Novel hinge insertions in South African HIV-1 subtype C protease:

implications for biochemical and biophysical properties

by

Zaahida Sheik Ismail

790012

A thesis

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

in

Molecular and Cell Biology

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

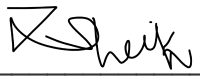
Supervisor: Professor Yasien Sayed

June 2022

DECLARATION

I, Zaahida Sheik Ismail, (Student number: 790012) am a student registered for the degree Doctor of Philosophy (PhD) in the year 2022. 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 ALL the work submitted for assessment for the above course is my own unaided work except where I have 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:  Date: 07 June 2022

RESEARCH OUTPUTS

Conferences:

Molecular Bioscience Research Thrust Research Day 2021:

- Oral presentation (2nd prize)
- Novel hinge region insertions in South African HIV-1 subtype C protease confer structural and functional changes affecting drug binding
- Zaahida Sheik Ismail
- University of the Witwatersrand, Johannesburg, South Africa
- 9 December 2021

South African Society of Biochemistry and Molecular Biology 2022:

- Oral presentation
- Effect of novel hinge insertions on drug binding and structurally stability of the South African HIV-1 subtype C protease
- Zaahida Sheik Ismail
- University of the Witwatersrand, Johannesburg, South Africa
- 24-27 January 2022

Publications:

Sherry, D., Worth, R., Sheik Ismail, Z. S. and Sayed, Y., (2021) “Cantilever-centric mechanism of cooperative non-active site mutations in HIV protease: Implications for flap dynamics.” *J Mol Graph Model*. doi: 10.1016/j.jmgm.2021.107931.

Submitted Manuscripts:

Sherry, D., Sheik Ismail, Z., Mokhantso, T. and Sayed, Y., (2022) “Chapter 12: HIV1 subtype C protease hinge region dynamics and flap flexibility in drug resistance.” *Proteolytic Enzymes and their Inhibitors in Infectious Pathogens*

Sheik Ismail, Z., Worth, R., Salerwe, M. and Sayed Y. “Amino Acid Insertions in the Hinge Region of HIV Protease Affect Enzyme Kinetics, Conformational Stability and Dynamics” *Biochemical Journal*

ABSTRACT

Although treatable, human-immunodeficiency virus infections remain distressingly high in sub-Saharan Africa, particularly due to the prevalence of the subtype C variant in this location. While advances have been made in tackling drug resistance globally, subtype C infections are majorly understudied specifically with regards to the effect of mutations in this subtype on drug resistance to currently available inhibitors. The HIV protease is responsible for the production of new infectious viral progenies. Cleavage of the Gag and Gag-Pol polyprotein precursors by the protease produces mature viral particles that can infect new host cells. The indispensable role of the HIV protease in the lifecycle of the virus makes it a major drug target for combatting this disease. Mutations in the protease accumulate both within and distal to the active site, often contributing to drug resistance. More recently, insertions within the hinge region of the protease have been identified. The effect of these hinge region insertions on the HIV protease is not fully understood. The study was based on a clinical isolate containing two insertions, Asn and Leu, at position 38 in the hinge region as well as four mutations: K20R, E35D, R57K and V82I (L38↑N↑L^{+4}). The aim of this study was to determine the direct effect of these hinge region insertions on the characteristics of the protease. As such, a variant subtype C protease containing only the double insertion of Asn and Leu was created (L38↑N↑L^{-4}). The variant L38↑N↑L^{-4} protease was successfully overexpressed and purified using a newly adapted ion-exchange chromatography purification protocol. Circular dichroism and size exclusion-high performance liquid chromatography experiments showed that the secondary and oligomeric structures of the variant protease was similar to the wild-type (WT). Active site titrations indicated that the purification of L38↑N↑L^{-4} resulted in a decrease of actively prepared sample in comparison to the WT. The specific activity and turnover number of L38↑N↑L^{-4} was significantly reduced in comparison to the WT while the catalytic efficiency was increased. Induced-fit docking studies showed that the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} do bind the protease inhibitors atazanavir (ATV), darunavir (DRV), ritonavir (RTV) and saquinavir (SQV) albeit with different affinities. *In vitro* displacement titration experiments confirmed that L38↑N↑L^{-4} does bind these inhibitors. The binding affinity of L38↑N↑L^{-4} to the protease inhibitors ATV, RTV and SQV was significantly reduced with changes to the overall binding energetics of these

reactions. Darunavir did not display any change in binding affinity to L38↑N↑L^{-4}; however, the overall energetics of this reaction was altered in comparison to the WT. In addition, the stability of these L38↑N↑L^{-4} complexes were significantly diminished implicating these insertions in drug binding interactions. Differential scanning calorimetry results showed that the transition temperature of the apo L38↑N↑L^{-4} protease was 5 °C higher than the apo WT protease indicating the role of these insertions in the stability of the protease. Molecular dynamic simulations confirmed the increased structural stability of the apo mutated proteases which exhibited more flexible hinge regions than the WT. Additionally, both hinge region mutants, L38↑N↑L^{-4} and L38↑N↑L^{+4}, exhibited a mainly closed conformation thus explaining the decreased catalytic properties of the protease. Furthermore, the presence of inhibitors altered the dynamics, conformations, and flexibility of the three proteases. The inhibitor bound mutated proteases, L38↑N↑L^{-4} and L38↑N↑L^{+4}, displayed more flexible flap and hinge regions in comparison to the WT. Except for SQV, the presence of inhibitors shifted the flap conformers of the mutated proteases from closed to semi-open as well. Further analysis indicated changes in the number of hydrophobic interactions, hydrogen bonding and water bridge formations between the proteases and the inhibitors. Altogether, these results implicate the hinge region insertions in drug binding affinities and interactions with protease inhibitors as well as structural stability, dynamics, and flexibility of the protease both in the absence and presence of protease inhibitors.

For my family –

أطلبوا العلم من المهد إلى اللحد

“Seek knowledge from the cradle to the grave”

Prophet Muhammed ﷺ

ACKNOWLEDGEMENTS

I would first and foremost like to thank the Almighty for blessing me with the opportunity to be able to contribute to the scientific community.

I would like to thank my supervisor, Prof. Yasien Sayed, for allowing me the opportunity and encouragement to grow and flourish into an independent, free-thinking scientist and human being. I wholeheartedly appreciate your willingness and confidence to let me explore the avenues of research. You are both a compassionate person and an inspirational mentor.

To my biggest supporter and soulmate - Raees van der Schyff – a very special thank you. Without you, this would not have been possible. Every single day you continuously challenge me to be a better version of myself and I am forever grateful. I look forward to achieving many more life goals together. I love you, bab.

A huge thank you to my moms, dads, and sisters. Your unwavering support, positive enforcements, and can-do attitudes have always provided comfort and encouragement when I needed it most.

Thank you to my friends and colleagues at the Protein-Structure Function Research Unit for the endless discussions and idea-bouncing conversations. Thank you to Dr. Sylvia Fanucchi and Dr. Ikechukwu Achilonu for always making themselves available if I needed a quick chat. A special thanks to our PhD French Toast group who managed to balance the science with loads of boardgames, breakfast dates and picnics.

Thank you to Dr. Salerwe Mosebi and Thato Manyaapelo from the University of South Africa for training and assisting me with the calorimetry experiments.

I am extremely grateful for the funding received from the National Research Foundation (NRF) without whom this degree would not have been possible.

CONTENTS

DECLARATION ii
ABSTRACT v
ACKNOWLEDGEMENTS viii
LIST OF FIGURES xii
LIST OF TABLES xiv
LIST OF ABBREVIATIONS xv
1 INTRODUCTION 19
1.1 Human immunodeficiency virus 19
1.1.1 HIV genome and life cycle 20
1.2 HIV/AIDS treatment 23
1.3 Gag and Gag-Pol Polyproteins 24
1.4 South African HIV-1 subtype C protease 25
1.4.1 Structure and function of protease 25
1.5 Catalytic mechanism of HIV-1 protease 30
1.6 Protease inhibitors 30
1.7 Mutations and drug resistance 34
1.7.1 Active site mutations 35
1.7.2 Non-active site mutations 35
1.7.3 Effects on drug binding thermodynamics 36
1.7.4 Effects on structure and stability 36
1.7.5 Effects on flap region 37
1.7.6 Insertion mutations 38
1.8 L38↑N↑L^{+4} 38
1.9 Aim and objectives 40
1.9.1 Aim 40
1.9.2 Objectives 40
2 METHODS 41
2.1 Bacterial Transformation 41
2.2 Expression of WT and L38↑N↑L^{-4} 43
2.2.1 Overexpression of the WT protease 44
2.2.2 Overexpression of L38↑N↑L^{-4} 44
2.3 Purification of WT and L38↑N↑L^{-4} 45

2.3.1 Purification of WT protease 45
2.3.2 Purification of L38↑N↑L^{-4} protease 47
2.4 Concentration and storage of purified proteins 51
2.5 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis 51
2.6 Protein concentration determination 52
2.7 Size exclusion-high pressure liquid chromatography 53
2.8 Far-UV circular dichroism spectropolarimetry 55
2.9 Protein crystallisation 56
2.10 Isothermal titration calorimetry 57
2.10.1 Active site concentration 57
2.10.2 Displacement titration calorimetry 59
2.11 Differential scanning calorimetry 60
2.12 Enzyme kinetics 60
2.13 Molecular dynamics simulations 62
2.14 Induced-fit docking studies 62
3 RESULTS 65
3.1 Confirmation of WT and L38↑N↑L^{-4} sequences 65
3.2 Overexpression of WT protease 66
3.3 Overexpression of L38↑N↑L^{-4} 66
3.3.1 Growth rates of L38↑N↑L^{-4} 66
3.3.2 Expression trials of L38↑N↑L^{-4} 69
3.3.3 Purification of the L38↑N↑L^{-4} protease 70
3.4 Determination of protein molecular mass 72
3.5 Protein concentration determination 72
3.5.1 Absorbance Spectra 72
3.5.2 Active site determination 75
3.6 Structural characterisation of WT and L38↑N↑L^{-4} 77
3.6.1 Secondary structure 77
3.6.2 Quaternary structure 78
3.6.3 Crystallisation 80
3.7 Isothermal titration calorimetry 82
3.8 Differential scanning calorimetry 86
3.9 Steady-state kinetics 88

3.10 Induced-fit docking studies 89
3.11 Molecular dynamics 91
3.11.1 Apo analysis 91
3.11.2 RMSF of protease inhibitor complexes 94
3.11.3 Ligand RMSF 97
3.11.4 Flap conformations 99
3.11.5 Radius of gyration and solvent accessibility of PIs 102
3.12 Protease inhibitor interactions 104
4 DISCUSSION 106
4.1 Purification of WT and L38↑N↑L^{-4} proteases 106
4.2 Structural characterisation of proteases 108
4.3 Effects of protease insertions on substrate binding 108
4.4 Effects of protease insertions on drug binding thermodynamics 111
4.5 Protease dimer stability 114
4.6 Molecular dynamics of apo proteases 117
4.7 Protein-ligand dynamics 119
4.8 Induced-fit docking studies 121
4.9 Binding interactions between proteases and inhibitors 123
4.10 Concluding remarks 130
4.11 Future Perspectives 131
5 SUPPLEMENTARY MATERIAL 132
Appendix 1. ProtParam of WT C-SA 132
Appendix 2. ProtParam of L38↑N↑L^{-4} sequence 134
6 REFERENCES 144

LIST OF FIGURES

Figure 1.1 Divisions of HIV-1. 20
Figure 1.2 HIV virion. 21
Figure 1.3 Life cycle of HIV-1. 22
Figure 1.4 Gag polyprotein of HIV. 24
Figure 1.5 Structure of HIV-1 protease. 26
Figure 1.6 Heterogenous flap conformations found in the HIV-1 protease. 28
Figure 1.7 Hydrophobic core residues of the HIV-1 protease. 29
Figure 1.8 Catalytic mechanism of the HIV-1 protease. 31
Figure 1.9. Three-dimensional structure of the nine clinically approved PIs. 32
Figure 1.10 Homology model of L38↑N↑L^{+4} protease. 39
Figure 2.1 Expression and purification of the WT protease. 48
Figure 2.2 Expression and purification of the L38↑N↑L^{-4} protease. 49
Figure 2.3 SE-HPLC procedure. 54
Figure 2.4 Basic principle of ITC. 58
Figure 2.5 Cleavage of fluorogenic substrate by the HIV-1 protease. 61
Figure 3.1 Sequencing results of WT (A) and L38↑N↑L^{-4} (B) plasmid inserts and alignment (C) with the C-SA (PDB:3U71) nucleotide sequence. 65
Figure 3.2 Cation exchange chromatography of WT protease. 67
Figure 3.3 Growth curves of *E. coli* Rosetta (A), T7 (B) and BL21 (DE3) pLysS (C) cell lines expressing the L38↑N↑L^{-4} protease. 68
Figure 3.4 Induction trials of the L38↑N↑L^{-4} protease in different *E. coli* cell lines. 69
Figure 3.5 Tricine (16%) SDS-PAGE gel of BL21 (DE3) pLysS cells expressing L38↑N↑L^{-4}. 70
Figure 3.6 Ion-exchange chromatography of the L38↑N↑L^{-4} protease. 71
Figure 3.7 Tricine (16%) SDS-PAGE standard curves used to calculate the molecular mass of the WT and L38↑N↑L^{-4} proteases. 73
Figure 3.8 Standard curves generated from absorbance spectrum readings for dilution series (1:10, 1:12, 1:14, 1:16 and 1:20). 74
Figure 3.9 Active site titrations of WT (A) and L38↑N↑L^{-4} (B) proteases. 76
Figure 3.10 Far-UV CD spectra of the WT (green) and L38↑N↑L^{-4} (purple) proteases. 78
Figure 3.11 SE-HPLC column calibration using standards of known molecular weights. 79
Figure 3.12 SE-HPLC elution profiles of WT (A) and L38↑N↑L^{-4} (B). 80
Figure 3.13 Image depicting trigonal bipyramidal WT crystals. 82
Figure 3.14 Dissociation binding constants for WT and L38↑N↑L^{-4} with ATV (A), DRV (B), RTV (C) and SQV (D). 83
Figure 3.15 Thermodynamic parameters (ΔG, ΔH and $T\Delta S$) of ATV (A), DRV (B), RTV (C) and SQV (D) binding to the WT and L38↑N↑L^{-4} proteases. 85
Figure 3.16 DSC thermograms of WT (green) and L38↑N↑L^{-4} (purple) proteases. 87

Figure 3.17 Thermal transition midpoints (T_m) of WT (green bars) and L38↑N↑L^{-4} (purple bars) for the apo forms and in the presence of ATV, DRV, RTV and SQV. . 88
Figure 3.18 IFDS of PIs binding to WT, L38↑N↑L^{-4} and L38↑N↑L^{+4}. 90
Figure 3.19 RMSD of apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} protease polypeptide backbones. 92
Figure 3.20 Radius of gyration of apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} protease polypeptide backbones. 93
Figure 3.21 RMSF of apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} regions with respect to amino acid residues. 95
Figure 3.22 L-RMSF ratio (mutant/WT) of PI atoms when in complex with L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases. 98
Figure 3.23 Flap conformations of WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases. 100
Figure 3.24 Flap conformations of WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} when bound to ATV (A), DRV (B), RTV (C) and SQV (D). 101
Figure 3.25 The average R_g of the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases when bound to ATV, DRV, RTV and SQV. 103
Figure 3.26 The average SASA of the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases when bound to ATV, DRV, RTV and SQV. 103
Figure 4.1 Hydrophobic amino acid side chains of the HIV-1 protease. 122
Figure 4.2. Amino acid interactions of proteases with ATV. 125
Figure 4.3 Amino acid interactions of proteases with DRV. 126
Figure 4.4 Amino acid interactions of proteases with RTV. 127
Figure 4.5 Amino acid interactions of proteases with SQV. 128

LIST OF TABLES

Table 1.1 HIV-1 protease mutations associated with inhibitor resistance..... 35
Table 2.1 Agar plates used for the selection of transformants. 42
Table 3.1 Nandrop values for WT and L38↑N↑L^{-4}. 75
Table 3.2 Corrected concentrations for WT and L38↑N↑L^{-4}. 77
Table 3.3 Hampton Research Index™ conditions that resulted in hits with the apo WT protease 81
Table 3.4 Hampton Research Index™ condition that resulted in hits with L38↑N↑L^{-4} bound to DRV..... 81
Table 3.5 Steady-state kinetics of the apo WT and L38↑N↑L^{-4} proteases. 89
Table 3.6 G_{score} values of three proteases in kcal/mol..... 90
Table 3.7 Average RMSD data for apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} collected over the course of 50 ns MD simulations. 92
Table 3.8 Average R_g values for apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} collected over the course of 50 ns MD simulations. 93
Table 3.9 Percentage change (increase/decrease) in average RMSF data for L38↑N↑L^{-4} and L38↑N↑L^{+4} with regards to WT during a 50 ns MD simulation..... 96
Table 3.10 Percentage change (increase/decrease) in average RMSF data for L38↑N↑L^{-4} and L38↑N↑L^{+4} regions with regards to WT during a 50 ns MD simulation. 96
Table 3.11 Major interactions formed between PIs and key residues in the active site and flaps of WT, L38↑N↑L^{-4} and L38↑N↑L^{+4}. 105
Table 4.1 Comparison of steady-state kinetic parameters. 109

LIST OF ABBREVIATIONS

°C	Temperature in degrees Celsius
µM	Concentration in micromolar
A_{280}	Absorbance at 280 nm
AIDS	Acquired immunodeficiency syndrome
AMP	Ampicillin
APV	Amprenavir
ARV	Antiretroviral
ASN	Asparagine
ASP25	Catalytic aspartic acid residue at position 25
ATV	Atazanavir
bis-THF	*bis*-tetrahydrofuranylurethane
CA	Capsid protein
CAM	Chloramphenicol
CD	Circular dichroism
CM	Carboxymethyl
CRFs	Circulating recombinant forms
C-SA	South African subtype C
DEAE	Diethylaminoethyl
DNA	Deoxyribonucleic acid
DNaseI	Bovine pancreatic deoxyribonuclease I
DRV	Darunavir

DSC	Differential scanning calorimetry
DTC	Displacement titration calorimetry
DTT	Dithiothreitol
E. coli	*Escherichia coli*
FDA	Food and drug administration
G_{score}	GlideScore
HAART	Highly active antiretroviral therapy
HAPA	Hydoxyaminopentane amide
HIV	Human immunodeficiency virus
HOD	Heats of dilution
IDV	Indinavir
IEC	Ion-exchange chromatography
IFDS	Induced-fit docking studies
ILE50	Flap tip isoleucine residue at position 50
IN	Integrase
INSTIs	Integrase strand transfer inhibitors
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
k_{cat}	Turnover number
k_{cat}/K_M	Catalytic efficiency
K_D	Dissociation binding constant
kDa	kilodaltons
LB	Luria-Bertani Broth

LEU	Leucine
LPV	Lopinavir
L-RMSF	Ligand root-mean-square fluctuation
MA	Matrix protein
MD	Molecular dynamics
mRNA	Messenger ribonucleic acid
MWCO	Molecular weight cut off
n	Stoichiometry
NC	Nucleocapsid protein
NFV	Nelfinavir
NNRTIs	Non-nucleoside reverse transcriptase inhibitors
NRTIs	Nucleoside reverse transcriptase inhibitors
OD_{600}	Optical density at 600 nm
pABA	*para*-aminobenzoic acid
PDB	Protein Data Bank
PIs	Protease inhibitors
PMSF	Phenylmethylsulfonyl fluoride
PR•PI	Protease inhibitor complex
R_g	Radius of gyration
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
RT	Reverse transcriptase
RTV	Ritonavir

SASA	Surface accessible solvent area
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SE-HPLC	Size exclusion-high pressure liquid chromatography
SOC	Super optimal broth with catabolite repression
SQV	Saquinavir
ssRNA	Single stranded ribonucleic acid
T_m	Transition temperature
TPV	Tipranavir
TΔS	Change in entropy
UV	Ultraviolet
WT	Wild-type
ΔC_p	Change in heat capacity
ΔG	Change in Gibbs free energy
ΔH	Change in enthalpy

1 INTRODUCTION

1.1 HUMAN IMMUNODEFICIENCY VIRUS

Acquired immunodeficiency syndrome (AIDS) is caused by the pathogenic Human Immunodeficiency Virus (HIV) and continues to remain a major global health concern. In 2020, 1.5 million new infections were reported with a total of 37.7 million people infected worldwide (UNAIDS, 2021). HIV is of particular concern in Eastern and Southern Africa as it is the most affected region contributing approximately 55% to the global number of infections in 2020 (UNAIDS, 2021). Disturbingly, South Africa has 7.8 million people infected with HIV, with 60% of the daily new cases occurring in sub-Saharan Africa. Even though adolescent females only represent 10% of the South African population, this demographic accounted for 25% of HIV infections in 2020. In addition, the recent global COVID-19 pandemic has drastically affected the UNAIDS 2020 targets for the HIV epidemic (UNAIDS, 2021).

The HI virus was first isolated in 1983 (Barre-Sinoussi *et al.*, 1983) and has since been shown to be genetically diverse with two types, HIV-1 and HIV-2, identified (Figure 1.1). HIV presents with a high degree of genetic diversity resulting in various distributions across the world. Generally, HIV-1 infections predominate while HIV-2 is mostly confined to West Africa (Clavel *et al.*, 1986). HIV-1 is separated into four groups - M, N, O and P of which group M accounts for 95% of the complete genome of HIV-1 (McCutchan, 2006). The major group M is further subdivided into nine subtypes - A, B, C, D, F, G, H, J, K and many circulating recombinant forms (CRFs) (Robertson *et al.*, 2000; Taylor *et al.*, 2008). These subtypes occur in different parts of the world with subtype C contributing to approximately 50% of HIV-1 infections worldwide (Buonaguro *et al.*, 2007). Subtype A is mainly found in East Africa, Eastern Europe and Central Asia, whereas subtype B is more prevalent in North America, Australia and Western Europe (McCutchan, 2006). Of particular interest, subtype C predominates in sub-Saharan Africa, China, India and Brazil, although an increase in the prevalence of this strain is being seen around the world (Hemelaar *et al.*, 2011). Due to the local distributions of these subtypes, majority of global research is focussed on the subtype B. Consequently, all current FDA-approved antiretroviral drugs (ARV) have been developed against the subtype B.

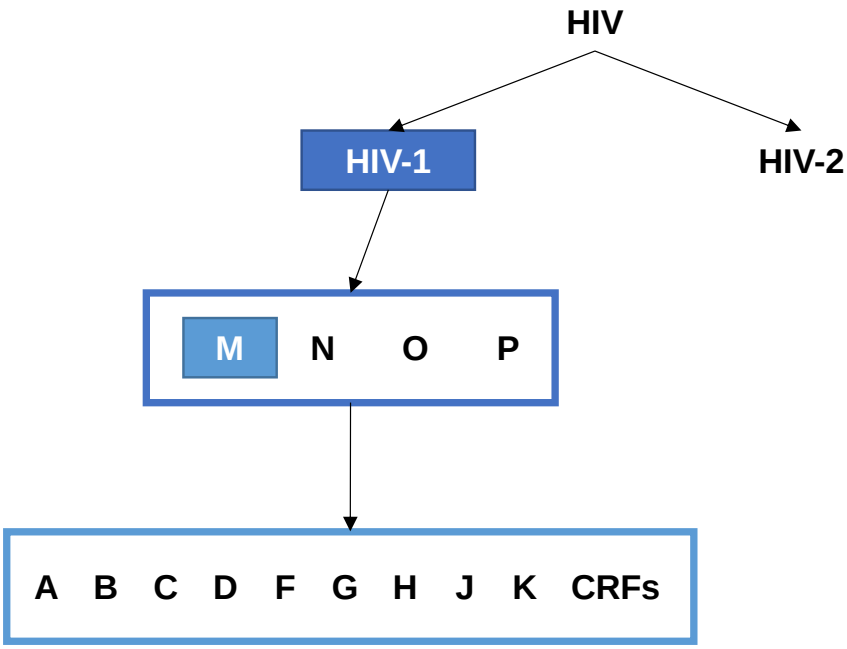


Figure 1.1 Divisions of HIV-1.

HIV-1 is divided into groups M, N, O and P with group M further subdivided into A, B, C, D, F, G, H, J, K and many CRFs (Robertson *et al.*, 2000; Taylor *et al.*, 2008). Image created with Microsoft PowerPoint, version 2204.

1.1.1 HIV GENOME AND LIFE CYCLE

HIV is a highly mutable retrovirus which is categorised into the *Lentivirus* genus. HIV is made up of a compact genome that contains two copies of single-stranded ribonucleic acid (ssRNA) (Cullen, 1991). The ssRNA, which codes for three large polyprotein precursors (Gag, Gag-Pol and Env) and several small regulatory proteins (nef, ref, tat, vif, vpr and vpu), must be reverse transcribed into deoxyribonucleic acid (DNA) prior to being incorporated into the host genome (Freed, 1998).

The HIV virion is encased by a lipid bilayer that is derived from the host cell (Figure 1.2). The glycoprotein, GP120, is found on the surface of the membrane and is anchored to the membrane by GP41 which is a transmembrane protein, both encoded by the *env* gene. The membrane is lined with matrix proteins (MA) which is encoded by the *gag* gene along with the nucleocapsid (NC) and capsid (CA) proteins. The conical capsid is made up of CA monomers which are assembled into hexamers and pentamers. These structures facilitate the curves at the bottom and top, closing the

capsid. The NC stabilises the two copies of unspliced viral ssRNA found within the capsid core. The *pol* gene encodes three enzymes: reverse transcriptase (RT), integrase (IN) and protease. Finally, the accessory proteins vif, vpr, vpu and nef are also found in the capsid core.

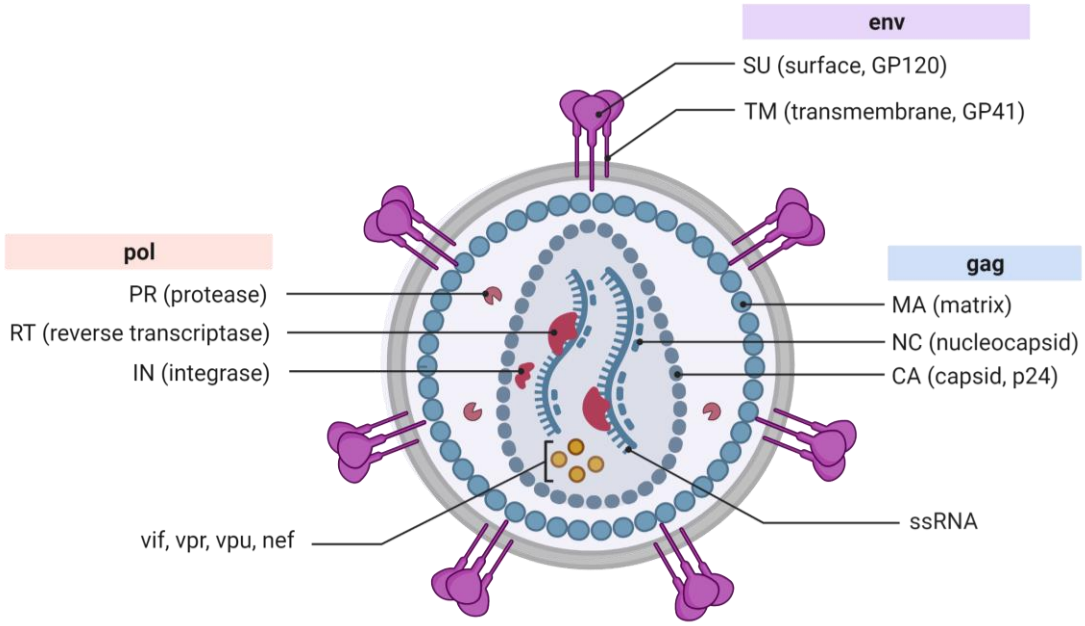


Figure 1.2 HIV virion.

The virus consists of the *pol/gag/env* genes which encode for various proteins required by the HIV to survive and replicate. The *pol* encodes the protease, reverse transcriptase and integrase (red). The *gag* encodes the MA, NC and CA proteins (blue). The *env* encodes the two glycoproteins: GP120 and GP41 (purple). Adapted from "HIV-1 Genome and Structure", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates.

The HIV life cycle is a complex multi-step process that requires various proteins and processes (Figure 1.3). The life cycle begins with the initial fusion of the virus with the host cell membrane through interactions between the viral GP120 and the host cell receptor CD4+. Thereafter, interactions between GP41 and CCR5 and CXCR4 allows entry into the host cell due to a change in the structure of the virion's envelope (Freed,

1998). The viral capsid enters the cell, is uncoated and two strands of RNA are released along with the three replication enzymes – reverse transcriptase, integrase and protease (Cullen, 1991; Freed, 1998).

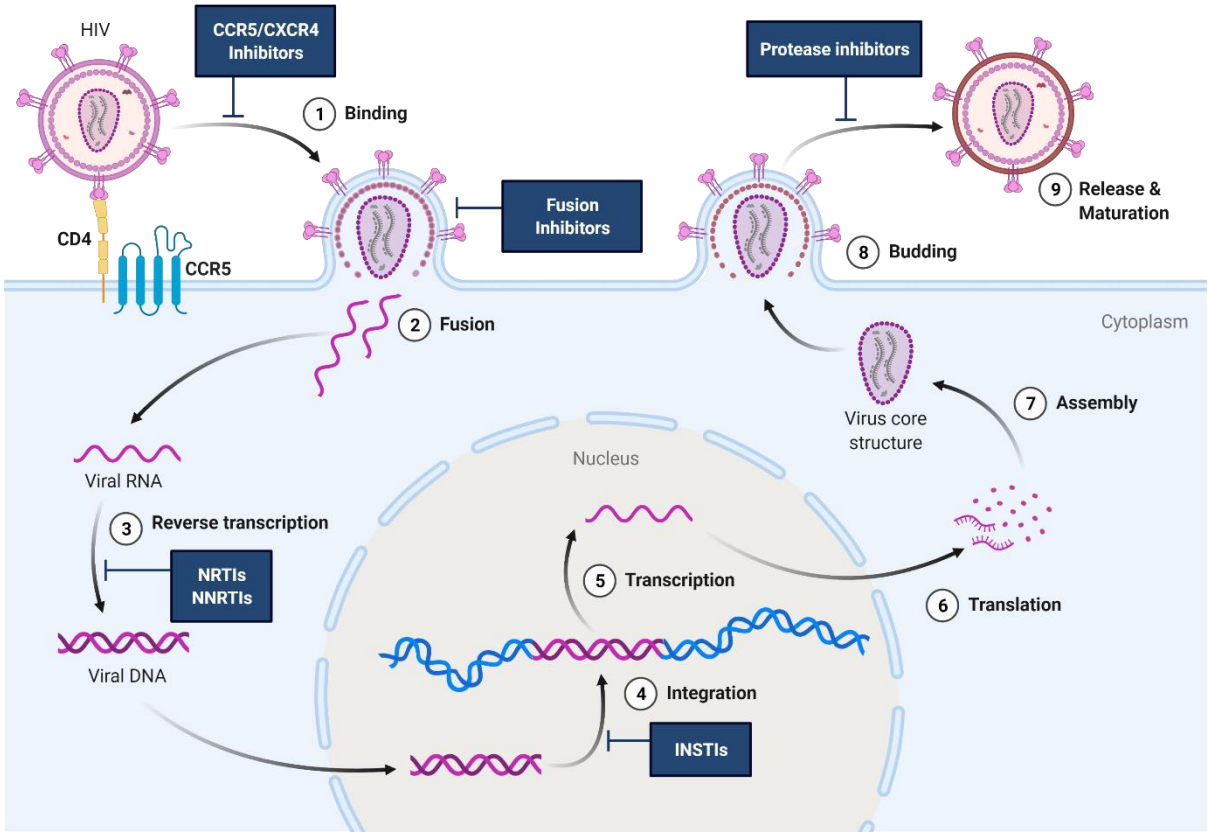


Figure 1.3 Life cycle of HIV-1.

The stages of the life cycle are numbered. Step 1: Binding of the HIV virion to the receptor on the host cell membrane. Step 2: Fusion of the virion to the membrane. Step 3: Reverse transcription of the viral ssRNA into double stranded RNA. Step 4: Entry of viral DNA into the nucleus and subsequent integration into the host cell genome. Step 5: Transcription of viral mRNA, ssRNA and genomic material. Step 6: Translation of viral mRNA into Gag and Gag Pol polyproteins. Step 7: Assembly of virus core structure. Step 8: Budding of virus core structure. Step 9: Release and maturation of mature, infectious HIV virion. Stages at which the life cycle can be inhibited is indicated by the navy boxes. Adapted from “HIV Sites for Therapeutic Intervention”, by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates.

The viral RNA is converted into DNA by RT which is inserted into the host genome by the IN enzyme (Emerman and Malim, 1998). The host cell's transcription machinery is then used to produce messenger RNA (mRNA) which is later translated into Gag, Gag-Pol, Env polyproteins and the accessory proteins. The viral RNA along with Gag and Gag-Pol precursors migrate to the cell membrane to form an immature virion that is able to bud off from the cell. The maturation stage occurs via the cleavage of Gag and Gag-Pol by HIV protease (Freed, 1998). The dimerisation of the Gag-Pol polyprotein activates the protease (Sawyer *et al.*, 1994). The embedded dimeric form is extremely unstable and initially cleaves intramolecularly until it is free and subsequently able to intermolecularly cleave the Gag polyprotein (Louis *et al.*, 1991). Once cleaved, the structural proteins (MA and CA) rearrange to produce an infectious virion. This fully functional virion can now attach to another cell membrane and repeat the HIV replication cycle.

1.2 HIV/AIDS TREATMENT

The multi-stage replication of HIV provides an abundance of processes at various stages of the life cycle that can be inhibited by ARV drugs. Currently, eight classes of anti-HIV drugs, which act on different stages of the life cycle, are available (WHO, 2003). These include nucleoside RT Inhibitors (NRTIs), non-nucleoside RT Inhibitors (NNRTIs), Protease Inhibitors (PIs), IN strand transfer inhibitors (INSTIs), fusion inhibitors and the entry inhibitors - CCR5 antagonists, CD4 T lymphocyte (CD4) post-attachment inhibitors and GP120 attachment inhibitors (Figure 1.3). Highly Active Antiretroviral Therapy (HAART) typically includes a combination of these drugs to control the replication of the virus. The treatment cannot fully eliminate the virus but rather allows for strengthening of the immune system by suppressing viral replication within the body (WHO, 2003).

As previously mentioned, the currently available antiretroviral drugs were designed against subtype B infections and, therefore, may not be as effective on other subtypes, such as subtype C (Mosebi *et al.*, 2008). The HIV protease is an attractive drug target because it is essential in the maturation phase and the production of functioning infectious virions (Freed, 1998) with recent data suggesting that PIs are the most effective course of therapy (Thompson *et al.*, 2010). Therefore, a complete

comprehension of the HIV protease, specifically the less studied subtype C protease, is required in order to produce effective drugs that specifically target the enzyme (Mosebi *et al.*, 2008).

1.3 Gag and Gag-Pol Polyproteins

The HIV genome codes for *env, gag and pol* which is typical of retroviruses. The Gag polyprotein is the main structural protein of HIV and is composed of the MA, CA, NC, p6 and two linker proteins – SP1 and SP2 (Figure 1.4). In order for the virus to assemble correctly, the order in which Gag is cleaved is essential. Cleavage of Gag begins with the separation of the NC from the CA which is located at the C-terminal end of the linker peptide SP1 (Clavel and Mammano, 2010). The CA protein is separated from the MA protein and the C-terminal sequence of p6 is released almost simultaneously (Clavel and Mammano, 2010). The linker proteins SP1 and SP2 are then trimmed from the CA and the NC, respectively. SP1 is vital for the incorporation of Gag and Pol into the virus while SP2 is vital for the regulation of ordered cleavage and maturation of Gag (Pettit *et al.*, 1994; Shehu-Xhilaga *et al.*, 2001; Clavel and Mammano, 2010).

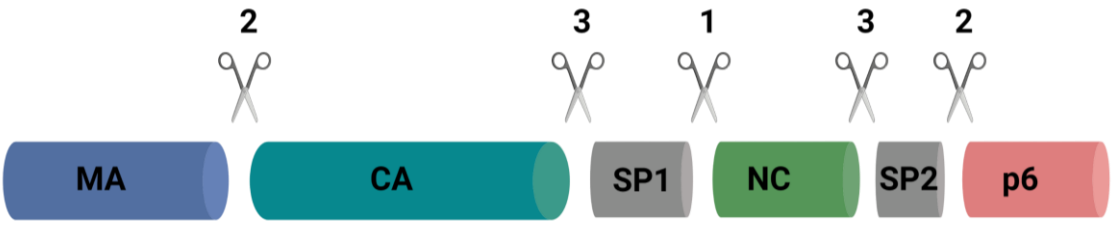


Figure 1.4 Gag polyprotein of HIV.

The order of cleavage of Gag is indicated by the numbered cleavage sites. Image created with BioRender.com.

1.4 South African HIV-1 subtype C protease

The HIV-1 protease (EC 3.4.23.16) is crucial for the production of new infectious virus particles and, therefore, provides an attractive target for drug therapy. The immature non-infectious virion contains mostly uncleaved Gag and Gag-Pol polyproteins. Maturation occurs through the cleavage of Gag at five specific cleavage sequences which results in the rearrangement of the virus core (Darke *et al.*, 1988; Briggs and Krausslich, 2011).

1.4.1 Structure and function of protease

HIV-1 protease is released from the Gag-Pol precursor through autoproteolytic processing. HIV-1 protease exists as a 22 kDa symmetrical homodimer with two monomers of 99 amino acids each (Coman *et al.*, 2008). HIV-1 protease belongs to a class of aspartic proteases and contains the signature aspartic acid, threonine and glycine (Asp25Thr26Gly27) conserved active site triad (Brik and Wong, 2003; Coman *et al.*, 2008). The catalytic triad is present in both monomers and is stabilised by a network of hydrogen bonds (Brik and Wong, 2003). The two monomers associate non-covalently through the dimer interface which results in the formation of the active site cavity (Wlodawer *et al.*, 1989). The Asp25/Asp25′ residue from each monomer contributes to formation of the active site, with the prime symbol indicating the amino acid belonging to the second monomer. The HIV-1 protease can only function in its dimeric form and requires the catalytic Asp25 residue to be enzymatically active (Seelmeier *et al.*, 1988; Naicker *et al.*, 2013).

Each dimeric subunit of HIV-1 protease consists of one α-helix and two β-sheets with a single binding pocket (Wlodawer *et al.*, 1989). The HIV-1 protease is structurally divided into the flap, hinge, cantilever, fulcrum and dimer interface regions (Figure 1.5). The dimer interface contains four short anti-parallel β-strands and is formed by residues 1-4 and 96-99 of the N- and C- termini from each monomer.

The flap region, residues 42-56, is important to the specificity and activity of the HIV-1 protease (Naicker *et al.*, 2013). The two glycine rich flaps are flexible and control access of substrates or inhibitors to the active site. These flexible flaps exist in the open, semi-open and closed conformations which can be defined by the distance

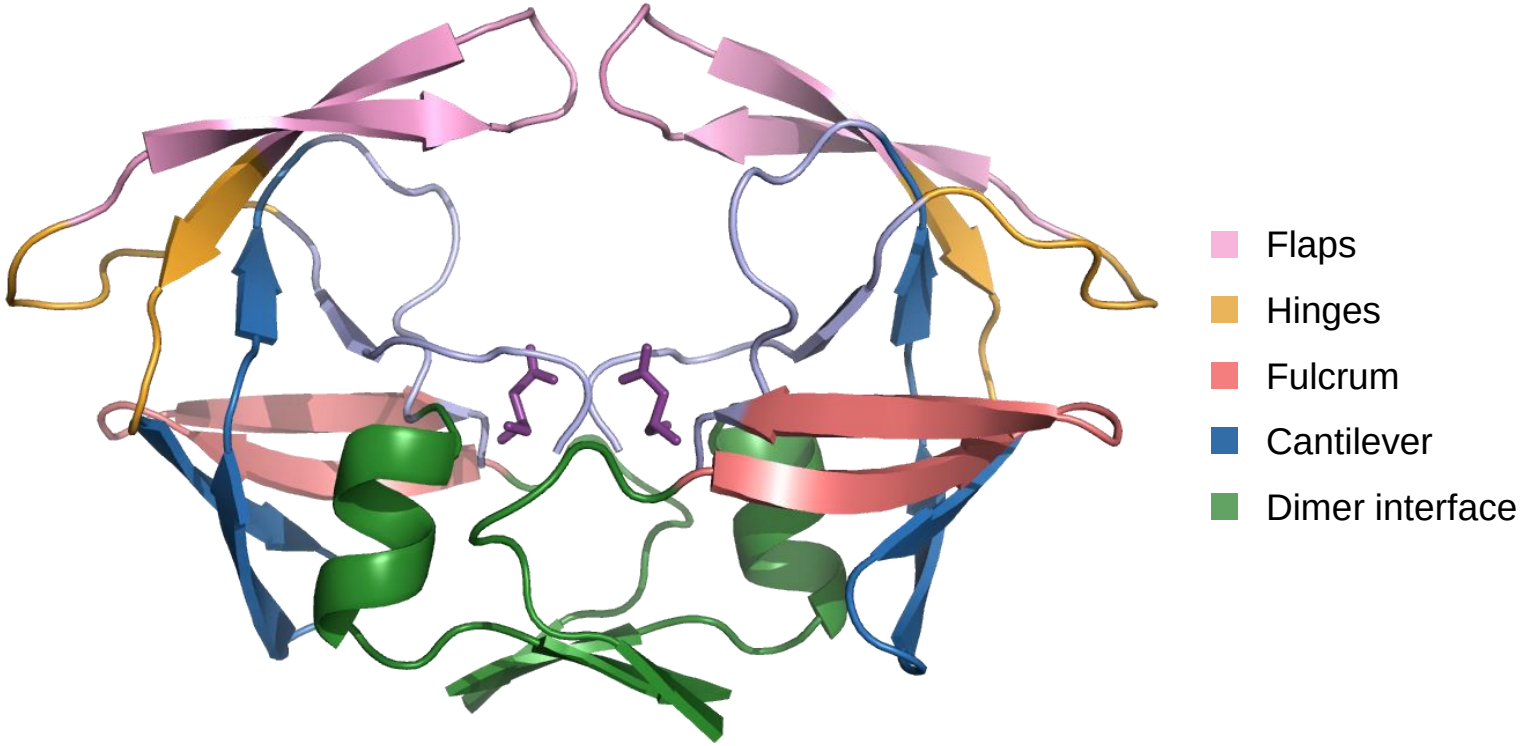


Figure 1.5 Structure of HIV-1 protease.

The regions of the wild-type (WT) HIV-1 South African subtype C (C-SA) are shown as follows: residues 10-22 indicating the fulcrum are shown in red, residues 35-42 and 57-61 indicating the hinge are shown in orange, residues 42-56 indicating the flaps are shown in pink, residues 62-78 indicating the cantilever are shown in blue and residues 1-9 and 86-99 indicating the dimer interface are shown in green. The purple sticks represent the Asp25 residues found in the active site cavity of the protease. The image was constructed using the PyMOL v0.99rc6 software using PDB ID: 3U71.

between the flap tip Ile50/50′ residues and the catalytic Asp25/25′ residues (Perryman *et al.*, 2006). A distance of <17 Å between these two residues represents the closed conformation, 17-22 Å represents the semi-open conformation and >22 Å represents the open conformation (Figure 1.6). The movement of the flaps allow entry of substrates (open conformation) and subsequent substrate binding (closed conformation) (Meiselbach *et al.*, 2007). This movement is largely facilitated by the hinge region, residues 35-42 and 57-61, which also contributes to the stability of the flaps (Naicker *et al.*, 2013). The opening of the flaps has also been linked to the movement of the cantilever (residues 62-78) and the fulcrum (residue 10-22) which acts with the hinge region in a concerted downwards motion (Hornak *et al.*, 2006; Foulkes-Murzycki *et al.*, 2007). The movement of these regions is referred to as the hydrophobic sliding mechanism and is described as the downward sliding of the hinge and cantilever regions across the surface of the fulcrum region. The hydrophobic sliding mechanism results in the upward and outward movement of the flaps and relies on the hydrophobic core of the protease which is stabilised by van der Waals contacts that can easily be exchanged between adjacent residues (Todd *et al.*, 2000).

The hydrophobic core of the protease has limited accessibility to solvent and is composed of twenty amino acids - L5, V11, I13, V15, I19, A22, L24, I26, L33, L38, I62, I64, I66, V75, V77, I85, M89, L90, L93, L97 (Figure 1.7) (Foulkes-Murzycki *et al.*, 2007; Naicker *et al.*, 2014). Several residues within the hydrophobic core are involved in the hydrophobic sliding mechanism. In the closed formation, the hinge region residues M36 and L38 and the cantilever residues I62, I64 and V75 form contacts with the fulcrum residue I15. Sliding of the hinge and cantilever regions across I15 and downwards toward the termini results in the transfer of the original contacts to I13 resulting in the opening of the flaps (Foulkes-Murzycki *et al.*, 2007). An increase in flap region flexibility could possibly reduce drug susceptibility and the rate of substrate proteolysis (Naicker *et al.*, 2013). Therefore, it is imperative to understand the role of amino acid substitutions, mutations and insertions in the flap and hinge regions and the resulting contribution to the activity of the HIV-1 protease.

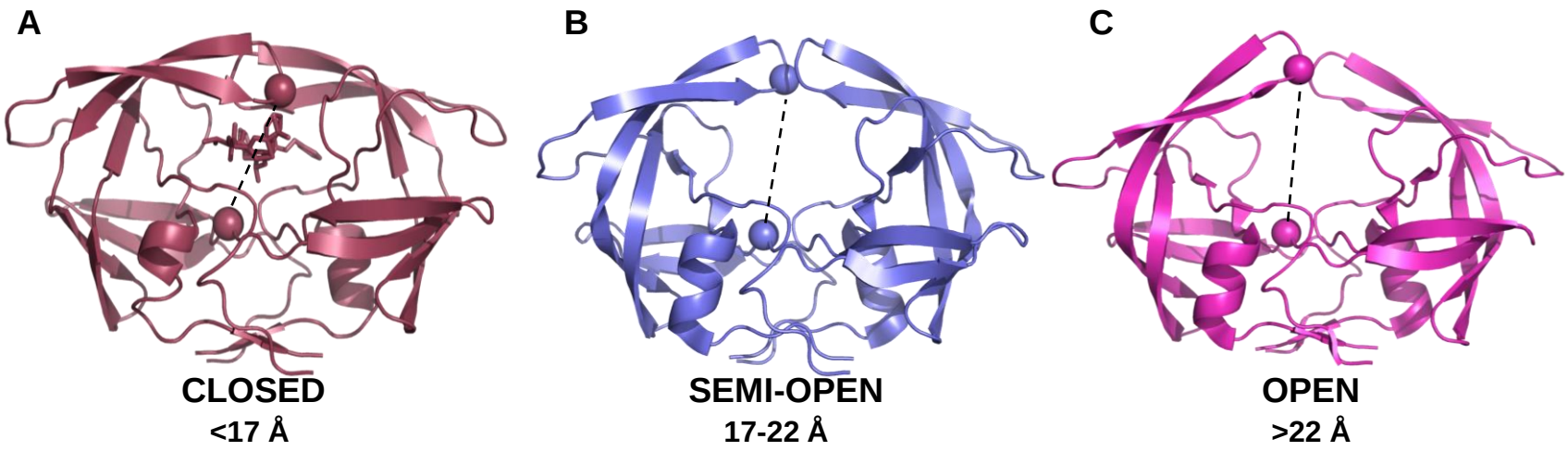


Figure 1.6 Heterogenous flap conformations found in the HIV-1 protease.

A: Closed conformation with Ritonavir shown in sticks (PDB: 1HXW). **B:** Semi-open conformation (PDB: 1HHP). **C:** Closed conformation (PDB: 3UF3). The flap tip Ile50 and active site Asp25 residues as displayed as spheres with the distance between these residues defining each conformation. The image was constructed using the PyMOL v0.99rc6 software.

