

https://dspace.org/

Biochemical and biophysical analysis of a hinge region protease variant in HIV-1 subtype C

Mokhantso, Tshele

2023-10
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/41666
Mokhantso, Tshele. (2023). Biochemical and biophysical analysis of a hinge region protease variant in HIV-1 subtype C. [PhD thesis, University of the Witwatersrand Johannesburg]. https://hdl.handle.net/10539/41666

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



Biochemical and biophysical analysis of a hinge region protease variant in HIV-1 subtype C

By

Tshele Mokhantso

1147592

A thesis

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

In

Molecular and Cell Biology

Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Prof. Yasien Sayed

Co-supervisor: Dr. Ikechukwu Achilonu

October 2023

Declaration

I, Tshele Mokhantso (1147592), am a student registered for the degree of Doctor of Philosophy (PhD) in the academic year 2023.
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.
- 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: Tshele Mokhantso **Date: 2023/10/01**

Research outputs:

Publications:

Submitted for publication (under review): Subtype-specific HIV-1 protease and the role of hinge and flap dynamics in drug resistance: a subtype C narrative. Sherry, D., Sheik-Ismail, Z., **Mokhantso, T**. and Sayed, Y. (Chapter 1)

Manuscripts in preparation:

i) Microsecond molecular dynamics simulations reveal consequential changes in the flap dynamics of novel variants HIV-1 protease. **Mokhantso, T**., Sherry, D., Achilonu, I., Sayed., Y. (Chapter 4)

ii) Contrasting the effect of hinge region insertions and non-active site polymorphisms on HIV protease-inhibitor interactions: insights from altered flap dynamics. **Mokhantso, T**., Worth, R., Sherry, D., Achilonu, I., Sayed., Y. (Chapter 5)

Conference outputs:

i) Molecular biosciences research Thrust Research Day. November 2021. *Grad flash.* The impact of hinge-region mutations on the dynamics and drug interactions of the South African HIV-1 subtype C protease. **Mokhantso, T.,** Worth, R., Sherry, D., Achilonu, I. and Sayed, Y.

ii) 27th Congress of the South African Society of Biochemistry and Molecular Biology. *Poster.* January 2022. Hinge-region mutations in HIV-1 protease affects dynamics and drug interactions of a south African HIV-1 subtype c protease. **Mokhantso, T**., Achilonu, I and Sayed., Y.

iii) IBM Quantum and WitsQ talk series-Introduction to Quantum Computing. May 2022. *Presentation.* Introduction to Protein Biochemistry and the HIV Protease. Mpho Setshedi, **Tshele Mokhantso** and Akeel Valli

iv) 7th Council of Scientific and Industrial Research Emerging Researchers Symposium. July 2022. *Oral presentation.* Analysis of the impact non-active site mutations in a South African HIV-1 subtype C protease have on conformational stability and drug interactions between HIV-1 protease and protease inhibitors. **Mokhantso, T**., and Sayed., Y.

v) 13th Cross-Faculty Postgraduate Symposium. July 2022. *Oral presentation.* Novel South African HIV-1 protease variant shows reduced conformational stability which may contribute to reduced drug binding. **Mokhantso, T**., and Sayed, Y.

Abstract

HIV-1 protease plays a crucial role in the maturation of the virus by cleaving gag and gag-pol polyproteins. Understanding the structural and functional consequences of mutations in this enzyme is essential for developing effective anti-HIV drugs, especially in the face of emerging drug-resistant variants. This study focused on the N37T↑V^{+10}•D25A mutant, a novel HIV-1 subtype C protease variant harbouring an insertion (↑V) and a substitution (N→T) at position 37, along with 10 naturally occurring polymorphisms. Mutations occurring distal to the active site have long been thought to contribute to drug resistance, with this in mind the study aimed to assess the impact these mutations have on the structure, stability, dynamics and drug binding of HIV-1 protease.

The N37T↑V^{+10}•D25A mutant and wild-type (WT•D25A) HIV-1 protease were overexpressed and purified from inclusion bodies formed in *Escherichia coli* cells using ion-exchange chromatography. Far-UV CD and SE-HPLC analysis showed that N37T↑V^{+10}•D25A exhibited a predominantly β-sheet secondary structure (218 nm trough) and had a homodimeric size of ~23 kDa, respectively, both similar to WT•D25A. Assessment of the local tertiary structure by intrinsic tryptophan fluorescence indicated that the protease retained its tertiary structure in the presence of mutations, with partial exposure of Trp residues (Trp 6, Trp 6', Trp 42/43, and Trp 42'/43'). Overall, the insertion and substitution mutations did not significantly alter the overall structure of HIV-1 protease. However, the conformational stability of N37T↑V^{+10}•D25A was found to be reduced compared to WT•D25A as determined by urea-induced equilibrium unfolding and thermal unfolding experiments. When denatured using urea and temperature, the mutant exhibited a two-state mechanism of unfolding without stable intermediates during the folding and unfolding process. Thermal unfolding experiment determined the melting temperature of N37T↑V^{+10}•D25A as 58 ± 1.2 °C, which is lower than that of the wild type of 62 ± 0.9 °C. This suggests a potential decrease in dimer stability due to the mutations present in N37T↑V^{+10}•D25A. Isothermal titration calorimetry with acetyl pepstatin as a model inhibitor was employed to examine the impact of mutations on drug binding. The enthalpy (ΔH) for WT•D25A and N37T↑V^{+10}•D25A were 35.94 and 36.02 kJ/mol, respectively. The entropy (ΔS) for WT•D25A and N37T↑V^{+10}•D25A was found to be

256.7 and 261.6 J.mol.K, respectively. These differences in thermodynamic parameters between the WT•D25A and N37T↑V^{+10}•D25A proteases, may indicate altered drug-protease interactions. Induced fit molecular docking predicted the binding strengths of both WT•D25A and N37T↑V^{+10}•D25A with nine protease inhibitors (atazanavir, darunavir, fosamprenavir, indinavir, lopinavir, nelfinavir, ritonavir, saquinavir and tipranavir), revealing that specific background mutations, such as P40S present in N37T↑V^{+10}•D25A, significantly decreased the binding energy of the HIV-1 protease to these inhibitors. Molecular dynamics simulations provided insights into the structural dynamics of N37T↑V^{+10}•D25A, showing reduced stability and increased dynamics, particularly in the flaps and hinges of the protease. An increase in flap tip curling and involvement of the cantilever tips were observed to be related to the flap opening mechanism of HIV-1 protease. Interactions between the HIV-1 protease and inhibitors were examined and the radius of gyration and solvent accessible surface area were calculated to evaluate protein compactness and solvent accessibility of the bound inhibitors, respectively. In general, the N37T↑V^{+10}•D25A mutant exhibited decreased compactness when bound to inhibitors, which correlated with the increased solvent exposure of PIs when bound to N37T↑V^{+10}•D25A. This may contribute to the drug resistance mechanisms of the protease, as inhibitors would have difficulty binding the active site and exhibit weaker binding. During the N37T↑V^{+10}•D25A and PI interactions, there was a decrease in hydrogen interactions, which form the basis for protease inhibitor drug-design and these were replaced by water-bridge interactions, which are weaker and can be easily broken. In conclusion, this study provides a comprehensive characterisation of the N37T↑V^{+10}•D25A mutant of HIV-1 protease, shedding light on its structural alterations, conformational stability, drug binding properties, and dynamic behaviour. These findings contribute to our understanding of drug resistance mechanisms in HIV-1 protease and offer valuable insights for the design of more effective inhibitors to combat HIV-1.

Dedication

I dedicate this thesis to my mother, Nthati Mokhantso, who instilled in me the importance of education and knowledge from a young age. Your teachings have shaped me into the person I am today, and I am forever grateful for your unwavering support, friendship, and guidance. Thank you for always believing in me and inspiring me to reach for the stars.

Acknowledgements

First and foremost, I would like to thank the Lord, God almighty for granting me the strength and opportunity to complete this journey. Even when I doubted Him, He did not give up on me.

I extend my heartfelt thanks to my family, and in particular, my brother, Clement Mokhantso, whose guidance and love continue to give me strength every day.

I am also indebted to my friends from Aliwal North, UFS, CRC, UFH, and Wits, who have contributed greatly to the person I am today.

I would like to thank my mentors, Dr. Dean Sherry, Dr. Zaahida Sheik-Ismail, and Dr. Monare Thulo, for their invaluable guidance, patience and support throughout my research journey.

A special thanks to my supervisor, Prof. Sayed, and co-supervisor, Dr. Achilonu, whose guidance, correction, and contribution have not only made me a better scientist but also a better person.

I would also like to thank the National Research Foundation and the Council of Scientific and Industrial Research for financial assistance throughout the course of my research journey.

Finally, I want to thank myself for believing in me, for doing all the hard work, and for never quitting. Thank you, Tshele, for pushing through even when the road was tough. It was all worth it.

Table of Contents

Declaration i

Research outputs: ii

Publications: ii

Manuscripts in preparation: ii

Conference outputs: ii

Dedication vi

Acknowledgements vii

List of figures xiii

List of tables xvi

List of equations xvii

List of abbreviations xviii

Chapter 1: Introduction 1

1.1 HIV/AIDS 1

1.2 Genetic diversity and global molecular epidemiology of HIV 2

1.3 HIV-1 subtype C 4

1.4 The organisation of the HIV genome 6

1.5 Replication cycle of HIV 7

1.6 HIV-1 protease 11

1.6.1 Catalytic mechanism of HIV-1 protease 15

1.6.2 Protease inhibitors 15

1.6.3 Mutations present in HIV-1 protease 19

1.6.4 Insertion mutations in HIV-1 protease 20

1.6.5 N37T↑V^{+10} subtype C protease 21

1.7 Aim and objectives 24

1.7.1 Aim 24

1.7.2 Objectives 25
Chapter 2: Methods 27
2.1 Bacterial transformation of Escherichia coli BL21 (DE3) pLysS cells 27
2.2 Heterologous overexpression of WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease 29
2.2.1 Optimisation of conditions for HIV-1 protease overexpression 30
2.2.2 Induction trials for N37T↑V^{+10}•D25A 30
2.2.3 Overexpression of WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease 31
2.3 Purification of the WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease 31
2.3.1 Purification of WT•D25A 32
2.3.2 Purification of N37T↑V^{+10}•D25A 33
2.4 Purity assessment of the proteases using 16% tricine-SDS-PAGE 34
2.5 Concentration determination of the proteases 35
2.6 Isothermal titration calorimetry 35
2.7 Enzyme activity assay 36
2.8 Structural characterisation 37
2.8.1 Secondary structure 37
2.8.2 Tertiary structure 38
2.8.3 Quaternary structure 41
2.9 Macromolecular crystallisation 42
2.10 Conformational stability and energetics 43
2.10.1 Thermal stability 44
2.10.2 Urea-induced equilibrium unfolding 44
2.10.3 Analysis of urea-induced equilibrium unfolding 45
Chapter 3: Results 48
3.1 Verification of WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease plasmid DNA 48

3.2 Growth curve analysis of WT•D25A and N37T↑V^{+10}•D25A 48
3.3 Expression trials for N37T↑V^{+10}•D25A 48
3.4 Purification of WT•D25A and N37T↑V^{+10}•D25A 52
3.4.1 Purification of WT•D25A 52
3.4.2 Purification of N37T↑V^{+10}•D25A 52
3.5 Purity and molecular concentration determination 52
3.6 Isothermal titration calorimetry 56
3.7 Activity assay 59
3.8 Structural characterisation 59
3.8.1 Secondary structure 59
3.8.2 Tertiary structure 62
3.8.3 Quaternary structure 64
3.8.4 Macromolecular crystallisation 66
3.9 Structural stability of the HIV-1 protease 66
3.9.1 Thermal unfolding 66
3.9.2 Urea-induced equilibrium unfolding 69
Chapter 4: Microsecond molecular dynamics simulations reveal consequential changes in the flap dynamics of novel variants HIV-1 protease 73
4.1 Introduction 73
4.2 Methods 76
4.2.1 Protein and ligand preparation 76
4.2.2 Molecular docking 77
4.2.3 Molecular dynamic simulations 77
4.3 Results and Discussion 78
4.3.1 Comparative conformational stability of WT PR and N37T↑V^{+10} PR 78
4.3.2 RMSD of apo PRs during simulations 79

4.3.3 RMSD of drug-bound PR complexes during simulations 79

4.3.4 Atomic fluctuations of the WT PR and N37T↑V^{+10} PR 81

4.3.5 The apo PR state 83

4.3.6 Drug-bound PR complexes 84

4.3.7 Altered flap dynamics 86

4.4 Conclusion 90

Chapter 5: Contrasting the effect of hinge region insertions and non-active site polymorphisms on HIV protease-inhibitor interactions: insights from altered flap dynamics 91

5.1 Introduction 91

5.2 Methods 94

5.2.1 Homology modelling 94

5.2.2 Ligand docking 94

5.2.3 Molecular dynamics 95

5.2.4 Computational data acquisition 95

5.3 Results and discussion 96

5.3.1 Non-active site mutations in N37T↑V PR and N37T↑V^{+10} PR reduce the predicted binding affinity of PIs 96

5.3.2 Individual non-active site polymorphisms have varying impacts on the binding affinity of several PIs 96

5.3.3 The N37T↑V PR and N37T↑V^{+10} PR can sample wider flap conformations for longer time periods 98

5.3.4 N37T↑V and N37T↑V^{+10} exhibit increased hinge and fulcrum region dynamics 100

5.3.5 Insertion mutations increase the dynamics key regions of the PR when complexed with PIs 102

5.3.6 Cantilever tip angle contraction is essential for flap conformational changes 103

5.3.7 N37T↑V PR and N37T↑V^{+10} PR are less compact and do not fully enclose PIs 106
5.3.8 Alteration in intramolecular salt bridge formation 107
5.4 Conclusion 109
Chapter 6: General summary and discussion 111
6.1 Overexpression of WT•D25A and N37T↑V^{+10}•D25A 112
6.2 Purification of WT•D25A and N37T↑V^{+10}•D25A 113
6.3 Active site titration 114
6.4 Structural characterisation of WT•D25A and N37T↑V^{+10}•D25A 115
6.4.1 Secondary structure 115
6.4.2 Tertiary structure 116
6.4.3 Quaternary structure 117
6.4.4 Macromolecular crystallisation 117
6.4.5 Hinge-region mutations have a minimal impact on PR structure 117
6.5 Conformational stability of N37T↑V^{+10}•D25A 118
6.5.1 N37T↑V^{+10}•D25A folds by a two-state model and has reduced conformational stability 118
6.6 Flap dynamics from a microsecond perspective 120
6.7 Hinge-region mutations result in less favourable binding of inhibitors both enthalpically and entropically 121
6.9 General conclusion 124
Supplementary: 126
References 135

List of figures

Figure 1: Radial phylogeny tree of HIV-1 subtypes under the consensus M group of HIV-1 3

Figure 2: The genome and structure of the HIV virion 5

Figure 3: The replication cycle of the Human Immunodeficiency Virus-type 1 9

Figure 4: A ribbon representation of the crystal structure of the South African HIV-1 subtype C protease. 12

Figure 5: HIV-1 protease active site nomenclature 13

Figure 6: The functional mechanism of HIV-1 protease 16

Figure 7: Chemical structures of FDA-approved HIV-1 protease inhibitors 18

Figure 8: Homology model of N37T↑V+10 protease 22

Figure 9: The pET-11a expression vector containing an HIV-1 protease sequence 28

Figure 10: Location of tryptophan residues in monomers A and B of WT•D25A and N37T↑V^{+10}•D25A 40

Figure 11: Sequence identity of N37T↑V^{+10}•D25A 49

Figure 12: Growth curves of transformed E. coli 50

Figure 13: Induction trials 16% SDS-PAGE tricine gel 51

Figure 14: Two-step purification profile of WT•D25A HIV-1 protease 53

Figure 15: Purification chromatogram of the optimised purification of N37T↑V^{+10}•D25A 54

Figure 16: Tricine 16% SDS-PAGE gel analysis of the WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease purity 55

Figure 17: Protein concentration determination 57

Figure 18: Calorimetric profile of active site titration by acetyl pepstatin 58

Figure 19: Activity assay for WT•D25A and N37T↑V^{+10}•D25A 60

Figure 20: Evaluation of secondary structure of WT•D25A and N37T↑V^{+10}•D25A mutants using far-UV CD. 61

Figure 21: Tertiary structure and the recoverability of WT•D25A and N37T↑V^{+10}•D25A after denaturation with urea (8 M). 63

Figure 22: SE-HPLC chromatogram of N37T↑V^{+10}•D25A PR against standard proteins. 65

Figure 23: Macromolecular crystallisation of N37T↑V^{+10}•D25A 67

Figure 24: Thermal stability of WT•D25A and N37T↑V^{+10}•D25APR. 68

Figure 25: Urea-induced equilibrium unfolding of WT•D25A and N37T↑V^{+10}•D25A monitored by intrinsic tryptophan fluorescence. 70

Figure 26: Urea-induced equilibrium unfolding of N37T↑V+10•D25A monitored by far-UV CD. 72

Figure 27: The HIV-1 protease enzyme 75

Figure 28: Root mean square deviation (RMSD) of WT PR and N37T↑V^{+10} PR 80

Figure 29: Backbone RMSF of WT PR (left) and N37T↑V^{+10} PR (right) 82

Figure 30: 2-D schematic diagram of interatomic interactions between protease inhibitors (ATV and DRV) and the WT PR and the N37T↑V^{+10} PR. 85

Figure 31: Per monomer alpha carbon interatomic distance between the D25 and I50(51) of the WT PR (purple/pink) and N37T↑V^{+10} PR (dark/light grey) 88

Figure 32: Average radius of gyration (Rg) of the WT PR (purple/pink) and N37T↑V^{+10} PR (dark/light grey) in the apo and drug-bound state. 89

Figure 33: A) Ribbon representation of N37T↑V PR (left) and N37T↑V+10 PR (right) and B) Amino acid sequence alignment of the WT subtype C PR, N37T↑V PR and N37T↑V^{+10} HIV PR. 93

Figure 34: Flap tip angle and flap conformations of WT PR (purple), N37T↑V (orange) PR and N37T↑V^{+10} PR (grey). 99

Figure 35: The average RMSF of the regions in the apo WT PR, N37T↑V PR and N37T↑V^{+10} PR HIV-1 protease. 101

Figure 36: Average RMSF of distinct regions of the WT PR, N37T↑V and N37T↑V^{+10} in the apo and drug-bound state..... 104

Figure 37: HIV PR cantilever tip contraction and flap conformational. 105

Figure 38: PI solvent accessible surface area (SASA) and radius of gyration (Rg) of WT PR (purple), N37T↑V PR (orange) and N37T↑V^{+10} PR (grey)..... 108

List of tables

Table 1: Genes and encoded proteins in the HIV genome 8

Table 2: Location of naturally occurring polymorphisms present in N37T↑V^{+10} HIV-1 protease and comparison of the residue charge to the HIV-1 subtype C wild type amino acid residues.. 23

Table 3: Summary of G_{score} of nine FDA-approved PIs in complex with WT PR, N37T↑V PR and N37T↑V^{+10} PR. 97

Table S1: The per region RMSF of the Apo and drug bound WT PR, N37T↑V, and N37T↑V+10 PR. 132

Table S2: Summarised induced fit docking parameters for WT, N37T↑V+10 and N37T↑V PR with clinical inhibitors 133

List of equations

Equation	Equation number	Page
$A = \varepsilon cl$	Equation 1	34
$\varepsilon=5550*Trp+1340*Tyr+150*$ Cys	Equation 2	34
$MRE=\theta x100/cnl$	Equation 3	36
$f_F + f_U = 1$	Equation 4	44
$y_{obs} = (yF \bullet f_F) +(y_U \bullet f_U)$	Equation 5	44
$f_U = (y_F-y_{obs})/(yF- y_U)$	Equation 6	45
$K = [U]^2/[N] = f_U^2/(1-f_U) = f_U/f_F$	Equation 7	45
$\Delta G = -RT \ln(K)$	Equation 8	45
$\Delta G = \Delta G_{water} - m(U)$	Equation 9	45

List of abbreviations

AIDS	Acquired immunodeficiency syndrome
ART	Antiretroviral therapy
ATV	Atazanavir
C_α	Alpha carbon
CM	Carboxyl-methyl
C-SA	South African subtype C
CD4	Cluster of differentiation 4
DEAE	Diethyl aminoethyl
kDa	Kilodalton
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DRV	Darunavir
EDTA	Ethylenediamineteraacetic acid
ES	Enzyme-substrate complex
Far-UV CD	Far-ultraviolet circular dichroism
f_F	Fraction folded
f_U	Fraction unfolded
FDA	Food and Drug Administration
FPV	Fosamprenavir
G_{score}	Glide score
HAART	Highly active antiretroviral treatment
HIV-1	Human immunodeficiency virus type-1
IDV	Indinavir
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
(k_{off})	Off-rate constant
LPV	Lopinavir
m-value	Cooperativity of protein unfolding

MD	Molecular dynamic
NICD	National Institute for Communicable Diseases
NFV	Nelfinavir
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NOP	Naturallly occuring polymorphism
NRTI	Nucleoside reverse transcriptase inhibitor
n	Stoichiometry
N	NATIVE/FOLDED
NVT	
ns	nanosecond
N37T↑V	HIV-1 subtype C protease containing N37T and ↑V mutations of the 37^{th} amino acid position
$N37T\uparrow V^{+10}$	N37T↑V with an additional background mutation
PDB	Protein Data Bank
PI	Protease inhibitor
PR	Protease
PR:PI	Protease-protease inhibitor complex
Rg	Radius of gyration
RMSD	Root-mean square deviation
RMSF	Root-mean square fluctuation
RT	Reverse transcriptase
RTV	Ritonavir
SASA	Solvent exposed surface area
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SE-HPLC	Size exclusion-high performance liquid chromatography
SQV	Saquinavir
TPV	Tipranavir
T_m	Melting/Transition temperature
U	Unfolded

WT	Wild type
ΔG	Change in Gibbs free energy
ΔG_{water}	Free energy in the absence of denaturant for equilibrium unfolding
ΔH	Change in enthalpy
ΔS	Change in entropy

Chapter 1: Introduction

1.1 HIV/AIDS

The human immunodeficiency virus (HIV) was first identified as an etiological agent of the acquired immunodeficiency syndrome (AIDS) in 1983 (Barré-Sinoussi *et al.,* 1983). It has since been described as a *lentivirus* that infects and slowly kills the human hosts' T-helper cells. This weakens the immune system until a state where the immune system is too weak to fend off infectious diseases (Cullen, 1991; Weiss, 1993). This condition is known as AIDS, a chronic disease that causes the human host to be vulnerable to opportunistic infections (Gallo and Montagnier, 1988). As of 2021, ~38.4 million people in the world are living with HIV (HIV positive) (UNAIDS, 2022; World Health Organization, 2022). It has been reported that approximately, 70% of this population is found in sub-Saharan Africa where resources and education are scarce. South Africa remains the epicenter of the HIV/AIDS pandemic, with 20% of the national population being HIV positive (UNAIDS, 2022). There is currently no available cure for HIV, however, therapeutic interventions, antiretroviral therapy (ART) in particular, have been developed to limit the damage caused by the virus.

With increased therapeutic interventions and education, the number of new infections in South Africa has drastically dropped from 510 000 in the year 2000 to 210 000 in the year 2021, (UNAIDS, 2022). This is highlighted by the fact that approximately 74% HIV positive South Africans use ART today. ART in South Africa has been valuable in decreasing AIDS related deaths from an estimated 280 000 in the year 2006 to 51 000 in 2021. However, the recent Covid-19 pandemic that was accompanied by the global lockdown has resulted in a decrease in education and access to ARVs. The World Health Organization, (2022) reports that Covid-19 derailed ART by six (6) months; this is estimated to lead to 500 000 more HIV/AIDS related deaths in sub-Saharan Africa by the end of the year 2023. HIV/AIDS continues to be a major health issue for millions of people worldwide. Since it was first identified finding an effective cure against HIV has long evaded researchers and this has primarily been attributed to the virus's large genetic diversity.

1.2 Genetic diversity and global molecular epidemiology of HIV

HIV has high genetic variability which has proven to be a major problem in the development of effective HIV treatment (Hu *et al.*, 1996; Konvalinka *et al.*, 1995). This is mainly due to the high replication F the virus and the low fidelity of the reverse transcriptase (RT) (Holec *et al.*, 2017; Shafer *et al.*, 2000). RT, an enzyme responsible for the transcription of viral ribonucleic acid (RNA) to viral deoxyribonucleic acid (DNA), lacks post-transcription proofreading mechanisms. This is assumed to drive the emergence of new mutations and consequently, novel variants as mutations become conserved (Shafer *et al.*, 2007; Winters and Merigan, 2005). Additionally, the viral recombination of two RNA molecules creates a mosaic of viral DNA, further altering the HIV genome and driving HIV evolution (McCutchan, 2006). It is thus of great interest to understand the genetic diversity of HIV, as it influences diagnosis as well as the design and development of antiretroviral drugs (Collier *et al.*, 1996; de Oliveira *et al.*, 2003; Sawant *et al.*, 2008).

HIV is classified as either type 1 or type 2 (HIV-1 and HIV-2, respectively) (Ali *et al.*, 2010; Taylor *et al.*, 2008). The predominant type, HIV-1, is classified into groups: group M (main), group N (non-M/non-O), group O (outlier) and group P (pending identification of other human cases). Group M is the most virulent and is responsible for the majority of global infections (Bessong, 2008). Within this group, nine subtypes (A, B, C, D, F, G, H, J, and K) and circulating recombinant forms (CRFs) exist (Robertson *et al., 2000; Taylor et al., 2008)*. The nine HIV-1 group M subtypes are shown in Figure 1 and the variants of each subtype are shown based on the country of origin.

Current HIV treatment was designed based on the viral proteins of a group M subtype, HIV-1 subtype B (Lv *et al.*, 2015; Mosebi *et al.*, 2008; Naicker and Sayed, 2014). HIV-1 subtype B is the prevalent subtype in North America, Western Europe and Australia. It is, however, not the most widespread subtype in the world. The most prevalent is HIV-1 subtype C which accounts for more than 50% of global HIV-1 infections (Mosebi *et al.*, 2008; Naicker and Sayed, 2014; Thomson and Nájera, 2001; UNAIDS, 2022; Velazquez-Campoy *et al.*, 2003). Furthermore, based on the phylogeny tree depicted in Figure 1*,* HIV-1 subtype C has evolved into more variants than the consensus

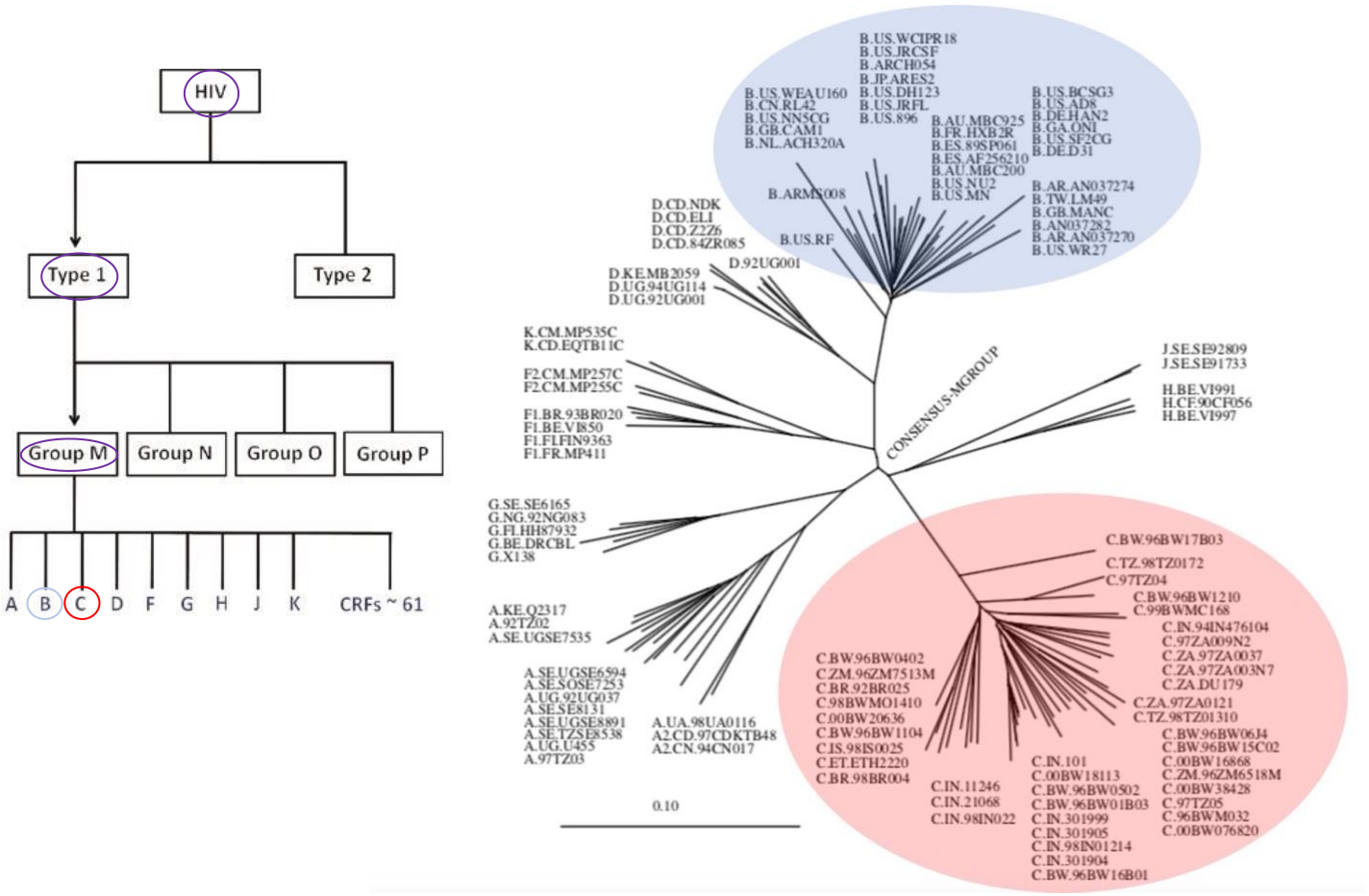


Figure 1: Radial phylogeny tree of HIV-1 subtypes under the consensus M group of HIV-1.

The identity code for each subtype is as follows. Subtype indicated by a letter, then country, then identity code. Made on www.hiv.lanl.gov. Phylogeny tree made by Tshele Mokhantso with NJ method and a F84 distance model with equal site rate, with 43 sequences and 7963 characters including insertions and gaps stripped.

subtype B. Subtype C is predominantly found in sub-Saharan Africa, Brazil, India and China. Although drugs that were developed based on HIV-1 subtype B have shown to improve the quality of life and decrease HIV infections for millions of people, there have been questions on how effective these drugs are on non-subtype B infected patients (Lv *et al.*, 2015; Mosebi *et al.*, 2008; Naicker and Sayed, 2014; Velazquez-Campoy *et al.*, 2002; Velazquez-Campoy *et al.*, 2003). It is troubling that all available drugs are based on a subtype (subtype B) that is mainly found in Western countries, which arguably have more access and financial capital for education, research, drug design and drug distribution in comparison to less developed countries. This imbalanced allocation of research and resources may give rise to drug-resistant variants from non-subtype B variants in less developed countries, which may, in time, affect the entire world.

1.3 HIV-1 subtype C

HIV-1 subtype C accounts for more than 56% of global HIV-1 infections (UNAIDS, 2022) as stated above. The HIV-1 subtype C was first identified in South Africa and Ethiopia and has since rapidly spread and become the predominant subtype in Southern and Eastern Africa. Today, subtype C is also the prevalent subtype in India, China and Brazil and a larger sense, the world (UNAIDS, 2022). The widespread nature of HIV-1 subtype C remains of great interest. This may be related to the infected host, the virus and socio-economic factors. Researchers have previously hypothesised that this subtype may have increased viral fitness, replicative advantage and is more easily spread (Bessong, 2008; Mosebi *et al.*, 2008; Naicker and Sayed, 2014; Taylor *et al.*, 2008; Velazquez-Campoy *et al.*, 2001).

HIV-1 subtype C possesses unique genetic and phenotypic characteristics that distinguish it from other HIV-1 subtypes. The HIV-1 subtype C virion (Figure 2) has an elongated nuclear factor-κB (NF-κB) binding site at the long terminal repeat (LTR) (De Arellano *et al.*, 2005; Rodenburg *et al.,* 2001; Rodriguez *et al.*, 2007). LTR is the region where transcription is initiated and houses binding sites for various transcription

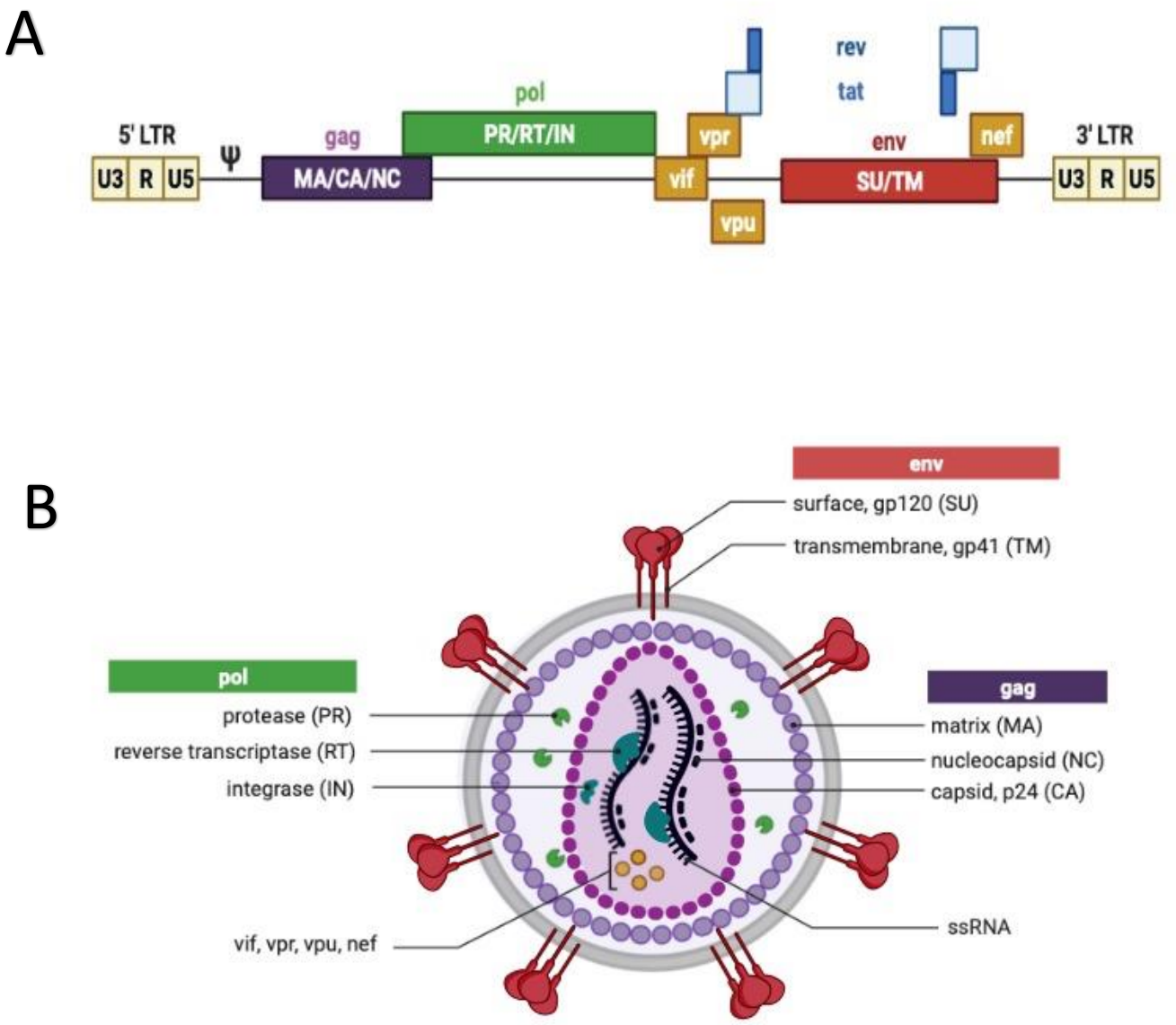


Figure 2: The genome and structure of the HIV virion

A) A schematic representation of the HIV genome reading frames. **B)** The structure of the HIV virion. The *gag* gene encodes for the structural proteins; matrix (MA), nucleocapsid (NC) and capsid (CA) proteins. The *pol* gene encodes the functional proteins; protease (PR), reverse transcriptase (RT) and integrase (IN). The env protein encodes the glycoproteins; surface (SU) (gp120) and transmembrane (TM) (gp41) proteins. In yellow are the regulatory proteins; namely, vif, vpr, vpu and nef proteins. Image adapted from Freed, (1998) to show the approximate location of HIV-1 viral proteins.

factors, such as NF-κB. LTR is responsible for the regulation of the expression of the viral genome. A longer LTR allows more NF-κB molecules, which are essential for the activation of HIV transcription initiation, to bind the LTR (Naghavi *et al.*, 1999; Rodenburg *et al.*, 2001; Nabel and Baltimore 1987; Gerritsen *et al.*, 1997; Mbonye and Karn, 2014). This strongly relates to an increase in gene expression and altered levels of transmissibility and pathogenesis of the subtype C virus (Gorden *et al.*, 2003; Wong and Jong, 2020; Gao *et al.,* 1998). HIV-1 subtype C has a viral protease that differs from other subtypes protease. This enzyme is essential for the hydrolysis of precursor polyproteins into functional enzymes and structural proteins. It has previously been shown that HIV-1 subtype C protease has an increased catalytic activity in comparison to other subtypes (Bessong, 2008; Velazquez-Campoy *et al.,* 2003). The increased catalytic activity may be essential for the observed viral superiority of HIV-1 subtype C protease.

It is unfortunate that HIV-1 subtype C, with all its distinct deviations from other subtypes, is still relatively under-studied. This is evident when accessing the Protein Data Bank (PDB) which contains many subtype B solved crystal structures, but less than 10 crystal structures of HIV-1 subtype C protease. Furthermore, all FDA-approved drugs and the majority of drug discovery is focused on HIV-1 subtype B. It is now essential to focus on the subtype found in the majority of patients with HIV.

1.4 The organisation of the HIV genome

HIV is classified as a retrovirus, belonging to the family *Retroviridae.* It stores its genetic information in the form of ribonucleic acid (RNA) typical of all lentiviruses (Cullen, 1991; Sawant *et al.*, 2008). Like most retroviruses, RNA must be reverse transcribed into viral DNA within the host cell for viral propagation. The HIV genome is mainly a diploid single-stranded RNA (dsRNA) molecule that is enclosed within a lipid bilayer. The lipid bilayer is synthesised using the host cell membrane and has a glycoprotein (gp120) on the surface. The gp120 protein is anchored onto the membrane by a transmembrane protein (gp41) (Prabakaran *et al.*, 2007). The membrane is lined by the matrix (MA) and capsid (CA) (Anderson *et al.*, 1986; Barré-Sinoussi, 1996; McCutchan, 2006). An assembled HIV virion is infective and is referred to as an HIV virion, *figure 2* (Rodriguez *et al.*, 2007).

The proviral DNA is flanked by two long terminal repeats (LTR) sequences on both the 5'-end and the 3'-end, *figure 2* (De Arellano *et al.*, 2005; Rodriguez *et al.*, 2007). The 5' LTR encodes for the promoter, which promotes transcription of HIV genes in the 5' to 3' direction of *gag*, *pol*, and then *env* genes. The genome consists of three genes namely: group-specific antigen (*gag*), polymerase (*pol*) and envelope (*env*) genes, which code for proteins that are necessary for the structure and function of HIV. The major structural components of the HIV virion (Figure 2) namely; the matrix (MA), capsid (CA) and nucleocapsid (NC) are encoded by the *gag* gene (Barré-Sinoussi, 1996). The *pol* gene encodes three enzymes: reverse transcriptase (RT), integrase (IN) and protease (PR). These three enzymes are essential components of the HIV viral replication cycle, and as such, drug discovery has focused on targeting and inhibiting the function of these enzymes (Shafer *et al.,* 2000). Neighbouring the pol gene is the *env* gene, which encodes for envelope glycoproteins gp120 and gp41. The HIV genome also includes regulatory and accessory/auxiliary proteins. The regulatory proteins are coded for by the *tat* and *rev* genes and the auxiliary proteins by the *nef, vif, vpr*, *vpu* and *vpx* genes. Table 1 gives a list of HIV-1 genes and the corresponding proteins that they encode.

1.5 Replication cycle of HIV

The replication of HIV is a complex multistep process in which the virus manufactures new copies of itself. The replication cycle (Figure 3) below gives a summary of the steps that make up the replication process. HIV replication starts with the attachment of the virion to the targeted host cells known as T lymphocytes or helper T cells. T cells also termed CD4+ cells because they possess CD4 antigens on the plasma membrane surface, and these serve as binding sites for the virion. This leads to the binding of HIV to the host cell by the interaction of the viral and cellular membranes. The viral envelope glycoproteins, gp120 and gp41 bind to the chemokine receptors, CCR5/CXCR4,

Table 1: Genes and encoded proteins in the HIV genome

Genes in the HIV genome along with the corresponding proteins that they encode (Faust *et al.*, 2017; Li *et al.*, 2015; Prabakaran *et al.*, 2007).

Gene	Protein	Protein function
GAG	Matrix (p24)	Important for viral assembly on the plasma membrane
	Capsid (p17)	Forms the conical capsid
	Nucleocapsid (p7)	Forms the nucleoprotein: RNA complex
	P6 proteins	Responsible for release of viral particles
POL	Protease (p10)	Cleaves the Gaga and Gag-Pol polyproteins to release activated viral components
	Reverse transcriptase (p51)	Transcribes HIV RNA into viral DNA (proviral DNA)
	Rnase H (p15) (p66)	Degrades viral RNA in the viral RNA/DNA replication complex
	Integrase (p32)	Integrates the viral DNA into host genome
ENV	Surface glycoprotein (gp120)	Binds virion to host cell receptors
	Transmembrane glycoprotein (gp41)	Mediates the fusion of viral membrane to the plasma membrane of the host cell
TAT	Transactivator protein (p14)	Regulates HIV gene expression by activating transcription and elongation from LTR promoter
REV	mRNA splicer regulator (p19)	Regulates nuclear export of viral mRNA
VIF	Viral infectivity factor (p23)	Promotion of infectivity of viral particles
VPR	Viral protein R (p14)	Regulates integration, cell growth and transactivation of cellular genes
VPU	Viral Protein U (p16)	Degrades CD4 cells in endoplasmic reticulum and enhances release of HIV-1 virion from the plasma membrane
NEF	Negative Factor (p27)	Down regulates CD4 cell production and increases viral infectivity and disease progression
VPX	Viral Protein X (p15)	Essential for efficient replication of HIV-2 and progression to AIDS
LTR	Long terminal repeat	Regulates initiation of transcription and polyadenylation

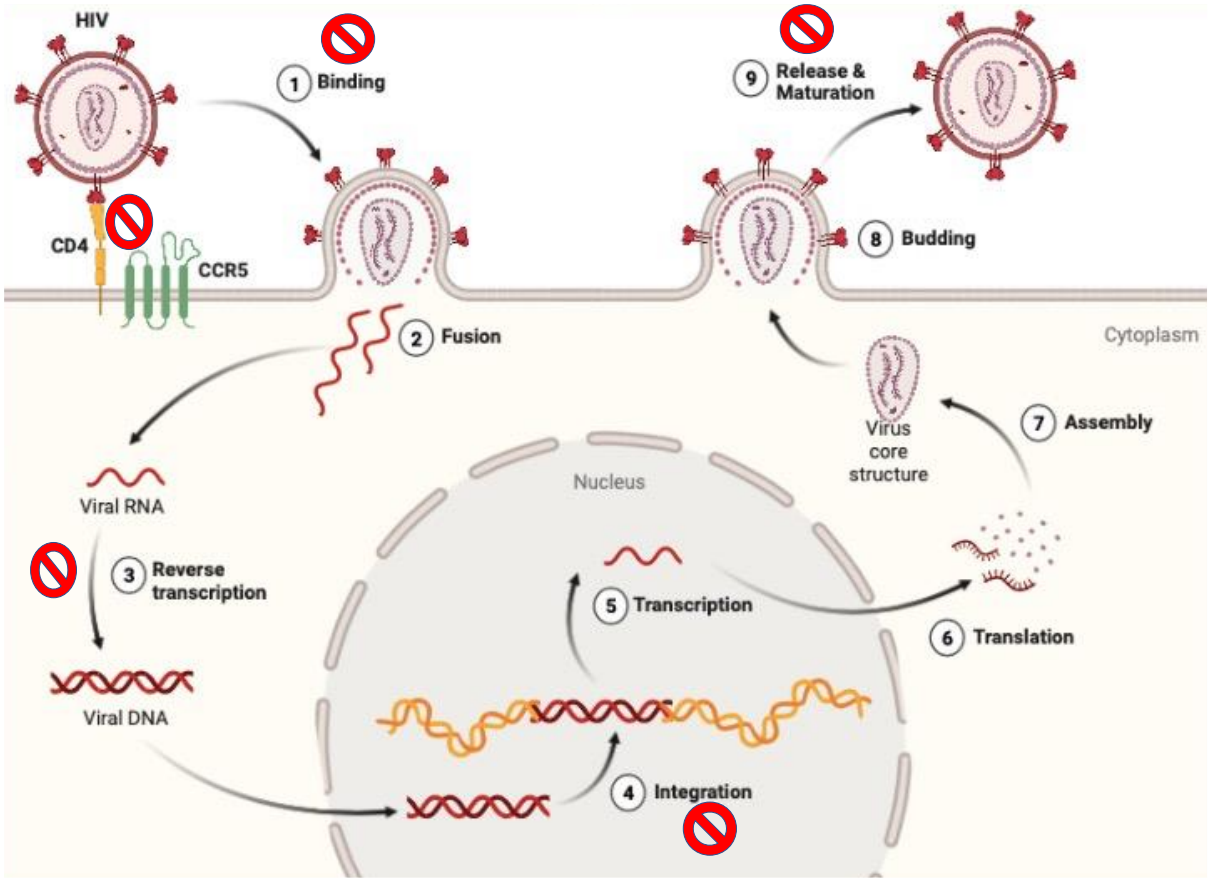


Figure 3: The replication cycle of the Human Immunodeficiency Virus-type 1

(1) The attachment of HIV to the host cell is the initial step of infection. (2) This is followed by endocytosis, allowing the virus into the cell. The protein envelopes are then removed exposing the viral RNA. (3) Reverse transcriptase reverse transcribes the viral RNA to form viral double-stranded DNA (dsDNA). (4) The viral dsDNA migrates to the nucleus, where it is integrated into the host cell genome by integrase. (5) The viral DNA is then transcribed to form viral RNA. (6) The viral RNA migrates out of the nucleus, into the cytoplasm where it is translated into polyproteins. (7) Assembly of new virions follows, which (8) leave the cell by budding. (9) The final step of HIV replication is maturation, here the protease cleaves precursor polyproteins to form mature viral components. The virion is now infective and seeks out other cells to infect to start the cycle again (Freed, 1998; Kirchhoff, 2013; Ratner, 1993).

located on the surface of the cell (Brik and Wong, 2002; Freed, 1998; Kirchhoff, 2013; Ratner, 1993; Dalai *et al.*, 2009; Taylor *et al.*, 2008). This binding renders the cell susceptible to HIV entry by fusion. Fusion is the merging of the plasma membrane and the viral membrane. Following fusion, the viral core which contains the HIV genetic material, single-stranded RNA, is then injected into the cell cytoplasm. HIV therapy has long targeted this step by blocking CCR5 co-receptor, preventing fusion and inhibiting HIV entry.

Once inside the cytoplasm, the viral single-stranded RNA is transcribed to a single-stranded DNA, which is in turn used as a template for the synthesis of a complementary DNA molecule, forming a double-stranded DNA (dsDNA). This process is an essential part of HIV replication and is mediated by HIV reverse transcriptase (Shafer *et al.,* 2000; Winters and Merigan, 2005). This process is called "reverse transcription" because transcription occurs in the opposite direction of the central dogma of life. Inhibiting reverse transcription has been the main aim of ART drugs such as NRTIs and NNRTIs.

The newly formed dsDNA, now termed proviral DNA, moves into the nucleus of the cell, where it is integrated into the host's genetic material (DNA) by integrase. Integrase cuts host cell DNA and inserts proviral DNA into different sites on the host cell DNA. The viral DNA is now part of the cell's genetic material and is called the provirus (Freed, 2001). The provirus at this stage may lay dormant for a long time until the cell is activated. This is a major concern in the fight against HIV as it is undetectable at this point. The provirus is then transcribed using host cell machinery into genomic mRNA. The viral mRNA is translated to viral proteins such as Gag, Gag-Pol and Env polyproteins (Freed, 1998; Shafer *et al.,* 2000). The newly formed viral RNA and proteins assemble to form new virions. New HIV virions leave the cell by budding-off the host-cell membrane. Maturation is the final step, in which HIV protease cleaves the polyproteins, releasing viral functional proteins. The virion is now in its mature, infective form and free to infect other cells (Freed, 1998).

