing to the reduced susceptibility to HIV protease inhibitor lopinavir. *Protein Sci* 17:1555–64. doi: 10.1110/ps.036079.108

127. Liu, F., Kovalevsky, A.Y., Louis, J.M., Boross, P.I., Wang, Y.-F., Harrison, R.W. and Weber, I.T. (2006) Mechanism of Drug Resistance Revealed by the Crystal Structure of the Unliganded HIV-1 Protease with F53L Mutation. *J Mol Biol* 358:1191–1199. doi: 10.1016/j.jmb.2006.02.076

128. Freedberg, D.I., Ishima, R., Jacob, J., Wang, Y., Kustanovich, I., Louis, J.M. and Torchia, D.A. (2009) Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci* 11:221–232. doi: 10.1110/ps.33202

129. Ishima, R., Torchia, D.A. and Louis, J.M. (2007) Mutational and structural studies aimed at characterizing the monomer of HIV-1 protease and its precursor. *J Biol Chem* 282:17190–17199. doi: 10.1074/jbc.M701304200

130. de Vera, I.M.S., Smith, A.N., Dancel, M.C.A., Huang, X., Dunn, B.M. and Fanucci, G.E. (2013) Elucidating a Relationship between Conformational Sampling and Drug Resistance in HIV-1 Protease. *Biochemistry* 52:3278–3288. doi: 10.1021/bi400109d

131. Liu, F., Kovalevsky, A.Y., Tie, Y., Ghosh, A.K., Harrison, R.W. and Weber, I.T. (2008) Effect of Flap Mutations on Structure of HIV-1 Protease and Inhibition by Saquinavir and Darunavir. *J Mol Biol* 381:102–115. doi: 10.1016/j.jmb.2008.05.062

132. Harte, W.E., Swaminathan, S., Mansuri, M.M., Martin, J.C., Rosenberg, I.E. and

Beveridge, D.L. (1990) Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. *Proc Natl Acad Sci U S A* 87:8864–8.

133. King, N.M., Prabu-Jeyabalan, M., Bandaranayake, R.M., Nalam, M.N.L., Nalivaika, E.A., Özen, A., Haliloğlu, T., Yılmaz, N.K. and Schiffer, C.A. (2012) Extreme Entropy–Enthalpy Compensation in a Drug-Resistant Variant of HIV-1 Protease. *ACS Chem Biol* 7:1536–1546. doi: 10.1021/cb300191k

134. Hornak, V., Okur, A., Rizzo, R.C. and Simmerling, C. (2006) HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc Natl Acad Sci* 103:915–920. doi: 10.1073/pnas.0508452103

135. Zondagh, J., Achilonu, I. and Sayed, Y. (2018) Molecular dynamics simulations: Video of the wild-type HIV-1 subtype C protease and N37T↑V protease. Retrieved from: https://youtu.be/V4aHhznWxW8

136. Kim, E.Y., Winters, M.A., Kagan, R.M. and Merigan, T.C. (2001) Functional correlates of insertion mutations in the protease gene of human immunodeficiency virus type 1 isolates from patients. *J Virol* 75:11227–33. doi: 10.1128/JVI.75.22.11227-11233.2001

137. Torbeev, V.Y., Raghuraman, H., Hamelberg, D., Tonelli, M., Westler, W.M., Perozo, E. and Kent, S.B.H. (2011) Protein conformational dynamics in the mechanism of HIV-1 protease catalysis. *Proc Natl Acad Sci U S A* 108:20982–7. doi: 10.1073/pnas.1111202108

138. Zhu, Z., Schuster, D.I. and Tuckerman, M.E. (2003) Molecular dynamics study of the connection between flap closing and binding of fullerene-based inhibitors of the HIV-1 protease. *Biochemistry* 42:1326–1333. doi: 10.1021/bi020496s

139. Trylska, J., Tozzini, V., Chang, C. a. and McCammon, J.A. (2007) HIV-1 Protease Substrate Binding and Product Release Pathways Explored with Coarse-Grained Molecular Dynamics. *Biophys J* 92:4179–4187. doi: 10.1529/biophysj.106.100560

140. Cai, Y., Myint, W., Paulsen, J.L., Schiffer, C.A., Ishima, R. and Kurt Yilmaz, N. (2014) Drug Resistance Mutations Alter Dynamics of Inhibitor-Bound HIV-1 Protease. *J Chem Theory Comput* 10:3438–3448. doi: 10.1021/ct4010454

141. Bronowska, A.K. (2011) Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design. In: Thermodynamics - Interaction Studies - Solids, Liquids and

Gases. InTech, pp 1–49

142. Mizuno, A., Matsui, K. and Shuto, S. (2017) From Peptides to Peptidomimetics: A Strategy Based on the Structural Features of Cyclopropane. *Chem - A Eur J* 23:14394–14409. doi: 10.1002/chem.201702119

143. Perryman, A.L., Lin, J.H. and McCammon, J.A. (2006) Optimization and computational evaluation of a series of potential active site inhibitors of the V82F/I84V drug-resistant mutant of HIV-1 protease: An application of the relaxed complex method of structure-based drug design. *Chem Biol Drug Des* 67:336–345. doi: 10.1111/j.1747-0285.2006.00382.x

144. Wodak, S.J. and Janin, J. (1978) Computer analysis of protein-protein interaction. *J Mol Biol* 124:323–342. doi: 10.1016/0022-2836(78)90302-9

145. Halperin, I., Ma, B., Wolfson, H. and Nussinov, R. (2002) Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins Struct Funct Genet* 47:409–443. doi: 10.1002/prot.10115

146. Tang, X., Wang, Z., Lei, T., Zhou, W., Chang, S. and Li, D. (2017) Importance of protein flexibility on molecular recognition: modeling binding mechanisms of aminopyrazine inhibitors to Nek2. Phys Chem Chem Phys. doi: 10.1039/c7cp07588j

147. Meng, X.-Y., Zhang, H.-X., Mezei, M. and Cui, M. (2011) Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des* 7:146–57. doi: 10.1109/TMI.2012.2196707

148. Elokely, K.M. and Doerksen, R.J. (2013) Docking Challenge: Protein Sampling and Molecular Docking Performance. *J Chem Inf Model* 53:1934–1945. doi: 10.1021/ci400040d

149. Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P. and Shenkin, P.S. (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem* 47:1739–1749. doi: 10.1021/jm0306430

150. Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T. and Banks, J.L. (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring.

2. Enrichment Factors in Database Screening. *J Med Chem* 47:1750–1759. doi: 10.1021/jm030644s

151. Jordan, S.P., Zugay, J., Darke, P.L. and Kuo, L.C. (1992) Activity and dimerization of human immunodeficiency virus protease as a function of solvent composition and enzyme concentration. *J Biol Chem* 267:20028–20032.

152. Agniswamy, J., Shen, C., Aniana, A., Sayer, J.M., Louis, J.M. and Weber, I.T. (2012) HIV-1 Protease with 20 Mutations Exhibits Extreme Resistance to Clinical Inhibitors through Coordinated Structural Rearrangements. *Biochemistry* 51:2819–2828. doi: 10.1021/bi2018317