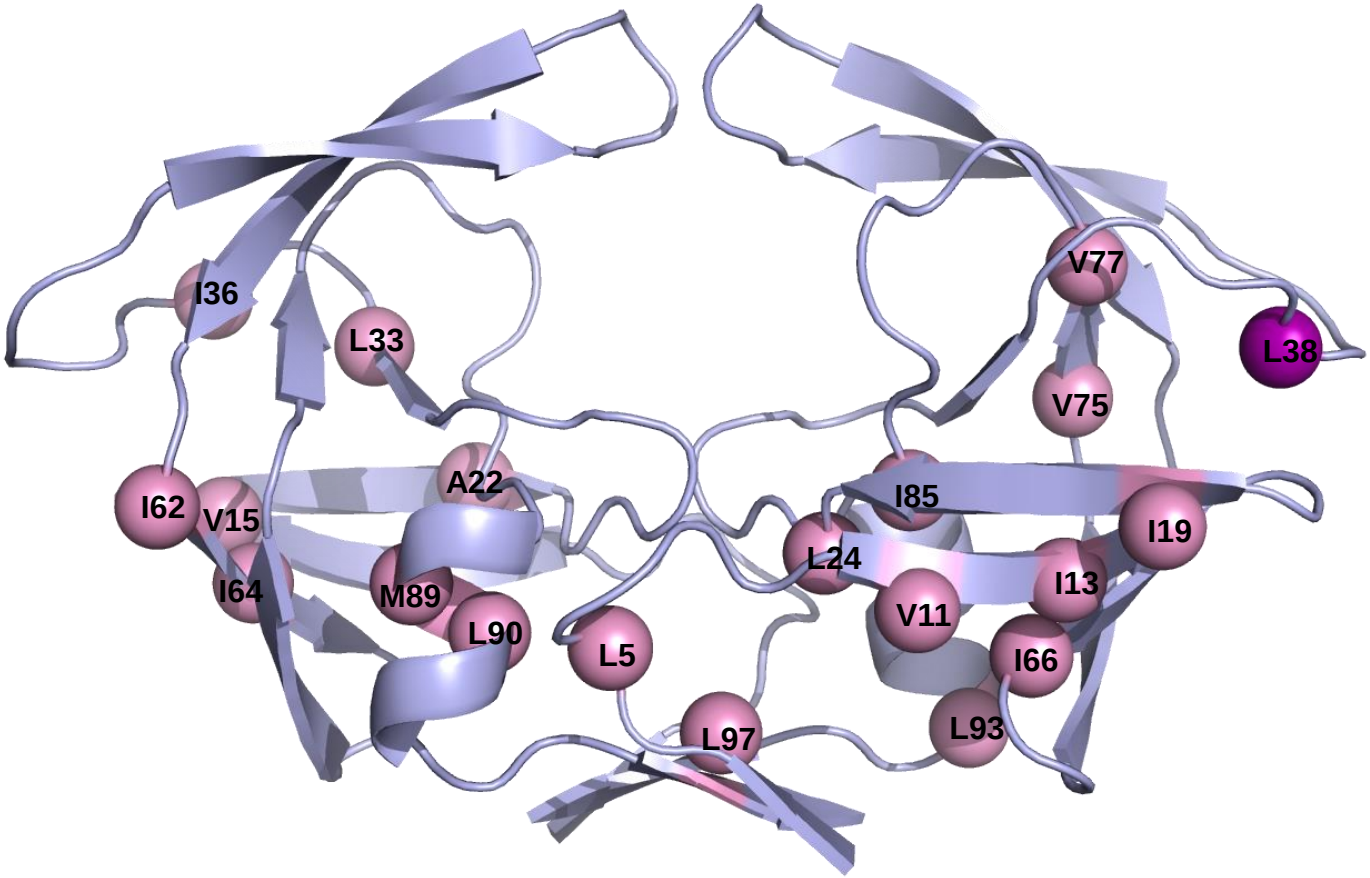


Figure 1.7 Hydrophobic core residues of the HIV-1 protease.

All twenty amino acid residues (L5, V11, I13, V15, I19, A22, L24, I26, L33, L38, I62, I64, I66, V75, V77, I85, M89, L90, L93, L97) involved in the hydrophobic core are shown as pink spheres (Foulkes-Murzycki *et al.*, 2007; Naicker and Sayed, 2014). The image was constructed using the PyMOL v0.99rc6 software using PDB ID: 3U71.

The South African HIV-1 subtype C (HIV-1 C-SA) protease accounts for approximately 95% of infections in sub-Saharan Africa (Mosebi *et al.*, 2008). The HIV-1 C-SA protease differs from the subtype B consensus sequence by eight polymorphisms – T12S, I15V, L19I, M36I, R41K, H69K, L89M AND I93L (Mosebi *et al.*, 2008). These polymorphisms occur in areas that do not affect catalytic activity, substrate binding affinity or structural stability (Naicker and Sayed, 2014). They occur distal to the active site and do not alter viral fitness. However, these polymorphisms influence the binding thermodynamics of protease inhibitors and may exacerbate drug resistance attributed to other known mutations.

1.5 Catalytic mechanism of HIV-1 protease

A water molecule, which can be activated either by a zinc atom or the two β-carboxylate groups of the Asp25 at the active site, acts as a nucleophile and attacks the amide bond of the substrate's scissile bond consistent with a general acid-base mechanism (Figure 1.8) (Suguna *et al.*, 1987). In the absence of a bound substrate, the active site of the protease is filled with water molecules. The two catalytic Asp25 residues interact with a water molecule which acts as a nucleophile. Therefore, when a substrate binds to the protease (ES), the nucleophile attacks the carbonyl carbon of the scissile bond resulting in the production of a gem-diol intermediate (ES*). The covalent linkages in the substrate are subsequently broken and results in the formation of two hydrolysis products, P_1 and P_2, which are sequentially released allowing for the hydrolysis of another substrate (Suguna *et al.*, 1987).

1.6 Protease inhibitors

The HIV-1 protease is an extremely attractive drug target due to its significant role in the replication process, with inactivation of the enzyme shown to produce immature non-infectious virions (Kohl *et al.*, 1988; Seelmeier *et al.*, 1988). The majority of PIs currently available mimic the transition state of Gag and Gag-Pol which is the natural substrate of the protease. There are currently nine available PIs - Saquinavir (SQV), Indinavir (IDV), Amprenavir (APV), Ritonavir (RTV), Nelfinavir (NFV), Lopinavir (LPV), Atazanavir (ATV), Tipranavir (TPV) and Darunavir (DRV) (Figure 1.9). These inhibitors

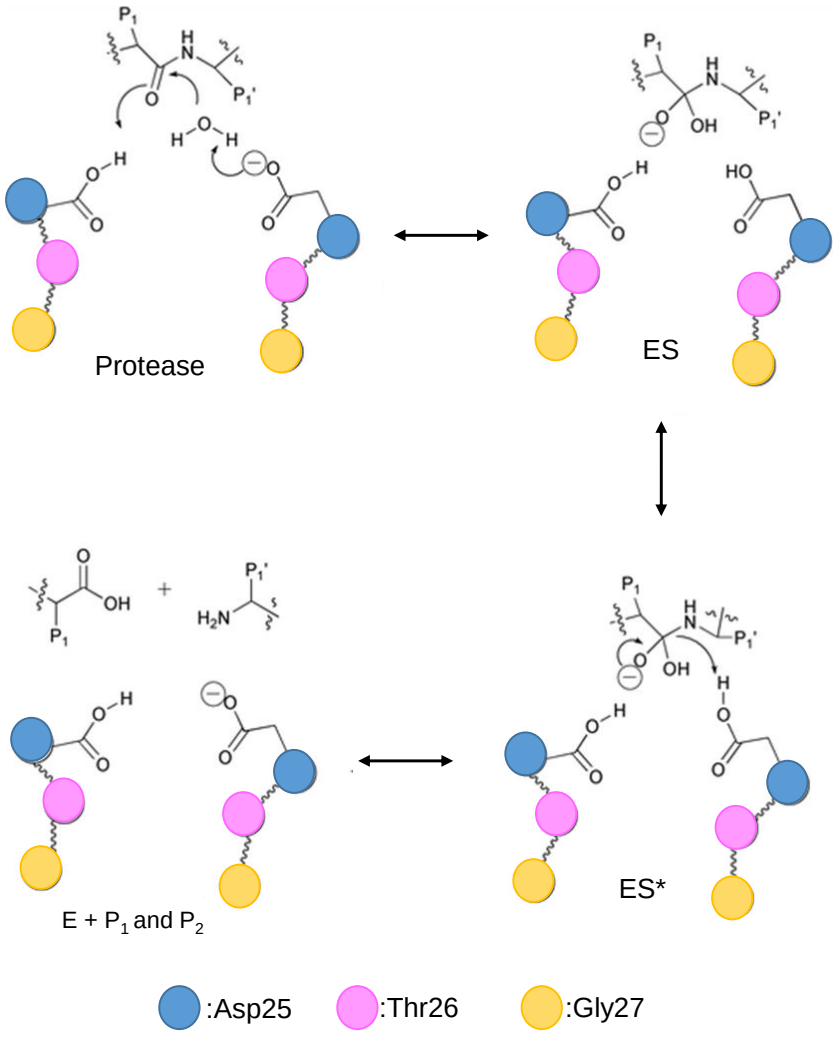


Figure 1.8 Catalytic mechanism of the HIV-1 protease.

The two catalytic Asp25 residues (blue spheres) of the protease activate the conserved water molecule, which then acts as a nucleophile for the cleavage of the peptide bond in the ES complex producing the ES* intermediate. The covalent linkages in the substrate are then broken resulting in the release of the enzyme (E) and products (P_1 and P_2). Figure adapted from Suguna *et al.* (1987).

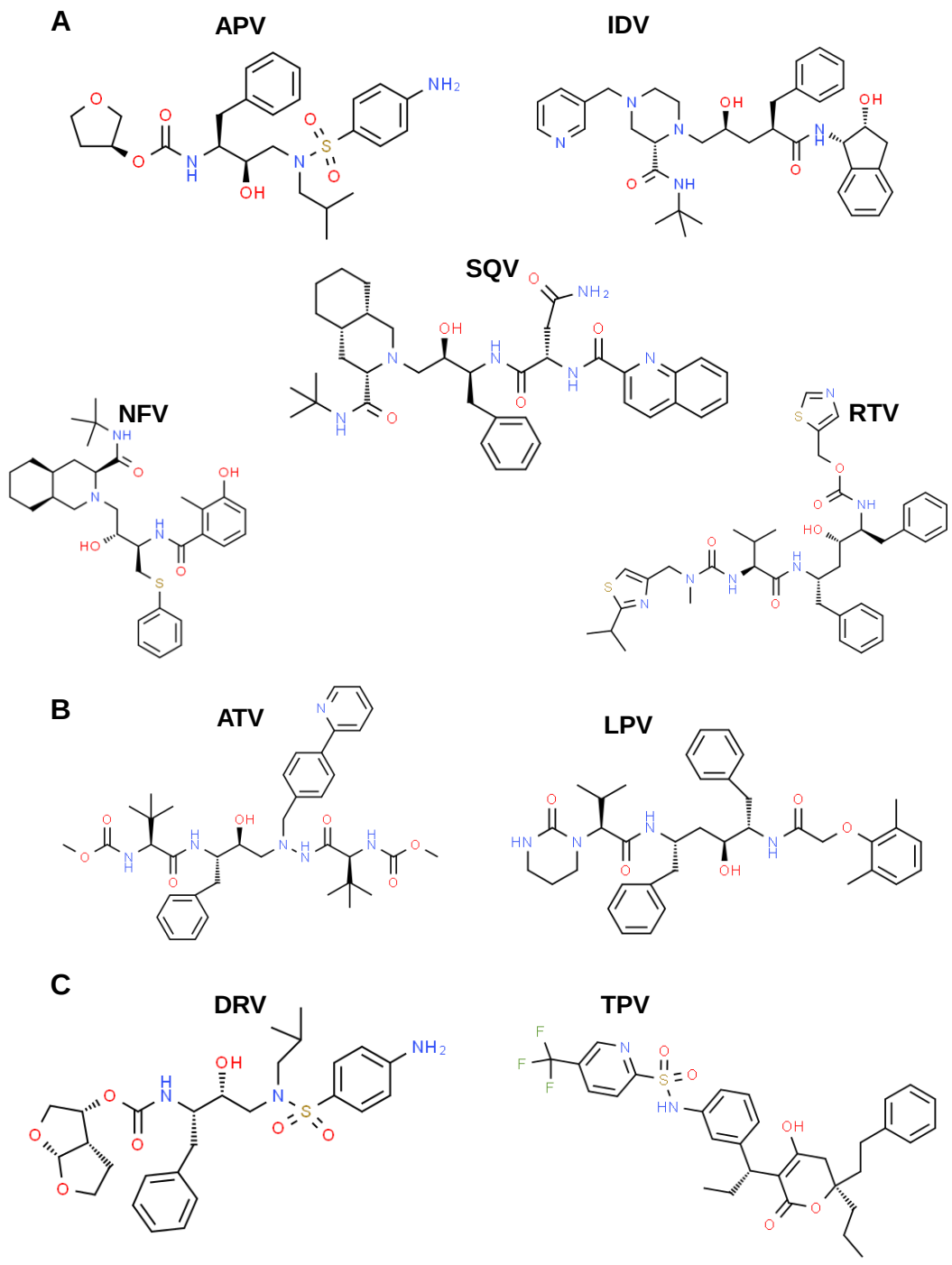


Figure 1.9. Three-dimensional structure of the nine clinically approved PIs.

A: First-generation PIs including APV (CID: 65016), IDV (CID: 5462355), NFV (CID: 64142), RTV (CID: 392622) and SQV (CID: 60787). **B:** Second-generation PIs including ATV (CID: 148192) and LPV (CID: 92727). **C:** Third generation PIs including DRV (CID: 213039) and TPV (CID: 4682461). The image was constructed using the ChemDraw online tool.

specifically target the HIV-1 protease by competitively binding to the active site with a higher affinity than their corresponding substrates (WHO, 2003).

The rapid development of drug resistance due to the low fidelity and lack of proofreading of the RT and the high mutation rate of the virus required a change in drug design. Second-generation PIs were designed with the central hydroxyl group but with less peptide backbone features. While the previous PIs were designed to contain a peptide-like carbonyl, the new PIs contained a sulfonamide instead. The second-generation PIs include ATV and LPV. LPV is only used in combination with RTV and its design was closely based off of this first-generation PI. ATV contains an aza-dipeptide core, an extended moiety and is a 2-hydroxyl-1,3-diaminopropane transition state analogue. The third-generation PIs, TPV and DRV, followed shortly after. TPV belongs to the dihydropyrones and is effective against isolated that are resistant to RTV. DRV is a broad spectrum, non-peptide inhibitor that is chemically related to the first-generation PI APV. DRV contains a novel *bis*-tetrahydrofuranylurethane (*bis*-THF) group which allows additional polar interactions with the protease making it a potent inhibitor.

The first-generation PIs include SQV, RTV, IDV, NFV and APV. SQV, a hydroxyethylamine transition state anoalogue, was the first PI to enter clinical trials. This was followed by the introduction of RTV which was derived from a C_2-symmetric peptidomimetic inhibitor. RTV is often used as a 'booster' with another PI in ARV treatment as it increases the blood plasma concentrations of the other PIs. IDV belongs to the hydoxyaminopentane amide (HAPA) compounds while NFV is a selective non-peptidic PI. APV has been shown to have activity against both HIV-1 and HIV-2.

As described by WHO, adolescent and adult first-line ART should contain two NRTIs and either an INSTI or a NNRTI. PIs are used as part of a second-line regimen in the case of NNRTI failure with LPV and ATV being preferred in South Africa. DRV is used as a recovery treatment after failure of the second line regimen (WHO, 2016).

During rational drug design, drug binding thermodynamics plays a crucial role in the development of new compounds. Compounds that achieve a high binding affinity towards its target combine favourable entropic and enthalpic contributions to the overall Gibbs free energy of binding (Ohtaka *et al.*, 2002). The contribution of entropy

and enthalpy towards Gibbs free energy originates from different types of interactions in the binding process. The binding enthalpy primarily indicates the energetic contribution of hydrogen bonds, polar and dipolar interactions and van der Waals interactions between the ligand and protein during the binding event. On the other hand, the binding entropy refers to the degree of disorder accompanying complex formation and reflects the solvation and conformational entropies. Some compounds bind targets with the same affinity but have different mechanisms and can be either entropically or enthalpically driven (Velazquez-Campoy *et al.*, 2001).

The binding of majority of the first generation PIs to the HIV-1 protease are entropically driven ($T\Delta S > 0$) whereas the second generation PIs bind the enzyme with both favourable enthalpic ($\Delta H < 0$) and entropic conditions (Velazquez-Campoy *et al.*, 2001). DRV has been shown to bind C-SA protease in enthalpically favourable and entropically disfavoured conditions whereas it binds to some HIV-1 proteases under entropically favourable conditions (Mittal *et al.*, 2013; Kozisek *et al.*, 2014). Additionally, ATV and RTV both bind exothermically to C-SA protease (Velazquez-Campoy *et al.*, 2001; Velazquez-Campoy *et al.*, 2003).

1.7 MUTATIONS AND DRUG RESISTANCE

Due to the highly error-prone nature of HIV-1 reverse transcriptase (Tang and Pillay, 2004), a number of mutations arise during the replication process that may result in failure of HAART. Specifically, the introduction of several mutations within the protease sequence with more than half of the 99 amino acids affecting drug susceptibility (Wu *et al.*, 2003). Depending on the mutations present and the number of mutations that occur, varying degrees of cross-resistance to PIs may occur. These mutations can be classed as either primary or secondary mutations (

Table 1.1). Primary mutations by themselves can directly affect resistance to a particular drug because they are specifically selected for each compound (Prabu-Jeyabalan *et al.*, 2002). Secondary mutations, however, occur distal to the active site and thus have little or no effect on resistance by themselves. They usually occur in combination with a primary mutation, either to enhance or compensate a primary resistance mutation (Ammaranond *et al.*, 2003).

Table 1.1 HIV-1 protease mutations associated with inhibitor resistance.

	Amino acid position														
Primary	D	V	M	I	G	I	F	V	N	I					
	30	32	46	47	48	50	53	82	83	84					
Secondary	L	V	K	L	L	E	K	I	Q	G	T	L	N	L	L
	10	11	20	24	33	35	43	54	58	73	74	76	88	89	90

1.7.1 ACTIVE SITE MUTATIONS

Primary mutations occur within the active site (Figure 1.5) and directly affect drug binding by altering the geometry of the active site without affect the polarity or charge (Chen *et al.*, 1995). Primary mutations typically occur at residues which are in direct contact with PIs and results in conformational changes which affect drug binding interactions (Velazquez-Campoy *et al.*, 2001; Ohtaka *et al.*, 2002; Kim and Baxter, 2008). These mutations often result in the enlargement of the active site which not only decreases inhibitor binding but also reduces the binding affinity of substrates (Croteau *et al.*, 1997; Nijhuis *et al.*, 1999; Mammano *et al.*, 2000). Therefore, while these mutations advantageously confer drug resistance to the virus, they also have to potential to diminish enzyme function and viral fitness (Clavel and Mammano, 2010). As a result, subsequent 'compensatory' mutations are introduced into the hydrophobic core as well as the Gag and Gag-Pol cleavage sites often resulting in the restoration of enzymatic activity (Piana *et al.*, 2002; Clavel and Mammano, 2010; Codoner *et al.*, 2017).

1.7.2 NON-ACTIVE SITE MUTATIONS

Secondary mutations are found outside of the active site and often occur as compensatory mutations to reduce the effect of primary mutations on enzyme activity. However, these mutations have been directly linked to inhibitor resistance as well, suggesting that secondary mutations may not be exclusively compensatory mutation (Clemente *et al.*, 2003; Muzammil *et al.*, 2003; Foulkes-Murzycki *et al.*, 2007; Rhee *et al.*, 2010). The accumulation of primary and secondary mutations occurring in multiple combinations results in differential drug resistance profiles depending on the mutations

present while still maintaining substrate specificity and turnover (Hirsch *et al.*, 2000; Dumans *et al.*, 2002; Prabu-Jeyabalan *et al.*, 2002).

1.7.3 Effects on drug binding thermodynamics

Drug binding affinities can be directly or indirectly affected by mutations that occur at specific amino acids in the HIV-1 protease with approximately 87 mutations been shown to confer resistance (Todd *et al.*, 2000). These polymorphisms occur in at least 47 positions of the protease and are active site and non-active site mutations. Drug binding thermodynamics of IDV, RTV, SQV and NFV on a HIV-1 protease containing a V82F/I84V double mutation were less favourable (Todd *et al.*, 2000). The binding of these four PIs to HIV-1 protease are mainly driven by a large positive entropy change. The V82F/I84V mutation, located at the edges of the active site, altered the binding enthalpy and entropy of these PIs with the less favourable change in entropy partially caused by the conformational change that occurs when the PIs bind the enzyme (Todd *et al.*, 2000). Notably, drug binding affinities for PIs were altered between subtype B and C-SA HIV-1 proteases revealing a decrease in affinity of DRV and to a lesser extent RTV for the C-SA protease in comparison to the subtype B protease (Naicker *et al.*, 2014). Additionally, the C-SA protease displayed higher vitalities to DRV and RTV suggesting that in the presence of these PIs the replication capacity of the protease is improved (Naicker *et al.*, 2014).

1.7.4 Effects on structure and stability

The most conserved regions of the HIV-1 protease includes the Asp25 catalytic site, the tip of the flap region (residues 49 to 51) and residues 80 to 91 (Zoete *et al.*, 2002). Notably, residues 82 and 84 which are usually in contact with the substrate or ligands are not conserved (Zoete *et al.*, 2002).

Although a number of structures for the subtype B protease can be found in the PDB, only a handful are reported for the subtype C protease. The first recently reported C-SA protease structure was resolved at 2.7 Å exhibiting the eight polymorphisms, mentioned previously, when compared to the subtype B protease (Naicker *et al.*, 2013). The structural analysis of HIV-1 C-SA protease indicated that a polymorphism

found at position 36 of the protease could affect the stability of the hinge region (Naicker *et al.*, 2013).

Furthermore, a subtype B protease containing the double V82F/I84V mutation was structurally more stable in comparison to the WT HIV-1 protease (Todd *et al.*, 2000). The HIV-1 protease D30N and L90M mutations occur separately and have often been linked to NFV drug resistant strains (Kozisek *et al.*, 2007). Although the L90M is a secondary mutation, it has been suggested that this mutation plays a role in the stability of the protease dimer (Mahalingam *et al.*, 1999). X-ray structures of the protease bound with NFV displayed no difference between wild type and D30N mutant (Kozisek *et al.*, 2007). However, when combined with a secondary mutation, a difference in the flap region is observed between the single and double mutants (Kozisek *et al.*, 2007). The decrease in binding affinity of the D30N mutant to NFV was suggested to be attributed to the decrease in conformational entropy of the PI upon binding rather than a change in structure of the protease.

1.7.5 Effects on flap region

Generally, a single amino acid mutation cannot produce significant inhibitor resistance and a combination of multiple polymorphisms are necessary to confer resistance. Mutations that arise in the protease are generally found at specific regions of the enzyme and do not occur randomly (Velazquez-Campoy *et al.*, 2003). Secondary mutations have been proposed to change the equilibrium between the flap conformers (Foulkes-Murzycki *et al.*, 2007; Mittal *et al.*, 2012). This presents a dilemma since PIs are designed to bind in the closed conformation and an equilibrium shift that favours one or other conformation may affect binding or dissociation rates. Secondary mutations within the hydrophobic core affects the internal hydrophobic surface area of the protease thereby altering van der Waals interactions within the core resulting in a change of flap dynamics and flexibility (Goldfarb *et al.*, 2015).

A C-SA protease containing a hinge region mutation and insertion (indicated by the upward arrow) at position 37 displayed the highest level of flexibility around the flap region. The mutation and insertion of the N37T↑V protease resulted in hinge region flexibility which, in turn, affected the rate and width of the active site opening (Zondagh *et al.*, 2018). Additionally, a C-SA protease containing a mutation and insertion at

position 36 displaying greater flexibility and movement of flaps when bound to NFV. In addition to this, the mutant displayed a decrease in affinity to DRV, NFV and TPV (Lockhat *et al.*, 2016).

1.7.6 INSERTION MUTATIONS

Though still rare (approximately 0.5% prevalence), amino acid insertions in the C-SA protease hinge region are being discovered more often and are frequently found in treatment-naïve patients (Ledwaba *et al.*, 2019). Amino acid insertions near residues 18, 25, 36, 70 and 95 in the HIV-1 protease have been identified and appear to be duplicates of the adjacent genetic sequence (Kim *et al.*, 2001; Sturmer *et al.*, 2003; Winters *et al.*, 2005). These insertions in the HIV-1 protease are caused by the stalling and slippage of the RT (Winters and Merigan, 2005) and can also be selected for during treatment with ARVs (Wensing *et al.*, 2010). The role of these insertions has not yet been explained but have been suggested to contribute to PI resistance in the presence of additional mutations usually found in the Gag or Gag-Pol proteins (Kim *et al.*, 2001; Ali *et al.*, 2010; Fun *et al.*, 2012).

The presence of the amino acid insertions has been shown to decrease PI susceptibility when found in combination with other protease mutations by affecting the flaps and binding pocket (Kozisek *et al.*, 2008). Notably, these insertions also moderately enhance viral replication (Kim *et al.*, 2001; Kozisek *et al.*, 2008). Insertions between residues 35-42 in the hinge region have become increasingly more prevalent (Adamson, 2012). These hinge region polymorphisms have been linked to impaired flap dynamics during inhibitor binding and most likely contribute to PIs resistance (Kozisek *et al.*, 2008; Naicker *et al.*, 2013; Lockhat *et al.*, 2016; Zondagh *et al.*, 2018).

1.8 L38↑N↑L^{+4}

A C-SA protease, designated L38↑N↑L^{+4}, carrying several polymorphisms was isolated from a drug naïve-infant whose mother was exposed to RT inhibitors (Department of Health Science, Africa) (Ledwaba *et al.*, 2019). The polymorphisms include a subset of mutations - K20R, E35D, R57K, V82I and a double insertion of asparagine and leucine in the hinge region at position 38 (Kuhn *et al.*, 2014) resulting

in each protease subunit of 101 amino acids (Figure 1.10). *In vitro* drug testing showed that the L38↑N↑L^{+4} variant displayed susceptibility to LPV, ATV and DRV when coupled with a WT Gag protein sequence (Williams *et al.*, 2019). However, phenotypic viral assays in the presence of the variant's naturally occurring Gag showed a decrease in susceptibility to DRV, while still susceptible to LPV and ATV (Williams *et al.*, 2019). In addition, L38↑N↑L displayed a reduction in turnover number (k_{cat}), Michaelis constant as well as a reduction in the specific activity (Williams *et al.*, 2019). Further investigation of the L38↑N↑L variant could provide a better understanding of the effect these insertions have on drug binding thermodynamics, structural changes and conformational stability of the protease.

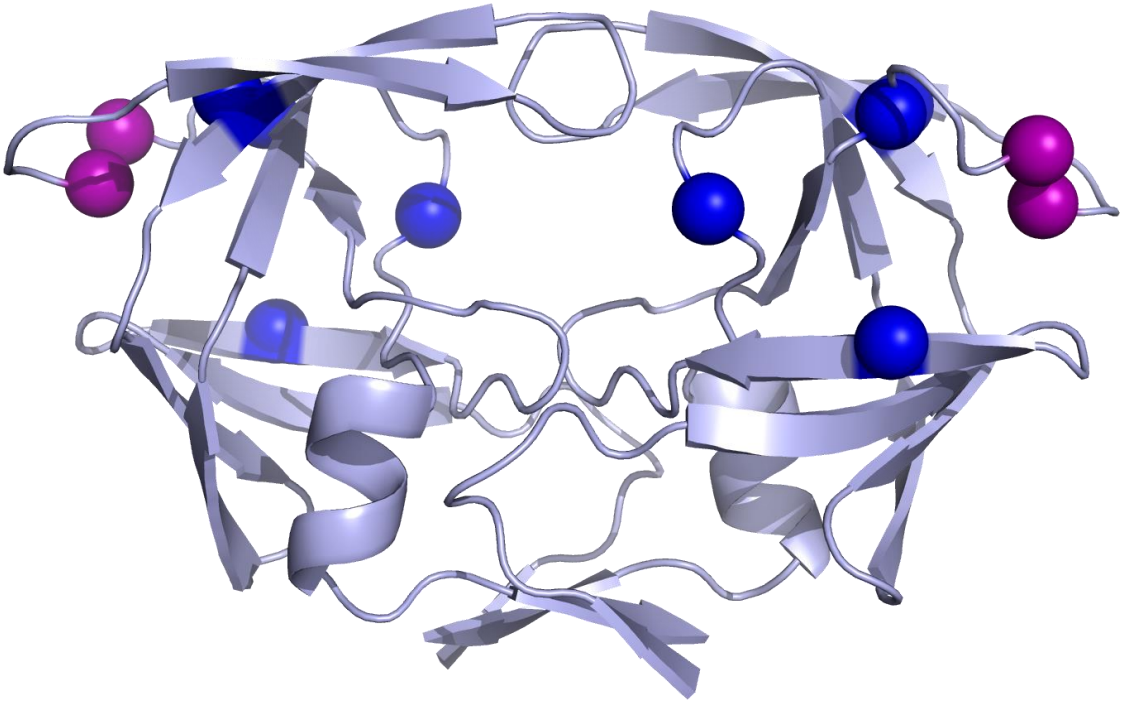


Figure 1.10 Homology model of L38↑N↑L^{+4} protease.

The hinge region double insertion of Asn and Leu is shown as purple spheres, with the accompanying subset of background mutations (K20R, E35D, R57K, V82I) shown as blue spheres. The image was constructed using the PyMOL v0.99rc6 software using PDB ID: 3U71.

1.9 AIM AND OBJECTIVES

1.9.1 AIM

While amino acid insertions in conjunction with other mutations have been shown to affect PI therapy, very little is known about how these insertions directly affect the protease. The aim of this study was to determine how these novel hinge region insertions affect structure and function of the HIV-1 C-SA protease. The L38↑N↑L^{+4} variant containing no other mutations was used to study the effect of hinge region insertions on the HIV-1 C-SA protease. To avoid any confusion this mutant will henceforth be referred to as L38↑N↑L^{-4}, indicating the absence of the four background mutations (K20R, E35D, R57K, V82I).

1.9.2 OBJECTIVES

1. Over-expression of both WT and L38↑N↑L^{-4} proteases through recombinant protein expression in *Escherichia coli*.
2. Purification of both proteases using ion-exchange chromatography.
3. Evaluate structural characteristics using far-ultraviolet circular dichroism, size exclusion chromatography high-performance liquid chromatography and X-ray crystallography.
4. Determine the active site concentration of purified proteases using isothermal titration calorimetry.
5. Determine the binding energetics between proteases and PIs using displacement titration calorimetry and selected PIs (ATV, DRV, RTV and SQV).
6. Evaluate the overall stability of both apo and PI-bound proteases using differential scanning calorimetry.
7. Determine the steady-state enzyme kinetics of L38↑N↑L^{-4}.
8. Elucidate the effect of these insertions on the binding of PIs through *in silico* induced-fit docking studies
9. Elucidate the effect of insertions of the dynamic movement of the proteases through *in silico* molecular dynamics simulations.

2 METHODS

2.1 BACTERIAL TRANSFORMATION

Both the wild-type (WT) and L38↑N↑L^{-4} protease sequence data were obtained from Prof Lynn Morris (AIDS Virus Research Unit, NICD of Johannesburg, South Africa). The pET-11b plasmid encoding the HIV-1 WT CSA protease was previously generated in our laboratory (Sorensen and Mortensen, 2005). The pET-11a plasmid containing the variant L38↑N↑L^{-4} gene was designed in-house and subsequently generated by GenScript (HongKong). Both plasmids were generated with the Q7K mutation in order to decrease autoproteolysis (Louis *et al.*, 1991).

Initially, 3 µl of the pET-11b vector (~100 ng/ul) in milliQ water (18.2 mΩ/cm at 25 °C, Millipore Direct-Q® 3 UV water purification system) was added to a 100 µl stock of competent *Escherichia coli* BL21 (DE3) pLysS cells and incubated on ice for 15 min. For L38↑N↑L^{-4}, 3 µl of the pET-11a vector (~100 ng/ul) was added to a 100 µl stocks of *E. coli* Rosetta, T7 express and BL21 (DE3) pLysS cells. The cells were then subjected to heat-shock at 42 °C for 45 s to ensure the uptake of DNA. Thereafter, the cells were stored on ice for 5 min and followed by the addition of 500 µl of super optimal broth with catabolite repression (SOC) media (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM sodium chloride, 10 mM magnesium chloride, 10 mM magnesium sulphate, 2.5 mM potassium chloride and 20 mM glucose. The glucose was autoclaved prior to be added to SOC media in order to prevent the Millard reaction. This mixture was incubated at 37 °C for 1-1.5 hours at 250 rpm in a New Brunswick™ Excella® E24/E24R incubator (Eppendorf, Hamburg, Germany).

Following incubation, the cells were spread across LB agar plates (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) sodium chloride, 1.5% (w/v) agar) containing 35 µg/ml chloramphenicol (CAM) and 100 µg/ml ampicillin (AMP). Resistance to these antibiotics was conferred by pLysS and pET11 plasmids respectively. A total of four spread plates were prepared (**Table 2.1**): two with transformed cells containing pET-11 plasmids and two with competent cells only. The diluted transformed plate contained 100 µl of the transformation reaction. For the concentrated transformed plate, the remaining transformation reaction (~500 µl) was centrifuged at 5000 × *g* for 1 min, resuspended in 100 µl MilliQ water and spread across the agar plate. For both

Table 2.1 Agar plates used for the selection of transformants.

Plate	Antibiotics	Resistance
Diluted transformed	CAM, AMP	CAM^R, AMP^R
Concentrated transformed	CAM, AMP	CAM^R, AMP^R
Competent cells only	CAM, AMP	CAM^R
Competent cells only	CAM only	CAM^R
Blank	None	None

competent cells only plates, 100 µl of untransformed cells were spread. A negative control blank plate containing agar only was also included to ensure that the agar and/or plates were not contaminated. All plates were incubated at 37 °C for 12-16 hours. A single, isolated bacterial colony was selected from one of the transformed plates and grown in LB media (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) sodium chloride) containing 35 µg/ml CAM and 100 µg/ml AMP at for 12-16 hours at 250 rpm in a New Brunswick™ Excella® E24/E24R incubator. Thereafter, 100 µl of cell culture was mixed with 100 µl glycerol (80% (v/v), autoclaved) and stored at - 80 °C.

2.2 Expression of WT and L38↑N↑L^{-4}

For the expression of WT and L38↑N↑L^{-4} genes, the T7 expression system was used. Briefly, T7 promoters found on the pET-11 vectors recognise the T7 polymerase whose gene is engineered into commercially available *E. coli* (DE3) strains. Additionally, the gene for T7 lysozyme is engineered into commercially available *E. coli* pLysS strains and prevents background expression of T7 RNA polymerase from prematurely expressing protein prior to induction. The Lac repressor protein encoded by the *lacI* gene binds the *lac* operator rendering the T7 promoter and the *lac* promoter inactive. Isopropyl β-D-1- thiogalactopyranoside (IPTG) binds to the Lac repressor and causes a conformational change that releases the repressor from the *lac* operator. This allows the host RNA polymerase to bind to the *lac* operator resulting in high numbers of T7 RNA polymerase being expressed. This T7 polymerase can then bind to the T7 promoter and consequently promote the transcription of the target gene (Smith and Johnson, 1988).

To confirm the successful transformation of *E. coli* BL21 (DE3) pLysS cells with the pET-11a and pET-11b vectors, the cDNA of the WT and L38↑N↑L^{-4} was isolated from the cells by using the commercially available GeneJet™ plasmid miniprep kit as per manufacturer's instructions (Thermo Fischer Scientific™, Massachusetts, USA). The cDNA was then sent to Inqaba Biotec (Pretoria, South Africa) to be sequenced and verified. Once the sequence was confirmed, the overexpression of WT and L38↑N↑L^{-4} was performed.

2.2.1 Overexpression of the WT protease

Initially, the transformed *E. coli* BL21 (DE3) pLysS cells containing the pET-11b vector was grown at 37 °C for 12-16 h at 200 rpm in 200 ml LB media with 35 µg/ml CAM and 100 µg/ml AMP. Thereafter, a 1:33 dilution of cell culture was done in 15 × 2 L Duran® Erlenmeyer flasks (Sigma Aldrich, Missouri, USA) with each flask containing 500 ml LB media and 35 µg/ml CAM and 100 µg/ml AMP. The cells were the left to grow at 37 °C in a LOM-400 orbital shaking incubator (MRC, London, UK) with shaking at 200 rpm until the early-log growth phase (OD_{600} ~0.4) was achieved. The protease was then overexpressed by the addition of 1 mM IPTG to each flask and allowed to grow for a period of 4 h. Thereafter, the cells were pelleted by using a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer Scientific™, Massachusetts, USA) at 5000 × *g* for 10 min, 4 °C. The resultant pellet was resuspended in 50 ml resuspension buffer (10 mM Tris, 2 mM ethylenediaminetetraacetic acid (EDTA), 1 mM phenylmethylsulfonyl fluoride (PMSF), pH 8.0) and stored at -20 °C.

2.2.2 Overexpression of L38↑N↑L^{-4}

Prior to the overexpression of L38↑N↑L^{-4}, the optimal conditions for maximum protein expression had to be determined.

2.2.2.1. Growth rate analysis

To determine the growth kinetics of the transformed *E. coli* Rosetta, T7 Express and BL21 (DE3) pLysS cells, growth curve analysis of each cell line was performed. A pre-culture containing 35 µg/ml CAM and 100 µg/ml AMP as well as transformed cells was incubated at 37 °C for 12-16 hours at 200 rpm. Thereafter, a 1:25 dilution of the overnight cell culture was done into 50 ml LB media and 35 µg/ml CAM and 100 µg/ml AMP. Growth rates were determined by recording the OD_{600} at 20 min intervals until the stationary phase was reached.

2.2.2.2. Expression trials of L38↑N↑L^{-4}

Once the optimal OD_{600} for expression was known, the induction time and IPTG concentration was determined. *E. coli* Rosetta, T7 Express and BL21 (DE3) pLysS

cells containing the pET-11a vector expressing L38↑N↑L^{-4}, was added to separate pre-cultures containing 35 µg/ml CAM and 100 µg/ml AMP and incubated at 37 °C for 12-16 hours at 200 rpm. Thereafter, a 1:25 dilution of the overnight cell culture was done into 50 ml LB media and 35 µg/ml CAM and 100 µg/ml AMP. The cells were the left to grow at 37 °C at 200 rpm until the early-log growth phase (OD_{600} ~0.5) was achieved. Once the OD_{600} was achieved, 1 mM of IPTG was added to each flask to induce overexpression of the L38↑N↑L^{-4} protease. A 1 ml aliquot was collected at various time points (0, 2, 4, 6 and 24 hours) and separated into soluble and insoluble fractions via centrifugation at 10 000 × *g* for 5 mins. The pellets were resuspended in 10 mM Tris-HCl. All samples were either stored at -20 °C until needed or electrophoresed on a tricine sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel.

2.3 Purification of WT and L38↑N↑L^{-4}

For the purification of WT and L38↑N↑L^{-4} proteins, ion-exchange chromatography (IEC) was used. Briefly, this technique exploits the surface charge of a protein which is specific to its amino acid composition and structure. Amino acids may possess either an overall negative, positive or zero-net charge depending on the pH of the surrounding environment. IEC is a form of adsorption chromatography whereby small ions attached to the stationary phase will be exchanged with biomolecules of a similar charge (Sheehan, 2009). Typically, ion-exchangers can be either cation- or anion-exchangers depending on whether they bind to cations or anions, respectively. Additionally, ion-exchangers can be further classified as either weak or strong depending on the chemistry of their functional groups. Purification of proteins is generally achieved using the weak ion-exchangers diethylaminoethyl (DEAE) or carboxymethyl (CM) columns which bind and purify proteins with a net negative or positive charge, respectively.

2.3.1 Purification of WT protease

Initially, the cell lysate was removed from -20 °C and thawed in a beaker of water at 20 °C. Once thawed, 50 µl of bovine pancreatic deoxyribonuclease I (DNase I) (10 mg/ml in milliQ water) and 250 µl of magnesium chloride (1 M) was added to 50 ml of

cell lysate and allowed to incubate at 20 °C for 30 min on a rotator. Following incubation, the bacterial cell membranes were disrupted using sonication at 70% amplitude for 8× 30 s intervals, with 30 s incubation on ice between each cycle (Figure 2.1). The cell lysate was then centrifuged at 23 000 × *g* for 30 min, 4 °C using a Sorvall LYNX 4000 Superspeed Centrifuge (ThermoFischer Scientific™, Massachusetts, USA) in order to separate the soluble and insoluble fractions. The HIV protease is expressed in inclusion bodies to prevent the protease from adopting its functional form and thereby minimising autoproteolysis. The supernatant was, therefore, discarded and the pellet containing the insoluble fraction was resuspended in wash buffer (10 mM Tris, 2 mM EDTA and 2% Triton X-100, 1 mM PMSF, pH 8.0) and was homogenised by using a Bio-Gen series PRO200 homogeniser (Pro Scientific Inc., Connecticut, USA). This mixture was then centrifuged at 23 000 × *g* for 30 min, 4 °C using a Sorvall LYNX 4000 Superspeed Centrifuge. A second wash step was performed by repeating the above process. To recover the protease from inclusion bodies, the pellet was resuspended in unfolding buffer (8 M urea, 10 mM Tris, 5 mM dithiothreitol (DTT), pH 8.0) and incubated at 20 °C for 30 min on a rotator. Following incubation, the unfolded protease was centrifuged at 23 000 × *g* for 30 min, 20 °C using a Sorvall LYNX 4000 Superspeed Centrifuge. The supernatant containing the protease was placed into SnakeSkin™ dialysis tubing (ThermoFisher Scientific™, Massachusetts, USA) with a MWCO of 3.5 kDa and refolded by dialysing against refolding buffer (0.01% (w/v) sodium azide, 10 mM formic acid containing 10% glycerol (v/v) at 4 °C for 4 h with gentle spinning. Unless otherwise stated, all dialysis steps were performed in SnakeSkin™ dialysis tubing with a MWCO of 3.5 kDa. The protease was then dialysed against equilibration buffer (10 mM sodium acetate, 0.01% (w/v) sodium azide, 2 mM DTT, pH 5.0) at 4 °C, overnight.

Prior to purification using a CM-Sepharose fast flow column (Sigma-Aldrich, Missouri, USA), the dialysate was centrifuged at 23 000 × *g* for 30 min, 4 °C using a Sorvall LYNX 4000 Superspeed Centrifuge (ThermoFischer Scientific™, Massachusetts, USA) to remove any aggregates. The 5 ml CM-Sepharose column was connected to an ÄKTAprime plus™ liquid chromatography purification system (GE Healthcare Life Sciences, Chicago, USA) and equilibrated with 5-10 column volumes of filtered equilibration buffer at a flowrate of 5 ml/min at 0.3 kPa. The supernatant was filtered through a 0.45 µM syringe filter (cellulose acetate membrane) and loaded onto the

pre-equilibrated 5 ml CM-Sepharose column at a flowrate of 2 ml/min at 0.3 kPa. To remove any contaminant or non-specifically bound proteins from the column, equilibration buffer was passed through the column until a stable baseline at A_{280} was observed. Since the theoretical pI of the protease is 9.1 (Wilkins *et al.*, 1999; Gasteiger *et al.*, 2005) and the pH of the unfolding buffer is 8.0, the net-charge of the protease is positive and therefore binds to the CM-Sepharose column which is negatively charged. Elution of the bound WT protease was performed using a salt gradient of 0-1 M NaCl and 3 ml fractions were collected and pooled. Lastly, the pooled protease was dialysed against storage buffer (10 mM sodium acetate, 0.01% (w/v) sodium azide, pH 5.0) at 4 °C overnight (Figure 2.1). Unless stated otherwise, all downstream experiments were performed in storage buffer.

2.3.2 Purification of L38↑N↑L^{-4} protease

The purification of the L38↑N↑L^{-4} protease was performed using a modified protocol published by Sherry *et al.* (2020) which uses a two-step purification method (Figure 2.2). Initially, the cell lysate was removed from -20 °C and thawed in a beaker of water at 20 °C. Once thawed, 50 µl of DNase I (10 mg/ml in milliQ water) and 250 µl of magnesium chloride (1 M) was added to 50 ml of cell lysate and allowed to incubate at 20 °C for 30 min on a rotator. Following incubation, the bacterial cell membranes were disrupted using sonication at 70% amplitude for 8x 30 s intervals, with 30 s incubation on ice between each cycle. The cell lysate was then centrifuged at 23 000 xg for 30 min, 4 °C using a Sorvall LYNX 4000 Superspeed Centrifuge (ThermoFischer Scientific™, Massachusetts, USA) in order to separate the soluble and insoluble fractions. The supernatant was discarded and the pellet containing the insoluble fraction was resuspended in wash buffer (10 mM Tris, 2 mM EDTA and 2% Triton X-100, 1 mM PMSF, pH 8.0) and homogenised using a Bio-Gen series PRO200 homogeniser (Pro Scientific Inc., Connecticut, USA). This mixture was then centrifuged at 23 000 × *g* for 30 min, 4 °C using a Sorvall LYNX 4000 Superspeed Centrifuge. A second wash step was performed by repeating the above process. To recover the protease from inclusion bodies, the pellet was resuspended in unfolding buffer (8 M urea, 10 mM Tris, 5 mM DTT, pH 8.0) and incubated at 20 °C for 30 min

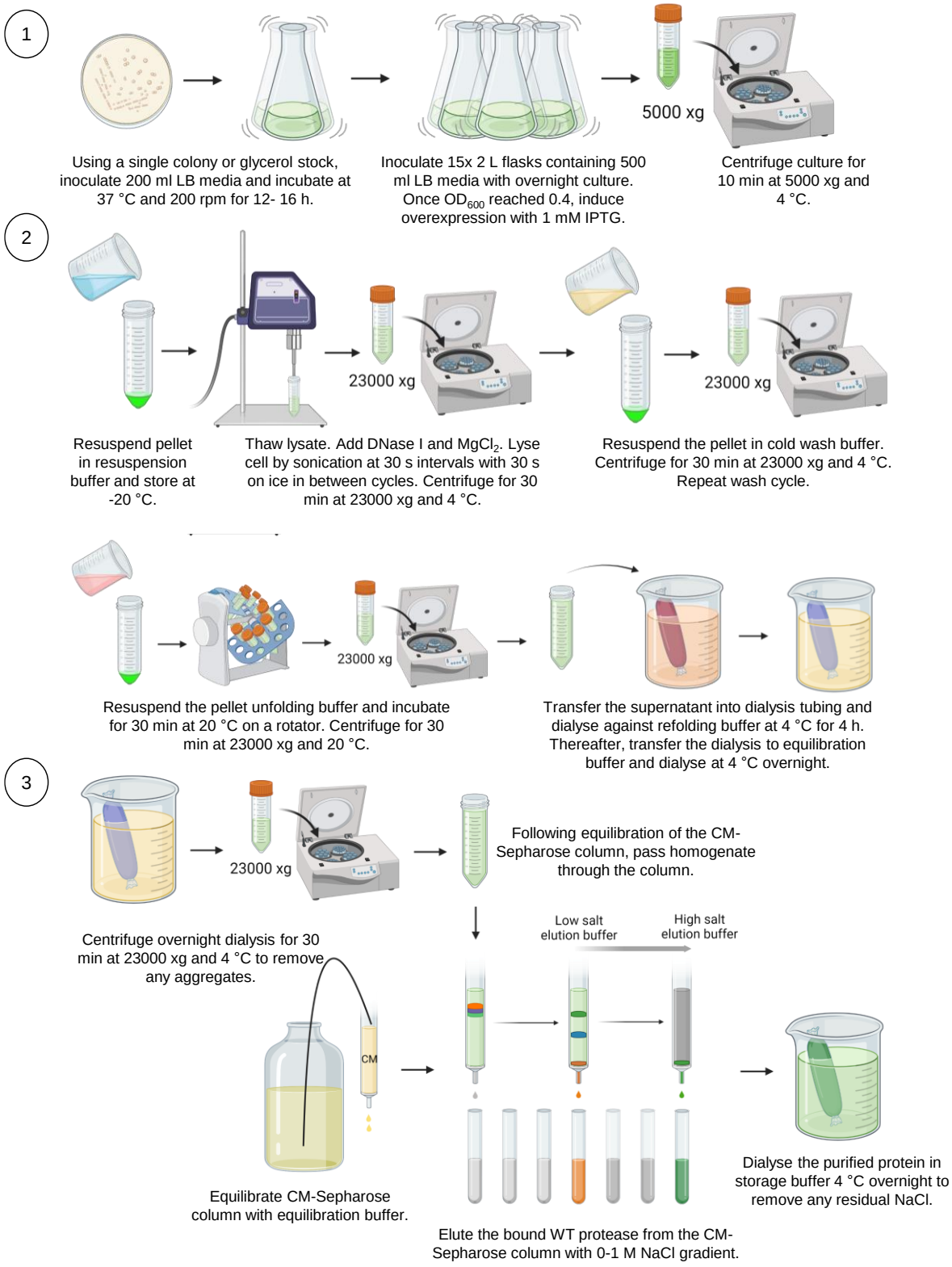


Figure 2.1 Expression and purification of the WT protease.

Step 1 illustrates the expression of the protease. Steps 2 illustrates the washing, unfolding and refolding of the lysate. Step 3 illustrates purification using cation exchange chromatography. Image created with BioRender.com.