The virions proceed to bud off the cell in order to be free to infect new cells, forming more copies of the virus and continuing the replication cycle. The infected cells genetic systems are damaged by the virus, and this leads to cellular apoptosis (cell death)

(Kirchhoff, 2013; Ratner, 1993). Antiretroviral drugs have been synthesised to inhibit the major steps in the virus's lifecycle. Current drugs mainly target the reverse transcriptase, integrase and protease, viral proteins that are responsible for the replication cycle of HIV (Sawant *et al.*, 2008). Drugs have been developed to inhibit the release of the virus (e.g., capsid inhibitor, lenacapavir, approved in August 2022 (Gilead Sciences Inc.)).

1.6 HIV-1 protease

HIV-1 protease (PR) is an aspartyl protease that is responsible for the post-translational cleavage of Gag and Gag-Pol precursor proteins (Debouck, 1992; Freed, 1998). This is an essential step in the replication cycle of HIV as it regulates the maturation of the HIV virion to a mature infective form (Peng *et al.*, 1989). HIV-1 PR cleaves the Gag and Gag-Pol polyproteins at nine distinct peptide sites and releases/activates structural and functional proteins. The cleavage of the Gag polyprotein results in the release of the matrix, capsid, nucleocapsid and p6 proteins. While the reverse transcriptase, integrase and protease are released by the cleavage of Gag-Pol polyprotein (Barrie *et al.*, 1996; Erickson-Viitanen *et al.*, 1989; Tritch *et al.*, 1991; Brik and Wong, 2002).

In its functional form, HIV-1 PR is a homodimer that is 22 kDa in size and is made up of two identical monomers that each contain 99 amino acid residues (Meek *et al.*, 1989; Weber and Agniswamy, 2009). Like all aspartyl proteases, the protease contains the characteristic catalytic triad, Asp25-Thr26-Gly27 in monomer A and Asp25'-Thr26'-Gly27' in monomer B (Mager, 2001; Spinelli *et al.,* 1991). The catalytic triad forms part of the active site where post-translational processing of viral polyproteins occurs (Baum *et al.*, 1990; Darke *et al.*, 1989). The HIV-1 protease active site nomenclature is shown in figure 5. The catalytic triad is central to the activity of the enzyme and any modification would deem the PR inactive. Asp25/25' is responsible for the hydrolysis of the substrate while Thr26/26' is responsible for the stabilisation of the conformational state of the active site. Gly27/27' role is to accommodate and bind the substrate in a manner that it is accessible to Asp25/25' (Mager, 2001). The catalytic triad is essential for the function of the protein and as

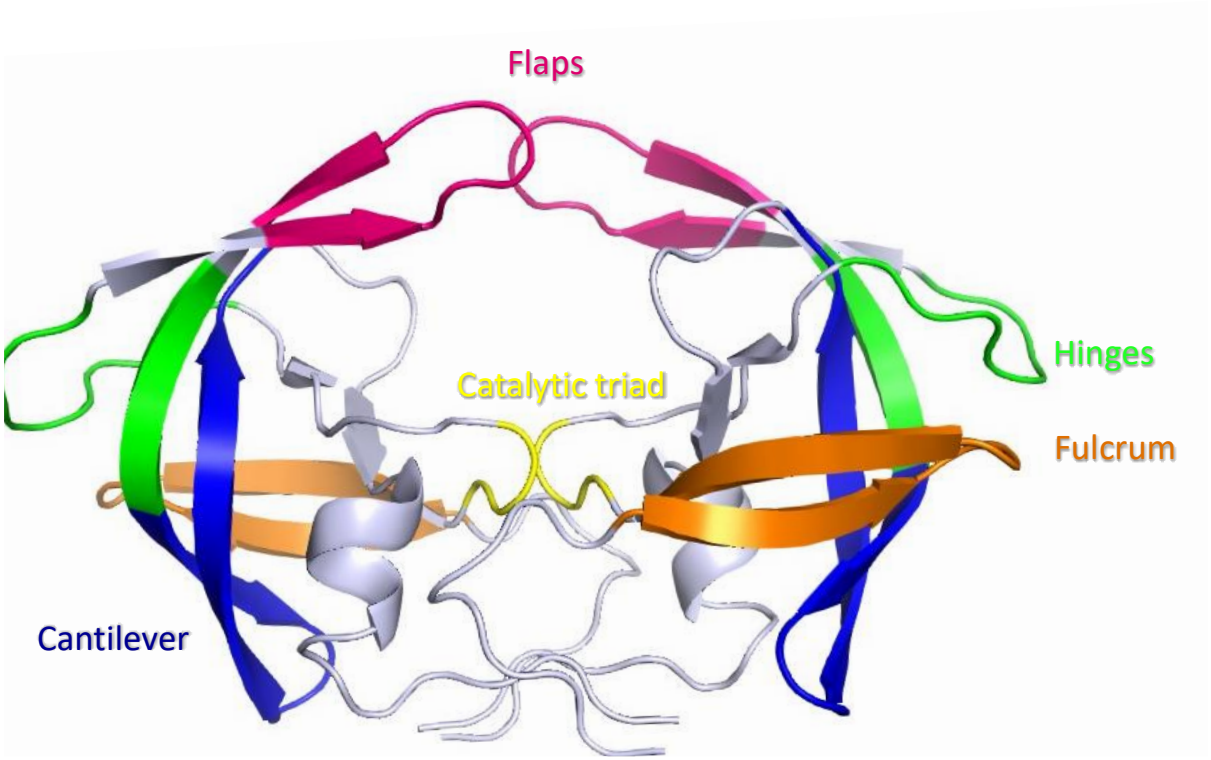


Figure 4: A ribbon representation of the crystal structure of the South African HIV-1 subtype C protease.

The flap region (residues 46-54) is shown in magenta; the hinges (residues 35-42 and 57-61) are in green. The cantilever (residues 62-75) and fulcrum (residues 10-23) are shown in blue and orange, respectively. The catalytic triad is shown in yellow (residue 25-27). The figure was generated using PyMOL Molecular Graphics Software, version 1.8 (Schrödinger, LLC), using data obtained from the Protein Data Bank (PDB ID: 3U71; Naicker *et al*., 2014).

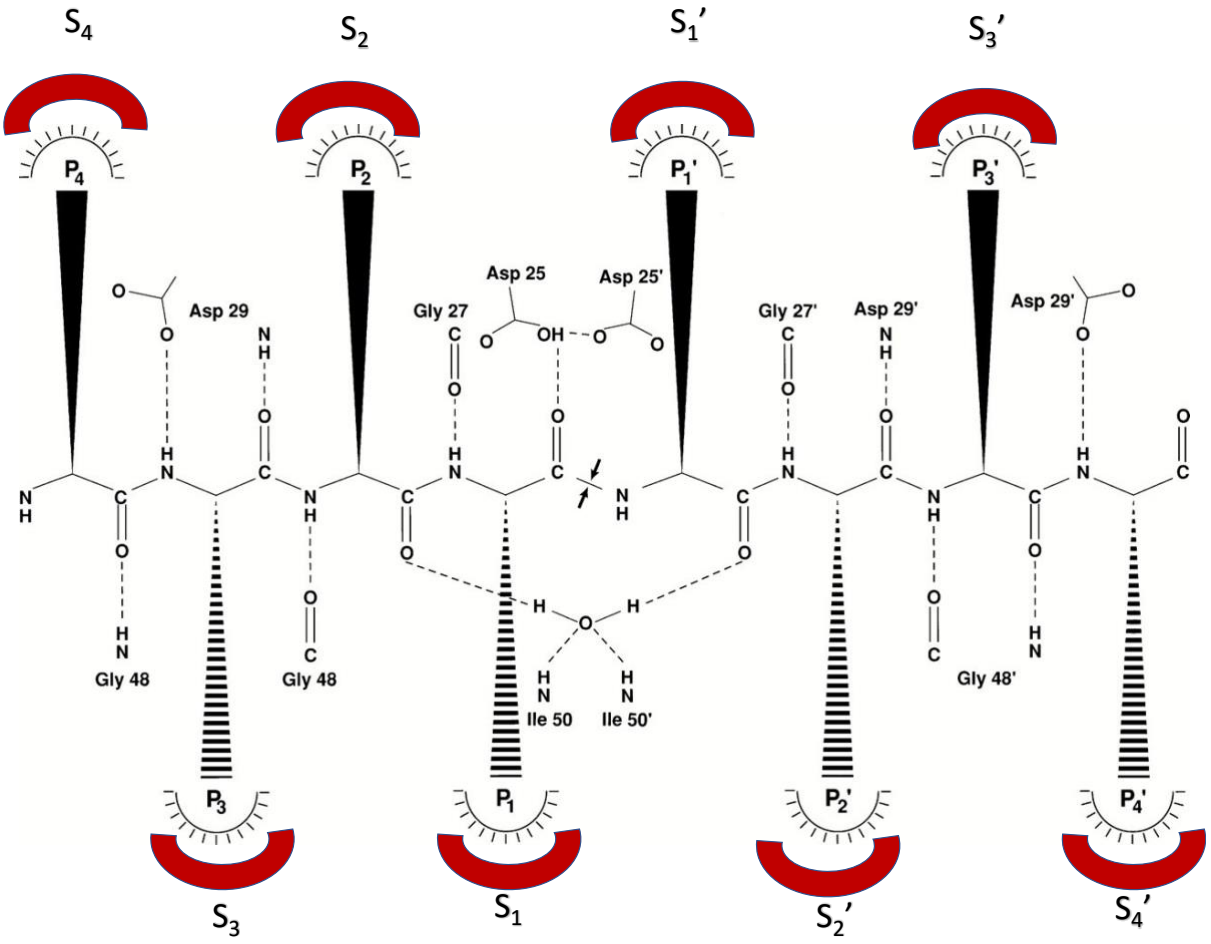


Figure 5: HIV-1 protease active site nomenclature

Schematic diagram depicting the nomenclature used to assign amino acid residues of peptide substrates (P) and the corresponding binding site (S) on HIV-1 protease molecule (Schechter and Berger, 1967; Wlodawer and Vondrasek, 1998). The active site of the enzyme contains subsites represented as S4 - S1 and S1' - S4' (red arc), which correspond to the regions towards the N- or C-terminus of the scissile bond, respectively. Similarly, the substrate sites are denoted as P4 - P1 and P1' - P4'. The P1 and P1' sites of the substrate bind in a lock-and-key relationship with the S1 and S1' subsites of the PR active site, fitting precisely into these specific regions.

such, most protease inhibitors (PIs) have been designed to target it within the binding pocket (Lv *et al.*, 2015).

The binding pocket (hydrophobic core) is made up of 20 amino acids in each monomer; namely, L5, V11, I13, V15, I19, A22, L24, I26, L33, L38, I62, I64, I66, V75, V77, I85, M89, L90, L93 and L97 (Mittal *et al.*, 2012; Sherry *et al.*, 2022). These residues do not touch the substrate or the inhibitors, however, alteration in these residues has previously been linked to drug resistance (Mittal *et al.*, 2012). This is because these residues take part in hydrophobic sliding, a multi-step process in which the residues rearrange through concerted sliding that is facilitated by the interdependent exchange of van der Waals contacts between hydrophobic sidechains to allow opening and closing of the binding pocket. It is thus essential to conserve the hydrogen bonding network within the core to ensure the uninterrupted opening and closing of HIV-1 protease (Foulkes-Murzycki *et al.*, 2007; Mittal *et al.*, 2012; Sherry *et al.*, 2022).

The general secondary structure of an HIV-1 PR monomer has two antiparallel ß-strands and one α-helix. Each monomer of the HIV-1 PR (Figure *4*) comprises the flap region (residue 46-54), flap hinges (residue 35-42), cantilever (residue 62-75) and fulcrum (residue 10-23) (Meher and Patel, 2013; Neuman and Liotta, 2008). These regions function together to bind and cleave the substrate, resulting in the release of mature viral proteins (Spinelli *et al.*, 1991). The flaps exist in a dynamic equilibrium between closed, semi-open and open conformations, with the latter allowing for substrates or inhibitors into the active site. The closed flap conformation covers the active site, containing the typical catalytic triad (Figure 1A), and promotes the formation of intermolecular interactions which are important in substrate recognition and hydrolysis (Liu *et al.*, 2008a; Scott and Schiffer, 2000). Opening and closing of the flap region are thus important for the entrance and exit of substrate and products, respectively. The flap region directly relies on the hinge region for both stability and movement (Naicker *et al.*, 2013; Scott and Schiffer, 2000; Zondagh *et al.*, 2018). It has been suggested that an increase in flap flexibility can significantly reduce substrate proteolysis as well as susceptibility to protease inhibitors (Naicker *et al.,* 2013; Zondagh *et al.,* 2018). It is thus, of paramount importance to investigate and

understand how mutations and insertions in the hinge region may affect the flexibility and structure of the flap region.

1.6.1 Catalytic mechanism of HIV-1 protease

The most widely accepted catalytic mechanisms for HIV-1 PR action involves the HIV-1 PR recognising the substrate scissile bond, binding it and forming an enzyme-substrate (ES) complex (Schechter and Berger, 1967; Suguna *et al.*, 1987). The two aspartic acid residues, Asp25 and Asp25', in HIV-1 PR exist as oppositely protonated amino acid residues in each monomer, respectively. These residues act as a general acid-base reaction and activate a nearby water molecule which then attacks the scissile bond as a nucleophile. The nucleophilic attack converts the ES complex into a gem-diol intermediate (ES*). This then leads to the breaking of covalent linkages present in the substrate and results in the formation of two products, P1 and P2 (Suguna *et al.*, 1987). These products a subsequently released, allowing the PR to hydrolyse another substrate. The schematic representation of the catalytic/functional mechanism of HIV-1 protease is shown on Figure 6 below.

1.6.2 Protease inhibitors

HIV-1 protease inhibitors (PIs) are compounds that function as competitive inhibitors of HIV-1 and compete with the natural substrate for the active site of the PR (Bardi *et al., 1997;* Lv *et al.,* 2015*;* Wlodawer and Vondrasek, 1998*)*. PIs are mostly tetrahedral carbon-containing compounds at a position that mimics a sequence found on the natural PR substrate. By binding to the PR active site, the PIs inhibit the enzymatic activity of the HIV-1 PR (Meek *et al.*, 1989). This prevents the substrate from being cleaved and thus limits infectivity and formation of viral particles, halting viral replication (Bessong, 2008; Lv *et al.*, 2015). PIs are one of the most widely used antiretroviral drugs and have improved the lives of many people worldwide. In highly active antiretroviral therapy (HAART), PIs are considered to be the most potent drugs against HIV and have been instrumental in the decrease of mortality in HIV-1 positive individuals (Bartlett *et al.*, 2001; Reddy *et al.*, 2007).

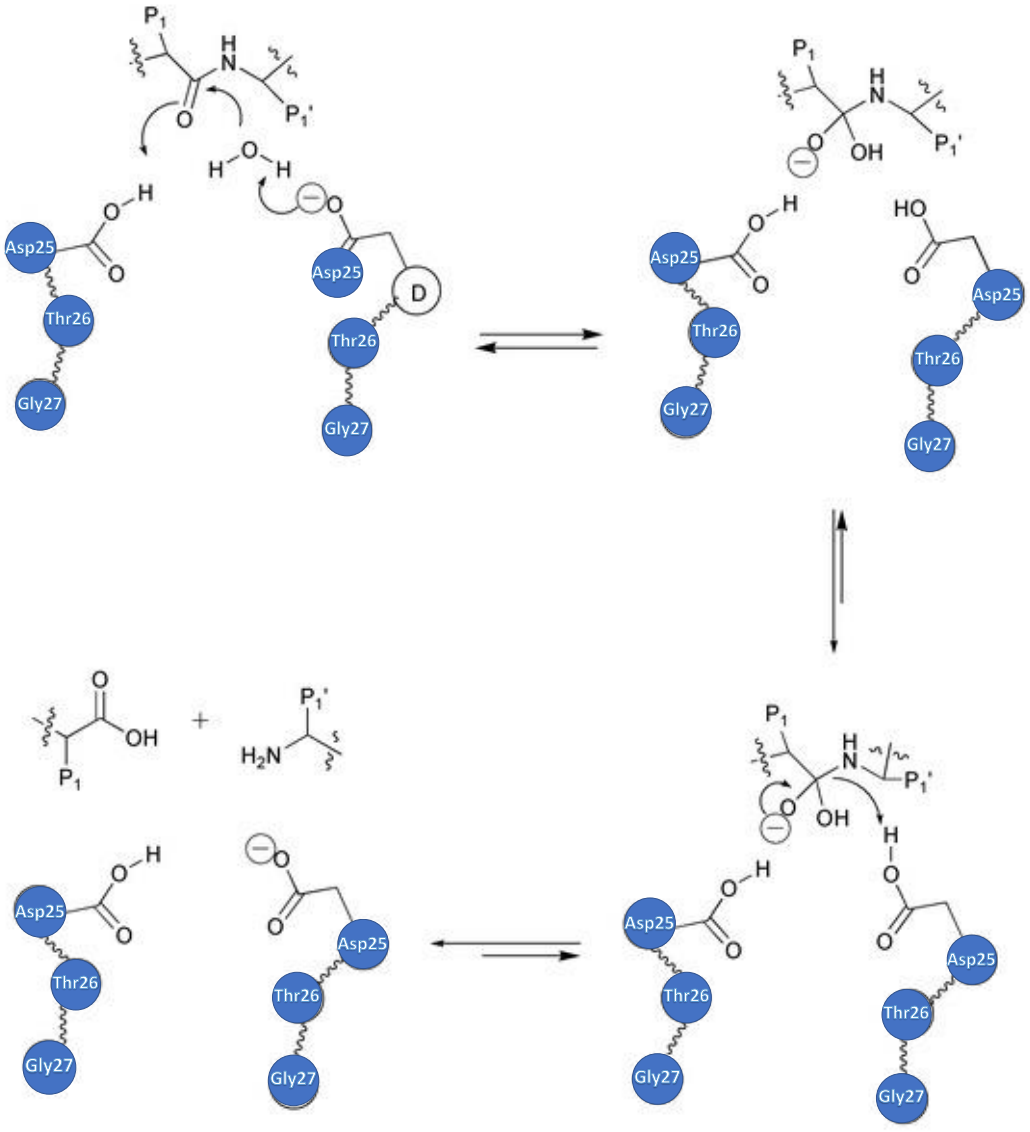


Figure 6: The functional mechanism of HIV-1 protease

Two Asp25 residues that are located on opposite sides of the peptide bond present in the active site. One Asp25 acts as a general base and activates a water molecule into a nucleophile which attacks the carbonyl carbon of the scissile amide bond found on the substrate. The other Asp25' acts as a general acid and protonates the leaving amine product. Image adapted from Brik and Wong (2002).

The first-generation of HIV-1 PIs were designed as peptidomimetic molecules that have polar groups that mimic those found on the natural substrate peptide chain (Lv *et al.*, 2015). They further contain hydroxyethylene or hydroxyethylamine isosteres which increase inhibitor specificity to HIV-1 PR, while not undergoing peptide cleavage (Reddy *et al.*, 2007; Wlodawer and Vondrasek, 1998). The first-generation PIs bind to HIV-1 PR in an entropically driven manner, while experiencing entropic resistance (Velazquez-Campoy *et al.*, 2001). Although effective, first-generation PIs had disadvantages due to their peptide-like structure. The PIs had low bioavailability, were not metabolically stable and had vicious side effects (Arts and Hazuda, 2012; Vlahov *et al.*, 2005; Wlodawer and Vondrasek, 1998). With the increased emergence of drug-resistant strains, it was then imperative to improving the drug design approach to focus on increasing the drug resistance barrier, improve bioavailability and decreasing drug toxicity (Clavel and Hance, 2004; Nijhuis *et al.*, 2007b). This led to the development of second-generation PIs.

Second-generation protease inhibitors were designed to have properties that bind the HIV-1 PR in a manner that is both enthalpically and entropically favourable. These inhibitors had fewer peptide backbone features but still contained the hydroxyl group. The carbonyl group in first-generation PIs was replaced with a sulphonamide group in second-generation PIs. These have been reported to be less likely to cause drug resistance in HIV-1 PR compared to first-generation PIs. This is because primary mutations do not alter the backbone of the HIV-1 PR and thus do not change the binding pocket structure. Therefore, it is still available for binding of second-generation PIs.

Structure-assessed inhibitor drug design has been the major method used for developing PIs. It has made use of crystallographic techniques, NMR and computational biochemistry to design effective PIs against the HIV-1 PR. This has resulted in the development of several FDA-approved inhibitors; namely, saquinavir, indinavir, ritonavir, nelfinavir, amprenavir, fosamprenavir, lopinavir, atazanavir, tipranavir, and darunavir (Lv *et al.*, 2015; Wlodawer and Vondrasek, 1998). However, due to polymorphisms and increased emergence of drug resistance, the effectiveness of the inhibitors is on the decline (Mosebi *et al.*, 2008). HIV-1 protease

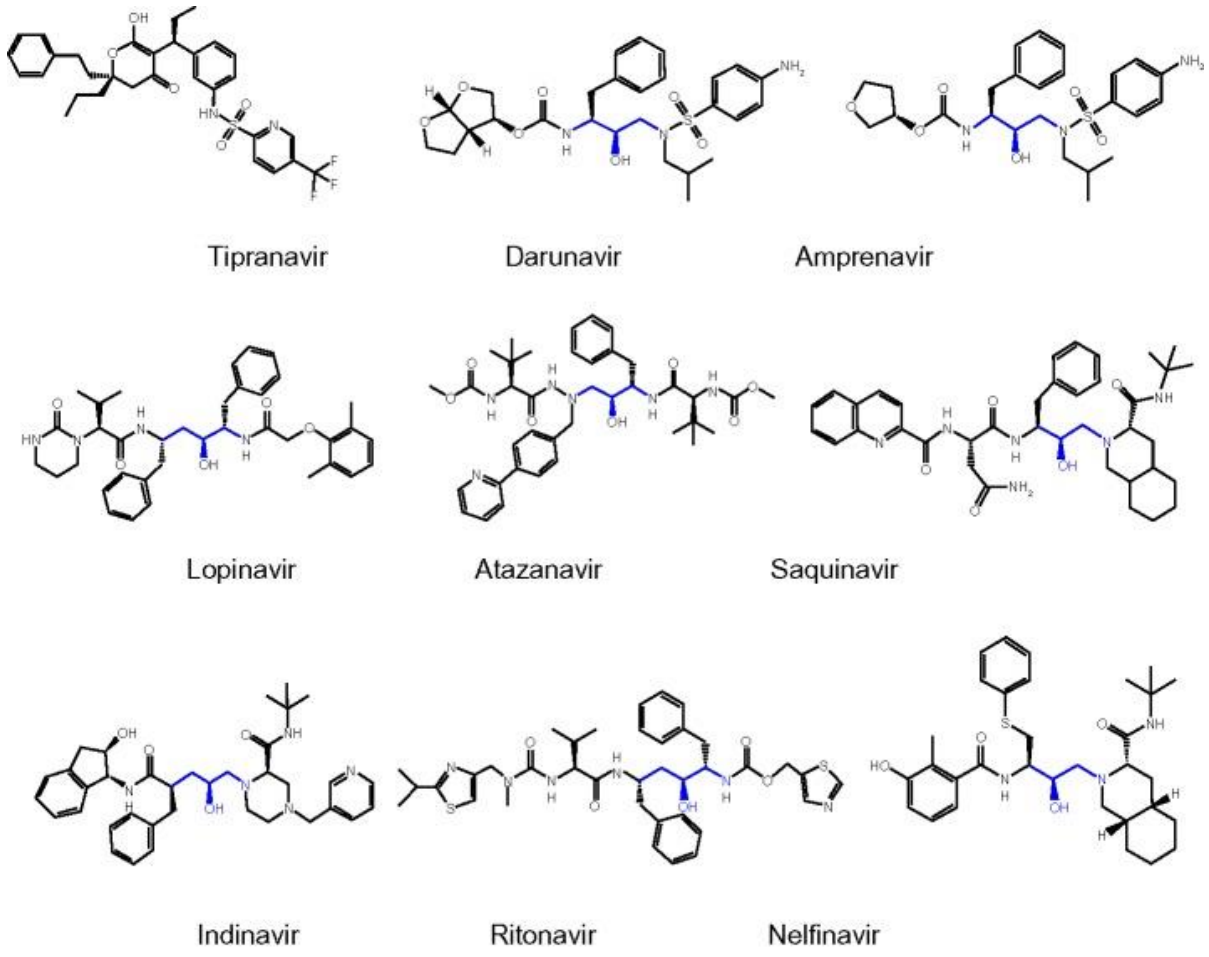


Figure 7: Chemical structures of FDA-approved HIV-1 protease inhibitors

The chemical structures of nine protease inhibitors that are used in the treatment of HIV-1 to target HIV-1 protease and prevent viral maturation an essential step in viral replication.

achieves resistance or evades PIs by following a “stepwise” pathway. Firstly, by acquiring primary resistant mutation in the protease gene and secondly by selection of secondary or compensatory protease mutation to restore the enzymatic activity and maintain viral fitness. And lastly, by selecting mutation in the cleavage sites of gag and gag-pol polyprotein precursors in order to restore protein processing and increase the yield of HIV-1 PR itself (Arts and Hazuda, 2012b; Clemente *et al.*, 2004; Condra *et al.*, 1995).

1.6.3 Mutations present in HIV-1 protease

The fast rate at which HIV-1 replicates, as well as the low fidelity and absence of the proof-reading ability nature of the reverse transcriptase, discussed above, has resulted in the emergence of HIV-1 protease variants that contain novel mutations. These mutations may be classified as either primary or secondary mutations. Primary mutations are usually conservative and occur within the active site of the PR. They directly alter drug binding and have been known to affect viral replication by altering the manner in which the PR’s active site binds the natural substrate (Konvalinka *et al.*, 1995; Shafer *et al.*, 2000). It is important to note that the nature and position of amino acid mutations are limited by the requirement of the PR to preserve its catalytic function. As such, mutations which alter the catalytic triad are near impossible to occur naturally. Mutations in the PR may also arise due to drug pressure. For a list of mutations occurring in the protease as a consequence of drug pressure, refer to https://hivdb.stanford.edu.

Secondary mutations compensate for primary mutations by improving proteolytic efficiency and increasing the rate at which the virus replicates (Liu *et al.*, 2008; Velazquez-Campoy *et al.*, 2002; Winters and Merigan, 2005). Additionally, amino acid mutations and/or insertions may arise at the natural cleavage sites of the PR, at the Gag and Gag-Pol polyproteins. This may allow the mutated PR to bind and cleave the polypeptides more effectively. Non-active site mutations, such as hinge-region mutations are regarded as secondary mutations because they do not directly alter the molecular interactions between the PR and PIs. Velazquez-Campoy *et al.,* (2001) reported that non-active site mutations may significantly lower the susceptibility of PR

to PIs. Additionally naturally occurring polymorphisms (NOPs) and non-active site mutations have been known to be incorporated within the protein sequence in order to evade drug binding while maintaining sufficient catalytic efficiency (Naicker and Sayed, 2014; Velazquez-Campoy *et al.*, 2001; Zondagh *et al.*, 2018).

A hinge region, HIV-1 subtype B PR mutation, M36I, was previously shown to reduce drug binding and induce PI resistance in HIV-1 subtype B patients (Costa *et al.,* 2014). The M36I mutation in subtype B PRs occurs as a response to drug pressure; however, it is naturally occurring in subtype C protease. The presence of M36I naturally in C-SA PR may contribute to the higher likelihood of increased PI resistance in subtype C patients. Costa *et al.,* (2014), further rationalised that the amino acid variations in the hinge region may affect the dynamics of the flaps and lead to reduced drug inhibition. Additionally, NOPs have previously been recognised to significantly alter the entropy and enthalpy of PR, which further drives drug resistance (Bessong, 2008; Velazquez-Campoy *et al.*, 2002).

1.6.4 Insertion mutations in HIV-1 protease

Amino acid insertions in HIV-1 PR are rather rare and usually occur at the hinge region between residues 32 and 41 (Winters and Merigan, 2005). It has been proposed that this region is more capable of accommodating extra amino acids without drastically changing the protein's overall structure (Winters and Merigan, 2005). These mutations, on their own, are not prone to induce drug resistance; however, when combined with other major mutations have been shown to reduce drug susceptibility to PIs (Liu *et al.,* 2016; Meher and Wang, 2015; Zondagh *et al.,* 2018). Insertions in the flap hinges can change the conformation of the active site and alter the flap movement and flap dynamics (Agniswamy *et al.,* 2016). The conformational change may allow the enzyme to block the access of the PIs to the active site (Todd and Freire, 1999). This in turn enhances the viral fitness of HIV, as the enzyme will develop resistance towards the PIs.

The South African HIV-1 subtype C protease (C-SA PR) contains eight amino acid polymorphic sites in comparison to the well-characterised HIV-1 subtype B PR (Mosebi *et al.*, 2008; Naicker *et al.*, 2013). These amino acid polymorphisms are;

T12S, I15V, L19I, M36I, R41K, H69K, L89M and I93L (Mosebi *et al.*, 2008). Although the polymorphisms are not within the active site, studies have shown that they may still play a role in drug resistance as secondary mutations (Liu *et al.,* 2016; Muzammil *et al.,* 2003; Naicker *et al.,* 2013). Therapeutic approaches have been focused on the subtype B PR, which is found predominantly in western countries. It has been suggested that the polymorphisms reduce the effectiveness of the protease inhibitors toward the C-SA PR (Mosebi *et al.*, 2008; Naicker and Sayed, 2014).

1.6.5 N37T↑V^{+10} subtype C protease

A blood sample was taken from a drug-naïve infant who had been infected with HIV-1. The sequence data for an HIV-1 positive, drug-naïve infant subtype C protease was acquired from Ledwaba *et al.,* (2019). The sequence revealed a variant of the South African HIV-1 subtype C protease, containing a substitution and insertion at the hinge regions. The substitution mutation was that of asparagine for threonine and the insertion of valine at position 37. The variant will, henceforth, be referred to as N37T↑V^{+10} and is illustrated in Figure 8. The N37T↑V^{+10} protease has 100 amino acids per monomer due to the insertion, which varies from the standard 99 amino acids per monomer of the wild type subtype C protease (Naicker *et al.,* 2013a). The variant protease contains an additional 10 NOPs: I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L, in each monomer. These NOPs and their polarity and charge are shown in Table 2 below.

Rationale

Previous studies suggest insertion mutations, particularly in the hinge region of HIV-1 protease may increase the flexibility of the flaps, and as such, result in weaker or altered binding of protease inhibitors to the active site (Liu *et al.,* 2016; Naicker *et al.,* 2013; Zondagh *et al.,* 2018). Weaker binding of the PIs may lead to decreased drug susceptibility and therefore, may predispose the protease to further mutations and drug resistance. It was, therefore, important to evaluate the impact that the hinge-region mutations and insertions have on the structure, stability and drug-binding

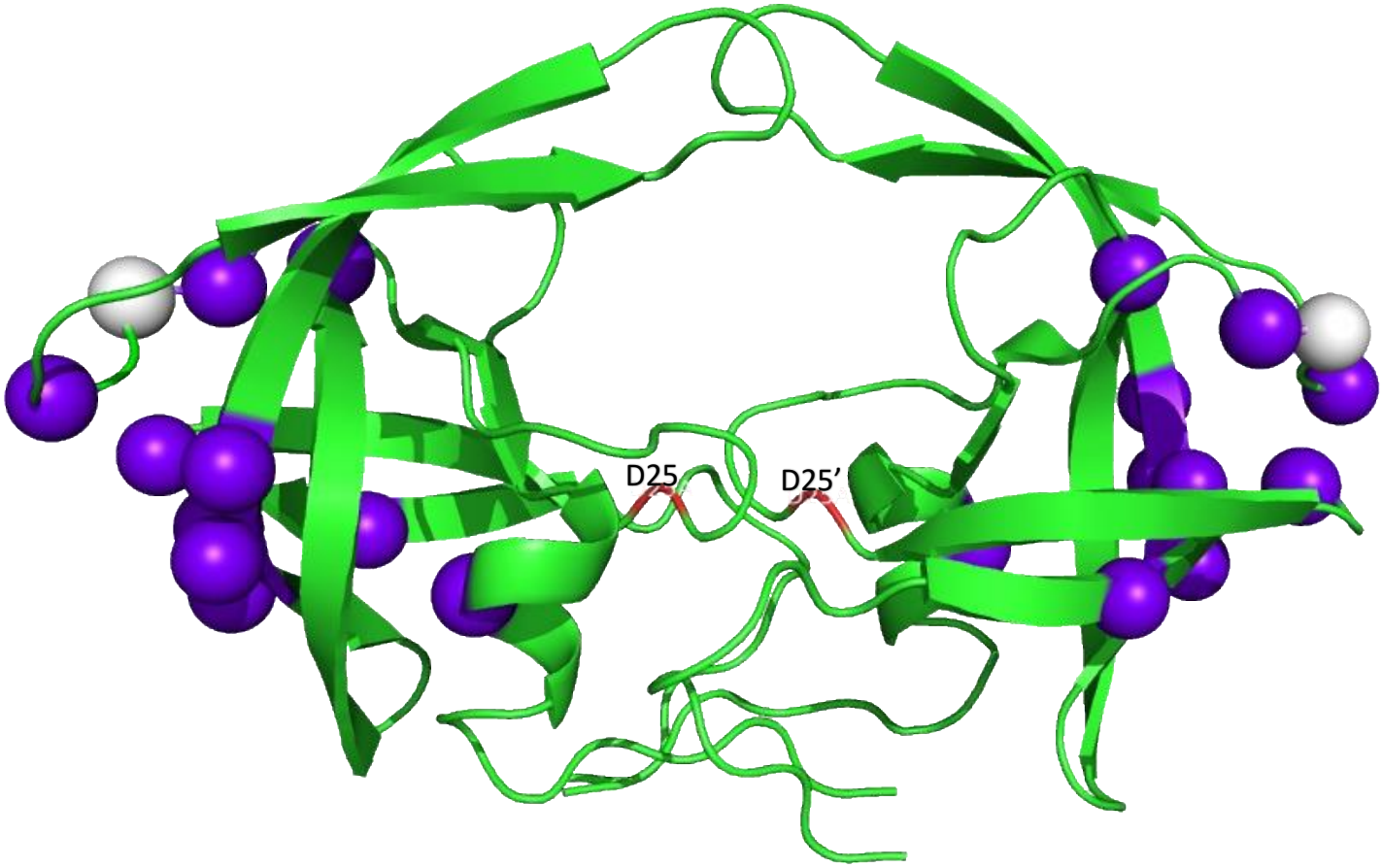


Figure 8: Homology model of N37T↑V+10 protease

The ribbon illustration of N37T↑V^{+10} HIV-1 subtype C protease. The background mutations which are shown in purple are I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L. The Asp25/25' active site residues are shown in red. The white sphere is the valine insertion. The homology model was generated using SWISS-MODEL (Biasini *et al.,* 2014) with the use of the wild type C-SA structure (PDB ID: 3u71; 2.72 Å resolution) as a template (Naicker *et al.,* 2013).

Table 2: Location of naturally occurring polymorphisms present in N37T↑V^{+10} HIV-1 protease and comparison of the residue charge to the HIV-1 subtype C wild type amino acid residues. (Information used in the table below was obtained from 3U71.PDB (Naicker *et al.,* 2013a; Zondagh *et al*., 2018).

Structural region	Associated residues	Polymorphism				
			WT C-SA		N37T↑V^{+10}	
Fulcrum region	10-22	1	Ile13	NP	Val13	NP
		2	Gly16	NP	Glu16	-VE
Hinge region	35-42	3	Ile36	NP	Thr36	P
		4	Pro39	NP	Ser39 (40)	P
	56-61	5	Asp60	-VE	Glu60 (61)	-VE
		6	Gln61	P	Glu61 (62)	-VE
Cantilever region	62-78	7	Ile62	NP	Val62 (63)	NP
		8	Leu63	NP	Pro63 (64)	NP
		9	Val77	NP	Ile77 (78)	NP
		10	Met89	NP	Leu89 (90)	NP

NP : non-polar
P : polar
-ve : negatively charged

properties of the HIV-1 protease. This will provide background into drug improvements and targets for drug design. According to Kim *et al.*, (2001); Ledwaba *et al.*, (2019), a threonine insertion at position 37 occurs in both subtype B and subtype C HIV-1 proteases. Comparative molecular assessment of N37T↑V^{+10} and C-SA protease showed that the insertion and substitution mutation altered the flap and hinge region dynamics of the protease (Zondagh *et al.*, 2018). This allowed for an increase in hinge region mobility and as such a greater opening of the flaps. The increased mobility may be, according to Liu *et al.*, 2016; Zondagh *et al.*, (2018) due to mutations within the hinge region eliminating
the salt-bridge interaction between Glu 35 and Arg 57. The absence of the salt bridge has been associated with the outward movement of the flaps, altered dynamics and altered conformational stability of HIV-1 proteases (Costa *et al.*, 2014; Liu *et al.*, 2020; Swairjo *et al.*, 1998).

Upon study of the crystal structure of an HIV-1 subtype B variant, it was revealed that the protease alters intramolecular interactions between the flaps and the hinge region to compensate for the loss of the salt bridge (Liu *et al.*, 2016). Consequently, this improves the stability of the HIV-1 protease. The altered stability of N37T↑V^{+10} protease remains to be investigated. This study made use of conformational stability assays, which include urea-induced equilibrium unfolding and thermal shift assays to give insight into the folding mechanism of N37T↑V^{+10} protease and reveal how altered molecular dynamics affect the folding and stability of the South African HIV-1 subtype C protease.

1.7 Aim and objectives

1.7.1 Aim

This study aimed to evaluate the implication of a hinge region mutation and insertion, on the stability, conformational dynamics and drug resistance of the HIV-1 subtype C protease.

1.7.2 Objectives

1. Overexpress and purify the WT•D25A and N37T↑V^{+10}•D25A HIV-1 proteases, with the goal of obtaining high yields of pure protein suitable for further structural and functional studies.
2.
3. Determine active-site concentration of WT•D25A and N37T↑V^{+10}•D25A using isothermal titration calorimetry.

4. Characterise the secondary, tertiary and quaternary structures of WT•D25A and N37T↑V^{+10}•D25A using far-UV CD, intrinsic tryptophan fluorescence and size-exclusion HPLC.

5. Optimise the crystallisation conditions for N37T↑V^{+10}•D25A protease, using a variety of macromolecular crystallisation techniques and screening conditions to identify suitable crystal forms for structure determination.

6. Characterise the thermodynamic stability of WT•D25A and N37T↑V^{+10}•D25A proteases using urea-induced equilibrium unfolding, to determine their relative stability and assess the impact of the mutations on the enzyme's folding and stability.

7. Investigate the thermal stability of WT•D25A and N37T↑V^{+10}•D25A proteases using thermal unfolding experiments, to determine their melting temperatures and the impact of the mutations on the enzyme's stability.

8. Use molecular docking and short simulations to predict the binding energies of WT, N37T↑V^{+10}, and N37T↑V proteases with ATV, DRV, FPV, IDV, NFV, RTV, SQV and TPV to gain insights into how background mutations affect the protease's binding affinity and specificity.

9. Conduct molecular dynamics simulations of WT and N37T↑V^{+10} HIV-1 proteases to investigate their structural dynamics, flexibility, and stability over

time, to gain a more comprehensive understanding of how the mutations affect the protease's function and stability.

Chapter 2: Methods

To study a protein, it is often necessary to first isolate the protein in its pure form. This process involves transformation, overexpression, purification and lastly quantification of the protein of interest. Once these steps are complete, structural and functional characterisation of a protein is possible. Isolation of high, correctly folded yields of HIV-1 protease has been a challenge for many years. An active HIV-1 protease undergoes auto-proteolysis, which is the self-cleavage of the protease into smaller polypeptides and amino acids (Darke *et al.*, 1989; Konvalinka *et al.*, 1995; Sayed *et al.*, 2008). Auto-proteolysis reduces the yield of protein that can be obtained experimentally.

Low protein yields have limited previous studies and according to Darke *et al.,* (1989), proteolytic activity of the protease may be eliminated using site-directed mutagenesis of the aspartic acid to asparagine or alanine at the 25th amino acid position. Recently, Sherry *et al.*, (2020) showed that the substitution of aspartic acid for alanine (D25A mutation) in HIV-1 subtype C disrupts the conserved catalytic triad and inactivates the protease while retaining its native structure. Taking this into consideration, a D25A mutation was introduced in each monomer to prevent the auto-proteolysis of both the wild type and N37T↑V^{+10}, to allow a greater yield of protease and an uncompromised analysis of the variant protease. The mutant with the D25A mutation will henceforth be referred to as N37T↑V^{+10}•D25A and the inactive wild type as WT•D25A. This study proposes the inactivation of N37T↑V^{+10} with a D25A mutation, as an improved method to increase the protein yield and allow for downstream experimentation. Furthermore, an optimised method for purification of inactive HIV-1 protease variant (N37T↑V^{+10}•D25A) is proposed.

2.1 Bacterial transformation of *Escherichia coli* BL21 (DE3) pLysS cells

For this study, gene sequences encoding WT•D25A and N37T↑V^{+10}•D25A were cloned into separate pET-11a vector (GenScript Biotech, United Kingdom) systems, Figure 9. The pET system is excellent for the cloning and purification of recombinant

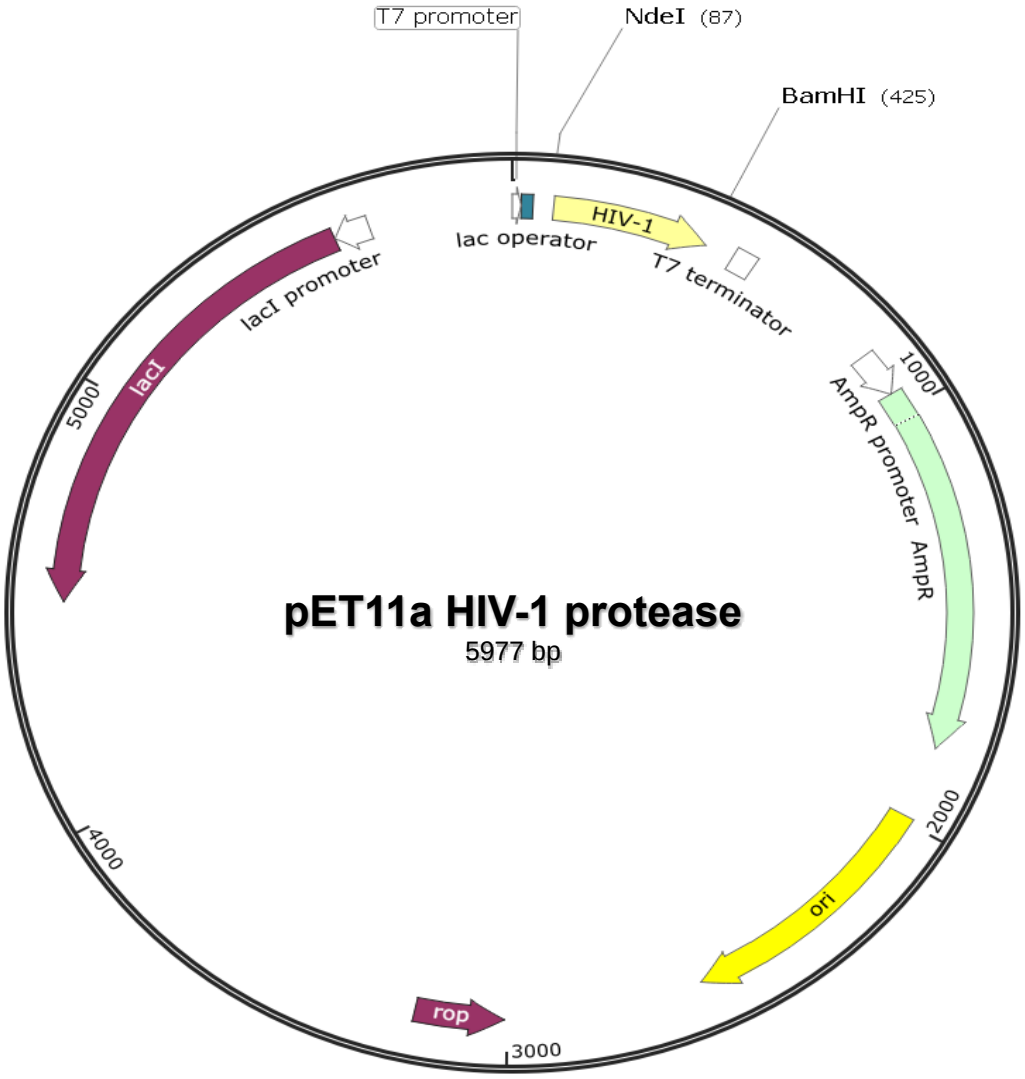


Figure 9: The pET-11a expression vector containing an HIV-1 protease sequence

The pET-11a vector (Novagen®) is widely used for heterologous protein expression in *E. coli* cells. The vector encodes the T7 promoter, which is essential for efficient transcription of the target gene (Mierendorf *et al.*, 1998; Novagen ®, 2003). The vector also conveys resistance to ampicillin, which allows for the selection of transformed cells. The sequence for HIV-1 protease was cloned into the vector and has NdeI and BamHI restriction enzyme sites. The image was adapted from the Addgene plasmid repository.

proteins in *Escherichia coli* (*E. coli*) (Novagen®, 2003). The pET-11a vector with the sequence for WT•D25A encoded was transformed into an *E. coli* BL21(DE3) pLysS cell line and the pET-11a vector that housed the N37T↑V^{+10}•D25A sequence was transformed into *E. coli* Rosetta cells using the heat shock technique (Fox and Hotchkiss, 1957). The successfully transformed cells were plated on agar plates containing 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) sodium chloride and 1.5% (w/v) agar. Additionally, 100 μg/ml of ampicillin and 35 μg/ml of chloramphenicol were spread on the agar plates as selective antibiotics. The pET-11a vector has a natural resistance to ampicillin; therefore, the successful growth of *E. coli* cells is indicative of the successful uptake of the vector by the cells (Novagen ®, 2003). Additionally, the pLysS of *E. coli* BL21(DE3) and *E. coli* Rosetta cells contain a selection marker that encodes for resistance to chloramphenicol. This allows for the selection for the growth of only *E. coli* cells in a chloramphenicol rich medium. From the transformed agar plates, a single, discrete bacterial colony was selected and inoculated into sterile LB media that was made up of 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) sodium chloride. Ampicillin (100 μg/ml) and chloramphenicol (35 μg/ml) were added to supplement the LB media. The cells were then incubated for 16-24 hours in a New Brunswick Excella® E24/E24R at 200 rpm at 37 ºC. Thereafter, 100 μl cell culture was mixed with 100 μl of 80% (v/v) glycerol to make glycerol stocks. The glycerol stocks were thereafter stored until use at -80 ºC.

It was then necessary to extract the DNA from the transformed cells to confirm successful transformation and ensure the correct sequence of HIV-1 protease was incorporated. The GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific™ Massachusetts, USA) was used to extract the plasmid DNA from the transformed cells. The purified DNA was, thereafter, sent to Inqaba Biotech® (Pretoria, South Africa) for Sanger sequencing to verify the sequence.

2.2 Heterologous overexpression of WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease

The conditions for protein expression require optimisation for optimal expression of the HIV-1 protease. This includes conducting growth curve studies to identify when the optimum induction period is. And secondly, it is essential to conduct induction trials to determine the right time and concentration of reagents to add for optimum induction of protein expression. The conditions for the overexpression of WT•D25A were previously identified in our lab (Sherry *et al.*, 2020), and these were used to express WT•D25A. The optimum conditions for the overexpression of N37T↑V^{+10}•D25A were yet to be established.

2.2.1 Optimisation of conditions for HIV-1 protease overexpression

To conduct growth curve studies for *E. coli* cells housing N37T↑V^{+10}•D25A PR, lysogeny broth (LB) media supplemented with antibiotics (ampicillin 100 µg/ml and chloramphenicol 35 µg/ml) was inoculated with transformed *E. coli* Rosetta cells using aseptic techniques. The cells were allowed to grow for 16-24 hours, at 37 ºC at 200 rpm in a New Brunswick Excella® E24/E24R incubator. Following this, a 1:25 dilution of overnight culture into 50 ml of LB media, containing the antibiotics was performed. The OD_{600} was recorded at 20 minutes intervals for 5 hours.

2.2.2 Induction trials for N37T↑V^{+10}•D25A

A 50 ml overnight culture was prepared in LB media (supplemented with ampicillin 100 µg/ml and chloramphenicol 35 µg/ml) for 16 hours, 230 rpm and 37 °C. A 1:25 dilution was then done by adding 2 ml of overnight culture in 5 Erlenmeyer flasks (vol: 250 ml) containing 50 ml of sterile LB media. The cells were allowed to grow until an OD_{600} of 0.4 at 37 °C and 230 rpm. Following this, isopropyl-ß-D-thiogalactopyranoside (IPTG) was added in each of the five flasks at concentrations of 0, 0.2, 0.5, 0.8 and 1 mM, respectively. IPTG is a molecular mimic of allolactose, which activates transcription of the lac operon (Lama and Carrasco, 1992). The activation of the transcription induces the overexpression of proteins by the *E. coli* cells. A 1 ml sample was collected from each flask after 0, 2, 4, 6 and overnight (16-24 hours) and saved for electrophoresis using sodium dodecyl sulfate-polyacrylamide

gel electrophoresis (SDS-PAGE). The samples were spun 10 000xg for 3 minutes and resuspended in 100 µl of 10 mM Tris-HCl before being stored at 20 °C until needed.