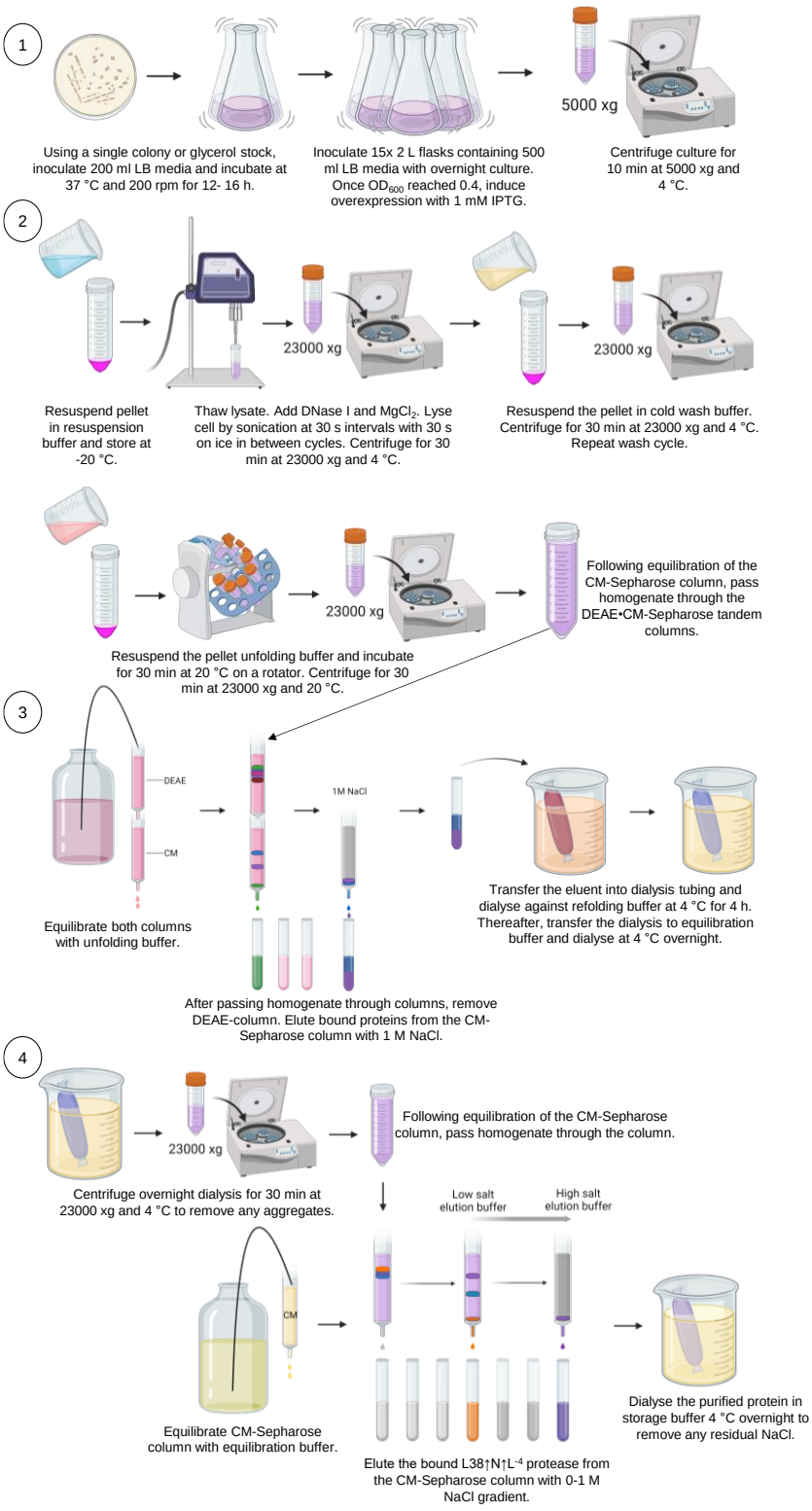


Figure 2.2 Expression and purification of the L38↑N↑L^{-4} protease.

Step 1 illustrates the expression of the protease. Steps 2 illustrates the lysis, washing and unfolding of the lysate. Step 3 illustrates the first purification with both anion and cation exchange chromatography. Step 4 illustrates the second purification with cation exchange chromatography. Image created with BioRender.com.

on a rotator. Following incubation, the unfolded protease was centrifuged at 23 000 × *g* for 30 min, 20 °C using a Sorvall LYNX 4000 Superspeed and filtered through a 0.45 µM syringe filter (cellulose acetate membrane). A 5 ml DEAE-column connected in tandem to a 5 ml CM-sepharose column was connected to an ÄKTAprime plus™ liquid chromatography purification system (GE Healthcare Life Sciences, Chicago, USA) and equilibrated with 5-10 column volumes of filtered unfolding buffer at a flowrate of 5 ml/min at 0.3 kPa. The filtered supernatant was loaded onto the equilibrated DEAE- and CM-sepharose columns at 2 ml/min at 0.3 kPa. To remove any contaminant or unbound proteins from the column, unfolding buffer was passed through the column until a stable baseline at A_{280} was observed. As previously mentioned, the protease is positively charged in unfolding buffer (pH 8.0) and, therefore, passes through the positively charged DEAE-column while binding to the negatively charged CM-Sepharose column. To avoid eluting any bound contaminant proteins from the DEAE-column, the column was removed prior to elution of the protease. The bound L38↑N↑L^{-4} protease was removed from the CM-sepharose column using a one-step elution with 1 M NaCl and the collected fraction was diluted 1:5 with unfolding buffer. The pooled protease was dialysed against refolding buffer (0.01% (w/v) sodium azide, 10 mM formic acid containing 10% glycerol (v/v) at 4 °C for 4 hours with gentle stirring. The protease was then dialysed against equilibration buffer (10 mM sodium acetate, 0.01% (w/v) sodium azide, 2 mM DTT, pH 5.0) at 4 °C overnight.

Prior to purification using a CM-Sepharose fast flow column, the dialysate was centrifuged at 23 000 × *g* for 30 min, 4 °C using a Sorvall LYNX 4000 Superspeed Centrifuge to remove any aggregates. The 5 ml CM-Sepharose column was connected to an ÄKTAprime plus™ liquid chromatography purification system (GE Healthcare Life Sciences, Chicago, USA) and equilibrated with 5-10 column volumes of filtered equilibration buffer at a flowrate of 5 ml/min at 0.3 kPa. The supernatant was filtered through a 0.45 µM syringe filter (cellulose acetate membrane) and loaded onto the pre-equilibrated 5 ml CM-Sepharose column at a flowrate of 2 ml/min at 0.3 kPa. To remove any contaminant or unbound proteins from the column, equilibration buffer was passed through the column until a stable baseline at A_{280} was observed. Since pH of the equilibration buffer is 5.0, the net-charge of the protease is positive and, therefore, binds to the negatively charged CM-Sepharose column. Elution of the bound

L38↑N↑L^{-4} protease was performed using a salt gradient of 0-1 M NaCl and 3 ml fractions were collected and pooled. Lastly, the pooled protease was dialysed against storage buffer (10 mM sodium acetate, 0.01% (w/v) sodium azide, pH 5.0) at 4 °C overnight. Unless stated otherwise, all downstream experiments were performed in storage buffer.

2.4 Concentration and storage of purified proteins

The collected fractions were concentrated to a final concentration of 40-45 µM using an Amicon® stirred cell (MilliporeSigma, Massachusetts, USA) that was fitted with a 3 kDa ultrafiltration membrane (MilliporeSigma, Massachusetts, USA). The protein was aliquoted into 500 µl fractions and stored at -20 °C for a maximum of one week without flash freezing.

2.5 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

It is imperative to continue all downstream experiments with pure protein samples, therefore the purity of the eluted proteins was evaluated during the purification process using SDS-PAGE. SDS-PAGE separates a mixture of proteins using an electric field which is applied to a polyacrylamide gel (Laemmli, 1970). The polyacrylamide gel matrix acts as a sieve and allows the separation of protein molecules according to their molecular weight. Since various proteins may have the same molecular weight but different native conformations, the anionic detergent SDS is used to denature and linearise proteins while conferring an overall negative charge to the protein (Laemmli, 1970). This ensures that the intrinsic charge of the protein is masked and migration of the sample through the gel is only influenced by the size of the molecule. There are various methods of SDS-PAGE with glycine (Laemmli, 1970) and tricine (Schagger, 2006) being the two most used, differing in the type of buffering systems used. Tricine SDS-PAGE was performed as described in Schagger (2006) as tricine allows for better resolution of proteins smaller than 30 kDa.

The purity of the purified proteases was verified on a 16% (w/v) separating gel and a 4% stacking gel. All protein samples were mixed with reducing sample buffer (10% (w/v) SDS, 30% (w/v) glycerol and 2% (v/v) 6% (w/v) beta-mercaptoethanol, 0.005%

Coomassie blue G-250 and 150 mM Tris-HCl, pH 7.5) in a 1:1 ratio. Thereafter, protein samples were boiled at 95 °C for 5 min and subsequently centrifuged to ensure the entire sample was at the bottom of the tube. Samples were electrophoresed in anode (1 M Tris, pH 8.9) and cathode (1 M Tris, 1 M Tricine and 1% (w/v) SDS, pH 8.25) buffer by using a Mini-PROTEAN® Tetra vertical electrophoresis cell system (Biorad, California, USA).

The gel was then removed from the system and stained using stain solution (0.25% (w/v) Coomassie Brilliant Blue R-250, 10% (v/v) glacial acetic acid and 50% (v/v) methanol) for 1-2 hours. This was followed by destaining in destain solution (10% (v/v) glacial acetic acid and 50% (v/v) methanol) for 8-12 hours. The relative migration distance (R_f) of the molecular weight marker protein standard (ThermoFisher Scientific, Massachusetts, USA) was determined using R_f = migration distance of the protein / migration distance of the dye front and plotted against their corresponding logarithmic molecular weight. The molecular weights of the target proteins were determined from the linear regression model based on the distance migrated.

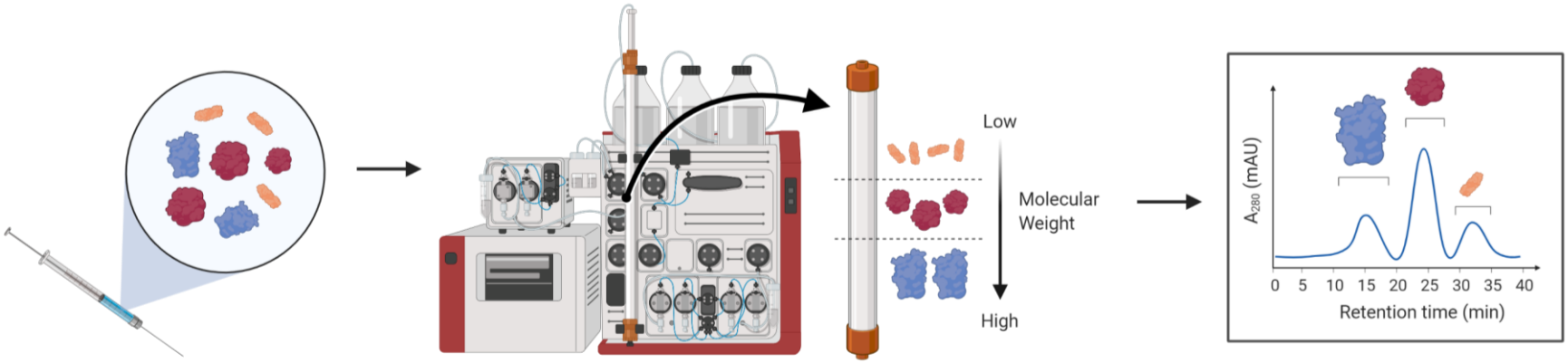
2.6 PROTEIN CONCENTRATION DETERMINATION

Proteins in solution absorb ultraviolet (UV) light at 280 nm due to aromatic amino acid residues such as tyrosine, tryptophan, and cysteine (Schmid, 2001) thus allowing the concentration of proteins in solution to be determined. The protein concentration for the target proteins were determined by absorbance spectroscopy using a Jasco V-630 Spectrophotometer (JASCO, USA) in accordance with the Beer-Lambert Law (Wilson and Walker, 2010). The molar absorptivity coefficients used for WT were 25 230 $M^{-1}cm^{-1}$ under non-reducing conditions or 24 980 $M^{-1}cm^{-1}$ under reducing conditions and the molar absorptivity coefficients used for L38↑N↑L^{-4} were 25 230 $M^{-1}cm^{-1}$ under non-reducing conditions or 24 980 $M^{-1}cm^{-1}$ under reducing conditions as predicted by the Expert Protein Analysis System (ExPASy) ProtParam tool (Wilkins *et al.*, 1999; Gasteiger *et al.*, 2005). For each target protein, a dilution series (1:10, 1:12, 1:14, 1:16 and 1:20) was set up using protein sample and storage buffer and the absorbance was measured in triplicate across a wavelength of 240-400 nm. The final protein concentration was corrected for light scatter and baseline noise (Winder and Gent, 1971).

The data obtained from the WT and L38↑N↑L^{-4} absorbance spectra were compared to those obtained using a NanoDrop™ One (Thermo Fisher Scientific, USA). Since the results obtained did not differ significantly (<5%), subsequent concentration determination was performed on the NanoDrop™ One (Thermo Fisher Scientific, USA) as it required a smaller volume of sample.

2.7 Size exclusion-high pressure liquid chromatography

In order to determine the oligomeric state of the purified proteases size exclusion-high pressure liquid chromatography (SE-HPLC) was performed on a TSKgel SuperSW2000 SEC column that had a resolution range of 5-150 kDa (Merck, Australia). Size exclusion chromatography (SEC) is a technique that separates molecules based on their molecular size with the larger molecules eluting first (Sheehan, 2009) (Figure 2.3). Combining SEC with high pressure pumps allows the mobile phase to move at a high flowrate through the column, ensuring optimal resolution, rapid separation and minimised band broadening (Sheehan, 2009; Wilson and Walker, 2010). The SEC column was attached with a TSKgel SWXL guard column (Merck, Australia) to a Shimadzu ultrafiltration liquid chromatography SPD-20A SE-HPLC system. The columns were equilibrated at 20 °C with salt buffer (10 mM sodium acetate, 0.01% (w/v) sodium azide, 200 mM NaCl, pH 5.0) using a flowrate of 0.2 ml/min and ~900 psi. Following equilibration, 20 µl of the purified protein (30 µM) was injected onto the column with a run time of 30 min. A set of known protein standards containing (thyroglobulin bovine (670 kDa), γ-globulin (150 kDa), albumin (44.3 kDa), ribonuclease A/myoglobin (13.7/17 kDa) and *para*-aminobenzoic acid/vitamin B (0.14/1.4 kDa) was passed through the column using the same parameters. Thereafter, a standard curve was constructed by plotting the logarithm of the molecular weight of the standards as a function of their corresponding retention times. The molecular weight and oligomeric state of the proteases could thus be determined using their corresponding retention times. The molecular size of the WT and L38↑N↑L^{-4} proteases, as predicted by the ExPASy ProtParam tool (Wilkins *et al.*, 1999; Gasteiger *et al.*, 2005), were 21.5 and 22 kDa, respectively.



Purified protein sample is loaded onto HPLC system.

Protein populations are separated by molecular weight.

Absorbance is measured for all eluted fractions and plotted to determine sizes.

Figure 2.3 SE-HPLC procedure.

Purified protein is added to a SE-HPLC system and separated according to the molecular weights. Adapted from “Protein Purification by Size-Exclusion Chromatography”, by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates.

2.8 FAR-UV CIRCULAR DICHROISM SPECTROPOLARIMETRY

Circular Dichroism (CD) is a spectroscopic technique in which the differential absorption of right- and left-handed circularly polarised light produced by chiral molecules is measured. CD is used to study protein structure since the 20 standard amino acids, except for glycine, are chiral molecules and confer chirality to the peptide bonds in the backbone of proteins. In far-UV CD, *n* to π^* and π to π^* are the dominant electronic transitions of peptide bonds that interact with circularly polarised light and are centred around 222 nm and 190 nm, respectively (Woody and Koslowski, 2002). These far-UV spectra produce specific patterns indicating secondary structural elements when interacting with the peptide backbone by reflecting the various phi (Φ) and psi (Ψ) angles. Proteins which are predominantly α-helical produce troughs at around 208 nm and 222 nm, and a peak near 190 nm (Woody and Koslowski, 2002; Greenfield, 2006). On the other hand, predominantly β-sheeted proteins produce a trough at approximately 218 nm and a peak close to 195 nm (Woody and Koslowski, 2002; Greenfield, 2006).

Far-UV CD was used to evaluate the secondary structures of the WT and L38↑N↑L^{-4} proteases. The spectra were recorded using a Jasco J-1500 spectropolarimeter connected to a Peltier temperature controller (Jasco, Pfungstadt, Germany) set to 20 °C. In a Suprasil® quartz cuvette with a path length of 1 mm (Hellma®, Müllheim, Germany), 10 μM of the target protein was measured using the continuous scanning mode with a bandwidth of 1 nm, data pitch of 0.2 nm and over a wavelength range of 180-250 nm. The resulting spectra were averaged over 10 accumulative scans with contributions of the buffer subtracted from the collected data and subsequently converted to mean residue ellipticity (MRE) in deg·cm^2·dmol^{-1} using the equation (Woody, 1995):

$$\theta_{\mathrm{MRE}} = \frac{\theta\ (\mathrm{mdeg}) \times 10^6}{\mathrm{pathelength}\ (\mathrm{mm}) \times [\mathrm{protein}](\mu\mathrm{M}) \times n}$$

where *n* is the number of residues in the target protein.

2.9 PROTEIN CRYSTALLISATION

X-ray crystallography is a method used to determine the structure of a protein from the molecular properties of its related crystal. Protein crystallisation is achieved by determining the correct buffers, precipitants, concentration of protein, temperature and other additives through various crystal trials. Once the optimal conditions are acquired, a crystal is formed by the precipitation of macromolecules or salts from the solution. The crystal is exposed to an X-ray beam and the resulting angles of diffraction and beam intensities are measured to accurately determine a three-dimensional representation of the protein.

Vapour diffusion, either hanging or sitting drop, is commonly used for the growth of protein crystals. A micro-droplet containing protein and precipitant solution is mixed and placed either on a glass coverslip for hanging drop or the well of a sitting drop plate. The coverslip is placed onto the hanging drop plate well which seals the mixture with the concentrated precipitant solution. Since the precipitant concentration within the drop is lower that the well, the sealed system allows for an equilibrium between the micro-droplet and the reservoir precipitant solution to be established through vapour diffusion. The concentration of the protein will increase as water diffuses from the micro-droplet until a nucleation event occurs which is required for the growth of a protein crystal (Rupp, 2009).

Crystallisation of the WT and L38↑N↑L^{-4} proteases, as an apo-form and in the presence of PIs was performed using the Hampton Research Index^{TM} screening kit (Hampton Research, California, USA) and the vapour diffusion method. Prior to setting up crystal trials, the purified proteases were further concentrated to approximately 3 mg/ml of active protein. Crystal trials were set up using the Oryx8^{TM} protein crystallisation robot (Douglas Instruments, East Garston, UK) in Qiagen EasyXtal^{TM} 15-well hanging-drop plate system (Qiagen®, Hilden, Germany) or Chryschem 24-well sitting drop plates (Hampton Research, California, USA). The wells of both plates contained 500 µl of reservoir screening solution.

2.10 ISOTHERMAL TITRATION CALORIMETRY

Isothermal titration calorimetry (ITC) is a technique used to quantify the thermodynamic parameters of a biomolecular reaction by studying its accompanying heat transfer (Sheehan, 2009). ITC allows for the quantification of high affinity binding constants within the range of 10^8-10^9 M^{-1} without any additional labels such as radioisotopes or fluorophores. By analysing interactions at a constant temperature, ITC can be used to obtain thermodynamic parameters such as dissociation constant (K_D), stoichiometry (n), change in enthalpy (ΔH), change in entropy ($T\Delta S$) and the change in Gibbs free energy (ΔG) (Figure 2.4) (Freyer and Lewis, 2008).

Therefore, ITC can be used evaluate the drug binding thermodynamics of enzyme-ligand reactions by measuring the heat that is absorbed or released during interactions (Tellinghuisen, 2004). A single ITC run produces the association constant (K_a), ΔH and n. The ΔG can then be determined using the following formula:

$$\Delta G = \mathrm{RT} ln\,(K_D)$$

where R = the universal gas constant (8.3145 J/mol/K) and T = the absolute temperature (in kelvin) (Andujar-Sanchez *et al.*, 2006). The ΔS can then be determined as follows:

$$\Delta G = \Delta H - \mathrm{T}\Delta S$$

2.10.1 ACTIVE SITE CONCENTRATION

Since the HIV-1 protease is prone to autolysis, it is important to quantify the concentration of active enzyme in a purified sample. To determine the active site concentration of the purified WT and L38↑N↑L^{-4} proteases, ITC was performed using an aspartyl protease inhibitor, acetyl pepstatin, which binds relatively weakly to HIV proteases. ITC experiments were performed in storage buffer at 20 °C using a NanoITC standard volume instrument (TA® Instruments, New Castle, USA) with each experiment requiring at least 2 ml of protein and 500 µl of ligand. A ligand into storage buffer titration was performed to monitor the heats of dilution (HOD). Protein and ligand samples were degassed under a vacuum of 0.3–0.5 atm for 15 min. Thereafter, 200 µM of acetyl

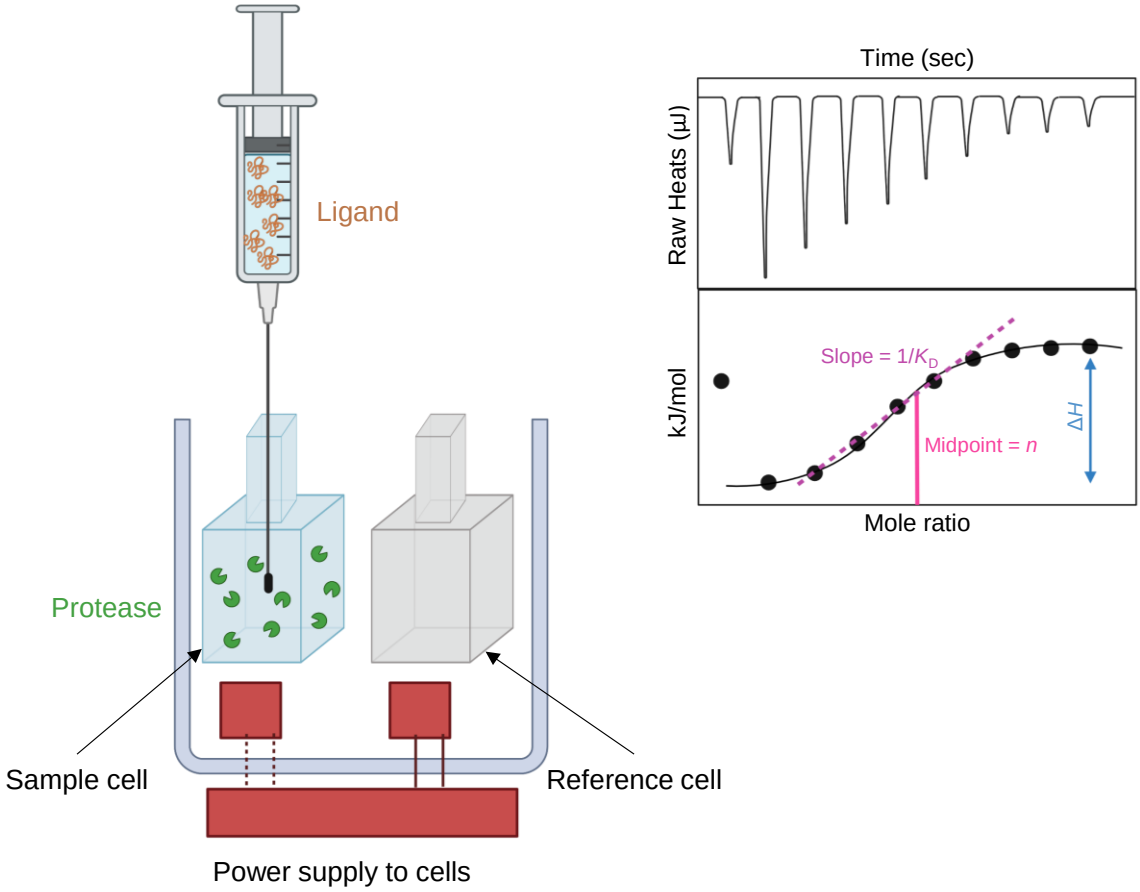


Figure 2.4 Basic principle of ITC.

A calorimeter contains two cells – reference cell (grey) and sample cell (blue). The reference cell receives a constant supply of power (solid red lines). As the ligand is titrated into the sample cell containing protease, heat is released or absorbed as the reaction occurs. Power is supplied to the sample cell to maintain the temperature at the same temperature as the reference cell containing storage buffer (dashed red lines). The data output is used to create a thermogram where the area under the curve for each injection is used to calculate the raw heats. This is plotted against the mole ratio and a model is fitted to determine the thermodynamic parameters of the reaction. Image created with BioRender.com.

pepstatin was titrated (5 µl injections) into a solution of 20-25 µM protease with a stir rate of 200 rpm, an initial and final baseline of 200 s and a duration period of 300 sbetween each injection. The HOD were subtracted from the resulting data set, the baseline was corrected and the change in heats were fitted to an independent binding model using the NanoAnalyze™ software. The percentage active protease in each sample was determined from the *n*-value with a value of 1 representing 100% active enzyme in sample preparations indicating all protease molecules are in the active form and no self-cleavage occurred. The *n*-value is, therefore, used as a correction factor for the concentration of active protease in a purified sample.

2.10.2 DISPLACEMENT TITRATION CALORIMETRY

To accurately determine the binding energetics of PIs which have high binding affinities (nM-pM range) (Velazquez-Campoy *et al.*, 2001). Displacement titration calorimetry (DTC) experiments were performed. During DTC, a ligand that binds with lower affinity to the protease, acetyl pepstatin, is displaced by the higher affinity ligand, PI (Velazquez-Campoy and Freire, 2006). The drug binging energetics of ATV, DRV, RTV or SQV were individually determined for both the WT and L38↑N↑L^{-4} proteases.

Prior to DTC, protein samples (30 µM) were incubated with 100 µM acetyl pepstatin and 1% (v/v) DMSO at 20 °C under a vacuum of 0.3–0.5 atm for 20-25 min. Since stock PIs were dissolved in 100% (v/v) DMSO, the protease-pepstatin mixture contained DMSO to avoid any buffer mismatch. Incubation with the aspartyl PI allowed for the equilibration of the mixture as well as the saturation of active sites with acetyl pepstatin. Thereafter, DTC experiments were performed at 20 °C using a NanoITC low volume instrument (TA® Instruments, New Castle, USA) in storage buffer with each experiment requiring at least 800 µl of protein and 250 µl of ligand. Briefly, 100 µM of PI (ATV, DRV, RTV or SQV) was titrated (4 µl injections) into a solution of ~30 µM protease with a stir rate of 100 rpm, an initial and final baseline of 200 s and a duration period of 300 s between each injection The HOD were subtracted from the resulting data set, the baseline was corrected and the change in heats were fitted to an independent binding model using the NanoAnalyze™ software in order to determine binding energetics of the PIs.

2.11 Differential scanning calorimetry

Differential Scanning Calorimetry (DSC) is a powerful tool used to determine the folding and stability of proteins (Bruylants *et al.*, 2005). It is capable of directly measuring the heat energy released or absorbed by a sample as it is cooled or heated at a constant rate as a function of temperature. During the unfolding of a protein, a heat capacity change (ΔC_p) is observed (Bruylants *et al.*, 2005). The sample and reference cells are maintained at the same temperature and the change in heat capacity observed is measured and plotted as a function of time or temperature.

The heat capacity function of the apo WT and L38↑N↑L^{-4} proteases and PI-complexed proteases were measured as a function of temperature using the Nano-DSC microcalorimeter (TA Instruments, Delaware, USA). Prior to DSC experiments, the protein samples (~30 µM) and reference solutions (storage buffer) were degassed under a vacuum of 0.3–0.5 atm for 15 min to avoid bubble formation during sample loading. For PI-complexed experiments, ~30 µM protease was incubated with 100 µM PI (ATV, DRV, RTV or SQV) at 20 °C under a vacuum of 0.3–0.5 atm for 20-25 min. The apo or PI-complexed proteases were heated linearly from 20-130 °C at a rate of 1 °C/min with the reference cell containing storage buffer.

Data were analysed using the NanoAnalyze™ software package according to the best fitted model. The WT data were fitted using a two-state mode and sigmoidal baseline. On the other hand, L38↑N↑L^{-4} was fitted using a two-state scaled model (1st order) and sigmoidal baseline for apo whereas for PI-complexed data were fitted using a two-state scaled model and polynomial baseline (3rd order).

2.12 Enzyme kinetics

To determine whether the insertions affected the kinetic properties of the HIV-1 protease the specific activity, k_{cat} and catalytic efficiency (k_{cat}/K_M) of the WT and L38↑N↑L^{-4} was determined. The hydrolysis of a fluorogenic substrate (Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2) which mimics the conserved capsid/p2 cleavage site within the HIV-1 Gag polyprotein was monitored (Figure 2.5). The Phe(NO_2) quenching amino group reduces the fluorescence emission of the fluorogenic 2-aminobenzoyl group (Abz). Cleavage of the Nle-Phe(NO_2) peptide bond by the

protease eradicates the quenching effect and allows Abz to fluoresce (Carmel and Yaron, 1978; Szeltner and Polgar, 1996). For all kinetic measurements, samples were exposed to an excitation wavelength of 337 nm and an increase in emission was detected at an emission wavelength of 425 nm at 20 °C. All assays were performed in triplicate using a Jasco FP-6300 spectrofluorometer (Molecular Dimensions, USA) and reaction buffer (50 mM sodium acetate, 1 M NaCl, pH 5).

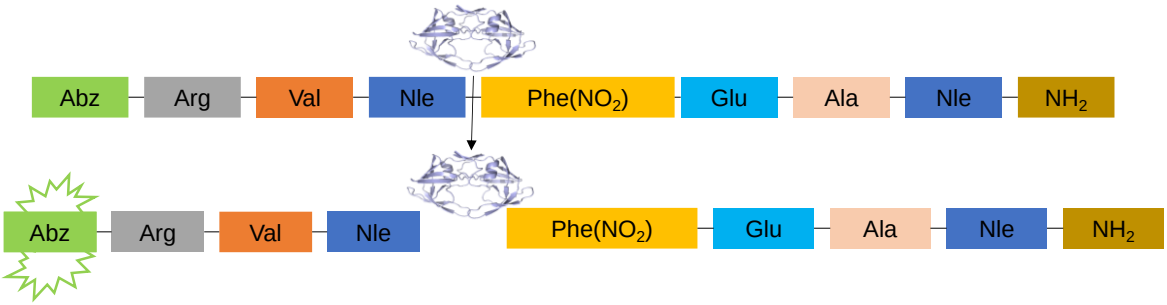


Figure 2.5 Cleavage of fluorogenic substrate by the HIV-1 protease.

The protease recognises and cleaves the bond between Nle-Phe(NO_2) allowing the Abz group to fluoresce. Image created with Microsoft PowerPoint, version 2204.

The specific activity and k_{cat} were determined using a constant substrate concentration of 50 µM and a range of active enzyme amounts from 1-10 pmol. The initial velocity versus amount of protease was plotted and the slope was determined from linear progress curves which represents the specific activity. k_{cat} can be determined by plotting the initial velocity against the amount of protease in µmol. The data was then fitted using the linear equation to determine k_{cat} values:

$$v_0 = V_{max} = k_{cat}\,[ES]$$

where v_0 = initial velocity, V_{max} = maximum velocity of enzyme, k_{cat} = catalytic turnover and [ES] = total concentration of assumed enzyme-substrate complex.

The catalytic efficiency (k_{cat}/K_M) was determined using a constant active enzyme concentration of 50 nM and a range of substrate concentrations from 1-10 µM.

2.13 MOLECULAR DYNAMICS SIMULATIONS

All atom molecular dynamics (MD) simulations is the standard method used to simulate the dynamic motion of protein molecules *in silico* (Kukol, 2008; Monticelli and Tieleman, 2013). MD simulations were performed on the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases using the Desmond software version 11.2 that was operated on LINUX architecture. All MD simulations were performed on two 3.68 GHz Intel core i7 5960x computers. The OPLS3 force field was used for all three models (Harder *et al.*, 2016). Using the TIP3 solvation model of the System Builder model, simulations were performed with explicit solvent in a cubic box universe (Mark and Nilsson, 2002). The cubic box universe was selected as previously defined whereby the protein was positioned 10 Å from the edges of a cubic box (Naicker *et al.*, 2013; Zondagh *et al.*, 2018). Following solvation, the systems were relaxed through energy minimisation by bringing the temperature to equilibrium using the constant number of particles, volume and temperature (NVT) ensemble. Thereafter, the system was left to equilibrate for 100 ps as the temperature increased from 10 K and plateaued at 300 K. The pressure of the system was then brought to equilibrium by using the constant number of particles, pressure and temperature (NPT) ensemble. The system was left to equilibrate for 100 ps until the density of the system was stable over time (Abraham *et al.*, 2017). Once the system was equilibrated, MD simulations were performed over 50 ns with data collected over 100 frames, every 0.05 ns. Analysis of the resultant trajectories was performed using the Simulation Quality Analysis Module in Maestro (Maestro, Schrodinger LLC 2020, USA).

2.14 INDUCED-FIT DOCKING STUDIES

Molecular docking is a method used to study protein-ligand interactions during which the induced-fit docked model is often applicable. Binding of a ligand to the protein's active site causes conformational changes in the side chain and/or backbone of the protein. These conformational changes allow the active site to fit closely to the shape of the ligand and is referred to as flexible docking. Induced-fit docking studies (IFDS) generates an optimised conformation of the protein-ligand complex which can then be used to assess the binding interaction and related data (Dror *et al.*, 2012; Du *et al.*, 2016).

IFDS studies were performed using the Glide software (Schrödinger LLC 2009, USA). The protease models were generated using PyMOL software (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC). Homology models of the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases were constructed using the coordinates of a high-resolution HIV-1 protease structure solved at 2.15 Å (PBD: 2HB4). This model was chosen as it was of higher resolution, 2.71 Å (PDB:3U71), to the available subtype C protease at the time. The models were generated using the built-in PyMOL mutagenesis plug-in. All three models contained the following mutations: T12S, I15V, L19I, M36I, R41K, H69K, L89M and I93L. The L38↑N↑L^{-4} and L38↑N↑L^{+4} models contained the addition of ASN and LEU at position 38. In addition, L38↑N↑L^{+4} contained the following mutations: K20R, E35D, R57K and V82I. All three models were validated using PROCHECK (Laskowski *et al.*, 1996).

Initially, the proteases were pre-processed using the Glide, Prime and Protein Preparation Wizard modules. The Protein Preparation Wizard was used to assign the correct bond orders and to add hydrogen atoms to the proteins. Thereafter, the solvent orientation was optimised and water molecules were removed. In addition, the Epik Prediction software tool (Shelley *et al.*, 2007; Greenwood *et al.*, 2010) was used to assign ionisable residues with a charge in accordance with an experimental pH of 5. The model was then subjected to energy minimisation by using the OPLS3 force-field (Harder *et al.*, 2016) until an average RMSD of 0.3 Å was achieved. This is the default RMSD threshold in the Protein Preparation Wizard module ensuring the optimal constraint of the receptor protein for IFDS by relaxing strained angles, bonds and clashes (Schrödinger LLC 2009, USA).

The two-dimensional structure of the ligands used for IFDS were obtained from PubChem (HTTPS://PUBCHEM.NCBI.NLM.NIH.GOV/). These included the four PIs used within this study: (i) ATV (CID: 148192), (ii) DRV (CID: 213039), (iii) RTV (CID: 392622) and (iv) SQV (CID: 60787). The LigPrep tool (Schrödinger LLC 2009, USA) was used to generate the ionisation state of the ligands at pH 5 by using Epik. Prior to energy minimisation of the ligand, hydrogens were added and bond lengths and angles were corrected.

IFDS were then performed using the energy minimised protein and ligand structures. The Asp25 residues at the active site were surrounded by a 20 Å cubic box to define

the region where IFDS would occur. During IFDS, residues of the protein that were within 5 Å of the ligand poses were refined and their corresponding side chains were optimised. Since Glide docking does not allow for receptor flexibility, the van der Waals radii of non-polar residues can be scaled to minimise the penalty scores for close contacts thereby providing some minor flexibility in the receptor and ligand. Consequently, a scaling factor of 0.5 was selected for the van der Waals radii of the protein and ligand during the docking. The IFDS simulations were performed using the OPLS3 forcefield (Harder *et al.*, 2016) and 20 poses were generated during Glide docking. The best docked structure was selected based on the top IFDS score for further investigation using MD simulations.

The GlideScore (G_{score}) is an empirical scoring function that can be used to evaluate the success of IFMD simulations. The G_{score} counts the number of various types of interactions between two binding partners and aims to replicate experimental affinity data (Guedes *et al.*, 2018) with a more negative value representing an interaction with a stronger binding affinity.

3 RESULTS

3.1 CONFIRMATION OF WT AND L38↑N↑L^{-4} SEQUENCES

The cells containing pET vectors expressing either WT or L38↑N↑L^{-4} were cultured overnight and the plasmids were extracted using a miniprep kit. The extracted plasmids were sequenced and compared to the C-SA (PDB: 3U71) sequence. As expected, the WT sequence matched the C-SA protease while the L38↑N↑L^{-4} sequence contained the insertions ASN and LEU at position 38 (Figure 3.1). Both pET plasmids were designed to contain the Q7K mutation which decreases auto-proteolysis without affecting the kinetic properties of the protease (Louis *et al.*, 1991).

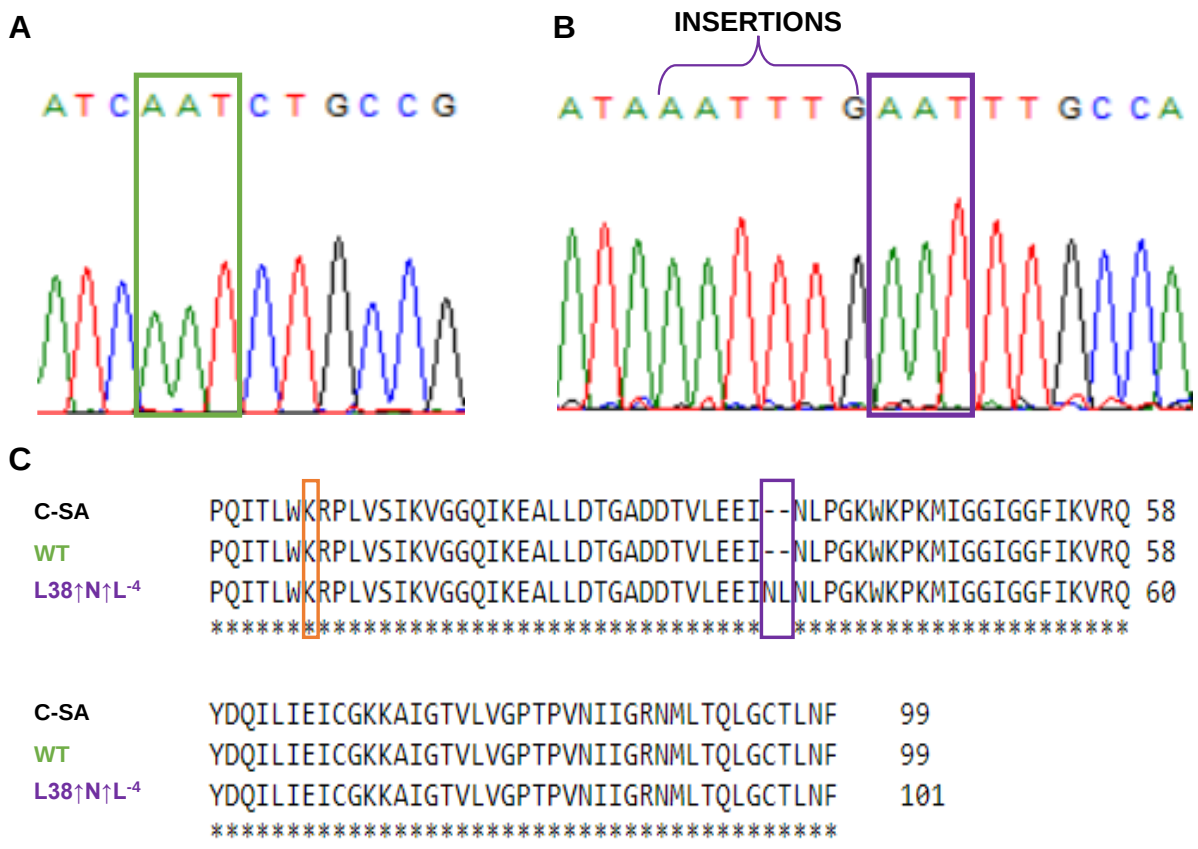


Figure 3.1 Sequencing results of WT (A) and L38↑N↑L^{-4} (B) plasmid inserts and alignment (C) with the C-SA (PDB:3U71) nucleotide sequence.

Figure legend continues on next page.

The green and purple blocks in A and B represent the original ASN codons at position 38 for both the WT and L38↑N↑L^{-4}, respectively. The ASN and LEU insertions for L38↑N↑L^{-4} are shown in B with a brace on top of the corresponding codons. Both the WT and L38↑N↑L^{-4} DNA sequences were translated to amino acids using ExPASy translate (Gasteiger *et al.*, 2005) and aligned with the C-SA protease amino acid sequence using NCBI BLAST (blast.ncbi.nlm.nih.gov). The double insertion in L38↑N↑L^{-4} is indicated by the purple box while the Q7K mutation is indicated by the orange box.

3.2 Overexpression of WT protease

In order to obtain a sufficient quantity of the WT protease to perform downstream characterisations, the protease was overexpressed in *E. coli* BL21 (DE3) pLysS cells for 4 hours at 37 °C with 1 mM IPTG. Prior to purification, the cell lysate was separated into insoluble and soluble fractions (Figure 3.2B, lanes 13 and 14). The insoluble fraction was then subjected to washes, urea unfolding and refolding by dialysis (Figures 3.2B, lanes 6-12). The overnight dialysis was then loaded onto a CM-Sepharose column and eluted using a 0-1 M NaCl gradient (Figure 3.2 A). All protease successfully bound to the column (Figures 3.2 B, lane 5) and the eluted protease was collected in the second peak (Figure 3.2 B, lane 2). The purified protease was then concentrated using an Amicon® stirred cell (Figure 3.2 B, lane 1).

3.3 Overexpression of L38↑N↑L^{-4}

Since the overexpression conditions for the L38↑N↑L^{-4} protease was not known, experiments had to be performed to determine the optimal conditions for expression.

3.3.1 Growth rates of L38↑N↑L^{-4}

To determine the optimal OD_{600} in which to induce maximal overexpression of the protease, 50 ml cultures of Rosetta, T7 and BL21 (DE3) pLysS cells containing the

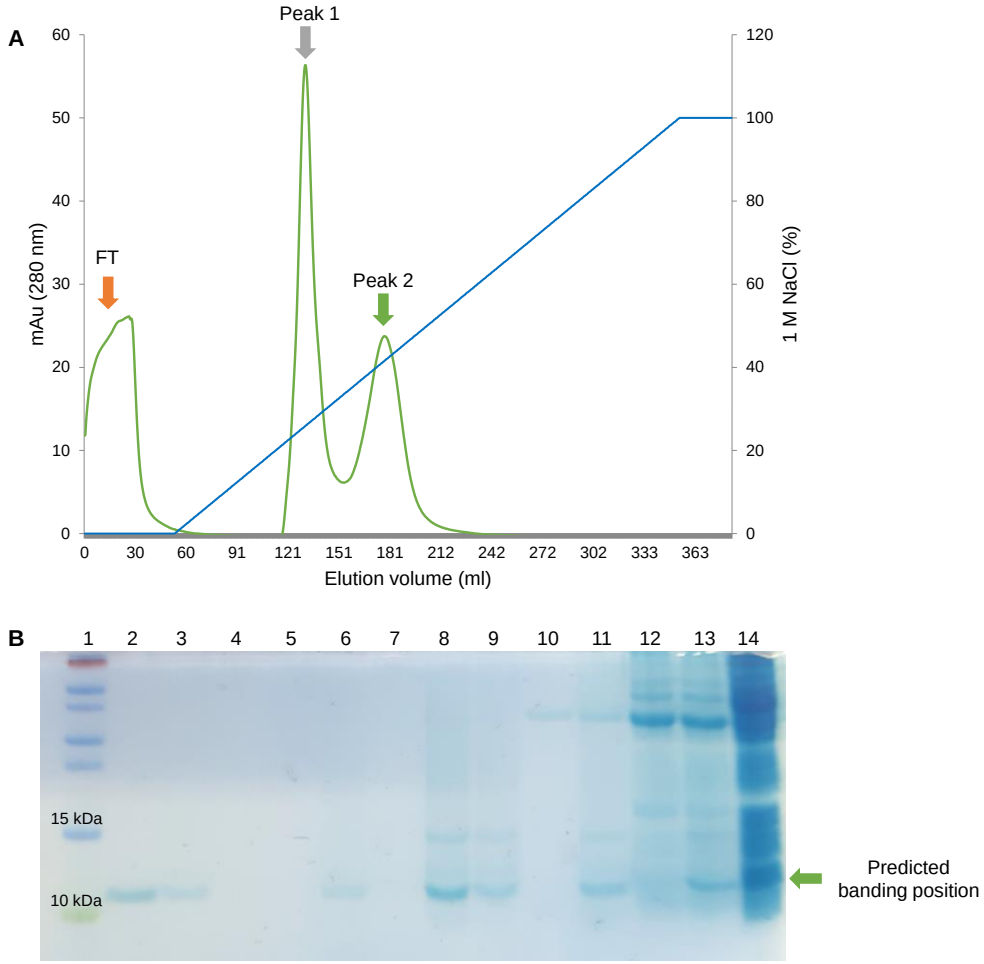


Figure 3.2 Cation exchange chromatography of WT protease.

A: The cation exchange chromatography elution profile of the WT protease. The blue line represents the 0-1 M NaCl gradient (%) while the green line represents the absorbance (mAu). The orange, grey and green arrows point to the flowthrough, peak 1 and peak 2, respectively. **B:** Tricine (16%) SDS-PAGE gel of samples collected throughout purification. Lane 1: Molecular weight marker, Lane 2: Concentrated protease, Lane 3: Peak 2 (green arrow), Lane 4: Peak 1 (grey arrow), Lane 5: Flowthrough (orange arrow), Lane 6: Dialysis, Lane 7: Unfolded pellet, Lane 8: Unfolded supernatant, Lane 9: Second wash pellet, Lane 10: Second wash supernatant, Lane 11: First wash pellet, Lane 12: First wash supernatant, Lane 13: Pellet and Lane 14: Supernatant. The green arrow in Figure 3.2B points to the WT protease throughout the purification. The PageRuler™ Prestained Protein Ladder, 10 to 180 kDa, was used.

pET-11a vector was grown and absorbance values were recorded over time. The early log phase, OD_{600} 0.4-0.8, of these cultures was determined to occur before 2 hours (Figure 3.3) and was used to perform induction trials.

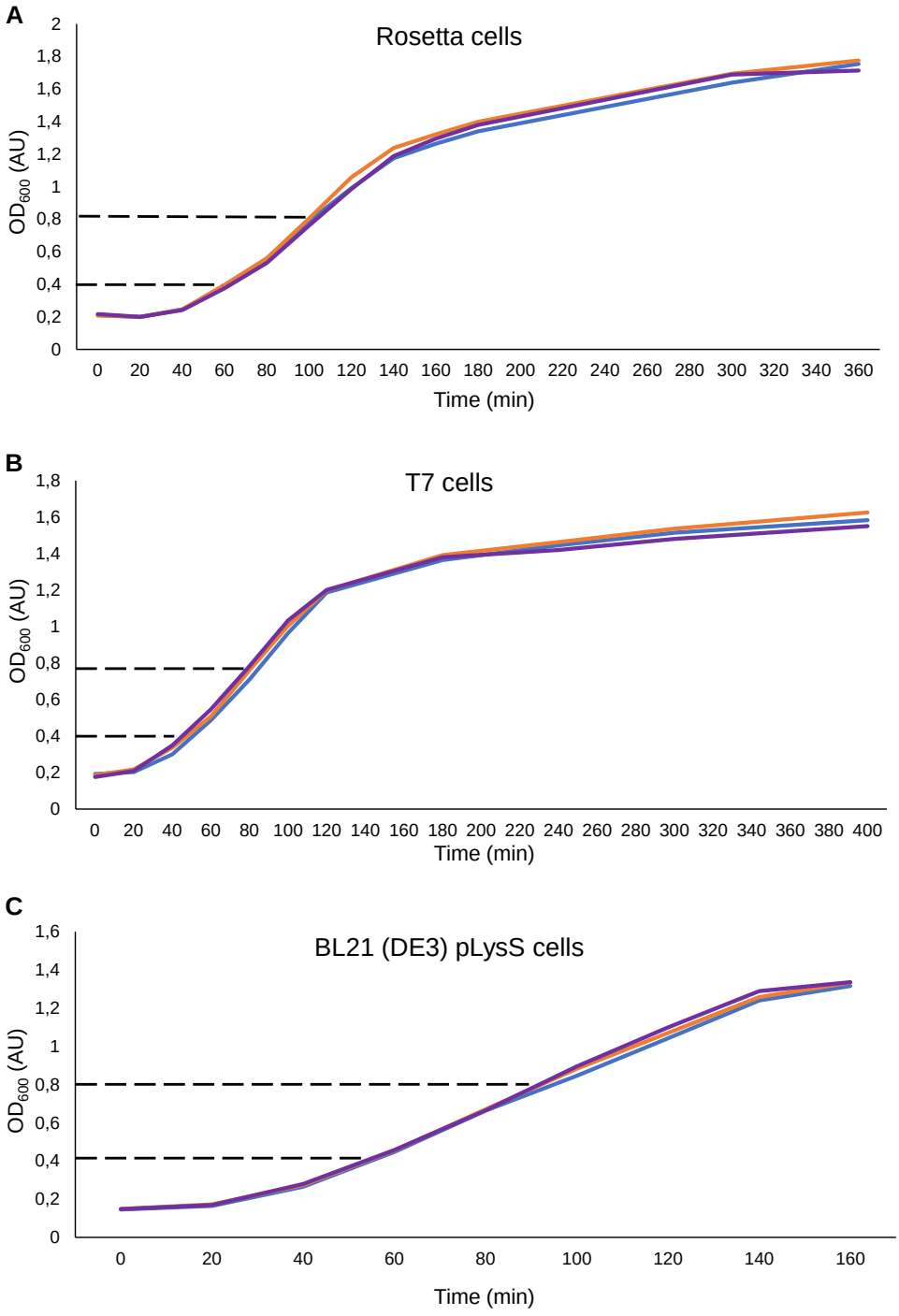


Figure 3.3 Growth curves of *E. coli* Rosetta (A), T7 (B) and BL21 (DE3) pLysS (C) cell lines expressing the L38↑N↑L^{-4} protease.