2.2.3 Overexpression of WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease

Following the identification of the optimal conditions for overexpression of both WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease growth of the two cell lines proceeded. *E. coli* BL21(DE3) pLysS and Rosetta cells were grown separately in LB media supplemented with antibiotics; ampicillin (100 µg/ml) and chloramphenicol (35 µg/ml), at 37 °C to express both WT•D25A and N37T↑V^{+10}•D25A, respectively. The cell growth was monitored by measuring the optical density at a wavelength of 600 nm (OD_{600}) at intervals of 20 minutes until it reached the early-log phase. Once cell growth was at the early-log phase it signaled for the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to induce overexpression. Overexpression was allowed to proceed for 4 hours. Following this, a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer Scientific™, Massachusetts, USA) was used to spin down the cells at 5 000xg, for 15 minutes at a temperature of 4 °C. The pellet was then resuspended in 30 ml of lysis buffer made of 10 mM Tris, 2 mM ethylenediaminetetraacetic acid (EDTA), 1 mM phenylmethylsulphonyl fluoride (PMSF), at pH = 8.0. The cell lysate was then stored at 20 °C and three freeze/thaw cycles were performed before protein purification could proceed.

2.3 Purification of the WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease

The purification of WT•D25A and N37T↑V^{+10}•D25A HIV-1 proteases from inclusion bodies made use of detergent-based methods that have previously been used in our laboratory (Mosebi *et al.*, 2008; Naicker *et al.*, 2013; Sherry *et al.*, 2020). The overexpressed cells were initially resuspended in a lysis buffer (10 mM Tris, pH = 8.0, 2 mM EDTA 2% (v/v) Triton X-100). Sterile 10 mM $CaCl_2$, 10 mM $MgCl_2$ and 1 mg/ml bovine pancreatic deoxyribonuclease I (DNase I) were added to the lysate which was thereafter mixed using a homogeniser. The mixture was then incubated at 20 ºC for 30 minutes to allow viscosity to decrease.

The solution was then mixed by ten sonication cycles of 30 seconds each, with 30 s incubation on ice between each sonication cycle. The cells were thereafter spun down at 23 000xg for 30 minutes at 4 ºC using a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer Scientific™, Massachusetts, USA). The pellet was resuspended in a wash buffer containing 2% Triton X-100 (v/v) and 1 mM PMSF. This was followed by centrifugation at 23 000×g for 30 minutes at 4 ºC. The resulting pellet was resuspended in unfolding buffer (8 M urea, 10 mM Tris, 5 mM DTT, pH = 8.0). The solution was then mixed using a homoginiser (Pro Scientific PRO 200) and incubated for an hour at room temperature. Following incubation, the solution was treated to centrifugation at 18 000xg for 30 minutes at 20 °C. At this point, the protein was expected to be in the supernatant and existing in its unfolded form. The purification of WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease then deviates. The purification of WT•D25A and N37T↑V^{+10}•D25A continues at **2.3.1** and **2.3.2,** respectively.

2.3.1 Purification of WT•D25A

A tandem setup of a diethylaminoethyl (DEAE) cellulose column (Cytiva, USA) and a CM cellulose column (Cytiva, USA) were connected to an ÄKTAprime plus™ liquid chromatography purification system (GE Healthcare Life Sciences, Chicago, USA) and used to puriy WT•D25A PR. The columns were equilibrated with the unfolding buffer (8 M urea, 10 mM Tris, 5 mM DTT, pH = 8.0), which was passed through the columns five times for the combined column volume (50 ml) to ensure system equilibrium. The unfolded WT•D25A, suspended in urea, was then passed through the columns. The negatively charged DEAE column bound any positively charged contaminating proteins while the positively charged WT•D25A protease bound to the negatively charged resin of the CM-Sepharose column. The DEAE column was removed, and the bound protein was eluted from the CM-Sepharose column using 20 ml of unfolding buffer containing 1 M NaCl. The fractions were pooled based on their absorbance at 280 nm.

The next step involved refolding and purifying the native HIV protease using anion exchange chromatography. The pooled fractions were diluted and dialysed against

refolding buffer (10 mM Tris, pH 7.5, 2 mM dithiothreitol (DTT) 0.02% (w/v) sodium azide 10% (v/v) glycerol and the resulting dialysate was passed through a DEAE column. The protein passed through the positively charged DEAE resin and was collected in the flowthrough based on absorbance at 280 nm. The sample was then concentrated using an Amicon (Millipore Sigma, Massachusetts, USA) stirred-cell concentrator, and the concentrated protein sample was dialysed against storage buffer (10 mM sodium acetate, pH 5.0 0.02% (w/v) sodium azide 150 mM sodium chloride) for 4-8 hours at 4 °C. The WT•D25A was then stored at -20 °C until further use.

2.3.2 Purification of N37T↑V^{+10}•D25A

The refolding of the now unfolded N37T↑V^{+10}•D25A HIV-1 protease was carried out by dialysis in SnakeSkin Dialysis Tubing (3.4 kDa cut off) using an ice-cold refolding buffer containing formic acid, 10% glycerol and sodium azide (0.01% w/v) at 4 °C for six hours, followed by overnight dialysis in a buffer containing 10 mM Tris, 5 mM DTT, 0.01% (w/v) sodium azide, pH=7.8, at 4 °C. The samples were centrifuged at 23 000xg for 30 minutes at 4 °C, and the supernatant was then filtered through a 0.45-micron filter (Sigma-Aldrich, USA) and stored. The filtered supernatant containing N37T↑V^{+10}•D25A HIV-1 protease was then purified using ion exchange chromatography. This made use of a cellulosic resin containing carboxyl methyl-Sepharose (CM-Sepharose) cation exchange column to separate it from other proteins expressed by the *E. coli* Rosetta cells. CM serves as a weak cationic exchanger used to separate basic and neutral proteins (Voet and Voet, 2010; Wilson and Walker, 2010).

A sodium acetate buffer, pH = 5, was used as it had a pH below the pI (9.08) of the HIV-1 protease and as such the enzyme obtained a net positive charge. The CM-Sepharose column separates based on charge, and the net charge may vary with pH (Maseko *et al.*, 2016). The CM-Sepharose column was equilibrated using a degassed sodium acetate buffer, and the filtered N37T↑V^{+10}•D25A sample was passed through the column. The HIV-1 protease was expected to bind to the column, while proteins with a net negative charge were not. The protein was eluted using a salt (NaCl) gradient of 0 – 1 M, using a sodium acetate buffer with 1 M NaCl. The eluted protein was collected and refolded by dialysis into a storage buffer (10 mM sodium acetate,

pH 5.0, 0.02% (w/v) sodium azide, 150 mM sodium chloride) overnight at 4 °C. The protein was then filtered through a 0.45-micron filter and stored at -80°C.

2.4 Purity assessment of the proteases using 16% tricine-SDS-PAGE

It was then necessary to evaluate the purity of the protein. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is an electrophoretic technique commonly used to separate proteins based on their size (Schägger and von Jagow, 1987). In this study, SDS-PAGE was used to determine the monomeric size and purity of WT•D25A and N37T↑V^{+10}•D25A HIV-1 proteases. When proteins are treated with SDS, the protein is denatured with the anionic detergent that binds it and imparts a negative charge that is proportional to the molecular mass (Nowakowski *et al.*, 2014). The protein is left denatured (unfolded) into the monomeric state, allowing for separation to continue based on their molecular size only.

To evaluate the purity of the purified proteases, a tricine-SDS-PAGE gel was used. This technique can resolve molecules smaller than 30 kDa and provides an accurate measurement of the purity of proteins (Schägger and von Jagow, 1987). The gel comprised of 16% (w/v) separating gel in combination with a 4% (w/v) stacking gel. Approximately 20 µL of purified protein samples were mixed in a 1:1 ratio with a loading buffer (10% (w/v) SDS, 30% (w/v) glycerol, 2% (v/v) 6% (w/v) beta-mercaptoethanol, 0.005% Coomassie blue G-250, and 150 mM Tris-HCl at pH 7.5). Subsequently, the protein samples were heated to 95 °C for 5 minutes to ensure complete binding to SDS. Thereafter, 5 µL of protein samples and 2 µL of a protein ladder, PageRuler™ Pre-stained Protein ladder protein standards (10 to 180 kDa; ThermoFisher Scientific, Massachusetts, USA) were loaded onto the gel. Electrophoresis was then carried out in an anode buffer (1 M Tris, pH 8.9) and cathode buffer (1 M Tris, 1 M Tricine, and 1% (w/v) SDS, pH 8.25) using a Mini-PROTEAN® Tetra vertical electrophoresis cell system (Biorad, California, USA). The gel was then removed and stained in stain solution (25% (w/v) Coomassie Brilliant Blue R-250, 10% (v/v) glacial acetic acid, and 50% (v/v) methanol) overnight (12-16 hours) to visualise the proteins. Destaining the gel then followed by submerging the gel in destaining

solution (10% (v/v) glacial acetic acid and 50% (v/v) methanol) for 8-12 hours. Pictures of the gels were taken and the migration of each protein was compared to proteins standards of the protein ladder (molecular weight marker).

2.5 Concentration determination of the proteases

The concentration of WT•D25A and N37T↑V^{+10}•D25A HIV-1 proteases was determined spectrophotometrically. The absorbance of the samples at 280 nm was recorded and the concentration was calculated. The concentration was calculated using the Beer-Lambert law (*equation 1*).

$$A = \varepsilon c l \qquad \textbf{(Equation 1)}$$

Where A is the absorbance (AU), ε is the molar extinction coefficient ($M^{-1}.cm^{-1}$), c is the molar concentration and l is the path length (1 centimetre). The theoretical extinction coefficient of both WT•D25A and N37T↑V^{+10}•D25A was determined using ProtParam (Geisteiger, 2005). The molar extinction coefficients for the dimeric WT•D25A and N37T↑V^{+10}•D25A were determined to be 25 480 $M^{-1}.cm^{-1}$ and 24 980 $M^{-1}.cm^{-1}$, respectively. This was confirmed by using *equation 2* (Nijhuis *et al.,* 2007).

$$\varepsilon = 5550 * \sum Trp + 1340 * \sum Tyr + 150 * \sum Cys \qquad \textbf{(Equation 2)}$$

2.6 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a powerful technique that allows for the determination of the thermodynamics of biomolecular interactions, including binding constants, reaction stoichiometry, enthalpy, and entropy (Leavitt and Freire, 2001); Jelesarov and Bosshard, 1999). In this study, ITC was used to perform an active-site titration of the wild-type and N37T↑V^{+10} proteases. Acetyl-pepstatin was used as the ligand, which is an inhibitor that binds specifically to the active site of aspartyl

proteases. The ligand was dissolved in a sodium acetate buffer (10 mM sodium acetate, 0.02% (w/v) sodium azide, 150 mM sodium chloride and pH 5.0), which is similar to the storage buffer, used to store the proteases. The 150 µM acetyl-pepstatin was titrated into the sample cell containing PR that had a concentration of 20 - 30 µM. The resulting heat change was recorded using the NanoITC standard volume instrument (TA® Instruments, New Castle, USA). The reference cell contained the storage buffer without the PR or ligand, was used to account for any heat change that was not due to the binding of the ligand to the protease.

The ITC instrument used for this study was equipped with a NanoAnalyze software package (TA Instruments) that allowed for the analysis of the data. The one-to-one binding model was used as the fitting model to obtain the thermodynamic parameters, including the binding constant, reaction stoichiometry, enthalpy, entropy, and Gibbs free energy. These parameters were obtained directly and/or calculated from the thermodynamic profile.

2.7 Enzyme activity assay

To confirm the inactivity of N37T↑V^{+10}•D25A protease due to the D25A mutation, an activity assay was conducted using a chromogenic substrate, Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2, which mimics the sequence KARVL/AEAM found on the capsid/p2 cleavage site of the Gag-polyprotein precursor. The quenching amino group, Phe(NO_2), and the fluorogenic 2-aminobenzoyl group (Abz) are held in close proximity. Upon cleavage of Nle-Phe(NO_2), the fluorophore, aminobenzoic acid (Abz) is released and fluoresces at 425 nm (Caramel and Yaron, 1978). The catalytic activity of both WT•D25A and N37T↑V^{+10}•D25A HIV-1 PR was investigated by measuring the ability of the PR to cleave the chromogenic substrate. Cleavage of the substrate would be indicated by an increase in fluorescence intensity as more substrate molecules were cleaved. The presence or fluctuation of fluorescence would signify an active PR.

The interaction of the enzyme with the substrate was monitored using a Jasco FP-6300 spectrofluorometer, and the sample was excited at an excitation wavelength of 337 nm and monitored at an emission wavelength of 425 nm. Enzyme concentrations

of 2 µM were mixed with 500 µM of the chromogenic substrate and a reaction buffer consisting of 1 M NaCl, 50 mM sodium acetate, at a pH of 5.0. The experiment was conducted at 20 °C. The emitted fluorescence was monitored over a period of 60 seconds. A linear curve of time (s) vs fluorescence intensity (AU) was constructed to observe activity fluctuation over the reaction time.

2.8 Structural characterisation

The local structure of the South African WT•D25A and N37T↑V^{+10}•D25A HIV-1 subtype C proteases have according to the researchers knowledge not yet been characterised. Sherry *et al.*, however, recently solved the crystal structure of an inactive subtype C protease, WT•D25A. This part of the current study aimed to investigate how the D25A mutation altered the local structure and further probe the impact the insertion and substitution mutations have on the secondary, tertiary, and quaternary structure of the protein. This would aid in understanding why HIV-1 protease selects and conserves these particular hinge region mutations. It was also necessary to identify macromolecular crystallization condition for the optimum growth of N37T↑V^{+10}•D25A HIV-1 protease crystals.

2.8.1 Secondary structure

Far-UV circular dichroism (CD) spectropolarimetry was used to determine the secondary structure content of the WT•D25A and N37T↑V^{+10}•D25A HIV protease. Circular dichroism is a powerful technique that measures the difference in absorption of left- and right-handed circularly polarised light by chiral molecules (Greenfield, 2006). The α-helices and β-sheets of the protease absorb polarised light in the far-UV region of 240-180 nm due to their chiral nature. By measuring the absorbance of the light, one can determine the secondary structural content of both the wild type and N37T↑V^{+10} HIV protease.

The protease samples were prepared in a buffer containing 20 mM sodium phosphate at pH 7.0 and a concentration of 0.1 mg/mL. The CD spectra were collected in triplicate on a Jasco J-815 CD spectrometer with a scan speed of 50 nm/min, a bandwidth of 1 nm, and a response time of 2 seconds. The spectra were analysed using the

software provided by the instrument manufacturer. The secondary structural content of the protease was determined by converting the raw signal to mean residue ellipticity (MRE) using equation 3:

$$\mathrm{MRE} = \frac{\theta \mathrm{x}\ 100}{\mathrm{cnl}} \qquad \textit{(Equation 3)}$$

where θ is the raw signal in millidegrees (mdeg), c is the protein concentration in mg/ml, l is the path length of light through a cuvette in cm, and n is the number of amino acid residues in the protein.

2.8.2 Tertiary structure

Fluorescence is a widely used technique in biochemistry and molecular biology to study protein structure and dynamics. It involves light that is emitted when an electron moves from a higher excited state to a lower state (Lakowicz, 2006). Fluorescence typically occurs from aromatic molecules which are known as fluorophores. In proteins, the aromatic amino acid fluorophores are tyrosine and tryptophan and a small contribution from phenylalanine. These residues have conjugated double bonds which can be excited with light of a specific wavelength to a higher excited electronic state. Upon relaxation, the molecules emit a photon of lower energy, resulting in fluorescence (Lakowicz, 1999).

Tryptophan emits intrinsic fluorescence that can be measured by fluorescence spectroscopy. Intrinsic fluorescence is highly dependent on an indole ring of the tryptophan residue (Callis, 1997). The fluorescence properties of the indole ring present in tryptophan make it a valuable tool for studying protein structure and dynamics. The indole ring has a maximum absorption wavelength of 295 nm and emits light in the range of 300-350 nm, depending on the polarity of its local environment (Carmel and Yaron, 1978; Lakowicz, 2006). Buried tryptophan residues display a blue-shifted emission peak (<320 nm), while solvent-exposed or unfolded tryptophan residues display a red-shifted peak (>340 nm). Thus, the emission wavelength of tryptophan can be used to determine whether it is buried or exposed to the solvent.

An emission peak greater than 350 nm indicates that it is solvent exposed, while a peak ranging from 310 nm to 320 nm indicates that it is buried (Chen *et al.*, 2008; Lakowicz, 2006; Osysko and Muíño, 2011). In the context of protein tertiary structural characterisation, the tryptophan residues in the 6th and 42nd positions of each HIV-1 protease monomer (Trp 6, Trp 6', Trp 42, and Trp 42') are particularly useful. These residues can be used to determine the local tertiary structure of the protein, providing valuable insights into conformational changes and dynamics of the protein (Davies, 1990).

The tertiary structure of the WT•D25A and N37T↑V^{+10}•D25A PR was analysed using intrinsic fluorescence spectroscopy. Specifically, the intrinsic fluorescence of the two tryptophan residues located in each monomer; at positions 6 and 42 of monomer A and 6' and 42' of monomer B (Figure 10), was measured. These fluorophores were excited at a wavelength of 295 nm using a Jasco FP-300 Spectrofluorometer, and the resulting emission spectra were measured between 280 nm and 500 nm. The parameters used for the measurement were an excitation bandwidth of 2.5 nm, an emission bandwidth of 2.5 nm, and a 0.5 nm data pitch. Overall, this method allowed for the determination of the local tertiary structure of WT•D25A and N37T↑V^{+10}•D25A PRs based on the fluorescence emission wavelength of the tryptophan residues.

The fluorescence emission spectra of both the WT•D25A and N37T↑V^{+10}•D25A PR were collected at concentrations of 2 µM in storage buffer, pH = 5. The experiments were conducted using a Jasco FP-6300 fluorescence spectrophotometer equipped with a Peltier temperature controller set at 20°C. A quartz cuvette was used for all fluorescence experiments, and the excitation and emission bandwidths were set at 5 nm and 2.5 nm, respectively. The spectra were recorded using a scanning speed of 200 nm/min. Intrinsic tryptophan fluorescence was monitored using excitation wavelengths of 295 nm, and emission was monitored between 290-500 nm. The data presented are the average of three accumulations per triplicate sample, and all spectra were corrected for buffer contributions. The WT•D25A and N37T↑V^{+10}•D25A were denatured/unfolded by resuspending in 8 M urea and allowed to incubate for 1 hour in 20 ^{o}C. The fluorescence emission was then monitored at 295 nm. The PR was refolded by diluting the urea out with a 10 mM sodium acetate buffer and incubating at room

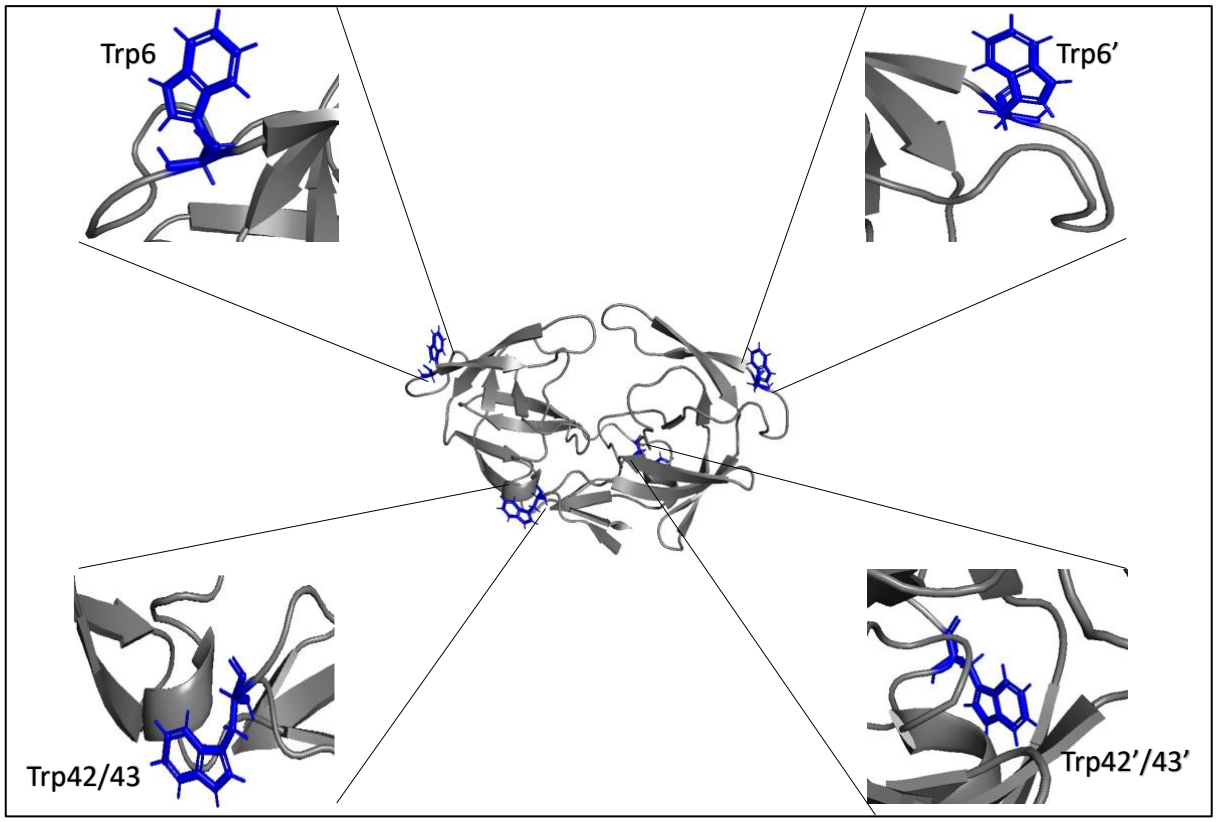


Figure 10: Location of tryptophan residues in monomers A and B of WT•D25A and N37T↑V^{+10}•D25A

On the WT•D25A the tryptophan residues are found at the 6th and 42nd amino acid residues on each monomer. On N37T↑V^{+10}•D25A tryptophan's are located on the 6th and 43rd residue on each monomer. The prime symbol indicates the residue on monomer B. This image was generated using PyMOL Molecular Graphics Software, version 1.8 (Schrödinger, LLC), using data obtained from the Protein Data Bank (PDB ID: 3U71) (Naicker et al., 2014).

temperature for 4 hours. The spectrum of the native, unfolded and refolded WT•D25A and N37T↑V^{+10}•D25A PR were then plotted using Microsoft Excel.

2.8.3 Quaternary structure

The quaternary structure of a protein refers to its complex arrangement of subunits, which can be critical to its function. One method used to assess the quaternary structure of a protein is size exclusion coupled with high performance liquid chromatography (SE-HPLC). SE-HPLC separates proteins based on their hydrodynamic volume, allowing the determination of the size and oligomeric states of proteins between 1 and 1000 kDa (Wilson and Walker, 2010). In this experiment, a Bio-Select SEC matrix (BIO-RAD, South Africa) was used as the matrix of HPLC as it can resolve proteins between 5 and 250 kDa in size. The HPLC system was operated at a constant flow rate of 1 $ml.min^{-1}$ with an isocratic pressure of 65 bar, and the experiment was conducted at a temperature of 20 °C. The equilibration buffer used in the experiment contained 10 mM sodium acetate, 1 mM NaCl, and 2 mM DTT, with an operational pH = 5. The buffer was filtered and degassed to remove any contaminants and air bubbles.

The equilibration buffer was then passed through the column to equilibrate the system and wash out any foreign molecules that may have been present in the column. The PR was prepared by dialysing 20 µM PR in a fresh storage buffer, filtering and then finally injected into the column. The absorbance at 280 nm was recorded using an absorbance detector connected to the system. The resultant chromatogram was collected and used to plot a graph of the elution time and absorbance. The molecular weight was then determined by plotting a standard curve of the log of the molecular weight versus the ratio of elution and void volume of the standard samples. Standard samples with known molecular weights; aprotinin (6.5 kDa), cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa), albumin (66 kDa) and blue dextran (2000 kDa) were used to deduce the molecular weight of the WT•D25A and N37T↑V^{+10}•D25A PR.

2.9 Macromolecular crystallisation

Due to the absence of an experimentally solved crystal structure of HIV-1 N37T↑V^{+10}•D25A PR in the Protein Data Bank (PDB), it was necessary to construct a homology model. To curb the future use of a homology model, macromolecular crystallography was employed. Crystallisation requires highly pure proteins, homogenised proteins capable of going through the three major steps of crystallography: namely, nucleation, growth and cessation of growth. The protein crystal forms when proteins in solution are brought to a supersaturated state which forces the macromolecules to a solid i.e., crystallise. Once obtained, the crystal structure may be used for drug studies, atomic and molecular studies.

There are several methods available for crystallisation, and the hanging-drop vapour diffusion method was the preferred method for this study. This is because it requires a less volume and concentration of samples compared to other techniques, it is easier to execute and allows for a large amount of flexibility during screening and optimisation. This method is further advantageous over sitting drop vapour diffusion, another widely used crystallography method, as the drop can be viewed through the glass without any optical interference from plastic. The hanging drop method is more flexible and has reduced chances of crystals sticking to hardware when compared to the sitting drop. The technique, however, requires substantially more reagents and time to complete.

Crystallisation trials required higher concentration of PR, therefore the apo-N37T↑V^{+10}•D25A PR samples were concentrated to a concentration range of 2 – 5 mg/ml and resuspended in a sodium acetate buffer (10 mM sodium acetate, pH 5.0, 0.02% (w/v) sodium azide, 150 mM sodium chloride). Thereafter, macromolecular crystallisation trials were conducted using the hanging-drop vapor diffusion method. This made use of the Hampton Research Index™ HR2-144 screening kit (Hampton Research, California, USA), which is a 96 well buffer system designed for the crystallisation of globular proteins. For the crystal trials, a 96 well plate was prepared by filling each reservoir well with 40 µL reservoir solution. The N37T↑V^{+10}•D25A PR was then mixed with the Index™ screening kit at a ratio of 1:2 and 2:1, with both having a final volume of 0.75 µL using the Oryx8™ protein crystallisation robot (Douglas

Instruments, East Garston, UK) and maintained at 20 °C. Additionally, a mixture of PR and Index^{TM} screening kit at a ratio of 1:1, with a final drop volume of 1 µL, was also prepared using the same robot and conditions. The plate was then carefully sealed with Crystal Clear Tape and incubated at 20°C for a screening time of one to four weeks. Conditions leading to protein crystal conditions were documented and images were taken using a microscope mounted eyepiece camera (Dino- Eye eyepiece camera, Dino-Lite).

2.10 Conformational stability and energetics

The conformational stability of a protein is an indication of how stable the native three-dimensional structure of a protein is (Pace *et al.*, 1996). The stability is given as the difference in Gibbs free energy (∆G) of the folded and unfolded state (Pace *et al.*, 1996; Zhou, 2004). Studies have suggested that a change in the protein's conformational stability may alter biological activity and drug binding (Ahmad *et al.,* 2012; Huang *et al.,* 2014; Noel *et al.,* 2009; Palese, 2017). Previous studies have investigated the stability, folding and thermodynamic properties of the HIV-1 protease using various biophysical techniques, including calorimetric studies, urea denaturation, differential scanning calorimetry and thermal shift assays (Noel *et al.*, 2009; Ohtaka *et al.*, 2002; Sherry *et al.*, 2022; Todd *et al.*, 1998). These investigations have allowed the analysis and elucidation of components such as enthalpy, entropy, T_m and heat capacity that govern the conformational stability of a protein and contribute to the Gibbs energy. It has previously been shown that hinge region mutations may alter the dynamics and conformational stability of the HIV-1 protease and it is important to look into how the insertion and substitution mutation in N37T↑V^{+10}•D25A protease affects the stability of the HIV-1 protease. Obtaining this may help us advance what it entails for drug binding and the activity of the HIV-1 protease.

This study focused on investigating the conformational stability and the unfolding mechanism of the WT•D25A and N37T↑V^{+10}•D25A proteases. Urea-induced equilibrium unfolding (using circular dichroism and fluorescence), and a thermal shift assay was used to assess the conformational stability and probe the unfolding of both the WT•D25A and N37T↑V^{+10}•D25A.

2.10.1 Thermal stability

The secondary structural characterisation results obtained in section 3.6.1, reveal a characteristic trough for HIV-1 PR at 216 nm. Therefore, 216 nm was used to monitor the thermal unfolding of WT•D25A and N37T↑V^{+10}•D25A using far-UV CD spectroscopy. Thermal unfolding of WT•D25A and N37T↑V^{+10}•D25A was conducted using a Jasco J-810 Spectropolarimeter over a temperature range of 20 – 90 °C, at increments of 1 °C/min. Altered thermal stability of the protein would give evidence of the impact the substitution and insertion mutations present in N37T↑V^{+10}•D25A have on HIV-1 protease.

2.10.2 Urea-induced equilibrium unfolding

Equilibrium unfolding is a technique used to characterise the folding/unfolding mechanism of a protein. It elucidates the cooperativity, conformational stability as well as the denaturation midpoint in terms of denaturant concentration. There are a variety of methods that can be used to unfold proteins, such as heat, pH, denaturants and enzymes. For this study a well-characterised chaotropic agent, urea, was used to denature the WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease.

Fluorescence emission and far-UV circular dichroism were used to measure the unfolding process. Intrinsic fluorescence spectrofluorometry measures the fluorescence intensity of fluorophores such as tyrosine and tryptophan. Tryptophan and tyrosine are commonly used as probes to study the local tertiary structure of a protein (Carmel and Yaron, 1978). Intrinsic fluorescence spectroscopy is highly sensitive, requires lower protein concentrations and allows for the inspection of specific motifs within the protein. When a protein is excited at 280 nm, tyrosine and tryptophan residues within the protein are excited and emit a fluorescence signal. When the protein is excited at 295 nm, only the tryptophan residue emits a fluorescence signal. The emitted fluorescence intensity depends on how exposed the fluorophores are to the environment. The more solvent exposed the fluorophores are, the higher the emission wavelength. Protein unfolding gradually exposes the fluorophores to the environment, if a probe at a particular wavelength is considered,

then the mechanism of folding as a function of the exposure of the fluorophore can be measured. Intrinsic tryptophan fluorescence was used to monitor the tertiary structure of the protein as it unfolds.

As stated above intrinsic tryptophan fluorescence technique is a highly sensitive technique; therefore, a final concentration of ~2 µM of HIV-1 protease (WT•D25A and N37T↑V^{+10}•D25A) was used for the experiment. The PR was suspended in 10 mM sodium acetate buffer of pH = 5. Each protein was treated with a urea concentration range of 0 - 8 M to induce unfolding of the protease. A Jasco FP-300 Spectrofluorometer was used to excite the protease at a wavelength of 295 nm, and the emission spectra was measured between 280 nm and 500 nm for five accumulations. Other experimental parameters include an excitation bandwidth of 2.5 nm, an emission bandwidth of 2.5 nm, and a 0.5 nm data pitch. The raw data obtained from the spectrofluorometer was used to generate a fluorescence spectrum with the aid of Microsoft® Office Excel 2010.

Circular dichroism is a technique used to investigate a protein's secondary structure. The α-helices and β-sheets of the HIV-1 protease can absorb polarised light in the far-UV region of 180-240 nm because of their chirality, which is a geometrical, asymmetrical property of a molecule. In this study, 10 µM was used to probe loss of secondary structure during the unfolding of N37T↑V^{+10}•D25A. N37T↑V^{+10}•D25A was treated with a concentration range of 0 – 8 M and allowed to incubate at 20 °C for a minimum of one hour in order to reach equilibrium. After the protein had reached equilibrium, far-UV CD was used to monitor the unfolding of the protein at a wavelength of 218 nm.

2.10.3 Analysis of urea-induced equilibrium unfolding

The fluorescence raw data had to be analysed and fitted. The fluorescence which was monitored at 325 nm (F_{325}), to monitor the folded conformation and 355 nm (F_{355}) to monitor the unfolded conformation was plotted against a range of urea concentrations. The ratio of unfolded and folded, F_{355}/F_{325} was then calculated. To identify the population of unfolded protein and folded protein at any point, the fluorescence data of the ratio F_{355}/F_{325} ($y_{observed}$) had to be converted to fraction unfolded (f_U) and fraction

folded (f_F). The conformational states can be represented by the equation:

$$f_F + f_U = 1 \qquad \textbf{Equation 4}$$

A straight line was constructed at the pre-transition and post-transition region and the equations were used to calculate the signal for folded (y_F) and unfolded (y_U) protein, respectively. To obtain the signal for the conformational state at any point, the equation:

$$y_{observed} = (y_F \bullet f_F) + (y_U \bullet f_U) \qquad \textbf{Equation 5}$$

was used. Using the information stated above, the equation for fraction unfolded becomes:

$$f_U = (y_F - y_{obs})/(y_F - y_U) \qquad \textbf{Equation 6}$$

The equilibrium constant, *K*, for every point in the transition region for a dimer, was calculated based on Bowie and Sauer, (1989) principles. And the equation for equilibrium is:

$$K = [U]^2/[N] = f_U^2/(1-f_U) = f_U/f_F \qquad \textbf{Equation 7}$$

According to Schellman, (1978) and Pace and Scholtz, 1997) the Gibbs free energy for unfolding is described by the equation:

$$\Delta G = -RT \ln(K) \qquad \textbf{Equation 8}$$

Where ΔG is the change in free energy, T is the absolute temperature, R is the universal gas constant of 8.314 J/mol/K and *K* is the equilibrium constant. If the free energy is calculated in the absence of denaturant the equation:

$$\Delta G = \Delta G_{water} - m[U] \qquad \textbf{Equation 9}$$

may be used, where ΔG_{water} is the free energy change in the absence of denaturant, m is the cooperativity of the reaction which relates to solvent accessible surface area of the protein during unfolding, and [U] refers to the concentration of the denaturant. Therefore, if equation 8 is substituted into equation 9, we get equation 10 as:

$$f_U = (y_N + y_U \bullet e^{-\Delta G \ = \Delta Gwater \ - m[U].RT}) / (1 + e^{-\Delta G \ = \Delta Gwater \ - m[U].RT})$$

f_U was plotted against concentration of urea. The data for unfolding fitted to this equation was then used to calculate ΔG for unfolding, the m-value and denaturation midpoint.

Chapter 3: Results

3.1 Verification of WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease plasmid DNA

Sanger sequencing confirmed that N37T↑V^{+10}•D25A HIV-1 PR cDNA, which was inserted in the pET-11a vector, successfully incorporated the nucleic acid sequence encoding for the insertion of valine and the substitution of asparagine for threonine at 37th amino acid position, as shown in Figure 11. The sequence results further confirmed the presence of ten background mutations: namely, I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L. The sequence alignment of the amino acid residues showed that N37T↑V^{+10}•D25A had 100 amino acid while WT•D25A had 99 amino acids. The D25A mutation, which was included to prevent autoproteolysis was also present in N37T↑V^{+10}•D25A. There was also the Q7K mutation which is normally incorporated in wet-lab studies to limit the autocatalytic function of the HIV-1 PR.

3.2 Growth curve analysis of WT•D25A and N37T↑V^{+10}•D25A

The growth rate of *E. coli* BL21 (DE3) pLysS cells housing WT•D25A PR and *E. coli* Rosetta cells with N37T↑V^{+10}•D25A HIV-1 PR were analysed using growth curves (Figure 12). The early-log phase period was determined in order to optimise the time for induction for each protein overexpression. For WT•D25A, early log-phase of bacterial growth was found to occur after 120 min, and here the optical density at 600 nm (OD_{600}) was between 0.5 and 0.8 AU (Figure 12A). For N37T↑V^{+10}•D25A, the OD_{600} at early-log phase was between 0.58 and 0.90 AU (Figure 12B). This was regarded as the optimum OD_{600} for the addition of IPTG resulting in the induction of protein overexpression.

3.3 Expression trials for N37T↑V^{+10}•D25A

The analysis of the induction trial samples ran on a tricine SDS-PAGE gel revealed that the N37T↑V^{+10}•D25A PR, in *E. coli* Rosetta cells, was primarily expressed in the insoluble fraction, as it was predominantly found in the pellet. Four (4) hours was identified as the minimum induction period required after induction with 0.8 mM IPTG for the successful overexpression of N37T↑V^{+10}•D25A PR. The four gels illustrating the expression trials are shown in Figure 13.

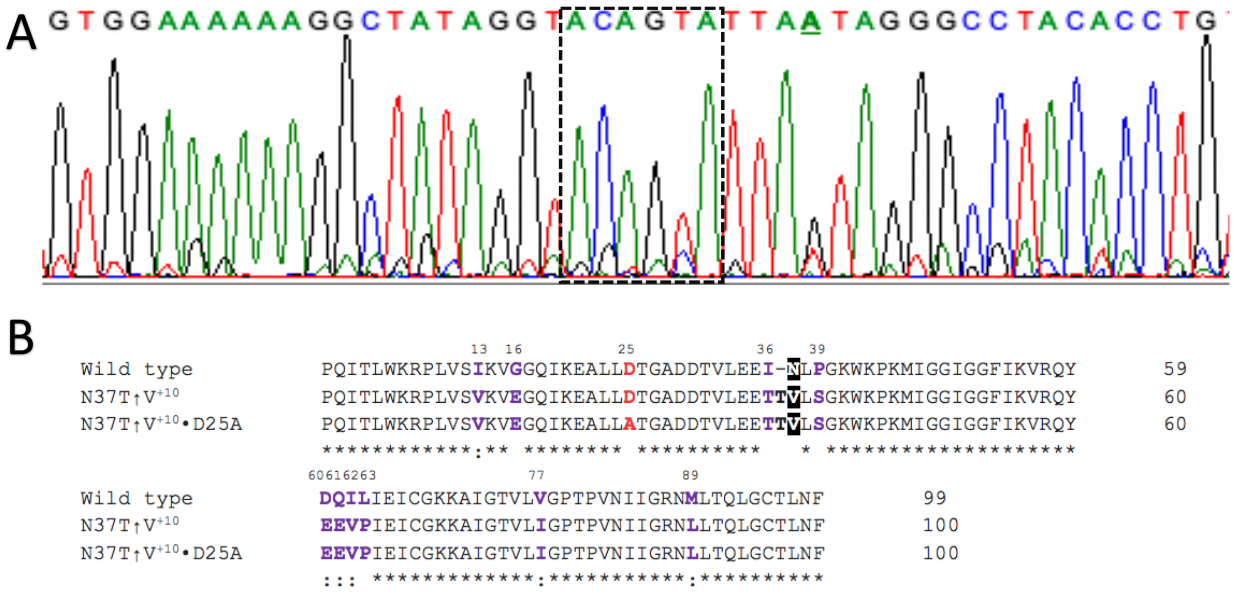


Figure 11: Sequence identity of N37T↑V^{+10}•D25A

A) Chromatogram of WT•D25A and N37T↑V^{+10}•D25A DNA sequenced by Sanger sequencing at Inqaba Biotech® (Pretoria, South Africa). The black box illustrates the DNA sequence of the N37T substitution and ↑V insertion. **B)** The sequence alignment of WT, N37T↑V^{+10} and N37T↑V^{+10}•D25A. The position of the D25A mutation is shown in orange letters, the valine insertion is highlighted in black and the 10 background mutations are shown as purple letters.

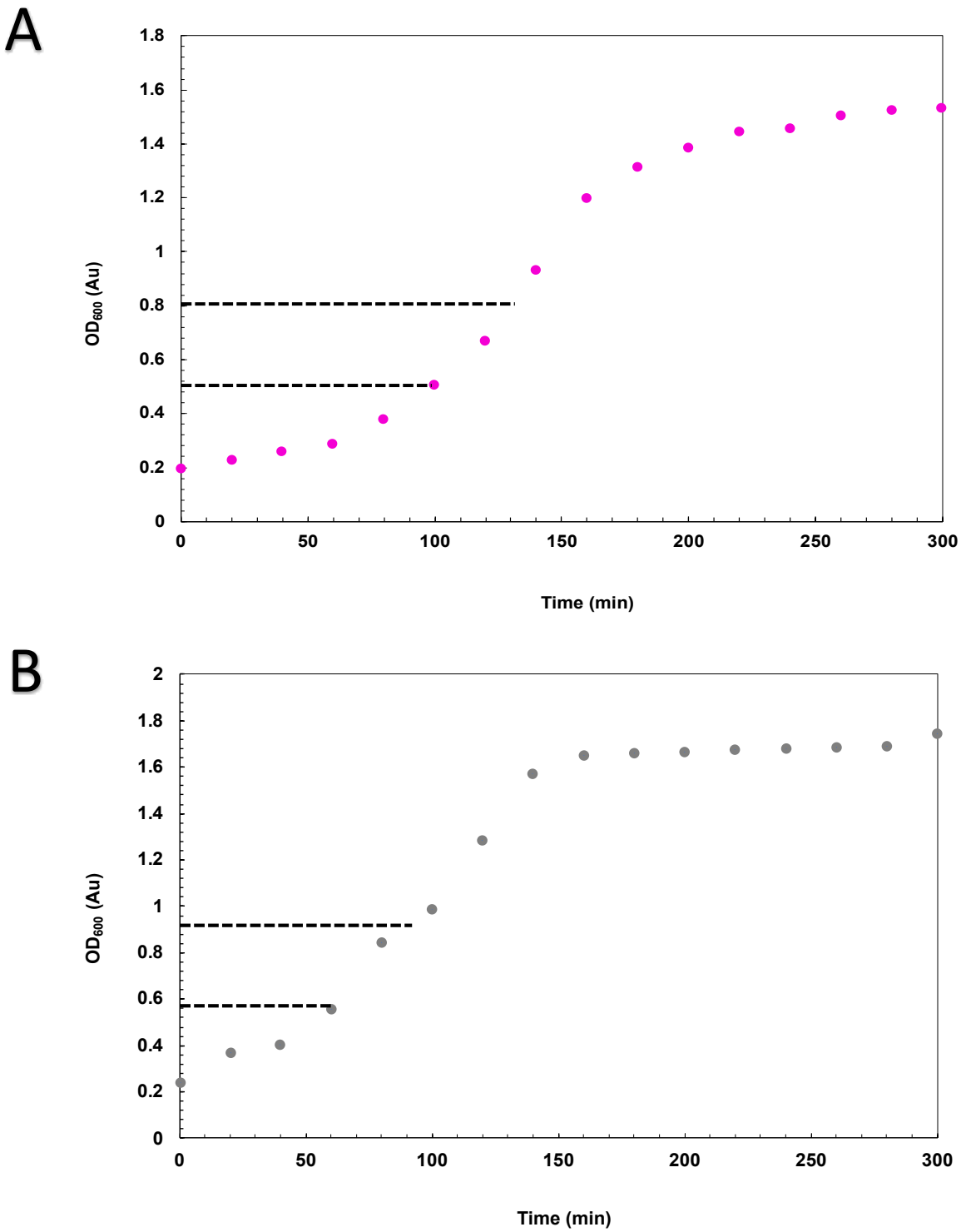


Figure 12: Growth curves of transformed *E. coli*

A) BL21(DE3) pLysS with vector encoding WT•D25A and B) Rosetta (DE3) cells with a vector encoding for N37T↑V^{+10}•D25A HIV-1 protease. The black dotted lines symbolise the early log phase region from the start of the log-phase to mid-log phase, between 70 and 100 min. Early-log phase is the optimum period for induction.

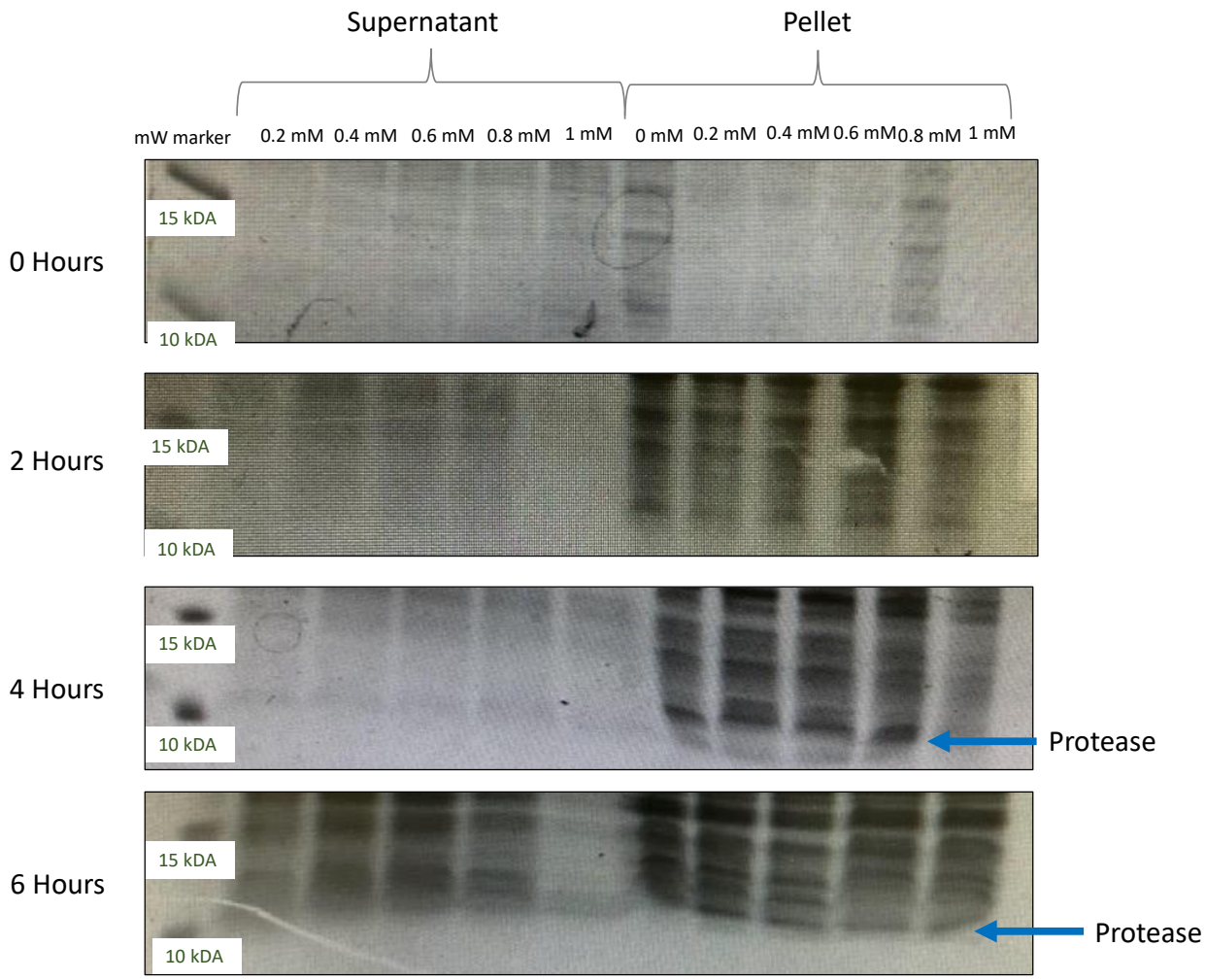


Figure 13: Induction trials 16% SDS-PAGE tricine gel

showing the induction trials from 0 to 6 hours, over a range of IPTG concentrations ranging from 0 to 1 mM. This was all conducted at a temperature of 37 ^{o}C. The HIV-1 protease is expressed in the pellet (insoluble fraction) and not the supernatant. This is because it is expressed as protein aggregates known as inclusion bodies. The HIV protease was expressed after 4 hours at a IPTG concentration of 0.8 mM at a constant temperature of 37 ^{o}C.

3.4 Purification of WT•D25A and N37T↑V^{+10}•D25A

3.4.1 Purification of WT•D25A

Following overexpression in *E. coli* cells the WT•D25A PR protease was purified using a cation-anion exchange chromatographic technique recently reported by Sherry *et al.*, (2018). The purification profile is illustrated in Figure 14. The crude sample was first collected by centrifugation, washed in wash buffer and then suspended in unfolding buffer containing 8 M urea. The now unfolded protein was passed through the CM-sepharose column that was in tandem with the DEAE column. The WT•D25A PR bound to the CM column and the contaminating proteins were either collected in the flow through or bound the DEAE column. The second step of the purification saw the protein being passed through the CM column alone and the PR bound the column after which it was eluted using a salt gradient of 0 – 1 M storage buffer with NaCl. The PR was collected, concentrated using a Macron spin column and stored at -20 °C.

3.4.2 Purification of N37T↑V^{+10}•D25A

An optimised cation exchange chromatography technique was used for the purification of N37T↑V^{+10}•D25A. The overexpressed protein was washed, unfolded, refolded and then passed through the CM-Sepharose column. Contaminating protein were collected at the flow through, and UV signal at 280 nm was used to collect the N37T↑V^{+10}•D25A fractions were then pooled, concentrated and stored at -20 °C. Figure 15 illustrates the elution chromatogram profile of the purification of N37T↑V^{+10}•D25A.