The black dashed lines indicate the early-log phase of each culture.

3.3.2 EXPRESSION TRIALS OF L38↑N↑L^{-4}

Expression trials were based on the WT overexpression conditions and were conducted at 37 °C with the addition of 1 mM IPTG at an OD_{600} of 0.5 over a period of 24 hours. Samples were collected at 0, 4, 6 and 24 hours for the Rosetta and T7 cell lines (Figure 3.4A and B) and 0, 2, 4, 6 and 24 hours for the BL21 (DE3) pLysS cell line (Figure 3.4C). No overexpression of the L38↑N↑L^{-4} protease occurred in the Rosetta and BL21 (DE3) pLysS cell lines. However, expression of the L38↑N↑L^{-4} protease can be seen in the pelleted fraction of the T7 cell line (Figure 3.4 B).

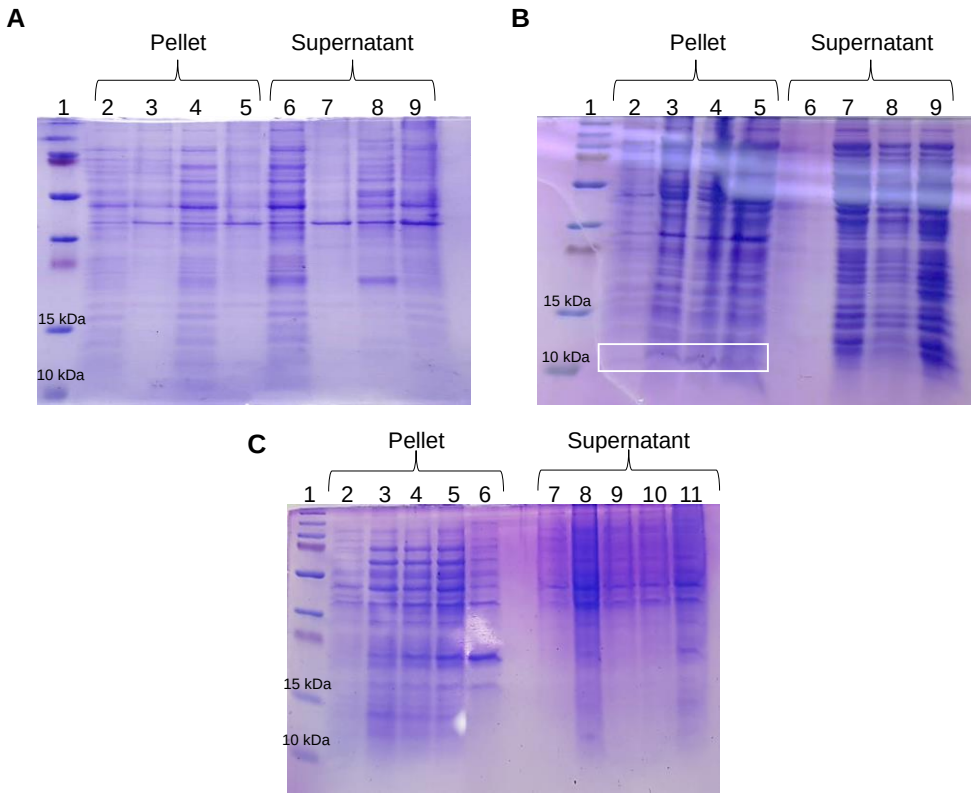


Figure 3.4 Induction trials of the L38↑N↑L^{-4} protease in different *E. coli* cell lines.

A: SDS-PAGE gel of Rosetta cell line. Lane 1: Molecular weight marker, Lanes 2 and 6: 0 hours, Lanes 3 and 7: 4 hours, Lanes 4 and 8: 6 hours and Lanes 5 and 9: 24 hours. **B**: SDS-PAGE gel of T7 cell line. Lane 1: marker, Lanes 2 and 6: 0 hours, Lanes 3 and 7: 4 hours, Lanes 4 and 8: 6 hours and Lanes 5 and 9: 24 hours. The white box indicates protease overexpression. **C:** SDS-PAGE gel of BL21 (DE3) pLysS cell line. Lane 1: marker, Lanes 2 and 7: 0 hours, Lanes 3 and 8: 2 hours, Lanes 4 and 9: 4 hours, Lanes 5 and 10: 6 hours and Lanes 6 and 11: 24 hours.

To avoid any bias down the line and since the overexpression of the WT was carried out in the BL21 (DE3) pLysS cell line, further induction studies were performed on this cell line. Expression trials were conducted at 37 °C with varying concentrations of IPTG at and OD_{600} of 0.4 over a period of 4 hours (Figure 3.5). The optimal conditions for the L38↑N↑L^{-4} protease was determined to be 1 mM IPTG for 4 hours at 37 °C in the BL21 (DE3) pLysS cell line.

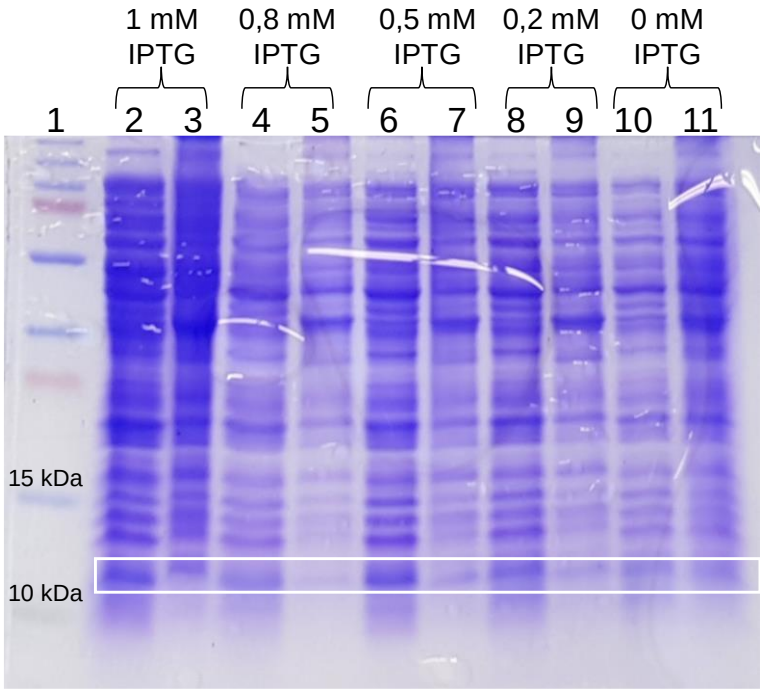


Figure 3.5 Tricine (16%) SDS-PAGE gel of BL21 (DE3) pLysS cells expressing L38↑N↑L^{-4}.

Lane 1: Molecular weight marker, Lanes 2, 4, 6, 8 and 10: pellet and Lanes 3, 5, 7, 9 and 11: supernatant. The white box indicates expression of protease.

3.3.3 Purification of the L38↑N↑L^{-4} protease

Following overexpression of the L38↑N↑L^{-4} protease, the cell lysate was separated into insoluble and soluble fractions (Figure 3.6 C, lanes 8 and 9). Prior to the first purification, the insoluble fraction was then subjected to washes and urea unfolding. The unfolded protein (Figure 3.6 C, lane 7) was loaded onto a tandem DEAE-CM-Sepharose column setup (Figure 3.6 A) and eluted from the CM-Sepharose column

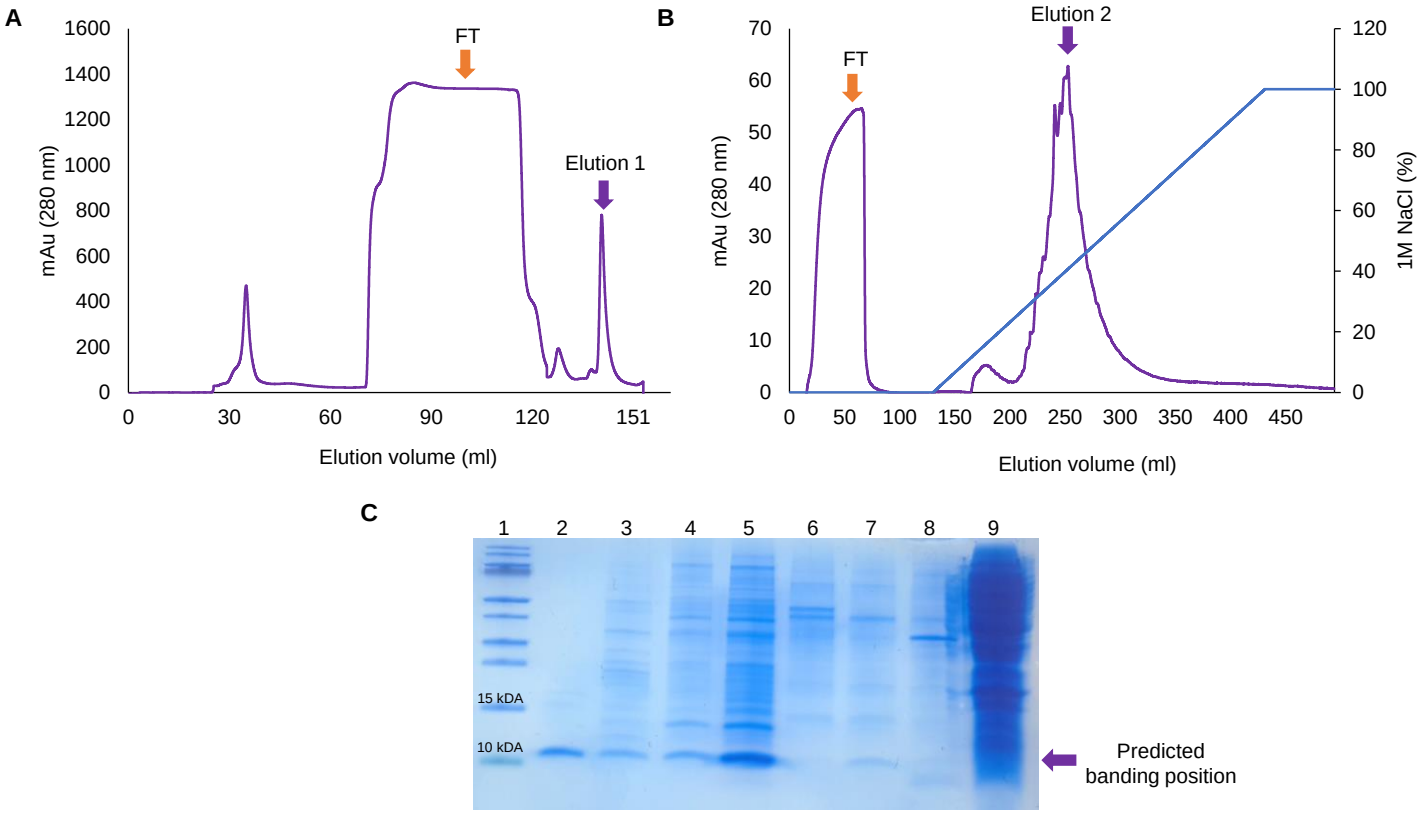


Figure 3.6 Ion-exchange chromatography of the L38↑N↑L^{-4} protease.

A: The IEC elution profile of the L38↑N↑L^{-4} protease. The purple line represents the absorbance (mAu). The orange and purple arrows point to the flowthrough and eluted protein, respectively. **B:** The cation exchange chromatography elution profile of the L38↑N↑L^{-4} protease. The purple line represents the absorbance (mAu) while the blue line represents the 1 M NaCl gradient (%). The orange and purple arrows point to the flowthrough and eluted protein, respectively. **C:** Tricine (16%) SDS-PAGE gel of samples collected throughout purifications. Lane 1: Molecular weight marker, Lane 2: elution 2, Lane 3: flowthrough 2, Lane 4: overnight dialysis, Lane 5: elution 1, Lane 6: flowthrough 1, Lane 7: unfolded protein, Lane 8: pellet and Lane 9: supernatant. The purple arrow in Figure 3.6C points to the L38↑N↑L^{-4} protease throughout the purification. The PageRuler™ Prestained Protein Ladder, 10 to 180 kDa, was used.

using a 1 M NaCl (Figure 3.6 C, lane 5). Proteins were dialysed overnight at 4 °C to allow for refolding. The overnight dialysis (Figure 3.6 C, lane 4) was then loaded onto a CM-Sepharose column and eluted using a 1 M NaCl gradient (Figure 3.6 B). The eluted L38↑N↑L^{-4} protease was collected (Figure 3.6 C, lane 2) and the purified protease was concentrated using an Amicon® stirred cell.

3.4 Determination of protein molecular mass

The molecular mass of the purified WT and L38↑N↑L^{-4} proteases were approximated using a molecular weight marker (Figure 3.7). Both gels (WT and L38↑N↑L^{-4}) produced a single band indicating that the samples were free from contamination. The size of the WT and L38↑N↑L^{-4} protease was determined to be 10.7 and 10.9 kDa, respectively which corresponds to the predicted monomer sizes determined by ExPASy ProtParam (Appendix 1 and Appendix 2) (Gasteiger *et al.*, 2005).

3.5 Protein concentration determination

3.5.1 Absorbance Spectra

The concentrations of the WT and L38↑N↑L^{-4} proteases were determined via absorbance spectroscopy and nanodrop readings. An absorbance spectrum of the proteases was measured from 230 – 400 nm using a Jasco V-360 spectrophotometer and a standard curve of the dilution factor vs the absorbance at 280 nm was plotted (Figure 3.8). The absorbance measurements at 280 nm were used to calculate the protease concentrations as this corresponds to the maximum absorbance of tryptophan residues in the protein samples. Using the Beer-Lambert law, the concentrations of the WT and L38↑N↑L^{-4} proteases were determined to be 46 and 45 µM, respectively.

Concentrations of the proteases were compared and confirmed using the Nanodrop (**Table 3.1**). The concentrations for the WT and L38↑N↑L^{-4} proteases were 0.98 and 1 mg/ml, respectively, which translated to 46 µM. This correlated with the concentrations calculated through the absorbance spectra. Additionally, the A_{280}/A_{260} ratio of 1.7 for WT and 1.73 for L38↑N↑L^{-4} indicating that the samples were free of DNA contamination.

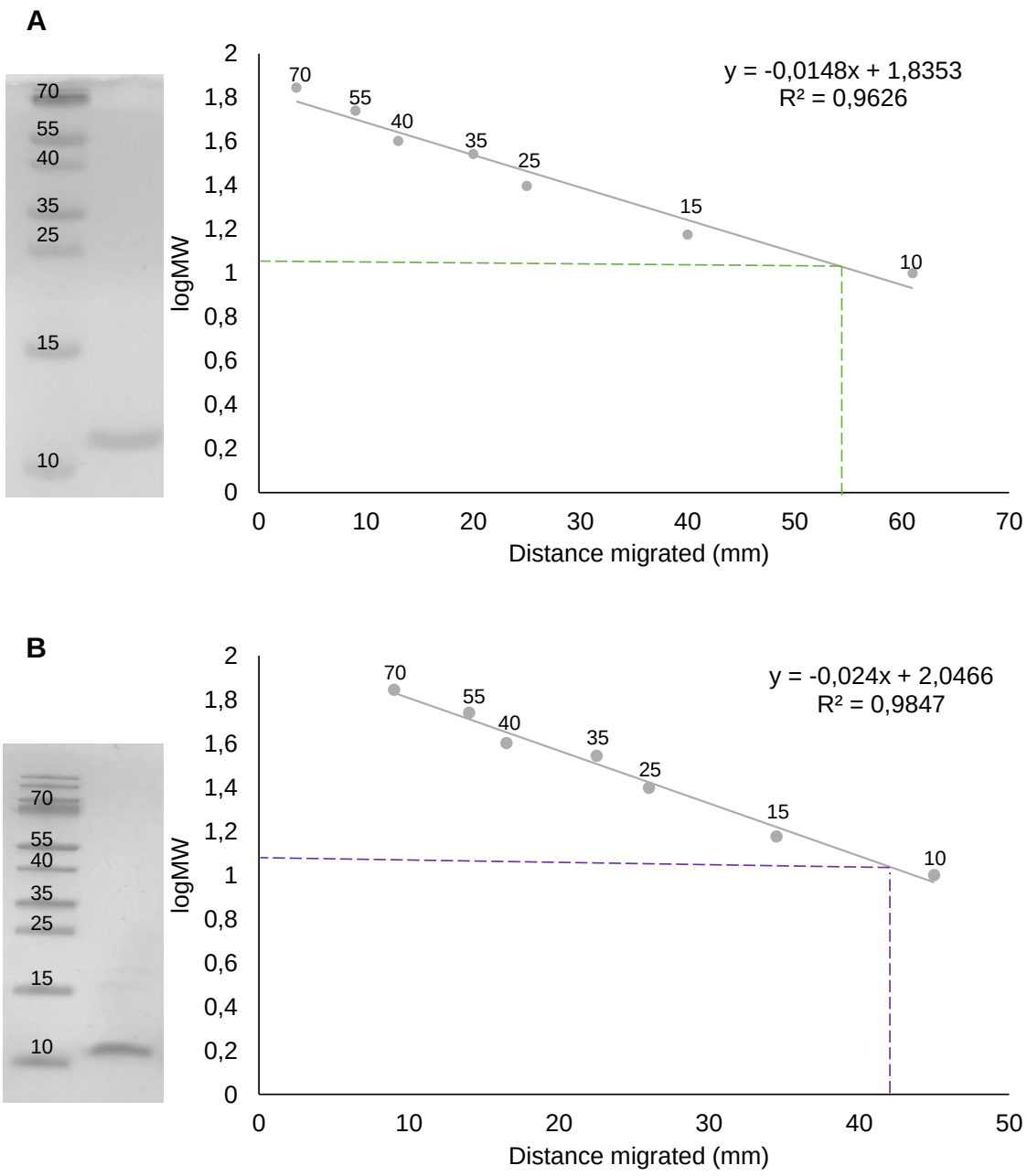


Figure 3.7 Tricine (16%) SDS-PAGE standard curves used to calculate the molecular mass of the WT and L38↑N↑L^{-4} proteases.

A: A sample of the purified WT protease alongside the molecular weight marker. **B:** A sample of the purified L38↑N↑L^{-4} protease alongside the molecular weight marker. The molecular mass of each standard in the marker is shown on the gel with the corresponding points shown on the graph. The PageRuler™ Prestained Protein Ladder, 10 to 180 kDa, was used.

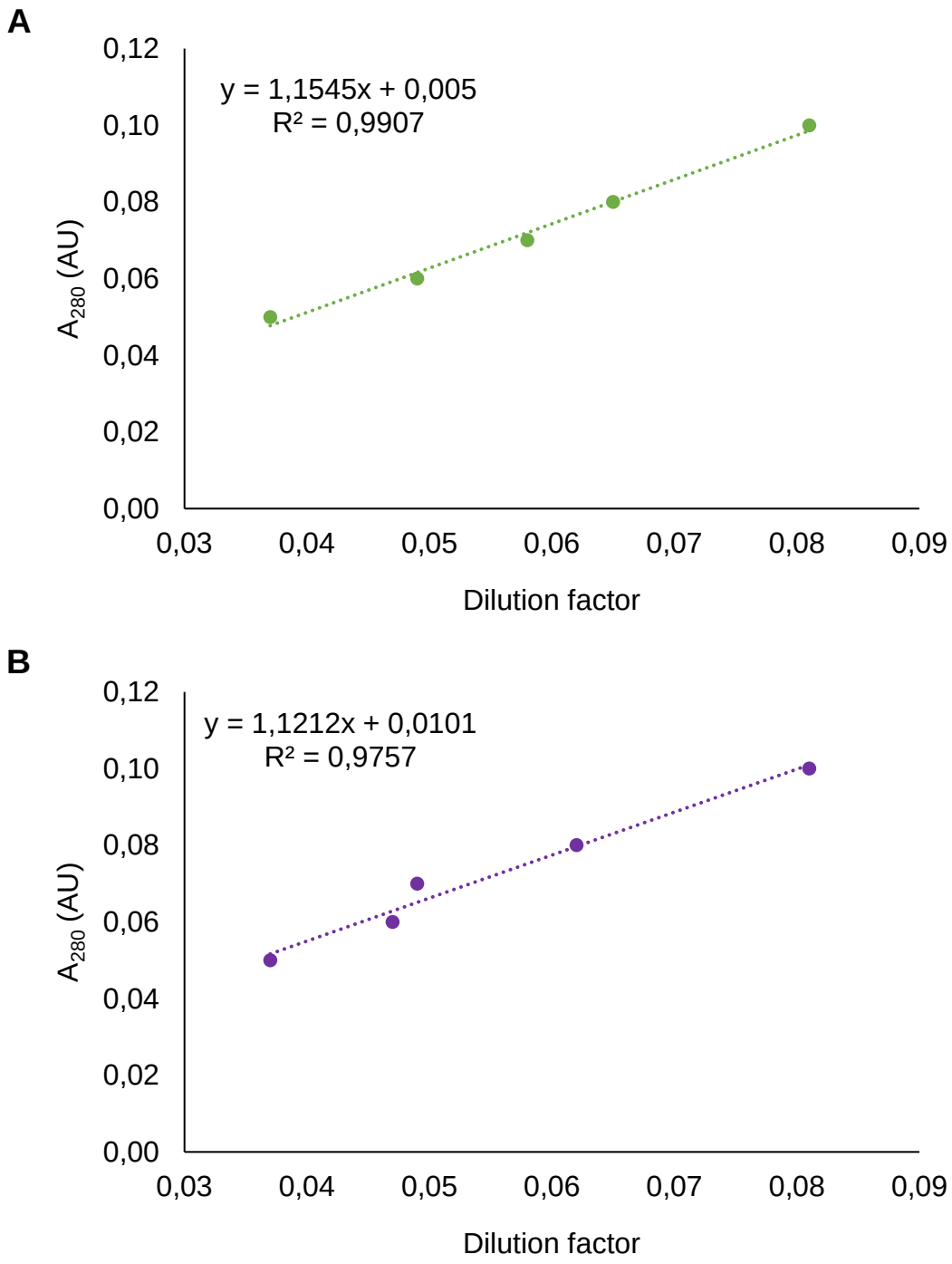


Figure 3.8 Standard curves generated from absorbance spectrum readings for dilution series (1:10, 1:12, 1:14, 1:16 and 1:20).

A: Standard curve of WT absorbance spectrum dilution series. **B:** Standard curve of L38↑N↑L^{-4} absorbance spectrum dilution series

Table 3.1 Nandrop values for WT and L38↑N↑L^{-4}.

Protein	A_{280}/A_{260}	Concentration	
		mg/ml	μM
WT	1.7	0.98	46
L38↑N↑L^{-4}	1.73	1	46

However, since the HIV-1 protease is prone to autoproteolysis and A_{280} values do not differentiate between native and misfolded or degraded protein, the Beer-Lambert law is insufficient in determining protein concentration. Therefore, it is imperative to determine the percentage of active enzyme in the protein sample before continuing to downstream experiments.

3.5.2 Active site determination

To determine the amount of correctly folded protease in solution, active site titrations using ITC and the aspartyl protease inhibitor, acetyl pepstatin, were performed on the WT and L38↑N↑L^{-4} proteases (Figure 3.9). The reported stoichiometry of acetyl pepstatin binding to the HIV-1 protease has previously been determined as 1:1 and can, therefore, be used to determine the concentration of active protease within a sample. By using the binding stoichiometry (*n*-value) produced during the titration of acetyl pepstatin into the HIV-1 protease, corrections for misfolded or degraded protein can be applied to the sample. Upon titration, the percentage of the WT and the L38↑N↑L^{-4} protease in the active conformation was 62% and 50%, respectively. Using these *n*-values, the concentrations of the proteases could be corrected for misfolding and/or degradation (**Table 3.2**). The corrected concentrations for the proteases for WT and L38↑N↑L^{-4} were 28.5 and 23 μM, respectively.

The free energies between the WT and L38↑N↑L^{-4} were relatively unchanged, with these interactions found to be -35.7 and -32.74 kJ/mol, respectively. The Δ*H* and Δ*S* values for the WT was 51.37 kJ/mol and 297 J/mol.K while L38↑N↑L^{-4} was 11.94 kJ/mol and 152.4 J/mol.K. While these values differed between the WT and L38↑N↑L^{-4} proteases, the overall result was similar with both producing positive Δ*H* and Δ*S* values.

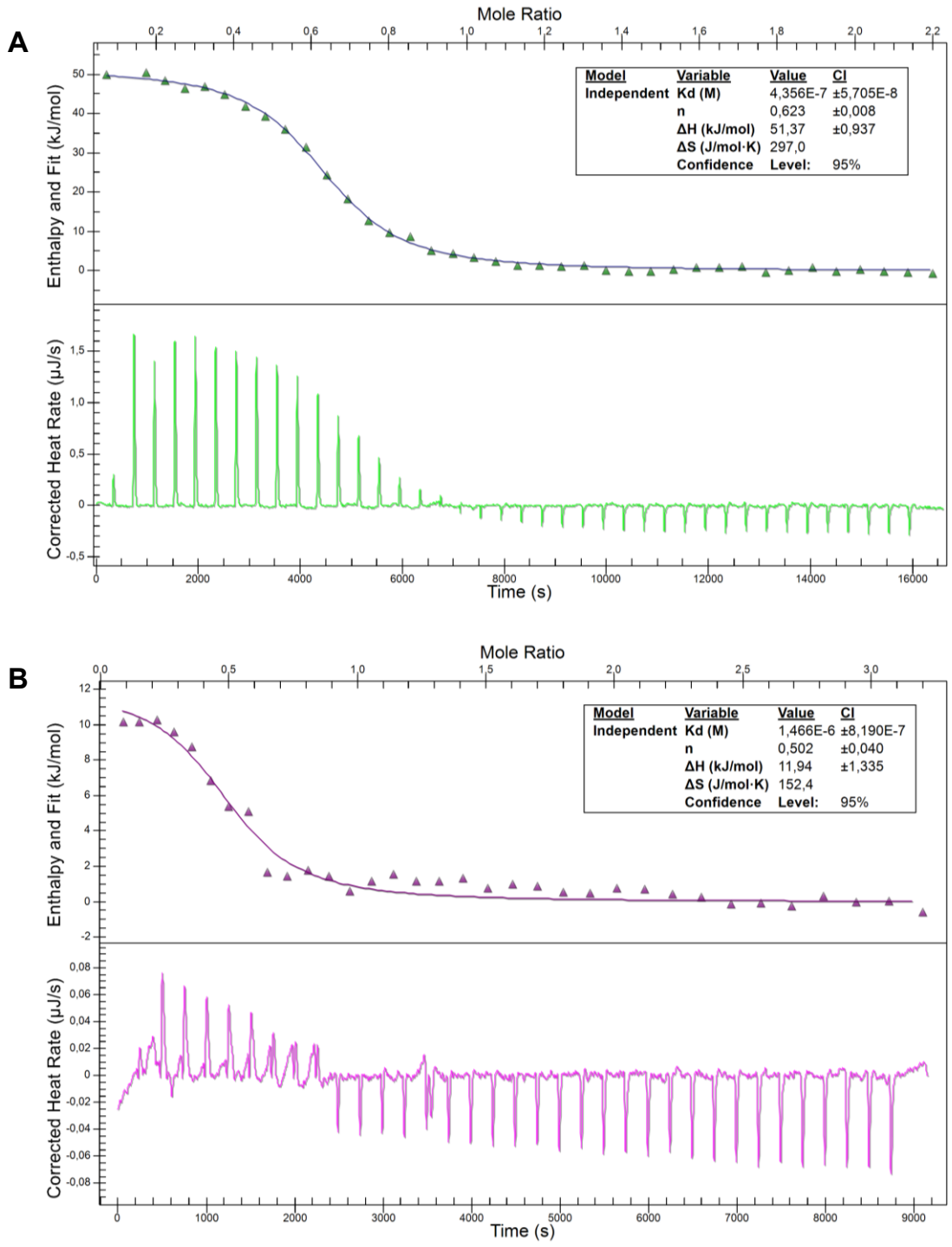


Figure 3.9 Active site titrations of WT (A) and L38↑N↑L^{-4} (B) proteases.

ITC titrations were performed at 20 °C using 200 µM of acetyl pepstatin and 46 µM of the respective protease. Injections of 5 µl were introduced into the protease in the sample cell with a duration of 300 seconds between each injection. The data were baseline corrected, HODs were subtracted and subsequently fitted using an independent model in the NanoAnalyze™ software.

Table 3.2 Corrected concentrations for WT and L38↑N↑L^{-4}.

Protein	[Initial] (μM)	*n*-value	[Corrected] (μM)
WT	46	0.62	28.5
L38↑N↑L^{-4}	46	0.5	23

The K_D values for the WT and L38↑N↑L^{-4} were 0.43 and 1.5 μM, respectively, indicating moderate binding between the inhibitor and proteases. However, it should be noted that the binding of L38↑N↑L^{-4} was slightly weaker in comparison to the WT as represented by the larger K_D value.

3.6 Structural characterisation of WT and L38↑N↑L^{-4}

Structural characterisations (far-UV CD and analytical SE-HPLC) were performed on the WT and L38↑N↑L^{-4} proteases in order to determine if any structural changes occurred with the protease.

3.6.1 Secondary structure

Far-UV CD was used to assess the secondary structures of the proteases. According to the crystal structure of the WT protease (3U71), the protease is a predominantly β-sheeted protein. Predominantly β-sheeted proteins will display a trough at approximately 218 nm and a peak close to 195 nm (Woody and Koslowski, 2002; Greenfield, 2006). Both the WT and L38↑N↑L^{-4} proteases displayed troughs at 218 nm although these spectra differed slightly (Figure 3.10). It is possible that the observed difference in the amount of active protease between the WT and L38↑N↑L^{-4} (62% vs 50%) contributed to the differences observed in the CD spectra. This may have resulted in the decreased signal observed with L38↑N↑L^{-4}, as well as the absence of a peak at 195 nm which can clearly be seen in the WT. However, the results obtained were in agreement with the apo C-SA crystal structure with both the WT and L38↑N↑L^{-4} proteases being predominantly β-sheeted proteins.

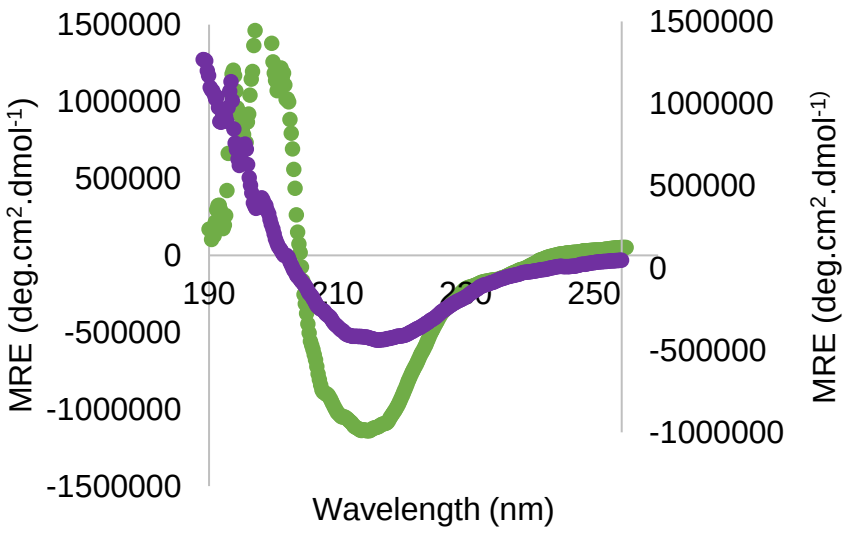


Figure 3.10 Far-UV CD spectra of the WT (green) and L38↑N↑L^{-4} (purple) proteases.

Spectra were measured on 10 µM protein in storage buffer using the Jasco J-1500 spectropolarimeter at 20 °C. The y-axis on the left corresponds to the WT and the y-axis on the right corresponds to L38↑N↑L^{-4}.

3.6.2 Quaternary structure

SE-HPLC was used to assess the quaternary structures of the proteases. The HIV-1 protease requires dimerization in order to be functional. Therefore, it was essential to check the oligomeric states of the proteases. Firstly, this would provide insight on whether the insertions shift the proportion from one state to another and secondly, it was essential to determine that the homodimeric state was present for downstream experiments. Analytical SE-HPLC was performed using a TSKgel SuperSW2000 SEC column and an elution buffer (10 mM sodium acetate, 0.01% (w/v) sodium azide, pH 5.0) containing 200 mM NaCl. The column was initially calibrated using standards of known molecular weights (Figure 3.11). Since the column had a resolution range of 5–150 kDa, only the standards that were in this range was used to plot the calibration curves. For the calibration curves, the logarithm of each appropriate molecular weight was plotted against its corresponding retention time (Figure 3.11). This allowed for the sizes of the WT and L38↑N↑L^{-4} proteases to be estimated relative to the sizes of the standings using the equation produced by the calibration curves. The equation for the

WT was y = -0.2038x + 8.2649 with an R^2 value of 0.9961. The equation for the L38↑N↑L^{-4} protease was y = -0.1892x + 7.9366 with an R^2 value of 0.9984.

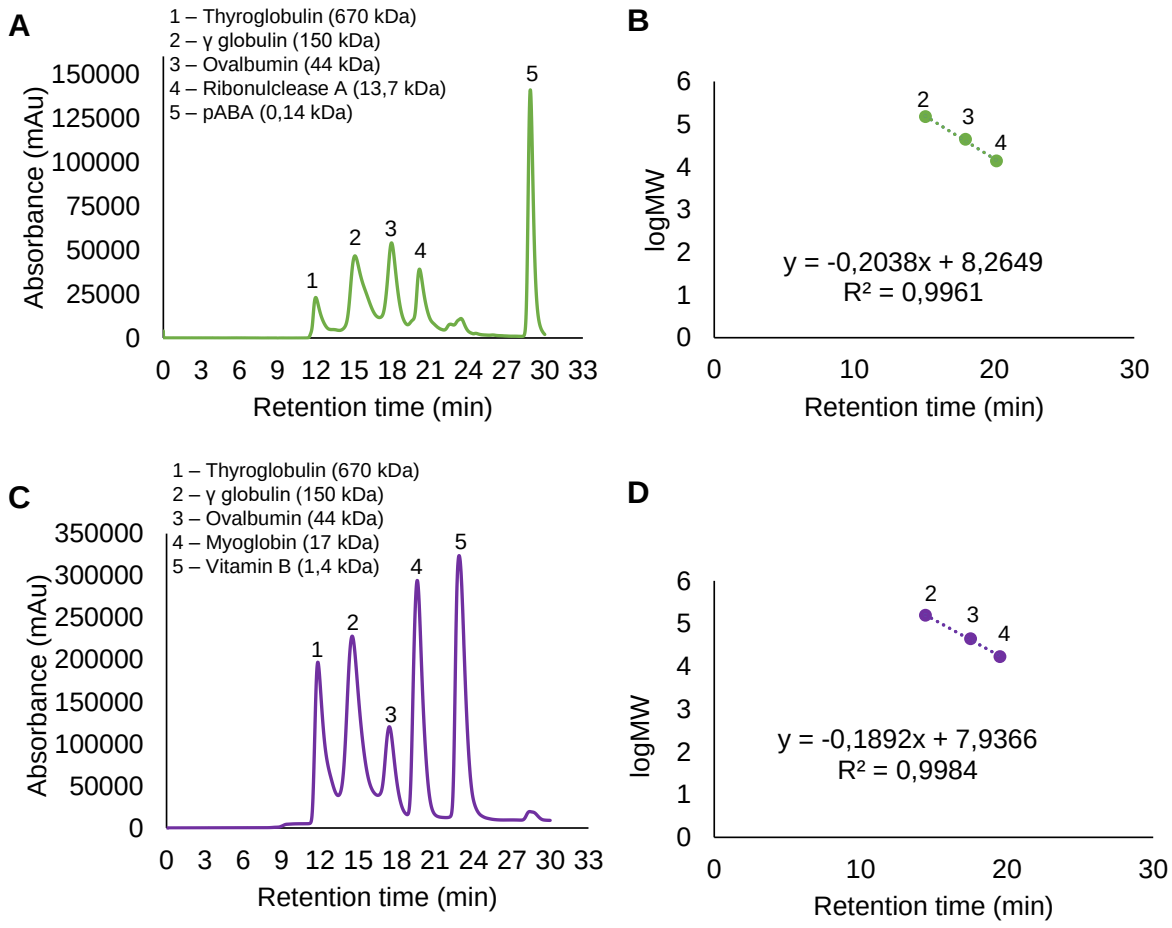


Figure 3.11 SE-HPLC column calibration using standards of known molecular weights.

A: Elution profile of the WT standards which was calibrated using thyroglobulin (670 kDa), γ globulin (150 kDa), ovalbumin (44 kDa), ribonuclease A (13,7 kDa) and pABA (0,14 kDA). **B:** Standard curve generated from elution profiles of WT standards. **C:** Elution profile of the L38↑N↑L^{-4} standards which was calibrated using thyroglobulin (670 kDa), γ globulin (150 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B (1,4 kDA). **D:** Standard curve generated from elution profiles of L38↑N↑L^{-4} standards.

The WT and L38↑N↑L^{-4} proteases were loaded separately onto the TSKgel SuperSW2000 SEC column using the same conditions as the standards at a concentration of 30 µM (Figure 3.12). The elution times of the peaks generated for each protein were substituted into the corresponding calibration curve equations. Both the WT and L38↑N↑L^{-4} proteases eluted as more than one single peak indicating a subpopulation of oligomeric states. Using the calibration curves, the calculated weights of the WT peaks were 10 and 16 kDa, while the L38↑N↑L^{-4} produced peaks of 7, 9 and 17 kDa. The WT consisted of a major peak and a shoulder which corresponded to the 10 and 16 kDa peaks, respectively. On the other hand, L38↑N↑L^{-4} consisted of a major peak, a shoulder and a smaller peak corresponding to the 7, 9 and 17 kDa peaks, respectively. While both proteases sampled the monomeric state of the protease (~10 kDa), there was no evidence of the dimeric states. It is possible that the presence of salt in the buffer increased the autoproteolytic activity of the proteases, therefore, resulting in cleaved states of the proteins.

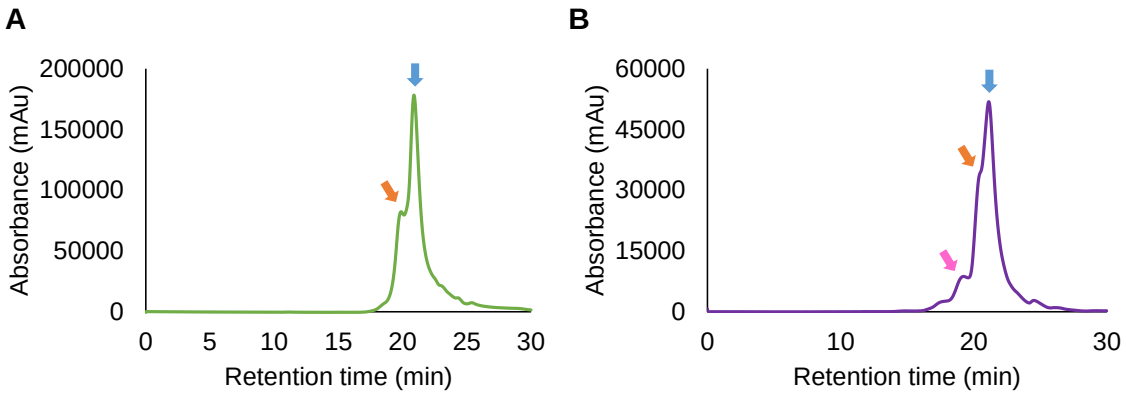


Figure 3.12 SE-HPLC elution profiles of WT (A) and L38↑N↑L^{-4} (B).

The blue, orange and pink arrows point to peak 1, shoulder and peak 2 respectively.

3.6.3 Crystallisation

In order to investigate the quaternary structure of the proteases, crystal trials were performed using the Hampton Research Index™ screening kit on the WT and L38↑N↑L^{-4} proteases. Attempts were made to crystalise proteases in both their apo

form and complexed to PIs. The WT protease produced apo crystals in a multitude of conditions (**Table 3.3**) which displayed the characteristic trigonal bipyramidal crystal structure (Figure 3.13). On the other hand, the L38↑N↑L^{-4} protease only produced crystals when complexed with DRV (**Table 3.4**).

X-ray diffraction of crystals were unsuccessful, and no further attempts were made to diffract WT or L38↑N↑L^{-4} crystals. Unsuccessful diffraction could be attributed to the export of crystals from South Africa to Diamond Light Source in the UK, which may have been damaged during transportation.

Table 3.3 Hampton Research Index^{TM} conditions that resulted in hits with the apo WT protease

Condition	# of crystals
0.1 M Bis-Tris (pH 5.5), 3 M Sodium chloride	4
0.2 M Ammonium sulphate, 0.1M Bis-Tris (pH 6.5), 25% (w/v) PEG 3350	6
0.2 M Ammonium sulphate, 0.1M Tris (pH 8.5), 25% (w/v) PEG 3350	2
0.2 M Sodium chloride, 0.1 M Bis-Tris (pH 5.5), 25% (w/v) PEG 3350	5
0.2 M Sodium chloride, 0.1 M Bis-Tris (pH 6.5), 25% (w/v) PEG 3350	3
0.2 M Magnesium chloride hexahydrate, 0.1 M Bis-Tris (pH 5.5), 25% (w/v) PEG 3350	7
0.2 M Magnesium chloride hexahydrate, 0.1 M Tris (pH 8.5), 25% (w/v) PEG 3350	>30
0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES (pH 7.5), 25% (w/v) PEG 3350	3
0.2 M Sodium malonate (pH 7), 20% (w/v) PEG 3350	6

Table 3.4 Hampton Research Index^{TM} condition that resulted in hits with L38↑N↑L^{-4} bound to DRV.

Condition	[DRV]	# of crystals
0.1 M Bis-Tris (pH 5.5), 3 M NaCl	10x	5
	15x	7
	20x	7

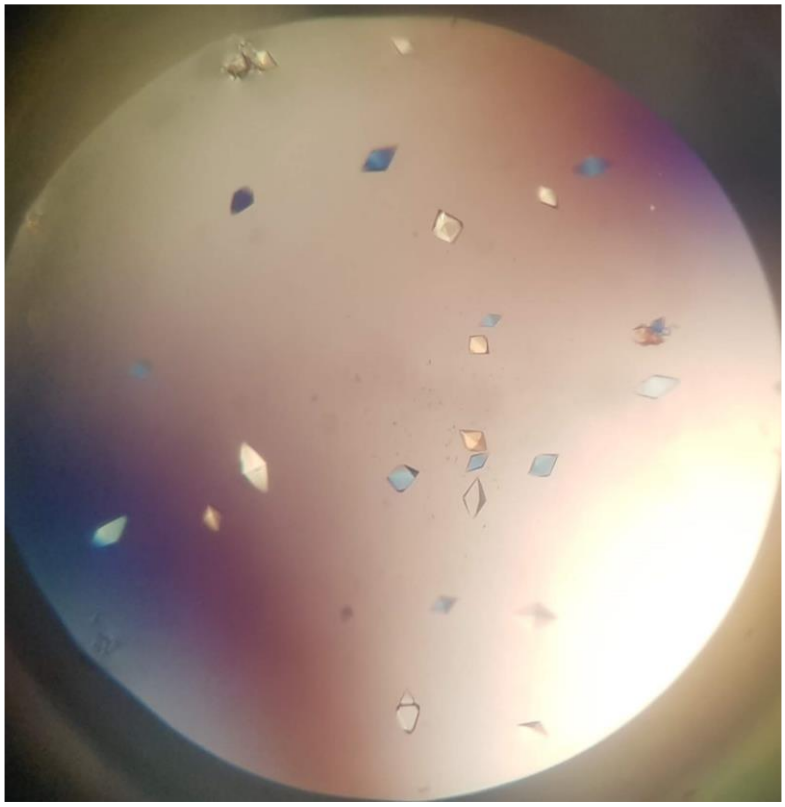


Figure 3.13 Image depicting trigonal bipyramidal WT crystals.

3.7 ISOTHERMAL TITRATION CALORIMETRY

Displacement titration calorimetry was used to determine the quantitative binding parameters of PIs to the WT and L38↑N↑L^{-4} proteases. The drug binging energetics of four PIs ATV, DRV, RTV and SQV were assessed. As previously reported, PIs bind the HIV-1 protease in nM-pM range (Velazquez-Campoy *et al.*, 2001). The dissociation constants (K_D) for all PIs were found to be in the nM range for both the WT and L38↑N↑L^{-4} proteases (Figure 3.14). The K_D values for DRV were 1 nM for both the WT and L38↑N↑L^{-4} proteases indicating that the binding affinity of this PI was unchanged for both proteases. The K_D values of the WT and L38↑N↑L^{-4} for ATV was 8 and 16 nM while SQV was 107 and 133 nM, respectively indicating a 2-fold and 1.2-fold increase. Interestingly, the K_D values of the WT and L38↑N↑L^{-4}, 1 and 49 nM for RTV, displayed a 49-fold increase between the binding affinities of the mutant and WT for this PI.

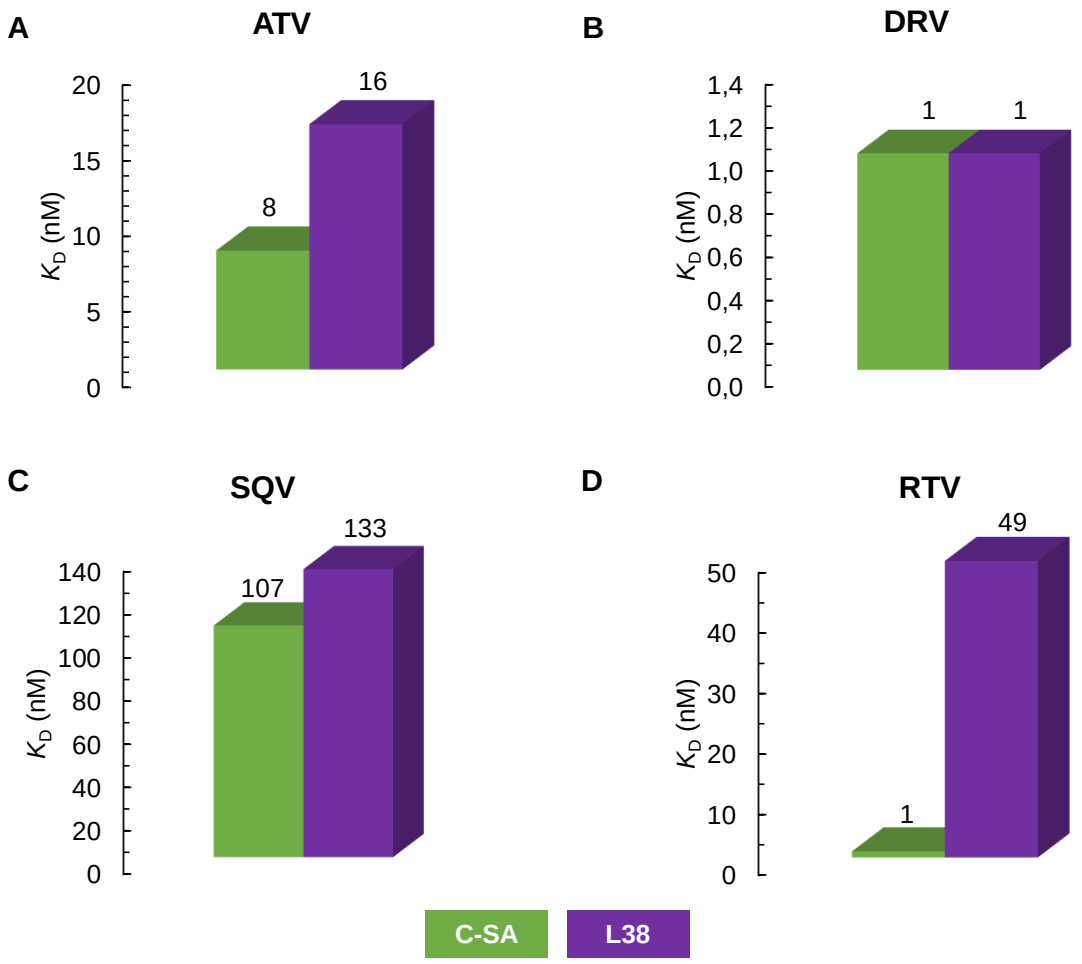


Figure 3.14 Dissociation binding constants for WT and L38↑N↑L^{-4} with ATV (A), DRV (B), RTV (C) and SQV (D).

The green bars correspond to the K_D values for the WT while the purple bars correspond to the K_D values for L38↑N↑L^{-4}.

The effect of the two insertions on the thermodynamic parameters (Δ*G*, Δ*H* and TΔ*S*) for each PI was examined (Figure 3.15). All isotherms displayed pre and post transitions allowing for the reliable observation of thermodynamic parameters (Figure S2-6). The raw data (Figure S2) generated by the displacement titrations were baseline corrected, HODs were subtracted and the change in heats were fitted to an independent binding model using the NanoAnalyze™ software. The corrected heat rates correspond to the raw data and the fits exhibit well-defined transitions (Figure S3-6), allowing for the extrapolation of meaningful K_D values from the slope. The

energetics of PI binding was examined by comparing the changes in Gibbs free energy (Δ*G*) between the WT and L38↑N↑L^{-4} proteases. All PIs produced negative Δ*G* values indicating that these interactions are spontaneous and energetically favourable (Figure 3.15). With the exception of DRV, which produced the same Δ*G* value for the WT and L38↑N↑L^{-4} (-51 kJ/mol), all other PIs produced lower Δ*G* values for L38↑N↑L^{-4} compared to the WT. Specifically, the change in Δ*G* was 10 and 5 kJ/mol for RTV and SQV, respectively. On the other hand, ATV did not show a significant change between the two proteases with the WT and L38↑N↑L^{-4} producing Δ*G* values of -45 and -44 kJ/mol, respectively. In order for protein-ligand interactions to be favourable, the Δ*G* of the system must be negative. Therefore, the Δ*G* decreases observed for RTV and SQV when binding to L38↑N↑L^{-4} in comparison to the WT indicates that the favourability of ligand is affected by the insertions.

Binding of all PIs to both the WT and L38↑N↑L^{-4} protease was exothermic and enthalpically driven as seen by the negative Δ*H* values produced for each interaction (Figure 3.15). The enthalpy change (Δ*H*) upon binding of ATV was similar in magnitude between the WT (-39 kJ/mol) and L38↑N↑L^{-4} (-38 kJ/mol) proteases. However, this was not the case with the other three PIs. While DRV and RTV showed an increase in Δ*H* values when comparing L38↑N↑L^{-4} to the WT, SQV exhibited a decrease in magnitude of the change in enthalpy when binding to the L38↑N↑L^{-4} in comparison to WT. The Δ*H* values for DRV were -62 and -80 kJ/mol for the WT and L38↑N↑L^{-4}, respectively while the Δ*H* values for RTV were -25 and -39 kJ/mol for the WT and L38↑N↑L^{-4}, respectively. SQV produced Δ*H* values of -21 and -13 kJ/mol for the WT and L38↑N↑L^{-4}, respectively. The effect of these insertions on L38↑N↑L^{-4} resulted in a ΔΔ*H* gain of 18 and 14 kJ/mol for DRV and RTV, respectively and a ΔΔ*H* loss of 8 kJ/mol for SQV.

The change in entropy (TΔ*S*) of binding between ATV, RTV and SQV all displayed positive values upon binding of the WT and L38↑N↑L^{-4} while DRV displayed a negative change in entropy (Figure 3.15). Once again, ATV did not show a significant change between the two proteases with the WT and L38↑N↑L^{-4} producing TΔ*S* values of 7 and 6 kJ/mol, respectively. Interestingly, an increase of 18 kJ/mol was observed in the change in entropy for DRV upon binding to L38↑N↑L^{-4} (-30 kJ/mol) in comparison to the WT (-12 kJ/mol). Moreover, the entropy changes upon binding of RTV and SQV showed significant differences between the two PIs as well as the WT and L38↑N↑L^{-4}

proteases. The TΔ*S* values of RTV upon binding the WT and L38↑N↑L^{-4} were 26 and 2 kJ/mol respectively while the TΔ*S* values of SQV upon binding the WT and L38↑N↑L^{-4} were 18 and 26 kJ/mol, respectively. The effect of these insertions on the protease resulted in a TΔΔ*S* loss of 24 kJ/mol and TΔΔ*S* gain of 8 kJ/mol for RTV and SQV respectively. The observed changes between the WT and L38↑N↑L^{-4} systems regarding the TΔΔ*S* of DRV, RTV and SQV are significant. Specifically, the changes indicate that the binding of DRV is more entropically unfavourable while SQV is more entropically favourable when binding to L38↑N↑L^{-4}. Noticeably, RTV displayed the largest change in entropy with binding to L38↑N↑L^{-4} being less entropically favourable.

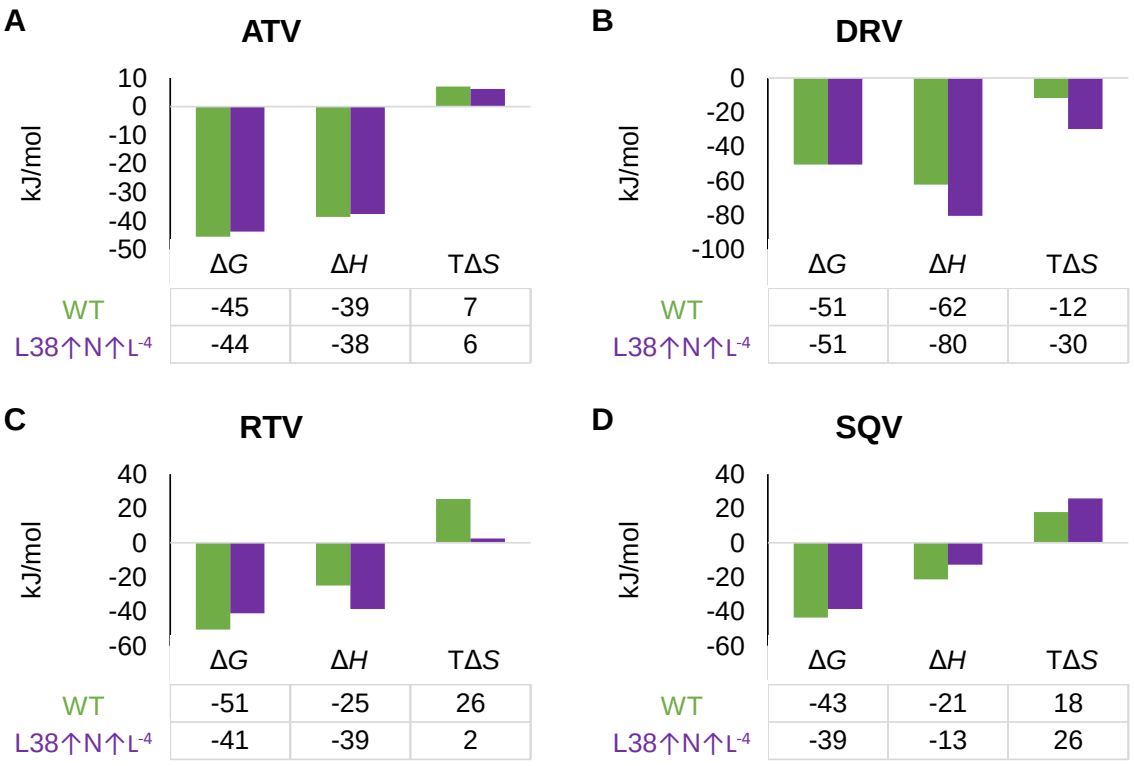


Figure 3.15 Thermodynamic parameters (Δ*G*, Δ*H* and TΔ*S*) of ATV (A), DRV (B), RTV (C) and SQV (D) binding to the WT and L38↑N↑L^{-4} proteases.

The green bars correspond to the values for the WT while the purple bars correspond to the values for L38↑N↑L^{-4}.

3.8 Differential scanning calorimetry

Differential scanning calorimetry (DSC) was used to determine the structural stabilities of the proteases by measuring the heat capacity of the proteins in solution as a function of temperature. The apo WT and L38↑N↑L^{-4} forms, as well as these proteins bound with ATV, DRV, RTV and SQV were compared (Figure 3.16). Noticeably, the stabilities of the two native proteases differed significantly with T_m values for the WT and L38↑N↑L^{-4} of 64 °C and 69 °C, respectively, suggesting that the native mutant is more stable than the WT (Figure 3.17). Interestingly, when complexed with PIs, L38↑N↑L^{-4} displayed a biphasic transition producing two distinct transition peaks while the WT exhibited only single transition peaks (Figure 3.16). While all the L38↑N↑L^{-4}•PI complexes produced two peaks, the area ratio of the peaks differed between PIs. The melting curves for L38↑N↑L^{-4}•ATV and L38↑N↑L^{-4}•DRV produced peaks with an area ratio of approximately 60/40 while L38↑N↑L^{-4}•RTV produced peaks with an area ratio of approximately 80/20. In contrast, L38↑N↑L^{-4}•SQV produced peaks with an area ratio of approximately 30/70.

The transition temperatures for the WT complexed with ATV, DRV, RTV and SQV were 88 °C, 93 °C, 86 °C and 87 °C, respectively (Figure 3.17). Due to the presence of two transition peaks for L38↑N↑L^{-4}, each PR•PI complex had two corresponding T_m values (Figure 3.17). The transition temperatures for L38↑N↑L^{-4}•ATV complex was 74 °C and 54 °C, L38↑N↑L^{-4}•DRV was 75 °C and 85 °C, L38↑N↑L^{-4}•RTV was 74 °C and 84 °C while L38↑N↑L^{-4}•SQV was 70 °C and 80 °C. All L38↑N↑L^{-4}•PI complexes exhibited lower values than the WT•PI complex counterparts indicating that when bound to an inhibitor, L38↑N↑L^{-4} was less stable than the WT.

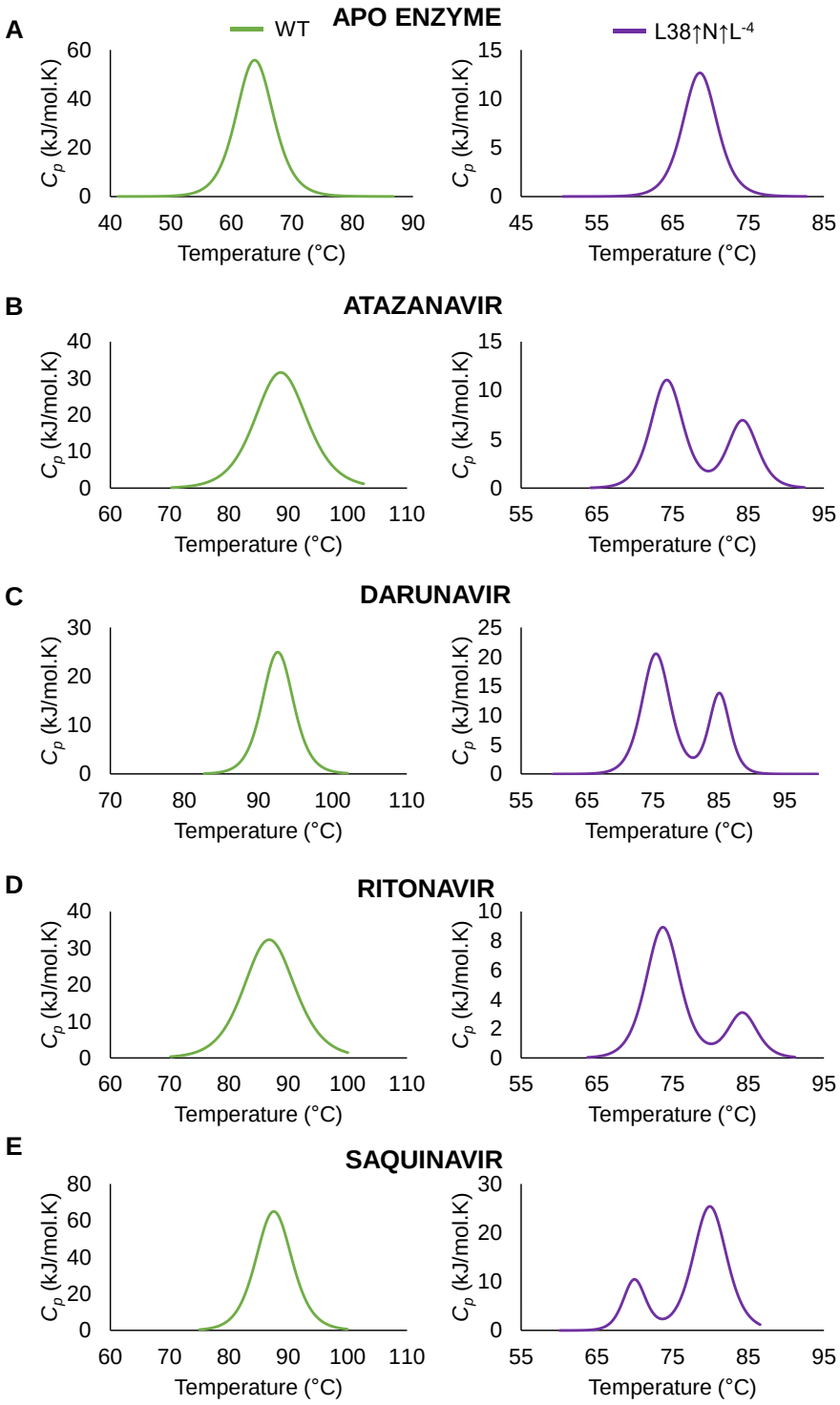
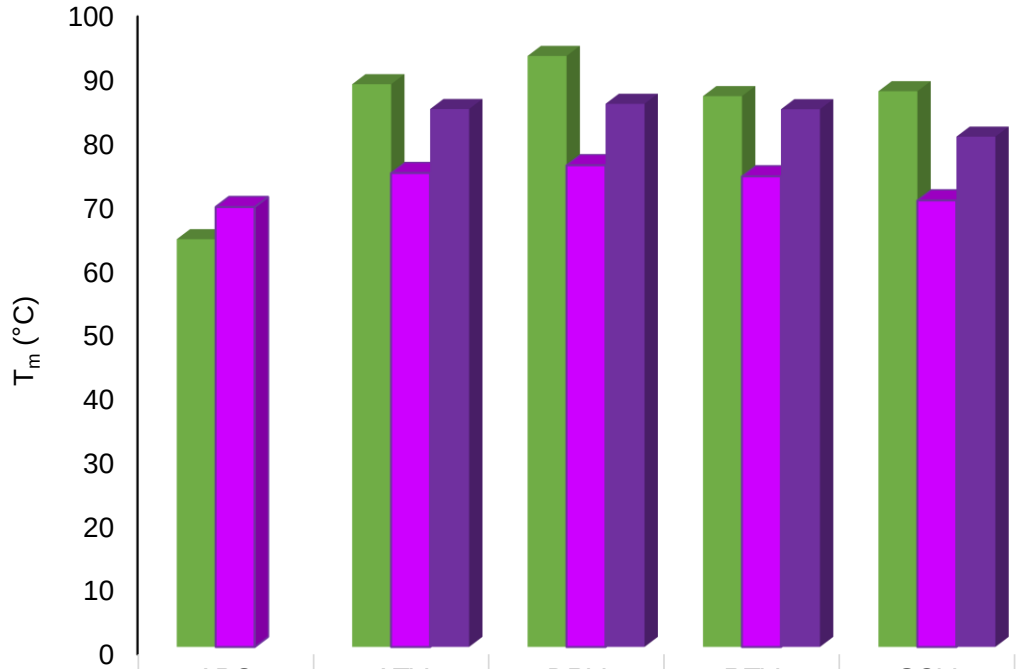


Figure 3.16 DSC thermograms of WT (green) and L38↑N↑L^{-4} (purple) proteases.

A: Profiles of proteases in their apo forms. **B:** Profiles of proteases complexed to ATV. **C:** Profiles of proteases complexed to DRV. **D:** Profiles of proteases complexed to RTV. **E:** Profiles of proteases complexed to SQV. All profiles were baseline corrected and normalised to molar hear capacity using the protease dimer concentrations.



	APO	ATV	DRV	RTV	SQV
WT	64	88	93	86	87
L38↑N↑L^{-4} P1	69	74	75	74	70
L38↑N↑L^{-4} P2		84	85	84	80

Figure 3.17 Thermal transition midpoints (T_m) of WT (green bars) and L38↑N↑L^{-4} (purple bars) for the apo forms and in the presence of ATV, DRV, RTV and SQV.

The light purple bars correspond to the first peak (P1) while the dark purple bars correspond to the second peak (P2) of L38↑N↑L^{-4}.

3.9 Steady-state kinetics

The steady-state kinetics of the apo WT and L38↑N↑L^{-4} protease was determined using a fluorogenic substrate that mimics the natural polypeptide's capsid/p2 cleavage site with a significant difference deemed to be any change more than 10%, (**Table 3.5**). The specific activity and k_{cat} of L38↑N↑L^{-4} was approximately 50% lower than the WT. On the other hand, the k_{cat}/K_M displayed a 60% increase between the WT and L38↑N↑L^{-4}. The r^2 values for the specific activity and k_{cat} plots (Figure S1A and B) were greater than 0.95 indicating reliable data for both the WT and L38↑N↑L^{-4} proteases. However, the r^2 values for the catalytic efficiency plots (Figure S1C) were 0.93 and 0.81 for the WT and L38↑N↑L^{-4}, respectively, indicating that this data is less reliable.

The changes in the kinetics of L38↑N↑L^{-4} suggests that the insertions in the hinge region affect the catalytic activity of the protease.

Table 3.5 Steady-state kinetics of the apo WT and L38↑N↑L^{-4} proteases.

	Specific activity (μmol^{-1}.min^{-1}.mg^{-1})	***k*$_{cat}$** (s^{-1})	***k*$_{cat}$/*K*$_M$** (s^{-1}.μM^{-1})
WT	22.6 ± 2	9.1 ± 0.8	1.38 ± 0.2
L38↑N↑L^{-4}	12.7 ± 1.2	4.97 ± 0.6	2.3 ± 0.6

3.10 Induced-fit docking studies

Induced-fit docking studies (IFDS) was used to select and compare the best poses for each of the PR•PI complexes. The pose which had the lowest IFDS score was selected and used to represent the most favourable complex. Since IFDS is performed *in silico*, the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} was included in this analysis. The GlideScore (G$_{score}$) provides an approximation of ligand binding free energy and, therefore, an approximation of ligand binding affinity (Freisner *et al*., 2004). The G$_{score}$ values produced from the IFDS of the three proteases with ATV, DRV, RTV and SQV were compared in order to determine if any changes in the calculated binding scores had occurred (Figure 3.18).

Noticeably, RTV produced the largest decrease in the G$_{score}$ between the WT and the mutant proteases. The decrease observed between the WT and L38↑N↑L^{-4} was 2.2 kcal/mol while L38↑N↑L^{+4} was 2.9 kcal/mol. ATV produced identical G$_{score}$ values of -9.5 kcal/mol for both the WT and L38↑N↑L^{+4} while L38↑N↑L^{-4} had a higher G$_{score}$ of -10.3 kcal/mol (**Table 3.6**). On the other hand, DRV produced a minor decrease of 0.2 kcal/mol between the WT and L38↑N↑L^{-4} while L38↑N↑L^{+4} displayed a larger decrease of 1 kcal/mol in comparison to the WT. Interestingly, binding of SQV differed between the two mutated proteases with L38↑N↑L^{-4} resulting in a decrease of 0.6 kcal/mol in comparison to the WT. In contrast, L38↑N↑L^{+4} displayed an increase of 0.6 kcal/mol when compared to the WT. With the exception of ATV, these G$_{score}$

calculations corresponds to the trend seen with the ITC data which shows that L38↑N↑L^{-4} binds these PIs with lower affinity than the WT.

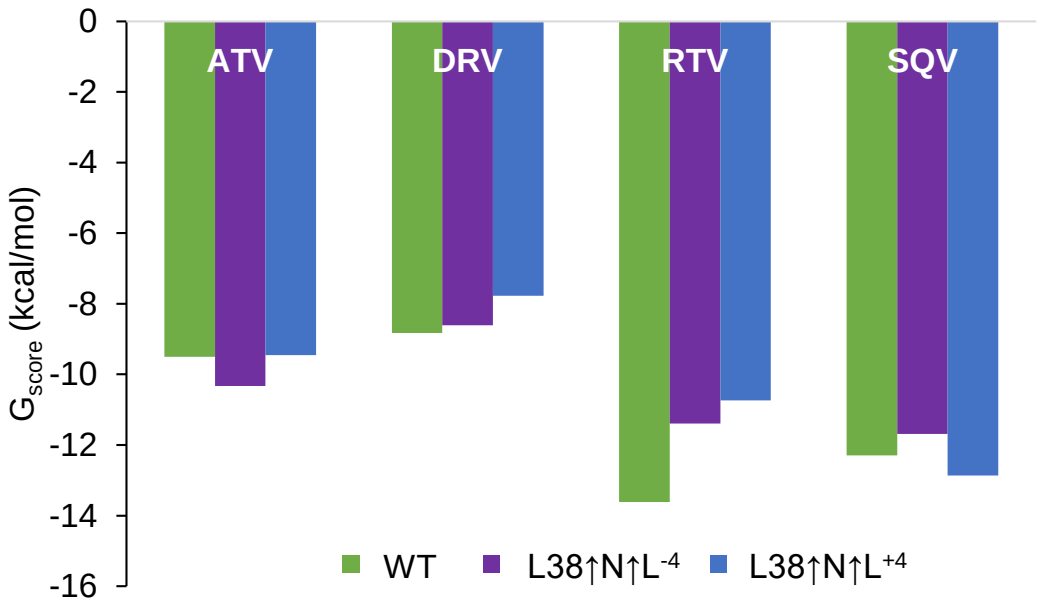


Figure 3.18 IFDS of PIs binding to WT, L38↑N↑L^{-4} and L38↑N↑L^{+4}.

The green, purple and blue bars represent the G$_{score}$ of the WT, L38↑N↑L^{-4} and L38↑N↑L+^{4}, respectively.

Table 3.6 G$_{score}$ values of three proteases in kcal/mol.

	ATV	DRV	RTV	SQV
WT	-9.5	-8.8	-13.6	-12.3
L38↑N↑L^{-4}	-10.3	-8.6	-11.4	-11.7
L38↑N↑L^{+4}	-9.5	-7.8	-10.7	-12.9

3.11 Molecular dynamics

Following IFDS, the best docked structure was selected based on the top IFDS score and was further investigated using MD simulations. Prior to MD simulations, the models were validated to ensure that results were accurate and reliable. A homology model of a known variant protease containing three mutations (I13V, I62V and V77I) was generated and superimposed on the available crystal structure (PBD: 6I45) to ensure that the modelled structures used in this study were accurate. The superimposition of the modelled variant protease and the crystal structure showed no visible differences (Figure S7), thereby confirming the accuracy of the generated models. In addition, the apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} protease as well as the complexed proteases were validated using PROCHECK. The Ramachandran plots (Figure S8) for all models showed that majority (>95%) of the atoms occurred in the allowed regions, indicating that the models were valid and could be used for further analyses.

3.11.1 Apo analysis

Molecular dynamics (MD) simulations were used to investigate the dynamics of the three proteases in complex with ATV, DRV, RTV and SQV. The essential dynamics, which refers to the motions of different regions within each protease over a specific time, of the apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} was investigated. The analyses of these studies were limited to the backbone C_α atoms of each protease.

MD simulations were performed on the solvated apo proteases for 50 ns with the starting structures aligned to the reference structure (PDB: 2HB4). All proteases were initially in the energy-minimised and closed conformation for MD simulations. The root-mean-square deviation (RMSD) of the apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} was assessed to determine the overall structural stability of the proteins during the simulations (Figure 3.19).

Noticeably, the RMSD data for the WT was higher than the L38↑N↑L^{-4} and L38↑N↑L^{+4} throughout the simulation. The L38↑N↑L^{-4} protease fluctuated constantly around approximately 1.2 Å. In contrast, the L38↑N↑L^{+4} protease started the simulation with the lowest RMSD value, however, it stabilised with the WT around 2 – 2.5 Å. The WT

displayed the highest RMSD with an average RMSD value of 2.22 Å (**Table 3.7**). The L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases exhibited average RMSD values of 1.45 and 1.81 Å respectively indicating that these mutated proteases were more stable than the WT.

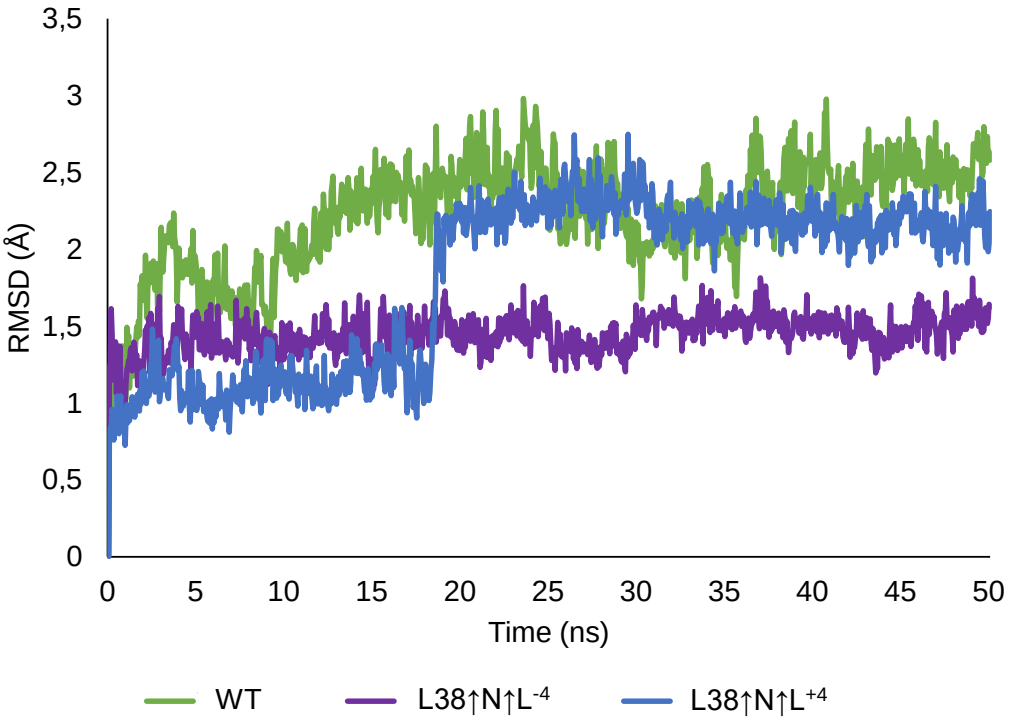


Figure 3.19 RMSD of apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} protease polypeptide backbones.

Table 3.7 Average RMSD data for apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} collected over the course of 50 ns MD simulations.

Protein	Average RMSD	Standard deviation
WT	2.22	±0.37
L38↑N↑L^{-4}	1.45	±0.13
L38↑N↑L^{+4}	1.81	±0.55

The radius of gyration (R_g) was used to assess the compactness of the apo proteases during the simulations. All three proteases started with the same degree of

compactness which then varied as the simulation progressed (Figure 3.20). After 20 ns, the L38↑N↑L^{-4} protease became more compact than both the WT and L38↑N↑L^{+4}. The L38↑N↑L^{+4} protease displayed a decrease in compactness around 15-20 ns while the WT was initially less compact than at the end of the simulation. Additionally, the average R_g values indicated that the L38↑N↑L^{-4} protease was the most compact, followed by the WT and then L38↑N↑L^{+4} (**Table 3.8**).

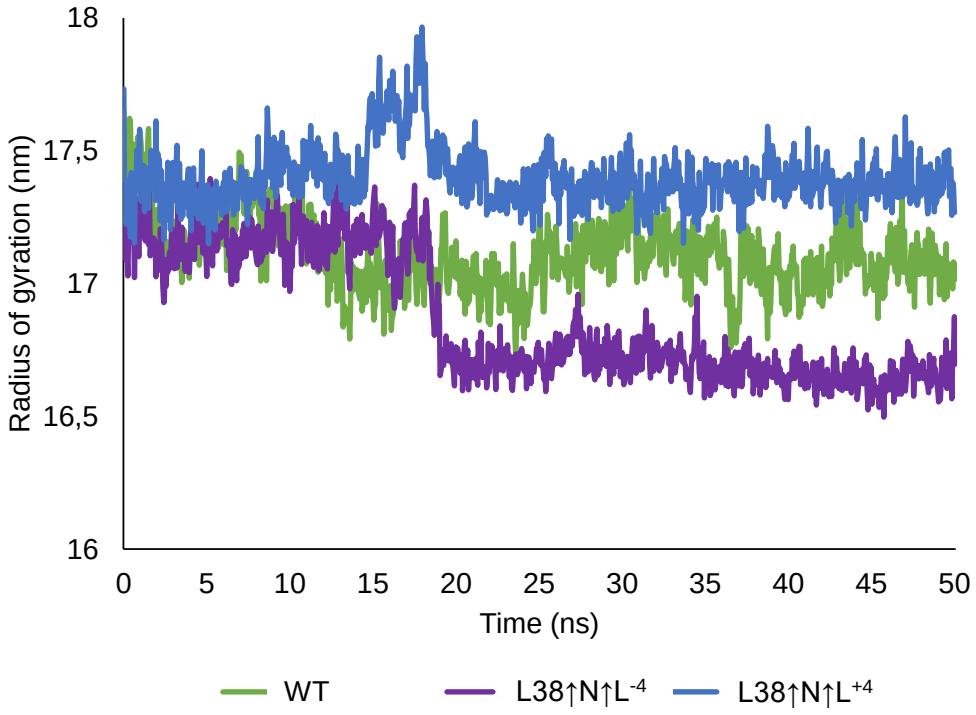


Figure 3.20 Radius of gyration of apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} protease polypeptide backbones.

Table 3.8 Average R_g values for apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} collected over the course of 50 ns MD simulations.

Protein	Average R_g	Standard deviation
WT	17.11	±0.14
L38↑N↑L-4	16.87	±0.24
L38↑N↑L+4	17.40	±0.11

Root-mean-square fluctuation (RMSF) describes the average deviation of the C_α of each amino acid residue relative to a reference point over time. RMSF was used to assess the local changes that occurred along the protein chains of the apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases (Figure 3.21). As previously mentioned, the starting conformation used for each simulation was the closed conformation of the protease. The WT residues exhibiting the greatest RMSF values were 42-56 and 62-78 which corresponds to the flap and cantilever regions, respectively (Figure 3.21). Interestingly, the fulcrum region, residues 10-22 of L38↑N↑L^{+4}, displayed the greatest RMSF values.

The most noticeable difference in the average RMSF of L38↑N↑L^{-4} occurred in the hinge and flap regions where the average RMSF of the hinge in the B chain increased by 49% while the flaps displayed a 44% decrease and 60% increase in chain A, respectively, when compared to the WT regions (**Table 3.9**). Noticeably, when compared to the WT, the cantilever region of L38↑N↑L^{-4} also showed a discrepancy between chains A and B with a 25% decrease and 60% increase, respectively. No significant difference was exhibited by the hinge region average RMSF of L38↑N↑L^{-4} in comparison to the WT. Interestingly, L38↑N↑L^{+4} exhibited a decrease in the average RMSF of both the flap (14%/21%) and cantilever (7%/25%) regions in comparison to the WT. The forward slash represents the respective changes in each chain (A/B) of the protease. In contrast, the hinge region of L38↑N↑L^{+4} showed a 28% and 33% increase in the average RMSF values in comparison to the WT. The L38↑N↑L^{+4} fulcrum regions of chain A and B exhibited a 12% increase and 19% decrease, respectively (**Table 3.9**).

3.11.2 RMSF OF PROTEASE INHIBITOR COMPLEXES

The RMSF of PR•PI complexes was also examined to characterise any changes in the degree of fluctuation of the protease regions when a PI was present. In comparison to the WT, both L38↑N↑L^{-4} and L38↑N↑L^{+4} exhibited decreases in the average RMSF values of the fulcrum region when bound with a PI (**Table 3.10**). Similarly, the cantilever region of both mutated proteases exhibited a decrease in RMSF values for majority of PIs when compared to the WT.

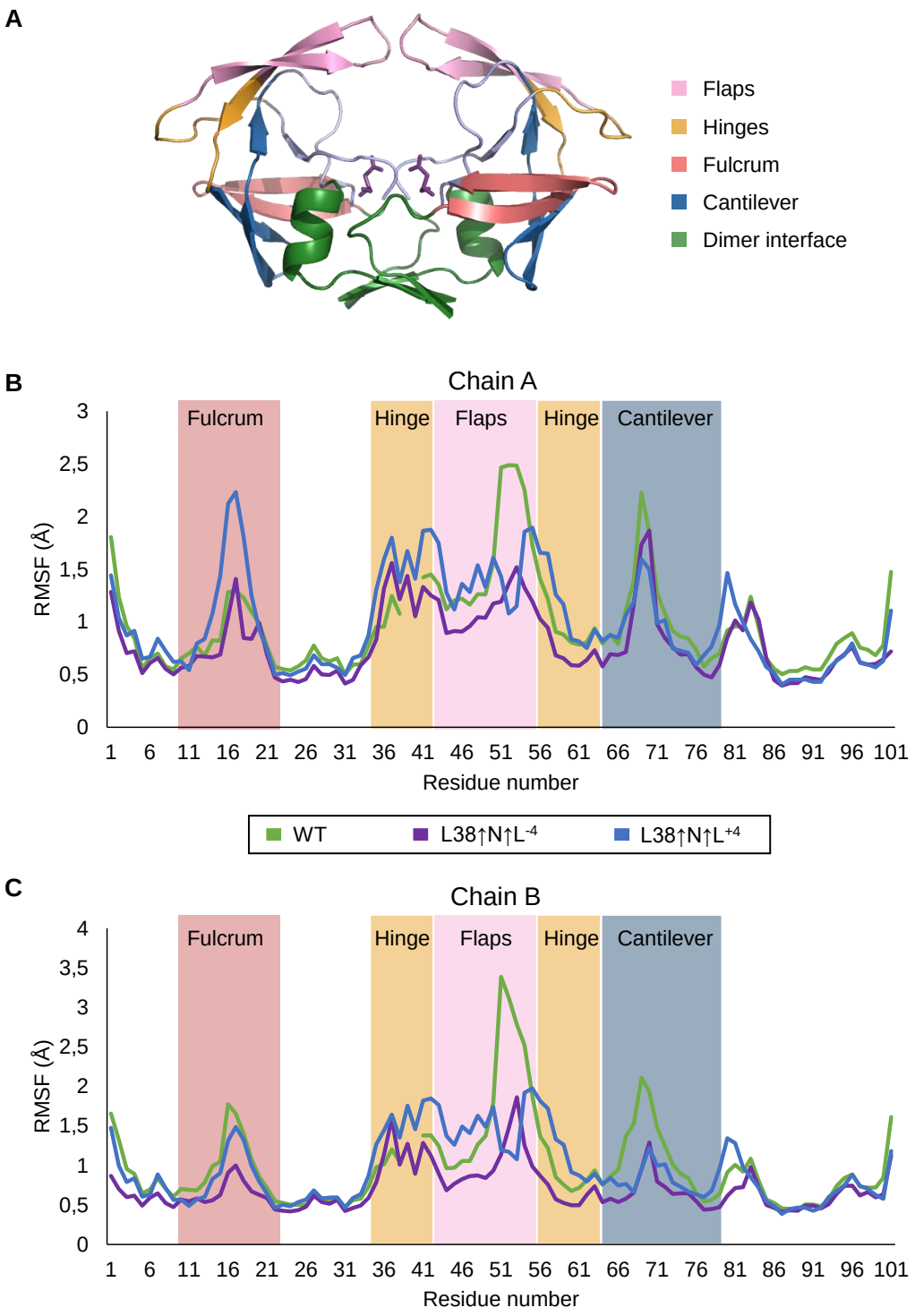


Figure 3.21 RMSF of apo WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} regions with respect to amino acid residues.

A: Structure of protease displaying flaps (pink), hinges (orange), fulcrum (red), cantilever (blue) and dimer interface (green). **B:** RMSF of chain A. **C:** RMSF of chain B.

Table 3.9 Percentage change (increase/decrease) in average RMSF data for L38↑N↑L^{-4} and L38↑N↑L^{+4} with regards to WT during a 50 ns MD simulation.

The grey shaded boxes represent a decrease in percentage while the orange shaded boxes represent an increase in percentage.

	Fulcrum		Hinge		Flap		Cantilever	
	A	B	A	B	A	B	A	B
L38↑N↑L^{-4}	-17	49	3	4	-44	69	-25	60
L38↑N↑L^{+4}	12	-19	29	33	-14	-21	-7	-25

Table 3.10 Percentage change (increase/decrease) in average RMSF data for L38↑N↑L^{-4} and L38↑N↑L^{+4} regions with regards to WT during a 50 ns MD simulation.

The grey shaded boxes represent a decrease in percentage while the orange shaded boxes represent an increase in percentage.

		Fulcrum		Hinge		Flap		Cantilever	
	PI	A	B	A	B	A	B	A	B
L38↑N↑L^{-4}	ATV	-7	-12	34	33	51	45	-2	-19
	DRV	-33	0	19	37	23	22	-14	-15
	RTV	-3	-14	23	22	16	29	-4	-40
	SQV	2	-14	26	12	-37	-33	-1	-18
L38↑N↑L^{+4}	ATV	-8	-19	19	2	12	-9	-6	-27
	DRV	-11	-7	20	21	4	14	-5	-7
	RTV	-7	-11	31	0	27	-2	3	-26
	SQV	21	-2	21	24	-42	10	8	-7

In contrast, the hinge region of both L38↑N↑L^{-4} and L38↑N↑L^{+4} exhibited varying degrees of increased fluctuations when compared to the WT. The most striking

increase was observed with DRV (19%/37%) and RTV (31%) for L38↑N↑L^{-4} and L38↑N↑L^{+4}, respectively. The RMSF values for the flap regions varied between proteases and PIs. The most striking differences observed for L38↑N↑L^{-4} was with ATV and SQV which exhibited a 51%/45% increase and 33%/37% decrease, respectively, when compared to the WT flap regions. Changes in the L38↑N↑L^{+4} flap region RMSF values were less prominent with SQV exhibiting the most striking difference. The differences observed in these regions of the proteases suggests that the insertions and mutations affect the behaviour of the flaps, hinges, cantilever and fulcrum both in the native and PI bound states.

3.11.3 LIGAND RMSF

L-RMSF was used to characterise changes in the ligand atom positions of ATV, DRV, RTV and SQV when bound to WT, L38↑N↑L^{-4} and L38↑N↑L^{+4}. In this case, the PR•PI complex was initially aligned to the reference structure and the L-RMSF was measured on the ligand heavy atoms. To determine the movement of the PIs within the active site of the protease, the ratios of the L-RMSF data between the mutated proteases with regards to the WT were analysed with ΔL-RMSF values <1 indicating reduced mobility and ΔL-RMSF values >1 indicating increased mobility of the PI (Figure 3.22). The ΔL-RMSF values of both L38↑N↑L^{-4} and L38↑N↑L^{+4} for DRV and RTV were <1 for 100% the PI atoms indicating that these inhibitors were largely restricted to the active site of these two proteases. The ΔL-RMSF values of L38↑N↑L^{-4} for ATV were <1 for 100% the PI atoms while L38↑N↑L^{+4} was similar to the WT fluctuating between a ΔL-RMSF value of 1. Conversely, the ΔL-RMSF values of L38↑N↑L^{-4} for SQV was similar to the WT fluctuating between 1 while the ΔL-RMSF values of L38↑N↑L^{+4} was <1 for 60% of the PI atoms. This indicates that ATV is more restricted to the active site of L38↑N↑L^{-4} while SQV is more restricted to the active site of L38↑N↑L^{+4}.

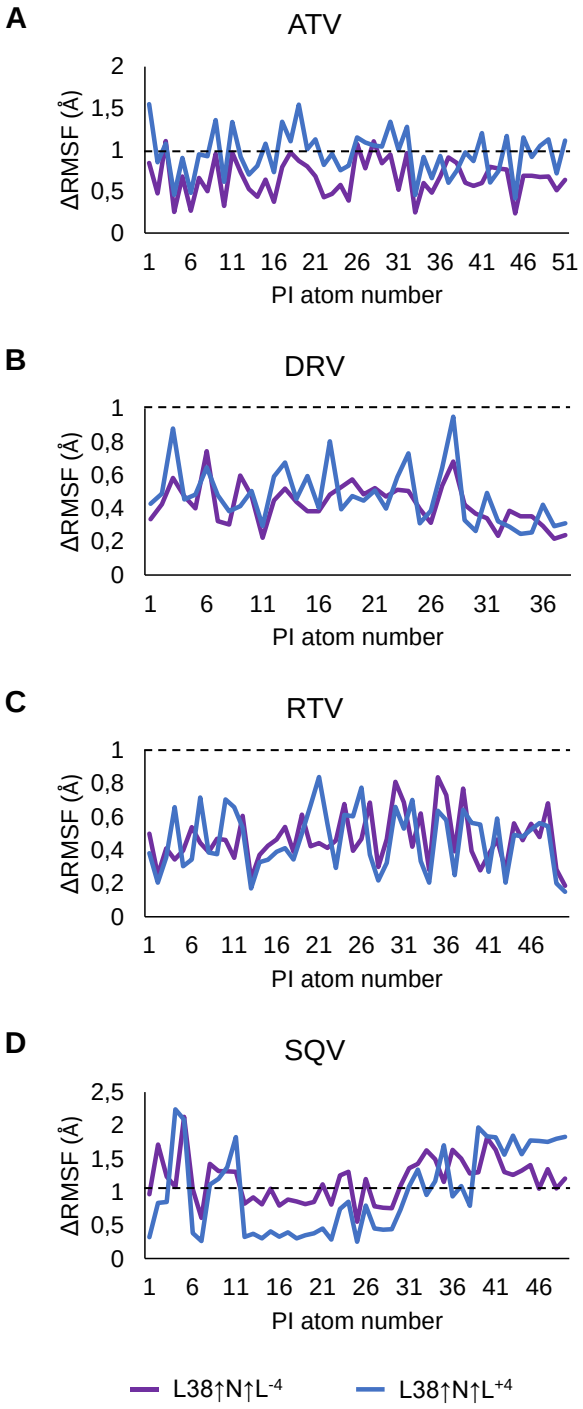


Figure 3.22 L-RMSF ratio (mutant/WT) of PI atoms when in complex with L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases.

A: L-RMSF ratios of mutant proteases in complex with ATV. **B**: L-RMSF ratios of mutant proteases in complex with DRV. **C**: L-RMSF ratios of mutant proteases in complex with RTV. **D**: L-RMSF ratios of mutant proteases in complex with SQV. The black dotted line represents the L-RMSF of the PIs in complex with the WT. The WT data was standardized to a value of 1 and serves as a reference point to which the mutants are compared.

3.11.4 FLAP CONFORMATIONS

The HIV-1 protease can exist in several flap conformations namely – open, semi-open and closed which exist in an equilibrium when the protease is in the apo state (Wlodawer *et al*., 1989). These conformations can be measured using the distance between the Ile50 residue in the flap tips and the Asp25 residue in the active site cavity. This provides a system of measurement of the exposure of substrates or PIs to the active site cavity. As mentioned previously these conformations are defined are <17 Å for the closed conformation, between 17-22 Å for the semi-open conformation and >22 Å for the open conformation.

The distance between Ile50 and Asp25 of each monomer was calculated for the WT (Figure 3.23). Due to the insertions at position 38, all residues following this position will have shifted two places. Therefore, for L38↑N↑L^{-4} and L38↑N↑L^{+4} the distance between Gly50 and Asp25 of each monomer was measured instead (Figure 3.23 A). The MD simulations were performed over 50 ns for each protease. The distance between the flap tips and the active site for chain A of all three proteases were similar at approximately 15 Å. However, the WT sampled the semi-open conformation at the beginning of the simulation (17-22 Å). In contrast, the distance between Ile50 and Asp25 of the WT chain B fluctuated and displayed all three conformations of the protease. Both mutated proteases measured <17 Å for the distance calculated between Gly50 and Asp25 of chain B indicating that these proteases maintained the closed conformation.

The binding of PIs to the protease active site is considered to shift the heterogenous conformation equilibrium to favour the closed conformation. As such, the distance between the flap tips and the active site was measured when PIs were bound to the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases (Figure 3.24). When in complex with ATV, chain A of the WT and L38↑N↑L^{-4} remained in the closed conformation (<17 Å) while L38↑N↑L^{+4} exhibited the semi-open conformation (17-22 Å) as well. In contrast, chain B of in complex with ATV remained solely in the closed conformation while the WT and sampled both the closed and semi-open conformations.

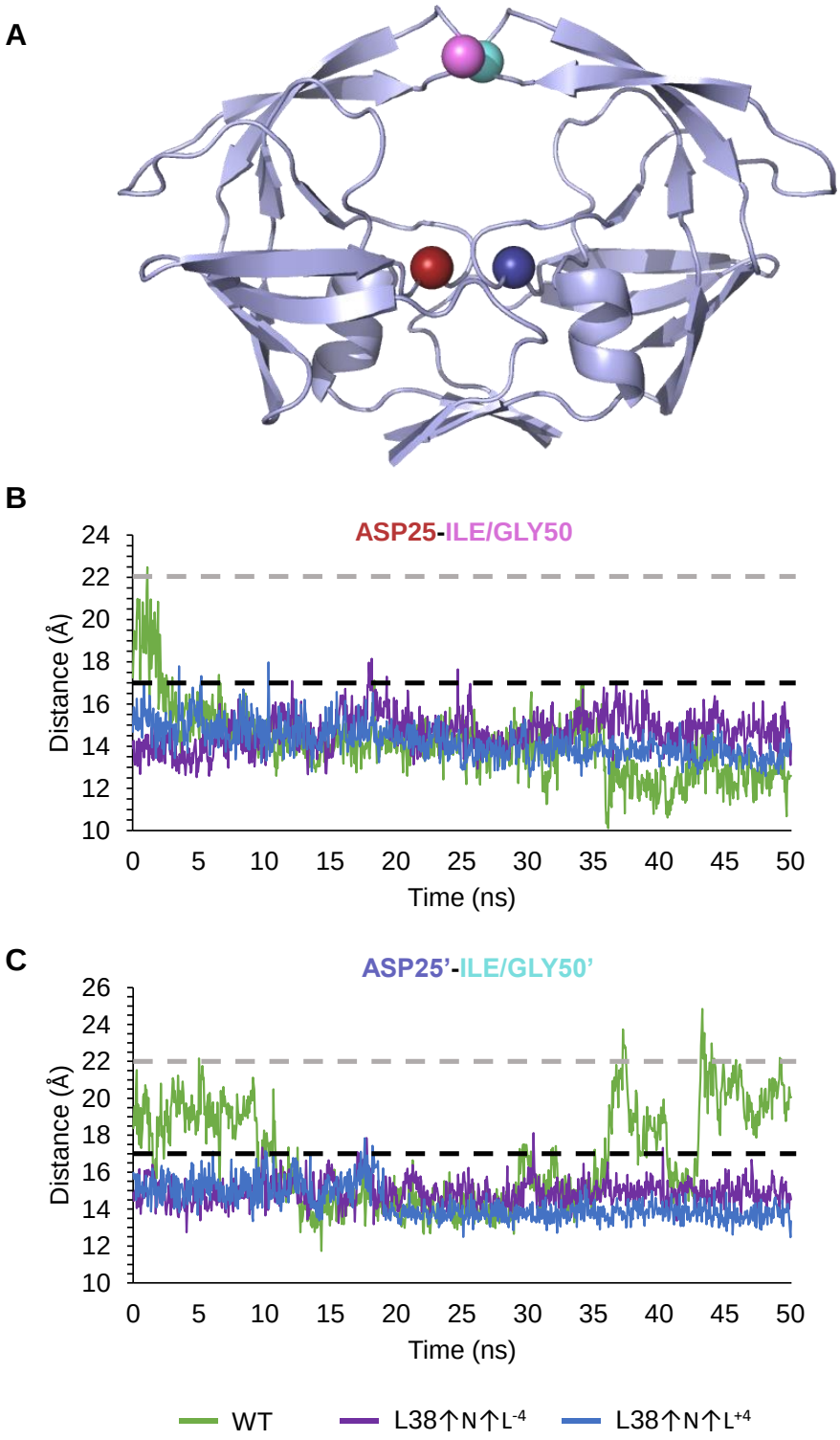


Figure 3.23 Flap conformations of WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases.