3.5 Purity and molecular concentration determination

The purity and monomeric size of the WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease were assessed using a 16% (w/v) Tricine-SDS-PAGE gel (Figure 16). The migration pattern of the protease through the gel was determined by comparing it with the Page Ruler™ Pre-stained Protein ladder protein standards (10 to 180 kDa). The approximate molecular weight of both WT•D25A and N37T↑V^{+10}•D25A proteases

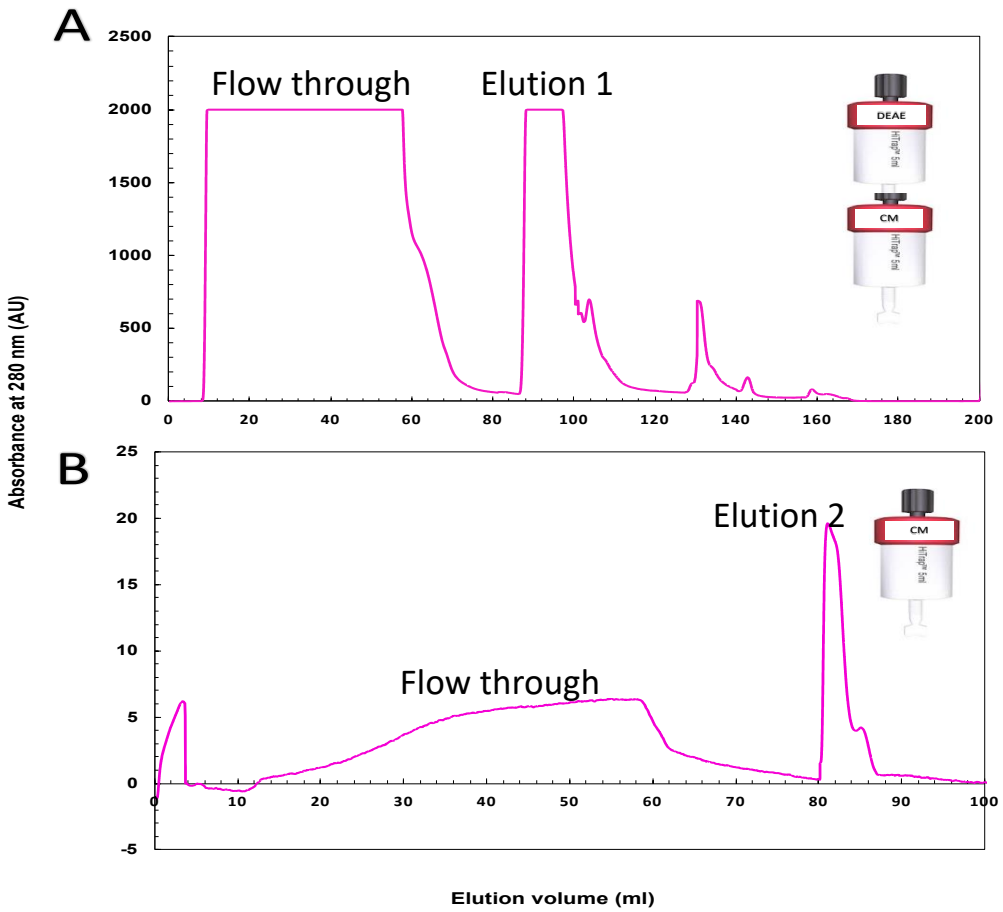


Figure 14: Two-step purification profile of WT•D25A HIV-1 protease

A) The first step of purification involved a tandem DEAE and CM ion exchange column. The lysate was passed through the column, and the elution profile is shown. The first peak corresponds to the flow-through fraction, containing proteins that did not bind to the column. The second peak represents the collected fraction, which includes the unfolded HIV protease along with a mixture of other proteins. **B)** The second step of purification employed cation exchange chromatography alone. The WT•D25A PR was eluted using a salt gradient of 0 - 1 M NaCl. The first wide peak observed from 10 ml to 80 ml corresponds to the flow-through fraction, consisting of contaminating proteins. The subsequent peak between 80 ml and 90 ml represents the purified HIV-1 protease, with the shoulder indicating different mobility properties of the unfolded protein. After refolding, a single species was observed thus indicating that these are the same species. This two-step purification strategy effectively isolates the WT•D25A HIV-1 protease from the initial lysate, with the final purification step achieving a higher degree of purity.

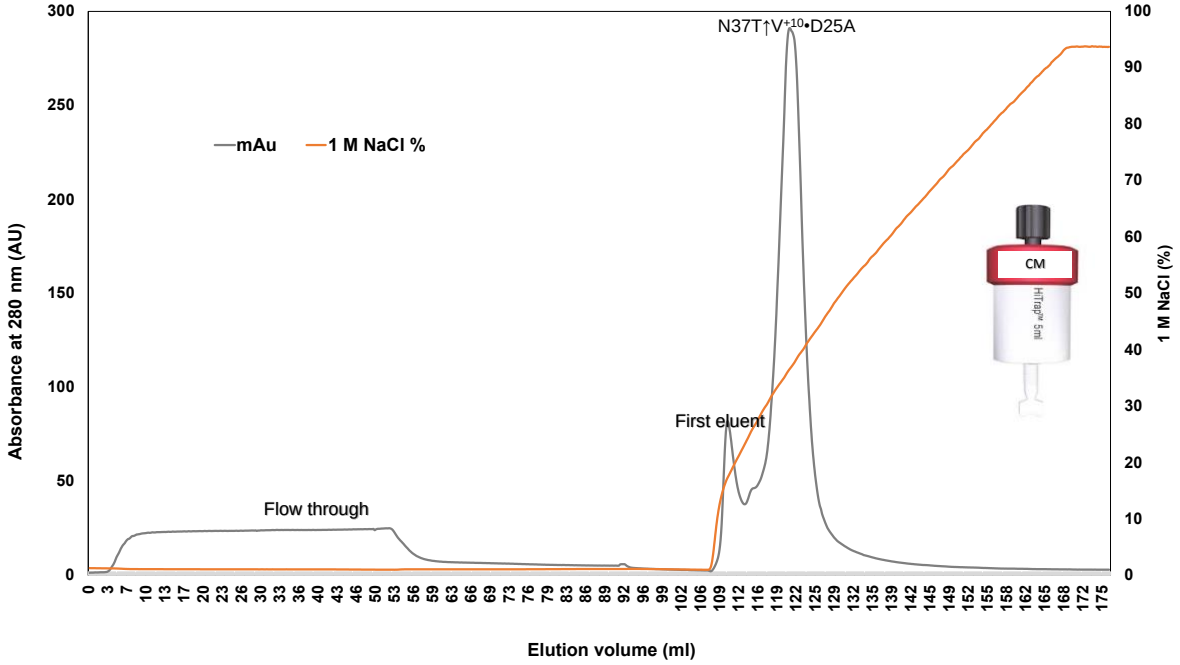


Figure 15: Purification chromatogram of the optimised purification of N37T↑V^{+10}•D25A

Cation exchange chromatography was employed to purify the protein, and a folded protein was eluted using a carefully controlled salt gradient. The chromatogram illustrates the elution profile, with the flow-through fraction observed between 3 and 56 ml. A contaminating peak is evident between 109 and 116 ml, while the target protease was successfully collected from 119 to 132 ml. This chromatogram demonstrates the effectiveness of the purification method in isolating the desired N37T↑V^{+10}•D25A protease.

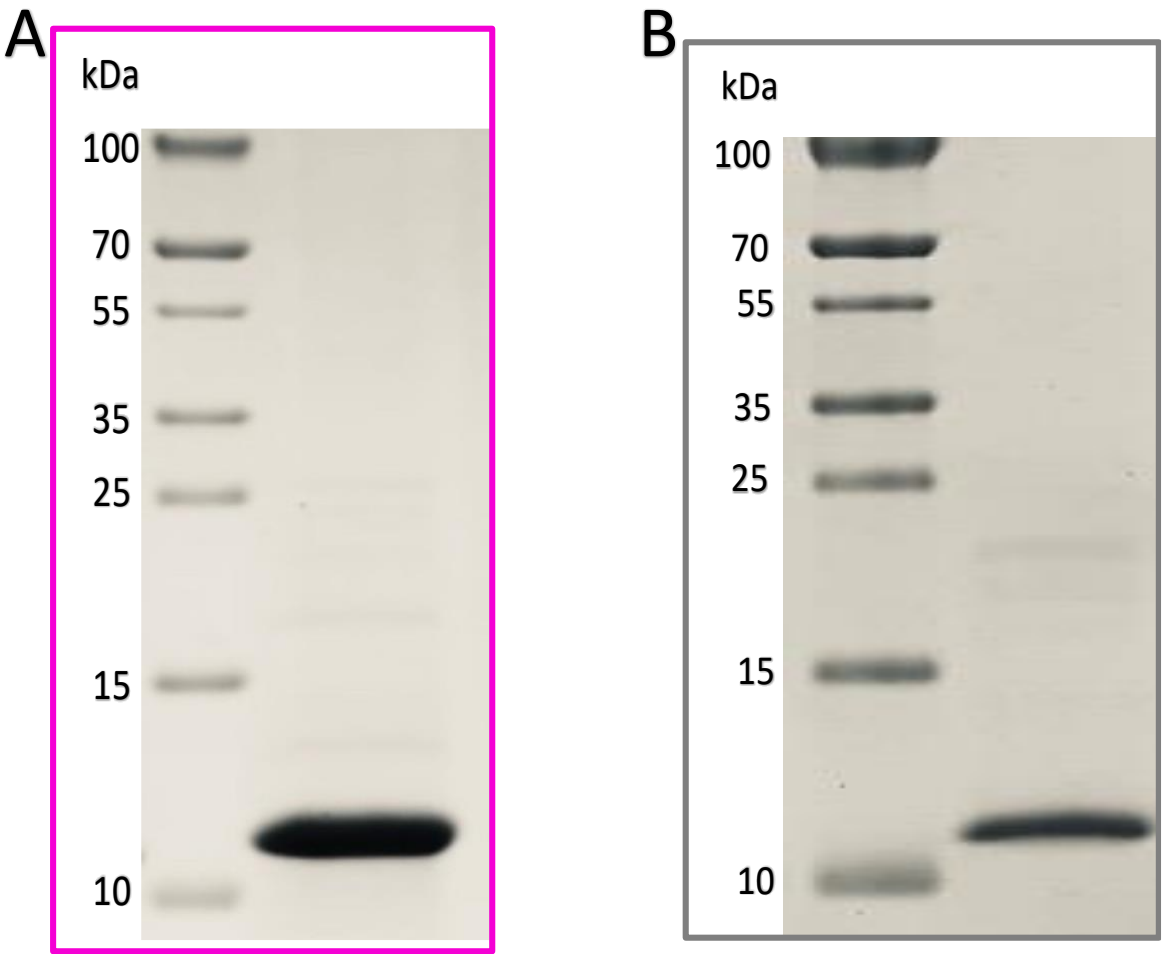


Figure 16: Tricine 16% SDS-PAGE gel analysis of the WT•D25A and N37T↑V^{+10}•D25A HIV-1 protease purity

SDS-PAGE gel electrophoresis was performed to evaluate the purity of the **A)** WT•D25A and **B)** N37T↑V+10•D25A HIV-1 protease. The gel analysis confirms the successful isolation of pure proteins for both WT•D25A and N37T↑V^{+10}•D25A PR. The SDS-PAGE results demonstrate the presence of a single band corresponding to the expected molecular weight of approximately 11 kDa for both proteases, indicating their purity.

was determined to be ~11 kDa, consistent with previous studies on HIV-1 protease and in agreement with the predicted monomeric size (Mosebi *et al.* 2008; Sherry *et al.,* 2020). The analysis of the gel revealed a lack of contaminating bands in the gel lanes, indicating a high level of purity for the protease samples. Additionally, there was no evidence of proteolysis, as evidenced by the absence of smaller bands below the HIV-1 protease band, which is typically observed during proteolysis.

To quantify the concentrations of WT•D25A and N37T↑V^{+10}•D25A proteases, absorbance spectroscopy was performed (Figure 17) along with the application of the Beer-Lambert equation. Serial dilution of the protease samples was carried out, and the absorbance at 280 nm was measured. By utilising the Beer-Lambert equation, the concentration of WT•D25A PR was determined to be 35 µM, while the concentration of N37T↑V^{+10}•D25A PR was found to be 27 µM. The accurate quantification of protease concentrations was achieved through absorbance spectroscopy in conjunction with the Beer-Lambert equation.

3.6 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was used to accurately determine the concentration of PR by correcting for its apparent concentration. This was accomplished by titrating the PR with acetyl pepstatin, a ligand that binds to the active site of the protease in a 1:1 ratio. A calorimetric profile was produced (Figure 18) and allowed for an accurate estimation of the HIV-1 PR concentration in the properly folded, functional form. Acetyl pepstatin specifically binds to aspartyl proteases, making it suitable for this study. By analysing the obtained thermodynamic parameters, valuable insights into the functional properties of the protein was gained.

The stoichiometry parameter (*n*) was used to determine the ratio of PR molecules capable of binding the ligand, which was then converted to percentages. For the WT•D25A PR variant, the *n*-value was determined to be 0.574, indicating that 57.4% of the WT•D25A PR molecules are correctly folded. Similarly, the N37T↑V^{+10}•D25A variant had an n-value of 0.901, indicating that 90.1% of the N37T↑V^{+10}•D25A PR molecules are correctly folded. By correcting for misfolding, the concentration was

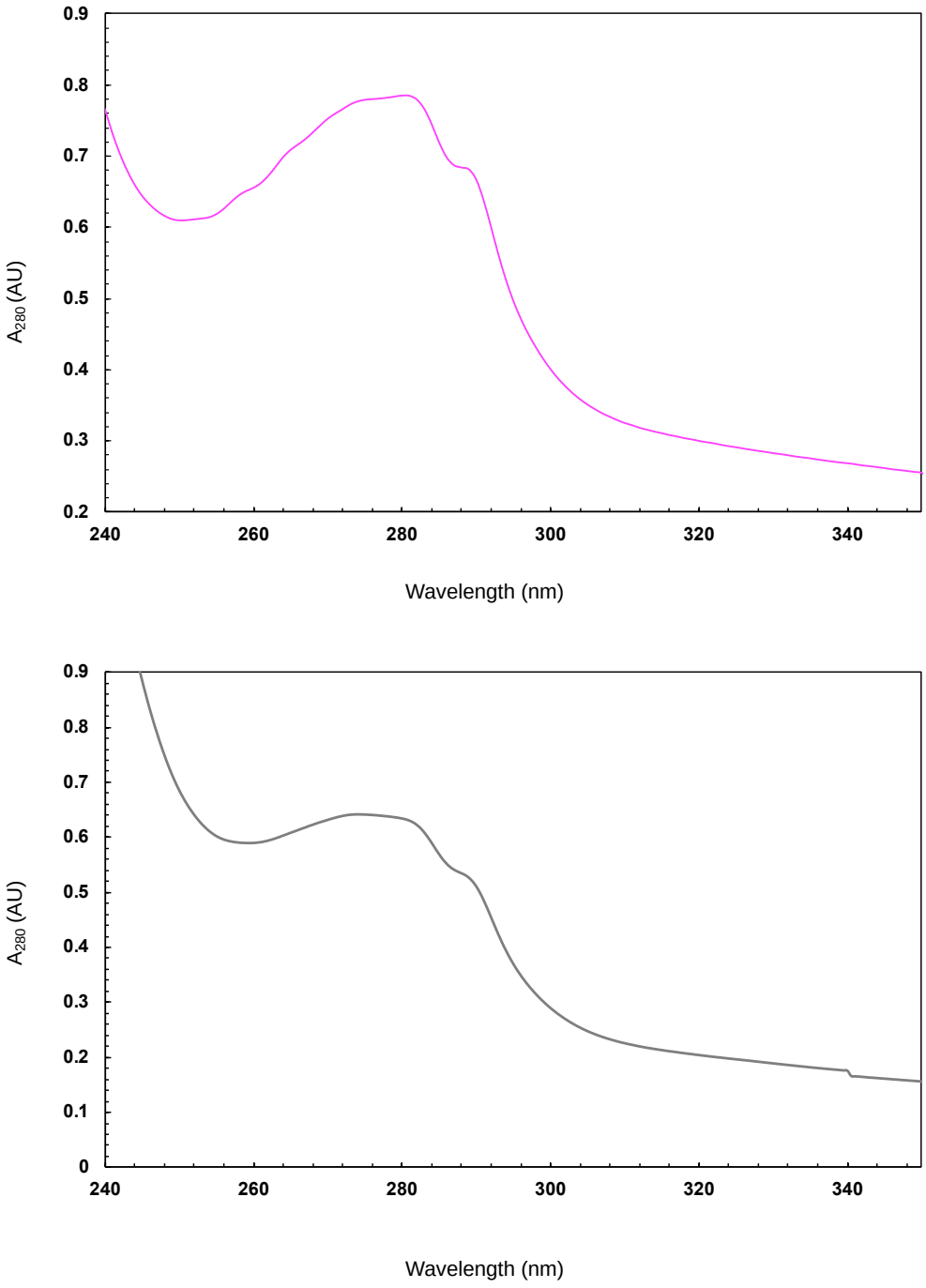
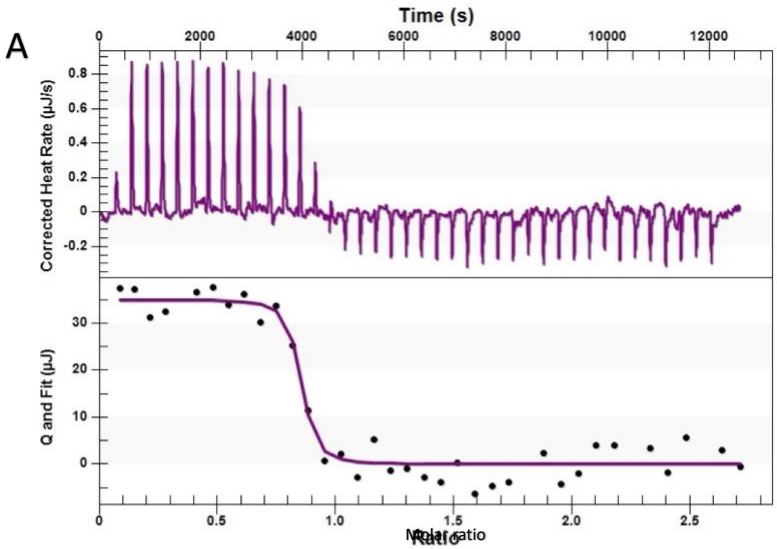
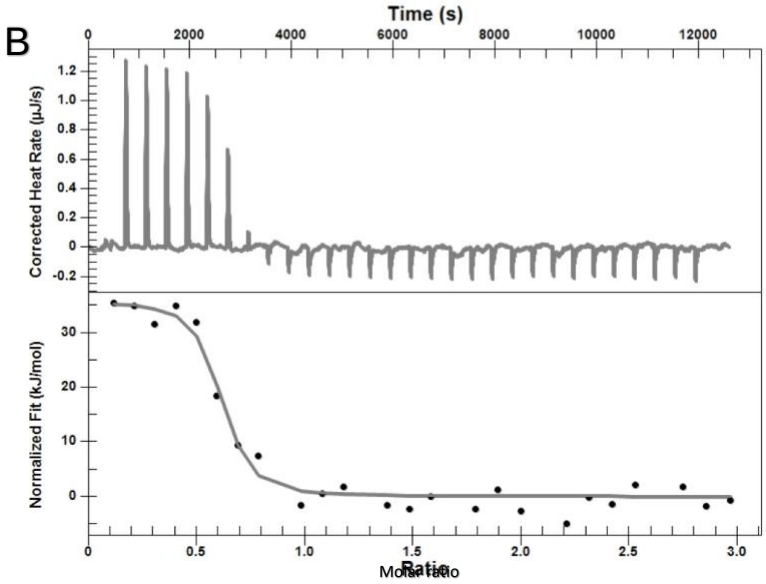


Figure 17: Protein concentration determination

A) WT•D25A and B) N37T↑V^{+10}•D25A HIV-1 protease using absorbance spectroscopy.. The Beer-Lambert law was used to calculate the concentration at 280 nm.





C

Protein	Initial concentration (µM)	n	Δ" (kJ/mol)	K_d (M)	Δ# (J/mol. K)	Corrected concentration (µM)	Confidence level
WT•D25A	35	0.574 ± 0.017	35.94 ± 2.72	$3.12 \times 10^{-7} \pm 6.52 \times 10^{-8}$	256.7	20	95%
N37T↑V^{+10}•D25A	27	0.904 ± 0.021	36.02 ± 1.437	$5.67 \times 10^{-7} \pm 3.637 \times 10^{-8}$	261.6	24	95%

Figure 18: Calorimetric profile of active site titration by acetyl pepstatin

A) WT•D25A and B) N37T↑V^{+10}•D25A, show the thermograms of the active site titration. The top panel is the raw data generated from the titration with acetyl pepstatin and the bottom panel shows the integrated heats for the above peaks plotted against a molar ratio of acetyl pepstatin PR molecule. **C)** The thermodynamic parameters of WT•D25A and N37T↑V^{+10}•D25A PR obtained from the titrations.

adjusted, resulting in corrected concentrations of 20 µM for WT•D25A and 24 µM for N37T↑V^{+10}•D25A. The change in enthalpy (ΔH) for both PRs when titrated with acetyl pepstatin was similar, $\Delta H_{WT•D25A}$ = 35.94 kJ/mol and $\Delta H_{N37T↑V^{+10}•D25A}$ = 36.02 kJ/mol. The change in entropy (ΔS) for WT•D25A was 256.7 J/mol·K and 261.6 J/mol·K for N37T↑V^{+10}•D25A. The dissociation constant (K_d) was calculated to be 3.12 x 10^{-7} M for WT•D25A and 5.67 x 10^{-7} M for N37T↑V^{+10}•D25A.

3.7 Activity assay

An activity assay was conducted by monitoring the cleavage of a chromogenic substrate, Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2. The substrate fluorophore group (Abz) was excited at 337 nm and monitored at the emission wavelength of 425 nm. Both WT•D25A and N37T↑V^{+10}•D25A had fluorescence intensities of 0 AU over the course of the reaction period. This is indicative that the chromogenic substrate was not cleaved and relates to an inactive enzyme incapable of hydrolysis. The D25A mutation is known to abolish activity (Sherry, *et al.* 2021). This assay was done to prove that both PRs were inactive. Therefore, no positive control was added.

3.8 Structural characterisation

3.8.1 Secondary structure

Recently, Sherry *et al.* (2020) conducted a crystallography study that showed that the HIV-1 subtype C WT•D25A secondary structure was primarily composed of β-sheets and had two alpha helices. Upon homology modelling (Figure 19), N37T↑V^{+10}•D25A PR was predicted to be predominantly composed of β-sheets. To further evaluate the secondary structure of N37T↑V^{+10}•D25A PR, far-UV CD was employed. This would give further insight in how mutations present in N37T↑V^{+10}•D25A PR alter/impact the secondary structure of HIV-1 PR.

A Jasco J-1500 spectropolarimeter was used to assess the absorbance of circularly polarised light at the far-UV range of 250 to 190 nm. The resulting spectrum in Figure 20A for WT•D25A demonstrated a trough at 216 nm, indicative of a protein that

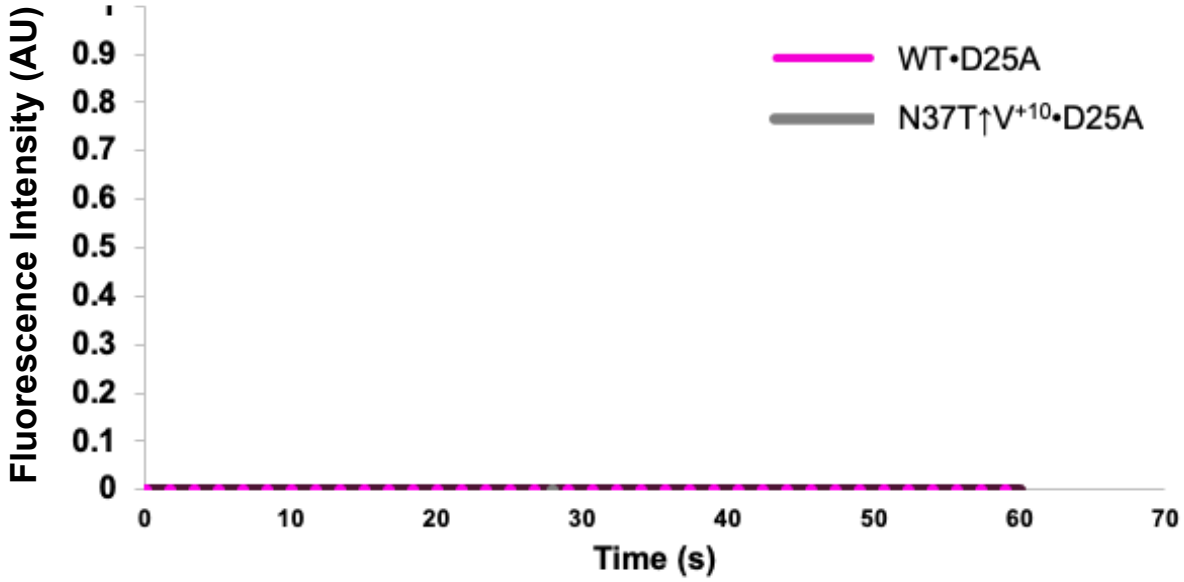


Figure 19: Activity assay for WT•D25A and N37T↑V^{+10}•D25A

Fluorescence intensity of 2 µM WT•D25A and N37T↑V^{+10}•D25A PR mixed separately with 500 µM of the chromogenic substrate, Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2 was excited at 337 nm and monitored at 425 nm over 60 seconds. There was no increase in fluorescence indicating that the chromogenic substrate was not cleaved.

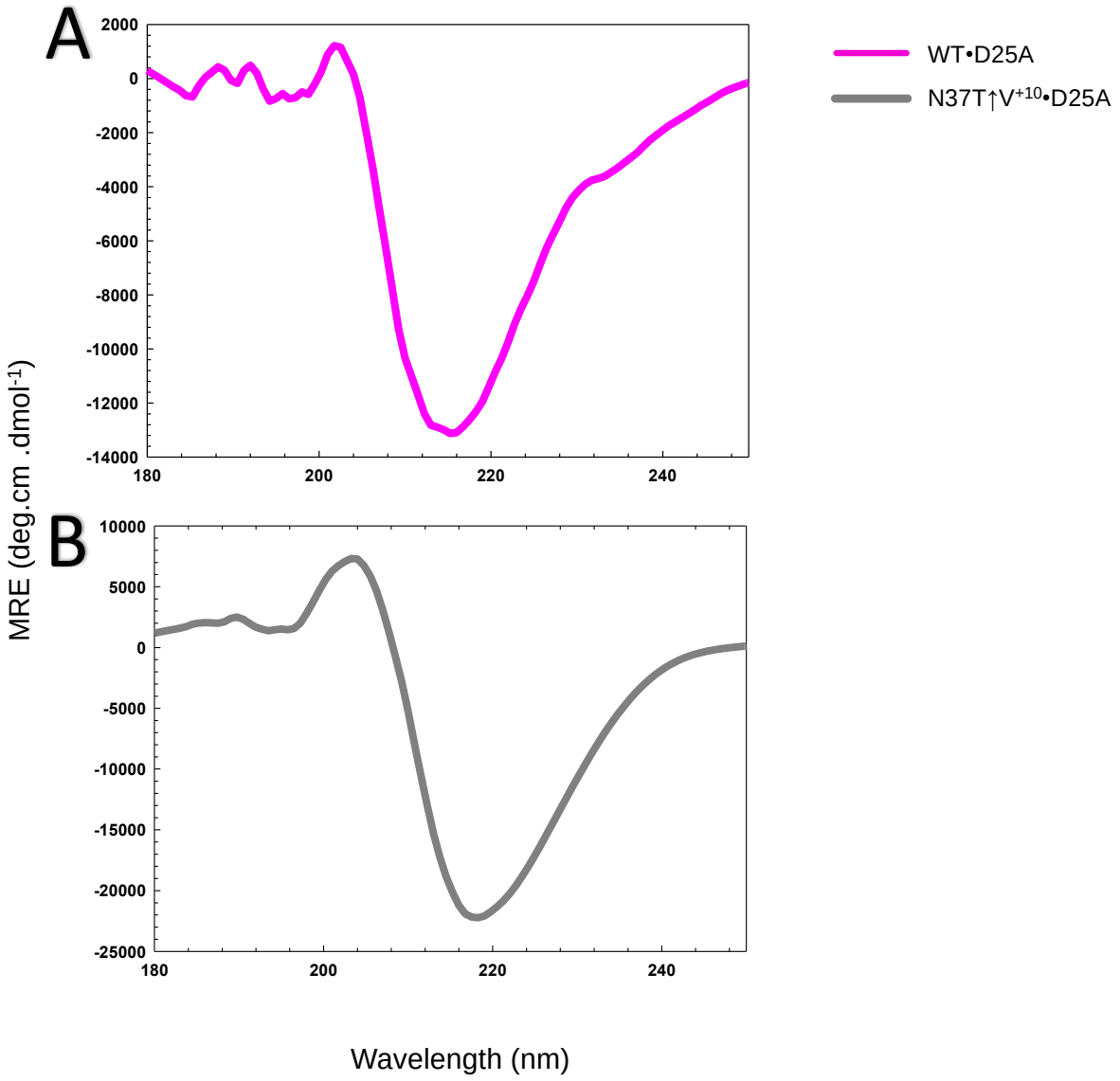


Figure 20: Evaluation of secondary structure of WT•D25A and N37T↑V^{+10}•D25A mutants using far-UV CD.

(A) Far-UV CD spectrum of WT•D25A exhibiting a trough at 216 nm and a peak at 204 nm, indicating a predominantly β-sheeted protein with random coils. **(B)** Far-UV CD spectrum of N37T↑V^{+10}•D25A mutant, primarily composed of β-sheets, with a less intense MRE compared to the wild type. The differences in MRE may be attributed to differences in concentration or protease folding.

is primarily β-sheeted (Böhm *et al.*, 1992; Greenfield, 2007), with a peak at 204 nm further confirming the presence of β-strands and random coils (Greenfield, 2007). Similarly, the far-UV CD spectrum for N37T↑V^{+10}•D25A in Figure 20B showed a protein that is predominantly β-sheeted, although the MRE was less intense compared to the wild type, possibly due to differences in concentration or folding. Overall, these results support the previous findings that both WT•D25A and N37T↑V^{+10}•D25A mutants are primarily composed of β-sheets.

3.8.2 Tertiary structure

The tertiary structure of the proteins was assessed using intrinsic fluorescence spectroscopy, utilizing four tryptophan residues (Trp6, Trp6', Trp 42/43, and Trp 42'/43') as probes. Fluorescence spectra of the native WT•D25A and N37T↑V^{+10}•D25A PR, excited at 295 nm, are presented in Figure 21. The emission peaks were observed at 346 nm and 347 nm for WT•D25A and N37T↑V^{+10}•D25A PR, respectively. These red-shifted peaks (>340 nm) suggest that the Trp residues are partially solvent-exposed.

To investigate the emission characteristics of unfolded PR, the proteins were suspended in 8 M urea. The resulting fluorescence spectra of the unfolded PR exhibited peaks at 350 nm and 352 nm for WT•D25A and N37T↑V^{+10}•D25A, respectively. These emission peaks are in the far, red-shifted region align with previously reported fluorescence intensity peaks of Trp residues when exposed to the solvent (Lakowicz, 1999). To determine if the unfolding of PR in urea is reversible, the recoverability of the native structure had to be assessed. Therefore WT•D25A and N37T↑V^{+10}•D25A denatured in 8 M urea was diluted in 10 mM sodium acetate, 0.02% (w/v) sodium azide, 150 mM NaCl at a pH = 5 and allowed to refold for 60 minutes. The recoverability of the tertiary structure was assessed using intrinsic tryptophan fluorescence. The refolded WT•D25A and refolded N37T↑V^{+10}•D25A had peaks of 346 and 347 nm, respectively. The recoverability of the tertiary structure for both WT•D25A and N37T↑V^{+10}•D25A following unfolding was calculated to be >90%, which is highly recoverable.

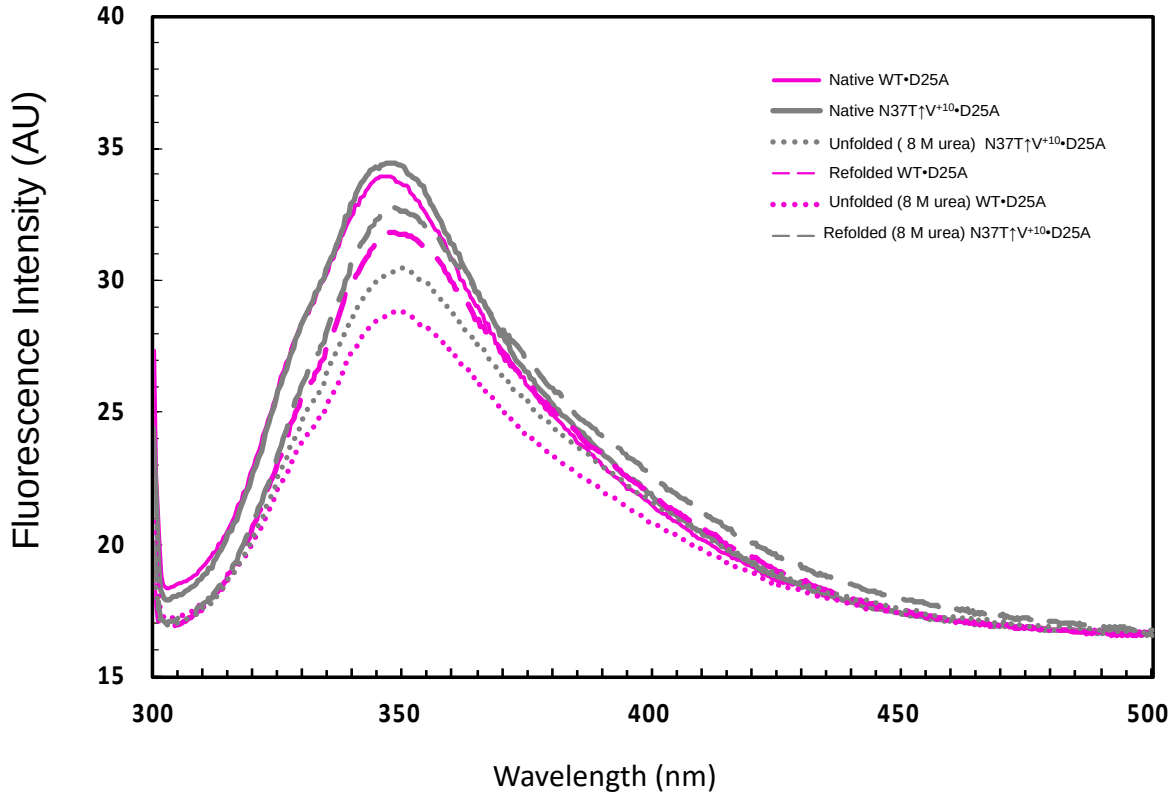


Figure 21: Tertiary structure and the recoverability of WT•D25A and N37T↑V^{+10}•D25A after denaturation with urea (8 M).

Tertiary structure was monitored by intrinsic tryptophan fluorescence at an excitation wavelength of 295 nm. Native WT•D25A and N37T↑V^{+10}•D25A PR had an emission peak of 346 nm and 347 nm, respectively. The unfolded WT•D25A and N37T↑V^{+10}•D25A PR, shown in dotted lines, had a more red-shifted spectrum with emission peaks at 351 and 352, respectively. This was accompanied by a decreased fluorescence intensity. The recoverability of the tertiary structure is shown in dashed lines and was calculated to be >98% for both proteins. Refolded WT•D25A had an emission peak of 346 nm and N37T↑V^{+10}•D25A had an emission peak of 347 nm, similar to the native PRs.

3.8.3 Quaternary structure

The hydrodynamic volume of a protein gives insights on the molecular weight and oligomeric state of a protein. Overall, this gives us information on the quaternary structure of the protein. HIV-1 PR is homodimeric in its functional form and mutations may alter folding and impact the formation of a homodimer. Therefore, it was necessary to assess the impact the insertion and substitution mutations have on the overall quaternary structure of the PR.

SE-HPLC with a TSKgel SuperSW2000 SEC column was used to assess the hydrodynamic volume of WT•D25A and N37T↑V^{+10}•D25A. An elution buffer containing: 10 mM sodium acetate, 0.01% (w/v) sodium azide, containing 200 mM NaCl was used to equilibrate the system. Protein standards of known molecular weights were used to calibrate the column (Figure 22A). A calibration curve was constructed using the logarithm of the molecular weight and retention time of each protein standard. This allowed for the molecular size of WT•D25A and N37T↑V^{+10}•D25A to be estimated relative to the protein standards sizes (Figure 22B).

The equation of the standard curve was y = -0.2198x + 5.5566 and an R^2 of 0.9842. The elution of WT•D25A and N37T↑V^{+10}•D25A was monitored over 30 minutes and the elution profile (Figure 22C) plotted. Using this standard curve equation, the apparent molecular weight of WT•D25A and was calculated as 21.7 kDa for WT•D25A and N37T↑V^{+10}•D25A was 22.6 kDa.

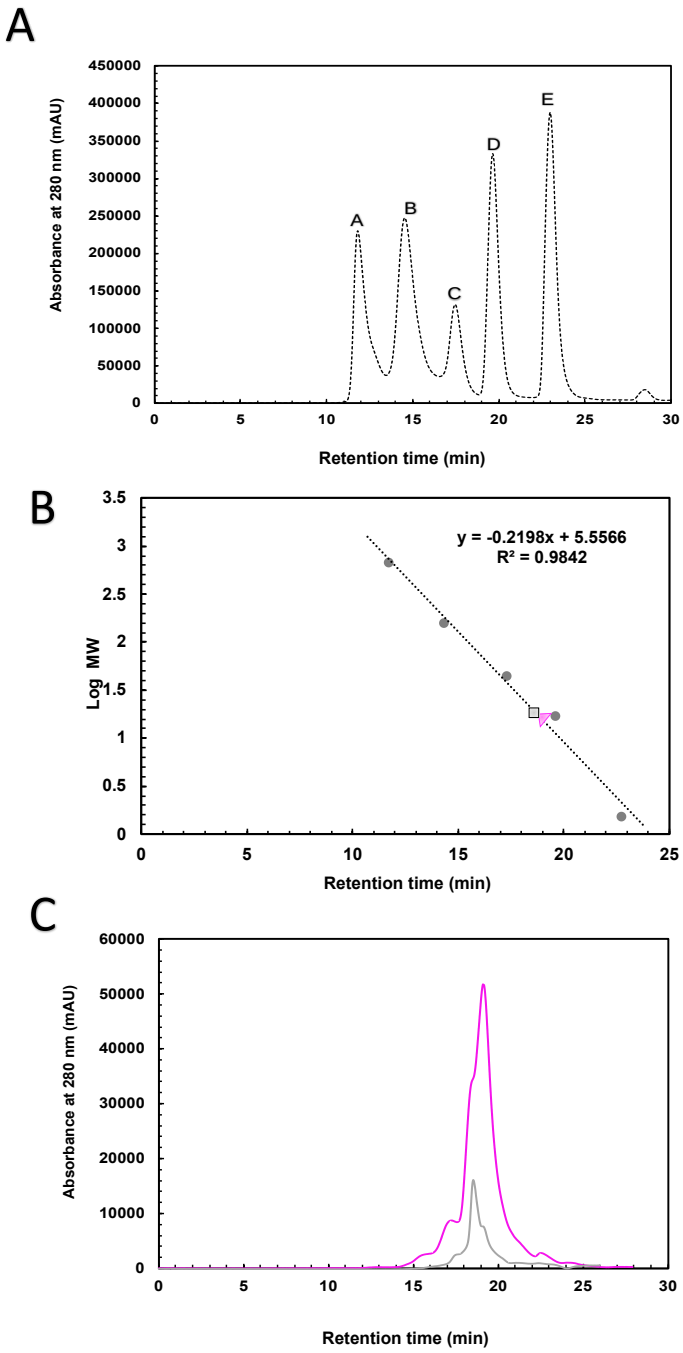


Figure 22: SE-HPLC chromatogram of N37T↑V^{+10}•D25A PR against standard proteins.

A) The elution profile of the protein standards; A. thyroglobulin (670 kDa), B. gamma (158 kDa), C. ovalbumin (44 kDa), D. myoglobulin (17 kDa) and E. vitamin B12 (1.5 kDa). **B)** Standard curve constructed using the retention time and the logarithm of the molecular weight of the protein standards, retention time of WT•D25A (purple triangle) and N37T↑V^{+10}•D25A (grey square) are annotated on the standard curve. **C)** Elution profile of WT•D25A (purple) and N37T↑V^{+10}•D25A (grey) HIV-1 PR.

3.8.4 Macromolecular crystallisation

The hanging drop vapour diffusion method was used for crystallisation trials to grow macromolecular crystals of N37T↑V^{+10}•D25A PR. The HR2-144 Hampton® Crystallisation screening kit was used to identify the optimal conditions for growth of N37T↑V^{+10}•D25A crystals. A concentration range of 2 – 5 mg/ml of N37T↑V^{+10}•D25A PR in sodium acetate buffer (pH = 5) was used in the experiment. Triagonal bi-pyramidal crystals (Figure 23) were found after a minimum of 4 days in 0.1 M Bis-Tris, pH = 6 with 10% w/v polyethylene glycol (PEG) MME 5000 and also in 0.1 M Tris pH = 8.5 with 25% PEG 3350. Various other crystal shapes were also observed and these are attributed to different space groups of N37T↑V^{+10}•D25A. The crystals were unable to be diffracted by x-ray diffraction due to time constraints; however, these conditions may be optimised and used in future to grow and diffract N37T↑V^{+10}•D25A crystals to solve the three-dimensional structure of N37T↑V^{+10}•D25A PR.

3.9 Structural stability of the HIV-1 protease

3.9.1 Thermal unfolding

Thermal unfolding of WT•D25A and N37T↑V^{+10}•D25A PR was monitored at a wavelength of 216 nm, which is the wavelength at which the most prominent trough for predominantly β-sheeted HIV-1 protease occurs. The data was collected at increments of 1 °C. The experiment was conducted using the biological triplicate approach to enhance the reproducibility and reliability of the results. Both WT•D25A and N37T↑V^{+10}•D25A PR unfolded from the native to the denatured states over a single transition of native (N) to the unfolded (U) state (N→ U) (Figure 24). The WT•D25A had an average unfolding transition region between 55 °C and 65 °C. Contrastingly, N37T↑V^{+10}•D25A had an unfolding transition between 48 °C and 68 °C.

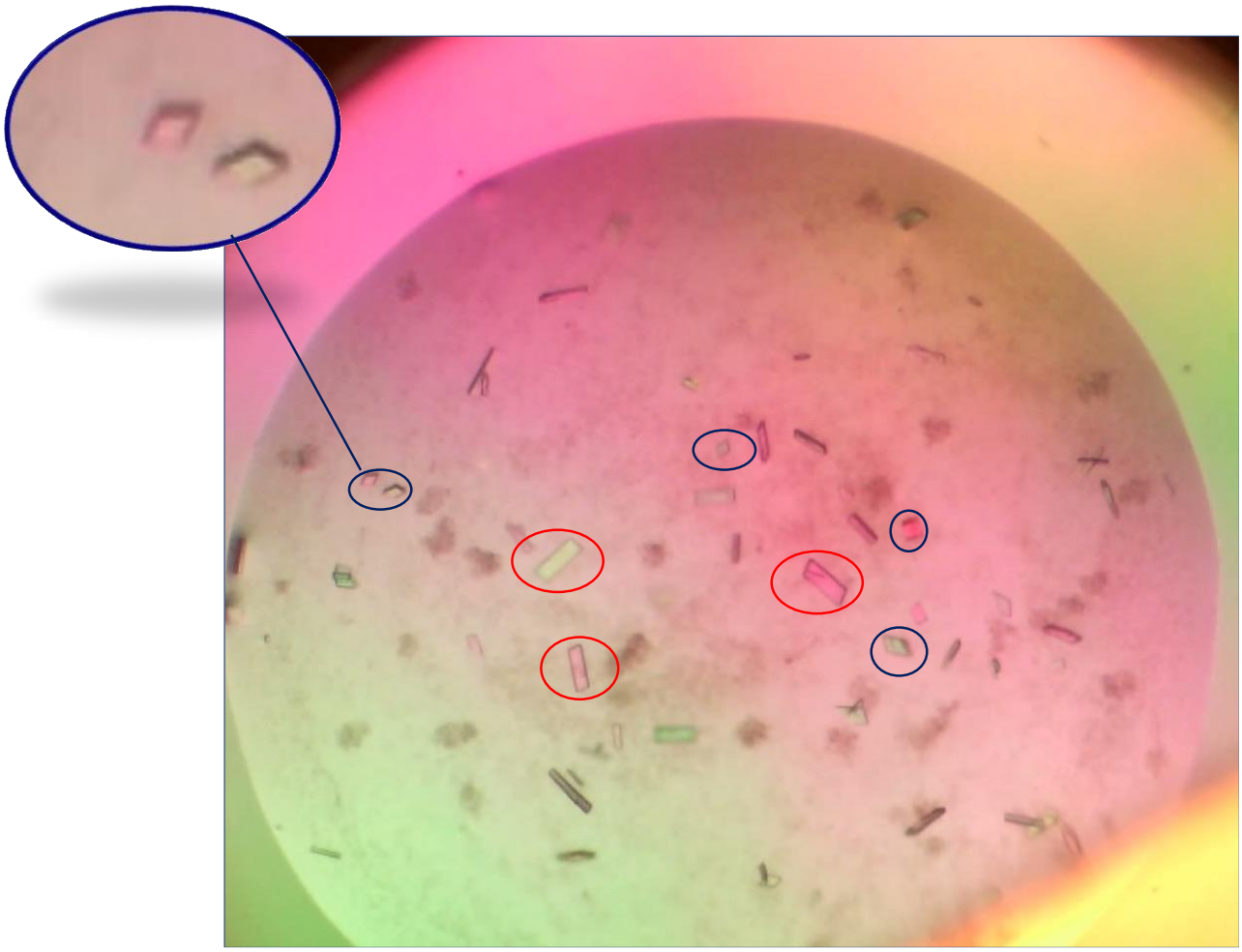


Figure 23: Macromolecular crystallisation of N37T↑V^{+10}•D25A

N37T↑V^{+10}•D25A crystals were grown using the hanging-drop vapour diffusion technique on 96 crystallisation agents from the HR2-144 Hampton® Crystallisation screening kit. The N37T↑V^{+10}•D25A crystals had the characteristic trigonal bi-pyramidal shape, indicated by the navy circles. Multiple crystals of different shapes, also grew, which are shown by the red circles.

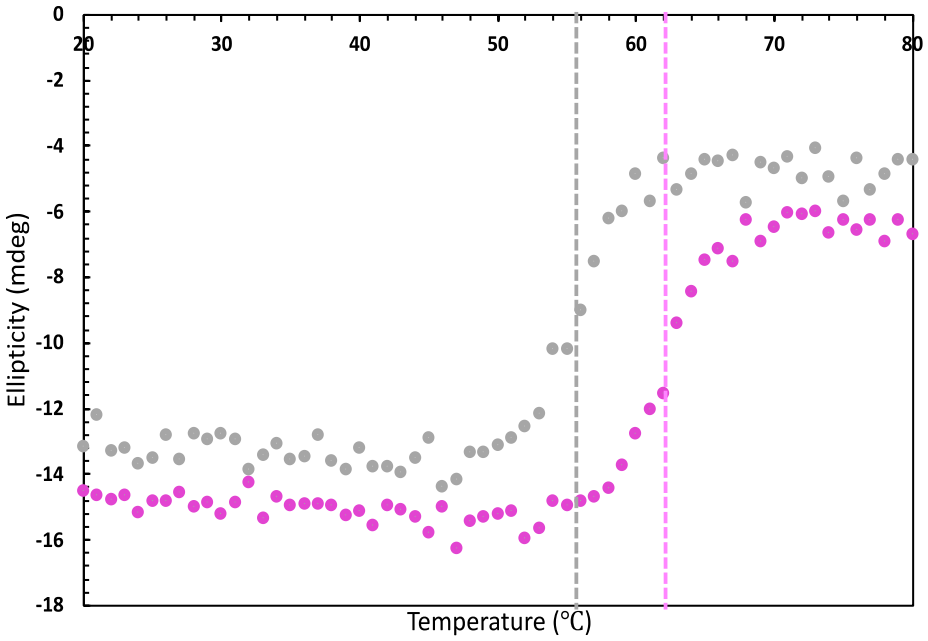


Figure 24: Thermal stability of WT•D25A and N37T↑V^{+10}•D25APR.

The thermal stability of WT•D25A (purple) and N37T↑V^{+10}•D25A (grey) PR were assessed using far-UV CD at a constant wavelength of 216 nm and an increase of temperature from 20 to 80 °C. The melting temperature (T_m) for WT•D25A was 62 ± 0.9 °C and N37T↑V^{+10}•D25A had a T_m of 58 ± 1.2 °C.

The average melting temperatures (T_m) for WT•D25A and N37T↑V^{+10}•D25A during thermal unfolding were 62 ± 0.9°C and 58 ± 1.2 °C, respectively. However, the thermal unfolding of both proteins was irreversible, therefore, it was not possible to determine the thermodynamic parameters of unfolding such as change in enthalpy (ΔH) and change in entropy (ΔS).

3.9.2 Urea-induced equilibrium unfolding

Urea-induced equilibrium unfolding was used to assess the structural unfolding behaviour of HIV-1 protease. The unfolding process of WT•D25A and N37T↑V^{+10}•D25A was monitored using intrinsic tryptophan fluorescence and far-UV CD. These methods have been extensively used to give insights on the secondary and tertiary structures of proteins.