A: Structure of protease displaying Asp25 (red) and Ile/Gly50 (pink) of chain A and Asp25' (blue) and Ile/Gly50' (cyan) of chain B. **B:** Distance between Asp25-Ile/Gly50 atoms in chain A. **C:** Distance between Asp25'-ILle/Gly50' atoms in chain B. The black and grey dashed lines indicate 17 Å and 22 Å, respectively.

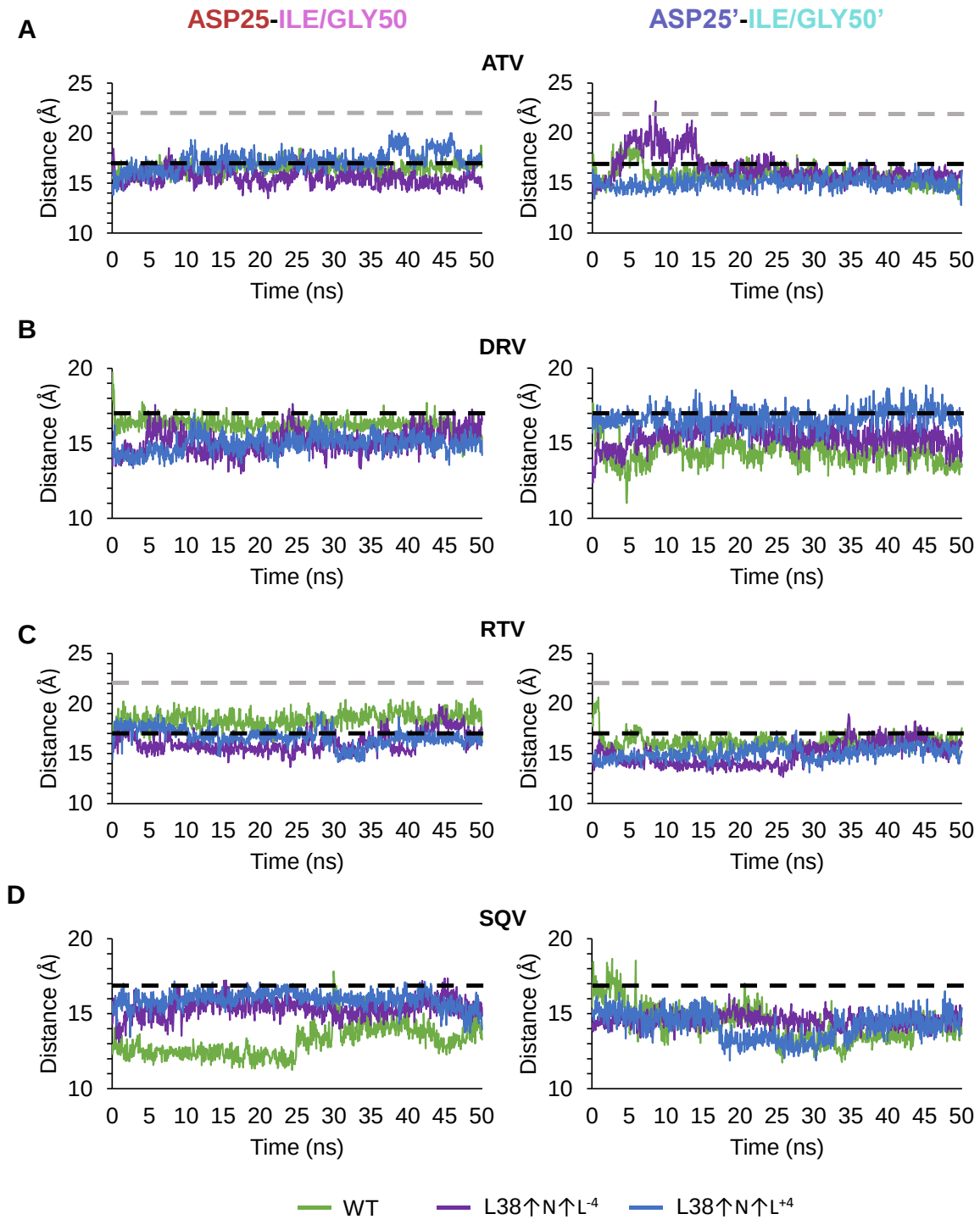


Figure 3.24 Flap conformations of WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} when bound to ATV (A), DRV (B), RTV (C) and SQV (D).

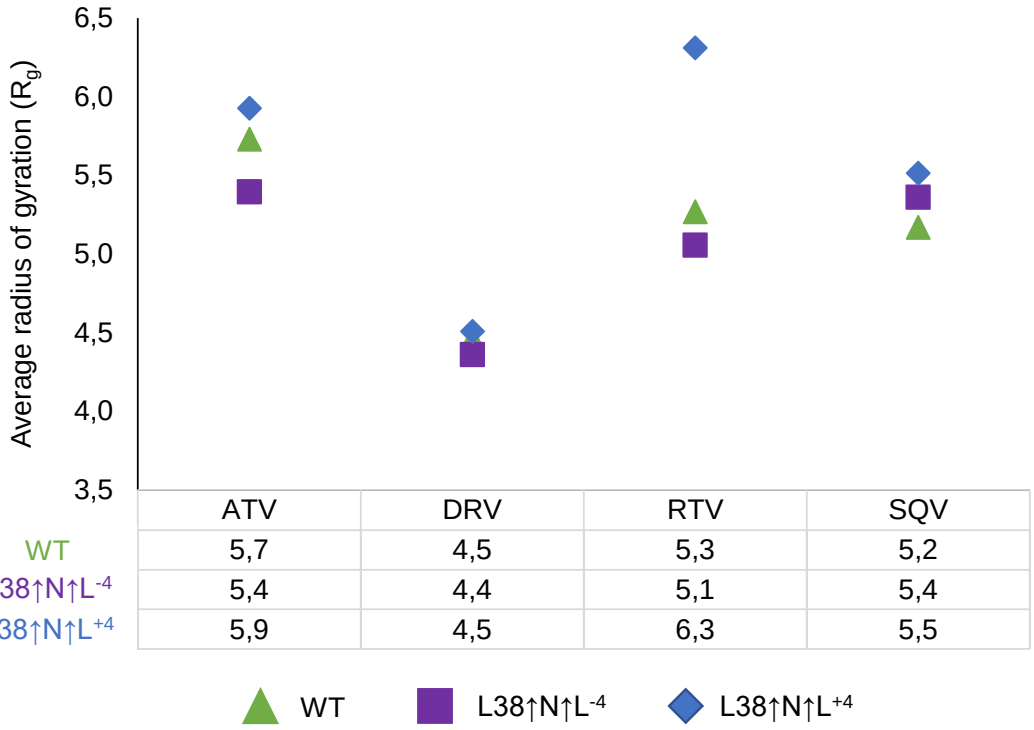
The black and grey dashed lines indicate 17 Å and 22 Å, respectively.

Both chain A and B of all three proteases remained mostly in the closed conformation when in complex with DRV and SQV. However, chain B of the WT and L38↑N↑L^{+4} also sampled the semi-open conformation with SQV and DRV, respectively. When in complex with RTV, chain B of all three proteases remained predominantly in the closed conformation. In contrast, chain A of the WT when in complex with RTV sampled the semi-open conformation exclusively while the mutated proteases sampled both the semi-open and closed conformations. Interestingly, none of the chains exhibited the open conformation (> 2 Å) for any of the proteases either unbound or when complexed with a PI.

3.11.5 Radius of gyration and solvent accessibility of PIs

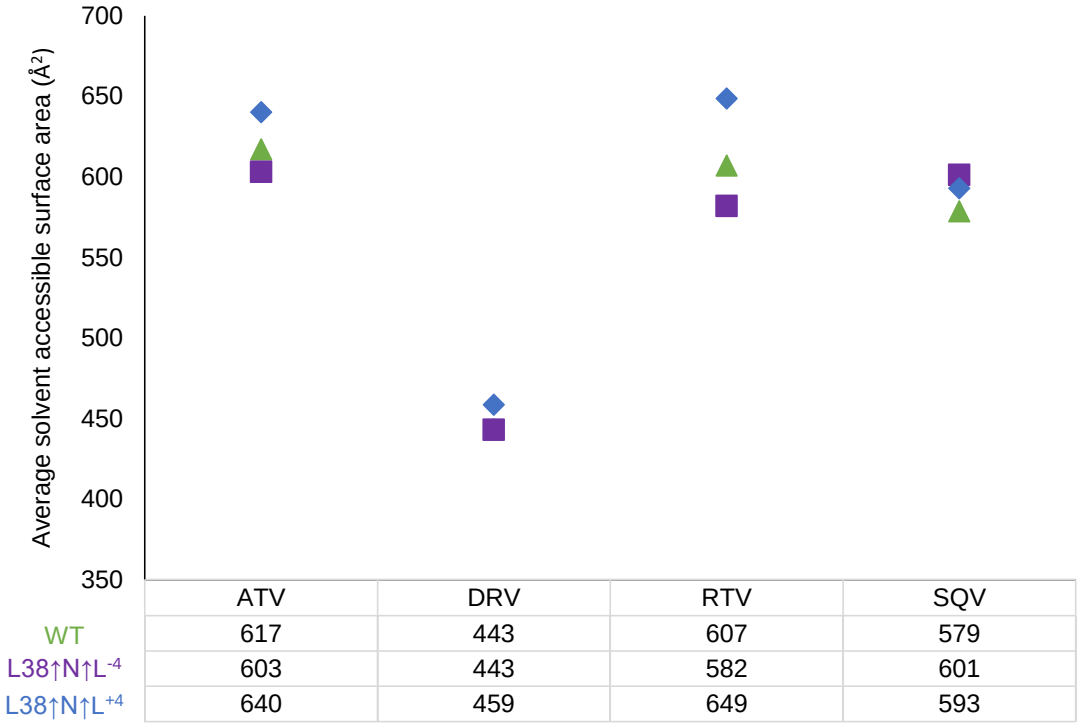
The radius of gyrations (R_g) and differences in the solvent accessible surface are (SASA) of PIs were analysed to assess the degree of compactness of the protease inhibitor complexes during the MD simulations. The average R_g values of the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} in complex with PIs indicated that the compactness of the proteases varied depending on the inhibitor bound (Figure 3.25). The average R_g values for DRV complexes were similar for all three proteases (~4.5 nm). The average R_g values for ATV complexes differed slightly with L38↑N↑L^{-4} displaying a 0.3 nm decrease in comparison to the WT indicating that L38↑N↑L^{-4}•ATV was slightly less dynamic. In contrast, both mutants complexed with SQV exhibited a slight increase in the average R_g value in comparison to the WT indicating that these protease complexes were slightly more dynamic. The L38↑N↑L^{-4}•RTV complex exhibited thelargest difference in the average R_g values indicating that this complex was significantly more dynamic that the WT•RTV complex.

The changes in the global SASA of PIs when bound to the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} were assessed (Figure 3.26). In comparison to the WT, the SASA of ATV and RTV decreased for L38↑N↑L^{-4} but increased for L38↑N↑L^{+4}. This suggests that ATV and RTV is more buried in the active site of L38↑N↑L^{-4} but is more solvent exposed in L38↑N↑L^{+4}. Interestingly, the SASA of DRV was unchanged for the L38↑N↑L^{-4} while it increased for L38↑N↑L^{+4}. The SASA of SQV increased for both mutated proteases when compared to the WT indicating a higher degree of solvent accessibility of the PI for these mutants.



	ATV	DRV	RTV	SQV
WT	5,7	4,5	5,3	5,2
L38↑N↑L^{-4}	5,4	4,4	5,1	5,4
L38↑N↑L^{+4}	5,9	4,5	6,3	5,5

Figure 3.25 The average R_g of the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases when bound to ATV, DRV, RTV and SQV.



	ATV	DRV	RTV	SQV
WT	617	443	607	579
L38↑N↑L^{-4}	603	443	582	601
L38↑N↑L^{+4}	640	459	649	593

Figure 3.26 The average SASA of the WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases when bound to ATV, DRV, RTV and SQV.

3.12 PROTEASE INHIBITOR INTERACTIONS

Protease inhibitors (PIs) are designed to bind at the active site and maintain the protease in the closed conformation thereby suppressing the flexibility of the flaps. Changes in the hinge regions could possibly alter flap dynamics and consequently drug binding interactions between the proteases and the PIs. Therefore, we investigated the interactions formed between ATV, DRV, RTV and SQV with key residues in the active site and flaps of the proteases (**Table 3.11**). The WT protease forms interactions with PIs through active site residues - Asp25, Gly27, Asp29 and/or Asp30 – as well as flap tip residues – Gly48, Gly49 and Ile50. Once again, due to the double insertion at position 38, the mutant proteases have different amino acids in the flap tips. Both L38↑N↑L^{-4} and L38↑N↑L^{+4} flap tip residues corresponding to the WT positions included Met38, Ile49 and Gly50. However, Gly51 and Ile52 was also included in the analysis as these corresponded to the amino acid residues present in the WT. All PIs form interactions with the key residues in the proteases (**Table 3.11**).

While the interaction between Asp25 of the WT protease and all PIs are exclusively hydrogen bonds, L38↑N↑L^{-4} exhibits a water bridge at this position when complexed with ATV and L38↑N↑L^{+4} exhibits the equal presence of hydrogen bonds and water bridges. Additionally, the interaction between Asp25 of L38↑N↑L^{+4} and DRV is a water bridge. Interestingly, Gly27 of L38↑N↑L^{-4} forms more interactions than the WT and L38↑N↑L^{-4}. The interactions formed with Asp29 were altered for all PIs and proteases – with majority of L38↑N↑L^{-4} interactions being water bridges. Interactions with Asp30 were altered as well with L38↑N↑L^{+4} having the least number of interactions. While the Gly48 of the WT interacted with all PIs, the Met38 of L38↑N↑L^{-4} and L38↑N↑L^{+4} showed no interactions between the proteases and any of the PIs. However, the Gly50 of both L38↑N↑L^{-4} and L38↑N↑L^{+4} formed either hydrogen bonds, water bridges or ionic contacts with the PIs. On the other hand, Gly49 of the WT formed minimal interactions with the PIs which was mirrored by L38↑N↑L^{+4} only forming one interaction with RTV. In contrast, Gly51 of L38↑N↑L^{-4} formed interactions with all PIs except for RTV. All isoleucine residues formed hydrophobic interactions with the proteases and in some cases formed hydrogen bonds as well.

Table 3.11 Major interactions formed between PIs and key residues in the active site and flaps of WT, L38↑N↑L^{-4} and L38↑N↑L^{+4}.

	ACTIVE SITE				WT				FLAPS							
	ASP25	ASP25'	GLY27	GLY27'	ASP29	ASP29'	ASP30	ASP30'	GLY48	GLY48'	GLY49	GLY49'	ILE50	ILE50'		
ATV	H	H	W	H	W	H	W	H	H	W	-	-	HI	HI		
DRV	H	H	-	-	W	H	W	H	H	-	-	H	HI	HI		
RTV	H	H	-	H	-	-	-	H	-	W	-	-	HI	HI		
SQV	H	H	H	-	W	-	-	-	I	-	W	-	H	HI		
	ACTIVE SITE				**L38↑N↑L^{-4}**				**FLAPS**							
	ASP25	ASP25'	GLY27	GLY27'	ASP29	ASP29'	ASP30	ASP30'	ILE49	ILE49'	GLY50	GLY50'	GLY51	GLY51'	ILE52	ILE52'
ATV	W	H	W	H	W	H	W	H	HI	HI	H	HI	H	HI	HI/H	HI
DRV	H	H	W	H	W	W	W	-	-	HI	H	H	H	HI	HI/H	HI
RTV	H	H	H	-	W	-	W	-	HI	HI	W	-	-	-	HI	HI/H
SQV	H	H	H	H	W	H	-	H	HI	HI	W	W	W	W	HI	HI
	ACTIVE SITE				**L38↑N↑L^{+4}**				**FLAPS**							
	ASP25	ASP25'	GLY27	GLY27'	ASP29	ASP29'	ASP30	ASP30'	ILE49	ILE49'	GLY50	GLY50'	GLY51	GLY51'	ILE52	ILE52'
ATV	H/W	H/W	H	-	W	-	-	-	HI	-	H	H	-	-	HI	HI
DRV	H	W	-	W	-	W	-	W	HI	HI	W	W	-	-	H	H/HI
RTV	H	H	H	-	H	I	W	-	HI	HI	H	H	-	W	HI	H
SQV	H	H	-	-	H	-	H/W	-	HI	-	W	-	-	-	H	HI

KEY: **Hydrogen bonds** | **Water bridges** | **Hydrophobic interactions** | **Ionic interactions**

4 DISCUSSION

In 2020, 680 000 deaths occurred due to AIDS-related causes emphasising that the HIV/AIDS epidemic remains a global health crisis. This is particularly more concerning for sub-Saharan Africa which contributes ~60% of all new daily infections (UNAIDS, 2021). The introduction of HAART therapy has significantly altered the course of HIV infections to a controllable and treatable long-term chronic disease. Using ARVs in combination with PIs have decreased mortality rates and prolonged life expectancy of HIV-positive individuals (UNAIDS, 2021). However, due to (1) high viral replication in HIV-positive individuals, (2) high mutation rates of HIV due to the low proofreading ability of RT and (3) selection of mutations due to exposure to ARVs, novel compounds to treat HIV are desperately required.

The HIV protease is an extremely attractive drug target considering the necessity of this enzyme to produce new infectious virions. Amino acid insertions selected by ARV therapy rarely occurs in the protease gene. While insertions in the RT are not unusual, they are rare in the proteases with a prevalence of 0.1% (Masquelier *et al.*, 2001; Kozisek *et al.*, 2008). More recently, these insertions have become frequently observed with double insertions at the same amino acid position being observed as well (Ledwaba *et al.*, 2019). Insertions are often repeats of the adjacent DNA sequences occurring in the hinge region. Resistance mutations in the protease are well-characterised and have been shown to play a role in drug resistance (Clavel and Hance, 2004). Limited knowledge about the effect of these insertion mutations on the overall structure, stability, dynamics and drug binding properties of the protease is currently available. Therefore, elucidating the effect of these insertion mutations on the protease could provide insight into the development of novel compounds to target the protease.

4.1 PURIFICATION OF WT AND L38↑N↑L^{-4} PROTEASES

The HIV-1 protease is notoriously difficult to express and purify due to its cytotoxic effects in various host cell lines. The proteases were both expressed in inclusion bodies to minimise autoproteolysis (Ido *et al.*, 1991). By expressing the proteins in inclusion bodies, the proteases cannot adopt its dimeric form thus preventing its

function. Proteins were recovered from inclusion bodies using 8 M urea and allowed to refold into its native state. Both proteases were successfully purified using different protocols of IEC and the monomers corresponded to the predicted sizes calculated by the ExPASy ProtParam tool (Gasteiger *et al.*, 2005). While the SDS-PAGE gel showed no evidence of autocatalytic activity, the protease is an obligate homodimer which can only function if it is folded correctly. Therefore, the sizes and oligomeric states of the proteases were experimentally determined using SE-HPLC. Once again, both proteases provided evidence of the monomeric states of the protein. However, the full dimeric state was absent for the WT and the L38↑N↑L^{-4} protease. An increase in catalytic activity and stability of the protease has been linked to an increase in ionic strength (Polgar *et al.*, 1994; Szeltner and Polgar, 1996). It is possible that the addition of NaCl in the salt buffer used to equilibrate and perform the SE-HPLC experiments increased autocatalysis thereby resulting in cleaved fragments of the proteases (Szeltner and Polgar, 1996).

On account of the autolytic activity of the protease and the refolding step during purification, it is crucial to determine the actual active site concentration of the purified samples. As previously mentioned, the protease only functions as a homodimer and since each monomer contributes an Asp25 residue to the active site, proper folding is essential for the activity of the protease. Active site titrations using ITC was performed on both proteases and the *n*-values were used to assess the percentage of correctly folded protein. The WT was renatured to at least 62% while L38↑N↑L^{-4} was renatured to 50%. The lower percentage observed for the mutant could be due to the prolonged periods of incubation and lengthier purification process of L38↑N↑L^{-4} (~4 days) in comparison to the WT (~3 days). It could be worthwhile to introduce L33I and L63I mutations in addition to the already present Q7K mutation for further studies. The triple Q7K/L33I/L63I mutant has been shown to decrease autolysis while still maintaining function (Mildner *et al.*, 1994). An alternative approach could include the D25A mutation which eliminates catalytic activity thereby eliminating autolysis and increasing protein yield (Darke *et al.*, 1989). However, this mutation could prove problematic if enzymatic experiments need to be performed. In addition, the D25A mutation has also been shown to affect the efficacy of ligand binding and may, therefore, limit the interpretation of data when studying clinical isolates (Sherry, Pandian and Sayed, submitted for review). For these reasons, the D25A mutation was

not introduced into the protein sequences and subsequent experiments were performed using the adjusted concentrations.

4.2 Structural characterisation of proteases

Far-UV CD was used to probe and compare the secondary structural content of the WT and L38↑N↑L^{-4} proteases. The HIV-1 protease is a predominantly β-sheeted protein containing roughly 50% β-sheets (Dash and Rao, 2001; Foulkes *et al.*, 2006; Naicker *et al.*, 2014). The far-UV spectra of both the WT and L38↑N↑L^{-4} proteins exhibited troughs at 218 nm, which is typical of a β-sheeted protein. This data indicates the two insertions in the hinge region do not affect the overall secondary structure of the protease.

While the PDB is rich in HIV-1 protease structures including a variety of inhibitor bound structures, only a handful of these structures are representative of the C-SA protease even though this subtype is the most prevalent worldwide (WHO, 2003). To obtain a better understanding of the structure and function of the C-SA protease, particularly those containing mutations and insertions, more crystal structures are required of both the apo and inhibitor-bound complexes. One of the drawbacks of macromolecular crystallisation is that it requires high concentrations of the target protein. The optimal protein concentration for most macromolecules is 8-20 mg/ml (McPherson and Cudney, 2014). A limitation of this study was the inability to acquire the protease at high concentrations thereby making crystallisation particularly challenging. A few successful crystallisation conditions for the apo WT were identified. The apo L38↑N↑L^{-4} was even more challenging and only produced crystals when in complex with DRV. Unfortunately, diffraction of these crystals was unsuccessful and due to the notoriously difficult nature of the protease and time constraints no further attempts were made to obtain crystals. In this case, it may be advantageous to introduce the D25A mutation in the sequences so that higher concentrations can be achieved.

4.3 Effects of protease insertions on substrate binding

The protease flap region plays a pivotal role in substrate binding with flap residues forming numerous interactions with the substrates (Gustchina and Weber, 1990). The

glycine rich flap tips have increased mobility and curl inward toward the active site before the flaps open providing ample space for the peptide chain to access the active site (Ishima *et al.*, 1999; Scott and Schiffer, 2000). For substrates to bind at the active site, the flaps of the protease needs to be in the open conformation, therefore, making flap movement a necessity for the efficient catalytic activity of the protease (Tozser *et al.*, 1997). The specific activity and catalytic turnover (k_{cat}) of the L38↑N↑L^{-4} protease was significantly reduced by 56% and 55%, respectively. On the other hand, the catalytic efficiency (k_{cat}/K_M) was significantly increased by 61% in comparison to the WT. The reported steady-state kinetic parameters for the WT were similar for the k_{cat}, k_{cat}/K_M and specific activity (Williams *et al.*, 2019). The catalytic efficiency value for L38↑N↑L^{-4} was unusually high as it has previously been shown that hinge region insertions reduce the catalytic efficiency of the protease (Kozisek *et al.*, 2008; Williams *et al.*, 2019). Interestingly, the addition of the four mutations that naturally occur in the L38↑N↑L^{+4} variant does not drastically lower the specific activity and k_{cat} of the enzyme (Table 4.1). However, the catalytic efficiency of L38↑N↑L^{+4} is significantly reduced suggesting that the mutations decrease the specificity of the proteases for the substrate.

Table 4.1 Comparison of steady-state kinetic parameters.

Parameter	WT	L38↑N↑L^{-4}	*WT	*L38↑N↑L^{+4}
Specific activity (µmol^{-1}.min^{-1}.mg^{-1})	22.6±2.0	12.7±1.2	21.0±1.1	12.1±1.1
k_{cat} (s^{-1})	9.1±0.8	4.97±0.6	7.7±0.4	4.5±0.4
k_{cat}/K_M (s^{-1}.µM^{-1})	1.38±0.2	2.3±0.6	1.6±0.4	1±0.004

* Indicates the parameters reported by Williams *et al.* (2019) for the WT and L38↑N↑L^{+4}.

Drug resistance is caused by polymorphisms that alter the balance of recognition to favour the substrate over the inhibitor. Amino acid polymorphisms in the protease along with mutations in the Gag cleavage sites have been shown to result in drug resistance. Often mutations that confer drug resistance results in a loss of replication capacity which is then compensated for by mutations in the Gag sequence. This was

confirmed with the co-evolution of the Gag substrate which restored a mutated protease to WT behaviour (Ozer *et al.*, 2015). Therefore, while the affinity for inhibitors is decreased due to mutations in the protease, substrate processing is maintained by the co-evolution of the Gag substrate (Prabu-Jeyabalan *et al.*, 2004). The specificity of the substrates is not due a particular amino acid sequence but rather a conserved shape defined as a consensus volume of substrates that is recognisable by the protease, termed the 'substrate envelope'. This envelope is created by the packing of substrate residues making substrate recognition interdependent (Prabu-Jeyabalan *et al.*, 2002; King *et al.*, 2004). As a result, a drug resistant mutation in the protease that negatively impacts one substrate position is often compensated by a mutation at a different position within the Gag substrate. This interdependency has been demonstrated through the co-evolution of Gag cleavage sites due to protease mutations, or vice versa (Doyon *et al.*, 1996; Bally *et al.*, 2000; Codoner *et al.*, 2017).

Most primary mutations within the protease occur where the PIs extend outside the substrate envelope and contacts the protease (King *et al.*, 2004). Compensatory mutations in the SP1/p6 and NC/SP1 cleavage sites increases the amount of substrate volume that remains in the dynamic substrate envelope (Ozen *et al.*, 2011; Ozer *et al.*, 2015). The ability of the protease to induce changes in the substrate sequence because of mutations that are caused by an exposure to ARVs demonstrates the interplay between substrate and protein and suggest that residues within the PR•substrate/PI complex system is conserved.

Unsurprisingly, analysis of the Gag precursor L38↑N↑L^{+4} demonstrated the co-evolution of the substrate due to the insertions and mutations present in this protease. The substrate contained mutations in the SP2/NC, NC/SP1 and SP1/p6 cleavage sites (Williams *et al.*, 2019). In addition, this Gag sequence contained a duplication in the SP1/p6 cleavage site and multiple polymorphisms in non-cleavage sites (Williams *et al.*, 2019). A relationship between the alteration in hinge region dynamics and the alteration in substrate cleavage site of the protease has also been established, suggesting that changes in the hinge region will be accompanied by substrate mutations (Ozer *et al.*, 2015). The enzyme assay in this study focused on the cleavage of the CA/SP2 site which is not mutated in the L38↑N↑L^{+4} protease. Therefore, while the cleavage of this specific site might be reduced, it is possible that the mutations in the other Gag cleavage sites may be unaffected or enhanced thereby rescuing activity

of the enzyme. Additionally, the K20R mutation present in the L38↑N↑L^{+4} protease increases viral fitness in mutants that has drug resistance to PIs (Nijhuis *et al.*, 1999) which may restore some functionality to the protease.

4.4 Effects of protease insertions on drug binding thermodynamics

To evade drug binding, the protease adapts by incorporating mutations throughout the protein sequence. Naturally occurring polymorphisms (NOPs) within the C-SA protease have been shown to contribute to drug resistance through reduced affinity to currently available PIs (Mosebi *et al.*, 2008). The hinge region of C-SA which aids the movement and flexibility of the flaps contains many NOPs such as E35D, M36I, S37N, R41K and R57K which may result in reduced binding free energy of PIs (Velazquez-Campoy *et al.*, 2001). Insertions in this hinge region have been suggested to play a role in drug resistance to PIs (Kozisek *et al.*, 2008; Pereira-Vaz *et al.*, 2009). However, due to the rarity at which these insertions occur within the protease their effect on drug binding has not been extensively studied. Amino acid insertions at positions 35, 37 and 38 have been previously identified in subtype B (Kim *et al.*, 2001) and have been reported to contribute to resistance but only in combination with major resistance mutations. These insertions alone have little effect on drug susceptibility *in vitro* but have been shown to affect PI binding to the subtype C protease (Amiel *et al.*, 2011; Grantz Saskova *et al.*, 2014; Zondagh *et al.*, 2018).

Protein-ligand interactions can occur through the non-covalent interactions of hydrogen bonding, electrostatic interactions, hydrophobic interactions and steric interactions. The binding of a ligand to its target is dependent on two essential components, Δ*H* and Δ*S*, which contribute to the overall system. The strength and thermodynamic favourability of an interaction is determined by the Gibbs free energy which can be used to determine the binding affinity:

$$\Delta G = \mathrm{RT}\ln K_D$$

The Δ*G* can also be defined by the enthalpy and entropy changes:

$$\Delta G = \Delta H - \mathrm{T}\Delta S$$

The contacts between inhibitor and the protease often have a substantial buried surface area that is shielded from the solvent contributes favourably towards the Gibbs free energy. Contributions of Δ*H* and Δ*S* are derived from different types of interactions that occurs during drug binding. Binding enthalpy reflects the energetic contribution of hydrogen bonds, van der Waals interactions and electrostatic interactions (polar and dipolar). A negative Δ*H* indicates that these interactions contribute favourably to the binding process and is often indicative of specific protein-ligand interactions (Chen *et al.*, 2016). The binding entropy represents the degree of disorder that is associated with the system. The Δ*S* is determined by the solvation and conformational entropies of the system. If the surfaces that are buried due to binding is mostly hydrophobic, then the change is solvation entropy is favourable and is indicated by a positive Δ*S* value. On the other hand, an unfavourable loss in conformational entropy is observed upon ligand binding due to the reduction in the number of rotameric states the complex can sample. The binding affinity of a ligand can be improved by minimising the unfavourable conformational entropy while maximising favourable enthalpy and solvation entropy.

Owing to the competitive inhibition of PIs, the effectiveness of these inhibitors is directly linked to the affinity at which the bind and inhibit the protease. The first-generation PIs, RTV and SQV, have been entropically optimised and can, therefore, bind the protease with high affinity contributing favourably to the Gibbs free energy of binding. The later-generation PIs, ATV and DRV, binds the protease with strongly favourable enthalpic energy and high affinity (nM range) (Velazquez-Campoy *et al.*, 2001). The stronger binding affinity of these inhibitors are primarily driven by their favourable enthalpic evolution. Therefore, polymorphisms that could reduce protein-inhibitor interactions would, in turn, reduce binding enthalpy resulting in diminished PI efficacy.

The K_D provides an indication of the binding affinity between the PIs and the proteases with a larger K_D indicating weaker binding of the ligand to the protein. The binding affinities of DRV (1 nM) was unchanged for the WT and the L38↑N↑L^{-4} protease. In comparison to the WT the K_D values for ATV, RTV and SQV were 2-, 49- and 1.2-fold larger for L38↑N↑L^{-4}, respectively, indicating that these PIs had weaker binding for the protease when the hinge insertions were present. Particularly, the binding of RTV to L38↑N↑L^{-4} was significantly increased in the presence of these insertions. These

changes in binding affinity suggest that the insertions in the hinge region affect inhibitor binding.

Binding of inhibitors to the proteases results in an unfavourable loss of conformational entropy in the PR•PI system due to the binding partners being unable to sample different rotameric states. Upon binding to the active site, the hydrophobic surface of the PI is desolvated. This results in the increase in the entropy of the surrounding solvent which compensates for the unfavourable loss of conformational entropy in the PR•PI system. Since the change in Gibbs free energy of the system needs to be negative in order for the protein-ligand interaction to be favourable, changes in Δ*S* are often complemented by changes in Δ*H*. This phenomenon is referred to as enthalpy-entropy compensation.

Since Δ*G* is inversely proportional to K_D, a decrease in the favourability of Δ*G* will be observed with an increase in K_D. The decrease in Δ*G* observed for RTV and SQV of L38↑N↑L^{-4} in comparison to the WT suggests that the favourability of ligand binding is affected by the presence of the insertions in the hinge regions. This decrease in Δ*G* observed for RTV and SQV indicates a decreased binding strength in L38↑N↑L^{-4} which is mirrored by the larger K_D values for these PIs. A significant difference was observed in Δ*G* (10 kJ/mol) with RTV as well as the contributions of Δ*H* and TΔ*S*. Binding of RTV was entropically driven (TΔ*S* = 26 kJ/mol) for the WT but enthalpically driven for L38↑N↑L^{-4} (Δ*H* = -39 kJ/mol) demonstrating a change in the initial binding mechanism when the insertions are present. It is possible that the binding energy for RTV mostly originates from the burial of a larger hydrophobic surface upon binding. The presence of water bridges between the PIs and the L38↑N↑L^{-4} protease is consistent with the burial of the hydrophobic area of RTV. This would result in a higher entropy loss which, in turn, would result in a lower binding affinity (49-fold decrease). While the Δ*G* difference in L38↑N↑L^{-4} for SQV was less significant (4 kJ/mol), contributions of Δ*H* and TΔ*S* were significantly altered. Binding of SQV was enthalpically driven for the WT (Δ*H* = -21 kJ/mol) but entropically driven (TΔ*S* = 26 kJ/mol) for L38↑N↑L^{-4}. It is possible that the insertions cause a loss of favourable contacts between PIs and the proteases. ATV was the least affected by the presence of the insertions with minimal changes observed in all three binding parameters between the WT and L38↑N↑L^{-4}. Majority of the binding energy of ATV and DRV for both the WT and L38↑N↑L^{-4} was

contributed by a negative Δ*H* indicating that these reactions were enthalpically driven. However, while the overall binding energetics of ATV was similar for both proteases the binding energetics of DRV was significantly different. Though both reactions with DRV produced the same Δ*G* (-51 kJ/mol), the contributions of Δ*H* and TΔ*S* were more negative for L38↑N↑L^{-4} indicating that this reaction was more enthalpically driven (Δ*H* = -80 kJ/mol) and more entropically opposed (TΔ*S* = -30 kJ/mol) in comparison to the WT (Δ*H* = -62 kJ/mol; TΔ*S* = -12 kJ/mol). The changes in binding affinities of these PIs are further complemented by the changes in their interactions with the protease. Mutations that affect the binding of PIs seems to alter the binding mechanisms between the proteases and the inhibitor. Specifically, insertions in the hinge region have been associated with variations in drug binding energies of PIs (Lockhat *et al.*, 2016).

Drug binding is a dynamic process which can be presented by the on/off rates of the inhibitor and the protease and can be defined as:

$$k_{on}[PR][PI] = k_{off}\,[PR \bullet PI]$$

Where the dissociation constant can be represented as:

$$K_{\mathrm{D}} = \frac{[\mathrm{PR}][\mathrm{PI}]}{[\mathrm{PR} \bullet \mathrm{PI}]} = \frac{k_{off}}{k_{on}}$$

Therefore, if the stability of the PR•PI complexes is lowered in the presence of the insertions, the energetic penalty associated with the structural changes necessary for displacing the inhibitor from the active site would, in turn, be lower. This could result in a higher off-rate of the PI from the active site. Consequently, an increase in the k_{off} would result in an increase in the K_{D}, shifting the equilibrium in favour of the unbound protein and, ultimately, in favour of resistance.

4.5 PROTEASE DIMER STABILITY

The relative stabilities of proteins can be evaluated using DSC, a technique that measures the stabilising forces within a protein (Todd *et al.*, 1998). DSC can also be used to screen the thermal stability of ligand binding. Most biological molecules will undergo phase transitions which can be studied using DSC. Ligand binding will either

increase or decrease the stability of the bound state and can be visualised by a shift in the transition temperature (Johnson, 2013).

The melting curve data for the apo proteases indicated that the two insertions were significant enough to stimulate a substantial change in the overall stability of the apo proteases. In the absence of inhibitors, the L38↑N↑L^{-4} protease was 5°C higher than the apo WT protease indicating increased stability of the apo mutated protease. This observation corresponds to literature, where increased T_m values have been linked to multiple drug resistant mutants (Todd *et al.*, 2000; Yanchunas *et al.*, 2005; Muzammil *et al.*, 2007). The T_m value here for the WT corresponds to the reported value of the C-SA protease which exhibits a higher degree of structural stability than the subtype B protease due to the NOPs present (Velazquez-Campoy *et al.*, 2003). Since the insertions occur in the hydrophobic core (Figure 1.7), the total hydrophobic surface area and side chain packing of the protease is altered. These alterations could be responsible for the changes in the number and strength of interactions that stabilise the native conformation of the protease.

Each monomer of the protease is stabilised by the hydrophobic interactions that occur within the hydrophobic core. The hydrophobic sliding mechanism plays an important role in the movement of the flaps and relies on these internal hydrophobic residues to slide over one another inducing changes in flap conformers (Foulkes-Murzycki *et al.*, 2007). This mechanism is strictly regulated and depends principally on the internal hydrophobic surface area as well as the number of van der Waals contacts involved. L38 is one of the residues involved with this mechanism and it is possible that changes in this hinge region position may alter the stability or flexibility of the protease. The increased stability caused by the insertions could be attributed to the resulting alterations within the hydrophobic core of the protease. Specifically, the addition of another Leu residue could increase the hydrophobicity of the core thereby stabilising interactions between residues involved in the hydrophobic sliding mechanism. These insertions could result in the transfer of van der Waals contacts between core residues facilitating conformational changes that stabilise the overall structure of the protease.

The thermal stability of the PR•PI complexes of both the WT and L38↑N↑L^{-4} are significantly higher than the apo proteases indicating that in the presence of inhibitors, these proteases are more stable. Binding of PIs to the protease brings the complex

into a more energetically stable state which is exhibited by the increase in the T_m values observed with both the WT and L38↑N↑L^{-4} proteases in comparison to the apo counterparts. The L38↑N↑L^{-4} protease exhibited a biphasic transition when melting in the presence of PIs. The existence of two transition peaks with different area ratios for PR•PI complexes were previously reported and has been attributed to the different orientations that PIs can adopt in the active site (Sayer *et al.*, 2008). The binding of DRV and SQV to a L76V mutant exhibited two orientations with decreased stability and higher K_D values (Louis *et al.*, 2011). Both transition peaks of L38↑N↑L^{-4}•PI were lower than the major transition peak of the WT•PI counterparts. The difference observed in the area ratios for each PI indicated that one of the bound orientations is less stable than the other, possibly due to altered contacts. This data indicated that when bound to inhibitors, L38↑N↑L^{-4} is less stable than the bound WT counterparts. Therefore, the insertions in the hinge region decrease the stability of the inhibitor-bound protease complexes. This decrease in stability is possibly attributed to the ease with which these PIs can dissociate from the variant protease which is echoed by the increase in K_D values. PIs bind the protease with high affinity which results in a lowered off-rate. Therefore, if the energetic penalty associated with the dissociation of PIs is decreased, the k_{off} would, in turn, be increased. The higher off-rate of the inhibitor would shift the equilibrium in favour of dissociation since the energy required to separate the protease from the PI is lowered. The accompanying decrease in Gibbs free energy implies an increase in dissociation constant therefore resulting in the reduced stability of these complexes. The decrease in the stability of the inhibitor-bound complexes could also be attributed to the changes occurring within the hydrophobic core which may stabilise the mutant in the apo form but instead minimise interactions with PIs. The structural analysis of PR•PI complexes have linked primary mutations to a change in the active site geometry which directly affects PR•PI interactions due to steric effects and/or lost interactions (Ohtaka *et al.*, 2002). Similarly, these insertions in the hinge region could play a role in drug resistance using an analogous mode of action.

4.6 MOLECULAR DYNAMICS OF APO PROTEASES

The hinges have been identified as the key mechanistic region of the protease controlling the conformational combinations. Often catalytic residues that are unique to certain structures are positioned close to the hinges (Yang and Bahar, 2005). Therefore, the hinges predominantly determine the flexibility and dynamics of the protease by triggering correlated movements of residues in the other regions (Zheng and Brooks, 2005). These hinges display a varying degree of fluctuation depending on the substrate bound (Ozer *et al.*, 2010).

The RMSD and R_g values indicated differences in the overall dynamic motion and degree of compactness between the apo proteases. The addition of the insertions and mutations resulted in increased overall structural stability of the mutated proteases (Figure 3.19, Figure 3.20). Overall, the L38↑N↑L^{-4} was the most compact and least dynamic protease. The WT displayed the highest average RMSD indicating decreased overall structural stability and, therefore, greater dynamic movements. Interestingly, addition of the mutations with the insertions further decreases the stability of the protease resulting in increased overall dynamics of L38↑N↑L^{+4}, but a lower degree of compactness.

RMSF analysis indicated that the hinge regions of both mutated proteases were more flexible than the WT. Notably, inclusion of the mutations further increased the flexibility of the hinge region. Furthermore, the cantilever and flaps of L38↑N↑L^{+4} displayed higher degrees of structural stability than the other two proteases. The cantilever and flaps of the WT protease displayed the highest degree of fluctuation. A reduction in flap flexibility can reduce the affinity of the protease for substrates and PIs (Collins *et al.*, 1995; Mahalingam *et al.*, 2001; Mahalingam *et al.*, 2002).

The various conformations that the flaps adopt in solution allows for substrate and inhibitor binding. These conformations; namely, open, semi-open and closed, can be measured using either the distance between the two Ile50/Ile50' residues or by using the distance between the Ile50 and Asp25 residues (Wlodawer *et al.*, 1989; Perryman *et al.*, 2004). PIs are substrate transition state analogues that competitively bind at the active site (Toth and Borics, 2006). When an inhibitor binds, the flaps close over the active site adopting the closed conformation. With the inhibitor bound, the flaps remain in the closed conformation due to the high energetic penalty required to change the

flaps to the open conformation (Louis and Roche, 2016). Mutations that shift the equilibrium of these conformations may, in turn, affect drug binding. Drug resistance could be acquired by either directly changing the relative stability of the apo and inhibitor bound protease conformations which correspond to the semi-open and closed conformations, respectively (Todd *et al.*, 2000). These drug resistance mutations stabilise the semi-open apo conformation which decreases drug binding affinity due to the high energetic barrier the protease would have to overcome to bring the flaps into the closed conformation to allow for adequate drug binding.

In the absence of inhibitors, both mutated proteases maintained the closed conformation while the WT sampled all three conformations. Evidently, the insertions alone are sufficient to shift the conformation equilibrium. The ability of both mutants to stabilise and maintain a more closed conformation would decrease the accessibility of PIs and substrates to the active site cavity. If the flaps are predominantly in the closed conformation, substrates cannot be processed efficiently. A more closed conformation would result in a decrease in substrate/protease association rates and, therefore, a decrease in k_{cat} and specific activity. This correlates with our steady-state kinetics data which shows that the enzymatic properties of the mutants are hindered.

Additionally, a decrease in RMSF for the flap regions of the mutants was observed in comparison to the WT suggesting that the insertions minimise the movement of the flaps to stabilise the closed conformation of the protease. The overall stability of the mutated protease is also increased by these insertions as indicated by the higher T_m values. The addition of the mutations further decreases the RMSF of the flap region, suggesting that these mutations play a cooperative role in maintaining the closed conformation of the protease. This stabilised closed conformation of the apo mutated proteases would require a substantial amount of energy to open the flaps, allow entry of the PIs and to subsequently close the flaps to allow for binding to occur. This would, therefore, result in decreased drug binding of PIs due to the increased energetic penalty associated with changing the conformations of the flaps. If the closed conformation is favoured, this presents a possible mechanism by which the mutated proteases could evade drug binding.

4.7 PROTEIN-LIGAND DYNAMICS

The movement and stability of the flaps is known to be closely associated and dependant on the hinge regions of the protease with changes in flexibility of the hinge impacting the flaps (Gustchina and Weber, 1990). The joint movement of the fulcrum, hinge and cantilever regions is proposed to shift the conformations of the flaps from closed to open conformations (Hornak *et al.*, 2006; Foulkes-Murzycki *et al.*, 2007). As a result, mutations that alter the stability of the hinge region may in turn lead to changes in the movement of the flaps.

The compactness of the L38↑N↑L^{+4} protease exhibited slightly greater overall dynamics compared to the WT while L38↑N↑L^{-4} exhibited lower overall dynamics when complexed with PIs. In the presence of PIs, both mutants displayed the greatest increase in RMSF in the hinge regions indicating increased mobility of the hinges when in complex with inhibitors. This translated to an increase in mobility of both L38↑N↑L^{-4} flaps when in complex with ATV, DRV and RTV. Interestingly, both flaps of L38↑N↑L^{-4} was less mobile than the WT when complexed with SQV. The two flaps of L38↑N↑L^{+4} displayed differential mobility owing to the asymmetric nature of PIs which results in different contacts of each monomer depending on the orientation of the PI in the active site. At least one of the L38↑N↑L^{+4} flaps displayed an increase in mobility when complexed to all four PIs. This data suggests the insertions increase the mobility of the hinge regions which correlates to an increase in the movement of flaps when inhibitors are present. While still maintaining increased mobility in the flap and hinge regions, the four mutations seem to stabilise these regions which may facilitate substrate binding while still evading drug binding. It is possible that the changes in flap conformation of these mutants could provide a mechanism for drug resistance due to the failure of the PIs to induce adequate flap closing (Hornak and Simmerling, 2007).

Consequently, the changes in the flap and hinge regions affected the movement of the PIs in the active site. The movement of ATV, DRV and RTV was reduced in the active site of L38↑N↑L^{-4}. The movement of SQV in the L38↑N↑L^{-4} active site was similar to the WT and could be attributed to the decreased flap mobility of this protease when complexed with this inhibitor. The presence of the mutations further altered the movement of the PIs. Interestingly, SQV along with DRV and RTV now displayed a decrease in mobility with the active site while movement of ATV was similar to the WT

suggesting that the mutations most likely contribute to protein-inhibitor interactions. The changes in the mobility of the PIs within the active site and the changes in flap flexibility will subsequently affect the nature and number of the intermolecular interactions within the active site.

According to the Stanford database, L38↑N↑L^{+4} does not contain any drug resistance mutations although it does contain the secondary mutation, V82I, which has not been linked to any PI resistance. This mutated protease also includes two hinge region mutations, E35D and R57K, as well as the double insertion of Asn and Leu in this region. Large changes in the hinge and loop regions have been associated with E35D, M36I and S57N indicating the importance of this region of the motion of the protease (Agniswamy *et al.*, 2016). The smaller side chain of E35D results in the loss of the ion pair between the side chains of Glu35 and Arg57 and instead Asp25 forms the ion pair with K20R (Agniswamy *et al.*, 2016). Therefore, it is possible that K20R, E35D, R57K and the insertions at L38 synergistically disrupt bonds resulting in the increased flexibility of the flaps and the enhanced ability to dissociate from inhibitors.

The SASA data further confirms the role of these insertions and mutations on ligand binding. The decrease in SASA of ATV and RTV in L38↑N↑L^{-4} correlates to the decreased mobility of these buried PIs within the active site as well as the decrease in favourable entropy for these PIs. The decreased SASA and diminished entropic favourability of binding indicates a decrease in exposure of these PIs to the surrounding solvent further supporting the change in binding mechanism in the presence of these insertions. The SASA of all PIs were increased in L38↑N↑L^{+4} indicating that more of the inhibitor was surrounded by water. It is possible that the mutations increase the entropic favourability of binding in this mutant which could explain the increase in SASA of these ligands. This could further be elucidated by the shift in flap conformations of the mutated proteases as well as the increase flexibility of the hinge and flap regions. *In vitro* ITC analysis of the L38↑N↑L^{+4} protease would be key in determining the entropic contributions of these PR•PI complexes.

While the apo mutated proteases sampled only closed conformations, when bound to ATV, DRV and RTV, the conformations of these mutated proteases included both the semi-open and closed flap conformers. During sizable movements of the protease, the network of hydrogen bonds in the backbone are maintained while the van der Waals

contacts freely exchange between residues in the hydrophobic core (Ragland *et al.*, 2014) The exchange of these contacts between residues in the hydrophobic core facilitates the movement of the flaps from one conformation to another. Therefore, the location of these insertions, K20R and R57K may provide insight into the nature of their function in reduced drug susceptibility.

Altogether, this data confirms that the insertions increase the flexibility of the flap and hinge regions of the protease which, in turn, changes the mobility of the PIs in the active site. Particularly, the location of these insertions in the protease may serve to stabilise the hydrophobic core of the protease which may complement the flexibility of the flap regions, Figure 4.1. Furthermore, the presence of the mutations serves to stabilise these regions and maintains the apo protease in the closed conformation possibly further reducing drug binding.

4.8 INDUCED-FIT DOCKING STUDIES

Induced-fit docking studies (IFDS) is thought to be representative of what occurs *in vitro* and is often the preferred method of docking as it includes the motions of both the protein and ligand. IFDS results indicated that all four PIs (ATV, DRV, RTV and SQV) bind the proteases albeit with different affinities as witnessed by the difference in the GlideScores (G_{score}) between the three proteases. With the exception of ATV, the data correlated to the ITC K_D values indicating that L38↑N↑L^{-4} binds RTV and SQV with weaker affinity while DRV was mostly unaffected by the presence of the insertions. The addition of the four mutations seems to further affect the binding affinity of all PIs. Notably, these additional mutations altered binding of DRV to L38↑N↑L^{+4} resulting in weaker binding of the inhibitor. Moreover, the presence of the mutated Gag sequence resulted in a decrease in susceptibility to DRV (Williams *et al.*, 2019). The same was found with RTV, where the insertions and mutations further decreased binding of this inhibitor to the protease. Surprisingly, these mutations resulted in tighter binding of SQV to the pro*tease which overshadowed both the WT and L38↑N↑L^{-4}.

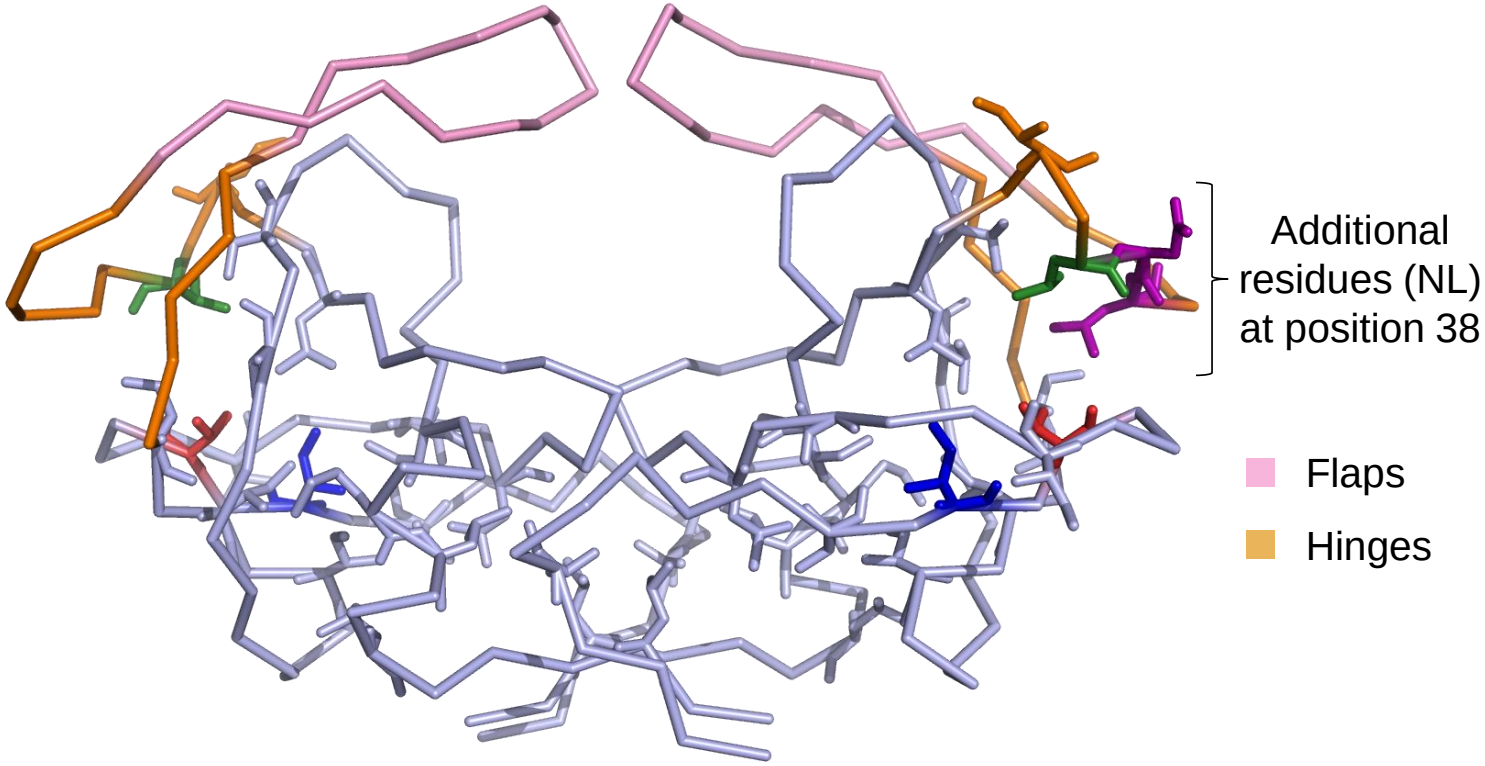


Figure 4.1 Hydrophobic amino acid side chains of the HIV-1 protease.

The flaps and hinges of the HIV-1 protease are shown in pink and orange, respectively. All twenty amino acid residue side chains (L5, V11, I13, V15, I19, A22, L24, I26, L33, L38, I62, I64, I66, V75, V77, I85, M89, L90, L93, L97) involved in the hydrophobic core are indicated as sticks (Foulkes-Murzycki *et al.*, 2007; Naicker and Sayed, 2014). The location of L38 is shown in green while the location of the NL double insertion is shown in purple. The I13 and V15 residues of the hydrophobic core are shown in blue and red, respectively. The image was constructed using the PyMOL v0.99rc6 software using PDB ID: 3U71.

4.9 BINDING INTERACTIONS BETWEEN PROTEASES AND INHIBITORS

The active site residues (Asp25, Gly27, Asp29 and/or Asp30) and the flap residues (Gly48, Gly49 and Ile50) are known to be associated with substrate and inhibitor binding (Wlodawer and Erickson, 1993; Prabu-Jeyabalan *et al.*, 2002; Nalam *et al.*, 2010). The binding of the substrate seems to be defined by the asymmetrical shape of the substrate rather than a specific amino acid sequence. The protease can bind a number of substrates with varying amino acid sequences that occupy a consensus volume within the active site. These substrates form conserved hydrogen bonds with Asp25, Gly27, Asp29, Gly48 and a partial hydrogen bond with Asp30. A conserved water molecule facilitates contacts with the flap tip residues, Ile50 (Prabu-Jeyabalan *et al.*, 2002).

While majority of the active site residues exhibit the conserved hydrogen bond between the PI and the above-mentioned residues, the L38↑N↑L^{-4} protease incorporates several water bridges between the active site residues and PIs. The incorporation of these transient water molecules between the PI and the active site would noticeably reduce the enthalpic favourability of drug binding thereby reducing the frequency of hydrogen bonds formed between the PI and the active site. Moreover, the addition of the mutations to the insertion mutant results in a decrease of overall contacts (hydrogen bonds and water bridges) between the PI and the active site.

Due to the insertions, a methionine amino acid was found in position 48 instead of Gly48 at the flap tips. This amino acid change resulted in no interaction between the mutated proteases and the PIs at this position while the WT exhibited water bridges for ATV, RTV and SQV at this position (**Table 3.11**). The water-mediated hydrophobic interaction with the two Ile50 residues is conserved in the WT for all inhibitors. The corresponding Gly50 in the mutated proteases show no evidence of hydrophobic interactions at this position. However, the corresponding Ile amino acid at position 52 in the mutated proteases showed a variation in the number of hydrophobic contacts. Overall, an increased number of hydrophobic interactions was observed at the flaps in the presence of the insertions. This, together with the multiple water bridges, could explain the low binding affinity of the inhibitors for the mutated protease. It is possible that the stronger flap interactions prevent the PI from forming the necessary contacts with the active site cavity. In addition, the increased mobility of the flaps and the

reduced number of hydrogen bonds could potentially diminish contact with the active site cavity. The WT protease requires a single water molecule for the PI to form interactions with the flap tips (Wlodawer and Erickson, 1993). The increase in flexibility of the flaps could be due to the coordination of multiple water bridges within the active site cavity.

The interaction between the carboxyl group of the active site Asp25 residues and the hydroxyl group of the inhibitor is often mediated by a hydrogen bond resulting in drug binding (Lv *et al.*, 2015). The interactions between these catalytic residues and the PIs are altered for the C-SA proteases (Figure 4.3, Figure 4.4, Figure 4.5, Figure 4.5). While the insertion alone does not drastically change the binding of ATV between the WT and L38↑N↑L^{-4}, the inclusion of the mutations changes the interactions with this PI resulting in a shared hydroxyl group between the two Asp25 residues and Gly27 (Figure 4.2). Interestingly, no contact is made between the hydroxyl group of DRV and the active site residues (Figure 4.3). Instead the mutants form either a hydrogen bond or water bridge. Both catalytic Asp25 residues bind to the same oxygen forming a water bridge with RTV in the presence of the insertions. The conserved hydrogen bond is seen with the WT and L38↑N↑L^{+4} when binding to RTV while the binding of SQV to the proteases seem to be mediated by the Gly27 amino acid (Figure 4.4, Figure 4.5). In addition to changes in the interaction between proteases and PIs, it is evident that the insertions and mutations alter the orientation of the inhibitors within the active site cavity. PIs are designed to provide a large degree of specificity for the active site of the protease becoming rigid peptides when bound. However, this rigidity poses a limitation on the ability of the PI to adapt to mutations that cause structural changes within the active site. The difference observed in orientations of the inhibitors could be attributed to a variation in the shape, charge and size of the active site cavity when the insertions and mutations are present. This in turn forces the PR•PI complexes into different orientations and conformations until the energy minima are established.

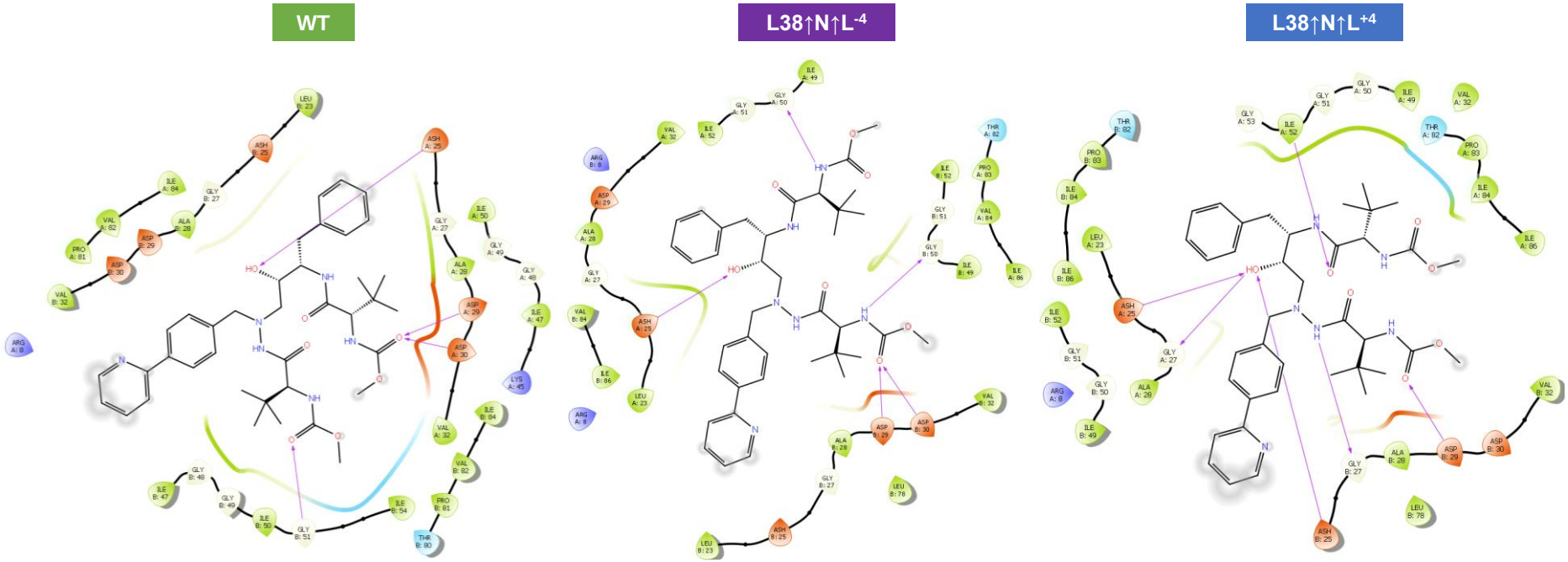


Figure 4.2. Amino acid interactions of proteases with ATV.

Green: hydrophobic amino acids, Turquoise: polar amino acids, purple arrow: hydrogen bonds, red: negatively charged amino acids and blue: positively charged amino acids. ASH indicates that the aspartic acid at position 25 in the active site is uncharged. The image was generated using the Glide software (Schrödinger LLC 2009, USA).

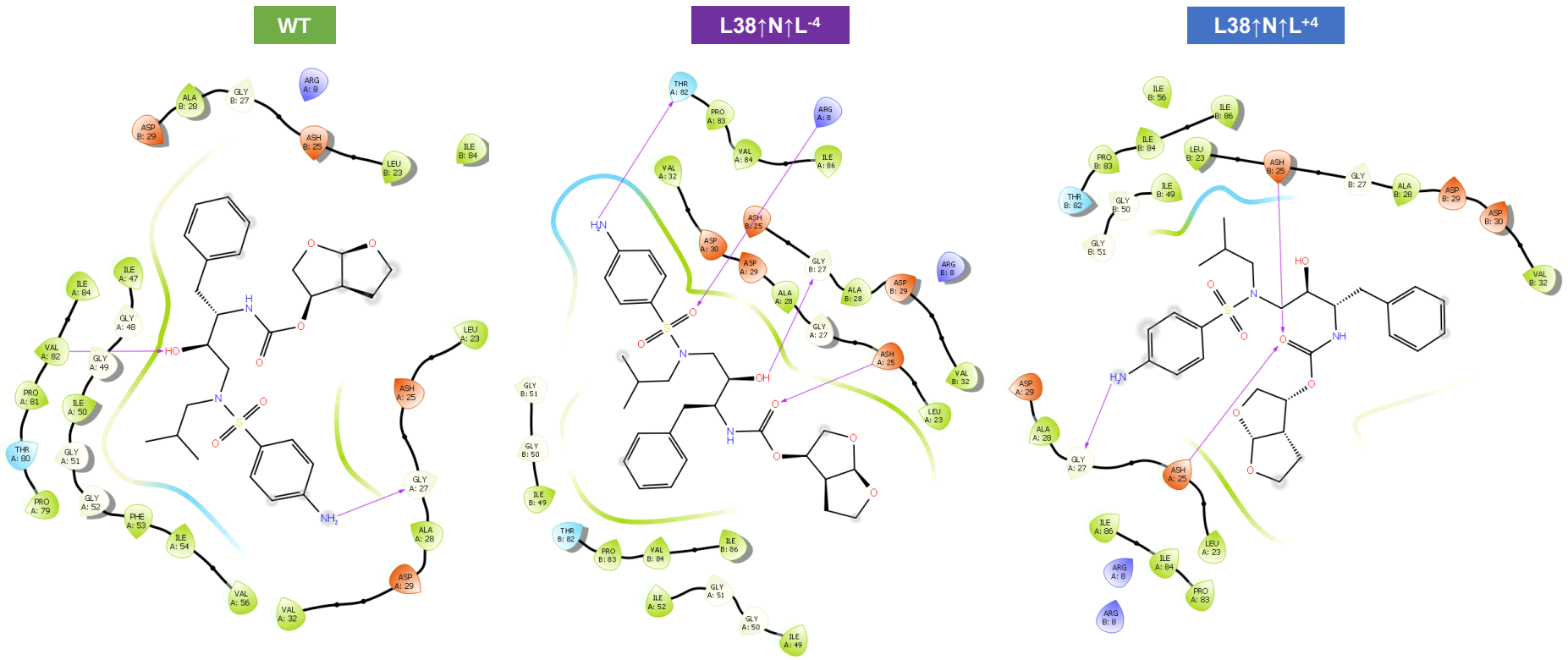


Figure 4.3 Amino acid interactions of proteases with DRV.

Green: hydrophobic amino acids, Turquoise: polar amino acids, purple arrow: hydrogen bonds, red: negatively charged amino acids and blue: positively charged amino acids. ASH indicates that the aspartic acid at position 25 in the active site is uncharged. The image was generated using the Glide software (Schrödinger LLC 2009, USA).

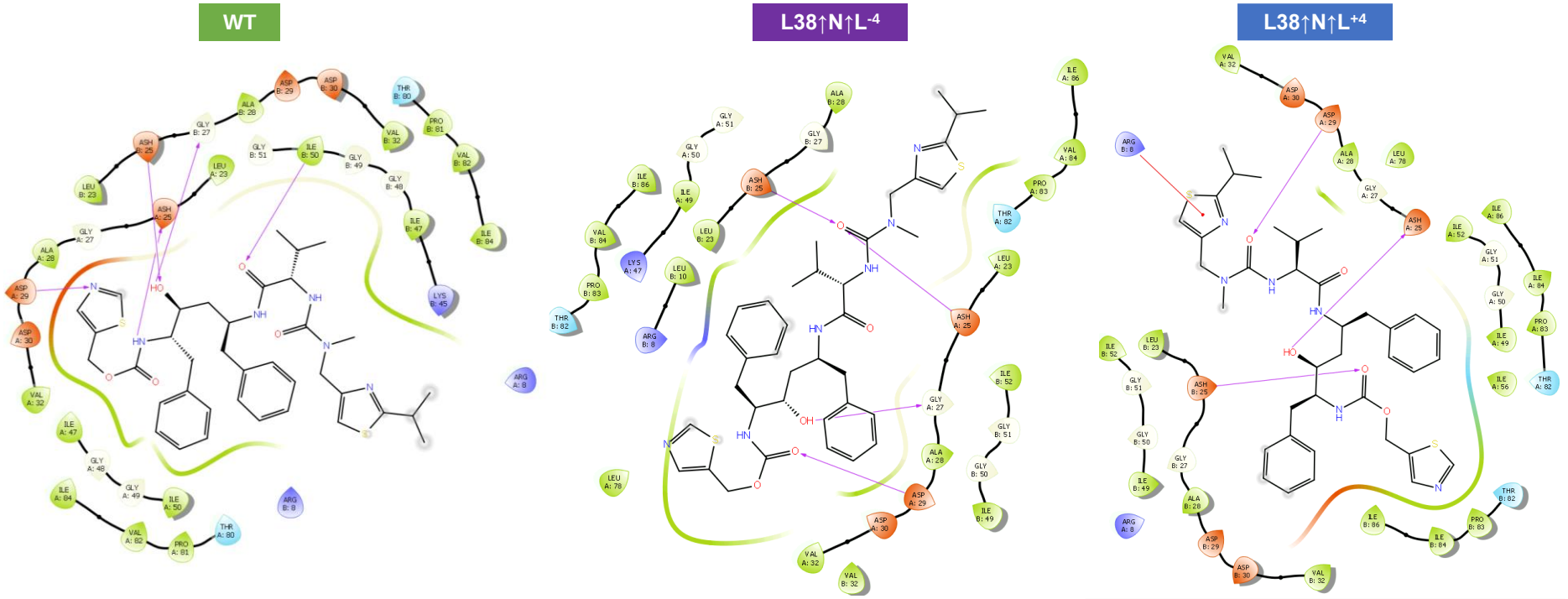


Figure 4.4 Amino acid interactions of proteases with RTV.

Green: hydrophobic amino acids, Turquoise: polar amino acids, purple arrow: hydrogen bonds, red: negatively charged amino acids and blue: positively charged amino acids. ASH indicates that the aspartic acid at position 25 in the active site is uncharged. The image was generated using the Glide software (Schrödinger LLC 2009, USA).

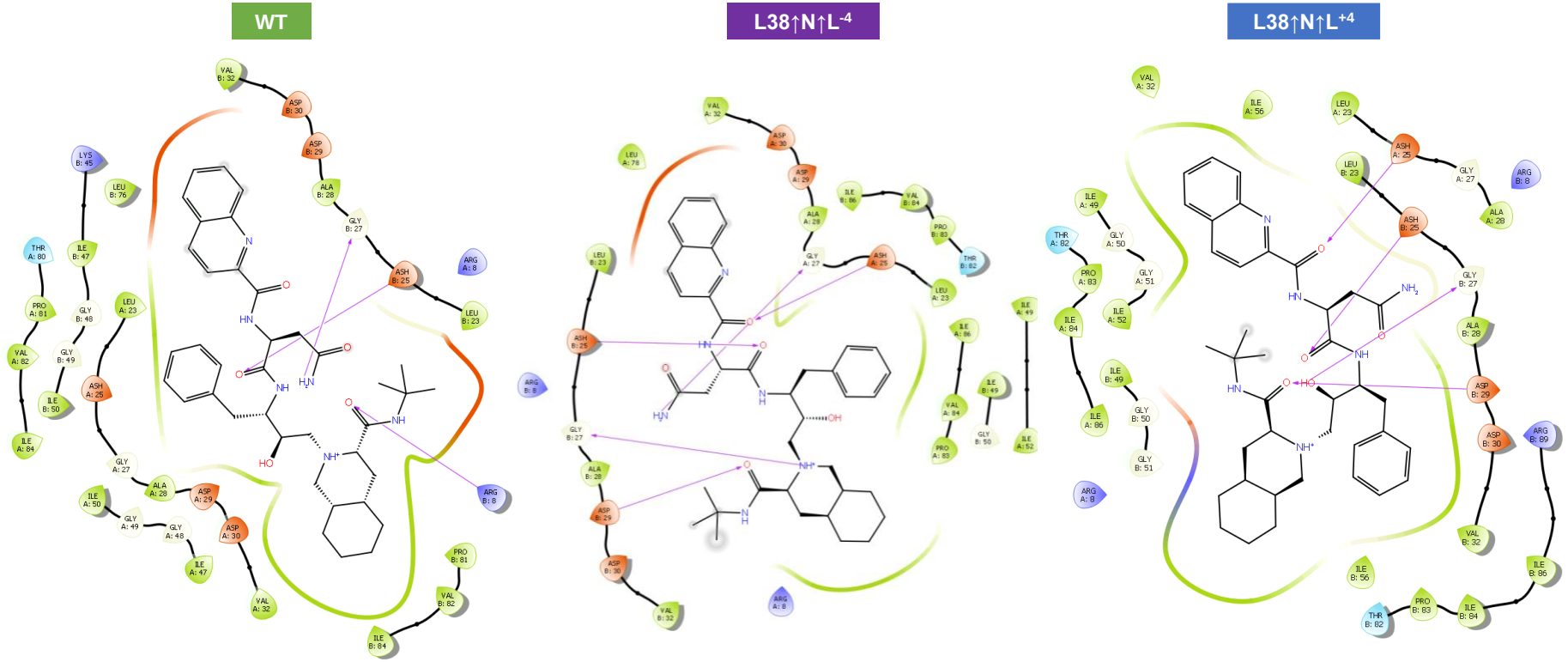


Figure 4.5 Amino acid interactions of proteases with SQV.

Green: hydrophobic amino acids, Turquoise: polar amino acids, purple arrow: hydrogen bonds, red: negatively charged amino acids and blue: positively charged amino acids. ASH indicates that the aspartic acid at position 25 in the active site is uncharged. The image was generated using the Glide software (Schrödinger LLC 2009, USA).

Since the hinge and flap regions play a significant role in binding, changes in their stability and dynamics could affect drug binding. Except for Asp25 and Asp29, majority of the amino acids that make up the core are hydrophobic and form hydrogen bonds with the main chain. The Arg8, Asp30 and Lys45 amino acids of the protease interact with the polar side chain of peptides (Weber and Agniswamy, 2009). PIs were designed to fit into the hydrophobic core and as such are mainly hydrophobic as well. A reduction in hydrophobic contacts is seen with drug resistant mutants (Weber and Agniswamy, 2009). Any changes to the hinge region could alter the hydrophobic interactions within the core and could potentially modify the accessibility of PIs to the active site (Goldfarb *et al.*, 2015). Moreover, L38 is directly involved in the hydrophobic sliding mechanism, making contacts with I15 and I13 to alter the flap conformation, Figure 4.1. The decrease of hydrophobic contacts observed with ATV, DRV and RTV in the presence of both the insertions and mutations could provide an explanation for the decrease in drug binding affinity. Conversely, the hydrophobicity in the active site pocket is increased when SQV is bound to L38↑N↑L^{+4} suggesting that this inhibitor may be the most suitable for ARV therapy. This notion is further supported by the increased binding affinity of the mutated protease for this inhibitor.

4.10 Concluding remarks

The impact that amino acid insertions have on the HIV-1 protease may have been significantly underestimated due to the scarcity of these polymorphisms. However, these polymorphisms clearly play a major role in drug binding thermodynamics and interactions between the currently available PIs and the protease. Additionally, a shift in conformational states of the protease is caused as a result of these insertions. Palpably, these specific hinge insertions do not require any additional primary or secondary mutations to elicit a response in protease stability or flexibility. It should be noted that these insertions primarily occur in the hinge region, specifically residues involved in the hydrophobic core. Changes to this hydrophobic core leads to a change in the hydrophobic sliding mechanism resulting in a more stabilised or compact core which may complement the flexibility of the flap regions. The changes also elicit an alteration in the interactions between current inhibitors and the active site. The mechanism by which these insertions could provide drug resistance lies in their ability to create stronger interactions at the flaps while incorporating water molecules around the active site thereby increasing the dissociation rate of the inhibitors. The correlation of higher K_D values, lower stability of PR•PI complexes, increased flap and hinge region dynamics and the reduced number of hydrophobic interactions could illustrate a mechanism of drug resistance of the insertion mutant proteases. It would be interesting to find out if the *in vitro* effects of the naturally occurring mutations of this mutated protease further affect drug binding and stability of the protease. The current data suggests that a patient affected with this specific L38↑N↑L^{+4} clinical isolate would highly likely benefit from a tailor-made treatment containing SQV.

This study highlights the major contribution of hinge region insertions on the HIV-1 protease and may provide an alternative, novel target for future inhibitor designs. A therapeutic that could target the hinge region would provide a number of advantages: (1) binding of the inhibitor could be targeted to the hinge residues found on the outside of the protease thereby avoiding issues related with flap conformations, (2) any changes to the flexibility of the flaps should not impose negatively on the binding of the inhibitor to the hinge region and (3) restricting the movement of the hinges will, in turn, limit flap movements thus preventing substrate processing and the subsequent proliferation of the virus. This study shows that it may be beneficial to target the hinge

region of the protease, whether insertions are present or not, given the importance of this region in the HIV-1 protease.

4.11 Future Perspectives

To eliminate the issue of autoproteolysis, the L33I and L63I mutations could be added as the triple Q7K/L33I/L63I mutant has been shown to decrease autolysis while still maintaining function (Mildner *et al.*, 1994). An alternative approach could be to include the D25A mutation which eliminates catalytic activity thereby eliminating autolysis and increasing protein yield (Darke *et al.*, 1989). An increase in the protein yield could be advantageous when attempting crystallisation of the proteases.

Since the IFDS showed that the inclusion of the four background mutations changed the binding affinity of the PIs to the mutant protease, displacement titrations of the L38↑N↑L^{+4} protease with ATV, DRV, RTV and SQV should be performed.

An additional study could look at the peptide fragments generated during autoproteolysis and interaction with PIs which could determine if the PIs bind to the peptide fragments and if this contributes to the data collected from ITC.

5 SUPPLEMENTARY MATERIAL

Appendix 1. ProtParam of WT C-SA

User-provided sequence:

```
        10          20          30          40          50          60
PQITLWKRPL   VSIKVGGQIK   EALLDTGADD   TVLEEINLPG   KWKPKMIGGI   GGFIKVRQYD

        70                                  80                          90
QILIEICGKK AIGTVLVGPT PVNIIGRNML TQLGCTLNF
```

Number of amino acids: 99

Molecular weight: 10754.83

Theoretical pI: 9.32

Amino acid composition: CSV format

Ala (A)	3	3.0%
Arg (R)	3	3.0%
Asn (N)	4	4.0%
Asp (D)	4	4.0%
Cys (C)	2	2.0%
Gln (Q)	5	5.1%
Glu (E)	4	4.0%
Gly (G)	13	13.1%
His (H)	0	0.0%
Ile (I)	13	13.1%
Leu (L)	11	11.1%
Lys (K)	9	9.1%
Met (M)	2	2.0%
Phe (F)	2	2.0%
Pro (P)	6	6.1%
Ser (S)	1	1.0%
Thr (T)	7	7.1%
Trp (W)	2	2.0%
Tyr (Y)	1	1.0%
Val (V)	7	7.1%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 8
Total number of positively charged residues (Arg + Lys): 12

Atomic composition:

Carbon	C	489
Hydrogen	H	810
Nitrogen	N	128
Oxygen	O	134
Sulfur	S	4

Formula: $C_{489}H_{810}N_{128}O_{134}S_4$
Total number of atoms: 1565

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 12615
Abs 0.1% (=1 g/l) 1.173, assuming all pairs of Cys residues form cystines

Ext. coefficient 12490
Abs 0.1% (=1 g/l) 1.161, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is P (Pro).

The estimated half-life is: >20 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
? (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 38.44
This classifies the protein as stable.

Aliphatic index: 118.08

Grand average of hydropathicity (GRAVY): 0.180

Appendix 2. ProtParam of L38↑N↑L^{-4} sequence

User-provided sequence:

```
        10          20          30          40          50          60
PQITLWKRPL   VSIKVGGQIK   EALLDTGADD   TVLEEINLNL   PGKWKPKMIG   GIGGFIKVRQ

        70                  80                  90                  100
YDQILIEICG KKAIGTVLVG PTPVNIIGRN MLTQLGCTLN F
```

Number of amino acids: 101

Molecular weight: 10982.09

Theoretical pI: 9.32

Submit

Amino acid composition:		
Ala (A)	3	3.0%
Arg (R)	3	3.0%
Asn (N)	5	5.0%
Asp (D)	4	4.0%
Cys (C)	2	2.0%
Gln (Q)	5	5.0%
Glu (E)	4	4.0%
Gly (G)	13	12.9%
His (H)	0	0.0%
Ile (I)	13	12.9%
Leu (L)	12	11.9%
Lys (K)	9	8.9%
Met (M)	2	2.0%
Phe (F)	2	2.0%
Pro (P)	6	5.9%
Ser (S)	1	1.0%
Thr (T)	7	6.9%
Trp (W)	2	2.0%
Tyr (Y)	1	1.0%
Val (V)	7	6.9%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 8
Total number of positively charged residues (Arg + Lys): 12

Atomic composition:

Carbon	C	499
Hydrogen	H	827
Nitrogen	N	131
Oxygen	O	137
Sulfur	S	4

Formula: $C_{499}H_{827}N_{131}O_{137}S_4$
Total number of atoms: 1598

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 12615
Abs 0.1% (=1 g/l) 1.149, assuming all pairs of Cys residues form cystines

Ext. coefficient 12490
Abs 0.1% (=1 g/l) 1.137, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is P (Pro).

The estimated half-life is: >20 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
? (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 37.87
This classifies the protein as stable.

Aliphatic index: 119.60

Grand average of hydropathicity (GRAVY): 0.179

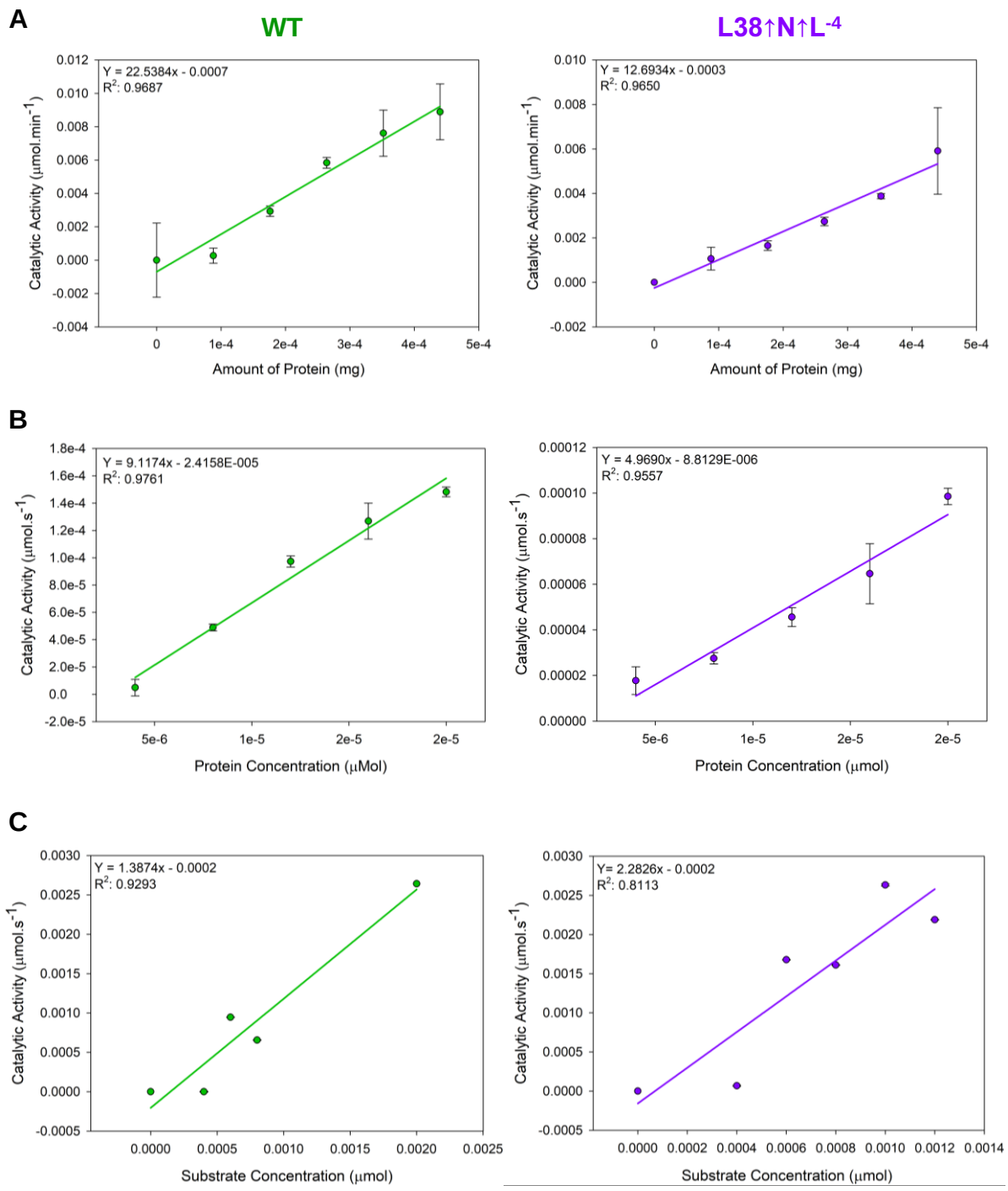


Figure S1 Steady-state kinetics data for the WT and L38↑N↑L^{-4} proteases.

A: Specific activity, **B:** k_{cat} and **C:** catalytic efficiency. The graphs are representative of three separate experiments. Graphs generated using SigmaPlot, version 12.

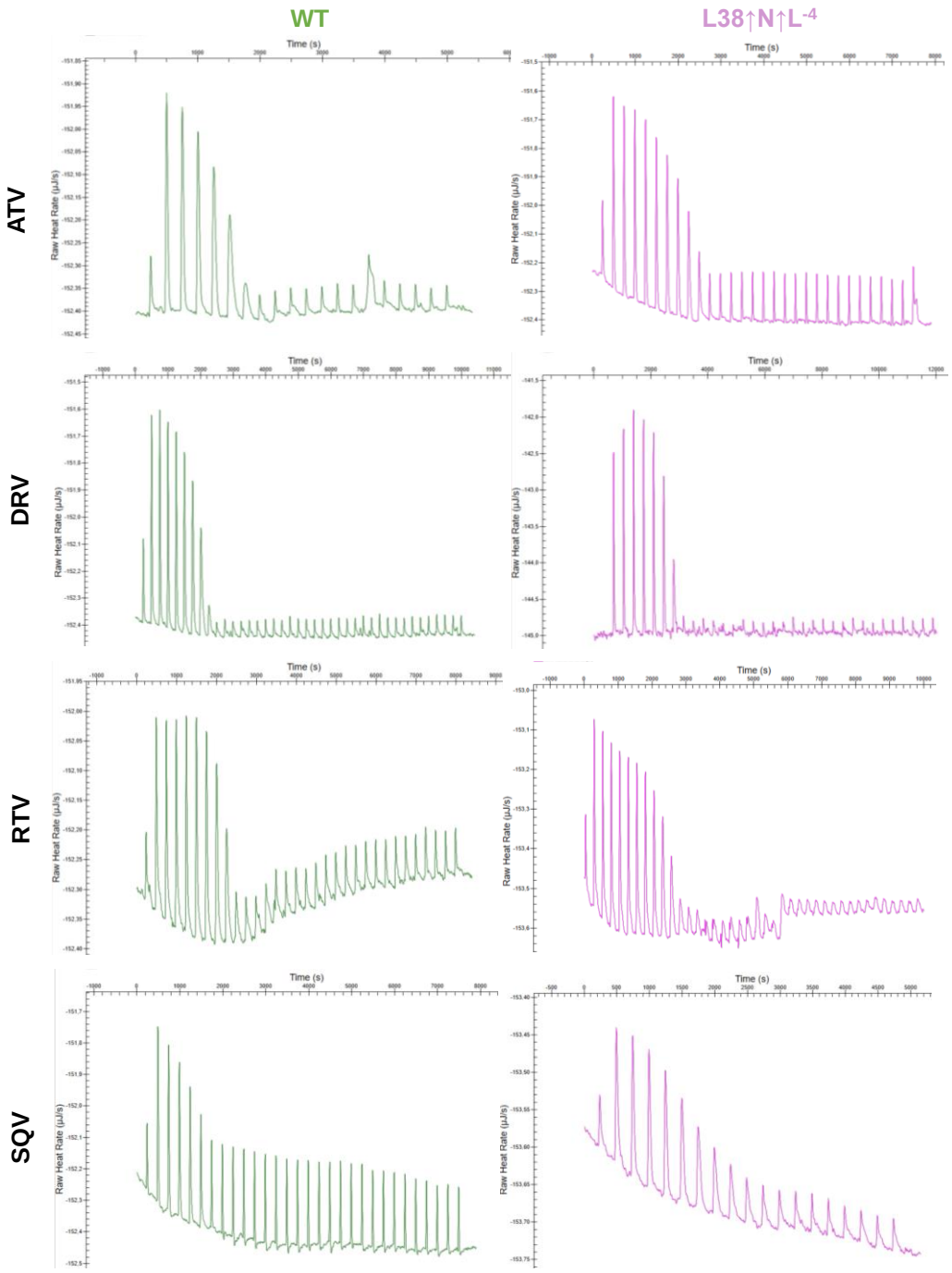


Figure S2 Raw calorimetric data for the displacement titrations of the WT and L38↑N↑L^{-4} proteases with ATV, DRV, RTV and SQV.

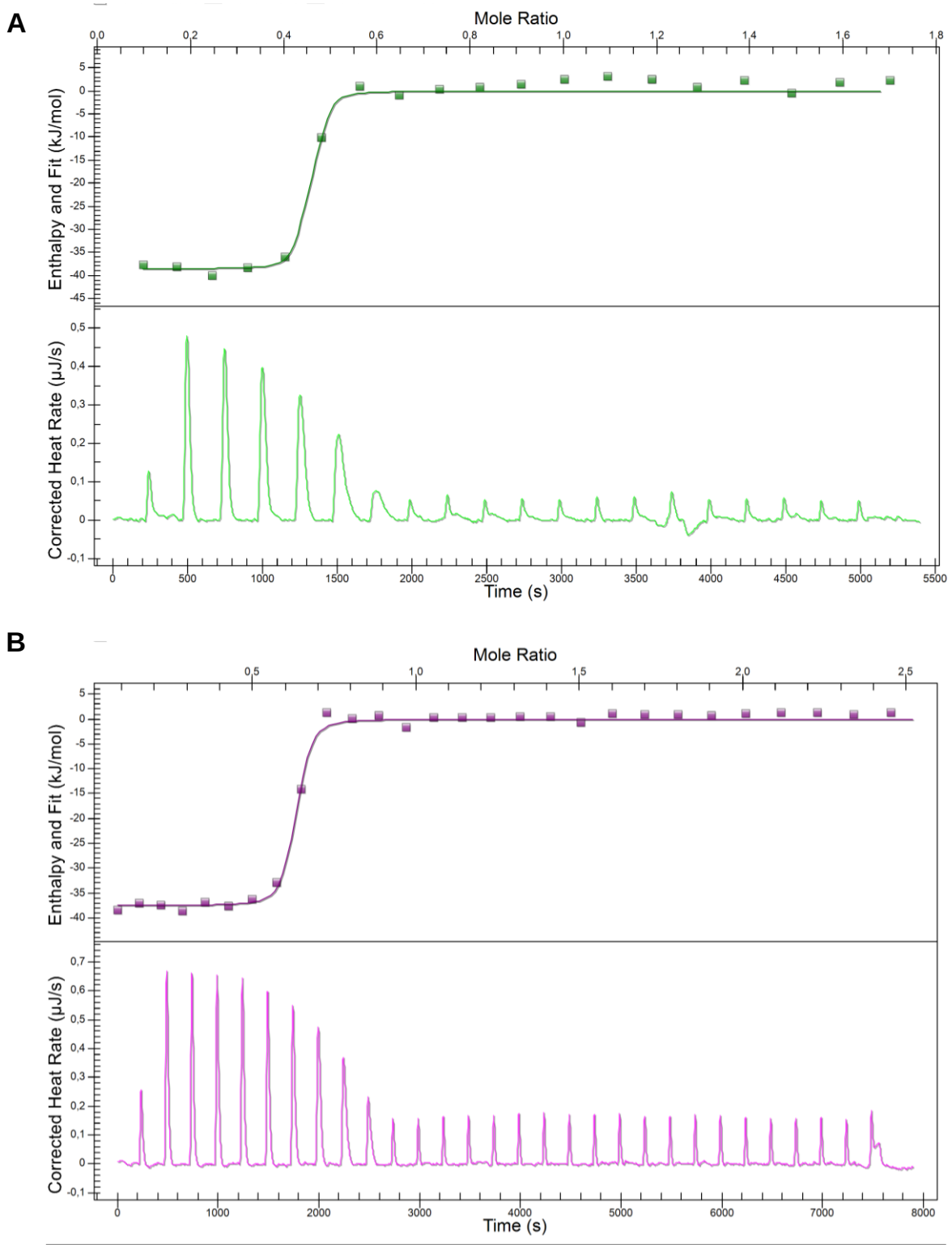


Figure S3 A representative calorimetric profile of the titration (A) WT and (B) L38↑N↑L^{-4} with ATV.

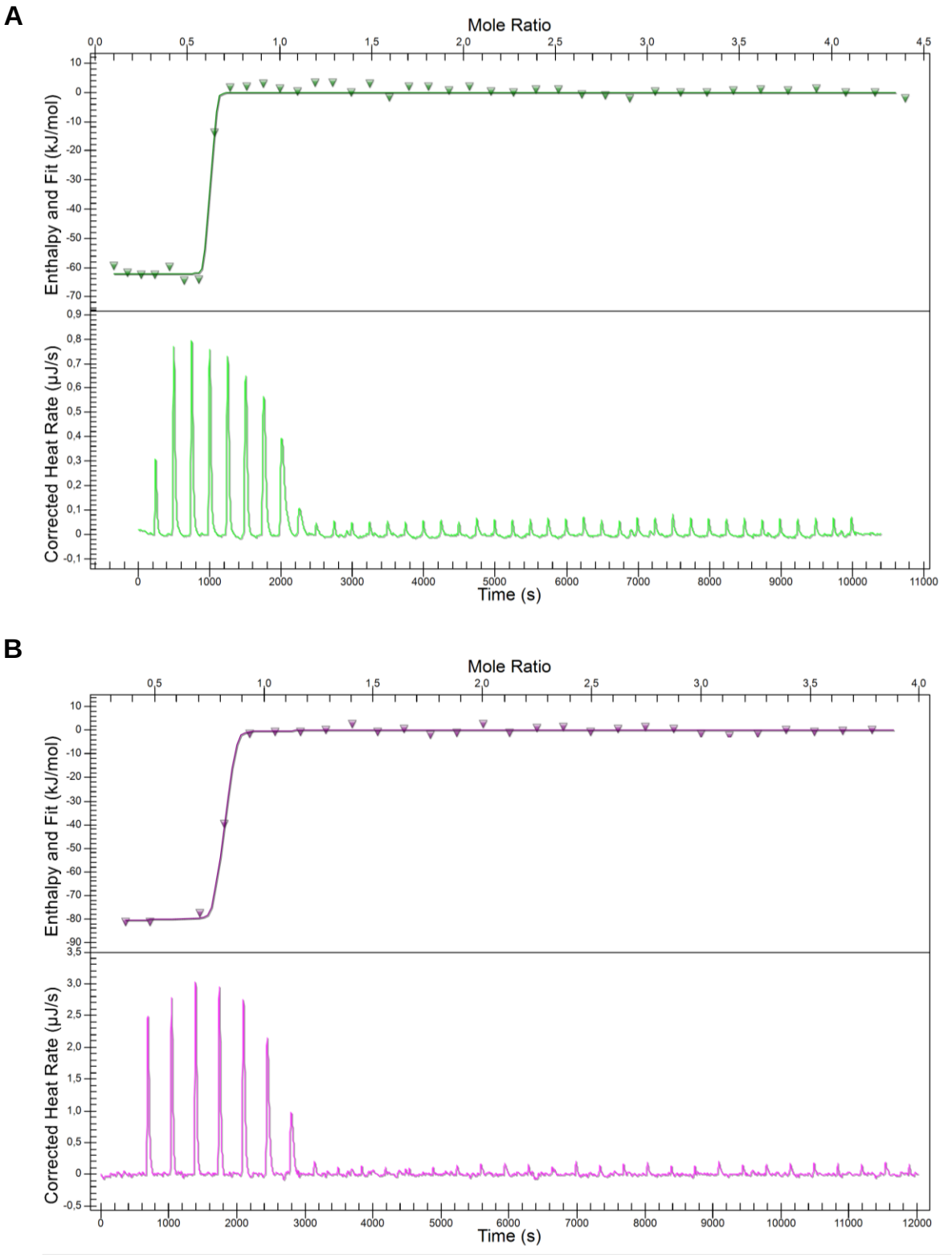


Figure S4 A representative calorimetric profile of the titration (A) WT and (B) L38↑N↑L^{-4} with DRV.

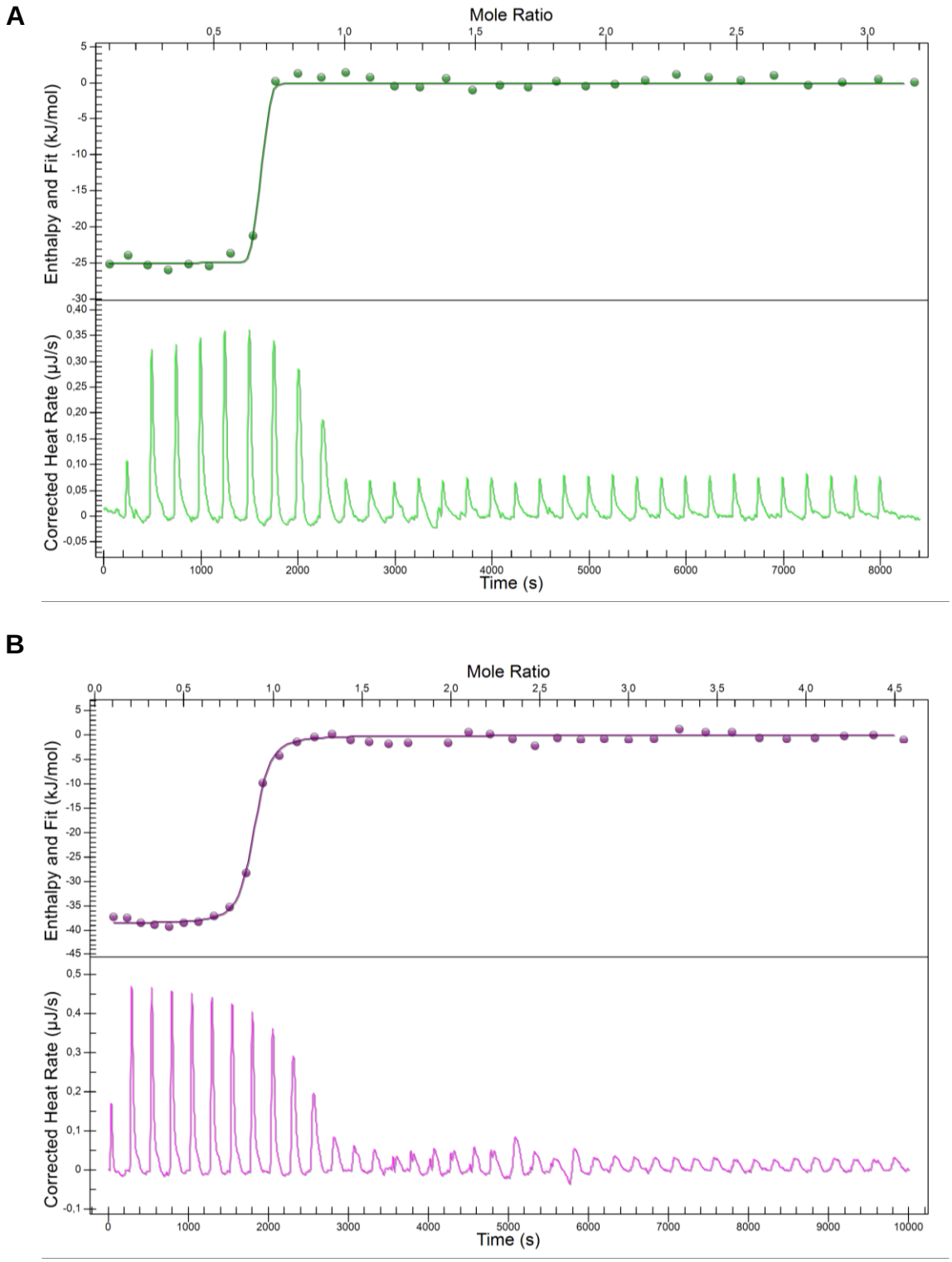


Figure S5 A representative calorimetric profile of the titration (A) WT and (B) L38↑N↑L^{-4} with RTV.