To assess the tertiary structural unfolding, final concentrations of 2 µM of WT•D25A and N37T↑V^{+10}•D25A triplicate (biological triplicates) samples were incubated in urea and allowed to reach equilibrium for a minimum of one hour at 20 °C. WT•D25A and N37T↑V^{+10}•D25A were excited at 295 nm and the fluorescence intensity was monitored at the emission wavelengths of 355 nm and 325 nm. Fluorescence intensity relies on the four tryptophan residues (Trp 6, Trp 6', Trp 42, and Trp 42') present in HIV-1 PR which become increasingly exposed to the solvent as the protein undergoes unfolding. This is used to characterise the unfolding tertiary structure as WT•D25A and N37T↑V^{+10}•D25A PR is denatured. The results (Figure 25) revealed that WT•D25A PR had a single transition event of N→U, between 2 M and 3.5 M urea treatment, with no observable stable intermediate. The Gibbs free energy (ΔG_{water}) obtained for WT•D25A and N37T↑V^{+10}•D25A were 12.815 kcal.mol^{-1} and 11.79 kcal.mol^{-1}, respectively. Furthermore, WT•D25A had an unfolding *m*-value of 2.63 ± 0.43 kcal.mol^{-1}.M^{-1}. Similar to the wild type, equilibrium unfolding for N37T↑V^{+10}•D25A resulted in a single transition event, N→U. Contrary to the wildtype, the transition for N37T↑V^{+10}•D25A occurred between 1.75 and 3 M urea concentrations. The *m*-value for N37T↑V^{+10}•D25A was 2.49 ± 0.23 kcal.mol^{-1}.M^{-1}.

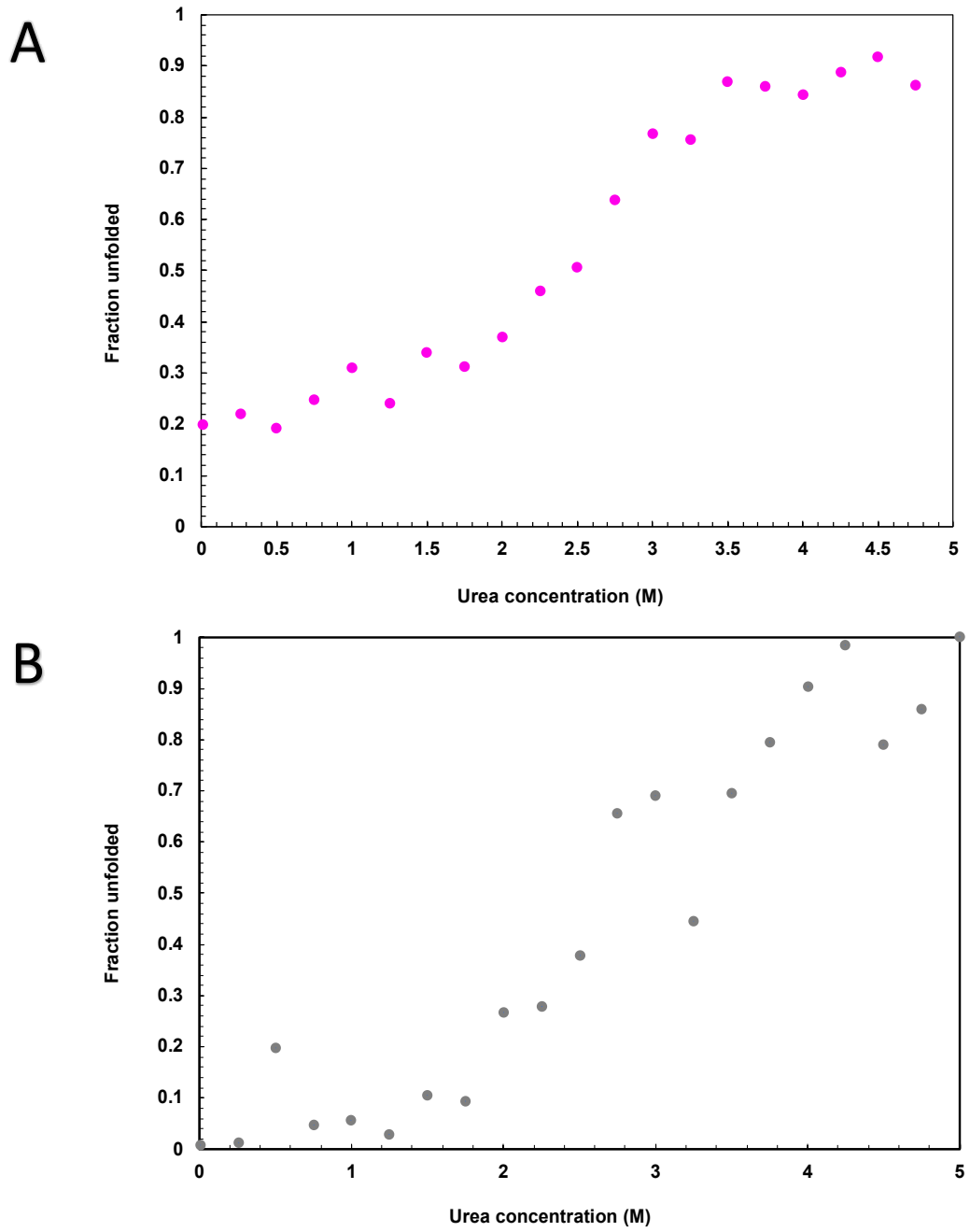


Figure 25: Urea-induced equilibrium unfolding of WT•D25A and N37T↑V^{+10}•D25A monitored by intrinsic tryptophan fluorescence.

The unfolding transition of A) WT•D25A (●) and B) N37T↑V^{+10}•D25A (●) excited at 295 nm and monitored at emission wavelengths of 355 nm and 325 nm. The average of the triplicate unfolding experiments is shown over a range of 0 – 5 M urea concentrations. Transitional from native to unfolded occurred between 2 M and 3.5 M for WT•D25A. A single transitional state was also observed for N37T↑V^{+10}•D25A between 1.75 M and 3 M urea concentrations.

The presence of a single transition state in both PRs is indicative of a cooperative unfolding process, suggesting a two-state unfolding mechanism.

The secondary structure unfolding of N37T↑V^{+10}•D25A PR was assessed using far-UV CD spectroscopy. The N37T↑V^{+10}•D25A PR was treated in urea concentration range of 0 – 8 M and the ellipticity of 10 µM of N37T↑V^{+10}•D25A PR was monitored at a wavelength of 218 nm. Data analysis revealed a sigmoidal curve (Figure 26) and gave insight on the loss of secondary structure as the PR undergoes denaturation and unfolds. A single transition event occurred at the range of 3 M and 4.3 M urea concentration. The single transition state supports the observation in Figure 25B above, of a two-state unfolding mechanism for the denaturation of N37T↑V^{+10}•D25A PR. However, N37T↑V^{+10}•D25A loses its tertiary structure between 1.75 M and 3 M urea concentration. This would suggest that a higher urea concentration is required to induce unfolding of the secondary structure compared to the tertiary structure. This implies that the PR is resilient in maintaining its secondary structures folded conformation.

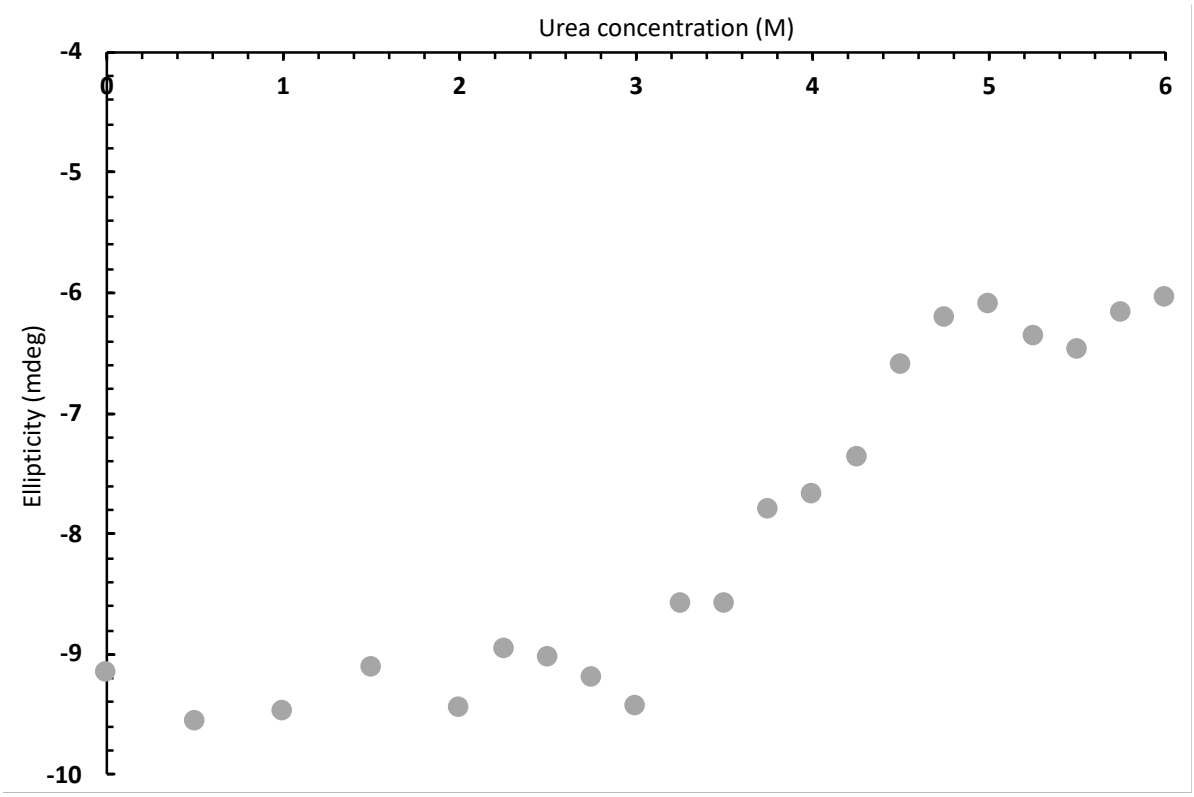


Figure 26: Urea-induced equilibrium unfolding of N37T↑V+10•D25A monitored by far-UV CD

Figure 26: Urea-induced equilibrium unfolding of N37T↑V^{+10}•D25A monitored by far-UV CD

The unfolding transition of N37T↑V^{+10}•D25A (●) monitored at a wavelength of 218 nm. The average of the triplicate unfolding experiments is plotted over a range of 0 – 5 M urea concentrations. A single transitional state occurred between urea concentrations of 3 M and 4.2 M and the C_m was 3.8 M.

Chapter 4: Microsecond molecular dynamics simulations reveal consequential changes in the flap dynamics of novel variants HIV-1 protease

4.1 Introduction

The HIV/AIDs pandemic has continued to impose a socioeconomic burden upon millions of individuals for over three decades and, yet the disease remains incurable**.** However, highly active antiretroviral treatment (HAART) has been a welcomed intervention to improve the quality and longevity of life for people infected with HIV (Collier *et al.*, 1996; Shafer and Schapiro, 2008). HAART incorporates first-line antiretroviral therapy (ART), which involves the combined use of two nucleoside reverse transcriptase inhibitors (NRTI) and a single non-nucleoside reverse transcriptase inhibitor (NNRTI), which target the viral reverse transcriptase (RT) (Detels *et al.*, 1998; Holec *et al.*, 2017; Paton *et al.*, 2017). These NRTI and NNRTI have previously been shown to reduce HIV viral load, improve prognosis of HIV-infected patients, increase human hosts CD4 cells, improve mortality of infected individuals and delay the development of AIDS (Barré-Sinoussi, 1996; Detels *et al.*, 1998; Murphy *et al.*, 2001; The Lancet, 2000; Vlahov *et al.*, 2005). In the event that this first-line treatment fails, second-line ART is employed. Second-line ART involves the use of protease inhibitors (PIs) of which nine have been approved by the FDA to date (Holec *et al.*, 2017; Paton *et al.*, 2017; World Health Organization, 2022).

PIs inhibit the HIV-1 protease (PR) enzymes, which is essential for the maturation of the HIV-1 virion during the viral lifecycle. It is noteworthy that even though the HIV-1 subtype C is most prevalent globally, all currently available PIs have been designed based on the sequence and structure of HIV-1 subtype B. Although PIs have been shown to inhibit non-subtype B HIV-1, there is evidence that suggests the inhibition is less effective and is attributable to naturally occurring polymorphisms (NOPs) between subtypes (Velazquez-Campoy *et al.*, 2002). Moreover, there is a growing body of evidence which suggests that emerging mutations in these non-B subtypes further contribute to drug resistance (De Arellano *et al.*, 2005; Mosebi *et al.*, 2008; Naicker *et al.*, 2013; Naicker and Sayed, 2014; Sherry *et al.*, 2022; Venkatachalam *et al.*, 2023;

Zondagh *et al.*, 2018). A subtype of particular interest, subtype C, is the leading cause of infections worldwide, infecting more than 50% of the global HIV-1 positive population (Bessong, 2008; Naicker *et al.*, 2013; UNAIDS, 2022). This subtype is the prevalent subtype in sub-Saharan Africa and, in South Africa, it accounts for more than 97% of the country's HIV infected population (UNAIDS, 2022).

The three-dimensional structure of the HIV-1 PR provides insight into the relationship between its structure, function and inhibition. The PR is a homodimeric aspartyl protease containing 99 amino acid residues in each monomer (Debouck, 1992). The structure includes the cantilever (residues 62–78), fulcrum (residues 10–22), flaps (residues 46-54) and hinges (residues 35–42 and 56–60) (Figure 27A). These regions function together in a process referred to as "hydrophobic sliding", which involves the synergistic movement of the fulcrum, cantilever, and hinges to facilitate large-scale movement of the flaps (Foulkes-Murzycki *et al.*, 2007; Sherry *et al.*, 2022). The flaps exist in a dynamic equilibrium between closed, semi-open and open conformations, with the latter allowing for substrates or inhibitors into the active site. The closed flap conformation covers the active site, containing a typical Asp25-Thr26-Gly27 catalytic triad (Figure 27A), and promotes the formation of intermolecular interactions which are important in substrate recognition and hydrolysis. PR-mediated cleavage of the Gag and Gag-Pol viral precursor polyproteins results in the significant internal rearrangement and subsequent maturation of the viral virion (Briggs and Kräusslich, 2011). PIs are designed to bind competitively to the active site and prevent the PR from accessing and successfully cleaving the substrates, thereby inhibiting the process of maturation. The aforementioned hydrophobic sliding mechanism is thought to be mediated by exchanging hydrophobic interactions between residues located in the hydrophobic core of the PR. As such, it has been hypothesised that significant alterations to the intramolecular hydrophobic interaction network, such as sequence mutations, are likely to alter the dynamics of the PR flaps.

The HIV-1 genome and, thus, by extension the PR itself are subject to accumulating mutations due to the error-prone nature of the RT enzymes which has a low fidelity and lacks a proof-reading mechanism. This leads to an increase in the genetic diversity of the HIV-1 genome. A phenomenon which is exacerbated by the added selective pressure applied by antiretroviral drug therapy.

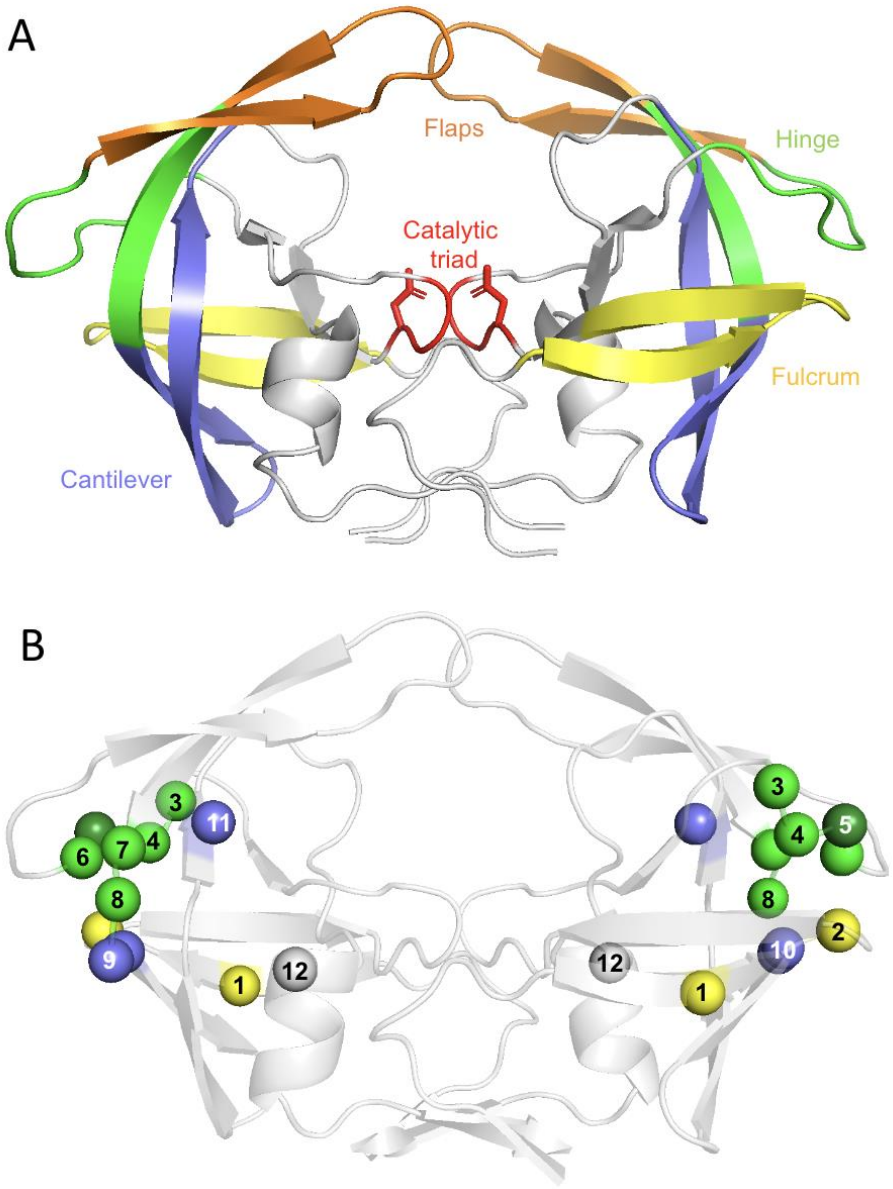


Figure 27: The HIV-1 protease enzyme

A) Ribbon representation of the HIV protease indicating the major structural elements: cantilever (blue; residues 62–78), fulcrum (yellow; residues 10–22), flaps (orange; residues 46-54) and hinges (green; residues 35–42 and 56–60). **B**) Ribbon representation of the N37T↑V^{+10} protease variant. The mutations present in N37T↑V^{+10} protease are as follows: 1. I13V; 2. G16E; 3. I36T; 4. N37T; 5. ↑V (valine insertion); 6. P39S; 7. D60E; 8. Q61E; 9. I62V; 10. L63P; 11. V77I; 12. M89L.

As such, new viral variants with resistance-conferring mutations are emerging regularly with mutations present in various viral proteins including the PR. One such variant of interest is a unique HIV-1 subtype C variant, N37T↑V^{+10} PR (Figure 27B), which contains an N37T polymorphism and an insertion of valine (↑V) at position 37 in the hinge region. The N37T↑V^{+10} PR contains 100 amino acids per monomer as opposed to the standard 99, as well as an additional 10 background polymorphisms (Table 2) which are distal to the active site (I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L). Previously, it has been shown that hinge region insertions may increase the dynamic motion of the hinges and, therefore, the flaps (Perryman *et al.*, 2004; Zondagh *et al.*, 2018). Moreover, it has been shown that insertions in the hinge region of the PR can lead to reduced drug susceptibility (Maseko *et al.*, 2019; Williams *et al.*, 2019; Zondagh *et al.*, 2018). However, previous computational studies of the dynamics of the PR flap regions have been conducted in a short simulation timescales between 20 and 700 nanoseconds (Eche *et al.*, 2021; Sherry *et al.*, 2021; Zondagh *et al.*, 2018). With the advancement of computational power, it is now necessary to investigate the dynamics of the PR flaps during longer timescales which may offer new and more detailed information about the drug binding or evading mechanisms of HIV-1 PR variants. In particular, we investigated two of the most commonly used PIs in South Africa, a country where 95% of the HIV positive population are infected with HIV-1 subtype C, namely, atazanavir (ATV) and darunavir (DRV).

4.2 Methods

4.2.1 Protein and ligand preparation

The sequence data for the N37T↑V^{+10} PR were obtained from Prof Lynn Morris (AIDS Virus Research Unit, NICD, Johannesburg, South Africa). The coordinates for the crystal structure of the South African HIV-1 subtype C PR wild type (PDB ID:3U71) which had previously been resolved (Naicker *et al.*, 2013) at resolution of 2.72 Å were extracted from RCSB PDB. The crystal structure of the variant, N37T↑V^{+10} PR is yet to be solved; therefore, a homology model was constructed. Since the HIV-1 subtype C wild type (WT) PR (PDB ID: 3U71) shares a high sequence similarity with N37T↑V^{+10} it was used as a template structure. The homology model for N37T↑V^{+10} was

constructed using the SWISS-model online tool (Arnold *et al.,* 2006) and the homology model was then validated using PROCHECK (Laskowski *et al.*, 1993). The Schrödinger models: Glide, Prime, Qsite, Liaison and Macromodel were subsequently used to prepare the homology models. The models underwent bond model correction, this involved assigning ionisable residues a charge which corresponds to an experimental pH of 5 (experimentally determined). The WT PR and N37T↑V^{+10} PR structures were subjected to energy minimisation using OPLS_2005 until the root mean square deviation was approximately 0.3 Å. The 2D structures of two protease inhibitors, atazanavir (ATV, PubChem CID: 14192) and darunavir (DRV, PubChem CID: 213039) were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/) in SDF format. The ligand preparation tool, LigPrep tool (Schrödinger LLC 2009, USA), was used to convert the PIs 2D structures to 3D structures and prepare the protease inhibitors for molecular dynamic simulations. Ligand preparation involved adding hydrogens, bond angles, bond lengths and choosing the appropriate chirality and subsequently energy minimisation was performed. Following this, the Epik-tool was used to select the lowest energy tautomers and ring structures.

4.2.2 Molecular docking

Induced fit docking was conducted on the energy minimised WT PR and N37T↑V^{+10} PR models with the two ligands, ATV and DRV. The induced fit docking (IFD) studies were performed using Glide software (Schrodinger LLC 2009, USA). This involved defining a region on the PR where induced fit docking of the PIs would occur by constructing a 20 Å cubic box around the protease active site residues, D25. For protein residues which fell within 5 Å of the ligand poses, refinement and sidechain optimisation was performed. OPLS3 force field was used to perform the IFD simulations and resulted in 20 conformational poses for each PR-PI complex. The top scoring conformations i.e., conformations with the highest IFD scores for each complex were selected for downstream MD simulations.

4.2.3 Molecular dynamic simulations

Molecular dynamic (MD) simulations were carried out using Maestro v12.2 and a GPU-enabled Desmond molecular dynamics simulations engine. The apo proteases and

protease:inhibitor (PR:PI) complexes; WT:ATV, WT:DRV, N37T↑V^{+10}:ATV and N37T↑V^{+10}:DRV were submitted to Linux (Ubuntu) architecture. A solvated cubic box universe was generated for each of the 6 systems (1. WT apo, 2. N37T↑V^{+10} apo, 3. WT:ATV, 4. WT:DRV, 5. N37T↑V^{+10}:ATV and 6. N37T↑V^{+10}:DRV) using the TIP3P solvation model within the System Builder model. The protein was placed at the centre of the cubic box, with the protein 10 Å from the cubic box edges. The apo, and DRV-bound systems were neutralised by the addition of 11 (charge −11) chloride ions while the ATV-bound systems required 12 chloride ions for neutralisation (charge − 12). Energy minimisation was conducted on the models until the system had an atomic force which was less than 100 kJ.mol^{-1}.nm^{-1} and an average root mean square deviation (RMSD) of approximately 0.3 Å using the NPT (constant number of particles (N), constant volume (V) and constant temperature (T)) ensemble (300 K, 1 atm). Following this, a 5 ns molecular dynamic simulation under NVT (constant number of particles (N), constant volume (V) and constant temperature (T)) conditions was conducted and the system pressure equilibrated. The temperature was linearly increased from 10 K to 300 K. To ensure a stable temperature while relaxing the models, the positional restraints were calculated and the original values annealed.

4.3 Results and Discussion

4.3.1 Comparative conformational stability of WT PR and N37T↑V^{+10} PR

The root mean square deviation (RMSD), is a measurement of how much the global structure of the protein deviates from a starting reference structure during the course of a molecular dynamics (MD) simulation. The RMSD can indicate the conformational stability of the structure of a protein, with a lower RMSD being indicative of a more rigid, stable structure and vice versa. Certain mutations have been shown to alter the global stability of HIV-1 PR. Moreover, in order to investigate how the mutations, present in N37T↑V^{+10} PR impact the global stability of the HIV-1 PR, we quantified the RMSD in the WT PR and N37T↑V^{+10} PR both in the apo state as well as when in complex with DRV and ATV.

4.3.2 RMSD of apo PRs during simulations

During the first 500 ns of the simulation, the RMSD of the WT PR varied considerably exhibiting two large scale conformational changes between 233 ns and 284 ns and another between 409 ns and 436 ns with an average RMSD of 2.34 Å (Figure 28). In contrast, during the first 500 ns of the simulation the RMSD of the N37T↑V^{+10} PR sharply increased during the first 180 ns before reaching a stable RMSD with an average of 3.78 Å. These results may suggest that the N37T↑V^{+10} PR is more structurally rigid than the WT PR. This is further reflected in the latter half of the simulation, where both PR variants appear to reach a stable RMSD. The WT PR exhibits a stable RMSD of 3.39 Å for half of the simulation while, in contrast, the N37T↑V^{+10} PR stabilises at an RMSD of approximately 3.7 Å for >80% of the simulation. It could be postulated that the N37T↑V^{+10} PR inhabits a reduced range of 3D conformations in the apo state in comparison to the WT PR. This may be as a result of the altered dynamics imparted by the mutations present in the sequence of this variant.

4.3.3 RMSD of drug-bound PR complexes during simulations

The WT:ATV complex exhibited an average RMSD of 2.27 Å in the first 500 ns of the simulation and 2.23 Å during the second 500 ns, which implies that the PR maintained a relatively stable RMSD throughout the simulation. This could suggest that ATV successfully binds to and stabilises the WT PR as expected and indicates that the WT PR was likely in the closed conformation for the majority of the simulation. However, two large conformational changes were observed in the WT:ATV complex between 219-284 ns and 293-328 ns where the RMSD spiked to 3.33 Å and 3.20 Å, respectively. Following this, the RMSD stabilised again, at approximately 2.4 Å. The N37T↑V^{+10} PR exhibited a relatively low average RMSD of 2.33 Å during the first 500 ns of the simulation which is similar to the WT PR. However, during the second half of the simulation the N37T↑V^{+10}:ATV exhibited an elevated average RMSD of 2.69 Å, an increase of ~18%. Moreover, between 560-658 ns the average RMSD of the N37T↑V^{+10}:ATV complex increased to 3.53 Å which is similar to the average RMSD of an unbound N37T↑V^{+10} PR (3.59 Å). This characteristic is observed again between

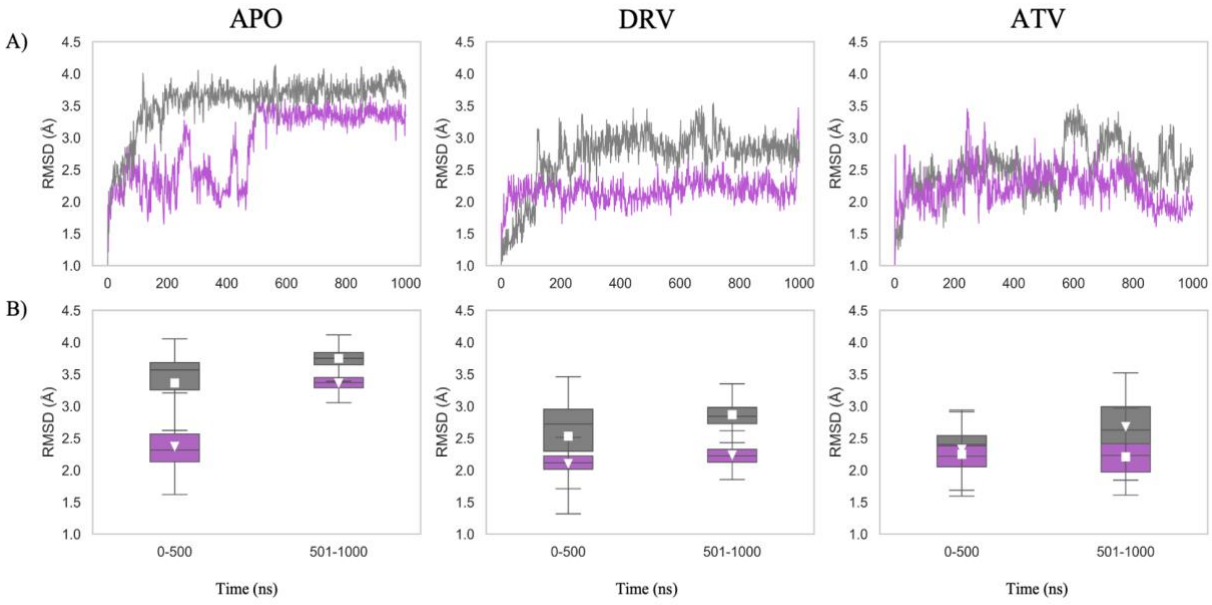


Figure 28: Root mean square deviation (RMSD) of WT PR and N37T↑V^{+10} PR

A) Line plots of the RMSD per system (Apo/DRV/ATV) over the course of the microsecond simulation and, **B**) Boxplots of the RMSD values per system, with mean values plotted (Mean RMSD: N37T↑V^{+10} PR - white square; WT PR - white triangle). WT PR data are coloured purple and N37T↑V^{+10} PR are coloured grey.

675-779 ns, where the complex samples an average RMSD 3.4 Å. This could mean that during this period the N37T↑V^{+10}:ATV complex behaves more like an unbound PR and that the flaps of the N37T↑V^{+10} PR may sample semi-open or open conformations when in complex with ATV. Together these data suggest that ATV may be unable to bring N37T↑V^{+10} into a stable closed conformation as readily as the WT PR complex.

When in complex with DRV the WT PR exhibited a peak RMSD of 2.9 Å with an average of approximately 2.7 Å, which is maintained with relative stability for the entirety of the simulation. Such characteristics indicate that it is likely that the WT PR is successfully brought into a stable, closed conformation by DRV. In contrast, when N37T↑V^{+10} PR is in complex with DRV, it exhibited an elevated average RMSD of 2.9 Å, with peaks as high as 3.55 Å, which is once again similar to the average RMSD of an unbound N37T↑V^{+10} PR (3.59 Å). These data suggest that the N37T↑V^{+10}:DRV complex is significantly less stable than that WT:DRV complex. This may be indicative of the inability of DRV to bring the N37T↑V^{+10} PR into the closed conformation. It is interesting to note that the drug bound N37T↑V^{+10} PR complexes appear to be less rigid in comparison to the WT PR complexes. These data suggest that the N37T↑V^{+10} PR may have a higher propensity for semi-open and open conformations when in complex with inhibitors. The mutations present in this variant may alter or even eliminate intermolecular interactions that facilitate inhibition by increasing the dynamics of the flap regions. Which is noteworthy, as an increase in the dynamics of the PR, especially the flaps, has been hypothesised to contribute to reduced drug susceptibility as a result of an increased off-rate (k_{off}) of the inhibitors binding to mutated PR (Sherry *et al.*, 2022).

4.3.4 Atomic fluctuations of the WT PR and N37T↑V^{+10} PR

The root mean square fluctuation (RMSF) is a measure of the average deviation of a protein from a reference structure over time and can provide insight into changes in the dynamics of specific regions of the PR. Peaks in a RMSF plot (Figure S1) are indicative of residues that exhibit the greatest atomic fluctuation over the course of the simulation. Figure 29 represents the backbone RMSF of both WT and N37T↑V^{+10} when uncomplexed and complexed to two protease inhibitors, ATV and DRV.

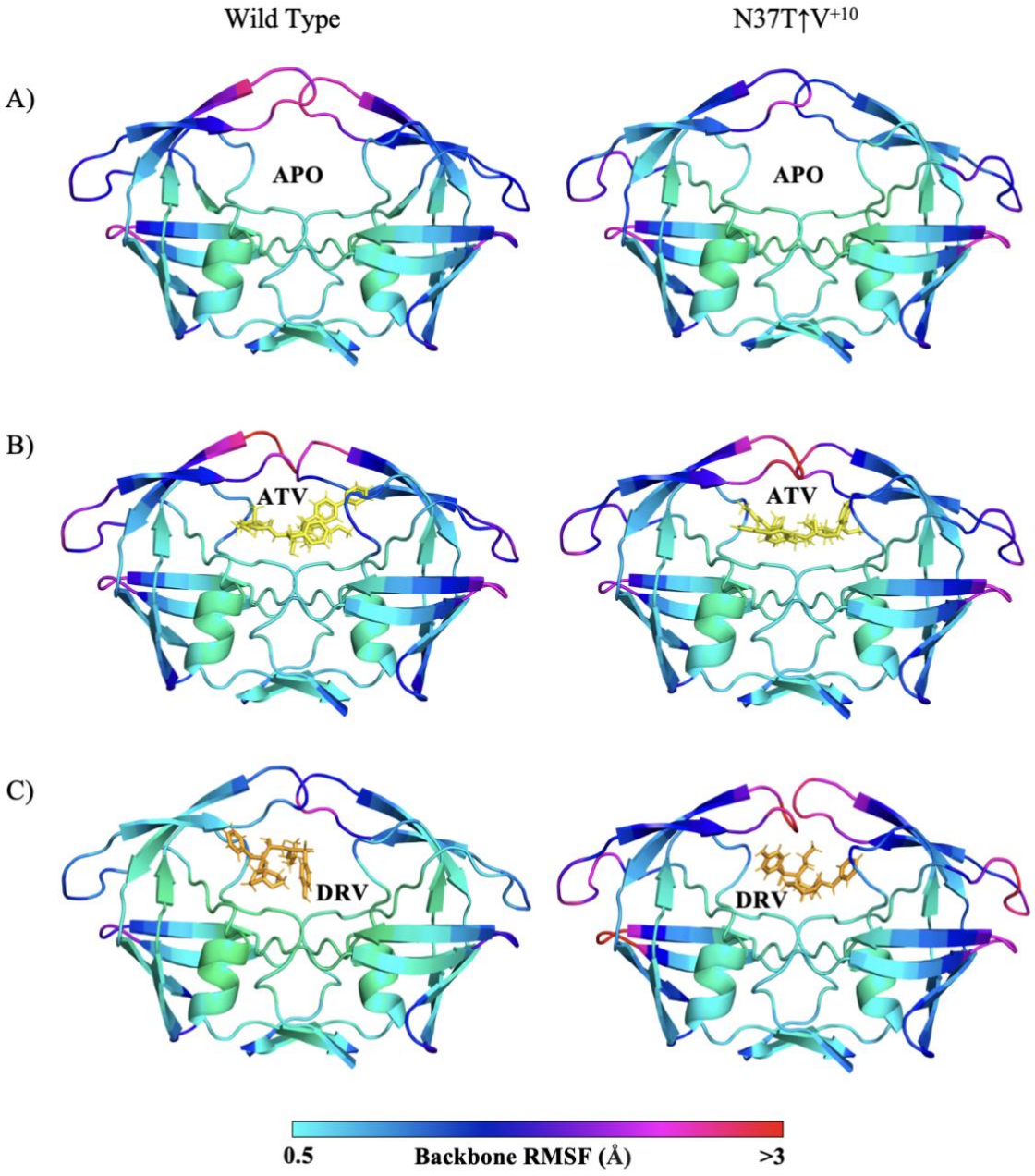


Figure 29: Backbone RMSF of WT PR (left) and N37T↑V^{+10} PR (right)

A) The apo (uncomplexed) state, **B**) In complex with ATV (yellow, stick representation), C) In complex with DRV (orange, stick representation). The ribbon representing the backbone of the PR has been colour coded according to the RMSF per residue with the colour scheme green-cyan-blue-magenta-red indicating the RMSF range from 0.5 Å to >3 Å.

4.3.5 The apo PR state

The backbone RMSF representations depicted in Figure 29A illustrate the differences in regional RMSF between the apo WT PR and apo N37T↑V^{+10} PR. It is evident that the RMSF of the flap regions (residues 46-54) of the N37T↑V^{+10} PR are significantly reduced in comparison to the WT PR represented by the colour shift from red, indicating high RMSF, to blue, indicating lower RMSF (Figure 29). More specifically, a decrease of between 10% and 46% was observed for the flap regions of the N37T↑V^{+10} PR.

These data illustrate that the flaps of the apo N37T↑V^{+10} PR are more stable than those of the WT PR (Figure S1). This is of particular importance when considering the nature of drug binding which requires the PR flaps to adopt the open conformation to allow inhibitor entry into the active site followed by the inhibitor bringing the flaps into a stable closed conformation. We previously hypothesised that increases in the rigidity of the flaps when in the apo state may contribute to diminished inhibitor binding affinity through a decreased on-rate of PIs and an increased energetic penalty associated with flap conformation changes (Sherry *et al.*, 2022). The hinge regions (residue 35-41 & residues 57-60) of the apo N37T↑V^{+10} PR exhibited a 16%-34% increase in RMSF with respect to the WT PR (Figure 29 A). This is noteworthy as the hinge region is where most of the mutations are found in this variant including the novel hinge region insertion (↑V), which suggests that these mutations alter the intramolecular interaction network in this region in a manner that alters the hinge dynamics. It has previously been noted that an increase in the hinges flexibility directly alters the flap dynamics (Zondagh *et al.*, 2018). Furthermore, it is postulated that increased flap dynamics plays a critical role in reduced drug susceptibility although the mechanism is not fully understood (Sherry *et al.*, 2022; Williams *et al.*, 2019; Zondagh *et al.*, 2018). In a parallel study we investigated the effect of removing the 10 background mutations present in N37T↑V^{+10} PR, generating N37T↑V PR (manuscript in preparation). Interestingly, the results of this study indicate that the presence of the background mutations serve to stabilise the core regions of the N37T↑V PR to ensure that with increased dynamics of the hinge regions, the mutant is likely to maintain its enzymatic function. We conclude that the cantilever region in particular is required to function as

a scaffold which compensates for increased hinge and flap dynamics in order to retain the enzymatic function of the PR. It is evident that the cantilever region is critical for the maintenance of the structural integrity of the PR. Interestingly the cantilever houses four NOPs, which may be present to reinforce the stability and structure of this region.

4.3.6 Drug-bound PR complexes

The backbone RMSF of the N37T↑V^{+10} PR and WT PR do not display significant differences when in complex with ATV (Figure 29B, Figure S1). The N37T↑V^{+10} PR does, however, exhibit slightly higher flexibility in the flap regions when bound by ATV. These results mirror the RMSD data (Figure 28) which illustrate that the N37T↑V^{+10} PR can sample semi-open or open conformations more often than the WT PR when in complex with ATV. Interestingly, the backbone RMSF of the N37T↑V^{+10} PR is substantially different from the RMSF of the WT PR when in complex with DRV. Firstly, the RMSF of the flap, hinge, cantilever and fulcrum regions of the WT PR are drastically decreased in comparison to the WT PR apo state (Figure 29A,C). Such results are expected given that the mechanism of action of PIs is considered to "lock-down" the PR into a stable closed conformation, which would reduce the movement of key regions of the PR. NMR investigations illustrate that the dynamics of the of the PR decrease from the nanosecond scale to microsecond and even a millisecond timescale regime upon inhibitor binding (Ishima *et al.*, 1999). It is noteworthy, therefore, that the same key regions of the N37T↑V^{+10} PR revert to a significantly more dynamic state (Figure 29C) when in complex with DRV when compared to the WT PR complex. Again, this trend is reflected in the RMSD data (Figure 28) and together these data suggest that DRV is unable to bring the N37T↑V^{+10} PR into a stable, fully closed state. Increased dynamics of the flaps of the PR are likely to contribute to reduced drug susceptibility through minimising the number of strong intermolecular interactions formed in the complex or introducing water-bridges into the active site (Sherry *et al.*, 2021). In fact, we observed a shift towards water-bridge formation and an increased transience of intermolecular interactions formed in the ATV-bound and DRV-bound N37T↑V^{+10} PR complexes, respectively (Figure 30). When in complex with ATV, the altered dynamics of the N37T↑V^{+10} PR necessitates the formation of water-bridges between the active site Asp25 residues and the polar atoms of the inhibitor. In

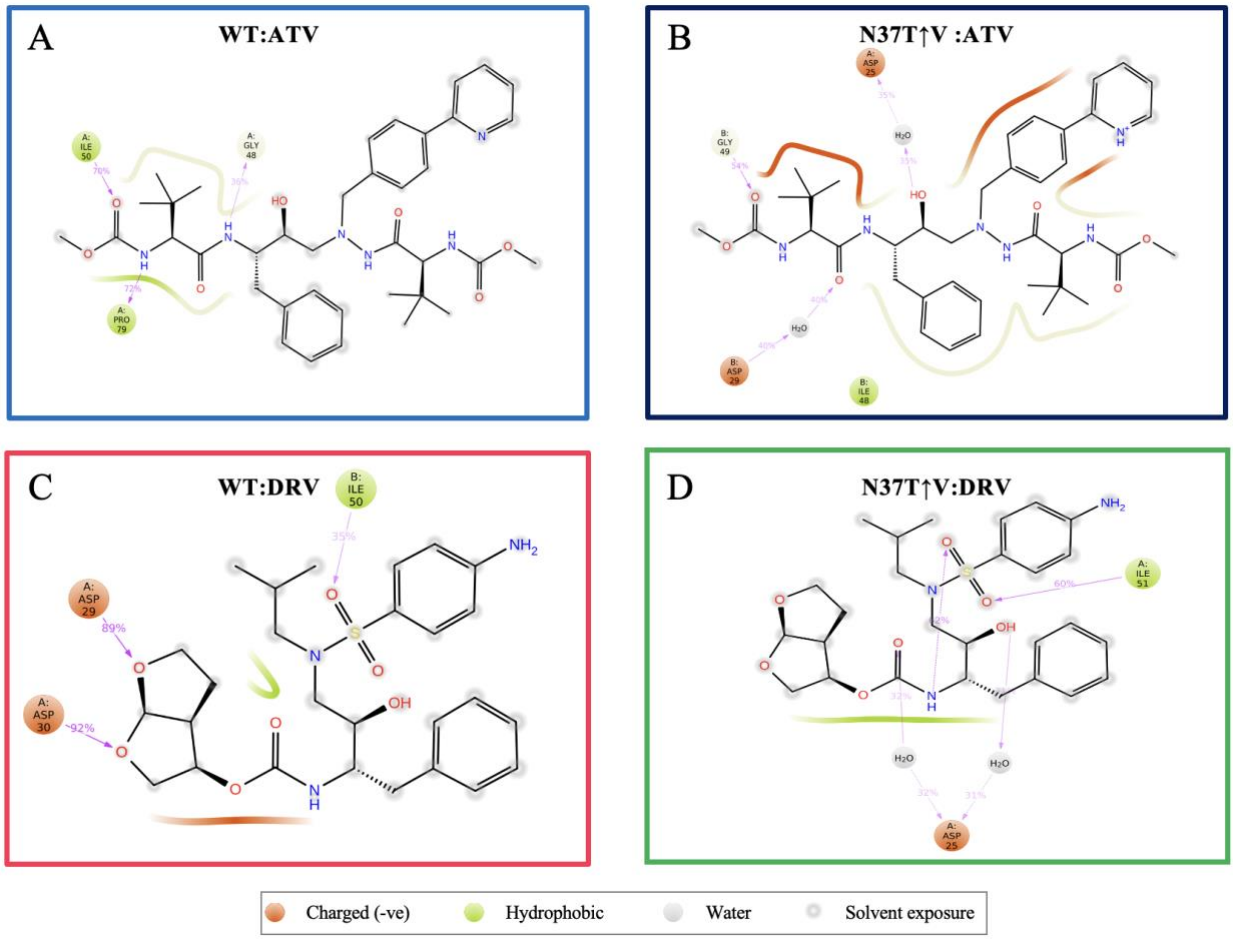


Figure 30: 2-D schematic diagram of interatomic interactions between protease inhibitors (ATV and DRV) and the WT PR and the N37T↑V^{+10} PR.

Interactions that occur for more than 30% of the 1 microsecond simulation (0-1000 ns) are shown.

fact, this characteristic is observable in several other key binding-associated residues including Gly-48 and Gly-49 of the flap tips, and Pro-79 (Figure S2). In the DRV-bound complex, the water bridges formed between the active site of the PR and the inhibitor become more transient in the N37T↑V^{+10} PR complex, decreasing by between 57% and 67% over the course of the simulation (Figure 30). Moreover, the N37T↑V^{+10} PR exhibits a complete loss of hydrogen bond formation with Asp-29 and Asp-30 which are located towards the bottom of the active site cavity (Figure S2).

4.3.7 Altered flap dynamics

A common mechanism of PIs binding to the PR involves intermolecular interactions formed with the active site residues (25, 27, 29, and 30) with simultaneous interactions formed with residues located in the flaps (Lv *et al* 2015). As such, any non-active site mutations that alter the ability of PIs to bring the flaps into a closed conformation would therefore contribute to inhibitor resistance by minimising contacts formed between one or both sets of the aforementioned residues. Similarly, mutations that make the flaps less flexible may introduce a higher energetic penalty associated with conformational changes in the flaps that are necessary for drug binding, thereby contributing to drug resistance. The PR flaps exist in dynamic equilibrium between the closed, semi-open and open conformations which have previously been defined by interatomic distances between Asp25 and Ile50. Where closed, semi-open and open conformations are defined as distances of <17 Å, 17–22 Å and >22 Å, respectively (Perryman *et al* 2004, Sherry *et al* 2021). Previous MD and crystallographic analyses indicate that the apo form of the PR favours the semi-open flap conformation, whereas inhibitor binding induces a closed flap conformation (Huang *et al.*, 2014). Contrastingly, when considering the apo state simulations of the WT PR and N37T↑V^{+10} PR, it is noteworthy that both PRs adopt the closed conformation for the majority of the simulation (Figure 31) which is reflected in the average radius of gyration (R_g) of the PRs (Figure 32). However, several studies have illustrated that the apo PR can adopt the closed conformation when in solution and have suggested that this may represent the ground state of the apo enzyme (Liu *et al.*, 2008). It is noteworthy, therefore, that the N37T↑V^{+10} PR appears to be more stable in the apo state than the WT PR, as evidenced by the N37T↑V^{+10} PR rapidly adopting the closed conformation and

sustaining this conformation for most of the simulations. While the flaps of the WT PR can be seen to oscillate between the closed and semi-open conformation during the first half of the simulation, before settling down into a stable closed conformation. These results mirror RMSF data (Figure 29A) which suggest that the N37T↑V^{+10} PR is more stable in the apo state in comparison to the WT PR.