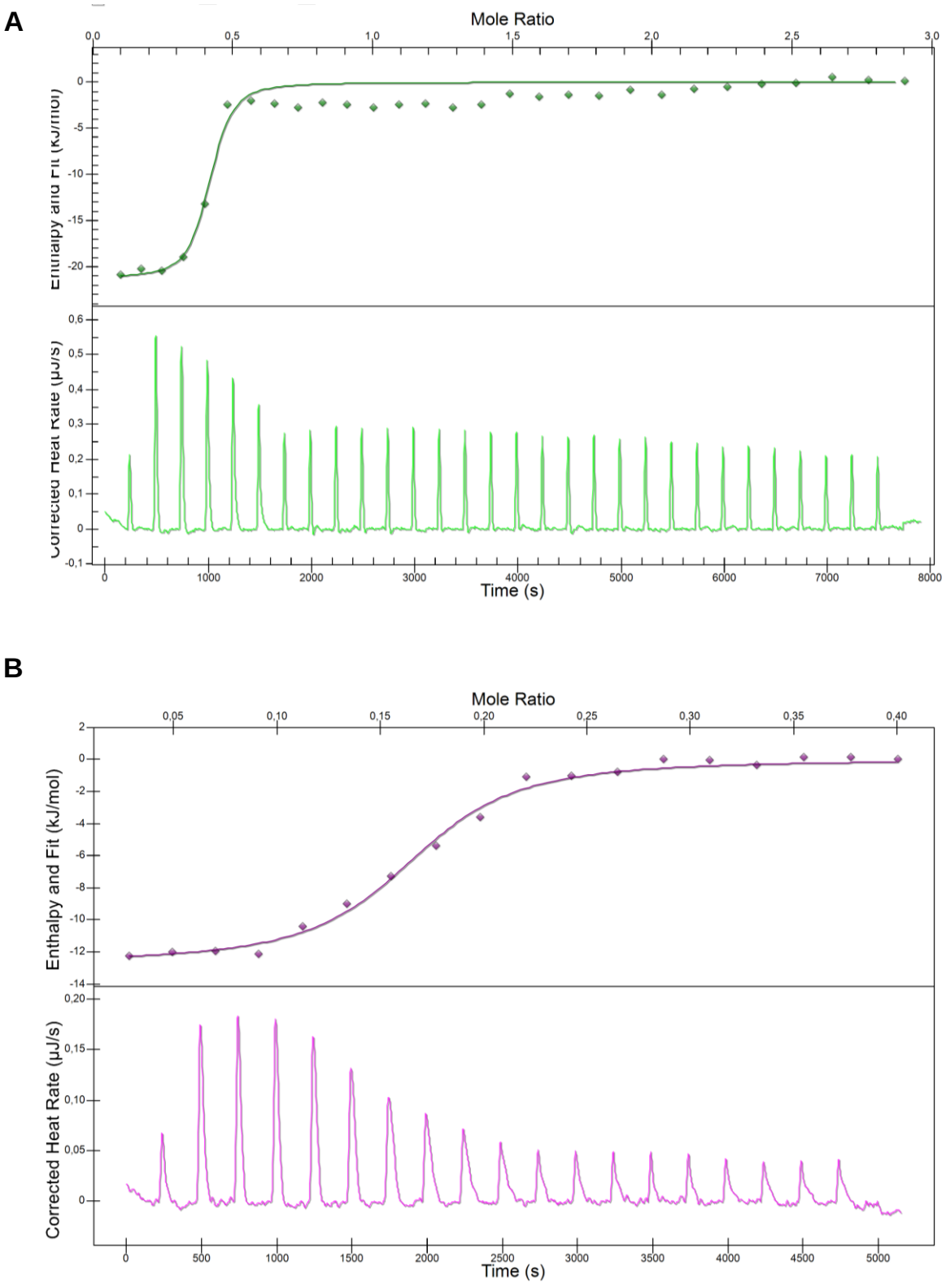


Figure S6 A representative calorimetric profile of the titration (A) WT and (B) L38↑N↑L^{-4} with SQV.

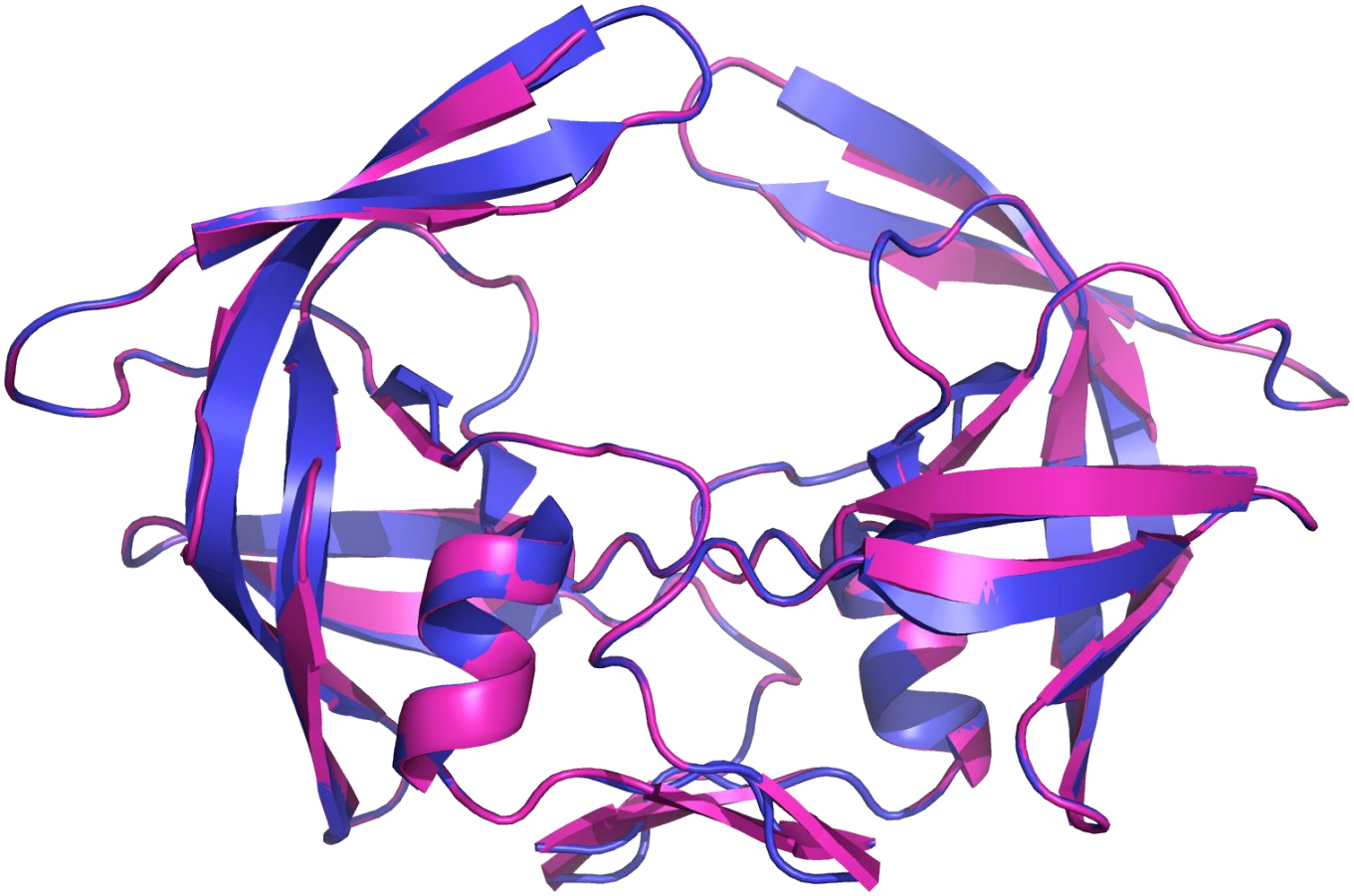


Figure S7 Overlay of I13V, I62V and V77I variant protease crystal structure (blue) and homology model (pink).

The coordinates of the variant protease (PDB: 6I45) was downloaded and viewed in PyMoL. A homology model of this variant protease was created using the built-in PyMOL mutagenesis plug-in and superimposed on the original coordinates. The image was constructed using the PyMOL v0.99rc6 software using PDB ID: 3U71.

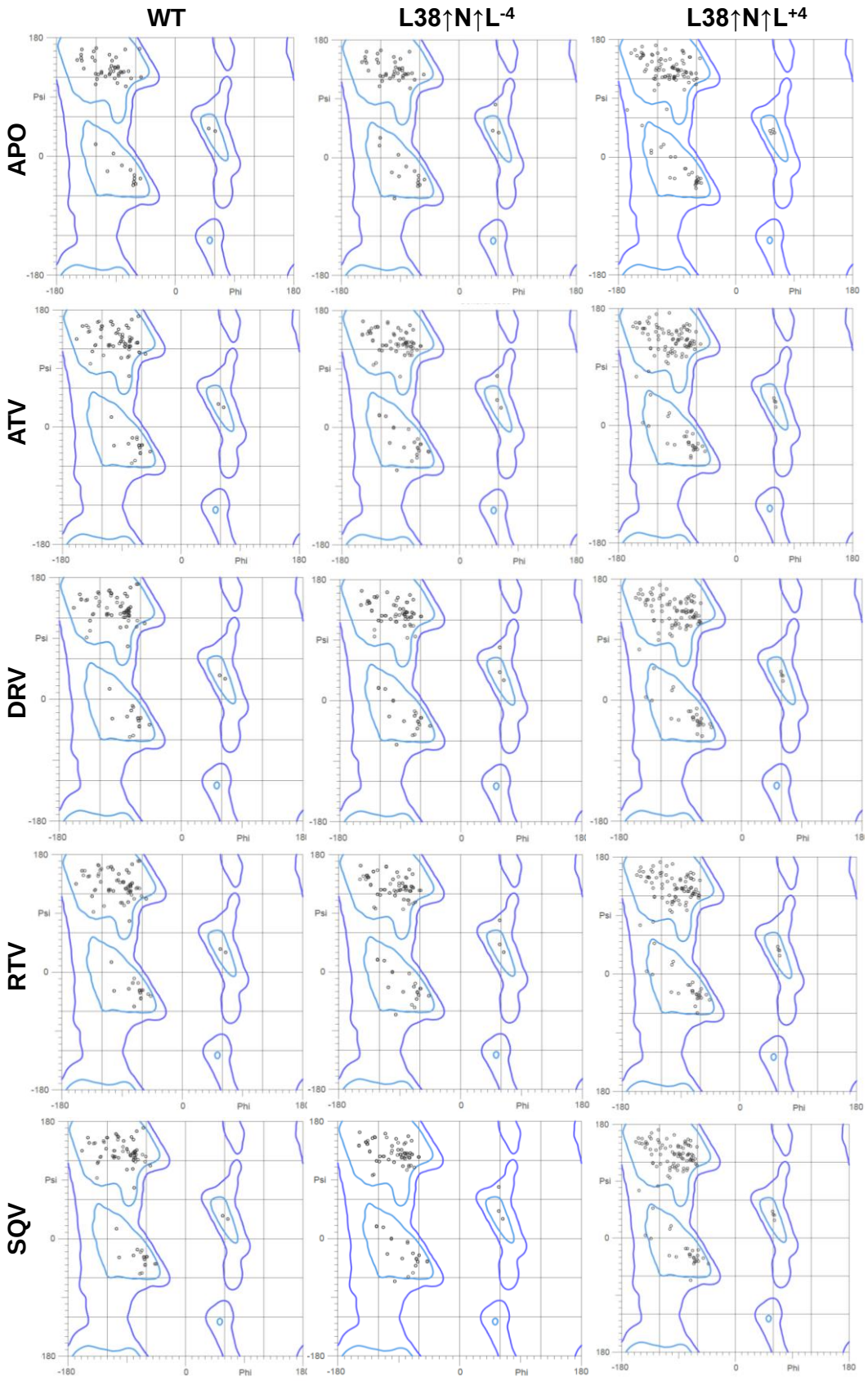


Figure S5. Ramachandran plots of the apo and complexed WT, L38↑N↑L^{-4} and L38↑N↑L^{+4} proteases.

6 REFERENCES

ABRAHAM, M. J., VAN DER SPOEL, D., LINDAHL, E. AND HESS, B. (2017). "GROMACS user manual version 2016.3. ." *The GROMACS development team, www.gromacs.org.*

ADAMSON, C. S. (2012). "Protease-Mediated Maturation of HIV: Inhibitors of Protease and the Maturation Process." *Mol Biol Int* **2012**: 604261.

AGNISWAMY, J., LOUIS, J. M., ROCHE, J., HARRISON, R. W. AND WEBER, I. T. (2016). "Structural Studies of a Rationally Selected Multi-Drug Resistant HIV-1 Protease Reveal Synergistic Effect of Distal Mutations on Flap Dynamics." *PLoS One* **11**(12): e0168616.

ALI, A., BANDARANAYAKE, R. M., CAI, Y., KING, N. M., KOLLI, M., MITTAL, S., MURZYCKI, J. F., NALAM, M. N., NALIVAIKA, E. A., OZEN, A., PRABU-JEYABALAN, M. M., THAYER, K. AND SCHIFFER, C. A. (2010). "Molecular Basis for Drug Resistance in HIV-1 Protease." *Viruses* **2**(11): 2509-2535.

AMIEL, C., CHARPENTIER, C., DESIRE, N., BONNARD, P., LEBRETTE, M. G., WEISS, L., PIALOUX, G. AND SCHNEIDER, V. (2011). "Long-term follow-up of 11 protease inhibitor (PI)-naive and PI-treated HIV-infected patients harbouring virus with insertions in the HIV-1 protease gene." *HIV Med* **12**(3): 138-144.

AMMARANOND, P., CUNNINGHAM, P., OELRICHS, R., SUZUKI, K., HARRIS, C., LEAS, L., GRULICH, A., COOPER, D. A. AND KELLEHER, A. D. (2003). "Rates of transmission of antiretroviral drug resistant strains of HIV-1." *J Clin Virol* **26**(2): 153-161.

ANDUJAR-SANCHEZ, M., LAS HERAS-VAZQUEZ, F. J., CLEMENTE-JIMENEZ, J. M., MARTINEZ-RODRIGUEZ, S., CAMARA-ARTIGAS, A., RODRIGUEZ-VICO, F. AND JARA-PEREZ, V. (2006). "Enzymatic activity assay of D-hydantoinase by isothermal titration calorimetry. Determination of the thermodynamic activation parameters for the hydrolysis of several substrates." *J Biochem Biophys Methods* **67**(1): 57-66.

BALLY, F., MARTINEZ, R., PETERS, S., SUDRE, P. AND TELENTI, A. (2000). "Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors." *AIDS Res Hum Retroviruses* **16**(13): 1209-1213.

BARRE-SINOUSSI, F., CHERMANN, J. C., REY, F., NUGEYRE, M. T., CHAMARET, S., GRUEST, J., DAUGUET, C., AXLER-BLIN, C., VEZINET-BRUN, F., ROUZIOUX, C., ROZENBAUM, W. AND MONTAGNIER, L. (1983). "Isolation of a T-Lymphotopic Retrovirus from a Patient at Risk for Aquired Immune Deficiency Syndrome (AIDS)." *Lancet* **220**: 868 - 871.

BRIGGS, J. A. AND KRAUSSLICH, H. G. (2011). "The molecular architecture of HIV." *J Mol Biol* **410**(4): 491-500.

BRIK, A. AND WONG, C. H. (2003). "HIV-1 protease: mechanism and drug discovery." *Org Biomol Chem* **1**(1): 5-14.

BRUYLANTS, G., WOUTERS, J. AND MICHAUX, C. (2005). "Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design." *Curr Med Chem* **12**(17): 2011-2020.

BUONAGURO, L., TORNESELLO, M. L. AND BUONAGURO, F. M. (2007). "Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications." *J Virol* **81**(19): 10209-10219.

CARMEL, A. AND 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**(2): 265-273.

CHEN, D., OEZGUEN, N., URVIL, P., FERGUSON, C., DANN, S. M. AND SAVIDGE, T. C. (2016). "Regulation of protein-ligand binding affinity by hydrogen bond pairing." *Sci Adv* **2**(3): e1501240.

CHEN, Z., LI, Y., SCHOCK, H. B., HALL, D., CHEN, E. AND KUO, L. C. (1995). "Three-dimensional structure of a mutant HIV-1 protease displaying cross-resistance to all protease inhibitors in clinical trials." *J Biol Chem* **270**(37): 21433-21436.

CLAVEL, F., GUYADER, M., GUETARD, D., SALLE, M., MONTAGNIER, L. AND ALIZON, M. (1986). "Molecular cloning and polymorphism of the human immune deficiency virus type 2." *Nature* **324**(6098): 691-695.

CLAVEL, F. AND HANCE, A. J. (2004). "HIV drug resistance." *N Engl J Med* **350**(10): 1023-1035.

CLAVEL, F. AND MAMMANO, F. (2010). "Role of Gag in HIV Resistance to Protease Inhibitors." *Viruses* **2**(7): 1411-1426.

CLEMENTE, J. C., HEMRAJANI, R., BLUM, L. E., GOODENOW, M. M. AND DUNN, B. M. (2003). "Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N." *Biochemistry* **42**(51): 15029-15035.

CODONER, F. M., PENA, R., BLANCH-LOMBARTE, O., JIMENEZ-MOYANO, E., PINO, M., VOLLBRECHT, T., CLOTET, B., MARTINEZ-PICADO, J., DRAENERT, R. AND PRADO, J. G. (2017). "Gag-protease coevolution analyses define novel structural surfaces in the HIV-1 matrix and capsid involved in resistance to Protease Inhibitors." *Sci Rep* **7**(1): 3717.

COLLINS, J. R., BURT, S. K. AND ERICKSON, J. W. (1995). "Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics." *Nat Struct Biol* **2**(4): 334-338.

COMAN, R. M., ROBBINS, A. H., GOODENOW, M. M., DUNN, B. M. AND MCKENNA, R. (2008). "High-resolution structure of unbound human immunodeficiency virus 1 subtype C protease: implications of flap dynamics and drug resistance." *Acta Crystallogr D Biol Crystallogr* **D64**(Pt 7): 754-763.

CROTEAU, G., DOYON, L., THIBEAULT, D., MCKERCHER, G., PILOTE, L. AND LAMARRE, D. (1997). "Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors." *J Virol* **71**(2): 1089-1096.

CULLEN, B. (1991). "Human Immunodeficiency virus as a prototypic comple retrovirus." *J. Virol.* **65**: 1053-1056.

DARKE, P. L., LEU, C. T., DAVIS, L. J., HEIMBACH, J. C., DIEHL, R. E., HILL, W. S., DIXON, R. A. AND SIGAL, I. S. (1989). "Human immunodeficiency virus protease. Bacterial expression and characterization of the purified aspartic protease." *J Biol Chem* **264**(4): 2307-2312.

DARKE, P. L., NUTT, R. F., BRADY, S. F., GARSKY, V. M., CICCARONE, T. M., LEU, C. T., LUMMA, P. K., FREIDINGER, R. M., VEBER, D. F. AND SIGAL, I. S. (1988). "HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins." *Biochem Biophys Res Commun* **156**(1): 297-303.

DASH, C. AND RAO, M. (2001). "Interactions of a novel inhibitor from an extremophilic Bacillus sp. with HIV-1 protease: implications for the mechanism of inactivation." *J Biol Chem* **276**(4): 2487-2493.

DOYON, L., CROTEAU, G., THIBEAULT, D., POULIN, F., PILOTE, L. AND LAMARRE, D. (1996). "Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors." *J Virol* **70**(6): 3763-3769.

DROR, R. O., DIRKS, R. M., GROSSMAN, J. P., XU, H. AND SHAW, D. E. (2012). "Biomolecular simulation: a computational microscope for molecular biology." *Annu Rev Biophys* **41**: 429-452.

DU, X., LI, Y., XIA, Y. L., AI, S. M., LIANG, J., SANG, P., JI, X. L. AND LIU, S. Q. (2016). "Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods." *Int J Mol Sci* **17**(2).

DUMANS, A. T., SOARES, M. A., PIENIAZEK, D., KALISH, M. L., DE VROEY, V., HERTOGS, K. AND TANURI, A. (2002). "Prevalence of protease and reverse transcriptase drug resistance mutations over time in drug-naive human immunodeficiency virus type 1-positive individuals in Rio de Janeiro, Brazil." *Antimicrob Agents Chemother* **46**(9): 3075-3079.

EMERMAN, M. AND MALIM, M. H. (1998). "HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology." *Science* **280**(5371): 1880-1884.

FOULKES-MURZYCKI, J. E., SCOTT, W. R. AND SCHIFFER, C. A. (2007). "Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease." *Structure* **15**(2): 225-233.

FOULKES, J. E., PRABU-JEYABALAN, M., COOPER, D., HENDERSON, G. J., HARRIS, J., SWANSTROM, R. AND SCHIFFER, C. A. (2006). "Role of invariant Thr80 in human immunodeficiency virus type 1 protease structure, function, and viral infectivity." *J Virol* **80**(14): 6906-6916.

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

FREYER, M. W. AND LEWIS, E. A. (2008). "Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions." *Methods Cell Biol* **84**: 79-113.

FUN, A., WENSING, A. M., VERHEYEN, J. AND NIJHUIS, M. (2012). "Human Immunodeficiency Virus Gag and protease: partners in resistance." *Retrovirology* **9**: 63.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S., WILKINS, M. R., APPEL, R. D. AND BAROCH, A. (2005). "The Proteomics Protocols Handbook. ." *New York (USA): Humana Press.*

GOLDFARB, N. E., OHANESSIAN, M., BISWAS, S., MCGEE, T. D., JR., MAHON, B. P., OSTROV, D. A., GARCIA, J., TANG, Y., MCKENNA, R., ROITBERG, A. AND 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**(2): 422-433.

GRANTZ SASKOVA, K., KOZISEK, M., STRAY, K., DE JONG, D., REZAOVA, P., BRYNDA, J., VAN MAARSEVEEN, N. M., NIJHUIS, M., CIHLAR, T. AND KONVALINKA, J. (2014). "GS-8374, a prototype phosphonate-containing inhibitor of HIV-1 protease, effectively inhibits protease mutants with amino acid insertions." *J Virol* **88**(6): 3586-3590.

GREENFIELD, N. J. (2006). "Using circular dichroism spectra to estimate protein secondary structure." *Nat Protoc* **1**(6): 2876-2890.

GREENWOOD, J. R., CALKINS, D., SULLIVAN, A. P. AND SHELLEY, J. C. (2010). "Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution." *J Comput Aided Mol Des* **24**(6-7): 591-604.

GUEDES, I. A., PEREIRA, F. S. S. AND DARDENNE, L. E. (2018). "Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges." *Front Pharmacol* **9**: 1089.

GUSTCHINA, A. AND WEBER, I. T. (1990). "Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap." *FEBS Lett* **269**(1): 269-272.

HARDER, E., DAMM, W., MAPLE, J., WU, C., REBOUL, M., XIANG, J. Y., WANG, L., LUPYAN, D., DAHLGREN, M. K., KNIGHT, J. L., KAUS, J. W., CERUTTI, D. S., KRILOV, G., JORGENSEN, W. L., ABEL, R. AND FRIESNER, R. A. (2016). "OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins." *J Chem Theory Comput* **12**(1): 281-296.

HEMELAAR, J., GOUWS, E., GHYS, P. D., OSMANOV, S., ISOLATION, W.-U. N. F. H. AND CHARACTERISATION (2011). "Global trends in molecular epidemiology of HIV-1 during 2000-2007." *AIDS* **25**(5): 679-689.

HIRSCH, M. S., BRUN-VEZINET, F., D'AQUILA, R. T., HAMMER, S. M., JOHNSON, V. A., KURITZKES, D. R., LOVEDAY, C., MELLORS, J. W., CLOTET, B., CONWAY, B., DEMETER, L. M., VELLA, S., JACOBSEN, D. M. AND RICHMAN, D. D. (2000). "Antiretroviral drug

resistance testing in adult HIV-1 infection: recommendations of an International AIDS Society-USA Panel." *JAMA* **283**(18): 2417-2426.

HORNAK, V., OKUR, A., RIZZO, R. C. AND SIMMERLING, C. (2006). "HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations." *Proc Natl Acad Sci U S A* **103**(4): 915-920.

HORNAK, V. AND SIMMERLING, C. (2007). "Targeting structural flexibility in HIV-1 protease inhibitor binding." *Drug Discov Today* **12**(3-4): 132-138.

IDO, E., HAN, H. P., KEZDY, F. J. AND TANG, J. (1991). "Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S." *J Biol Chem* **266**(36): 24359-24366.

ISHIMA, R., FREEDBERG, D. I., WANG, Y. X., LOUIS, J. M. AND TORCHIA, D. A. (1999). "Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function." *Structure* **7**(9): 1047-1055.

JOHNSON, C. M. (2013). "Differential scanning calorimetry as a tool for protein folding and stability." *Arch Biochem Biophys* **531**(1-2): 100-109.

KIM, E. Y., WINTERS, M. A., KAGAN, R. M. AND MERIGAN, T. C. (2001). "Functional correlates of insertion mutations in the protease gene of human immunodeficiency virus type 1 isolates from patients." *J Virol* **75**(22): 11227-11233.

KIM, R. AND BAXTER, J. D. (2008). "Protease inhibitor resistance update: where are we now?" *AIDS Patient Care STDS* **22**(4): 267-277.

KING, N. M., PRABU-JEYABALAN, M., NALIVAIKA, E. A. AND SCHIFFER, C. A. (2004). "Combating susceptibility to drug resistance: lessons from HIV-1 protease." *Chem Biol* **11**(10): 1333-1338.

KOHL, N. E., EMINI, E. A., SCHLEIF, W. A., DAVIS, L. J., HEIMBACH, J. C., DIXON, R. A., SCOLNICK, E. M. AND SIGAL, I. S. (1988). "Active human immunodeficiency virus protease is required for viral infectivity." *Proc Natl Acad Sci U S A* **85**(13): 4686-4690.

KOZISEK, M., BRAY, J., REZACOVA, P., SASKOVA, K., BRYNDA, J., POKORNA, J., MAMMANO, F., RULISEK, L. AND KONVALINKA, J. (2007). "Molecular analysis of the HIV-1 resistance development: enzymatic activities, crystal structures, and thermodynamics of nelfinavir-resistant HIV protease mutants." *J Mol Biol* **374**(4): 1005-1016.

KOZISEK, M., LEPSIK, M., GRANTZ SASKOVA, K., BRYNDA, J., KONVALINKA, J. AND REZACOVA, P. (2014). "Thermodynamic and structural analysis of HIV protease resistance to darunavir - analysis of heavily mutated patient-derived HIV-1 proteases." *FEBS J* **281**(7): 1834-1847.

KOZISEK, M., SASKOVA, K. G., REZACOVA, P., BRYNDA, J., VAN MAARSEVEEN, N. M., DE JONG, D., BOUCHER, C. A., KAGAN, R. M., NIJHUIS, M. AND KONVALINKA, J. (2008). "Ninety-nine is not enough: molecular characterization of inhibitor-resistant human immunodeficiency virus type 1 protease mutants with insertions in the flap region." *J Virol* **82**(12): 5869-5878.

KUHN, L., HUNT, G., TECHNAU, K. G., COOVADIA, A., LEDWABA, J., PICKERILL, S., PENAZZATO, M., BERTAGNOLIO, S., MELLINS, C. A., BLACK, V., MORRIS, L. AND ABRAMS, E. J. (2014). "Drug resistance among newly diagnosed HIV-infected children in the era of more efficacious antiretroviral prophylaxis." *AIDS* **28**(11): 1673-1678.

KUKOL, A. (2008). "Molecular modelling of proteins. ." *New York (USA): Humana Press.*

LAEMMLI, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." *Nature* **227**(5259): 680-685.

LASKOWSKI, R. A., RULLMANNN, J. A., MACARTHUR, M. W., KAPTEIN, R. AND THORNTON, J. M. (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." *J Biomol NMR* **8**(4): 477-486.

LEDWABA, J., SAYED, Y., PILLAY, V., MORRIS, L. AND HUNT, G. (2019). "Low Frequency of Protease Inhibitor Resistance Mutations and Insertions in HIV-1 Subtype C Protease Inhibitor-Naive Sequences." *AIDS Res Hum Retroviruses* **35**(7): 673-678.

LOCKHAT, H. A., SILVA, J. R., ALVES, C. N., GOVENDER, T., LAMEIRA, J., MAGUIRE, G. E., SAYED, Y. AND KRUGER, H. G. (2016). "Binding Free Energy Calculations of Nine FDA-approved Protease Inhibitors Against HIV-1 Subtype C I36T upward arrowT Containing 100 Amino Acids Per Monomer." *Chem Biol Drug Des* **87**(4): 487-498.

LOUIS, J. M., MCDONALD, R. A., NASHED, N. T., WONDRAK, E. M., JERINA, D. M., OROSZLAN, S. AND MORA, P. T. (1991). "Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in Escherichia coli." *Eur J Biochem* **199**(2): 361-369.

LOUIS, J. M. AND ROCHE, J. (2016). "Evolution under Drug Pressure Remodels the Folding Free-Energy Landscape of Mature HIV-1 Protease." *J Mol Biol* **428**(13): 2780-2792.

LOUIS, J. M., ZHANG, Y., SAYER, J. M., WANG, Y. F., HARRISON, R. W. AND WEBER, I. T. (2011). "The L76V drug resistance mutation decreases the dimer stability and rate of autoprocessing of HIV-1 protease by reducing internal hydrophobic contacts." *Biochemistry* **50**(21): 4786-4795.

LV, Z., CHU, Y. AND WANG, Y. (2015). "HIV protease inhibitors: a review of molecular selectivity and toxicity." *HIV AIDS (Auckl)* **7**: 95-104.

MAHALINGAM, B., BOROSS, P., WANG, Y. F., LOUIS, J. M., FISCHER, C. C., TOZSER, J., HARRISON, R. W. AND WEBER, I. T. (2002). "Combining mutations in HIV-1 protease to understand mechanisms of resistance." *Proteins* **48**(1): 107-116.

MAHALINGAM, B., LOUIS, J. M., HUNG, J., HARRISON, R. W. AND WEBER, I. T. (2001). "Structural implications of drug-resistant mutants of HIV-1 protease: high-resolution crystal structures of the mutant protease/substrate analogue complexes." *Proteins* **43**(4): 455-464.

MAHALINGAM, B., LOUIS, J. M., REED, C. C., ADOMAT, J. M., KROUSE, J., WANG, Y. F., HARRISON, R. W. AND WEBER, I. T. (1999). "Structural and kinetic analysis of drug resistant mutants of HIV-1 protease." *Eur J Biochem* **263**(1): 238-245.

MAMMANO, F., TROUPLIN, V., ZENNOU, V. AND CLAVEL, F. (2000). "Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug." *J Virol* **74**(18): 8524-8531.

MARK, P. AND NILSSON, L. (2002). "Structure and dynamics of liquid water with different long-range interaction truncation and temperature control methods in molecular dynamics simulations." *J Comput Chem* **23**(13): 1211-1219.

MASQUELIER, B., RACE, E., TAMALET, C., DESCAMPS, D., IZOPET, J., BUFFET-JANVRESSE, C., RUFFAULT, A., MOHAMMED, A. S., COTTALORDA, J., SCHMUCK, A., CALVEZ, V., DAM, E., FLEURY, H., BRUN-VEZINET, F. AND SIDA, A. A. R. S. G. F. A. N. D. R. S. L. (2001). "Genotypic and phenotypic resistance patterns of human immunodeficiency virus type 1 variants with insertions or deletions in the reverse transcriptase (RT): multicenter study of patients treated with RT inhibitors." *Antimicrob Agents Chemother* **45**(6): 1836-1842.

MCCUTCHAN, F. E. (2006). "Global epidemiology of HIV." *J Med Virol* **78 Suppl 1**: S7-S12.

MEISELBACH, H., HORN, A. H., HARRER, T. AND STICHT, H. (2007). "Insights into amprenavir resistance in E35D HIV-1 protease mutation from molecular dynamics and binding free-energy calculations." *J Mol Model* **13**(2): 297-304.

MILDNER, A. M., ROTHROCK, D. J., LEONE, J. W., BANNOW, C. A., LULL, J. M., REARDON, I. M., SARCICH, J. L., HOWE, W. J., TOMICH, C. S., SMITH, C. W. AND ET AL. (1994). "The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties." *Biochemistry* **33**(32): 9405-9413.

MITTAL, S., BANDARANAYAKE, R. M., KING, N. M., PRABU-JEYABALAN, M., NALAM, M. N., NALIVAIKA, E. A., YILMAZ, N. K. AND SCHIFFER, C. A. (2013). "Structural and thermodynamic basis of amprenavir/darunavir and atazanavir resistance in HIV-1 protease with mutations at residue 50." *J Virol* **87**(8): 4176-4184.

MITTAL, S., CAI, Y., NALAM, M. N., BOLON, D. N. AND SCHIFFER, C. A. (2012). "Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease." *J Am Chem Soc* **134**(9): 4163-4168.

MONTICELLI, L. AND TIELEMAN, D. P. (2013). "Force fields for classical molecular dynamics." *Methods Mol Biol* **924**: 197-213.

MOSEBI, S., MORRIS, L., DIRR, H. W. AND SAYED, Y. (2008). "Active-site mutations in the South african human immunodeficiency virus type 1 subtype C protease have a significant impact on clinical inhibitor binding: kinetic and thermodynamic study." *J Virol* **82**(22): 11476-11479.

MUZAMMIL, S., ARMSTRONG, A. A., KANG, L. W., JAKALIAN, A., BONNEAU, P. R., SCHMELMER, V., AMZEL, L. M. AND FREIRE, E. (2007). "Unique thermodynamic response of tipranavir to human immunodeficiency virus type 1 protease drug resistance mutations." *J Virol* **81**(10): 5144-5154.

MUZAMMIL, S., ROSS, P. AND FREIRE, E. (2003). "A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance." *Biochemistry* **42**(3): 631-638.

NAICKER, P., ACHILONU, I., FANUCCHI, S., FERNANDES, M., IBRAHIM, M. A., DIRR, H. W., SOLIMAN, M. E. AND SAYED, Y. (2013). "Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance." *J Biomol Struct Dyn* **31**(12): 1370-1380.

NAICKER, P. AND SAYED, Y. (2014). "Non-B HIV-1 subtypes in sub-Saharan Africa: impact of subtype on protease inhibitor efficacy." *Biol Chem* **395**(10): 1151-1161.

NAICKER, P., SEELE, P., DIRR, H. W. AND SAYED, Y. (2013). "F99 is critical for dimerization and activation of South African HIV-1 subtype C protease." *Protein J* **32**(7): 560-567.

NAICKER, P., STOYCHEV, S., DIRR, H. W. AND SAYED, Y. (2014). "Amide hydrogen exchange in HIV-1 subtype B and C proteases--insights into reduced drug susceptibility and dimer stability." *FEBS J* **281**(24): 5395-5410.

NALAM, M. N., ALI, A., ALTMAN, M. D., REDDY, G. S., CHELLAPPAN, S., KAIRYS, V., OZEN, A., CAO, H., GILSON, M. K., TIDOR, B., RANA, T. M. AND SCHIFFER, C. A. (2010). "Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance." *J Virol* **84**(10): 5368-5378.

NIJHUIS, M., SCHUURMAN, R., DE JONG, D., ERICKSON, J., GUSTCHINA, E., ALBERT, J., SCHIPPER, P., GULNIK, S. AND BOUCHER, C. A. (1999). "Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy." *AIDS* **13**(17): 2349-2359.

OHTAKA, H., VELAZQUEZ-CAMPOY, A., XIE, D. AND 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* **11**(8): 1908-1916.

OZEN, A., HALILOGLU, T. AND SCHIFFER, C. A. (2011). "Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope." *J Mol Biol* **410**(4): 726-744.

OZER, N., OZEN, A., SCHIFFER, C. A. AND HALILOGLU, T. (2015). "Drug-resistant HIV-1 protease regains functional dynamics through cleavage site coevolution." *Evol Appl* **8**(2): 185-198.

OZER, N., SCHIFFER, C. A. AND HALILOGLU, T. (2010). "Rationale for more diverse inhibitors in competition with substrates in HIV-1 protease." *Biophys J* **99**(5): 1650-1659.

PEREIRA-VAZ, J., DUQUE, V., TRINDADE, L., SARAIVA-DA-CUNHA, J. AND MELICO-SILVESTRE, A. (2009). "Detection of the protease codon 35 amino acid insertion in sequences from treatment-naive HIV-1 subtype C infected individuals in the Central Region of Portugal." *J Clin Virol* **46**(2): 169-172.

PERRYMAN, A. L., LIN, J. H. AND 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* **13**(4): 1108-1123.

PERRYMAN, A. L., LIN, J. H. AND MCCAMMON, J. A. (2006). "Restrained molecular dynamics simulations of HIV-1 protease: the first step in validating a new target for drug design." *Biopolymers* **82**(3): 272-284.

PETTIT, S. C., MOODY, M. D., WEHBIE, R. S., KAPLAN, A. H., NANTERMET, P. V., KLEIN, C. A. AND SWANSTROM, R. (1994). "The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions." *J Virol* **68**(12): 8017-8027.

PIANA, S., CARLONI, P. AND ROTHLISBERGER, U. (2002). "Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations." *Protein Sci* **11**(10): 2393-2402.

POLGAR, L., SZELTNER, Z. AND BOROS, I. (1994). "Substrate-dependent mechanisms in the catalysis of human immunodeficiency virus protease." *Biochemistry* **33**(31): 9351-9357.

PRABU-JEYABALAN, M., NALIVAIKA, E. AND SCHIFFER, C. A. (2002). "Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes." *Structure* **10**(3): 369-381.

PRABU-JEYABALAN, M., NALIVAIKA, E. A., KING, N. M. AND SCHIFFER, C. A. (2004). "Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease." *J Virol* **78**(22): 12446-12454.

RAGLAND, D. A., NALIVAIKA, E. A., NALAM, M. N., PRACHANRONARONG, K. L., CAO, H., BANDARANAYAKE, R. M., CAI, Y., KURT-YILMAZ, N. AND 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**(34): 11956-11963.

RHEE, S. Y., TAYLOR, J., FESSEL, W. J., KAUFMAN, D., TOWNER, W., TROIA, P., RUANE, P., HELLINGER, J., SHIRVANI, V., ZOLOPA, A. AND SHAFER, R. W. (2010). "HIV-1 protease mutations and protease inhibitor cross-resistance." *Antimicrob Agents Chemother* **54**(10): 4253-4261.

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. AND KORBER, B. (2000). "HIV-1 nomenclature proposal." *Science* **288**(5463): 55-56.

RUPP, B. (2009). "Biomolecular crystallography-principles, practice and application to structural biology. ." *1st ed. New York (USA): Garland Science.*

SAWYER, J. R., SCHLOM, J. AND KASHMIRI, S. V. (1994). "The effects of induction conditions on production of a soluble anti-tumor sFv in Escherichia coli." *Protein Eng* **7**(11): 1401-1406.

SAYER, J. M., LIU, F., ISHIMA, R., WEBER, I. T. AND 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**(19): 13459-13470.

SCHAGGER, H. (2006). "Tricine-SDS-PAGE." *Nat Protoc* **1**(1): 16-22.

SCHMID, F. X. (2001). "Biological macromolecules: UV-Visible spectrophotometry. ." *New York (USA): Macmillan Publishers Ltd.*

SCOTT, W. R. AND SCHIFFER, C. A. (2000). "Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance." *Structure* **8**(12): 1259-1265.

SEELMEIER, S., SCHMIDT, H., TURK, V. AND VON DER HELM, K. (1988). "Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A." *Proc Natl Acad Sci U S A* **85**(18): 6612-6616.

SHEEHAN, D. (2009). "Physical biochemistry: Principles and applications. ." *England (UK): John Wiley and Sons Ltd.*

SHEHU-XHILAGA, M., KRAEUSSLICH, H. G., PETTIT, S., SWANSTROM, R., LEE, J. Y., MARSHALL, J. A., CROWE, S. M. AND MAK, J. (2001). "Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation." *J Virol* **75**(19): 9156-9164.

SHELLEY, J. C., CHOLLETI, A., FRYE, L. L., GREENWOOD, J. R., TIMLIN, M. R. AND UCHIMAYA, M. (2007). "Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules." *J Comput Aided Mol Des* **21**(12): 681-691.

SHERRY, D., WORTH, R. AND 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**(1): e106.

SMITH, D. B. AND JOHNSON, K. S. (1988). "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase." *Gene* **67**(1): 31-40.

SORENSEN, H. P. AND MORTENSEN, K. K. (2005). "Advanced genetic strategies for recombinant protein expression in Escherichia coli." *J Biotechnol* **115**(2): 113-128.

STURMER, M., STASZEWSKI, S., DOERR, H. W. AND HERTOGS, K. (2003). "A 6-base pair insertion in the protease gene of HIV type 1 detected in a protease inhibitor-naive patient is not associated with indinavir treatment failure." *AIDS Res Hum Retroviruses* **19**(11): 967-968.

SUGUNA, K., PADLAN, E. A., SMITH, C. W., CARLSON, W. D. AND 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**(20): 7009-7013.

SZELTNER, Z. AND POLGAR, L. (1996). "Rate-determining steps in HIV-1 protease catalysis. The hydrolysis of the most specific substrate." *J Biol Chem* **271**(50): 32180-32184.

TANG, J. W. AND PILLAY, D. (2004). "Transmission of HIV-1 drug resistance." *J Clin Virol* **30**(1): 1-10.

TAYLOR, B. S., SOBIESZCZYK, M. E., MCCUTCHAN, F. E. AND HAMMER, S. M. (2008). "The challenge of HIV-1 subtype diversity." *N Engl J Med* **358**(15): 1590-1602.

TELLINGHUISEN, J. (2004). "Volume errors in isothermal titration calorimetry." *Anal Biochem* **333**(2): 405-406.

THOMPSON, M. A., ABERG, J. A., CAHN, P., MONTANER, J. S., RIZZARDINI, G., TELENTI, A., GATELL, J. M., GUNTHARD, H. F., HAMMER, S. M., HIRSCH, M. S., JACOBSEN, D. M., REISS, P., RICHMAN, D. D., VOLBERDING, P. A., YENI, P., SCHOOLEY, R. T. AND INTERNATIONAL, A. S.-U. S. A. (2010). "Antiretroviral treatment of adult HIV infection: 2010 recommendations of the International AIDS Society-USA panel." *JAMA* **304**(3): 321-333.

TODD, M. J., LUQUE, I., VELAZQUEZ-CAMPOY, A. AND FREIRE, E. (2000). "Thermodynamic basis of resistance to HIV-1 protease inhibition: calorimetric analysis of the V82F/I84V active site resistant mutant." *Biochemistry* **39**(39): 11876-11883.

TODD, M. J., SEMO, N. AND FREIRE, E. (1998). "The structural stability of the HIV-1 protease." *J Mol Biol* **283**(2): 475-488.

TOTH, G. AND BORICS, A. (2006). "Closing of the flaps of HIV-1 protease induced by substrate binding: a model of a flap closing mechanism in retroviral aspartic proteases." *Biochemistry* **45**(21): 6606-6614.

TOZSER, J., YIN, F. H., CHENG, Y. S., BAGOSSI, P., WEBER, I. T., HARRISON, R. W. AND OROSZLAN, S. (1997). "Activity of tethered human immunodeficiency virus 1 protease containing mutations in the flap region of one subunit." *Eur J Biochem* **244**(1): 235-241.

UNAIDS (2021). "UNAIDS data 2021."

VELAZQUEZ-CAMPOY, A. AND FREIRE, E. (2006). "Isothermal titration calorimetry to determine association constants for high-affinity ligands." *Nat Protoc* **1**(1): 186-191.

VELAZQUEZ-CAMPOY, A., KISO, Y. AND FREIRE, E. (2001). "The binding energetics of first- and second-generation HIV-1 protease inhibitors: implications for drug design." *Arch Biochem Biophys* **390**(2): 169-175.

VELAZQUEZ-CAMPOY, A., TODD, M. J., VEGA, S. AND FREIRE, E. (2001). "Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes." *Proc Natl Acad Sci U S A* **98**(11): 6062-6067.

VELAZQUEZ-CAMPOY, A., VEGA, S., FLEMING, E., BACHA, U., SAYED, Y., DIRR, H. W. AND FREIRE, E. (2003). "Protease inhibition in African subtypes of HIV-1." *AIDS Rev* **5**(3): 165-171.

WEBER, I. T. AND AGNISWAMY, J. (2009). "HIV-1 Protease: Structural Perspectives on Drug Resistance." *Viruses* **1**(3): 1110-1136.

WENSING, A. M., VAN MAARSEVEEN, N. M. AND NIJHUIS, M. (2010). "Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance." *Antiviral Res* **85**(1): 59-74.

WHO (2003). "The use of antiretroviral drugs for teating and preventing HIV infection." *Geneva, Switzerland: WHO Press*.

WHO (2016). "Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach." *WHO Press* **Geneva, Switzerland**.

WILKINS, M. R., GASTEIGER, E., BAIROCH, A., SANCHEZ, J. C., WILLIAMS, K. L., APPEL, R. D. AND HOCHSTRASSER, D. F. (1999). "Protein identification and analysis tools in the ExPASy server." *Methods Mol Biol* **112**: 531-552.

WILLIAMS, A., BASSON, A., ACHILONU, I., DIRR, H. W., MORRIS, L. AND SAYED, Y. (2019). "Double trouble? Gag in conjunction with double insert in HIV protease contributes to reduced DRV susceptibility." *Biochem J* **476**(2): 375-384.

WILSON, K. AND WALKER, J. (2010). "Principles and techniques of biochemistry and molecular biology. ." *7th ed. England (UK): Cambridge University Press.*

WINDER, A. F. AND GENT, W. L. (1971). "Correction of light-scattering errors in spectrophotometric protein determinations." *Biopolymers* **10**(7): 1243-1251.

WINTERS, M. A., KAGAN, R. M., HESELTINE, P. N. AND MERIGAN, T. C. (2005). "New two-amino acid insertion near codon 70 of the HIV type 1 protease gene." *AIDS Res Hum Retroviruses* **21**(4): 311-313.

WINTERS, M. A. AND MERIGAN, T. C. (2005). "Insertions in the human immunodeficiency virus type 1 protease and reverse transcriptase genes: clinical impact and molecular mechanisms." *Antimicrob Agents Chemother* **49**(7): 2575-2582.

WLODAWER, A. AND ERICKSON, J. W. (1993). "Structure-based inhibitors of HIV-1 protease." *Annu Rev Biochem* **62**: 543-585.

WLODAWER, A., MILLER, M., JASKOLSKI, M., SATHYANARAYANA, B. K., BALDWIN, E., WEBER, I. T., SELK, L. M., CLAWSON, L., SCHNEIDER, J. AND KENT, S. B. (1989).

"Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease." *Science* **245**(4918): 616-621.

WOODY, R. W. (1995). "Circular dichroism." *Methods Enzymol* **246**: 34-71.

WOODY, R. W. AND KOSLOWSKI, A. (2002). "Recent developments in the electronic spectroscopy of amides and alpha-helical polypeptides." *Biophys Chem* **101-102**: 535-551.

WU, T. D., SCHIFFER, C. A., GONZALES, M. J., TAYLOR, J., KANTOR, R., CHOU, S., ISRAELSKI, D., ZOLOPA, A. R., FESSEL, W. J. AND SHAFER, R. W. (2003). "Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments." *J Virol* **77**(8): 4836-4847.

YANCHUNAS, J., JR., LANGLEY, D. R., TAO, L., ROSE, R. E., FRIBORG, J., COLONNO, R. J. AND DOYLE, M. L. (2005). "Molecular basis for increased susceptibility of isolates with atazanavir resistance-conferring substitution I50L to other protease inhibitors." *Antimicrob Agents Chemother* **49**(9): 3825-3832.

YANG, L. W. AND BAHAR, I. (2005). "Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes." *Structure* **13**(6): 893-904.

ZHENG, W. AND BROOKS, B. (2005). "Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model." *J Mol Biol* **346**(3): 745-759.

ZOETE, V., MICHIELIN, O. AND KARPLUS, M. (2002). "Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility." *J Mol Biol* **315**(1): 21-52.

ZONDAGH, J., BALAKRISHNAN, V., ACHILONU, I., DIRR, H. W. AND 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.