When bound to ATV, the WT PR exhibits a fully closed conformation in monomer B for the majority of the simulation while monomer A adopts a semi-open conformation before adopting a closed conformation during the final 200 ns (Figure 31). The observed difference in flap conformations between monomers is likely due to the asymmetrical nature of PI binding. The N37T↑V^{+10} PR appears to be more dynamic when in complex with ATV, as evidenced by monomer B adopting a semi-open conformation for the entire simulation while monomer A initially adopts a closed conformations for the first 450 ns before adopting the semi-open conformation as well (Figure 32). These data suggest that ATV is unable to bring the flaps of the N37T↑V^{+10} PR into a stable, closed conformation as readily as the WT PR. Rather paradoxically, the R_g of the WT PR in complex with ATV is greater than that of the N37T↑V^{+10} PR complex (Figure 32). However, it is interesting that the R_g of the WT PR is significantly lowered when in complex with DRV when compared to the WT PR in complex with ATV while the R_g of the N37T↑V^{+10} PR remains essentially unchanged. These data may suggest that the N37T↑V^{+10} PR may be more well adapted towards maintaining its inherent conformational dynamics when bound to PIs. Indeed, the flaps of the N37T↑V^{+10} PR do exhibit large conformational changes when in complex with DRV (Figure 31) and the maintained conformational dynamics of the mutant PR flaps is exemplified in the increased RMSF with respect to the WT PR complex (Figure 29C). Since the flaps of the N37T↑V^{+10} PR appear to be more dynamic when in complex with the PIs, particularly DRV, it follows that the flaps likely sample wider conformations than the corresponding WT PR complexes. Wider flap conformations would place the inhibitor higher up in the active site cavity, beyond the bonding distance of several key active site residues such as Asp25, Asp29 and Asp30. In fact, this phenomenon has been observed previously, where the wider flap conformations of mutated PR variants necessitate water-bridge formation in order to simultaneously maintain contacts with these and residues of the flap (Sherry *et al.*, 2021)

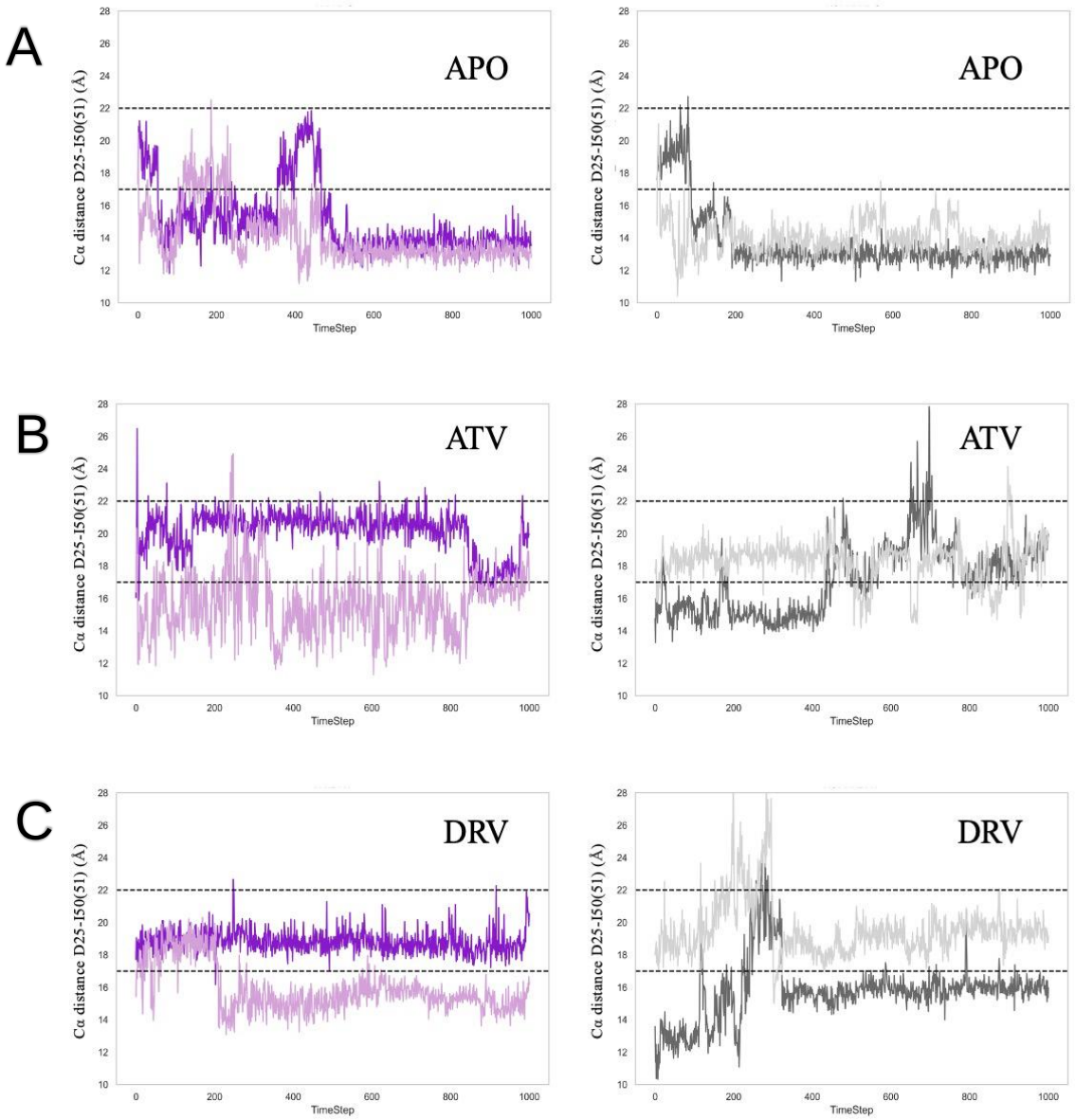


Figure 31: Per monomer alpha carbon interatomic distance between the D25 and I50(51) of the WT PR (purple/pink) and N37T↑V^{+10} PR (dark/light grey)

A) Apo, **B**) bound to atazanavir (ATV) and, **C**) bound to darunavir (DRV). Each graph has two horizontal lines, one at 17 Å and the second at 22 Å; which correspond to closed flap conformations being defined as distances of <17 Å, semi-open conformations as 17–22 Å and the open conformation being anything >22 Å.

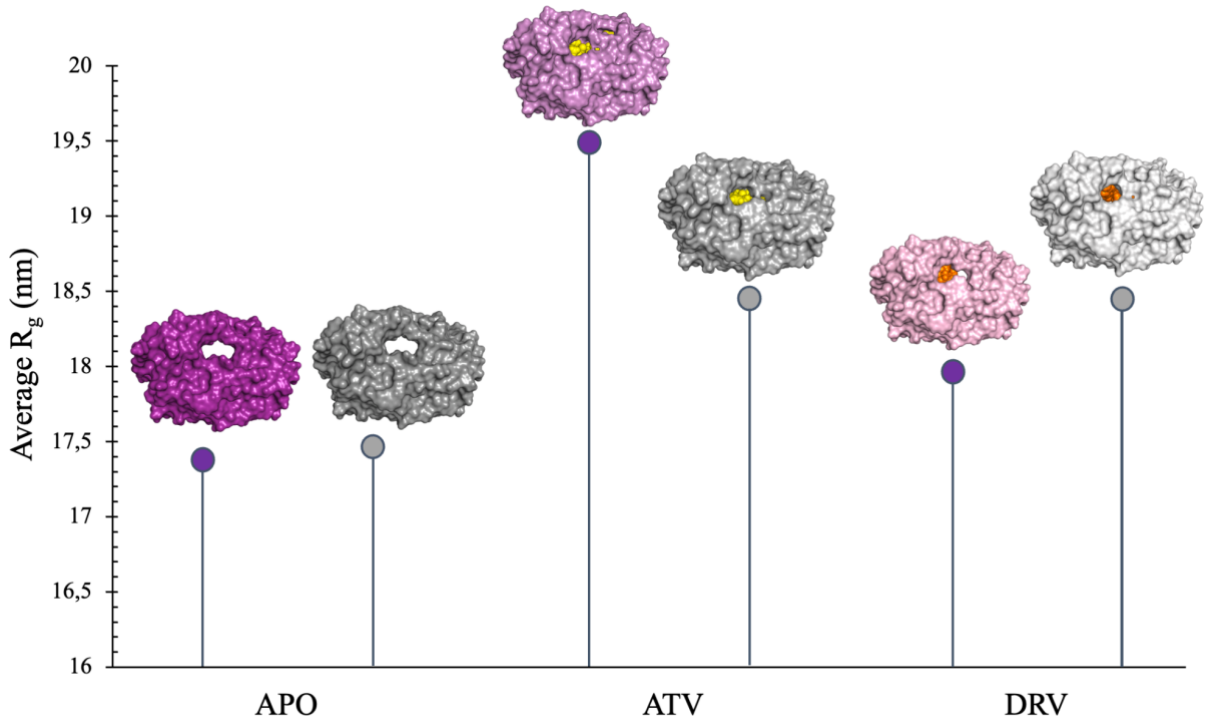


Figure 32: Average radius of gyration (Rg) of the WT PR (purple/pink) and N37T↑V^{+10} PR (dark/light grey) in the apo and drug-bound state.

The radius of gyration of the PR systems was measured of the course of the entire 1000 ns simulation carried out in explicit solvent.

4.4 Conclusion

The HIV-1 PR is an important target for antiretroviral therapy due to its key role in viral maturation. The accumulation of mutations, both within and distant from the active site, in response to drug regimens imposes selective pressure, which leads to disruptions in the protein-drug interface. While the exact impact of non-active site mutations on reduced drug susceptibility remains partially understood, the current study on a variant, N37T↑V^{+10} PR, sheds light on the molecular dynamics and stability implications of non-active site mutations. The mutations present in N37T↑V^{+10} PR were found to increase flap rigidity in the apo state, raising the energetic cost of flap opening when compared to the WT PR. Conversely, in the presence of the two protease inhibitors, ATV and DRV, N37T↑V^{+10} PR exhibits greater flap mobility, reducing the longevity of intermolecular interactions and, in turn, diminishing the binding affinity toward the inhibitors. This dual mechanism suggests that mutations in N37T↑V^{+10} PR contribute to reduced drug susceptibility by altering the PRs dynamics, both in the absence and presence of inhibitors, providing valuable insights for future therapeutic strategies.

Chapter 5: Contrasting the effect of hinge region insertions and non-active site polymorphisms on HIV protease-inhibitor interactions: insights from altered flap dynamics

5.1 Introduction

Human Immunodeficiency Virus (HIV) protease (PR) plays a crucial role in the viral replication cycle and is a major target for antiretroviral therapy (Debouck, 1992). While the use of antiretroviral drugs has led to significant improvements in the treatment and management of HIV infection, the development of drug resistance remains a significant challenge (Bossard *et al.*, 2021; Murphy *et al.*, 2001). HIV protease inhibitors (PIs) have been used extensively in the treatment of HIV infection; however, the emergence of drug resistance mutations presents a major limitation to their long term efficacy (Detels *et al.*, 1998; Shafer and Schapiro, 2008; Vlahov *et al.*, 2005).

Resistance begins with genetic changes that arise in the viral genome following exposure to antiretroviral drugs (Carter *et al.*, 2014; Weber and Agniswamy, 2009). Resistance starts with primary mutations which accumulate in the active site cavity of the PR (Lin *et al.*, 1995; Shafer and Schapiro, 2008). Primary mutations directly impact the protein:drug interface in a manner that can significantly diminish binding affinity. However, primary resistance mutations can also deleteriously affect the ability of the PR to bind to and hydrolyse its natural Gag and Gag-Pol protein substrates (Condra *et al.*, 1995; Mosebi *et al.*, 2008; Williams *et al.*, 2019). As such, secondary mutations tend to arise in regions distant to the active site cavity where they lead to changes in the viral phenotype, including restoring normal or enhanced catalytic activity and resistance to PIs (Ohtaka *et al.*, 2002; Velazquez-Campoy *et al.*, 2001). Secondary mutations in the PR are known to play a critical role in the development of drug resistance, and there is a growing body of research investigating the mechanisms underlying these mutations (Bessong, 2008; Liu *et al.*, 2008; Sherry *et al.*, 2022; Venkatachalam *et al.*, 2023). Secondary mutations can impact the conformational stability and activity of the PR (Eche *et al.*, 2021; Mosebi *et al.*, 2008) and can lead to changes in the dynamics of the enzyme in a manner that can compensate for the negative effects of primary mutations on the protease activity and viral fitness (Sherry *et al.*, 2022; Velazquez-Campoy *et al.*, 2001).

The important role that non-active site mutations play in resistance is clear when considering that cross-resistance to multiple PIs is associated with a plethora of non-active site mutations that augment the mechanism of one or two active site mutations (Clemente *et al.*, 2004). Additionally, the effect of active site mutations on drug resistance is radically reduced when all accompanying non-active site mutations are reverted to their corresponding wild type of residue. Indeed, non-active site mutations are thought to accumulate to re-stabilise enzyme activity, a phenomenon that has been observed in oseltamivir resistance in influenza (Bloom *et al.*, 2010).

Although it is not fully understood how these non-active site mutations impart their effect, some explanation may lie in their effect on the conformational flexibility of mutant PRs (for an in-depth review see Sherry *et al.*, 2022). More specifically, non-active site mutations are likely to affect the molecular dynamics of the flap regions of the PR through modulating the stability of other distinct regions of the PR such as the cantilever, fulcrum and hinge regions. Changes in the dynamics of these regions have been shown to work in a cooperative manner to influence the protein:drug interface (Foulkes-Murzycki *et al.*, 2007; Goldfarb *et al.*, 2015; Sherry *et al.*, 2021). Moreover, mutations and insertions located in the hinge region of the PR have been shown to affect drug susceptibility (Liu *et al.*, 2016; Ragland *et al.*, 2014; Williams *et al.*, 2019; Zondagh *et al.*, 2018). Therefore, further understanding how insertions and mutations influence the conformational stability and function of HIV PR is critical to understanding how these non-active site mutations bring about drug resistance.

In this context, the current study aimed to assess how secondary background mutations outside the active site affect the conformational stability and drug binding of a novel HIV-1 protease variant using molecular dynamic simulations and molecular docking. This novel HIV subtype C PR variant contained an amino acid polymorphism at position 37 as well as a hinge-region insertion (↑V) (Figure 33). Furthermore, this PR variant contained 10 additional background amino acid polymorphisms. In order to dissect the independent effects of these mutations on the dynamics of the PR these

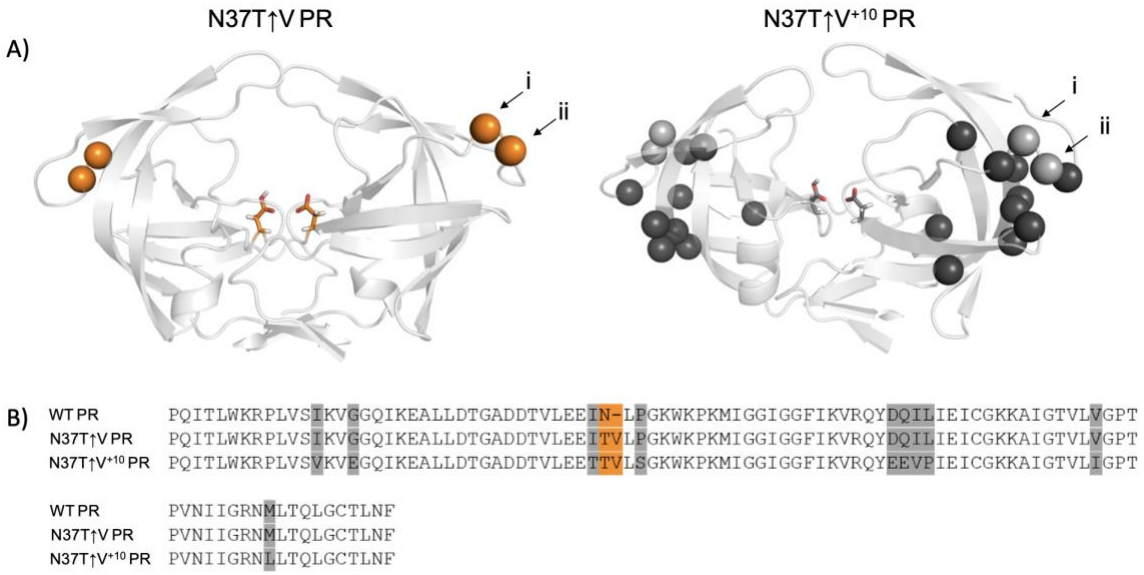


Figure 33: A) Ribbon representation of N37T↑V PR (left) and N37T↑V+10 PR (right) and B) Amino acid sequence alignment of the WT subtype C PR, N37T↑V PR and N37T↑V^{+10} HIV PR.

The arrows indicate (i) the N37T substitution and (ii) insertion mutations present in N37T↑V. The 10 background mutations present in N37T↑V^{+10} are represented in dark grey. The Asp25/25′ catalytic residues are represented as sticks.

mutations were investigated independently. Which is to say the hinge-region insertion was analysed with and without the 10 additional background mutations.

5.2 Methods

5.2.1 Homology modelling

Two homology models were constructed in order to study the effect of non-active site mutations on the dynamics of the PR. The first PR model, termed N37T↑V (Figure 33A), contained a single polymorphism (N37T) and an insertion of a valine at position 37 (↑V). The second PR model, N37T↑V^{+10} (Figure 33A), contained the same mutations present in N37T↑V but with the additional 10 background mutations: I13V, G16E, I36T, P39S, D60E, Q61E, I62V, L63P, V77I and M89L. The homology models were constructed using the SWISS-model online tool (Arnold *et al.*, 2006) with the atomic coordinates of the wild type (WT) PR (PDB ID: 3U71) used as a template. The homology models were then validated using PROCHECK (Laskowski *et al.*, 1993)

5.2.2 Ligand docking

The induced-fit docking studies were performed using Glide software (Schrödinger LLC 2009, USA) implemented on Windows architecture. The protein preparation step involved using the Schrödinger modules; Glide, Prime, and Protein Preparation Wizard (Schrödinger LLC 2009, USA). The 2D structures files of the nine HIV protease inhibitors (PIs) approved by the FDA were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/), namely, atazanavir (ATV, CID: 148192), darunavir (DRV, CID: 213039), fosamprenavir (FPV, CID: 131536), indinavir (IDV, CID: 5462355), lopinavir (LPV, CID: 92727), nelfinavir (NFV, CID: 64142), ritonavir (RTV, CID: 392622), saquinavir (SQV, CID: 60787) and tipranavir (TPV, CID: 54682461). The LigPrep tool (Schrodinger LLC 2009, USA) was used to convert the 2D ligand structure data into 3D structures by adding hydrogens, bond angles, and lengths, choosing the correct chirality, and performing energy minimization. The Epik tool was used to select the lowest energy tautomers and ring structures. Energy-minimised WT PR, N37T↑V PR, and N37T↑V^{+10} PR were docked with each of the nine PIs by defining the active site Asp25 residues as the centre of the docking region. The

induced fit docking was performed using the OPLS3 force field. The van der Waals radii of non-polar atoms of the PRs were scaled by a factor of 0.50, and 20 conformational poses were calculated for each complex. The docking score, Glide energy, and Glide E-model were used to choose the best conformation for further analyses using molecular dynamics simulations. Hydrophobic interactions and hydrogen bonds formed between the ligand and proteases were visualised using Maestro v12.2 (Schrödinger LLC 2009, USA).

5.2.3 Molecular dynamics

Molecular dynamics (MD) simulation studies were performed on Desmond software version 11.2 (Schrödinger LLC 2009, USA) using the OPLS3 force field for all the PR systems. The simulations were conducted with explicit solvent generated using the TIP3P solvation model within the System Builder module. The solvated systems were subjected to relaxation and minimisation using the steepest descent method until the system reached an atomic force of less than 100 $kJ.mol^{-1}.nm^{-1}$. This process performs a series of minimisations and short MD simulations in order to equilibrate each system before the experimental simulation commences. Briefly, a 5 ns MD simulation is performed under the NVT ensemble (constant number of particles (N), constant volume (V) and constant temperature (T)). Thereafter, a 5 ns simulation is performed under the NPT ensemble (constant number of particles (N), constant pressure (P) and constant temperature (T)) at a temperature of 10 K and a pressure of 1 atm. Following equilibration, the MD simulations were performed for 200 ns at a constant temperature of 300 K and an average pressure of 1 atm maintained by the Parrinello-Rahman barostat algorithm (Okumura *et al.*, 2007).

5.2.4 Computational data acquisition

The quality of the trajectories was assessed using the Simulation Quality Analysis Module in Maestro. The trajectory analyses such as atomic RMSF, RMSD, ligand interaction-diagrams, SASA and radius of gyration (R_g) were performed using the Simulation Event Analysis and Simulation Interaction Diagram modules in Maestro. The simulation trajectories were visualised and analysed using software UCSF Chimera version 1.11.2 and PyMOL (The PyMOL Molecular Graphics System, Version

1.8 Schrödinger, LLC.). Sequence alignments were performed using the Clustal Omega tool (EMBL-EBI).

5.3 Results and discussion

5.3.1 Non-active site mutations in N37T↑V PR and N37T↑V^{+10} PR reduce the predicted binding affinity of PIs

Induced fit docking (IFD) was used to systematically predict the interactions that occurred between the PR variants and nine FDA-approved PIs utilised in this study. The summarised induced fit docking parameters for each PR when bound to nine clinical inhibitors is shown in Table S3. The calculated free energy score (G_{score}) provides an approximation of the binding affinity of each protease variant for the nine PIs (Table 3). All of the inhibitors exhibit negative binding free energies for the WT PR and mutant PR variants. However, the efficiency of binding varied significantly when comparing the WT PR and the mutated PR variants. Notably, calculated G_{score} was not significantly different between the three PRs when in complex with IDV, LPV, and SQV. In contrast, the calculated G_{score} for the PRs when in complex with ATV, DRV, FPV, NFV, RTV and TPV exhibited marked decreases when in complex with N37T↑V PR and N37T↑V^{+10} PR (Table 3). The observed decrease in the G_{score} suggests that these PIs may bind the PR mutants less effectively. As such these complexes were selected for further analyses by MD simulations since IFD is limited to a static representation of the complexes which are inherently dynamic in nature.

5.3.2 Individual non-active site polymorphisms have varying impacts on the binding affinity of several PIs

In order to gain further insight into the contribution of each of the background mutations to the changes observed in the predicted affinity it was necessary to analyse the effect that reverting each mutation would have on the in G_{score} while retaining the other nine background mutations. From these analyses it is clear that the P40S mutation consistently provides the largest change in the calculated G_{score} when reverted to the wild type proline residue (Figure S3).

Table 3: Summary of G_{score} of nine FDA-approved PIs in complex with WT PR, N37T↑V PR and N37T↑V^{+10} PR.

Glide score (G_{score})			
Protease inhibitor	Wild Type	N37T↑V	N37T↑V^{+10}
ATV	-12.20	-11.15	-10.32
DRV	-7.26	-7.65	-8.61
FPV	-8.26	-10.01	-9.18
IDV	-11.36	-11.68	-11.09
LPV	-11.10	-11.33	-10.49
NFV	-8.74	-9.86	-9.37
RTV	-12.20	-12.10	-10.80
SQV	-12.25	-12.19	-12.23
TPV	-9.60	-10.63	-8.57

The P40S mutation occurs in the hinge region of the PR and the substitution of the proline residue for a more flexible serine may increase the mobility of this region in a manner that affects the flaps during binding. Future studies could be directed towards investigating the role that the proline residue plays in the dynamics of this region. To a lesser extent, reversion of the I13V, V77I and M89L mutations significantly improved the calculated binding affinity of ATV, DRV and TPV (Figure S3). Notably, it has been shown that the I13V and V77I mutations play a role in increased flap, and hinge region dynamics resulting in the PR favouring the semi-open conformation when in complex with these inhibitors (Sherry *et al.*, 2021). Constellations of non-active site mutations are required to reach high levels of drug resistance in the HIV PR. However, it is important to consider the contribution of individual mutations, or smaller sets of mutations, and how these contribute cooperatively to drug resistance mechanisms.

5.3.3 The N37T↑V PR and N37T↑V^{+10} PR can sample wider flap conformations for longer time periods

It is understood that flap tip curling promotes flap opening in the HIV PR (Wlodawer and Vondrasek, 1998). Therefore, both of these phenomena were analysed in order to determine the effect of the non-active site mutations on flap tip curling and flap dynamics. The angle between the adjacent alpha carbons (C_α) of the flap tips (Gly-49, Ile-50 and Gly-51) has been widely used as a metric to measure flap tip curling in HIV-1 protease (Scott and Schiffer, 2000). Where the flap tips are considered curled in and curled out when the angle between the adjacent alpha carbons is ~100˚ and ~140˚, respectively. While the interatomic distance between Cα atoms of Asp-25 and Ile-50 have been used to define closed, semi-open and fully open flap conformations as <17 Å, 17-22 Å and > 22 A, respectively (Perryman *et al.*, 2004). It should be noted that the hinge insertion present in both mutant PRs increased the residue numbering by one and that this was considered during analysis.

In monomer A (Figure 34) it is clear that the flap tip of the WT PR samples a larger angle of ~140˚ which is directly correlated with the monomer adopting a closed flap conformation with an average interatomic Asp25-Ile50 distance of <17 Å. Notably, the same monomer of both the N37T↑V PR and N37T↑V^{+10} PR exhibited curling-in of the flap tips at an angle of ~100˚ for the entire simulation which corresponded with the

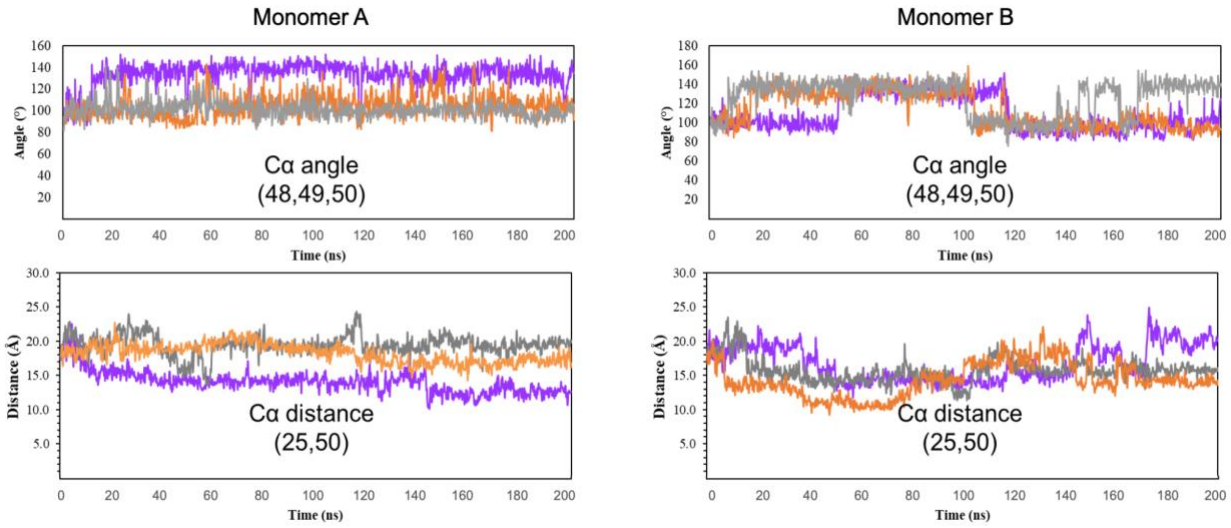


Figure 34: Flap tip angle and flap conformations of WT PR (purple), N37T↑V (orange) PR and N37T↑V^{+10} PR (grey).

Flap tip angles were measured as the angle between the Cα atoms of Gly-49, Ile-50 and Gly-51. Flap conformations were monitored by measuring the distance between the Cα atoms of Asp-25 and Ile-50.

flaps adopting the semi open conformation (17-22 Å). Monomer B of WT PR initially adopted a curled-in flap tip with an angle of 100˚, followed by curling-out of the flap tip to ~140˚ and then finally reverting to 100˚. The curling of the flap tips was, once again, mimicked in the conformation of the flaps. Here, the WT monomer initially adopted a semi-open conformation, followed by the closed conformation when the flap tips curled out towards ~140˚ and then moved back into the semi-open conformation when the flap tips curled in, to 100˚. These results show strong correlation between the conformation of the flaps and curling of the flap tip residues as has been suggested previously (Sherry *et al.*, 2021). Monomer B of N37T↑V PR and N37T↑V^{+10} PR immediately adopted curled-out flap tip conformations (~ 140˚) for the first half of the simulation before curling inwards for the latter half of the simulation. Interestingly, the mutant PRs were able to adopt the curled-out flap tip conformation for longer time periods than the WT PR. Together, these findings may suggest that the flaps of the N37T↑V^{+10} PR are more likely to sample stable conformations over longer timescales than the WT PR. This is evidenced by the longer periods of sustained flap opening and curled in flap tip conformations by the N37T↑V^{+10} PR. The background mutations in the hinges may increase the dynamics of the flaps, while mutations in the cantilever region may provide internal stability enabling the flaps to maintain open and closed conformations for longer durations. This is of importance when considering that altered flap dynamics and stability in the mutated PRs may have implications for decreased inhibitor susceptibility. Such mutations could allow the PR to exhibit an increased rate of flap opening and closing.

5.3.4 N37T↑V and N37T↑V^{+10} exhibit increased hinge and fulcrum region dynamics

The root mean square fluctuation (RMSF) allows for quantification of the average deviation of the PR backbone atoms over the course of the simulations. These analyses revealed significant differences in the dynamic behaviour of the apo N37T↑V PR and N37T↑V^{+10} PR in comparison to the apo WT PR. Specifically, the fulcrum and hinge regions of the N37T↑V and N37T↑V^{+10} mutants exhibited increased RMSF fluctuations, indicating increased flexibility in these regions (Figure 35). This increased fluctuation in the fulcrum was noticeably larger in the N37T↑V^{+10} PR relative to the N37T↑V PR. The N37T↑V^{+10} PR contains two mutations in the fulcrum region, I13V

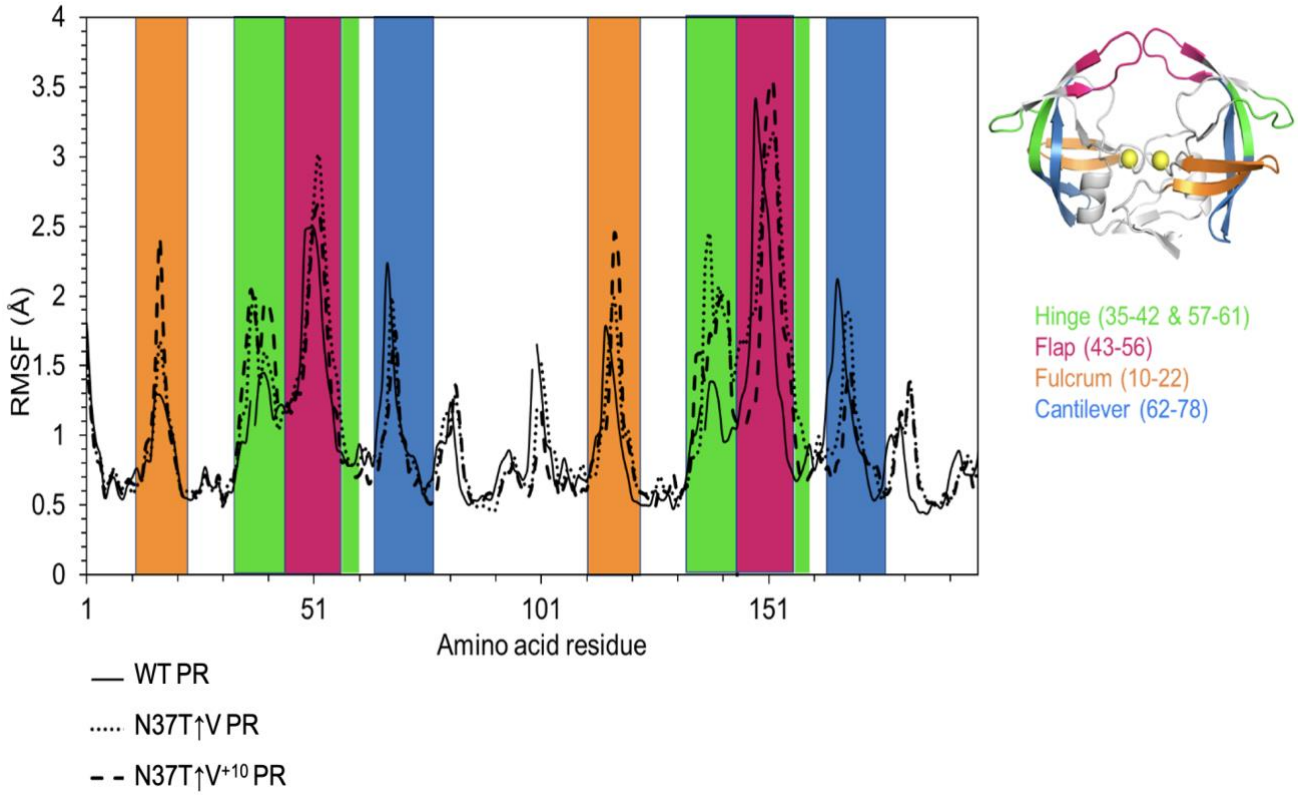


Figure 35: The average RMSF of the regions in the apo WT PR, N37T↑V PR and N37T↑V^{+10} PR HIV-1 protease.

and G16E, which may alter the network of intramolecular interactions in a manner that increases the flexibility of this region.

Interestingly, while the hinge regions of the mutants were more dynamic, the cantilever region was found to be more stable compared to the wild type, as indicated by a reduced RMSF in this region (Table S1). This finding deviated to a previous study where it was found that the I13V mutation increased flexibility in the cantilever region (Sherry *et al.*, 2021). The difference is most likely due to the presence of several additional secondary mutations present in the cantilever region of the N37T↑V^{+10} protease, namely, D60E, D61E, I62V, L63P, and V77I. Naturally occurring polymorphisms between subtype B and subtype C, such as L63P, have been shown to affect the conformational flexibility of the HIV-1 protease, with L63P increasing the rigidity of subtype B (Carter *et al.*, 2014). Residues 13, 62, 63, 77 and 89 form part of the twenty amino acids comprising the hydrophobic core of the PR (Foulkes-Murzycki *et al.*, 2007). The hydrophobic core plays a critical role in the flexibility and conformational changes of the PR and, specifically, the flap regions through the hydrophobic sliding mechanism (Goldfarb *et al.*, 2015; Mittal *et al.*, 2012; Ragland *et al.*, 2014). Significant alterations in the hydrophobic sliding mechanism have been postulated to result in reduced drug susceptibility. The presence of the 10 background mutations may be essential for the stabilisation of the cantilever in N37T↑V^{+10} PR in a manner that alters the intramolecular hydrophobic interactions of the PR. This stabilisation may be introduced to accommodate for the wider conformation of the flap regions that result from the increased dynamics of the hinges and fulcrum in the N37T↑V mutant. Overall, the results suggest that the N37T↑V PR and N37T↑V^{+10} PR mutants have altered dynamics in core regions of the PR which are used to coordinate drug binding which may, as a result, affect drug susceptibility.

5.3.5 Insertion mutations increase the dynamics key regions of the PR when complexed with PIs

The formation of stable interactions between the PR and PIs is crucial for strong binding interactions required for sufficient inhibition. Therefore, it follows that increased dynamics and flexibility of PR residues that contact PIs can weaken binding interactions and reduce drug susceptibility. Here, it was observed that the distinct

regions of the WT PR consistently exhibited decreased RMSF when in complex with several PIs, namely, ATV, DRV, FPV, NFV, RTV and TPV relative to the apo state (Figure 36). However, for both the N37T↑V PR and N37T↑V^{+10} PR, the RMSF of the hinge and flap regions were significantly increased. Particularly when in complex with ATV and DRV and to a lesser extent when in complex with FPV and TPV. Increased flexibility in these regions relative to the WT PR may be significant when considering inhibition. As such, higher mobility of the flaps may lead to weaker or more transient interactions formed with PIs in the active site since PIs are inherently rigid and unable to accommodate perturbations in the active site. Interestingly, we observed a significant decrease in fluctuations of these regions when the mutants were in complex with ritonavir (RTV), indicating that in the presence of RTV these regions become less dynamic. The study demonstrated that N37T↑V^{+10} mutations significantly increase the molecular dynamics of the flap and hinge regions in HIV protease when in complex with PIs (manuscript in preparation).

5.3.6 Cantilever tip angle contraction is essential for flap conformational changes

Recently, Sherry *et al.* (2021) highlighted the importance of cantilever tip angle contraction for facilitating conformational changes in flaps of the HIV-1 protease. While flap tip curling has traditionally been viewed as the primary mechanism for initiating flap movement, there is now evidence to suggest that the cantilever tip residues (Ile-66, Cys-67, Gly-68) may also play a critical role in regulating flap mobility. Previously it has been shown that when the cantilever tip Cα angle is curled-in (≤100˚) the flaps adopt a semi-open conformation and when this angle is curled-out (>100˚), the flaps adopt the closed conformation (Sherry *et al.*, 2021). In this study, the WT PR exhibited an average cantilever tip angle of 102.25˚, indicative of a relaxed (curled-out) cantilever tip which correlated with a closed flap conformation (<17 Å) throughout the simulation (Figure 37). In contrast, the N37T↑V and N37T↑V^{+10} mutants displayed cantilever tip angles of 96.73 and 94.76, respectively, indicating a more curled-in conformation. As such, it is noteworthy that the flaps of the mutant PRs were maintained in a semi-open flap conformation (17-22 Å). These findings provide further evidence to suggest that cantilever tip angle contraction is a critical mechanism for regulating flap mobility and conformational change in HIV-1 PR. Further illustrating,

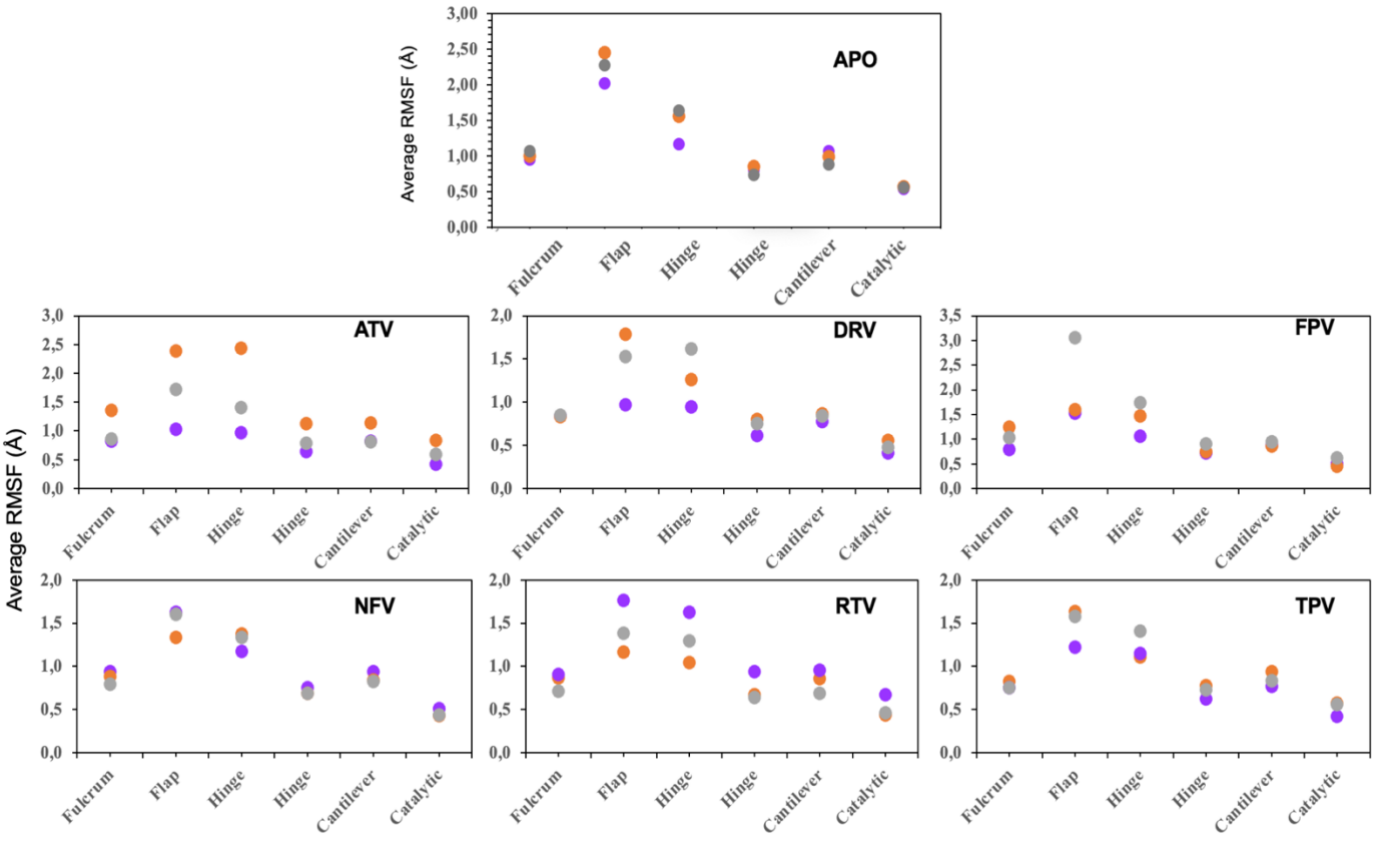


Figure 36: Average RMSF of distinct regions of the WT PR, N37T↑V and N37T↑V^{+10} in the apo and drug-bound state.

The distinct regions investigated include the fulcrum, flap, hinge, cantilever and catalytic region. WT PR (purple), N37T↑V (orange) and N37T↑V^{+10} (grey).

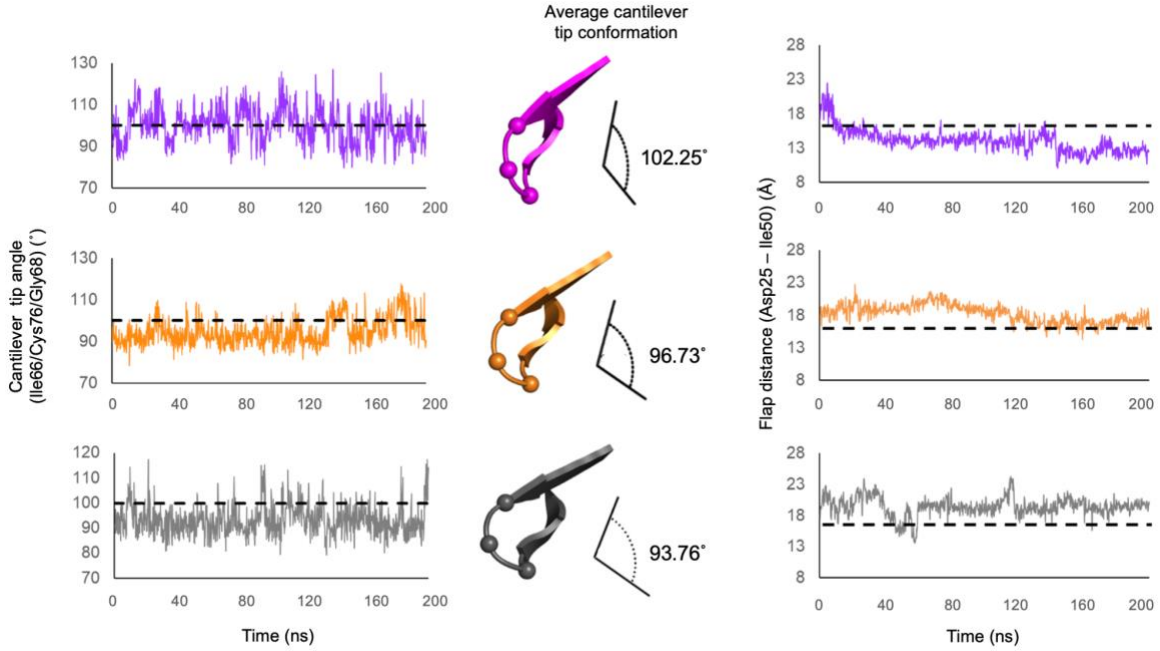


Figure 37: HIV PR cantilever tip contraction and flap conformational.

Left: cantilever tip Cα angle between Ile-66, Cys-67, Gly-68. The dotted line indicates the threshold between curled-in (≤100˚) and curled out (>100˚) as per Sherry *et al.*, 2021/ Centre: the average cantilever tip conformation observed during molecular dynamics simulations C_α atoms of Ile66/Cys-67/Gly-68 are represented as spheres. Right: flap distance, measured as the interatomic distance of the C_α from Ile-50 (flap) to Asp-25 (active site). Dotted line indicates the threshold between closed (<17 Å) and semi-open (17-22 Å) flap conformations.

therefore, that the global dynamics of the PR must be considered when considering mechanisms of PR catalysis, inhibition and resistance.

5.3.7 N37T↑V PR and N37T↑V^{+10} PR are less compact and do not fully enclose PIs

The flap regions of the PR can adopt a variety of conformational states which exist in equilibrium, however, inhibition via PIs brings the flaps into a more stable closed conformation. This stabilisation results from the binding free energy associated with intermolecular interactions formed between the PR and the inhibitor (Todd *et al.*, 1999). Hence, non-active site mutations that increase the flexibility of the flaps when complexed to inhibitors may increase the dissociation constant of the complex. A phenomenon that has been characterised in saquinavir resistance (Maschera *et al.*, 1996). The radius of gyration (R_g) of the PR can quantify how compact a protein is during the MD simulation and, thus, a more dynamic system would be reflected in an increased average R_g. Notably, the N37T↑V PR and N37T↑V^{+10} PR both exhibit higher average R_g values than those of the WT PR when in the apo state as well as when bound to all six PIs tested (Figure 38). When comparing the N37T↑V PR and N37T↑V^{+10} PR, the latter more consistently exhibits an increased R_g in comparison to N37T↑V PR. This may suggest that the additional 10 background mutations allow the N37T↑V^{+10} PR to sample a wider ensemble of dynamic states through increased internal rigidity. These R_g values indicate that the PIs are unable to fully “lock down” the mutant PRs as effectively as the WT PR. A similar trend is observed with the solvent accessible surface area (SASA) (Figure 38) for each PI which is elevated in the mutant PR systems, particularly when in complex with ATV, DRV and RTV. These results are further supported by the observation that the closed flap conformation for the WT PR facilitate tighter interactions and therefore a decreased SASA, but the semi-open conformation for the mutants results in less stable interactions and as a result this increase in dynamics is accompanied by an increase in the SASA. This increased solvent exposure would likely lead to a decrease in the entropic favourability of burying the hydrophobic PIs in the active site cavity for a favourable interaction, and as a result this would potentially contribute to drug resistance.

Moreover, such alterations in the dynamics of the PR flaps as well as the increased solvent exposure of the PIs are likely to affect the quantity and quality of the intermolecular interactions formed. The number and strength of hydrogen bonds formed between the protease and inhibitors plays a critical role in determining the binding affinity and specificity of the inhibitors (Rhee *et al.*, 2010). Our analyses reveal that the N37T↑V PR and N37T↑V^{+10} PR form fewer hydrogen bonds with ATV, DRV, and RTV (Figure S4). In these instances, the number of water-mediated contacts, or water bridges, that are formed in these complexes increases significantly. We have characterised a similar phenomenon in other PR variants containing non-active site mutations (Sherry *et al.*, 2021).This study postulates that the increase in flexibility of the flaps causes the inhibitor to be beyond hydrogen bonding distance with the active site Asp25/25′ residues when in contact with the flaps. This increased distance necessitates the formation of water-bridges to simultaneously maintain contacts with the active site and the flaps. Furthermore, the lability of water-bridges in comparison to hydrogen bonds formed directly between the PR and inhibitor could diminish the enthalpic favourability of inhibitor binding. Therefore, it can be concluded that the differences observed in the dynamics of the flaps of the N37T↑V PR and N37T↑V^{+10} PR may decrease inhibitor susceptibility by simultaneously decreasing the entropic and enthalpic favourability of inhibitor binding.

5.3.8 Alteration in intramolecular salt bridge formation

The HIV PR contains several structurally important salt-bridges involving residues of the flap and hinge regions of the PR involving the following residue pairs: Lys-20/Glu-34, Asp-30/Lys-45 and Glu-35/Arg-57 and Lys-43/Asp-60 (Figure S5. These salt bridges have been shown to play an important role in the dynamics of the flap regions (York *et al.*,1993, Swairjo *et al.* 1998, Zondagh *et al* 2018). Furthermore, changes in the strength and occurrence of salt bridges in and surrounding the flaps have been hypothesised to play a role in drug resistance (Naicker *et al.*, 2013, Zondagh *et al.*, 2018). Analyses of the apo and drug-bound complexes of the N37T↑V PR and N37T↑V^{+10} PR revealed interesting changes in the interatomic distances of several salt bridges. Specifically, an increased interatomic distance was observed for two

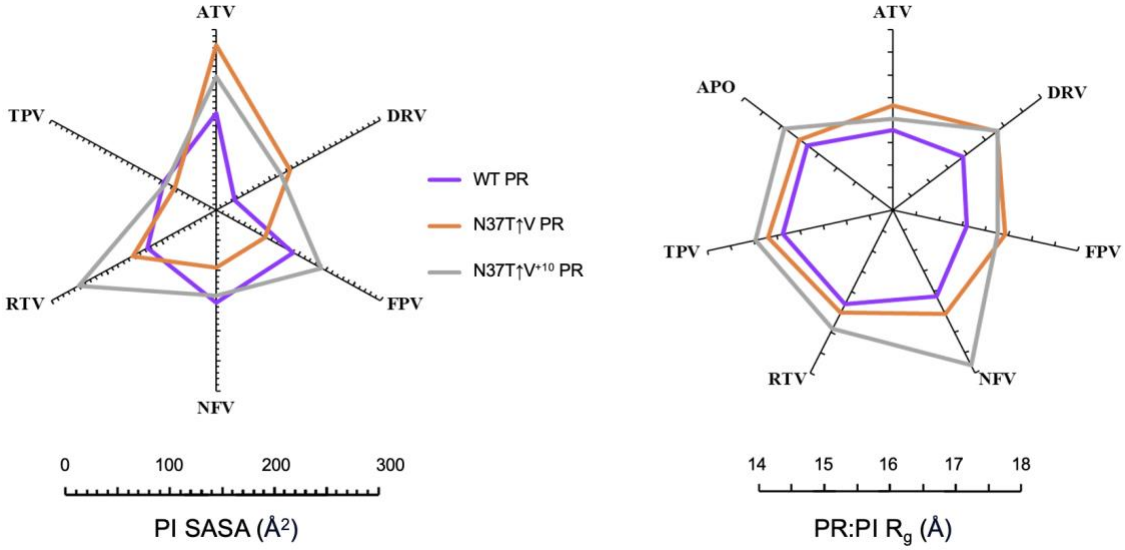


Figure 38: PI solvent accessible surface area (SASA) and radius of gyration (Rg) of WT PR (purple), N37T↑V PR (orange) and N37T↑V^{+10} PR (grey).

The PI SASA was calculated for the PR when in complex with six inhibitors: ATV, DRV, FPV, NFV, RTV and TPV. The R_g values were calculated for the same six systems, as well as for the apo PR systems.

state and in complex with all six PIs (Figure S6). The Lys20/Glu34 salt bridge anchors the hinge region to the fulcrum region, while Asp30/Lys45 anchors the flap region to the active site. An increase in the average interatomic distance observed in these anchoring salt bridges is likely to result in increased dynamics of the flap and hinge regions relative to the WT PR. Conversely, salt bridges between the flap and hinge region exhibited a decreased average interatomic distance in the mutant PRs (Figure S6). Both the Lys43/Asp60 and Lys20/Glu35 salt bridges connect the hinge region to the flap region (Figure S5). The strengthening of these interactions may provide increased receptivity of the flap regions to alterations in the dynamics of the hinges. It is well understood that the Glu-35/Arg-57 bridge plays a crucial role in maintaining the closed conformation of the flaps (Swairjo *et al.* 1998, Zondagh *et al.*, 2018). Therefore, it could be postulated that changes in the interatomic distances of several other key salt bridges could significantly alter flap dynamics of the PR. As such, future work should investigate the importance of these specific salt bridges involving the flap and hinge regions of the PR.

5.4 Conclusion

The dynamics of the hinge and flap regions of the PR play an important role in successful function as well as inhibition of the enzyme. Non-active site mutations that alter the flexibility of the hinge and flap regions may contribute to drug resistance though altering the dynamics of key regions of the HIV-1 PR. Novel PR variants with hinge region insertions have been increasingly characterised in literature (Maseko *et al.*, 2019; Ragland *et al.*, 2014; Sherry *et al.*, 2022, 2021; Venkatachalam *et al.*, 2023; Williams *et al.*, 2019; Zondagh *et al.*, 2018). Understanding the mechanism of action of hinge region insertions may shed light on how such mutations could lead to reduced drug susceptibility.

This study aimed to investigate how background mutations of N37T↑V^{+10} PR augment the effect of the insertion and mutation present in the hinge region of the N37T↑V PR. The insertion mutation and non-active site polymorphisms result in significant changes in the flap dynamics and flexibility of the enzyme (Figure 34, Figure 35). The apo mutant PRs adopt a semi-open conformation with wider flap conformations, while the

apo WT PR is more compact and adopts a more closed conformation. It would appear that the hinge region mutations present in N37T↑V PR introduce significant flexibility which translates to more dynamic flaps. However, the excessive flexibility of the hinges may be detrimental without the stabilising effects of the background mutations present in N37T↑V^{+10} PR. We have also observed more flexible hinges and fulcrum in the mutants, while the cantilever region showed reduced flexibility. These conformational changes were also associated with an increase in the curling of the cantilever Ile67/Cys68/Gly69 tip residues which has been correlated with a more open flap conformation in the mutated PR variants (Figure 37).

These data also indicate that the flap regions of the N37T↑V PR and N37T↑V^{+10} PR fluctuate more readily when in complex with certain PIs in comparison to the WT PR (Figure 36). In addition, the mutant PRs display higher R_g when drug bound and, in some instances, the PIs were more solvent exposed in comparison to the corresponding WT PR complexes (Figure 38). The mutations present in N37T↑V PR and N37T↑V^{+10} PR caused increased solvent exposure of several PIs, suggesting that the complex formation would be less entropically favourable. Moreover, the mutant proteases formed less hydrogen bonds with inhibitors and, rather, formed less-stable water-bridges. This is likely a result of the ability of the mutated PRs to adopt wider flap conformations for longer time periods (Figure 34).

In conclusion, this study proposes that the insertion mutation present in N37T↑V PR and the additional background mutations present in N37T↑V^{+10} PR result in less favourable binding of inhibitors both enthalpically and entropically. The increased flap dynamics and reduced drug binding interactions introduced by these non-active site mutations likely lead to resistance through altering the dynamics of the PR. These results have important implications for understanding how non-active site mutations lead to the development of drug resistance.

Chapter 6: General summary and discussion

HIV-1 continues to affect millions of people worldwide and an effective cure is yet to be discovered (UNAIDS, 2022). HIV-1 PR, an aspartyl PR, plays a critical role in the maturation of HIV-1 and, as such, is a major drug target for antiretroviral therapy (Darke *et al.*, 1989; Debouck, 1992; Kiran Kumar Reddy *et al.*, 2007; Peng *et al.*, 1989). Inhibiting this enzyme prevents HIV-1 maturation and halts the replication cycle of the virus. Traditional drug design methods have relied on the lock-and-key paradigm and involved designing PIs that bind tightly to the active site (binding site) of HIV-1 PR (Wlodawer and Vondrasek, 1998). The binding site is where the natural substrates, Gag and Gag-Pol polyproteins are bound and cleaved to release mature viral components (Erickson-Viitanen *et al.*, 1989). Alteration of the binding site in shape or biochemistry may disrupt PR:PI interactions, decreasing the binding strength of the PIs. This may decrease the drug efficacy and lead to drug resistance.

HIV-1 PR has previously been found to accumulate mutations within and distal to the active site in response to selective drug pressure imposed by drug regimens. These mutations were found to disrupt the PR:PI interface and alter drug binding (Goldfarb *et al.*, 2015; Mosebi *et al.*, 2008; Ragland *et al.*, 2014; Sherry *et al.*, 2022). The role that such non-active site mutations play in reduced drug susceptibility is not fully understood, however, it is hypothesised that their presence in the protease sequence alters the dynamic ensemble of the protease both in the apo- and drug-bound states. Understanding the HIV-1 PR structure and dynamics may contribute to the design of drugs that are more effective and less susceptible to drug resistance. With the emergence of HIV-1 subtype C hinge-region variants in mind (Naicker and Sayed, 2014; Velazquez-Campoy *et al.*, 2001), this study aimed to assess how a hinge region insertion and substitution impact the overall biophysical and biochemical characteristics of an HIV-1 subtype C variant. With the use of a novel hinge region variant, N37T↑V^{+10}, obtained from a drug naïve infant, the study assessed the impact the hinge region mutations have on the structure, stability and drug-binding characteristics of N37TV.

6.1 Overexpression of WT•D25A and N37T↑V^{+10}•D25A

A major challenge in the in vitro study of HIV-1 protease is the ability to overexpress and purify high yields of correctly folded HIV-1 PR. HIV-1 PR is inherently proteolytic and in the absence of a substrate, will undergo self-hydrolysis, leading to a decrease in the yield. For this study, a D25A mutation was incorporated into the sequence of both WT and N37T↑V^{+10} PRs, resulting in the inactive WT•D25A and N37T↑V^{+10}•D25A, respectively. The D25A mutation disrupts the catalytic triad of an aspartyl PR making it inactive and preventing the cytotoxic and autoproteolytic activity of the enzyme. This leads to an increase in the yield of PR and allows for the structural and functional characterisation of HIV-1 PR. Aspartic acid and alanine have similar geometry, as their main chain atoms (N-C$_\alpha$-C) have similar bond angles and lengths. Therefore, the incorporation of the D25A mutation does not significantly alter the shape of the HIV-1 PR active site. While previous studies included only a Q7K mutation to limit autoproteolysis, in this study a D25A mutation was included in WT•D25A and N37T↑V^{+10}•D25A to completely block PR activity, as confirmed by an activity assay. The sequence of WT•D25A and N37T↑V^{+10}•D25A were cloned into two separate pE11-a vectors, respectively. Sanger sequencing of the cDNA confirmed the inclusion of the insertion and substitution mutations at the 37th amino acid position, as well as the D25A mutation in N37T↑V^{+10}•D25A. Additionally, N37T↑V^{+10}•D25A was shown to have ten background mutations, which differ from the consensus South African HIV-1 subtype C PR.

The pET-11a constructs containing the sequences for WT•D25A and N37T↑V^{+10}•D25A were successfully transformed into BL21 (DE3) pLysS and Rosetta *E. coli* cells. *E. coli* cells were used in this study for heterologous protein expression. The short replication time and ability to produce high protein yields make *E. coli* an attractive expression system (Darke *et al.*, 1989). BL21 (DE3) pLysS and Rosetta *E. coli* cells have previously been reported to efficiently express cytotoxic proteins, such as HIV-1 protease (Lama and Carrasco, 1992; Rosano and Ceccarelli, 2014; Zondagh *et al.*, 2018). Overexpression of N37T↑V^{+10}•D25A in Rosetta cells was induced with 0.8 M IPTG over 4 hours. It is worth noting that N37T↑V^{+10}•D25A was primarily expressed in the insoluble fraction as inclusion bodies, which are protein aggregates (Ido *et al.*, 1991). This is consistent with previous studies of the expression of HIV-1

PR in *E. coli* (Maseko *et al.*, 2016; Nguyen *et al.*, 2015). Inclusion bodies are important as they limit the functionality of HIV-1 protease and preventing it from folding into its functional, cytotoxic form. Ordinarily, it would be ideal to express a protein in a soluble form, however, due to the cytotoxic nature of HIV-1 protease, expression in the insoluble fraction is ideal.

6.2 Purification of WT•D25A and N37T↑V^{+10}•D25A

The purification of HIV-1 PR from inclusion bodies has historically proven to be a challenge, primarily due to insufficient lysis of cells containing the inclusion bodies, resulting in the incorporation of whole cells during purification. To overcome this and improve the downstream purification and characterisation of HIV-1 PR, an additional cell lysis step was employed. This method, supported by the findings of Walker (2009), involved treating the cell suspension with several freeze-thaw cycles. Freezing the cells forms ice crystals, which cause the cell to swell until they rupture, while thawing causes the cells to contract. Treating the cell suspension with three freeze-thaw cycles improved cell lysis and enhanced the release of proteins from the cell cytoplasm.

Isolating the protein from inclusion bodies is important for the downstream characterisation of that protein. Here, two purification techniques were used to isolate WT•D25A and N37T↑V^{+10}•D25A. For WT•D25A isolation, the two-step purification method, recently reported by Sherry *et al.*, (2020) was employed. For the N37T↑V^{+10}•D25A variant, cation-exchange chromatography was used. The successful purification of WT•D25A and N37T↑V^{+10}•D25A was confirmed by tricine SDS-PAGE. The PRs both had a monomeric size of ~11 kDa, confirming the predicted size. However, WT•D25A yields were significantly decreased due to aggregation of the PR. For future experiments, aggregation may be decreased by adding smaller dialysis steps between refolding to gradually remove the denaturant (urea) over a longer time. This would give the protein time to fold at a slower rate minimising the chances of misfolding and aggregation. It may also be beneficial to revisit the use of fusion tags such as Trx, which was recently reported to reduce aggregation compared to other expression systems (Maseko *et al.*, 2019).

6.3 Active site titration

The results obtained through ITC provide important insights into the functional properties of the studied protein, HIV-1 protease (PR). By using acetyl pepstatin as a ligand that binds to the active site of the protease, the concentration of properly folded PR molecules in their functional form was estimated. Acetyl pepstatin specifically binds to aspartyl proteases, making it suitable for this study. The stoichiometry parameter (n) was determined to assess the ratio of PR molecules capable of binding the ligand and converted to percentages. For the WT•D25A PR variant, 57.4% of the molecules were found to be correctly folded, while the N37T↑V^{+10}•D25A variant exhibited 90.1% correctly folded molecules.

By correcting for misfolding, the concentrations of the PR variants were adjusted to 20 µM for WT•D25A and 24 µM for N37T↑V^{+10}•D25A. The change in enthalpy (ΔH) observed during the titration with acetyl pepstatin was similar for both PR variants, indicating comparable binding behaviour. The change in entropy (ΔS) was determined to be 256.7 J/mol·K for WT•D25A and 261.6 J/mol·K for N37T↑V^{+10}•D25A, suggesting differences in the dynamics and flexibility of the protein structures. Acetyl pepstatin is a serine protease inhibitor and binds PRs. The dissociation constant (K_d) was calculated to be 3.12 x 10^{-7} M for WT•D25A and 5.67 x 10^{-7} M for N37T↑V^{+10}•D25A. This parameter provides information on the strength of the interaction between the PR and the ligand, with a lower K_d indicating a higher affinity between to the PR and acetyl pepstatin. What was observed in this experiment is that the K_d obtained for both WT•D25A and N37T↑V^{+10}•D25A PR are similar in magnitude and previous studies have shown that the value for inactive WT•D25A was found to be 3.761 x 10^{-7} M (Sherry, *et al.*, 2020) and for active WT was 4.4 x 10^{-7} M (Sheikh-Ismail, *et al.,* 2023).

These results provide valuable insights into the concentration, folding status, thermodynamic properties, and binding characteristics of the WT•D25A and N37T↑V^{+10}•D25A variants of HIV-1 PR. They contribute to a better understanding of the functional behaviour of the protein and its potential implications in drug binding and catalytic activity.

6.4 Structural characterisation of WT•D25A and N37T↑V^{+10}•D25A

6.4.1 Secondary structure

By using far-UV circular dichroism, the determination of the predominance of α-helices, β-sheets or random coils within the protease was possible and this allowed the characterisation of the secondary structure of the wild type and N37T↑V^{+10} subtype C protease. The wild type subtype C protease has been previously shown to be predominantly β-stranded exhibiting a peak at approximately 416 nm (Liu *et al.*, 2016; Mosebi *et al.*, 2008; Naicker *et al.*, 2013). The effect of the mutation and insertion was assessed to probe how these mutations alter the secondary structure of the protease.

The far-UV circular dichroism technique has been widely used to study the secondary structure of proteins by measuring the absorbance of circularly polarised light by the amino acid components. The secondary structure of a protein is composed of α-helices, β-sheets or random coils and this technique allows researchers to determine which structural component predominates in the protein. Recent research has shown that even small alterations caused by polymorphisms can significantly impact the secondary structural folding of proteins. Therefore, the impact insertion, substitution, and background mutations in N37T↑V^{+10}•D25A have on the secondary structure of the HIV-1 protease was assessed by comparing the CD of both PR.

These results show that both WT•D25A and N37T↑V^{+10}•D25A produce CD spectra that are characteristic of proteins predominantly composed of β-sheets. The trough for WT•D25A and N37T↑V^{+10}•D25A was at 216 nm and 218 nm, respectively, falling within the range of a predominantly β-sheeted protein. These findings are consistent with previous studies on the secondary structure of HIV-1 protease, which suggest that the enzyme is comprised of 4% alpha helices, 48% β-sheets, and 48% random coils (Foulkes *et al.*, 2006; Naicker *et al.,* 2014). Moreover, the data obtained from our study also aligns with the predicted structural content found by ProtParam tool, indicating

that both the WT•D25A and N37T↑V^{+10}•D25A are predominantly β-sheeted. The data suggests that the hinge region insertion and substitution do not significantly alter the overall secondary structural folding of HIV-1 protease.

6.4.2 Tertiary structure

Intrinsic tryptophan fluorescence spectroscopy has long been used to characterise the local tertiary structure of proteins. This makes use of the indole ring present in tryptophan's ability to absorb UV light and emit fluorescence at longer wavelengths. Using fluorescence spectroscopy the tryptophan residue can be excited at a maximum wavelength of 295 nm (Carmel and Yaron, 1978). The emission wavelength of tryptophan depends on the location of the fluorophore. When tryptophan residues are completely buried within the protein structure, the residues emit fluorescence at shorter wavelengths (less than 320 nm). Conversely, when tryptophan residues are free and exposed to the polar environment, the residues emit fluorescence at wavelengths at longer wavelengths (between 350 nm and 355 nm) (Lakowicz, 1999). Tryptophan is large in size, relatively scarce and has a particular sensitivity to the polarity of the microenvironment (Demchenko, 2013; Osysko and Muíño, 2011). This makes it a reliable probe to assess the local tertiary structure of proteins.

In this study, both WT•D25A and N37T↑V^{+10}•D25A were evaluated using intrinsic fluorescence spectroscopy. Both WT•D25A and N37T↑V^{+10}•D25A displayed red-shifted fluorescence peaks, with emission maxima of 346 nm and 347 nm, respectively. This red-shifted fluorescence may be attributed to the particular tryptophan residues (Trp 42/43 and Trp 42'/43) present in WT•D25A and N37T↑V^{+10}•D25A. After the PRs were treated with 8 M urea and unfolded the emission wavelength shifted to 351 nm and 352 nm, respectively. This additional red-shift of the emission wavelength is attributed to additive fluorescence emission from the tryptophan residues (Trp 6 and Trp 6') which are usually buried but become exposed as the protein unfolds. Interestingly, there was a reduction in fluorescence intensity for the unfolded PR. This reduction in fluorescence intensity is suggested to be due to the solvent ability to stabilise the transfer of electric charge from the excited indole ring to

the protein's amide backbone (Fidy *et al.* 2001; Lakowicz, 2006). This leads to a decrease in fluorescence intensity for unfolded proteins.

6.4.3 Quaternary structure

The hydrodynamic volume of WT•D25A and N37T↑V^{+10}•D25A was measured using SE-HPLC. Protein standards were used to construct a standard curve and the molecule size of both proteins was determined. This molecular size of WT•D25A was found to be 20 kDa and N37T↑V^{+10}•D25A molecular size was determined to be 22 kDa, respectively. The molecular size of both proteins is similar. Additionally, the results showed the presence of two oligomeric states for both proteins. Because HIV-1 PR is homodimeric in its native, function form the two observed oligomeric states were understood to be the dimer and monomer.

6.4.4 Macromolecular crystallisation

Macromolecular crystallisation is an essential step in determining the three-dimensional structure of a protein using X-ray crystallography. The protein data bank (PDB) is populated by a large amount of solved protein crystal structures, however, crystal structures of subtype C HIV-1 PRs are still in the minority when compared to subtype B PRs. Obtaining the crystal structure of PRs provides valuable information which may aid in drug discovery. Conditions for crystal growth were successfully identified. These need to be optimised in order to grow diffraction quality crystals. This would require high concentrations of PR. When protein was concentrated it was not stable at high concentrations of 5 mg/ml and formed aggregates. This was a limitation in the study and as a result diffraction of N37T↑V^{+10}•D25A crystals was not possible.

6.4.5 Hinge-region mutations have a minimal impact on PR structure

The structure of the protein is closely related to its biological function. By studying the structure, we may get more information on the function of the protein. Because HIV-1 PR is prone to mutations, it was essential to investigate if these mutations distal from the active site affect the structural integrity of the protein. The secondary, tertiary and quaternary structures of the WT•D25A and N37T↑V^{+10}•D25A behaved similarly, suggesting the hinge-region mutations did not have an observable impact on the

overall structure of the PR. It was necessary to further investigate the role played by hinge-region mutations in HIV-1 PR. Two complementary techniques were used to achieve this. Firstly, unfolding studies were conducted to investigate the conformation stability of N37T↑V^{+10}•D25A compared to WT•D25A. Differences in the unfolding could shed light on the impact these mutations have on the overall thermal and chemical stability of the mutations. Secondly, computational methods were employed to investigate the impact hinge region mutants have on the intermolecular interactions, molecular dynamics and drug interactions of HIV-1 PR.

6.5 Conformational stability of N37T↑V^{+10}•D25A

The conformational stability of a protein reflects how stable a protein's native three-dimensional structure is (Pace *et al.*, 1996; Zhou, 2004). This can be determined by the difference in Gibbs free energy (∆G) between its folded and unfolded states. Alterations in a protein's conformational stability can affect its biological activity and drug binding (Ahmad *et al.*, 2012; Huang *et al.*, 2014; Noel *et al.*, 2009; Palese, 2017). Mutations have been reported to change molecular inter/intra-interactions and altering the PR conformational stability. With this in mind, the conformational stability of WT•D25A and N37T↑V^{+10}•D25A was investigated. By assessing the unfolding of the secondary and tertiary structures, the study was able to provide valuable insights into the impact of mutations in N37T↑V^{+10}•D25A and the effect on conformational stability and integrity of the structures. This study made use of thermal denaturation and urea-induced unfolding to assess the secondary and tertiary structure stability of the PRs.

6.5.1 N37T↑V^{+10}•D25A folds by a two-state model and has reduced conformational stability

The urea-induced unfolding of N37T↑V^{+10}•D25A secondary structure was monitored using far-UV CD. This produced a sigmoidal curve, which reflects the cooperative unfolding of the secondary structure. This relates to a protein that has a single well-defined transition event between the native (folded) and unfolded state. Moreover, this suggests that the secondary structure maintains a stable compact conformation. Cooperative unfolding has previously been used to highlight the stability and

robustness of the secondary structure of WT•D25A. The transition event for N37T↑V^{+10}•D25A occurs in a range from 3 M to 4.3 M urea concentration. This is a higher range than previously obtained for an active WT HIV-1 PR (Noel *et al.,* 2009). This result may be due to protease inactivity caused by the D25A mutation. It would be interesting to compare the stability of active and inactive proteins by urea-induced equilibrium unfolding in order to make these values comparable. The transition event is important when studying the stability of the secondary structure. It is here where the PR proceeds to unfold and loses its characteristic conformation. Understanding this transition provides insights into the critical threshold at which the secondary structure begins to unfold and highlights the stability limits of this structure.

It is noteworthy that the unfolding of the tertiary structure of the N37T↑V^{+10}•D25A PR variant occurs at a lower urea concentration range of 1.75 M to 3 M compared to the secondary structure. This contrast implies that the secondary structure is more stable and resilient to unfolding compared to the tertiary structure. The differential unfolding behaviour of the secondary and tertiary structures suggests that the secondary structure plays a crucial role in maintaining the overall stability and integrity of the protein. The *m*-value, also known as the denaturant molar concentration midpoint, is a measure of the cooperativity of protein unfolding. It represents the change in denaturant concentration required to shift the population of folded and unfolded protein states from 10% to 90%. The *m*-value provides information about the extent to which denaturant-induced unfolding occurs in a cooperative manner. Proteins with higher *m*-values exhibit greater cooperativity, meaning that their unfolding transition is more abrupt and occurs over a narrower range of denaturant concentrations. The *m*-value is a useful parameter in studying protein stability and unfolding. It provides insights into the thermodynamic and kinetic aspects of protein folding and can help in characterising the effects of environmental conditions, mutations, or ligand binding on protein stability. When the urea-induced equilibrium was monitored by intrinsic tryptophan fluorescence, *m*-value for WT•D25A and N37T↑V^{+10}•D25A were 2.63 ± 0.04 kcal.mol^{-1}.M^{-1} and 2.49 ± 0.09 kcal.mol^{-1}.M^{-1}, respectively. A higher *m*-value is related to a greater stability, this means that mutations present in N37T↑V^{+10}•D25A may reduce the stability as there is less urea required to unfold the protein and the *m*-value is also less. These *m*-values are in agreement with previous values of 2.89 ±

0.08 kcal.mol^{-1}.M^{-1} HIV-1 PR. The ΔG_{water} for WT•D25A and N37T↑V^{+10}•D25A were 12.815 kcal.mol^{-1} and 11.79 kcal.mol^{-1}, respectively. Further affirmation of reduced stability introduced by the hinge region mutations.

These findings have broader implications in the field of protein folding and stability. The differential unfolding behaviour of the secondary and tertiary structures provides insights into the folding pathways and stability landscape of the N37T↑V^{+10}•D25A PR variant. Understanding these folding characteristics can contribute to a deeper understanding of the protein's function, stability, and potential therapeutic applications. By elucidating the structural stability of the N37T↑V^{+10}•D25A PR variant, this research paves the way for further investigations into the protein's behaviour and its interactions with drugs or inhibitors.

Thermal shift assay was monitored using far-UV CD at a constant wavelength of 216 nm. Thermal unfolding was irreversible as high temperatures led to the permanent denaturation of the PR. The study reports a denaturation temperature (T_m) was used to measure and compare the stabilities of the PR. At this temperature, the unfolded PR population is greater than the folded population. The observed T_m for WT•D25A was 62 °C, while N37T↑V^{+10}•D25A had a slightly lower T_m of T_m= 58 °C. The decrease in T_m observed in N37T↑V^{+10}•D25A relates to a decrease in thermal stability, indicated by the decrease in T_m by 4 °C. Todd *et al.*, (1998) reported a denaturation temperature of 59 °C for WT. This suggests that the hinge-region mutations may alter interactions within the protein, which results in a slightly reduced conformational stability.

6.6 Flap dynamics from a microsecond perspective

Molecular dynamics simulations can provide atomic-level insights into the motions of several key regions of the protease associated with flap mobility and stability. However, due to the large computational cost associated with performing such simulations, many studies are limited in their length which may obscure dynamic movements which occur over microsecond timescales. Here, a microsecond-long molecular dynamics simulation of a wild-type subtype C protease and a novel protease variant containing an insertion in the hinge region, named N37T↑V^{+10} PR is presented.

This study showed how mutations present in the N37T↑V^{+10} PR impacted the stability and dynamics of the PR and several key regions of the enzyme associated with flap mobility.

These data illustrate that the flaps of the apo N37T↑V^{+10} PR are more conformationally rigid than those of the apo WT PR. There is evidence to suggest that the "ground-state" of the apo PR is, in fact, the closed flap conformation. As such, mutations that increase the rigidity of the flaps of the PR when in the apo state could be associated with an increased energetic penalty associated with flap opening, thereby reducing the favourability of this conformational change. In contrast, when in complex with ATV and DRV, the N37T↑V^{+10} PR exhibits exaggerated flap mobility, particularly when in complex with DRV. These data suggest that the flaps of the N37T↑V^{+10} PR are not fully "locked-down" by these inhibitors as is the case with the WT PR. Furthermore, it was foud that the increased dynamics of the flap regions when in complex with PIs decreases the longevity of intermolecular interactions formed between the inhibitors and several key residues of the active site. The minimisation of intermolecular interactions would likely diminish the enthalpic favourability of complex formation which would decrease the binding affinity of the PIs. As such, it would appear that the mutations present in the N37T↑V^{+10} PR may contribute to reduced drug susceptibility in two ways: i) through increasing rigidity of the flaps in the apo state, thereby increasing the energetic penalty associated with binding and, ii) by increasing the flexibility of the flap regions when in complex with PIs so as to minimise intermolecular contacts and prevent inhibitors from locking the protease into a stable, closed conformation.

6.7 Hinge-region mutations result in less favourable binding of inhibitors both enthalpically and entropically

The dynamics of the hinge and flap regions of the PR play an important role in successful function as well as inhibition of the enzyme. Non-active site mutations that alter the flexibility of the hinge and flap regions may contribute to drug resistance through altering the dynamics of key regions of the HIV-1 PR. Novel PR variants with hinge region insertions have been increasingly characterised in literature (Maseko *et*

al., 2019; Ragland *et al.*, 2014; Sherry *et al.*, 2021, 2022; Venkatachalam *et al.*, 2023; Williams *et al.*, 2019; Zondagh *et al.*, 2018). Understanding the mechanism of action of hinge region insertions may shed light on how such mutations could lead to reduced drug susceptibility.

The study used molecular dynamic simulations to investigate how background mutations of N37T↑V^{+10} PR augment the effect of the insertion and mutation present in the hinge region of the N37T↑V PR. The insertion mutation and non-active site polymorphisms result in significant changes in the flap dynamics and flexibility of the enzyme (Figure 34, Figure 35). The apo mutant PRs adopt a semi-open conformation with wider flap conformations, while the apo WT PR is more compact and adopts a more closed conformation. It would appear that the hinge region mutations present in N37T↑V PR introduce significant flexibility which translates to more dynamic flaps. However, the excessive flexibility of the hinges may be detrimental without the stabilising effects of the background mutations present in N37T↑V^{+10} PR. The study also found that there are more flexible hinges and fulcrum in the mutants, while the cantilever region showed reduced flexibility. These conformational changes were also associated with an increase in the curling of the cantilever Ile67/Cys68/Gly69 tip residues which has been correlated with a more open flap conformation in the mutated PR variants (Figure 37).

These data also indicate that the flap regions of the N37T↑V PR and N37T↑V^{+10} PR fluctuate more readily when in complex with certain PIs in comparison to the WT PR (Figure 36). In addition, the mutant PRs display higher R_g when drug bound and, in some instances, the PIs were more solvent exposed in comparison to the corresponding WT PR complexes (Figure 38). The mutations present in N37T↑V PR and N37T↑V^{+10} PR caused increased solvent exposure of several PIs, suggesting that the complex formation would be less entropically favourable. Moreover, the mutant proteases formed fewer hydrogen bonds with inhibitors and, rather, formed less stable water-bridges. This is likely a result of the ability of the mutated PRs to adopt wider flap conformations for longer time periods (Figure 34).

In conclusion, the researcher postulates that the insertion mutation present in N37T↑V PR and the additional background mutations present in N37T↑V^{+10} PR lead to less

favourable binding of inhibitors both enthalpically and entropically. The increased flap dynamics and reduced drug binding interactions introduced by these non-active site mutations likely lead to resistance by altering the dynamics of the PR. These results have important implications for understanding how non-active site mutations lead to the development of drug resistance.

6.8 Potential impact of hinge-region mutations on the hydrophobic sliding mechanism of HIV-1 protease

The N37T↑V^{+10}•D25A mutant of HIV-1 protease displays notable structural alterations and increased dynamics compared to the WT•D25A. These observed changes have the potential to impact the hydrophobic sliding mechanism, a crucial process involved in the opening of the flaps of HIV-1 protease. The flaps, flexible regions located near the active site, play a critical role in substrate binding and catalytic activity by undergoing conformational changes. During the opening of the flaps, specific hydrophobic interactions between residues within the flaps and the substrate are disrupted, enabling access to the active site.

In the case of the N37T↑V^{+10}•D25A mutant, molecular dynamics simulations provided insights into its dynamics and motion, particularly within the flaps and hinge regions. The observed increased mobility suggests a potential alteration in the conformational dynamics of the flaps. Furthermore, the simulations indicated an increased curling of the flap tips, indicating possible changes in the overall behavior of the flaps. These findings suggest that the structural alterations and increased dynamics of the mutant could have implications for the hydrophobic sliding mechanism. Additionally, the decreased compactness and increased solvent exposure observed in the N37T↑V^{+10}•D25A mutant may lead to modifications in the hydrophobic environment surrounding the flaps. Hydrophobic residues within the flaps and their surrounding regions play a crucial role in maintaining stability and facilitating hydrophobic interactions. The altered structural dynamics and solvent exposure in the mutant could disrupt these hydrophobic interactions, potentially affecting the proper opening and closing of the flaps during substrate binding. However, it is important to acknowledge that the specific impact of the N37T↑V^{+10}•D25A mutant on the hydrophobic sliding

mechanism requires further investigation. Detailed analyses focusing on specific hydrophobic residues and their interactions would provide a more precise understanding of how the mutant influences the hydrophobic sliding mechanism in HIV-1 protease. These future studies would contribute to our comprehension of the functional consequences of the mutant and aid in the design of targeted strategies to counteract drug resistance in HIV-1 protease.

6.9 General conclusion

Based on the extensive analysis conducted in this study, it is evident that the hinge-region mutations, specifically in the N37T↑V^{+10}•D25A mutant of HIV-1 protease, have significant implications for the protein's structure, stability, and dynamics. The investigation of secondary, tertiary, and quaternary structures revealed that the mutations did not have an observable impact on the overall structure of the protease. However, further exploration into the conformational stability highlighted a reduced stability in the N37T↑V^{+10}•D25A mutant, indicating potential alterations in molecular interactions and inter/intra-interactions. Furthermore, the study delved into the unfolding of secondary and tertiary structures that revealed the cooperative unfolding behaviour of the PR further emphasising the stability and compactness of the secondary structure. Notably, the unfolding of the tertiary structure occurred at a lower urea concentration range compared to the secondary structure, suggesting that the secondary structure is more resilient to unfolding.

Examining the implications of these findings, it became apparent that the hinge-region mutations in the N37T↑V^{+10}•D25A mutant could potentially disrupt the hydrophobic sliding mechanism, a vital process involved in the opening of the flaps of HIV-1 protease. Molecular dynamics simulations revealed increased dynamics and structural alterations, particularly within the flaps and hinge regions of the mutant. These changes in dynamics, along with decreased compactness and increased solvent exposure, could modify the hydrophobic environment surrounding the flaps, ultimately affecting their proper opening and closing during substrate binding.

In conclusion, this thesis provides valuable insights into the structural and functional consequences of hinge-region mutations in HIV-1 protease. The observed alterations in conformational stability, dynamics, and potential impact on the hydrophobic sliding mechanism highlight the significance of these mutations in drug resistance and overall protease activity. Further investigations into specific hydrophobic residues and their interactions are warranted to gain a deeper understanding of the mutant's influence on the hydrophobic sliding mechanism and to facilitate the development of targeted strategies against drug resistance in HIV-1 protease. Ultimately, this research contributes to the broader field of protein folding, stability, and therapeutic interventions in combating HIV-1 infection.

Supplementary:

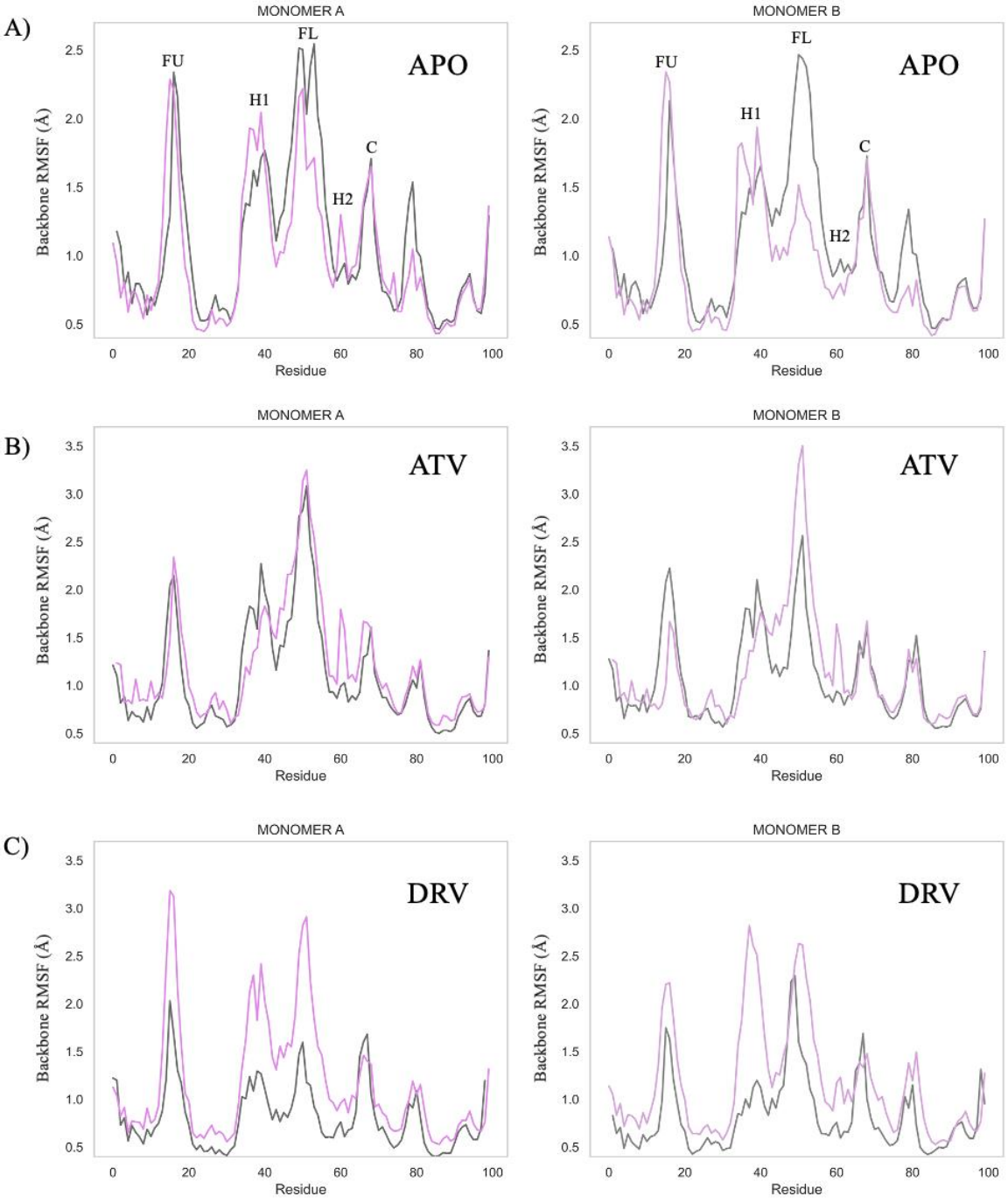


Figure S1: Backbone root mean square fluctuation (RMSF) plots per monomer (A/B) of WT PR (purple) and N37T↑V^{+10} PR (grey) over the course of the 1000 ns simulation.

A) Apo proteases. For reference the peaks in RMSF have been labelled according to the corresponding PR region (FU = fulcrum, H1 = hinge, H2 = hinge continued, FL = flap, C = cantilever). **B**) bound to atazanavir (ATV) and, **C**) bound to darunavir (DRV).

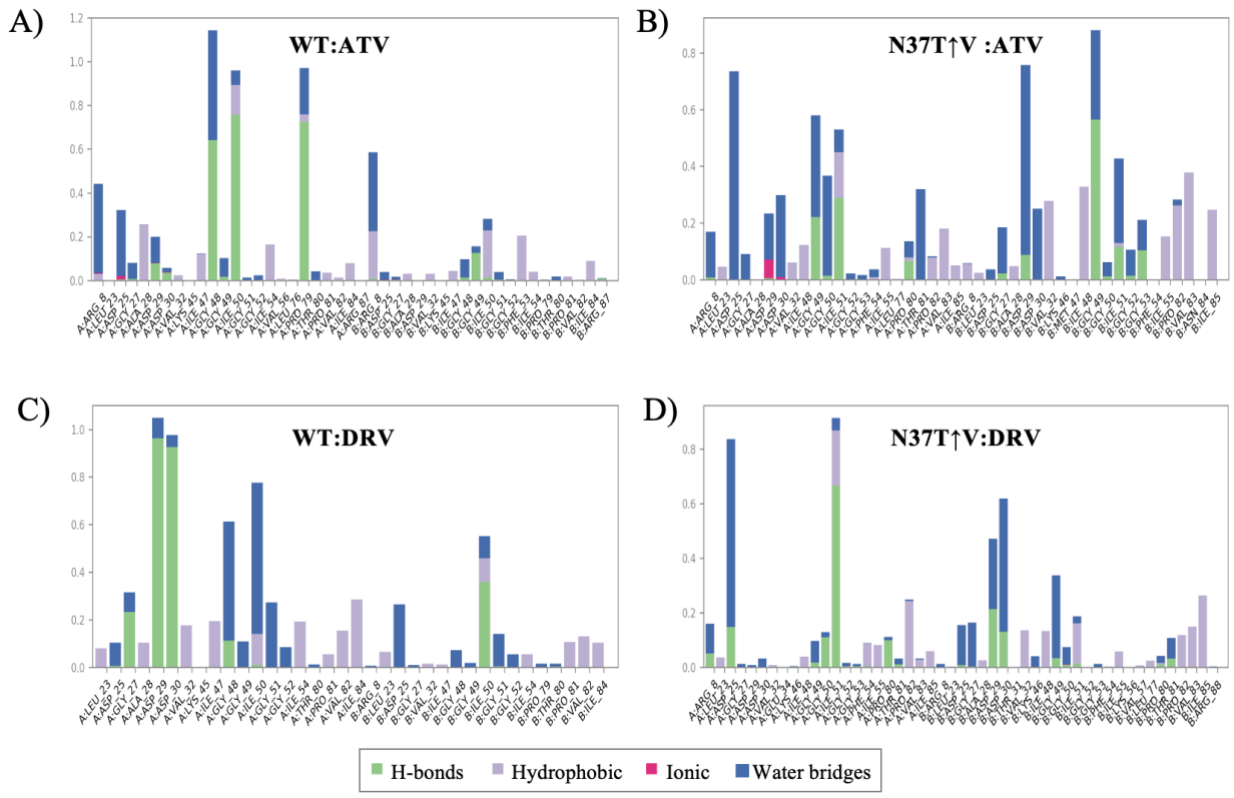


Figure S2: Types of intermolecular interactions formed between the PR:PI complexes over the course of the 1000 ns molecular dynamics simulations.

A) WT PR:ATV complex, **B**) N37T↑V^{+10} PR:ATV complex, **C)** WT PR:DRV complex, **D**) N37T↑V^{+10} PR:ATV complex.

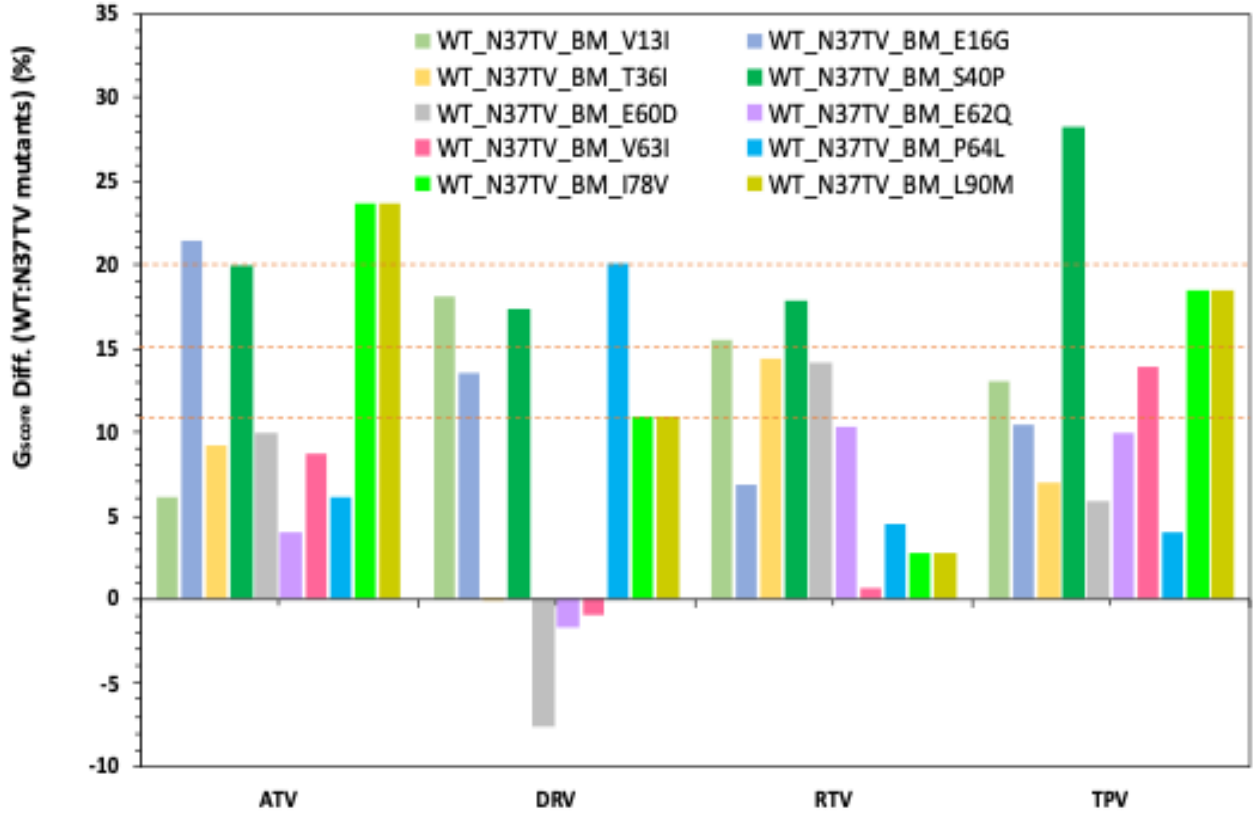


Figure S3: Difference in G_{score} of PIs as a result of systematically reverting individual background mutations present in N37T↑V^{+10} PR.

The effect of reverting each of the 10 background mutations present in N37T↑V^{+10} is represented as a percentage (%) change in G_{score} (N37T↑V:N37T↑V^{+10}_mutant). Here, a positive value is indicative of increased predicted binding affinity upon reversion.

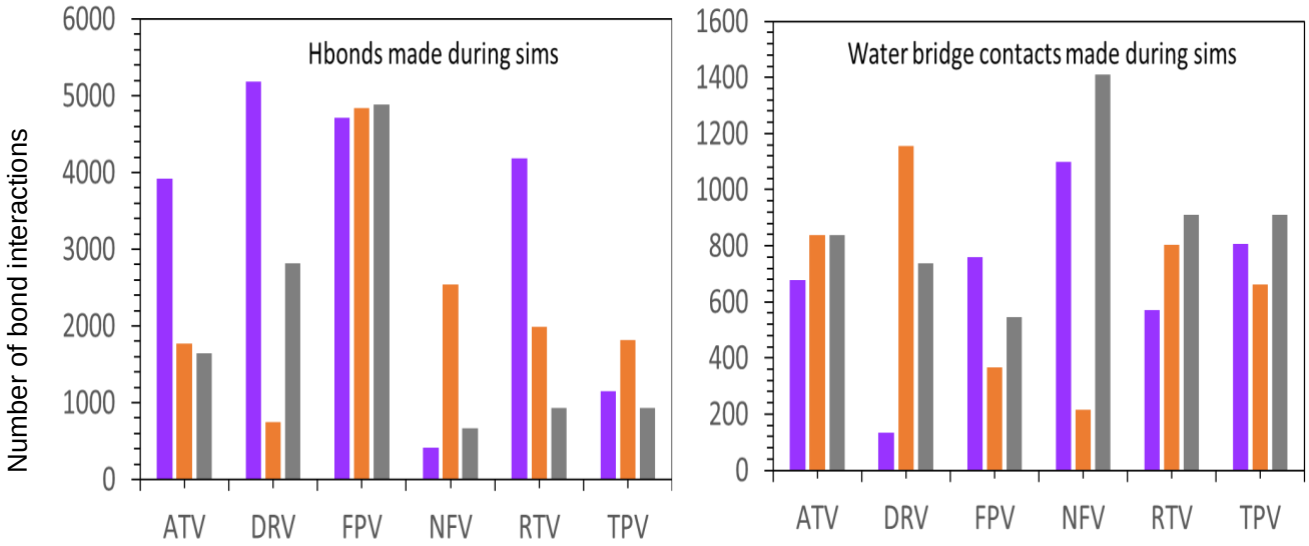


Figure S4: Total number of intermolecular interactions formed between WT PR (purple), N37T↑V PR (orange) and N37T↑V^{+10} PR (grey) during molecular dynamics simulations.

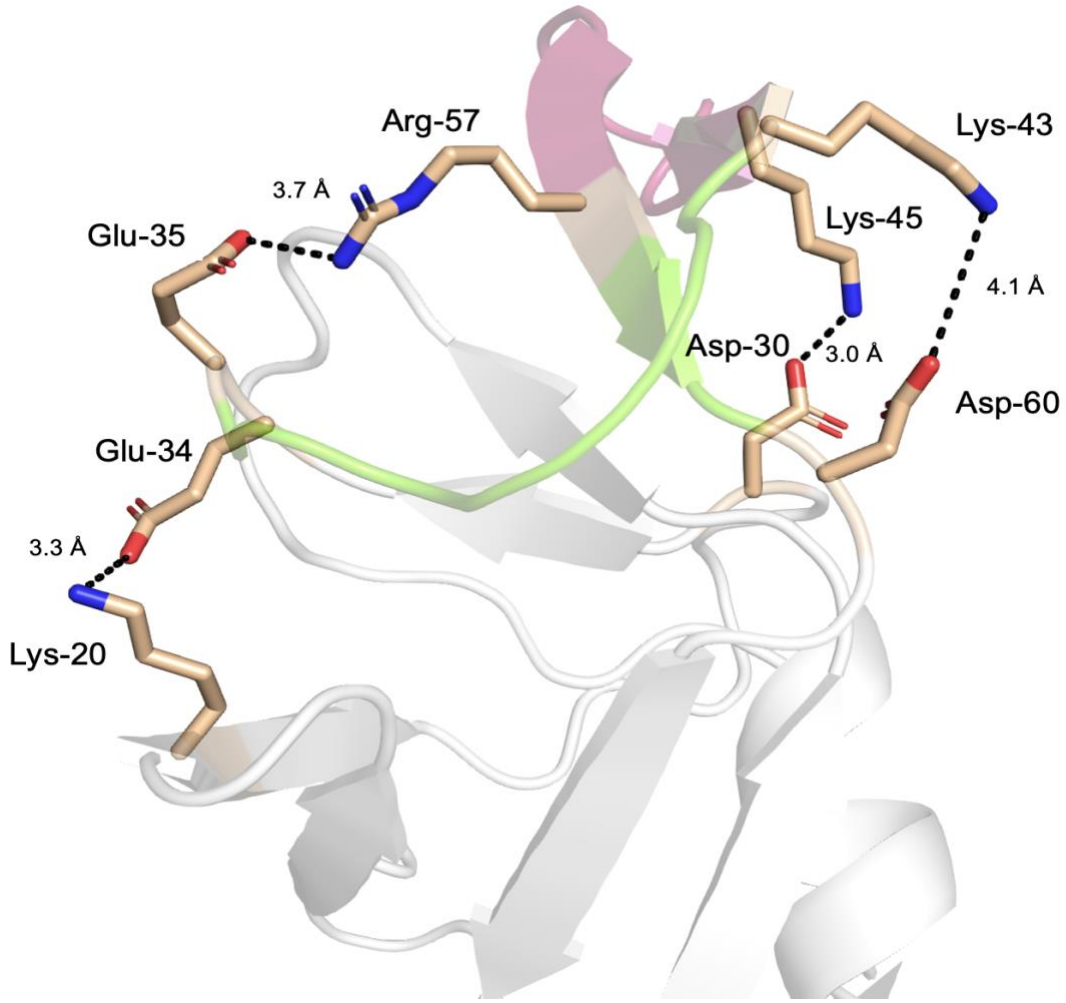


Figure S5: Intramolecular salt bridges of the HIV-1 PR.

The distance (Å) between the charged atoms of each amino acid is indicated. The hinge region (residues 35-42 & 57-61) and flap region (residues 43-56) are coloured green and red, respectively.

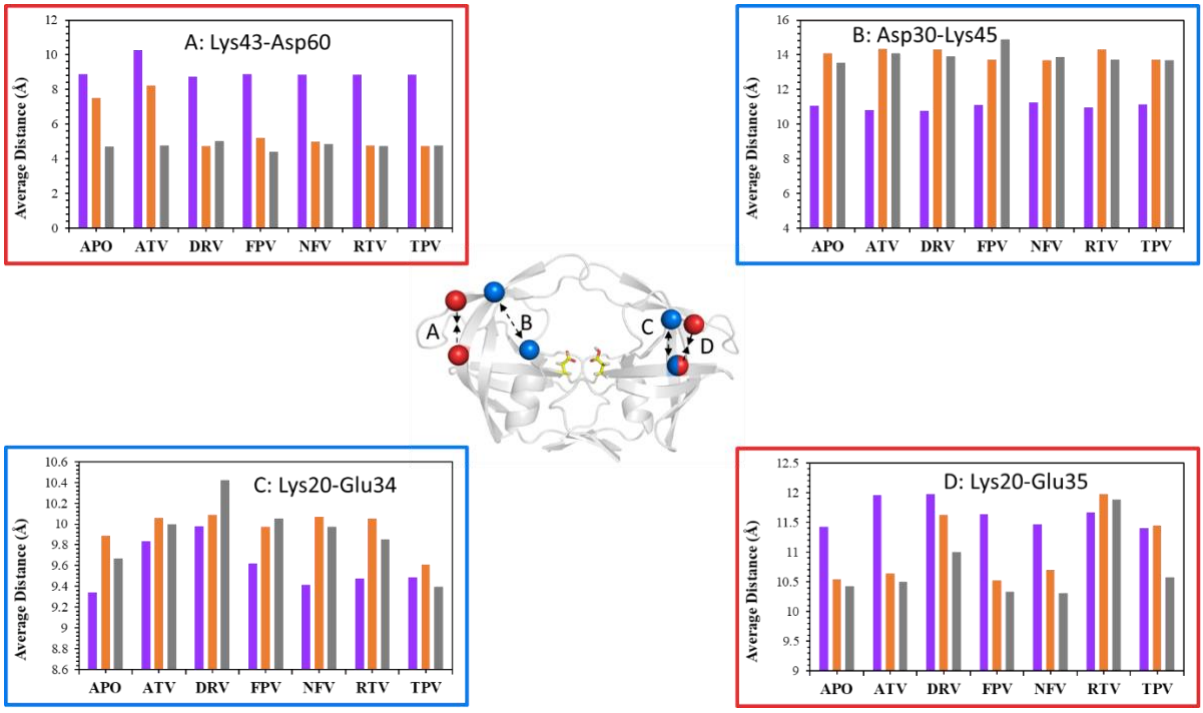


Figure S6: Interatomic distances between salt bridges of WT PR (purple), N37T↑V PR (orange) and N37T↑V+10 PR (grey). A) Lys43/Asp60, B) Asp30/Lys45, C) Lys20/Glu34 and D) Lys20/Glu35.

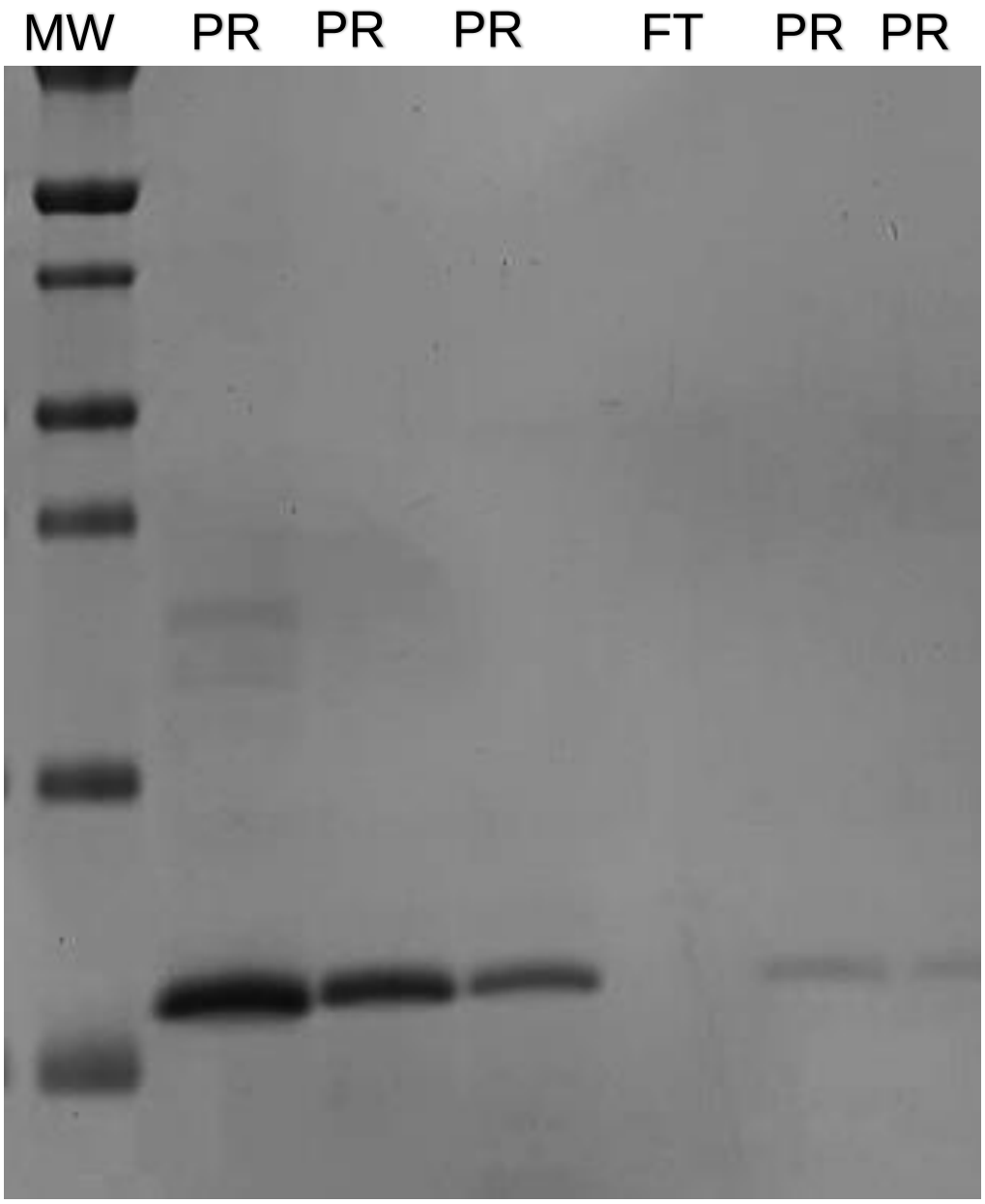
The measurements were performed for the apo PR variants as well as when in complex with ATV, DRV, FPV, NFV, RTV and TPV.

Table S1: The per region RMSF of the Apo and drug bound WT PR, N37T↑V, and N37T↑V+10 PR.

		PR variant		
Complex type	**PR region**	**HIV-1 WT**	**N37T↑V**	**N37T↑V^{+10}**
Apo	Fulcrum	0.89	1.12	0.99
	Flap	1.83	2.06	1.33
	Hinge	1.16	1.55	1.37
	Hinge	0.77	0.89	0.70
	Cantilever	0.96	1.03	0.87
	Catalytic	0.52	0.52	0.53
ATV	Fulcrum	0.82	1.36	0.86
	Flap	1.03	2.39	1.72
	Hinge	0.97	2.44	1.41
	Hinge	0.64	1.13	0.79
	Cantilever	0.83	1.14	0.81
	Catalytic	0.42	0.83	0.59
DRV	Fulcrum	0.84	0.83	0.85
	Flap	0.97	1.78	1.52
	Hinge	0.95	1.26	1.61
	Hinge	0.61	0.80	0.75
	Cantilever	0.78	0.86	0.84
	Catalytic	0.41	0.56	0.48
FPV	Fulcrum	0.79	1.24	1.03
	Flap	1.53	1.59	3.06
	Hinge	1.06	1.47	1.74
	Hinge	0.72	0.74	0.90
	Cantilever	0.88	0.86	0.94
	Catalytic	0.51	0.46	0.63
NFV	Fulcrum	0.94	0.88	0.79
	Flap	1.63	1.34	1.60
	Hinge	1.17	1.38	1.33
	Hinge	0.76	0.69	0.69
	Cantilever	0.94	0.84	0.83
	Catalytic	0.51	0.43	0.44
RTV	Fulcrum	0.87	0.91	0.71
	Flap	1.16	1.76	1.38
	Hinge	1.04	1.63	1.30
	Hinge	0.67	0.94	0.64
	Cantilever	0.85	0.95	0.69
	Catalytic	0.44	0.67	0.46
TPV	Fulcrum	0.83	0.75	0.75
	Flap	1.64	1.22	1.57
	Hinge	1.11	1.15	1.40
	Hinge	0.78	0.62	0.73
	Cantilever	0.94	0.77	0.83
	Catalytic	0.58	0.42	0.56

Table S2: Summarised induced fit docking parameters for WT, N37T↑V+10 and N37T↑V PR with clinical inhibitors

	Protease inhibitor	Glipo	Ghbond	Gevdw	Gcoul	Gligand efficiency
	ATV	-4.826	-1.034	-69.508	-17.055	-0.846
		-5.372	-0.717	-70.801	-10.179	-0.767
		-4.258	-0.724	-67.047	-14.003	-0.731
	DRV	-2.702	-1.052	-40.025	-11.493	-0.642
		-2.408	-0.929	-41.058	-17.353	-0.677
		-3.288	-1.003	-45.424	-16.096	-0.762
	FPV	-2.74	-0.051	-36.303	-27.706	-0.493
		-2.873	-1.224	-48.455	-25.548	-0.646
		-2.771	-0.98	-37.31	-30.49	-0.573
	IDV	-4.079	-0.956	-56.735	-23.098	-0.701
		-6.398	-0.52	-64.813	-10.224	-0.726
		-4.09	-1.082	-53.135	-23.215	-0.765
	LPV	-4.991	-0.657	-60.923	-15.888	-0.865
		-5.021	-0.48	-53.889	-20.605	-0.883
		-4.606	-0.48	-52.859	-18.414	-0.817
	NFV	-3.433	-0.72	-43.516	-18.881	-0.747
		-3.985	-1.077	-45.964	-18.788	-0.843
		-3.609	-1.024	-40.181	-20.601	-0.801
	RTV	-6.534	-0.456	-76.311	-9.637	-0.899
		-6.31	-0.304	-69.691	-13.054	-0.694
		-5.13	-0.717	-71.402	-10.484	-0.796
	SQV	-6.265	-0.758	-68.888	-11.663	-0.915
		-6.595	-0.456	-68.509	-9.573	-0.91
		-5.464	-0.419	-59.976	-18.521	-0.913
	TPV	-4.875	-0.16	-57.497	-3.727	-0.795
		-4.982	-0.471	-60.237	-10.194	-0.88
		-4.309	-0.16	-55.881	-4.117	-0.71


MW
PR
PR
PR
FT
PR
PR

References

Ahmad, S., Kumar, V., Ramanand, K.B., Rao, N.M., 2012. Probing protein stability and proteolytic resistance by loop scanning: A comprehensive mutational analysis. Protein Sci. Publ. Protein Soc. 21, 433–446. https://doi.org/10.1002/pro.2029

Anderson, R.M., Medley, G.F., May, R.M., Johnson, A.M., 1986. A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. Math. Med. Biol. 3, 229–263. https://doi.org/10.1093/imammb/3.4.229

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. Science 181, 223–230. https://doi.org/10.1126/science.181.4096.223

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinforma. Oxf. Engl. 22, 195–201. https://doi.org/10.1093/bioinformatics/bti770

Arts, E.J., Hazuda, D.J., 2012. HIV-1 antiretroviral drug therapy. Cold Spring Harb. Perspect. Med. https://doi.org/10.1101/cshperspect.a007161

Bardi, J.S., Luque, I., Freire, E., 1997. Structure-based thermodynamic analysis of HIV-1 protease inhibitors. Biochemistry 36, 6588–6596. https://doi.org/10.1021/bi9701742

Barré-Sinoussi, F., 1996. HIV as the cause of AIDS. The Lancet 348, 31–35. https://doi.org/10.1016/S0140-6736(96)09058-7

Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220, 868–871. https://doi.org/10.1126/science.6189183

Barrie, K.A., Perez, E.E., Lamers, S.L., Farmerie, W.G., Dunn, B.M., Sleasman, J.W., Goodenow, M.M., 1996. Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. Virology 219, 407–416. https://doi.org/10.1006/viro.1996.0266

Bartlett, J.A., DeMasi, R., Quinn, J., Moxham, C., Rousseau, F., 2001. Overview of the effectiveness of triple combination therapy in antiretroviral-naive HIV-1 infected adults. AIDS Lond. Engl. 15, 1369–1377. https://doi.org/10.1097/00002030-200107270-00006

Bessong, P.O., 2008. Polymorphisms in HIV-1 subtype C proteases and the potential impact on protease inhibitors. Trop. Med. Int. Health 13, 144–151. https://doi.org/10.1111/j.1365-3156.2007.01984.x

Blond, A., Ennifar, E., Tisné, C., Micouin, L., 2014. The Design of RNA Binders: Targeting the HIV Replication Cycle as a Case Study. ChemMedChem 9, 1982–1996. https://doi.org/10.1002/cmdc.201402259

Bloom, J.D., Gong, L.I., Baltimore, D., 2010. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. Science 328, 1272–1275. https://doi.org/10.1126/science.1187816

Bossard, C., Schramm, B., Wanjala, S., Jain, L., Mucinya, G., Opollo, V., Wiesner, L., van Cutsem, G., Poulet, E., Szumilin, E., Ellman, T., Maman, D., 2021. High Prevalence of NRTI and NNRTI Drug Resistance Among ART-Experienced, Hospitalized Inpatients. JAIDS J. Acquir. Immune Defic. Syndr. 87, 883. https://doi.org/10.1097/QAI.0000000000002689

Briggs, J.A.G., Kräusslich, H.-G., 2011. The molecular architecture of HIV. J. Mol. Biol. 410, 491–500. https://doi.org/10.1016/j.jmb.2011.04.021

Callis, P.R., 1997. [7] 1La and 1Lb transitions of tryptophan: Applications of theory and experimental observations to fluorescence of proteins, in: Methods in Enzymology, Flourescence Spectroscopy. Academic Press, pp. 113–150. https://doi.org/10.1016/S0076-6879(97)78009-1

Carmel, A., Yaron, A., 1978. An Intramolecularly Quenched Fluorescent Tripeptide as a Fluorogenic Substrate of Angiotensin-I-Converting Enzyme and of Bacterial Dipeptidyl Carboxypeptidase. Eur. J. Biochem. 87, 265–273. https://doi.org/10.1111/j.1432-1033.1978.tb12375.x

Carmel, Amoes, Yaron, A., 1978. An Intramolecularly Quenched Fluorescent Tripeptide as a Fluorogenic Substrate of Angiotensin-I-Converting Enzyme and of Bacterial Dipeptidyl Carboxypeptidase. Eur. J. Biochem. 87, 265–273. https://doi.org/10.1111/j.1432-1033.1978.tb12375.x

Carter, J.D., Gonzales, E.G., Huang, X., Smith, A.N., de Vera, I.M.S., D'Amore, P.W., Rocca, J.R., Goodenow, M.M., Dunn, B.M., Fanucci, G.E., 2014. Effects of PRE and POST therapy drug-pressure selected mutations on HIV-1 protease conformational sampling. FEBS Lett. 588, 3123–3128. https://doi.org/10.1016/j.febslet.2014.06.051

Chen, J., Toptygin, D., Brand, L., King, J., 2008. Mechanism of the efficient tryptophan fluorescence quenching in human gammaD-crystallin studied by time-resolved fluorescence. Biochemistry 47, 10705–10721. https://doi.org/10.1021/bi800499k

Clavel, F., Hance, A.J., 2004. HIV drug resistance. N. Engl. J. Med. 350, 1023–1035. https://doi.org/10.1056/NEJMra025195

Clemente, J.C., Moose, R.E., Hemrajani, R., Whitford, L.R.S., Govindasamy, L., Reutzel, R., McKenna, R., Agbandje-McKenna, M., Goodenow, M.M., Dunn, B.M., 2004. Comparing the accumulation of active- and nonactive-site mutations in the HIV-1 protease. Biochemistry 43, 12141–12151. https://doi.org/10.1021/bi049459m

Collier, A.C., Coombs, R.W., Schoenfeld, D.A., Bassett, R.L., Timpone, J., Baruch, A., Jones, M., Facey, K., Whitacre, C., McAuliffe, V.J., Friedman, H.M., Merigan, T.C., Reichman, R.C., Hooper, C., Corey, L., 1996. Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. AIDS Clinical Trials Group. N. Engl. J. Med. 334, 1011–1017. https://doi.org/10.1056/NEJM199604183341602

Condra, J.H., Schleif, W.A., Blahy, O.M., Gabryelski, L.J., Graham, D.J., Quintero, J.C., Rhodes, A., Robbins, H.L., Roth, E., Shivaprakash, M., 1995. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. Nature 374, 569–571. https://doi.org/10.1038/374569a0

Cowgill, R.W., 1970. Fluorescence and protein structure: XVII. On the mechanism of peptide quenching. Biochim. Biophys. Acta BBA - Protein Struct. 200, 18–25. https://doi.org/10.1016/0005-2795(70)90038-3

Darke, P. L., Leu, C.T., Davis, L.J., Heimbach, J.C., Diehl, R.E., Hill, W.S., Dixon, R.A., Sigal, I.S., 1989. Human immunodeficiency virus protease. Bacterial expression and characterization of the purified aspartic protease. J. Biol. Chem. 264, 2307–2312.

Darunavir, 2021. aidsmap.com. URL: https://www.aidsmap.com/about-hiv/arv-background-information/darunavir

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

De Arellano, E.R., Soriano, V., Holguín, A., 2005. Genetic analysis of regulatory, promoter, and TAR regions of LTR sequences belonging to HIV type 1 Non-B subtypes. AIDS Res. Hum. Retroviruses 21, 949–954. https://doi.org/10.1089/aid.2005.21.949

de Oliveira, T., Engelbrecht, S., Janse van Rensburg, E., Gordon, M., Bishop, K., zur Megede, J., Barnett, S.W., Cassol, S., 2003. Variability at human immunodeficiency

virus type 1 subtype C protease cleavage sites: an indication of viral fitness? J. Virol. 77, 9422–9430. https://doi.org/10.1128/JVI.77.17.9422-9430.2003

Debouck, C., 1992. The HIV-1 protease as a therapeutic target for AIDS. AIDS Res. Hum. Retroviruses 8, 153–164. https://doi.org/10.1089/aid.1992.8.153

Debouck, C., Gorniak, J.G., Strickler, J.E., Meek, T.D., Metcalf, B.W., Rosenberg, M., 1987. Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the gag precursor. Proc. Natl. Acad. Sci. 84, 8903–8906. https://doi.org/10.1073/pnas.84.24.8903

Demchenko, A.P., 2013. Ultraviolet Spectroscopy of Proteins. Springer Science & Business Media.

Detels, R., Muñoz, A., McFarlane, G., Kingsley, L.A., Margolick, J.B., Giorgi, J., Schrager, L.K., Phair, J.P., 1998. Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV infection duration. Multicenter AIDS Cohort Study Investigators. JAMA 280, 1497–1503. https://doi.org/10.1001/jama.280.17.1497

Dunn, B.M., 2007. HIV-protease, in: Enna, S.J., Bylund, D.B. (Eds.), XPharm: The comprehensive pharmacology reference. Elsevier, New York, pp. 1–6. https://doi.org/10.1016/B978-008055232-3.60560-4

Eche, S., Kumar, A., Sonela, N., Gordon, M.L., 2021. Acquired HIV-1 Protease conformational flexibility associated with lopinavir failure may shape the outcome of darunavir therapy after antiretroviral therapy switch. Biomolecules 11, 489. https://doi.org/10.3390/biom11040489

Erickson-Viitanen, S., Manfredi, J., Viitanen, P., Tribe, D.E., Tritch, R., Hutchison, C.A., Loeb, D.D., Swanstrom, R., 1989. Cleavage of HIV-1 gag polyprotein synthesized in vitro: Sequential cleavage by the viral protease. AIDS Res. Hum. Retroviruses 5, 577–591. https://doi.org/10.1089/aid.1989.5.577

Esparza, J., Bhamarapravati, N., 2000. Accelerating the development and future availability of HIV-1 vaccines: why, when, where, and how? The Lancet 355, 2061–2066. https://doi.org/10.1016/S0140-6736(00)02360-6

Faust, T.B., Binning, J.M., Gross, J.D., Frankel, A.D., 2017. Making sense of multifunctional proteins: human immunodeficiency virus type 1 accessory and regulatory proteins and connections to transcription. Annu. Rev. Virol. 4, 241–260. https://doi.org/10.1146/annurev-virology-101416-041654

Ferguson, M.R., Rojo, D.R., Lindern, J.J. von, O'Brien, W.A., 2002. HIV-1 replication cycle. Clin. Lab. Med. 22, 611–635. https://doi.org/10.1016/S0272-2712(02)00015-X

Fidy, J., Laberge, M., Ullrich, B., Polgar, L., Szeltner, Z., Gallay, J., Vincent, M., 2001. Tryptophan rotamers that report the conformational dynamics of proteins: Pure Appl. Chem. 73, 415–419. https://doi.org/10.1351/pac200173030415

Foulkes-Murzycki, J.E., Scott, W.R.P., Schiffer, C.A., 2007. Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. Structure 15, 225–233. https://doi.org/10.1016/j.str.2007.01.006

Fox, M.S., Hotchkiss, R.D., 1957. Initiation of bacterial transformation. Nature 179, 1322–1325. https://doi.org/10.1038/1791322a0

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

Freed, E.O., 2001. HIV-1 Replication. Somat. Cell Mol. Genet. 26, 13–33. https://doi.org/10.1023/A:1021070512287

Gallo, R.C., Montagnier, L., 1988. AIDS in 1988. Sci. Am. 259, 40–51.

Geddes, C.D. (Ed.), 2010. Reviews in fluorescence 2008, reviews in fluorescence. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1260-2

Gerritsen, M.E., Williams, A.J., Neish, A.S., Moore, S., Shi, Y., Collins, T., 1997. CREB-binding protein/p300 are transcriptional coactivators of p65. Proc. Natl. Acad. Sci. U. S. A. 94, 2927–2932.

Goldfarb, N.E., Ohanessian, M., Biswas, S., McGee, T.D., Mahon, B.P., Ostrov, D.A., Garcia, J., Tang, Y., McKenna, R., Roitberg, A., Dunn, B.M., 2015. Defective hydrophobic sliding mechanism and active site expansion in HIV-1 protease drug resistant variant Gly48Thr/Leu89Met: mechanisms for the loss of saquinavir binding potency. Biochemistry 54, 422–433. https://doi.org/10.1021/bi501088e

Haase, A.T., 1986. Pathogenesis of lentivirus infections. Nature 322, 130–136. https://doi.org/10.1038/322130a0

HIV Databases. URL: https://www.hiv.lanl.gov/content/index

Holec, A.D., Mandal, S., Prathipati, P.K., Destache, C.J., 2017. Nucleotide Reverse Transcriptase Inhibitors: A Thorough Review, Present Status and Future Perspective as HIV Therapeutics. Curr. HIV Res. 15, 411–421. https://doi.org/10.2174/1570162X15666171120110145

Hu, D.J., Dondero, T.J., Rayfield, M.A., George, J.R., Schochetman, G., Jaffe, H.W., Luo, C.C., Kalish, M.L., Weniger, B.G., Pau, C.P., Schable, C.A., Curran, J.W., 1996. The emerging genetic diversity of HIV. The importance of global surveillance for diagnostics, research, and prevention. JAMA 275, 210–216.

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., Fanucci, G.E., 2014. The Role of Select Subtype Polymorphisms on HIV-1 Protease Conformational Sampling and Dynamics. J. Biol. Chem. 289, 17203–17214. https://doi.org/10.1074/jbc.M114.571836

Ido, E., Han, H.P., Kezdy, F.J., Tang, J., 1991. Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S. J. Biol. Chem. 266, 24359–24366.

https://doi.org/10.1016/S0021-9258(18)54237X

Ishima, R., Freedberg, D.I., Wang, Y.X., Louis, J.M., Torchia, D.A., 1999. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. Struct. Lond. Engl. 1993 7, 1047–1055. https://doi.org/10.1016/s0969-2126(99)80172-5

Jaenicke, R., 1984. Protein Folding and Protein Association. Angew. Chem. Int. Ed. Engl. 23, 395–413. https://doi.org/10.1002/anie.198403953

Jelesarov, I., Bosshard, H.R., 1999. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. J. Mol. Recognit. 12, 3–18. https://doi.org/10.1002/(SICI)1099-1352(199901/02)12:1<3::AID-JMR441>3.0.CO;2-6

K.K. Reddy, G.S., Ali, A., Nalam, M.N.L., Anjum, S.G., Cao, H., Nathans, R.S., Schiffer, C.A., Rana, T.M., 2007. Design and Synthesis of HIV-1 Protease Inhibitors Incorporating Oxazolidinones as P2/P2′ Ligands in Pseudosymmetric Dipeptide Isosteres. J. Med. Chem. 50, 10.1021/jm070284z. https://doi.org/10.1021/jm070284z

Kirchhoff, F., 2013. HIV Life Cycle: Overview, in: Hope, T.J., Stevenson, M., Richman, D. (Eds.), Encyclopedia of AIDS. Springer New York, New York, NY, pp. 1–9. https://doi.org/10.1007/978-1-4614-9610-6_60-1

Koch, N., Ndihokubwayo, J.-B., Yahi, N., Tourres, C., Fantini, J., Tamalet, C., 2001. Genetic Analysis of HIV Type 1 Strains in Bujumbura (Burundi): Predominance of Subtype C Variant. AIDS Res. Hum. Retroviruses 17, 269–273. https://doi.org/10.1089/088922201750063205

Konvalinka, J., Litterst, M.A., Welker, R., Kottler, H., Rippmann, F., Heuser, A.-M., Usslich, H.-G.K., 1995. An active-site mutation in the human immunodeficiency virus type 1 proteinase (PR) causes reduced PR activity and loss of PR-mediated cytotoxicity without apparent effect on virus maturation and infectivity. J VIROL 69, 7.

Korpimäki, T., Kurittu, J., Karp, M., 2003. Surprisingly fast disappearance of β-lactam selection pressure in cultivation as detected with novel biosensing approaches. J. Microbiol. Methods 53, 37–42. https://doi.org/10.1016/s0167-7012(02)00213-0

Kräusslich, H.G., Ingraham, R.H., Skoog, M.T., Wimmer, E., Pallai, P.V., Carter, C.A., 1989. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. Proc. Natl. Acad. Sci. U. S. A. 86, 807–811.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685. https://doi.org/10.1038/227680a0

Lakowicz, J.R. (Ed.), 2006. Instrumentation for fluorescence spectroscopy, in: principles of fluorescence spectroscopy. Springer US, Boston, MA, pp. 27–61. https://doi.org/10.1007/978-0-387-46312-4_2

Lakowicz, J.R., 1999. Introduction to fluorescence, in: Lakowicz, j.r. (ed.), principles of fluorescence spectroscopy. Springer US, Boston, MA, pp. 1–23. https://doi.org/10.1007/978-1-4757-3061-6_1

Lama, J., Carrasco, L., 1992. Expression of poliovirus nonstructural proteins in Escherichia coli cells. Modification of membrane permeability induced by 2B and 3A. J. Biol. Chem. 267, 15932–15937. https://doi.org/10.1016/S0021-9258(19)49623-3

Landmarks of the HIV genome
https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark

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

Leavitt, S., Freire, E., 2001. Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr. Opin. Struct. Biol. 11: 560-566.

Li, G., Piampongsant, S., Faria, N.R., Voet, A., Pineda-Peña, A.-C., Khouri, R., Lemey, P., Vandamme, A.-M., Theys, K., 2015. An integrated map of HIV genome-wide variation from a population perspective. Retrovirology 12, 18. https://doi.org/10.1186/s12977-015-0148-6

Lin, Y., Lin, X., Hong, L., Foundling, S., Heinrikson, R.L., Thaisrivongs, S., Leelamanit, W., Raterman, D., Shah, M., 1995. Effect of point mutations on the kinetics and the inhibition of human immunodeficiency virus type 1 protease: relationship to drug resistance. Biochemistry 34, 1143–1152. https://doi.org/10.1021/bi00004a007

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

Liu, Z., Huang, X., Hu, L., Pham, L., Poole, K.M., Tang, Y., Mahon, B.P., Tang, W., Li, K., Goldfarb, N.E., Dunn, B.M., McKenna, R., Fanucci, G.E., 2016. Effects of Hinge-region Natural Polymorphisms on Human Immunodeficiency Virus-Type 1 Protease Structure, Dynamics, and Drug Pressure Evolution. J. Biol. Chem. 291, 22741–22756. https://doi.org/10.1074/jbc.M116.747568

Lv, Z., Chu, Y., Wang, Y., 2015. HIV protease inhibitors: a review of molecular selectivity and toxicity. HIVAIDS Auckl. NZ 7, 95–104. https://doi.org/10.2147/HIV.S79956

Mager, P.P., 2001. The active site of HIV-1 protease. Med. Res. Rev. 21, 348–353. https://doi.org/10.1002/med.1012

Maseko, S., Padayachee, E., Maphumulo, S., Govender, T., Sayed, Y., Maguire, G., Lin, J., Naicker, T., Baijnath, S., Gerhardus, K.H., 2019. Kinetic and thermodynamic characterisation of HIV-protease inhibitors against E35D↑G↑S mutant in the South

African HIV-1 subtype C protease. J. Enzyme Inhib. Med. Chem. 34, 1451–1456. https://doi.org/10.1080/14756366.2019.1636234

Maseko, S.B., Govender, D., Govender, T., Naicker, T., Lin, J., Maguire, G.E.M., Kruger, H.G., 2019. Optimized Procedure for Recovering HIV-1 Protease (C-SA) from Inclusion Bodies. Protein J. 38, 30–36. https://doi.org/10.1007/s10930-018-9805-7

Maseko, S.B., Natarajan, S., Sharma, V., Bhattacharyya, N., Govender, T., Sayed, Y., Maguire, G.E.M., Lin, J., Kruger, H.G., 2016. Purification and characterization of naturally occurring HIV-1 (South African subtype C) protease mutants from inclusion bodies. Protein Expr. Purif. 122, 90–96. https://doi.org/10.1016/j.pep.2016.02.013

Mata-Munguía, C., Escoto-Delgadillo, M., Torres-Mendoza, B., Flores-Soto, M., Vázquez-Torres, M., Gálvez-Gastelum, F., Viniegra-Osorio, A., Castillero-Manzano, M., Vázquez-Valls, E., 2014. Natural polymorphisms and unusual mutations in HIV-1 protease with potential antiretroviral resistance: a bioinformatic analysis. BMC Bioinformatics 15, 72. https://doi.org/10.1186/1471-2105-15-72

McCutchan, F.E., 2006. Global epidemiology of HIV. J. Med. Virol. 78, S7–S12. https://doi.org/10.1002/jmv.20599

Meek, T.D., Dayton, B.D., Metcalf, B.W., Dreyer, G.B., Strickler, J.E., Gorniak, J.G., Rosenberg, M., Moore, M.L., Magaard, V.W., Debouck, C., 1989. Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease. Proc. Natl. Acad. Sci. U. S. A. 86, 1841–1845.

Meher, B.R., Patel, S., 2013. Chapter Eight - Structural and Dynamical Aspects of HIV-1 Protease and Its Role in Drug Resistance, in: Karabencheva-Christova, T. (Ed.), Advances in Protein Chemistry and Structural Biology, Dynamics of Proteins and Nucleic Acids. Academic Press, pp. 299–324. https://doi.org/10.1016/B978-0-12-411636-8.00008-0

Mierendorf, R.C., Morris, B.B., Hammer, B., Novy, R.E., 1998. Expression and Purification of Recombinant Proteins Using the pET System. Methods Mol. Med. 13, 257–292. https://doi.org/10.1385/0-89603-485-2:257

Mittal, S., Cai, Y., Nalam, M.N., Bolon, D.N.A., Schiffer, C.A., 2012. Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. J. Am. Chem. Soc. 134, 4163–4168. https://doi.org/10.1021/ja2095766

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

Muiño, P.L., Callis, P.R., 2009. Solvent effects on the fluorescence quenching of tryptophan by amides via electron transfer. Experimental and computational studies. J. Phys. Chem. B 113, 2572–2577. https://doi.org/10.1021/jp711513b

Murphy, E.L., Collier, A.C., Kalish, L.A., Assmann, S.F., Para, M.F., Flanigan, T.P., Kumar, P.N., Mintz, L., Wallach, F.R., Nemo, G.J., Viral Activation Transfusion Study Investigators, 2001. Highly active antiretroviral therapy decreases mortality and morbidity in patients with advanced HIV disease. Ann. Intern. Med. 135, 17–26. https://doi.org/10.7326/0003-4819-135-1-200107030-00005

Naghavi, M.H., Schwartz, S., Sonnerborg, A., Vahlne, A., 1999. Long Terminal Repeat Promoter/Enhancer Activity of Different Subtypes of HIV Type 1. AIDS Res. Hum. Retroviruses 15, 1293–1303. https://doi.org/10.1089/088922299310197

Naicker, P., Achilonu, I., Fanucchi, S., Fernandes, M., Ibrahim, M.A.A., Dirr, H.W., Soliman, M.E.S., Sayed, Y., 2013a. 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. https://doi.org/10.1080/07391102.2012.736774

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

Naicker, P., Seele, P., Dirr, H.W., Sayed, Y., 2013b. F99 is Critical for Dimerization and Activation of South African HIV-1 Subtype C Protease. Protein J. 32, 560–567. https://doi.org/10.1007/s10930-013-9517-y

Naicker, P., Stoychev, S., Dirr, H.W., Sayed, Y., 2014. Amide hydrogen exchange in HIV-1 subtype B and C proteases insights into reduced drug susceptibility and dimer stability. FEBS J. 281, 5395–5410. https://doi.org/10.1111/febs.13084

Neuman, A.W., Liotta, D.C., 2008. Human Immunodeficiency Viruses: Antiretroviral Agents, in: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), Encyclopedia of Virology (Third Edition). Academic Press, Oxford, pp. 505–517. https://doi.org/10.1016/B978-012374410-4.00429-5

Nguyen, H.-L.T., Nguyen, T.T., Vu, Q.T., Le, H.T., Pham, Y., Trinh, P.L., Bui, T.P., Phan, T.-N., 2015. An efficient procedure for the expression and purification of HIV-1 protease from inclusion bodies. Protein Expr. Purif. 116, 59–65. https://doi.org/10.1016/j.pep.2015.07.011

Nijhuis, M., van Maarseveen, N.M., Boucher, C.A.B., 2007a. HIV protease resistance and viral fitness. Curr. Opin. HIV AIDS 2, 108–115. https://doi.org/10.1097/COH.0b013e32801682f6

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., Boucher, C.A.B., 2007b. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. PLoS Med. 4, e36. https://doi.org/10.1371/journal.pmed.0040036

Noel, A.F., Bilsel, O., Kundu, A., Wu, Y., Zitzewitz, J.A., Matthews, C.R., 2009. The Folding Free Energy Surface of HIV-1 Protease: Insights into the Thermodynamic Basis for Resistance to Inhibitors. J. Mol. Biol. 387, 1002–1016. https://doi.org/10.1016/j.jmb.2008.12.061

Novagen(R), 2003. pET System Manual.

Nowakowski, A.B., Wobig, W.J., Petering, D.H., 2014. Native SDS-PAGE: High Resolution Electrophoretic Separation of Proteins With Retention of Native Properties Including Bound Metal Ions. Met. Integr. Biometal Sci. 6, 1068–1078. https://doi.org/10.1039/c4mt00033a

Ohtaka, H., Velázquez-Campoy, A., Xie, D., Freire, E., 2002. Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. Protein Sci. Publ. Protein Soc. 11, 1908–1916. https://doi.org/10.1110/ps.0206402

Okumura, H., Itoh, S.G., Okamoto, Y., 2007. Explicit symplectic integrators of molecular dynamics algorithms for rigid-body molecules in the canonical, isobaric-isothermal, and related ensembles. J. Chem. Phys. 126, 084103. https://doi.org/10.1063/1.2434972

Osysko, A.P., Muíño, P.L., 2011. Fluorescence quenching of tryptophan and tryptophanyl dipeptides in solution. J. Biophys. Chem. 2, 316–321. https://doi.org/10.4236/jbpc.2011.23036

Pace, C., Shirley, B., McNutt, M., Gajiwala, K., 1996. Forces Contributing to the Conformational Stability of Proteins. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 10, 75–83. https://doi.org/10.1096/fasebj.10.1.8566551

Palese, L.L., 2017. Conformations of the HIV-1 protease: A crystal structure data set analysis. Biochim. Biophys. Acta BBA - Proteins Proteomics 1865, 1416–1422. https://doi.org/10.1016/j.bbapap.2017.08.009

Paton, N.I., Kityo, Cissy, Thompson, Jennifer, Nankya, Immaculate, French, N., Matenga, J., Cloherty, G., Wyk, J. van, Norton, M., Lehrman, S., Lamba, P., Malik, K., Rooney, J., Snowden, W., Villacian, J., 2017. Nucleoside reverse-transcriptase inhibitor cross-resistance and outcomes from second-line antiretroviral therapy in the public health approach: an observational analysis within the randomised, open-label, EARNEST trial. Lancet HIV 4, e341–e348. https://doi.org/10.1016/S2352-3018(17)30065-6

Peng, C., Ho, B.K., Chang, T.W., Chang, N.T., 1989. Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. J. Virol. 63, 2550–2556.

Perkins, S.J., 1986. Protein volumes and hydration effects. Eur. J. Biochem. 157, 169–180. https://doi.org/10.1111/j.1432-1033.1986.tb09653.x

Perryman, A.L., Lin, J.H., McCammon, J.A., 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. Protein Sci. Publ. Protein Soc. 13, 1108–1123. https://doi.org/10.1110/ps.03468904

Prabakaran, P., Dimitrov, A.S., Fouts, T.R., Dimitrov, D.S., 2007. Structure and Function of the HIV Envelope Glycoprotein as Entry Mediator, Vaccine Immunogen, and Target for Inhibitors. Adv. Pharmacol. San Diego Calif 55, 33–97. https://doi.org/10.1016/S1054-3589(07)55002-7

Ragland, D.A., Nalivaika, E.A., Nalam, M.N.L., Prachanronarong, K.L., Cao, H., Bandaranayake, R.M., Cai, Y., Kurt-Yilmaz, N., Schiffer, C.A., 2014. Drug Resistance Conferred by Mutations Outside the Active Site through Alterations in the Dynamic and Structural Ensemble of HIV-1 Protease. J. Am. Chem. Soc. 136, 11956–11963. https://doi.org/10.1021/ja504096m

Protease. J. Am. Chem. Soc. 136, 11956–11963. https://doi.org/10.1021/ja504096m

Ratner, L., 1993. HIV life cycle and genetic approaches. Perspect. Drug Discov. Des. 1, 3–22. https://doi.org/10.1007/BF02171653

Rhee, S.Y., Taylor, J., Fessel, W.J., Kaufman, D., Towner, W., Troia, P., Ruane, P., Hellinger, J., Shirvani, V., Zolopa, A., Shafer, R.W., 2010. HIV-1 protease mutations and protease inhibitor cross-resistance. Antimicrob. Agents Chemother. 54, 4253–4261. https://doi.org/10.1128/AAC.00574-10

Robertson, D.L., Anderson, J.P., Bradac, J.A., Carr, J.K., Foley, B., Funkhouser, R.K., Gao, F., Hahn, B.H., Kalish, M.L., Kuiken, C., Learn, G.H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P.M., Wolinsky, S., Korber, B., 2000. HIV-1 nomenclature proposal. Science 288, 55–56. https://doi.org/10.1126/science.288.5463.55d

Rodenburg, C.M., Li, Y., Trask, S.A., Chen, Y., Decker, J., Robertson, D.L., Kalish, M.L., Shaw, G.M., Allen, S., Hahn, B.H., Gao, F., UNAIDS and NIAID Networks for HIV Isolation and Characterization, 2001. Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. AIDS Res. Hum. Retroviruses 17, 161–168. https://doi.org/10.1089/08892220150217247

Rodriguez, M.A., Shen, C., Ratner, D., Paranjape, R.S., Kulkarni, S.S., Chatterjee, R., Gupta, P., 2007. Genetic and functional characterization of the LTR of HIV-1 subtypes A and C circulating in India. AIDS Res. Hum. Retroviruses 23, 1428–1433. https://doi.org/10.1089/aid.2007.0152

Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in Escherichia coli: advances and challenges. Front. Microbiol. 5, 172. 1-17 https://doi.org/10.3389/fmicb.2014.00172

Sayer, J.M., Liu, F., Ishima, R., Weber, I.T., Louis, J.M., 2008. Effect of the Active Site D25N Mutation on the Structure, Stability, and Ligand Binding of the Mature HIV-1 Protease. J. Biol. Chem. 283, 13459–13470. https://doi.org/10.1074/jbc.M708506200

Schägger, H., von Jagow, G., 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. 166, 368–379. https://doi.org/10.1016/0003-2697(87)90587-2

Schechter, I., Berger, A., 1967. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. 27, 157–162. https://doi.org/10.1016/S0006-291X(67)80055-X

Scott, W.R.P., 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. https://doi.org/10.1016/S0969-2126(00)00537-2

Shafer, R.W., Schapiro, J.M., 2008. HIV-1 Drug Resistance Mutations: an Updated Framework for the Second Decade of HAART. AIDS Rev. 10, 67–84.

Sherry, D., Pandian, R., Sayed, Y., 2022. Non-active site mutations in the HIV protease: Diminished drug binding affinity is achieved through modulating the hydrophobic sliding mechanism. Int. J. Biol. Macromol. 217, 27–41. https://doi.org/10.1016/j.ijbiomac.2022.07.033

Sherry, D., Worth, R., Ismail, Z.S., Sayed, Y., 2021. Cantilever-centric mechanism of cooperative non-active site mutations in HIV protease: Implications for flap dynamics. J. Mol. Graph. Model. 106, 107931. https://doi.org/10.1016/j.jmgm.2021.107931

Sherry, D., Worth, R., Sayed, Y., 2020. Two-Step Preparation of Highly Pure, Soluble HIV Protease from Inclusion Bodies Recombinantly Expressed in Escherichia coli. Curr. Protoc. Protein Sci. 100, e106. https://doi.org/10.1002/cpps.106

Sindrewicz, P., Li, X., Yates, E.A., Turnbull, J.E., Lian, L.-Y., Yu, L.-G., 2019. Intrinsic tryptophan fluorescence spectroscopy reliably determines galectin-ligand interactions. Sci. Rep. 9, 11851. https://doi.org/10.1038/s41598-019-47658-8

Spinelli, S., Liu, Q.Z., Alzari, P.M., Hirel, P.H., Poljak, R.J., 1991. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie 73, 1391–1396. https://doi.org/10.1016/0300-9084(91)90169-2

Suguna, K, Padlan, E.A., Smith, C.W., Carlson, W.D., Davies, D.R., 1987. Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action. Proc. Natl. Acad. Sci. U. S. A. 84, 7009–7013.

The Cascade Collaboration. Survival after introduction of HAART in people with known duration of HIV-1 infection, 2000. The Lancet 355, 1158–1159. https://doi.org/10.1016/S0140-6736(00)02069-9

Taylor, B.S., Sobieszczyk, M.E., McCutchan, F.E., Hammer, S.M., 2008. The Challenge of HIV-1 Subtype Diversity. N. Engl. J. Med. 358, 1590–1602. https://doi.org/10.1056/NEJMra0706737

The Lancet, 2000. Survival after introduction of HAART in people with known duration of HIV-1 infection. The Lancet 355, 1158–1159. https://doi.org/10.1016/S0140-6736(00)02069-9

Dill., K.A., McCallum, J.L. The Protein-Folding Problem, 50 Years On. 2012. Science, 338, 1042-1046. https://doi.org/10.1126/science.121902

Thomson, M.M., Nájera, R., 2001. Travel and the Introduction of Human Immunodeficiency Virus Type 1 Non-B Subtype Genetic Forms into Western Countries. Clin. Infect. Dis. 32, 1732–1737. https://doi.org/10.1086/320764

Todd, M.J., Freire, E., 1999. The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. Proteins Struct. Funct. Bioinforma. 36, 147–156. https://doi.org/10.1002/(SICI)10970134(19990801)36:2<147::AIDPROT2>3.0.CO;23

Todd, M.J., Semo, N., Freire, E., 1998. The structural stability of the HIV-1 protease11Edited by P. E. Wright. J. Mol. Biol. 283, 475–488. https://doi.org/10.1006/jmbi.1998.2090

Tomasselli, A.G., Hui, J.O., Sawyer, T.K., Staples, D.J., Bannow, C., Reardon, I.M., Howe, W.J., DeCamp, D.L., Craik, C.S., Heinrikson, R.L., 1990. Specificity and inhibition of proteases from human immunodeficiency viruses 1 and 2. J. Biol. Chem. 265, 14675–14683. https://doi.org/10.1016/S0021-9258(18)77355-9

Torbeev, V.Yu., Kent, S.B.H., 2012. Ionization state of the catalytic dyad Asp25/25' in the HIV-1 protease: NMR studies of site-specifically 13C labelled HIV-1 protease prepared by total chemical synthesis. Org. Biomol. Chem. 10, 5887–5891. https://doi.org/10.1039/c2ob25569c

Tritch, R.J., Cheng, Y.E., Yin, F.H., Erickson-Viitanen, S., 1991. Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 gag polyprotein. J. Virol. 65, 922–930. https://doi.org/10.1128/JVI.65.2.922-930.1991

UNAIDS, 2022. Global HIV & AIDS statistics — Fact sheet. URL https://www.unaids.org/en/resources/fact-sheet

Using circular dichroism spectra to estimate protein secondary structure | Nature Protocols. URL https://www.nature.com/articles/nprot.2006.202

Velazquez-Campoy, A., Todd, M.J., Vega, S., Freire, E., 2001. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. Proc. Natl. Acad. Sci. 98, 6062–6067. https://doi.org/10.1073/pnas.111152698

Velazquez-Campoy, A., Vega, S., Fleming, E., Bacha, U., Sayed, Y., Dirr, H., Freire, E., 2003. Protease Inhibition in African Subtypes of HIV–1. AIDS Rev. 5, 165–71.

Velazquez-Campoy, A., Vega, S., 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. https://doi.org/10.1021/bi020160i

Venkatachalam, S., Murlidharan, N., Krishnan, S.R., Ramakrishnan, C., Setshedi, M., Pandian, R., Barh, D., Tiwari, S., Azevedo, V., Sayed, Y., Gromiha, M.M., 2023. Understanding Drug Resistance of Wild-Type and L38HL Insertion Mutant of HIV-1 C Protease to Saquinavir. Genes 14, 533. https://doi.org/10.3390/genes14020533

Vlahov, D., Galai, N., Safaeian, M., Galea, S., Kirk, G.D., Lucas, G.M., Sterling, T.R., 2005. Effectiveness of Highly Active Antiretroviral Therapy among Injection Drug Users with Late-Stage Human Immunodeficiency Virus Infection. Am. J. Epidemiol. 161, 999–1012. https://doi.org/10.1093/aje/kwi133

Voet, D., Voet, J.G., 2010. Biochemistry, 4th ed. Wiley, Hoboken.

Weber, I.T., Agniswamy, J., 2009. HIV-1 Protease: Structural Perspectives on Drug Resistance. Viruses 1, 1110–1136. https://doi.org/10.3390/v1031110

Weiss, R.A., 1993. How Does HIV Cause AIDS? Science 260, 1273–1279. https://doi.org/10.1126/science.8493571

Williams, A., Basson, A., Achilonu, I., Dirr, H.W., Morris, L., Sayed, Y., 2019. Double trouble? Gag in conjunction with double insert in HIV protease contributes to reduced DRV susceptibility. Biochem. J. 476, 375–384. https://doi.org/10.1042/BCJ20180692

Wilson, K., Walker, J., 2010. Principles and Techniques of Biochemistry and Molecular Biology. Cambridge University Press, Cambridge.

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

World Health Organization, 2022. Fact Sheet: HIV. https://www.who.int/news-room/fact-sheets/detail/hiv-aids

Zhou, H.-X., 2004. Loops, Linkages, Rings, Catenanes, Cages, and Crowders: Entropy-Based Strategies for Stabilizing Proteins. Acc. Chem. Res. 37, 123–130. https://doi.org/10.1021/ar0302282

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